Continence Outcomes Measurement Suite

together with

Review of Patient Satisfaction Measures

2006

www.toiletmap.gov.au

Continence Outcomes Measurement Suite by La Trobe University

Review of Patient Satisfaction Measures by University of Melbourne Private

LA TROBE UNIVERSITY

MELBOURNE
UNIVERSITY PRIVATE

University of Wollongong

Continence Outcomes Measurement Suite

2006

By
Professor Shane Thomas
Professor Rhonda Nay
A/Professor Kate Moore
A/Professor David Fonda
A/Professor Graeme Hawthorne

With
Nicholas Marosszeky and Jan Sansoni

Review of Patient Satisfaction Measures – Page 225

Acknowledgements for the main report

The project team wishes to thank Jan Sansoni for her kind and expert input to the project process and report. Her support was invaluable.

We also thank Cath Stoove, Elizabeth Kearney and Amanda Weeks for their assistance in creating the project report.

Table of Contents

List of Tables

List of Figures

How to use this document

This document is intended to be read and used by a range of different audiences including primary care and non-specialist practitioners, specialist practitioners, funders and researchers in the area of incontinence. It is obvious that the need for technical detail concerning the outcome measurement tools for these different groups varies considerably. It is a challenge to meet al.l of these diverse needs.

To this end, an extended executive summary has been developed that outlines the main activities undertaken in the project and summarises its major findings. The recommendations for adoption of the various measures have been summarised in this section of the report. It is recommended that this part of the report is used by primary care practitioners and specialist practitioners who want to use the recommendations without detailed analysis of the evidence base underpinning them. There is no pressing need for such practitioners to read the entire report. It is suggested that occasional forays into the main body of the report to read about the details of the recommended tools would be the most appropriate use of the report for this group.

The main body of the report is intended for a technical audience including practitioners with advanced research skills and also measurement specialists who wish to read the detail of the recommendations and the evidence base underpinning them. The report contains considerable detail of the best available tools that we have chosen. The field of outcomes measurement is a dynamic one in which new measures are constantly being published. This is a field, however, in which there is a major lack of follow through when new measures are developed. Some survive, but many are used in only a handful of studies. The project team has attempted to weed out these tools and also those tools that are obviously not suited to the measurement tasks at hand.

Appended to this report are detailed instrument reviews using the format provided by the Australian Health Outcomes Collaboration. These reviews contain the bulk of the evidence used to develop the instrument ratings and recommendations.

Abbreviations and Acronyms

15D	Fifteen-Dimensional measure of health-related quality of life
AMS	American Medical Systems Faecal Incontinence Scale
AQoL	Assessment of Quality of Life
AUASI	American Urological Association Symptom Index
Barthel	Barthel ADL Index
BFLUTS	Bristol Female Lower Urinary Tract Symptom assessment
Dan-PSS-1	No English Acronym
EQ5D	European Quality of Life Measure – 5D (formerly the EUROQOL)
FIM	Functional Independence Measure
FIQLS	Faecal Incontinence Quality of Life Scale
GIQLI	Gastro Intestinal Quality of Life Index
HRQoL	Health Related Quality of Life
HUI3	Health Utilities Index – Version 3
ICIQ	International Consultation on Incontinence Questionnaire
ICIQ-SF	International Consultation on Incontinence Questionnaire Short Form
ICS	International Continence Society
ICS-male short	International Continence Society Male short form assessment
ICS-male	International Continence Society Male assessment
ICUD	International Consultation on Urological Diseases
IFFGD	International Foundation for Functional Gastrointestinal Disorders
I-QOL	Urinary Incontinence-Specific Quality of Life Instrument
ISI	Incontinence Severity Index
Katz	Katz ADL Scale
KHQ	King's Health Questionnaire
LUTS	Lower Urinary Tract Symptoms
MAU	Multi-Attribute Utility
ME	Magnitude estimation
NHP	Nottingham Health Profile
Pescatori	Pescatori Faecal Incontinence Symptom Severity Measure
PTO	Person Trade-Off
QALY	Quality adjusted life year
QoL	Quality of Life
QWB	Quality of Well Being
Rosser	Rosser Quality of Life Index
SF 12	SF-12® Health Survey (Version 1.0)
SF 36 (II)	SF-36® Health Survey (Version 2.0)
SF 36 (I)	SF-36® Health Survey (Version 1.0)
SF6D	No acronym available
SG	Standard Gamble
SIP	Sickness Impact Profile
St Marks	St Marks Faecal Incontinence Grading System
TTO	Time Trade-Off
UDI	Urogenital Distress Inventory
UDI Long	Urogenital Distress Inventory Long Form
UDI Short	Urogenital Distress Inventory Short Form
VAS	Visual Analog Rating Scale
Wexner	Wexner/Cleveland Clinic Faecal Incontinence Score
WHO	World Health Organisation
WHO-QOL 100	World Health Organisation Quality of Life Assessment 100 item version
WHO-QOL Bref	World Health Organisation Quality of Life Assessment Bref (26 item) version

Executive Summary of the Continence Outcomes Measurement Suite Project

Background

This document is the final report of the Continence Outcomes Measurement Suite project. The project was conducted by A/Professor David Fonda, A/Professor Kate Moore, Professor Rhonda Nay, Professor Shane Thomas and a team of technical experts (Jan Sansoni, A/Professor Graeme Hawthorne and Nicholas Marosszeky) and support staff (Amanda Weeks, Cath Stoove and Elizabeth Kearney). The project was commissioned by the Australian Government Department of Health and Ageing, National Continence Management Strategy Research Program with the goal of recommending a suite of continence outcome measures to be used by clinicians and researchers in Australia.

This project is also linked to a number of other concurrent research projects on continence conditions undertaken for the Department of Health and Ageing. These include the *Development of a Framework for Economic and Cost Evaluation of Continence Conditions* (Moore et al. 2004) which, when used in conjunction with this report (refer to Chapter 3 and the Appendix), will be relevant to those wishing to undertake economic and cost evaluations of interventions for continence conditions. There is also the *Review of Patient Satisfaction Measures for Continence Conditions* (Hawthorne, forthcoming report). A number of the recommended outcomes measures have been trialed in the 2004 South Australian Health Omnibus Survey (Harrison Health Research, 2004) and a report concerning these analyses is in preparation by Hawthorne and Sansoni.

Project Activities

The project drew its recommendations from a series of data gathering activities designed to create an evidence base to meet the project's objectives. These activities included:

- Expert review of the continence outcomes measurement literature
- Consultations with and survey of expert continence practitioners and researchers
- Reference group deliberations/expert panel of measurement specialists

The emphasis in this study was upon the identification of measures that are suitable for use in clinical settings for clinical practice as well as research settings, so there was a substantial degree of consultation with practitioners and researchers. These consultations were intended to complement the process of identification and formal rigorous review of the scientific literature in continence outcomes measurement.

In order to ensure that the recommended tools were applicable to a wide variety of both patient/client groups and practitioners, a content framework for the outcome measurement suite recommendations was devised. In this framework it has been assumed that there are three main identifiable groups who may use incontinence measures including:

- Primary care practitioners
- Specialist practitioners, and
- Researchers

This framework recognised that different groups have different requirements and that therefore a "one size fits all" approach was not appropriate. Researchers may have higher needs for measurement accuracy, for example, and less stringent requirements concerning practicability issues such as time and cost. The recommendations are intended to cross professional boundaries as it is acknowledged that workers in all three of the above categories may come from a wide variety of health professions including medicine, nursing, physiotherapy, occupational therapy, psychology, public health, epidemiology and range of other allied health professions. For instance, Primary Care practitioners can include general practitioners, community nurses, nurse practitioners, continence advisors and physiotherapists.

Outcomes of the continence outcomes measurement literature

The main body of the report reviews aspects of the very large research literature in this area. This includes a review of definitions and terminology concerning continence and incontinence that are found in the research literature. *(NB: In this work the terms continence and incontinence are used*

interchangeably.) The review also focuses on the psychometric properties of the many tools that have been developed to measure incontinence symptoms and quality of life. However, while this literature is voluminous, it is noted that it is characterised by many large gaps.

Outcomes of the review of definitions and terminology concerning incontinence

Defining incontinence

The vast majority of research articles in the area of incontinence do not define what they specifically mean by either urinary or faecal incontinence.

Definition of Urinary Incontinence

It is clear from the literature that urinary incontinence is a broad-ranging term used to describe a variety of conditions. Objective evidence of volume of urine loss or frequency of urine leakage may be required before a diagnosis of urinary incontinence is made; similarly, the urine loss may be required to be of a volume which impacts hygienically or socially on the sufferer. Questions arise regarding the appropriateness of classifying incontinence with terms to frequency while ignoring volume when the volume of leaked urine may be substantially different (1-2 drops compared with loss requiring a change of outer-clothes) (see for example Burgio et al. 1991). Other definitions of urinary incontinence focus on whether a need to change clothing or bed linen exists.

While some authors discuss (for example) "urge", "stress" or "mixed" incontinence as conditions with their own set of symptoms, related diagnosis, treatment protocols and outcome measures, other suggest these "conditions" are more properly viewed as symptoms of incontinence (compare, for example, O'Brien et al. 2001 with Blaivas, 1998 and Fonda et al. 1998). Importantly, however, the International Continence Society (ICS) standard definitions usefully isolate symptoms, signs and conditions.

The problem created by lack of precision is recognised by the ICS. In 2001 the ICS produced a report – *The standardisation of terminology in lower urinary tract function* – (see Abrams et al. 2002b), which defines lower urinary tract symptoms, signs suggestive of lower urinary dysfunction, urodynamic observations and conditions, conditions and treatment (referred to as the "new" definitions). The following definitions of symptoms, signs and conditions from the ICS report are particularly pertinent to this report:

- Symptoms are the subjective indicator of a disease or change in conditions as perceived by the patient, carer or partner and may lead him/her to seek help from health care professionals.
- Signs are observed by the physician to verify symptoms and quantify them. Conditions are defined by the presence of urodynamic observations associated with characteristic symptoms or signs and/or non-urodynamic evidence of relevant pathological processes.

This discussion is included to give an indication of the scope of definitions employed by researchers and clinicians for urinary incontinence. The range of definitions for "urinary incontinence" includes variations on the past definition developed by the ICS to those that are similar to the new definition:

- ICS New definition: *Urinary incontinence is the complaint of any involuntary leakage of urine.*
- ICS Old definition: *Urinary incontinence is the involuntary loss of urine that is a social or hygienic problem.*

The ICS recommended that both definitions be employed:

"Two definitions of incontinence are necessary firstly to reflect the true prevalence of incontinence (NEW definition) and secondly to describe the prevalence of troublesome incontinence that is likely to lead the sufferer to seek advice and treatment" (Abrams et al. 2002b).

The Outcomes Project Team agrees with the ICS recommendation that both definitions be employed. The first definition ("any incontinence") is appropriate for researchers, such as epidemiologists, and for research studies of incontinence prevention. The second definition ("incontinence provoking a social or hygienic problem") should be used by clinicians and primary care providers.

The mix of objective measurements with subjective (such as "social or hygienic problem") further exacerbates the development and adoption of a single, precise definition of urinary incontinence (Swithinbank et al. 1999). For example, Jackson's (1997a) findings that the "bother" of urine loss for women did not correlate with the volume of loss challenges definitions incorporating objective and subjective values to more precisely weight these values. Jackson's paper was written in relation to

patients suffering from urge incontinence, however more explicit guidance regarding the use of the two ICS definitions is necessary; it is clear that some people would be diagnosed with incontinence by one definition, but as continent with reference to the other.

Definition of cure and failure for Stress Urinary Incontinence outcomes

Lack of clarity with outcome definitions such as "cure" or "improvement" are also problematic. This situation is further clouded by researchers' and clinicians' claims to "cure" incontinence when one classification of incontinence is successfully treated or managed (such as stress incontinence) but another still exists (such as urge or overflow incontinence). The varieties of definitions of cure and improvement confound efforts to compare research and indeed analysis within individual studies.

The following working definition of cure and failure for stress urinary incontinence outcomes was adopted from Weber et al. (2001):

Cure of stress urinary incontinence is defined as:

(1) Resolution of the stress incontinence symptoms;

(2) Resolution of the sign (negative full bladder cough stress test, performed under the same conditions as before treatment); and

(3) No new symptoms or side effects. New symptoms or side effects should be specifically described and could include new urinary symptoms such as urinary urgency, frequency, urge incontinence, with or without urodynamic changes of detrusor overactivity (detrusor instability); change in sexual function; development or worsening of pelvic organ prolapse; adverse effect on bowel function; onset of urinary tract infections; surgical complications, such as foreign-body reaction to grafts, the development of fistula or diverticula; osteitis or osteomyelitis; neuropathy; and others. In studies using urodynamics after intervention the absence of genuine stress incontinence should be documented.

Failure of treatment of stress urinary incontinence is defined as persistent stress symptoms with the number of incontinent episodes unchanged, or worse, by voiding diary, plus a positive full bladder cough stress test (performed under the same conditions as before treatment), or genuine stress incontinence confirmed by urodynamic studies, with or without new symptoms or side effects.

Weber, Abrams, Brubaker et al. (2001)

Definition of cure and failure for Urge Urinary Incontinence outcomes

The following working definition of cure and failure for urge urinary incontinence outcomes was adopted from Weber et al. (2001):

Cure and failure for urge urinary incontinence outcomes is defined as:

For outcomes related to symptoms, cure is defined as the patient's statement that the symptom(s) is no longer present. In the case of frequency, there are seven or fewer micturitions per 24 hours. Failed treatment is defined as the patient's statement that the symptom(s) is no better or worse, with objective data from a urinary diary. As discussed above, the category of improvement cannot be specifically defined at present and requires further research to define outcomes of value to patients.

Improvement could include the patient's statement that the symptom(s) is less frequent or less troublesome with evidence from a urinary diary.

Detrusor overactivity (detrusor instability and hyperre-fiexia) is an urodynamic diagnosis characterized by involuntary phasic detrusor contractions during the filling phase of cystometry, which may be spontaneous or provoked, and which the patient cannot completely suppress. The detrusor contractions may be provoked by rapid filling, alterations of posture, coughing, walking, jumping, or other provocative manoeuvres. Outcomes for detrusor overactivity should be defined separately for symptoms, as described above, and for urodynamic findings. Cure of detrusor overactivity is defined as the absence of involuntary phasic detrusor contractions on filling cystometry. **Failure is defined as unimproved or worsened detrusor overactivity on urodynamics.** Again, the definition of improvement cannot be specified but, if used, the method of measurement should be precisely defined.

Weber, Abrams, Brubaker et al. (2001)

Definition of Faecal Incontinence

Very few articles addressing faecal incontinence define the term either generally or within the parameters of the research. Research articles specifically addressing terminology are rare. Those articles that do define their terms are more recent than those of urinary incontinence, perhaps reflecting the lengthier neglect of this area. Perhaps the development of new surgical methods to treat faecal incontinence has enabled interventions to be more readily measured. Again, similar to the discussion of urinary incontinence terminology, it is widely recognised that the epidemiology and impact of faecal incontinence can only be accurately discerned if the problem is accurately defined (see, for example, Reilly et al. 2000).

The 2nd International Consultation on Incontinence held in July, 2001 in Paris proposed the following on page 987 of their report (see Abrams et al. 2002a).

"Faecal incontinence has been variously defined and there are no internationally accepted or accredited definitions available. The Royal College of Physicians has proposed "the involuntary or inappropriate passage of faeces" [1]. An international panel of experts has defined "functional faecal incontinence" as "recurrent uncontrolled passage of faecal material for at least one month, in an individual with a developmental age of at least four years" [2].

> [1] Royal College of Physicians. Incontinence. Causes, management and provision of services. A Working Party of the Royal College of Physicians. *Journal of the Royal College of Physicians of London* 29, 272-274. 1995.

> [2] Whitehead, W.E., Wald, A., Diamant, N.E., Enck, P., Pemberton, J.H. and Rao, S.S.C. Functional disorders of the anus and rectum. *Gut* 45, 1155-1159. 1999.

The variety of faecal incontinence definitions suggests a similar scope to those of urinary incontinence. Definitions with reference to symptom, physiology, and frequency dominate. In addition to definitions of faecal incontinence are widely used severity ratings which incorporate symptoms and frequency. The most popular of these is the "Wexner Scale" (Jorge and Wexner, 1993). While not strictly a definition, projects have relied on the Wexner Scale to measure the impact of interventions and therapies and determining "what counts" as faecal incontinence and so is relevant to both this topic and the wider project. The scale seeks to "evaluate frequency and type of incontinence (solid, liquid, and gas and the need for pads) and its impact of lifestyle" (Yoshioka, Ogunbiyi and Keighley, 1999, p.253).

Cure and improvement also lack clear definitions within the literature, being largely dependent upon individual research parameters. Definitions of cure and/or success of intervention may be related to the prediction of outcomes. A study of pelvic floor retraining defined cure as a greater than 80% improvement (Rieger et al. 1997) and an operative intervention (graciloplasty) uses a decrease of 50% in incontinent episodes from baseline (Baeten et al. 2000). In contrast, Weber et al. (2001) proposed that cure be defined by a patient statement that the involuntary loss (of formed stool, liquid stool or gas) be absent within the relevant time frame.

The problem with faecal incontinence (compared to urinary leakage) is that even a small occasional loss of faeces or flatus can be associated with tell tale odour or noise. The patients are thus more frightened of the risk of leakage because any leakage provokes embarrassment and degradation. In contrast, a small occasional urine leak may not cause any smell and there is no noise component. Hence the definition of "cure" for anal incontinence may need to be stricter. Patients may not be able to achieve a normal lifestyle if there is residual fear of anal incontinence. This issue has not been addressed in any of the literature about anal incontinence.

Definition of cure and failure for faecal incontinence outcomes

The following working definition of cure and failure for faecal incontinence outcomes was adopted:

Cure and failure for urge faecal incontinence outcomes is defined as:

In defining the impact of interventions on faecal incontinence, cure was defined as complete resolution of the symptoms. Failed treatment (persistence or recurrence) was defined as no improvement or a worsening of symptoms. As with the other pelvic floor disorders, improvement could not be specifically defined but could include some favourable change in symptoms that may be based on quality of life measures, frequency of symptoms and consistency of loss. Further research is needed to develop clinically meaningful levels of improvement after intervention.

Weber, Abrams, Brubaker et al. (2001)

Outcomes of consultations with continence practitioners and researchers

Associated with this project a national meeting with incontinence clinicians and researchers was held in Melbourne on the 23rd of September, 2002. A second workshop was held in Adelaide on the 16th of October, 2002 associated with an incontinence scientific meeting. Here is a summary of the consultation outcomes.

Symptom Severity Indexes

The participants did not recommend any index.

Quality of Life Measures

The participants made no specific recommendations except that a recommendation based upon the literature would be useful.

Pad Tests

For primary care practitioners 24 hours monitoring should be employed as the minimum duration in pad tests but this was sometimes difficult to implement. For specialist practitioners and researchers, a minimum of 48 hours duration and preferably 72 hours should be employed in pad tests but that this was practically difficult.

Bladder Diary/Frequency Volume Chart

Frequency Volume Charts provide important insights but are difficult to implement. Bladder Diaries require the person to be cognitively intact.

Faecal Symptoms

The Wexner Faecal Incontinence Score ought to be considered but more basic tools would be useful.

Other tools

The participants did not endorse the wet check and paper towel test because they were considered not to provide sufficiently reliable and valid measures. These recommendations mainly accord with the ones derived from review of the published literature. The main additional findings from the consultations were:

- There is currently considerable variability in practice in the use of incontinence outcome measurement tools.
- The practicability of most measurement tools in the clinical context is problematic because of the length of the time required to complete and score the tools.
- There is considerable interest in the development of realistic and practical recommendations for outcomes measurement.

Outcomes of the reference group deliberations by the expert panel of measurement specialists

The panel reviewed the evidence collected by the other activities and then conducted several lengthy meetings to consider these materials. Each tool was considered with respect to the following evaluation criteria and associated scoring system.

Table 1   Criteria and weights used to assess instruments

Evaluation Criteria	Scoring system	Weight
Availability of comparison data	1 = not widely used in Australian and international incontinence research/clinical settings, 2 = some use in Australian and international incontinence research/clinical settings, 3 = wide use in Australian and international incontinence research/clinical settings	3
Length, ease and time to complete	1 = long instrument, 2 = medium length instrument, 3 = short instrument	2
Method of administration	1 = interviewer required, 2 = self-completion	2
Translations readily available	1 = available only in English, 2 = some translations available, 3 = many translations available	1
Ease of scoring	1 = scoring complex, 2 = scoring reasonably straightforward, 3 = scoring easy with computer code available	2
Sensitivity to incontinence	1 = not known to be sensitive, 2 = sensitive to incontinence status, 3 = good sensitivity to incontinence status	3
Reliability evidence available	1 = no or little published evidence identified, 2 = evidence suggests moderate reliability, 3 = evidence suggests good reliability	3
Validity evidence available	1 = no published validity evidence identified, 2 = evidence suggests moderate validity, 3 = evidence suggests good validity	3
Aderence to psychometric axioms	1 = does not meet basic axioms, 2 = partially meets basic axioms, 3 = mostly meets basic axioms	3
Cost of using the instrument	1 = costs charged for using instrument, 2 = costs charged for commercial use, 3 = instrument available free of charges	2

In all tables where the term "weighted total" appears this refers to the weighting scheme described in Table 1 .

Sansoni, Hawthorne and Thomas developed the weighting and scoring scheme for the criteria used in the technical evaluations of the measurement tools. The weighted total scores for each instrument were calculated by assigning a score for each criterion based on the expert panel judgments, multiplying the score by the assigned weight and then summing the scores over all criteria. In all cases, a higher score indicated the instrument might provide better measurement. A cut-off score at the mid-point of the scoring range, 47 points (the maximum possible score was 70 points), was selected as the level at or above which tools could be recommended for use. Tools with total weighted scores below this level were not recommended, as they were considered not to have sufficiently strong psychometric and associated properties.

The top tools were selected for review, provided that they scored 47 points or above. Tools for which there was no adequate published research were not included in the analysis. The types of measures reviewed include:

- Faecal Incontinence Symptom Measures
- Faecal Incontinence Health-Related Quality of Life Measures
- Urinary Incontinence Symptom Measures
- Urinary Incontinence Health-Related Quality of Life Measures

- Pad Tests
- Frequency Volume Charts and Bladder Diaries
- Multi-Attribute Utility Instruments
- Generic Health-Related Quality of Life Measures (General Health Status)
- Functional Measures

The following summary tables show the results of the expert psychometric panel deliberations concerning tool suitability for use in incontinence outcome measurement.

Table 2   Summary of ratings for faecal incontinence symptom measures

Criteria	Tool			
	Wexner	**Pescatori**	**St Marks**	**AMS**
Availability of comparison data/usage	3	2	1	1
Length, ease and time to complete	3	3	3	3
Method of administration	2	2	2	2
Translations available	3	2	2	1
Ease of scoring	3	3	2	2
Sensitivity to incontinence	1	1	1	1
Reliability evidence available	2	1	2	2
Validity evidence available	2	1	1	1
Adherence to psychometric axioms	2	2	2	2
Cost of using the instrument	3	3	3	3
Weighted Total	**55**	**45**	**43**	**42**

Wexner	Wexner/Cleveland Clinic Faecal Incontinence Score
Pescatori	Pescatori Faecal Incontinence Symptom Severity Measure
St Marks	St Marks Faecal Incontinence Grading System
AMS	American Medical Systems Faecal Incontinence Scale

The Wexner/Cleveland Clinic Faecal Incontinence Score was the only measurement tool to reach the required 47-point score in this category. Thus the only recommended Faecal Incontinence Symptom Severity Measure is the Wexner. It is noted that all measures save for the St Mark's Grading System do not address the issue of faecal urgency. The Wexner Scale and some items concerning faecal urgency have been included in the autumn 2004 South Australian Health Omnibus Survey (Harrison Health Research, 2004) and following psychometric analysis of these items it may be possible to further refine the Wexner Scale to address this issue.

Table 3 **Summary of ratings for faecal incontinence health-related quality of life measures**

Criteria	Tool	
	FIQLS	GIQLI
Availability of comparison data/usage	1	1
Length, ease and time to complete	1	1
Method of administration	2	2
Translations available	1	1
Ease of scoring	2	2
Sensitivity to incontinence	1	1
Reliability evidence available	2	1
Validity evidence available	1	1
Adherence to psychometric axioms	2	2
Cost of using the instrument	3	3
Weighted Total	**38**	**35**

FIQLS   Faecal Incontinence Quality of Life Scale
GIQLI   Gastro Intestinal Quality of Life Index

The available Faecal Incontinence Quality of Life measures are in an early stage of development. No tools reached the required 47-point cut-off score in this category and hence none are recommended.

Table 4 **Summary of ratings for urinary incontinence symptom measures**

Criteria	Tool								
	Kings HQ	UDI Long	UDI Short	ISI	BFLUTS	AUASI	ICS-male	ICS-male short	Dan-PSS-1
Availability of comparison data/usage	3	2	2	1	3	2	1	1	1
Length, ease and time to complete	2	2	3	3	2	3	2	2	2
Method of administration	2	2	2	2	2	2	2	2	2
Translations available	3	2	2	3	2	1	2	2	2
Ease of scoring	2	2	2	3	2	2	2	2	2
Sensitivity to incontinence	3	3	3	2	2	1	2	2	1
Reliability evidence available	3	3	2	3	2	2	2	2	2
Validity evidence available	3	3	3	3	2	2	2	2	2
Adherence to psychometric axioms	3	3	3	2	2	2	2	2	2
Cost of using the instrument	3	3	3	3	3	3	3	3	3
Weighted Total	**66**	**62**	**61**	**58**	**53**	**48**	**47**	**47**	**44**

KHQ   King's Health Questionnaire

UDI Long	Urogenital Distress Inventory Long Form
UDI Short	Urogenital Distress Inventory Short Form
ISI	Incontinence Severity Index
BFLUTS	Bristol Female Lower Urinary Tract Symptom assessment
AUASI	American Urological Association Symptom Index
ICS-male	International Continence Society Male assessment
ICS-male short	International Continence Society Male short form assessment
Dan-PSS-1	No English Acronym

Of the tools studied in the Urinary Incontinence Symptom Measures category, the King's Health Questionnaire, the Urogenital Distress Inventory (UDI) Short Form, the Urogenital Distress Inventory (UDI) Long Form, the Incontinence Severity Index, the Bristol Female Lower Urinary Tract Symptom assessment, the American Urological Association Symptoim Index, the International Continence Society Male assessment, and the International Continence Society Male short form assessment, made the required 47-point score threshold. Given their high ratings, the first two measures, the King's Health Questionnaire and the UDI (in either form) are the recommended tools in this category and the ISI is recommended for use in primary care and public health settings.

No recommendations were made for Urinary Incontinence Health-Related Quality of Life Measures.

Table 5   Recommendations for condition specific health-related quality of life measures in urinary incontinence

User category	Recommended Condition Specific Health-Related Quality of Life Measures
Primary Care Practitioner	None
Specialist Practitioner	None
Researcher	None

Table 6   Recommendations for pad tests

User category	Recommended use of Pad Test Measures
Primary Care Practitioner	24-hour pad test
Specialist Practitioner	24-hour pad test
Researcher	24-hour or 48-hour pad test

Table 7   Recommendations for frequency volume charts, bladder charts and diaries

User category	Recommended use of Frequency Volume Charts Bladder Charts and Diaries
Primary Care Practitioner	ICS/WHO templates for 3 days
Specialist Practitioner	ICS/WHO templates for 3 days
Researcher	ICS/WHO templates for 3 to 7 days

Table 8   Summary of ratings for multi-attribute utility instruments

Criteria	Tool						
	EQ5D	AqoL	HUI3	15D	SF6D	QWB	Rosser
Availability of comparison data/usage	3	2	2	1	1	2	2
Length, ease and time to complete	3	2	2	2	1	1	1
Method of administration	3	2	2	2	3	1	1
Translations available	3	1	2	1	3	1	1
Ease of scoring	3	3	3	3	3	2	1
Sensitivity to incontinence	1	2	2	3	1	1	1
Reliability evidence available	2	2	2	1	1	1	1
Validity evidence available	2	2	2	1	1	1	1
Adherence to psychometric axioms	2	3	3	1	1	1	1
Cost of using the instrument	2	3	1	3	3	3	3
Weighted Total	**55**	**54**	**51**	**42**	**38**	**33**	**31**

EQ5D   European Quality of Life Measure – 5D (formerly the EUROQOL)
AqoL   Assessment of Quality of Life
HUI3   Health Utilities Index – Version 3
15D   Fifteen-Dimensional measure of health-related quality of life
SF6D   Short Form 6 Dimensions
QWB   Quality of Well Being
Rosser   Rosser Quality of Life Index

Of the tools studies in the Multi-Attribute Utility instruments category, three obtained the requisite 47-point score or higher. These were the Assessment of Quality of Life (AQoL), the European Quality of Life Measure-5D (EQ5D) and the Health Utilities Index (HUI3). All three tools are recommended.

Additional Multi-Attribute Utility instrument recommendations

Given the uncertainties in the literature, a number of options could be considered either individually or collectively when using Multi-Attribute Utility instruments for incontinence conditions:

1.   A single MAU-instrument could be recommended as the preferred instrument of choice for routine use at the clinician- and specialist-levels. This instrument should be short, easy to administer and score and population norms could be made available for easy reference. If such a policy was adopted, it would be in light of the limitations outlined in this report and there would be no guarantee that results obtained would be comparable with results obtained elsewhere using another instrument. Indeed, where QALYs were computed as the result of a treatment, it is likely these would reflect instrument choice as much as treatment effect. Where two MAU-instruments were recommended as the preferred measures, these difficulties would be compounded if some studies included one of the instruments and other studies opted for the other instrument.

2.   To overcome this uncertainty, it could be recommended that two MAU-instruments be included in any particular research or evaluation study, and that researchers be encouraged to provide both sets of results. One of the recommended instruments should be that recommended for clinician use. This strategy would have the benefit of reducing the bias inherent in a one-instrument strategy, and it would produce a range of estimated benefits from interventions, thus acknowledging the limitations of relying upon any particular existing MAU-instrument. Given that, inevitably, comparisons will be

made with incontinence studies overseas, this strategy would have the further benefit of enabling cross-cultural comparisons.

3. Several instruments could be trialled in 3-4 large incontinence studies for the explicit purpose of identifying the instrument to be recommended for future use. Whilst this would impose an immediate burden for, say, 3 to 5 years, it would enable many of the questions raised in this report regarding the validity of MAU-instruments to be thoroughly investigated in an Australian context. This would place Australia in a position of world leadership in incontinence and utility research; it would enable a fully informed decision to be made regarding instrument selection; and it is likely the Australian model would become the world standard in the immediate future given the paucity of current research in the field. Should this latter scenario eventuate, it is likely this would enhance international cooperation in the field.

4. As an alternative to #3, consideration could be given to including the instruments under consideration in the 2004 South Australian Health Omnibus Survey (HOS), together with suitable questions on incontinence and incontinence-related health sequelae. This would enable the rapid collection of data and its analysis leading to instrument selection and recommendation. Since the HOS involves drawing a weighted population sample, the findings could be used to establish population norms against which future work could be interpreted. (NB: This work is currently in progress.)

5. A specific study could be funded to develop an incontinence module for attachment to a generic MAU-instrument descriptive system. This recommendation arises from the consideration that there are HRQoL areas of concern to those with incontinence that are not addressed with fully generic instruments. If an incontinence module for an existing instrument were constructed, researchers would be in a position to report both incontinence-specific HRQoL effects and generic utility scores. This model has been followed by the SF-36, for which there are now many disease-specific modules, and it is being followed by the World Health Organisation Quality of Life Group for the WHOQOL-OLD (being specifically developed for use with older adults) (see Murphy & Hawthorne, 2001), and also in Australia in the area of visual impairment and the AQoL. The chief difficulty lies in selecting the base instrument.

Table 9   Summary of ratings for generic health-related quality of life measures (general health status)

Criteria	Tool						
	SF 36 (II)	SF 36 (I)	SIP	WHO-QOL Bref	SF 12	NHP	WHO-QOL 100
Availability of comparison data/usage	2	3	2	2	2	2	1
Length, ease and time to complete	2	2	1	2	2	1	1
Method of administration	2	2	2	2	2	2	2
Translations available	2	2	2	2	2	2	2
Ease of scoring	3	3	2	2	3	2	2
Sensitivity to incontinence	3	3	2	2	2	2	3
Reliability evidence available	3	2	2	2	2	2	2
Validity evidence available	3	2	2	1	1	1	1
Adherence to psychometric axioms	3	3	2	2	2	2	2
Cost of using the instrument	1	1	3	3	1	3	3
Weighted Total	**60**	**57**	**48**	**47**	**45**	**45**	**45**

SF 36 (II)	SF-36® Health Survey (Version 2.0)
SF 36 (I)	SF-36® Health Survey (Version 1.0)
SIP	Sickness Impact Profile
WHO-QOL Bref	World Health Organisation Quality of Life Assessment Bref (26 item) version
SF 12	SF-12® Health Survey (Version 1.0)
NHP	Nottingham Health Profile
WHO-QOL 100	World Health Organisation Quality of Life Assessment 100 item versionThe SF-36® Health Survey (Version 1.0),

The SF-36® Health Survey (Version 1.0) the SF-36® Health Survey (Version 2.0), the Sickness Impact Profile, and the World Health Organisation Quality of Life Assessment Bref were the only tools in the Generic Health-Related Quality of Life Measures category to reach the required 47-point cut-off score. Given their high ratings, the first two measures, the SF-36® Health Survey (Version 1.0) and the SF-36® Health Survey (Version 2.0), are the recommended tools in this category. Both tools are recommended with a preference for the new Version 2.0 especially when forthcoming Australian normative data becomes available. The SF-36 Version 2.0 is less prone to ceiling and floor effects on the role functioning scales, which may be an important consideration for patients with incontinence.

Table 10   Summary of ratings for functional measures

	Tool		
Criteria	**FIM**	**Barthel**	**Katz**
Availability of comparison data / usage	3	2	1
Length, ease and time to complete	1	2	2
Method of administration	1	1	1
Translations available	2	2	2
Ease of scoring	1	3	2
Sensitivity to incontinence	3	2	1
Reliability evidence available	3	2	1
Validity evidence available	3	2	1
Adherence to psychometric axioms	3	2	1
Cost of using the instrument	1	3	3
Weighted Total	**55**	**50**	**33**

FIM	Functional Independence Measure
Barthel	Barthel ADL Index
Katz	Katz ADL Scale

The functional measure category had two tools that reached the 47-point cut-off score. These were the Functional Independence Measure and the Barthel.

Overall outcome measurement tool recommendations

The following overall recommendations for faecal and urinary incontinence outcome measures are based upon the research literature evidence reviewed by the project team, expert panel deliberations and consultations with practitioners. Where multiple tools are recommended within a category the first ranked tool appears first in the numbering.

Table 11   Summary of recommended instruments

Tool Content Domain	User Category		
	Primary Care Practitioner	**Specialist Practitioner**	**Researcher**
Faecal Incontinence Symptom Measures	Wexner Symptom Scoring System	Wexner Symptom Scoring System	Wexner Symptom Scoring System
Faecal Incontinence Health-Related Quality of Life Measures	None	None	None
Urinary Incontinence Symptom Measures	ISI	1. King's Health 2. UDI	1. King's Health 2. UDI
Urinary Incontinence Health-Related Quality of Life Measures	None	None	None
Pad Tests	24-hour pad test	24-hour pad test	24 or 48-hour pad test
Frequency Volume Charts and Bladder Diaries	ICS/WHO templates for 3 days	ICS/WHO templates for 3 days	ICS/WHO templates for 7 days
Multi-Attribute Utility Instruments	None	1. EQ5D 2. AQoL 3. HUI3	1. EQ5D 2. AQoL 3. HUI3
General Health-Related Quality of Life Measures (General Health Status)	1. SF-36 V2 2. SF-36 V1	1. SF-36 V2 2. SF-36 V1	1. SF-36 V2 2. SF-36 V1
Functional Measures	Barthel	1. FIM 2. Barthel	1. FIM 2. Barthel

The contribution of this work is twofold. First, a rigorous methodology for tool evaluation has been developed. It includes criteria that are important to practitioners and researchers alike in the selection of measurement tools. This is considered to be an important innovation as it means that the process of formulation of these recommendations has been made transparent for all to see. This is very different from the approaches taken in previous endeavours in this field to date where the criteria for tool evaluation have not been explicitly stated nor weighted. Second, a practical set of recommendations for tools to be used by practitioners and researchers has been made.

Durability of the recommendations arising from the present study

As has been noted throughout this paper, many areas in the measurement of outcomes for continence are in a state of flux. For example, it has been noted that only recently have credible tools been developed in the area concerning faecal incontinence health related quality of life measures and that these await proper psychometric evaluation. However, in the other areas such as the development of urinary incontinence symptom measures, it may well be the case that this period will see significant development of the evidence base.

In other areas, such as the generic quality of life measures and the multi-attribute utility measures, there is a much lengthier and sounder research base upon which to make recommendations. In these areas it is unlikely that within the next two to three years that there will be such significant developments so as to require the revision and overturning of current recommendations.

Thus, in formulating recommendations concerning the durability of the recommended suite of outcome measures, it is important to consider the differential states of development of the respective areas of the relevant literatures. For this reason, we suggest that a review period of two to three years would be satisfactory after which the evidence base should be re-examined and the recommendations updated if necessary. For instance, during the finialisation of this publication, a number of important papers emerged in the literature. These include: Cockell et al. (2003) on the Postpartum Flatal and Fecal Incontinence Quality of Life Scale; the development of a new set of measures for urinary incontinence by Avery et al. (2004); a new paper by Rockwood (2004) on severity and quality of life scales for faecal incontinence; and two reviews of urinary incontinence quality of life meaures (Corcos et al. 2002 and Naughton et al. 2004). Also the recent South Australian Health Ominbus Survey (Harrison Health Research, 2004) will provide updated epidemological data about incontinence for the Australian population.

Recommendation: It is suggested that a review period of two to three years would be satisfactory after which the evidence base should be re-examined and the recommendations updated if necessary.

Recommendations concerning research priorities in continence outcome measurement research

Having made recommendations for the selection of continence outcome measurement tools for the various categories of practitioners in the various content domains, it is considered that there are three major research priorities that need to be addressed in subsequent studies:

- There is a need for some large-scale omnibus studies using the recommended tools in the Australian population to further explore validity, reliability, and applicability and practicability considerations within targeted populations with the full range of incontinence problems. The methodology employed by Vaizey, Garapeti, Cahill, & Kamm (1999) in their omnibus study provides a useful model for the proposed study or studies. The recommended tools should be trialled together in a study or linked studies to determine which are most suitable and also to develop appropriate Australian norms. Some tools such as the SF tools are in a better position in this regard than others, but most do not have convincing large-scale benchmark studies to date. There are nevertheless, opportunities to conduct them that ought to be pursued (refer also to the recommendations concerning multi-attribute utility instruments).

- There is a need for development of short form tools and common minimum data sets for use by primary care and specialist incontinence practitioners so that the needs of such practitioners can be better met. Although there are many tools in existence, many have poor or undemonstrated psychometric properties with lengthy formats that are unsuited to the generalist primary care practitioner. In various jurisdictions, some tools have been developed to capture the more general patient characteristics. Such tools include the Initial Needs Identification tool developed for the Victorian Primary Care Partnerships initiatives. There are also various coding dictionaries such as those developed by the Australian Institute for Health and Welfare, which are helpful in the design of common coding systems for minimum data sets. However, there is a dearth of usable tools in the area of incontinence, which are practical for primary care practitioners. As such practitioners provide the bulk of services to people with incontinence this is a significant deficiency in current capacity.

- Examination of the use of outcome measures for incontinence conditions across different cultural and lingustic groups is lacking in Australia. This is especially the case for the Aboriginal community. Here perhaps, the work of Dr. Kate A. Senior could be used as a guide (Senior, 2003).

NB: for instruments referred to in the executive summary can be found in the full technical report.

The Full Technical Report

Chapter 1   Introduction to the Full Technical Report

Background

This document is the final report of the Continence Outcomes Measurement Suite project. The project was conducted by A/Professor David Fonda, A/Professor Kate Moore, Professor Rhonda Nay, Professor Shane Thomas and a team of technical experts (Jan Sansoni, A/Professor Graeme Hawthorne and Nicholas Marosszeky) and support staff (Amanda Weeks, Cath Stoove and Elizabeth Kearney). The project was commissioned by the Australian Government Department of Health and Ageing, National Continence Management Strategy Research Program with the goal of recommending a suite of continence outcome measures to be used by clinicians and researchers in Australia.

This project is also linked to a number of other concurrent research projects on continence conditions undertaken for the Department of Health and Ageing. These include the *Development of a Framework for Economic and Cost Evaluation of Continence Conditions* (Moore et al. 2004) which, when used in conjunction with this report (refer to Chapter 3 and the Appendix), will be relevant to those wishing to undertake economic and cost evaluations of interventions for continence conditions. There is also the *Review of Patient Satisfaction Measures for Continence Conditions* (Hawthorne, forthcoming report). A number of the recommended outcomes measures have been trialed in the 2004 South Australian Health Omnibus Survey (Harrison Health Research, 2004) and a report concerning these analyses is in preparation by Hawthorne and Sansoni.

(NB: In this work the terms continence and incontinence are used interchangeably.)

What is the purpose of the Continence Outcomes Measurement Suite project?

It is widely accepted that outcomes measurement is an integral part of effective evidence based clinical practice. Sansoni (2002) noted that a focus on health outcomes is an important extension of previous work on health inputs and throughputs. She argued that this focus on health outcomes in the Australian context has derived from a series of interrelated factors including increased expenditure, recognition of the limited information about the effects of many services and treatments, concerns about quality of care and concerns about whether new technologies are actually improving patient well-being. The pre-occupation with outcomes measurement is a natural progression of the now strongly supported evidence based framework used widely in clinical and health services research.

Within this framework the ability to measure the effectiveness of interventions is essential. While this principle is widely agreed, there is considerable variation in the measurement of continence and the outcomes of continence interventions. Some of the measures used in the published literature do not have established norms or established levels of reliability and validity. Variation in measurement protocols means that the impact of different interventions and study population characteristics and needs cannot be directly compared. These limitations contribute to the lack of uniform standards for the use of measurement tools in this area and are a threat to the use of outcome measurement tools in an efficient and effective manner.

Both Sansoni (2002) and Eagar (2000) advocate the definition of health outcomes provided by the Australian Health Minister's Advisory Council. This is:

> "A health outcome is a change in the health of an individual, or a group of people which is wholly or partly attributable to an intervention or series of interventions".
> (Australian Health Ministers Advisory Council (AHMAC) February 1993; modified by the National Health Information Group, AHMAC, 1996.)

As Eagar notes, this definition has several ideas embedded within it including sensitivity to change, the measurement of health or health status and the ability to attribute that change to an intervention. This latter feature is an issue that relates to intervention study research design rather than the design of a measurement tool (Polgar and Thomas, 2004). In this project the emphasis has been placed strongly upon the measurement component of the definition.

Roessner's review (2002) of outcome measurement in the United States provides an important overview of the use of outcome measurement in government programs and policy. Roessner provides a useful typology of research and evaluation questions as follows and illustrates the central importance of outcome measures to address these questions.

- Should this program be continued or terminated?
- What are the economic payoffs from this program? (Return-on-investment; benefit/cost).
- How are the outcomes and impacts from this program realized? (Process or formative evaluation).
- Is the program generating results that are in accord with accepted standards of scientific quality? (Merit or peer review).
- What are the (measurable) outputs from this program? (Monitoring or performance measurement).

The purpose of this project was to give thorough consideration to the various available measures and to recommend a suite of continence outcome measures for use by Australian clinicians and researchers. During the course of the project and in consultations with clinicians and researchers, it became evident that no single set of indicators could fulfil such a broad purpose. Clinicians working in the field of continence, as in most other fields, are heavily constrained by time and resources. Their needs and objectives are different from those of researchers who deal with much smaller volumes of participants with much higher levels of resources that can be devoted to the measurement process. Thus, it was decided early in the project to provide separate advice concerning appropriate outcome measures for clinicians, specialist clinicians and researchers rather than to attempt a "one size fits all" approach. It is considered that this targeted approach provides a useful framework for projects involving recommendations for the standardisation of measures in any health context.

The evidence philosophy of the project

Prior to describing the process and outcomes of the project, it is important to articulate the approach taken in the gathering and use of evidence. Evidence in the clinical research context is frequently taken as meaning the use of the Cochrane collaboration typology of research evidence with randomised control trials sitting at the apex of the evidence pyramid. This is all very well for an intervention evidence framework, but the present project is about the identification of technically sound measurement tools that also meet acceptability criteria by different user groups. Technical soundness as traditionally measured by psychometric properties, such as reliability and validity are necessary but not sufficient conditions for inclusion in a measurement suite to be used in clinical and research settings. Measurement tools have four properties that need consideration including reliability and validity (this is where most technical evaluations stop) and applicability and practicability (where most clinicians make their selection decisions). In order to deliver a technically sound and useful set of recommendations that is likely to be adopted by clinicians and clinical researchers, all four of these characteristics need to be taken into account. In ventures such as expert advisory panels sponsored by the International Continence Society, these are exactly the criteria that are taken into account by the panel, although the deliberations concerning how judgments are made of the levels of practicability and applicability issues have generally not been well documented. This stands in contrast with the present work where the criteria used and the judgments made for the formulation of these recommendations have been explicitly expressed.

The evidence base for reliability and validity of continence outcome measures is reasonable although it is patchy in some areas for some tools. Many tests and measures in this field are developed and used without the reporting of a systematic psychometric evaluation. However, the evidence base for applicability and practicability is actually even poorer. It is remarkable that many papers focussing on measurement tools neglect to include basic data such as the administration and scoring times of the tools in order that evidence based judgments may be made about practicability, for example.

The consequence of evidence philosophy and the need to explicitly and systematically consider issues such as practicability in formulating our recommendations led to an emphasis upon expert consultation in this project to complement the technical evaluations that have been conducted in this review of the published literature.

Recent relevant work in incontinence outcome measurement

The World Health Organisation in association with the International Continence Society hosted two international consultations on incontinence in 1998 and 2001 in which important recommendations were made concerning outcomes measurement. The work of this project was informed by these important international benchmark works.

The 2nd International Consultation on Incontinence

The 2nd International Consultation on Urinary Incontinence was held in July, 2001 in Paris. The consultation was organised by the International Continence Society (ICS) and the International Consultation on Urological Diseases (ICUD) in collaboration with the World Health Organisation (WHO). The purpose of the consultation was to develop recommendations for the diagnostic evaluation and treatment of urinary and faecal incontinence and pelvic organ prolapse. A series of sub-committee reports was developed and peer reviewed. The Scientific Committee, consisting of the Chairs of all the committees then refined the final recommendations. Two of the present project team members were leading participants in the consultations (see Abrams et al. 2002a).

The content of the recommendations from the various sub-committees that focused on outcome measurement have been thoroughly analysed. Of course, many of the recommendations relate to other matters and while these are of general interest these have not been included in this discussion.

The Research Methodology in Urinary Incontinence Sub-Committee made the following recommendations to this work:

1. Clinician's observations of anatomy should be recorded using standardized, reproducible measurements.
2. Pelvic muscle and voluntary sphincter function should be reported using a quantifiable scale.
3. These measures should be repeated after intervention and correlated with primary clinical outcome measures.
4. Clinical trials of incontinence and LUTS should include bladder diaries as an essential baseline and outcome measure.
   a. The diary should include measured voided volume (for at least one day if a multi-day diary is employed).
   b. 24-hour diaries are adequate for most studies.
5. Clinical trials of incontinence and LUTS should include a pad test as an essential baseline and outcome measure.
6. At this time, clinical studies should enrol subjects by carefully defined symptoms, not urodynamic findings.
7. In all trials, standardized urodynamic protocols (based on ICS recommendations) are defined at the outset. In multicenter trials, urodynamic tests are interpreted by a central reader to minimize bias.
8. Research in incontinence and LUTS should include both generic and condition-specific HRQOL instruments.
9. Changes in HRQOL after therapy should be correlated with changes in individual symptoms, and with physiologic and anatomic outcome measures to learn how the particular therapy is working.

LUTS = Lower Urinary Tract Symptoms

The committee also made recommendations concerning specific clinical groups including, men, women, children and the frail elderly. Essentially the recommendations were at a fair degree of generality and are not particularly useful in terms of this task for the selection of specific measures. Moreover, it should be noted that the report of the committee is replete with notes concerning the lack of research literature to provide a sound evidence base for the selection of appropriate tools.

The Quality of Life Sub Committee of the Consultation in contrast made a series of specific recommendations for measurement tools (see Abrams et al. 2002a). These were as follows:

> GRADE A: HIGHLY RECOMMENDED
>
> Urogenital Distress Inventory (UDI)
>
> UDI-6
>
> Urge-UDI
>
> King's Health Questionnaire
>
> (Women only) Incontinence Severity Index
>
> (Men only) DAN-PSS-1
>
> (Men only) ICS male SF

GRADE B: RECOMMENDED
Bristol Lower Urinary Tract Symptoms
Symptom Severity Index

This is a useful starting list. However there were several major issues associated with it. The first was the length of the recommended list. While freedom of choice is a good thing, it also means that there would be little commonality in data across different facilities and researchers. This lack of commonality would potentially impact adversely upon the ability to benchmark and to construct large multi-centre data sets. The second issue was that of the imprecise specification of the criteria used to select the measures. The emphasis in this work seemed to be very squarely upon researchers, which contrast with the requirement of the current study to provide advice to primary care practitioners and specialist practioners, as well as researchers.

The Economics of Incontinence Sub-Committee provided a methodological checklist for the conduct of economic studies (see Abrams et al. 2002a). The component of this checklist that directly related to outcome measurement included the following elements:

- Explicitly stated
- Appropriate
- Utility of measure (source of data, appropriate methods)

As with many of the other Consultation sub-committees the committee discussed the current lack of standardisation of measurement tools but did not make detailed and specific recommendations for particular tools.

The Urodynamics Sub-Committee also made similarly developmental but non-specific recommendations (see Abrams et al. 2002a). The ones pertinent to outcome measurement included:

- Development of improved techniques for assessing bladder sensation.
- Minimization of the invasiveness of urodynamics techniques (e.g. refinement of ultrasonic imaging techniques) so that they are more widely applicable and can be carried out in more natural settings, with a minimum of personnel and expense.
- Development of methods of quantification of urodynamic observations, to improve reliability, inter-changeability, and clinical relevance.

These recommendations reflected the committee's views and evidence that urodynamics measurement tools were at a relatively early stage of development and that those in current use were not practicable in the general clinical setting.

The outcomes of the 2nd International Consultation on Urinary Incontinence are of particular pertinence to the present project (see Abrams et al. 2002a). However, detailed analysis of the recommendations shows that there was general acknowledgement that outcomes measurement in incontinence was at an early stage of development, notwithstanding the wide range of alternative tools in some measurement domains such as Health Related and Condition Specific Quality of Life. The focus was upon basic reliability and validity criteria and there was limited evidence available concerning these in a wide range of tools. These deficiencies in the evidence base constrained the specificity of the recommendations that were possible.

UK Urinary Incontinence Health Outcome Indicators Working Party

In the United Kingdom, the National Centre for Health Outcomes Development presented a major report to the Department of Health concerning urinary incontinence health outcome indicators. The approach taken by the working group was quite different from the various other working groups and meetings that had been hosted by the WHO/ICS in that they had much more of a population health and hospital based service perspective (see Brocklehurst et al. 1999).

UK Urinary Incontinence Health Outcome Indicators Working Party Recommendations

This work was on acute health settings and is of limited applicability to our work. Moreover, in terms of the specifics of outcome measures, in many of the discussions in the Working Party's report it was noted that there were no existing tools available that would meet appropriate psychometric criteria. Thus, while the British report is of some use to clinical services involved in the provision of acute urinary incontinence services, recommendations are probably not of wider use. Indeed, it is arguable as to

whether many of the indicators listed would be considered as outcome measures in the conventional sense. In some typologies the information would be to categorise many of these indicators as in fact being process, rather than being outcome measures. Thus, the utility of this work for the current project is somewhat limited but it is important to note the context of this previous work.

Project methodology

As outlined above, because the emphasis in this study was upon the identification of measures that are suitable for use in clinical settings for clinical practice as well as research settings, there was a substantial degree of consultation with practitioners and researchers. These consultations were intended to complement the process of identification and formal rigorous review of the scientific literature in continence outcomes measurement.

The project included the following activities:
- Expert review of continence outcomes measurement literature
- Consultations with and survey of expert continence practitioners and researchers
- Reference group deliberations/expert panel of measurement specialists

Each activity undertaken in the project is described:

Expert review of incontinence outcomes measurement literature

This involved the collation and review of the incontinence outcomes measurement literature. A large body of literature is available with a large number of tools developed with varying attributes. A search for continence or incontinence on Medline generates in excess of 30,000 "hits". A thorough search for and analysis of relevant incontinence outcomes measurement literature was conducted.

The evaluation of suitability of measurement tools for their purpose requires a consistent and articulated set of criteria. Philosophies concerning the evaluation of tests and measures in Chapter 2 of this report are outlined. Jan Sansoni, A/Professor Graeme Hawthorne and Nicholas Marosszeky assisted with expert psychometric review of the relevant measures.

Consultations with and survey of expert continence practitioners and researchers

Notwithstanding the importance of adequately incorporating the findings of the scientific literature, local expertise and practice were also incorporated into the deliberations. The following topics were raised within the expert groups:
- Frequency Volume Chart
- Bladder Chart
- Wet Checks in older people
- Urilos and other electronic devices
- Pad Tests
- Other home tests
- Scoring systems of measures
- Quality of Life instruments
- Generic
- Disease specific
- Utility measures and QALYS
- Outcome measures currently used and advocated

The purpose of the consultations was to elicit expert views and document them thoroughly in order that they may inform, along with the research literature, the choice of continence outcomes measures. Several meetings were conducted and for those who were unable to attend the meetings, a structured interview was offered in order to maximise the inclusiveness of the process.

Reference group deliberations

The reference group was established to assist the research team in its deliberations concerning the outcomes measurement suite to be recommended. The task was both technical and to act as a reality check to ensure acceptability of the recommendations.

A content framework for the outcome measurement suite recommendations

In order to ensure that the recommended tools were applicable to a wide variety of both patient/client groups and practitioners, we devised a content framework for the outcome measurement suite recommendations. In this framework we assumed that there are three main identifiable groups who may use incontinence measures including:

- Primary care practitioners
- Specialist practitioners
- Research workers

This framework recognised that different groups have different requirements and that therefore a "one size fits all" approach was not appropriate. Researchers may have higher needs for measurement accuracy, for example, and less stringent requirements concerning practicability issues such as time and cost requirements. These recommendations are intended to cross professional boundaries as it is acknowledged that workers in all three of the above categories may come from a wide variety of health professions including medicine, nursing, physiotherapy, occupational therapy, psychology, public health, epidemiology and range of other allied health professions. For instance, Primary Care professionals can include general practitioners, community nurses, nurse practitioners, continence advisors and physiotherapists.

Defining incontinence

The vast majority of research articles in the area of incontinence do not define what they specifically mean by either urinary or faecal incontinence.

Definition of Urinary Incontinence

It is clear from the literature that urinary incontinence is a broad-ranging term used to describe a variety of conditions. Objective evidence of volume of urine loss or frequency of urine leakage may be required before a diagnosis of urinary incontinence is made; similarly, the urine loss may be required to be of a volume which impacts hygienically or socially on the sufferer. Questions arise regarding the appropriateness of classifying incontinence with terms to frequency while ignoring volume when the volume of leaked urine may be substantially different (1-2 drops compared with loss requiring a change of outer-clothes) (see for example Burgio et al. 1991). Other definitions of urinary incontinence focus on whether a need to change clothing or bed linen exists.

While some authors discuss (for example) "urge", "stress" or "mixed" incontinence as conditions with their own set of symptoms, related diagnosis, treatment protocols and outcome measures, other suggest these "conditions" are more properly viewed as symptoms of incontinence (compare, for example, O'Brien et al. 2001 with Blaivas, 1998 and Fonda et al. 1998). Importantly, however, the ICS standard definitions usefully isolate symptoms, signs and conditions.

The problem created by lack of precision is recognised by the ICS. In 2001 the ICS produced a report – The standardisation of terminology in lower urinary tract function – (see Abrams et al. 2002b), which defines lower urinary tract symptoms, signs suggestive of lower urinary dysfunction, urodynamic observations and conditions, and conditions and treatment (referred to as the "new" definitions). The following definitions of symptoms, signs and conditions from the ICS report are particularly pertinent to this report:

- Symptoms are the subjective indicator of a disease or change in conditions as perceived by the patient, carer or partner and may lead him/her to seek help from health care professionals.
- Signs are observed by the physician to verify symptoms and quantify them. Conditions are defined by the presence of urodynamic observations associated with characteristic symptoms or signs and/or non-urodynamic evidence of relevant pathological processes.

This discussion is included to give an indication of the scope of definitions employed by researchers and clinicians for urinary incontinence. The range of definitions for "urinary incontinence" includes variations on the past definition developed by the ICS and those that predict the new definition:

- ICS New definition: *Urinary incontinence is the complaint of any involuntary leakage of urine.*
- ICS Old definition: *Urinary incontinence is the involuntary loss of urine that is a social or hygienic problem.*

The ICS recommended that both definitions be employed:

"Two definitions of incontinence are necessary firstly to reflect the true prevalence of incontinence (NEW definition) and secondly to describe the prevalence of troublesome incontinence that is likely to lead the sufferer to seek advice and treatment" (Abrams et al. 2002b).

The Outcomes Project Team agrees with the ICS recommendation that both definitions be employed. The first definition ("any incontinence") is appropriate for researchers, such as epidemiologists, and for research studies of incontinence prevention. The second definition ("incontinence provoking a social or hygienic problem") should be used by clinicians and primary care providers.

The mix of objective measurements with subjective (such as "social or hygienic problem") further exacerbates the development and adoption of a single, precise definition of urinary incontinence (Swithinbank et al. 1999). For example, Jackson's (1997) findings that the "bother" of urine loss for women did not correlate with the volume of loss challenges definitions incorporating objective and subjective values to more precisely weight these values. Jackson's paper was written in relation to patients suffering from urge incontinence, however more explicit guidance regarding the use of the two ICS definitions is necessary; it is clear that some people would be diagnosed with incontinence by one definition, but as continent with reference to the other.

Lack of clarity with outcome definitions such as "cure" or "improvement" are also problematic. This situation is further clouded by researchers' and clinicians' claims to "cure" incontinence when one classification of incontinence is successfully treated or managed (such as stress incontinence) but another still exists (such as urge or overflow incontinence). Similarly the varieties of definitions of cure and improvement confound efforts to compare research and indeed analysis within individual studies. Chakrin et al. (1999) raise this concern within their study:

> "Outcomes following incontinence surgery may vary depending on how the analysis was performed, patient selection, definition of success and so forth ... Cure implies a restoration to normal yet this is usually not the case."

They suggested that the terms "responders" and "non-responders" or improvement measured as a percentage of normal would provide more meaningful measurement information. This may also address the discrepancy between clinical measurements of cure and improvement and patients' own perceptions.

Groutz et al. (2000) describe the minimal standards proposed by the Urodynamic Society to assess and report on the efficacy for therapy for urinary incontinence therapy. These include:

- The patient's opinion of treatment outcome
- Micturation questionnaire
- Voiding diary
- Pad-test
- Physical examination
- Uroflowmetry, and
- Estimation of post-void residual urine

Very recently, Robinson et al. (2003) surveyed 156 urogynaecologists (33) urologists (29) gynaecologists (21) about their concept of "cure". Overall, 86% felt that a "good improvement in symptoms that allowed a return to normal lifestyle" was a realistic outcome. In clinical practice, occasional small leak was accepted as within the definition of "cure" by 54% of these clinicians. In the research setting, 30% defined cure on Quality of Life and pad test results, and 31% defined cure on QoL and either urodynamic or pad testing results.

Definition of cure and failure for Stress Urinary Incontinence outcomes

The following working definition of cure and failure for stress urinary incontinence outcomes was adopted:

Cure of stress urinary incontinence is defined as:

(1) Resolution of the stress incontinence symptoms;

(2) Resolution of the sign (negative full bladder cough stress test, performed under the same conditions as before treatment); and

(3) No new symptoms or side effects. New symptoms or side effects should be specifically described and could include new urinary symptoms such as urinary urgency, frequency, urge incontinence, with or without urodynamic changes of detrusor overactivity (detrusor instability); change in sexual function; development or worsening of pelvic organ prolapse; adverse effect on bowel function; onset of urinary tract infections; surgical complications, such as foreign-body reaction to grafts, the development of fistula or diverticula; osteitis or osteomyelitis; neuropathy; and others. In studies using urodynamics after intervention the absence of genuine stress incontinence should be documented.

Failure of treatment of stress urinary incontinence is defined as persistent stress symptoms with the number of incontinent episodes unchanged, or worse, by voiding diary, plus a positive full bladder cough stress test (performed under the same conditions as before treatment), or genuine stress incontinence confirmed by urodynamic studies, with or without new symptoms or side effects.

Weber, Abrams, Brubaker et al. (2001)

Definition of cure and failure for Urge Urinary Incontinence outcomes

The following working definition of cure and failure for urge urinary incontinence outcomes was adopted:

Cure and failure for urge urinary incontinence outcomes is defined as:

For outcomes related to symptoms, cure is defined as the patient's statement that the symptom(s) is no longer present. In the case of frequency, there are seven or fewer micturitions per 24 hours. Failed treatment is defined as the patient's statement that the symptom(s) is no better or worse, with objective data from a urinary diary. As discussed above, the category of improvement cannot be specifically defined at present and requires further research to define outcomes of value to patients.

Improvement could include the patient's statement that the symptom(s) is less frequent or less troublesome with evidence from a urinary diary.

Detrusor overactivity (detrusor instability and hyperre-fiexia) is an urodynamic diagnosis characterized by involuntary phasic detrusor contractions during the filling phase of cystometry, which may be spontaneous or provoked, and which the patient cannot completely suppress. The detrusor contractions may be provoked by rapid filling, alterations of posture, coughing, walking, jumping, or other provocative manoeuvres. Outcomes for detrusor overactivity should be defined separately for symptoms, as described above, and for urodynamic findings. Cure of detrusor overactivity is defined as the absence of involuntary phasic detrusor contractions on filling cystometry. **Failure is defined as unimproved or worsened detrusor overactivity on urodynamics.** Again, the definition of improvement cannot be specified but, if used, the method of measurement should be precisely defined.

Weber, Abrams, Brubaker et al. (2001)

Definition of Faecal Incontinence

Very few articles addressing faecal incontinence define the term either generally or within the parameters of the research. Research articles specifically addressing terminology are rare. Those articles that do define their terms are more recent than those of urinary incontinence, perhaps reflecting the lengthy neglect of this area. Perhaps the development of new surgical methods to treat faecal incontinence has enabled

interventions to be more readily measured. Again, similar to the discussion of urinary incontinence terminology, it is widely recognised that the epidemiology and impact of faecal incontinence can only be accurately discerned if the problem is accurately defined (see, for example, Reilly et al. 2000).

The World Health Organisation 2nd International Consultation on Urinary Incontinence held in July, 2001 in Paris proposed the following on page 987 of their report.

"Faecal incontinence has been variously defined and there are no internationally accepted or accredited definitions available. The Royal College of Physicians has proposed "the involuntary or inappropriate passage of faeces" [1]. An international panel of experts has defined "functional faecal incontinence" as "recurrent uncontrolled passage of faecal material for at least one month, in an individual with a developmental age of at least four years" [2].

[1] Royal College of Physicians. Incontinence. Causes, management and provision of services. A Working Party of the Royal College of Physicians. *Journal of the Royal College of Physicians of London* 29, 272-274. 1995.

[2] Whitehead, W.E., Wald, A., Diamant, N.E., Enck, P., Pemberton, J.H. and Rao, S.S.C. Functional disorders of the anus and rectum. *Gut* 45, 1155-1159. 1999.

The variety of faecal incontinence definitions suggests a similar scope to those of urinary incontinence. Definitions with reference to symptom, physiology, and frequency dominate. In addition to definitions of faecal incontinence are widely used severity ratings which incorporate symptoms and frequency. The most popular of these is the "Wexner Scale". While not strictly a definition, projects have relied on the Wexner Scale to measure the impact of interventions and therapies and determining "what counts" as faecal incontinence and so is relevant to both this topic and the wider project. The scale seeks to "evaluate frequency and type of incontinence (solid, liquid, and gas and the need for pads) and its impact of lifestyle" (Yoshioka, K., Ogunbiyi and Keighley, 1999, p.253).

The Wexner Scale has been employed to measure the outcomes in a number of surgical interventions (see, for example Gee and Durdey, 1997; Sielezneff et al. 1999; and Rao & Patel, 1997). It is widely used, and appears to be largely accepted by the relevant disciplines.

Other definitions also rely on symptoms or a combination of symptoms and frequency. Dziki & Bartos (1998) measured prevalence in their study for the following categories: control of solid stool, control of pasty stool, control of liquid stool, control of wind, frequent soiling of underwear. They raise questions regarding the last category (frequent soiling of underwear) noting that hirsuteness and differences in sphincter may account for the higher prevalence of this category for men, suggesting that for some groups the category is inappropriate.

Similarly, assessment of severity also impacts on both outcome measure and definition. In a study exploring biofeedback therapy Rao et al. (2002) measure prevalence with a severity rating which also defines faecal incontinence according to "grades":

Grade 1 = anal soiling (staining of underclothes without loss of formed stool) ± incontinence to flatus
Grade 2 = grade 1 + incontinence to liquid stool
Grade 3 = grade 1 + grade 2 + incontinence to formed stool

Other efforts to define faecal incontinence using a graded system include Körsgen et al. (1997) whose system ranged from 1 (continent) to 6 (daily incontinence to solids, aware of defecation), and Hool et al. (1999) who used four categories (category A – continent to category D – no control over solid or liquid stool or flatus). None of these approaches stand out but Wexner's approach is by far the most universally used.

Cure and improvement also lack clear definitions within the literature, being largely dependent upon individual research parameters. Definitions of cure and/or success of intervention may be related to the prediction of outcomes. A study of pelvic floor retraining defined cure as a greater than 80% improvement (Rieger et al. 1997) and an operative intervention (graciloplasty) uses a decrease of 50% in incontinent episodes from baseline (Baeten et al. 2000). In contrast, Weber et al. (2001) proposed that cure be defined by a patient statement that the involuntary loss (of formed stool, liquid stool or gas) be absent within the relevant time frame.

The problem with faecal incontinence (compared to urinary leakage) is that even a small occasional loss of faeces or flatus can be associated with tell tale odour or noise. The patients are thus more frightened of the risk of leakage because any leakage provokes embarrassment and degradation. In contrast, a small occasional urine leak may not cause any smell and there is no noise component. Hence the

definition of "cure" for anal incontinence may need to be stricter. Patients may not be able to achieve a normal lifestyle if there is residual fear of anal incontinence. This issue has not been addressed in any of the literature about anal incontinence.

Definition of cure and failure for faecal incontinence outcomes

The following working definition of cure and failure for faecal incontinence outcomes was adopted:

Cure and failure for urge faecal incontinence outcomes is defined as:

In defining the impact of interventions on faecal incontinence, cure was defined as complete resolution of the symptoms. Failed treatment (persistence or recurrence) was defined as no improvement or a worsening of symptoms. As with the other pelvic floor disorders, improvement could not be specifically defined but could include some favourable change in symptoms that may be based on quality of life measures, frequency of symptoms and consistency of loss. Further research is needed to develop clinically meaningful levels of improvement after intervention.

Weber, Abrams, Brubaker et al. (2001)

Chapter 2   Psychometric properties and the selection and development of incontinence outcome measures

Psychometrics is the discipline of measurement, instrument construction and validation. This section of the report is a review of basic psychometric issues and their relevance to the selection and development of incontinence outcome measures.

Measurement tool development protocols

In developing measurement tools, it is recommended that specified and systematic protocols be used. The methodology that is advocated for the development and validation of measurement tools in this document draws upon the principles and protocols proposed by De Vellis (2003). There are others, for example like Streiner & Norman (1995) but the framework outlined below is by De Vellis.

De Vellis (2003) advocates an eight-step process of scale development. These are:
- "Determine clearly what it is you want to measure"
- "Generate an item pool"
- "Determine the format for measurement"
- "Have initial item pool reviewed by experts"
- "Consider inclusion of validation items"
- "Administer the items to a development sample"
- "Evaluate the items"
- "Optimise scale length"

1. "Determine clearly what it is you want to measure"

Prior to the development of any tools, it is necessary to clarify exactly what it is that is going to be measured, including the different aspects (domains or dimensions) and what the purpose of that measurement is.

2. "Generate an item pool"

The generation of an item pool can occur in a variety of ways. A good way is to derive each item directly from the statement of what the tool is intending to measure. This process can also be very usefully informed by analysis of the literature surrounding the target concepts. This review process may also reveal other similar tools, that are similar to the new tool under construction but that are not exactly what is desired from a content or methodological viewpoint. Another particularly effective way of developing items can be to do a content analysis of several related tools to identify the full range of the items included in similar tools and then use these content guides for the new tool.

3. "Determine the format for measurement"

An important design decision is the scale format of the items in the tool under development. As discussed in Polgar and Thomas (2004), there are virtues associated with different types of response categories. The types of scale formats that are usually considered for use with measurement tool items are:
- Guttman
- Likert
- Multiple response
- Semantic differential
- Numerical rating scale (visual analog scale)

4. "Have initial item pool reviewed by experts"

Review of the item pool that has been generated using the previous steps in the process is a particularly useful procedure.

5. "Consider inclusion of validation items"

In De Vellis' exposition, the inclusion of validation items may be achieved in several different ways. Some psychological tests such as the MMPI have truth sub-scales where items that are demonstrably true or false are scored to form a measure of the truth of the responses. In many circumstances, there

is no gold standard measure of the attributes in question. This is why the test development process has occurred in the first case.

6. "Administer the items to a development sample"

It should be accepted that the first run of the scale is a development exercise. No matter how carefully the earlier steps have been implemented, there are statistical issues and logistic issues that cannot always be predicted at the design stage. It is useful to pilot the new instrument with a small group of people drawn from the target population. The procedure is to administer it and then to interview participants about ease of completion, understanding or confusion about items and so on.

7. "Evaluate the items"

Once the measurement tool has been administered to a sample of respondents, there is a wide range of statistical techniques available to analyse the psychometric properties of the items of a scale and the scale overall. These are described below:

Item difficulty. Item difficulty is the proportion of respondents who "pass" a particular item. Generally, item difficulty refers to the responsiveness of the item response scale, which is its ability to discriminate between those cases with the condition of interest and those without the condition.

Item scale correlations. In many instances, it is possible to derive a total score from the scale, through combining the numerical values assigned to the scale points on each item. If the score of each item is correlated with this total score, then this gives an indication of the contribution of the item to the variability in the overall scale. Low item-scale total correlations are undesirable. An axiom of psychometric measurement is that all items contributing to a scale should "belong" to that scale. This is usually demonstrated through factor analysis (see below).

Coefficient alpha. Coefficient alpha or Cronbach's alpha is a measure of the homogeneity of the scale items. Generally, for scales to be reliable for group measurement, the literature recommends that Cronbach's alpha should be within the range 0.70 – 0.90. For reliability of clinical diagnosis of individual cases it is often recommended that alpha should exceed 0.90. Alphas within the conventional range can be obtained either through the inclusion of many items in a scale or by including only items that are highly correlated.

Criterion related validity correlations. Each item score is correlated with a criterion variable. High correlations are generally desirable.

Factor analysis. Factor analysis is a means of exploring the internal structure of multi item scales. First the correlation matrix of all items with all other items is constructed. This is then subjected to examination to identify whether there are sets of items that are more highly correlated with each other than with other items. The assumption is that there are latent variables underlying the items that are partly measured by them. For example, it might be hoped that within a particular scale that there is a single latent variable measured by all the items. Factor analysis reveals how many latent variables or factors (referred to as vectors) are present in a set of items. Thus factor analysis is widely used to study the internal structure of scales. The factor structure may then be studied to see which items "load" on which vectors. Generally it is acceptable that a scale should comprise only those items loading on the scale vector.

8. "Optimise scale length"

The shortening of a scale to optimal length is achieved by deletion of poorly performing or redundant items using the outcomes of the various item analyses described above.

The purpose of the use of standardiased scale development protocols is to maximise the chances of obtaining desirable properties in the developed tools. The next section includes a discussion of these properties.

Desirable psychometric properties in incontinence measures

It is generally acknowledged that tests need to have the following intrinsic properties:
- Reliability
- Validity
- Sensitivity
- Applicability
- Practicability

Reliability refers to the capacity for a test or measure to give the same result consistently, *ceteris peribus.* There are several different ways of measuring reliability including test-retest reliability (where the same test is administered on two occasions to the same participants by the same assessors usually at 2-week intervals to control for memory and history confounders), inter-rater reliability (where the same participants are assessed by different assessors and the results compared) and internal consistency where test item values are compared with each other. Obviously if a test is administered on two occasions close in time, it is desirable for the same results to be obtained for the same individuals. Similarly, it would be unacceptable if different assessors arrived at widely discrepant results for the same individuals while using the same test. The basic, widely reported form of reliability is internal consistency. For multi-purpose scales this is generally reported in the form of Cronbach's alpha.

Regarding test-retest, the degree of correlation between repeated administrations has traditionally been represented by the Pearson correlation coefficient although there has been a move towards the use of the Intra Class Correlation (ICC) in recent times. The reason for this is that the Pearson correlation assures linearity between the items over time, in contrast the ICC takes variability within groups and over time into account. There is no specific standard for evaluation of reliability coefficients although values of 0.8 would be considered to be robust and 0.9 would be considered high.

Validity refers to an instrument measuring what it purports to measure. In clinical contexts, validity is frequently measured by reference to specificity and sensitivity, referring to a test's capacity to correctly detect those with the target attribute and to correctly filter out those who do not have it. It is also assessed through correlations with other measures which are now to measure the construct of interest (concurrent validity).

Applicability refers to the ability to the test to be applied to a specific target group. Tests work best within prescribed groups and may work less well when used in other groups. For example, a test developed for women may not work well for men.

Practicability refers to the ability of a test to be practically applied in given contexts. For example, if the test requires several hours to be administered or expensive equipment then it may not be suitable for a clinical context.

Sensitivity and specificity of tests and measures

The concepts of sensitivity and specificity apply to diagnostic or screening tests where the goal of the test is to determine whether the person has the target condition/need or not. There are essentially four situations that can arise with test results as follows:

True positive   The test is positive and the person actually has the condition/need

False positive   The test is positive but the person does not actually have the condition/need ("false alarm")

True negative   The test is negative and the person actually does not have the condition/need

False negative   The test is negative and the person actually has the condition/need ("miss")

These outcomes can be presented in a table as shown below:

Table 12   Determining the sensitivity and specificity of test and measures

Test results	True Situation		Totals
	Person has condition (+)	Person does not have condition (–)	
Positive (+)	True positive (a)	False positive (b)	a + b
Negative (–)	False negative (c)	True negative (d)	c + d
Totals	a + c	b + d	a + b + c + d

Sensitivity is the proportion of people who really have the condition who are detected as such by the test.

> Using the table above,
> Sensitivity = a/(a+ c)

Specificity is the proportion of people without the condition who are detected as such by the test.

> Using the table above,
> Specificity = b/(b+d)

Sensitivity and specificity provide important information as to how to interpret a test result and the confidence that should be placed on a positive and a negative test result. Given that important actions can be taken based upon a clinical test result, test designers strive for high levels of sensitivity and specificity. For example, a false positive test result could result in high levels of needless intervention whereas a false negative result could result in life threatening risks through delayed or no action taken to address a serious health problem. The types of classification or decision errors made in diagnostic tests can generally be adjusted by changing the cut-off or decision points on the score continuum. But such adjustments to the cut-off score while decreasing one type of error consequently increase the other.

Conclusions concerning psychometric properties and incontinence outcome measures

As a general principle, tests should not be used without an educated and evidence based view of their accuracy.

How were the tests and measures rated in this study?

As previously discussed, the World Health Organisation in association with the International Continence Society hosted two international consultations on incontinence in 1998 and 2001 in which important determinations were made concerning outcomes measurement. These important international works have informed the reviews collected. However, although the studies represent major achievements in the field, surprisingly, the basis for the recommendations made by these studies was not as clearly documented as was desirable. Recommendations were made without systematically specifying what criteria employed, and how instruments were scored and weighted. In this study we have used a more systematic procedure using an evidence-based weighted scoring procedure. This was implemented through the use of an expert panel consisting of measurement methodologists. The expert panel consisting of Jan Sansoni, Graeme Hawthorne and Shane Thomas met to discuss the criteria and associated scoring and weighting system. The weights assigned to each criterion were assigned on the basis of the extent to which they were considered by the panel to be pertinent to the Australian incontinence outcomes measurement context. The scoring system used was as follows:

Table 13   Criteria and weights used to assess instruments

Evaluation Criteria	Scoring system	Weight
Availability of comparison data	1 = not widely used in Australian and international incontinence research/clinical settings, 2 = some use in Australian and international incontinence research/clinical settings, 3 = wide use in Australian and international incontinence research/clinical settings	3
Length, ease and time to complete	1 = long instrument, 2 = medium length instrument, 3 = short instrument	2
Method of administration	1 = interviewer required, 2 = self-completion	2
Translations readily available	1 = available only in English, 2 = some translations available, 3 = many translations available	1
Ease of scoring	1 = scoring complex, 2 = scoring reasonably straightforward, 3 = scoring easy with computer code available	2
Sensitivity to incontinence	1 = not known to be sensitive, 2 = sensitive to incontinence status, 3 = good sensitivity to incontinence status	3

Reliability evidence available	1 = no or little published evidence identified, 2 = evidence suggests moderate reliability, 3 = evidence suggests good reliability	3
Validity evidence available	1 = no published validity evidence identified, 2 = evidence suggests moderate validity, 3 = evidence suggests good validity	3
Adherence to psychometric axioms	1 = does not meet basic axioms, 2 = partially meets basic axioms, 3 = mostly meets basic axioms	3
Cost of using the instrument	1 = costs charged for using instrument, 2 = costs charged for commercial use, 3 = instrument available free of charges	2

In all tables where the term "weighted total" appears this refers to the weighting scheme described in the above table.

Sansoni, Hawthorne and Thomas developed the weighting and scoring scheme for the criteria used in the technical evaluations of the measurement tools. The weighted total scores for each instrument were calculated by assigning a score for each criterion based on the expert panel judgments, multiplying the score by the assigned weight and then summing the scores over all criteria. In all cases, a higher score indicated the instrument might provide better measurement. A cut-off score at the mid-point of the scoring range, 47 points (the maximum possible score was 70 points), was selected as the level at or above which tools could be recommended for use. Tools with total weighted scores below this level were not recommended, as they were considered not to have sufficiently strong psychometric and associated properties.

The top tools were selected for review, provided that they scored 47 points or above. Tools for which there was no adequate published research were not included in the analysis.

Issues in the validity of the ratings

It is important to note that the ratings used to determine recommendations are consensus evidence based panel judgments of the selected tools on the relevant criteria. Thus it is not a matter of our judgments being valid or invalid but whether they are reasonable given the available evidence. A different team may arrive at different ratings because of the following:

1. Use of a different scoring and weighting scheme

The weighting scheme reflects the panel's views of what criteria should be used in the assessment of the measurement tools. While these criteria are not particularly controversial, different weights may be assigned by different raters. The ratings used here have been made explicit and transparent.

2. Disagreement over the criteria used

The criteria employed in the panel ratings were based on panellists' consideration of the important criteria applicable to instrument review and selection. They closely relate to the four criteria (reliability, validity, applicability, practicability) outlined in the review of psychometric qualities.

3. Differencess over interpretation of the evidence to arrive at the ratings

Consensus ratings were used and there was no rating on which the raters did not agree. The ratings were made on the best available literature at the time. Of course, the research evidence supporting various tools is dynamic and also in many cases currently insufficient.

Thus while there are legitimate reasons for disagreement with these recommendations in terms of advocating different criteria, the scoring protocols are clear and transparent.

Chapter 3   Review of the incontinence outcome measures

This chapter reviews the evidence available concerning the range of tools and measures that were considered in our study. The types of measures reviewed include:

- Faecal Incontinence Symptom Measures
- Faecal Incontinence Health-Related Quality of Life Measures
- Urinary Incontinence Symptom Measures
- Urinary Incontinence Health-Related Quality of Life Measures
- Pad Tests
- Frequency Volume Charts and Bladder Diaries
- Multi-Attribute Utility Instruments
- Generic Health-Related Quality of Life Measures (General Health Status)
- Functional Measures

The main body of the review is included in the body of the report and additional information is also provided in the selected instrument reviews that are appended to this report.

Faecal Incontinence Outcome Measures

The need for valid, reliable, applicable and practical measures of faecal incontinence symptoms and intervention outcomes is widely acknowledged within the clinical research literature. Notwithstanding the recognition of this need, many commentators argue that the current state of knowledge in this area is such that specific recommendations for the adoption of particular tools are not currently feasible. The specialist group charged with the task of providing advice concerning anal incontinence at the Second International Consultation held under the auspices of the ICS/WHO reached the following conclusions:

> "At present there are no widely utilized validated anal incontinence specific bowel symptoms questionnaires. The Faecal Incontinence Questionnaire (FIQ) although lengthy may be a useful instrument in the initial assessment of the patient with anal incontinence." (Abrams et al. 2002a)

Similarly, Nancy Norton, the President of the International Foundation for Functional Gastrointestinal Disorders in her November 2002 address at the IFFGD Symposium made the following observations:

> "There are numerous scoring systems for the severity of faecal incontinence – Wexner, St Marks, Pescatori, Rockwood, among others … and there are numerous scoring systems for quality of life of these patients … If research cannot determine which among these different sets of scales best serve patients and clinicians alike then professional organisations like the American Society of Colon & Rectal Surgeons should insist on a standard adoption across institutions of a single severity scale on a single quality of life scale." (2002)

Other commentators, such as Byrne et al. (2002) have made similar observations:

> "Many incontinence scores have been devised but no universal quantitative measure has been adopted." (2002)

Notwithstanding this rather negative assessment by these commentators there are several anal incontinence symptom grading and scoring systems that are in widespread use. These include the Pescatori, Wexner, American Medical Systems and St Marks Hospital faecal incontinence measures. In addition, there are some anal incontinence specific quality of life measures and symptom measures that may show promise but that will require future development. These include the Rockwood Faecal Incontinence Severity Index, the Eyepasch Gastrointestinal Quality of Life Index and the Byrne et al. scale based upon the Direct Questioning of Objectives framework, amongst others. Firstly, the major contenders for the measurement of faecal incontinence symptoms are examined.

Faecal Incontinence Symptom Measures

Vaizey, Carapeti, Cahill & Kamm (1999) reported the outcomes of an important paper in which they compared the psychometric performance of the Pescatori score, the Wexner score[1], the American Medical Systems score and their newly developed scale at St Marks Hospital in the United Kingdom.

---

[1] Note that Wexner has started calling his measure the Cleveland Clinic Florida Faecal Incontinence Score but this has not achieved wide usage as yet. For the purposes of this review we have adhered to the Wexner nomenclatur to maximise recognition and understanding of our recommendations.

Unfortunately, as with many other such studies, the numbers of participants in the Vaizey study was very small (n = 33 patients). Notwithstanding these small study numbers, the study included a detailed analysis of the concurrent validity of the tools with each other as well as independent expert judgment concerning the status of each of the patients involved in the study. It is also useful to note that the study was prospective in nature and hence allowed for the assessment of test retest reliability for each of the four scales. The test retest reliabilities as measured by Pearson's correlation coefficient obtained in the study sample for the four scales are shown in the Table 14.

Table 14   Test-retest reliabilities of faecal incontinence measures

	Test Retest Reliability
St Marks	0.87
Pescatori	0.58
Wexner	0.75
American Medical Systems	0.84

It is not specifically reported in the Vaizey et al. (1999) paper whether the differences between the obtained reliabilities are statistically significant but given the small number of participants in the Vaizey study it would be unusual if this were found to be the case. Nevertheless, as the proportion of variance explained is a function of the square of the correlations, this means that the St Marks measure has 76% of shared variance between the two administrations compared to 34% of shared variance for the Pescatori measure across the two administrations. This is a robust difference. The differences between the St Marks measure and the Wexner and American Medical Systems measures (56% and 71%) are less remarkable but if effects of this magnitude were preserved in large studies they would still be clinically useful.

The contents and the scoring systems for each of the scales are shown in the tables below.

Table 15 shows the composition of the Pescatori et al. (1992) score. The Pescatori index was initially reported in Pescatori, Anastasio, Bottini and Mentasi (1992). As can be seen from the table, essentially the Pescatori et al. score is a weighted combination where, for each of the three items, a weighted score is applied according to the frequency with which the symptoms occur for the particular patient, thus the anal incontinence score is the sum of the anal incontinence degree and the anal incontinence frequency for each of the areas.

The Vaizey et al. paper is the most convincing psychometric concurrent reliability and validation study. A test-retest reliability of 0.58 is towards the lower end of what would normally be considered to be psychometrically acceptable.

Table 15   The Pescatori et al. score

A	Incontinence for flatus/mucous	Less than once a week	1
		At least once a week	2
		Every day	3
B	Incontinence for liquid stool	Less than once a week	1
		At least once a week	2
		Every day	3
C	Incontinence for solid stool	Less than once a week	1
		At least once a week	2
		Every day	3

Anal Incontinence degree	Points	Anal Incontinence frequency	Points	Anal Incontinence score
A	1	1	1	2
A	1	2	2	3
A	1	3	3	4
B	2	1	1	3
B	2	2	2	4
B	2	3	3	5
C	3	1	1	4
C	3	2	2	5
C	3	3	3	6

Pescatori et al. Anal Incontinence score = Anal Incontinence Degree + Anal Incontinence Frequency

Table 16 shows the structure of the Wexner score. The score is derived from a frequency rating of the type of incontinence with the addition of whether the patient has to wear a pad and whether the patient's lifestyle has been altered by the incontinence. In psychometric terms it is unconventional to include measure of the frequency of occurrence of symptoms combined with measures of the frequency of the impacts of those symptoms in the one index. There are two reasons for this. First, it is an axiom of measurement that items with different conceptual bases should not be combined in the one score, and, second the wearing of a pad and the effect on a person's lifestyle are both functions of frequency of the condition, thereby leading to double-counting where these are included in the score.

It is also noted that the Wexner Scale does not address the issue of faecal urgency. The Wexner Scale and some items concerning faecal urgency have been included in the autumn 2004 South Australian Health Omnibus Survey (Harrison Health Research, 2004) and following psychometric analysis of these items it may be possible to further refine the Wexner Scale to address these issues.

Subject to these caveats, the Wexner score is very simple, straightforward and intuitively appealing since it is amenable to completion by both clinicians and patients. Perhaps it is the case that simplicity of ease of use and design will always win in terms of user adoption over psychometric purity. Jorge and Wexner's (1993) paper outlined the structure of the measurement tool. Under the scoring system a score of 0 means no incontinence nor impact and a score of 20 means the worst possible incontinence and impact. The Vaizey et al. study's finding of a test-retest reliability value of 0.75 for the Wexner score in their study sample is within the conventional range.

Table 16   The Wexner Faecal Incontinence Score

Type of incontinence	Frequency				
	Never 0	Rarely < 1/month	Sometimes < 1/week ≥ 1/month	Usually < 1/day ≥ 1/week	Always ≥ 1/day
Solid	0	1	2	3	4
Liquid	0	1	2	3	4
Gas	0	1	2	3	4
Requires pad	0	1	2	3	4
Lifestyle	0	1	2	3	4

Rarely = less than once per month; sometimes = between once per week and once per month; usually = between once per day and once per week; always = at least once per day

Table 17 shows the scoring system associated with the American Medical Systems Faecal Incontinence Scale. It is also a weighted composite of reports of various types of faecal incontinence symptoms.

Table 17   The American Medical Systems Faecal Incontinence Score

Over the past four weeks, how often:	*Never*	*Rarely*	*Sometimes*	*Weekly*	*Daily*	*Several times daily*
Did you experience accidental bowel leakage of gas?	0	1	2	3	4	5
Did you experience minor bowel soiling or seepage?	0	1	2	3	4	5
Did you experience significant accidental bowel leakage or liquid stool?	0	1	2	3	4	5
Did you experience significant accidental bowel leakage and solid stool?	0	1	2	3	4	5
Has this accidental leakage affected your lifestyle?	0	1	2	3	4	5

The Vaizey et al. finding of a test-retest reliability value of .84 for the American Medical Systems faecal incontinence scale is good. However the weighting method for the scoring system does not have a well developed rationale in the documentation that has been accessed.

Table 18 shows the incontinence score developed by Vaizey et al. at St Marks Hospital.

Table 18   The faecal incontinence score by Vaizey et al. at St Marks

	Never	*Rarely*	*Sometimes*	*Weekly*	*Daily*
Incontinence for solid stool	0	1	2	3	4
Incontinence for liquid stool	0	1	2	3	4
Incontinence for gas	0	1	2	3	4
Alteration in lifestyle	0	1	2	3	4
				No	*Yes*
Need to wear a pad or plug				0	2
Taking constipating medicines				0	2
Lack of ability to defer defecation for 15 minutes				0	4

The total is reached by adding one score from each row: minimum score = 0 perfect continence; maximum score = 20 = totally incontinent.

At this point there is little in psychometric terms to separate the various measures of Pescatori, Wexner, American Medical Systems and St Marks Faecal Incontinence Measures, although the Pescatori index stands out from the others in the Vaizey et al. study in terms of having significantly lower test-retest reliability. These scoring systems are the most widely used in the relatively few studies of faecal incontinence that have used systematic measurement tools, but, there remains much work to do in terms of the psychometric development of these tools. This is reflected in the ICS/WHO sponsored Anal Incontinence Committees' recommendations for the non-adoption of any specific measure over others as a result of the lack of convincing evidence for the psychometric superiority of any of the tools over others. Essentially the project team concurs with the Committee's finding in the recommendations, because of the widespread use of the Wexner and its acceptability amongst clinicians. Its interim adoption is recommended pending the broadening of the evidence base as to the psychometric properties of the various tools. Table 19 shows the expert ratings of the tools assessed in this category supporting the selection of the Wexner.

Table 19   Summary of ratings for faecal incontinence symptom measures

Criteria	Tool			
	Wexner	Pescatori	St Marks	AMS
Availability of comparison data/usage	3	2	1	1
Length, ease and time to complete	3	3	3	3
Method of administration	2	2	2	2
Translations available	3	2	2	1
Ease of scoring	3	3	2	2
Sensitivity to incontinence	1	1	1	1
Reliability evidence available	2	1	2	2
Validity evidence available	2	1	1	1
Adherence to psychometric axioms	2	2	2	2
Cost of using the instrument	3	3	3	3
Weighted Total	**55**	**45**	**43**	**42**

Wexner   Wexner/Cleveland Clinic Faecal Incontinence Score
Pescatori   Pescatori Faecal Incontinence Symptom Severity Measure
St Marks   St Marks Faecal Incontinence Grading System
AMS   American Medical Systems Faecal Incontinence Scale

The Wexner/Cleveland Clinic Faecal Incontinence Score was the only measurement tool to reach the required 47-point score in this category. Thus the only recommended faecal incontinence symptom severity measure is the Wexner. It is noted that all measures save for the St Mark's Grading System do not address the issue of faecal urgency. The Wexner Scale and some items concerning faecal urgency have been included in the autumn 2004 South Australian Health Omnibus Survey (Harrison Health Research, 2004) and following psychometric analysis of these items it may be possible to further refine the Wexner Scale to address this issue.

Faecal Incontinence Health-Related Quality of Life Measures

The review of faecal incontinence health-related quality of life measures showed that there has been little co-coordinated effort in the development of outcome measurement tools to date. The literature is characterized by a number of single studies, or perhaps sequences of two or three studies, where the tool has been developed and used and then fallen into disuse.

The SF-36 has been used in a range of studies to investigate the impact of faecal incontinence upon health status and some of these reviews are presented in the section of this report dealing with general health status measures. There is reasonable evidence that the SF-36 is useful in describing people with faecal incontinence and as in many clinical conditions should be considered for inclusion as an outcome measurement tool. It is considered that there is little advantage to using condition specific health-related quality of life measures where a generic measure could be used. The advantages of generic tools include the very important ability to benchmark with other groups not to mention avoiding yet another expensive test development exercise.

In terms of condition specific measures of health-related quality of life associated with faecal incontinence, Rockwood et al. have reported several relevant studies in their 1999 paper. They reported the development of a faecal incontinence severity index involving patient and surgeon ranking of the severity of symptoms associated with faecal incontinence. The scale used the familiar frequency matrix as used in the Wexner, American Medical Systems and Pescatori scales.

The second paper published by Rockwood's group in 2000 was based upon the same sample of 118 patients reported in the earlier paper and reports the early stages of the development of the Faecal Incontinence Quality of Life Scale. This is a self-administered questionnaire consisting of 29 items. The scale consists of four sub-scales including Embarrassment, Lifestyle, Coping and Depression scored on a 5 point Likert style rating scale. However there is little psychometric data reported in either of the publications and the scoring method is still under development. More recently Byrne et al. (2002) have pointed out that although Rockwood's work is promising they consider that there are flaws in the use of data from clinicians' ratings in the development of weighting schemes for patient self-report measures.

Reilly, Talley, Pimberton and Zinsmeister (2000) reported the outcomes of a validation study for the Faecal Incontinence Questionnaire. Ninety-four adult patients took part in this study. Thirty-four patients repeated the questionnaire via a mail-out and 41 were re-tested over the telephone. The Kappa correlation co-efficient was used to assess test-retest reliability, and K = 0.68 for the mail-out questionnaire. The data reported by the researchers concerning this new scale indicate acceptable levels of reliability and validity, although the reliability coefficients are towards the lower end of what would normally be considered to be psychometrically acceptable. This tool, like the many others available within the current published research literature, is at an early stage of development. It will require further development and the conduct of a range of reliability and validation studies in order to determine its ultimate acceptability as a viable conditions specific measure of faecal incontinence symptoms and quality of life.

Osterberg, Graph, Karlbom and Pahlman (1996) reported an evaluation study of a self-administered questionnaire in the area of faecal incontinence and constipation. The study sample consisted of 16 control subjects, 36 patients with faecal incontinence and 38 with constipation. Overall, test-retest reliability was found to be 0.57 for the faecal incontinence group, 0.60 for the constipation group and 0.95 for the controls. It is likely that the variability in the test-retest reliability across the different groups accurately reflects the variability of their conditions. Thus controls that are unlikely to have high changeability in "symptoms" over time have been found to have highly reliable scores, whereas the clinical patients from the two clinical groups demonstrated high variability. Once again, while this measurement tool shows some promise, it will require considerably more development through the conduct of a range of validation and reliability studies.

Bug, Kiff and Hosker (2001) have reported a validation study for a new condition specific health-related quality of life questionnaire for the assessment of women with faecal incontinence in the United Kingdom. A postal survey approach was used in the surveying of 220 women who had diagnosed anal incontinence. The new tool, entitled the Manchester Health Questionnaire, was derived from the Kings Health Questionnaire. This is a condition specific health-related quality of life questionnaire for women with faecal incontinence. The psychometric data reported in this one-off study appear to demonstrate acceptable levels of internal inconsistency and test-retest reliability but one study is not sufficient basis to rate the tool with respect to the full range of evaluation criteria.

When further studies are published this tool will be revisited. However, as with other measurement tools further work and research will be needed to demonstrate the robustness of the tool across different groups and settings.

Two other papers have been published: neither, however, report sufficient detail for review of the measures. Eypasch, Williams, Wood-Dauphinee, Ure, Schmulling, Neugebaurer & Troidl (1995) reported basic psychometric properties of the Gastrointestinal Quality of Life Index in a series of staged validation studies. The scale consists of 36 questions with a Likert type five point rating scale summed over the questions.

Byrne, Pager, Rex, Roberts & Solomon (2002) have recently reported their own contribution to the development of quality of life measurement tools for patients with faecal incontinence. This group has used the Direct Questioning of Objectives methodology to develop their own measures. Unfortunately, this paper does not provide a sufficiently detailed report of the basic psychometric properties of the tool.

The following table shows the expert panel's ratings of the Faecal Incontinence Quality of Life Scale and the Gastro Intestinal Quality of Life Index, these being the scales for which there was sufficient research evidence to complete the ratings task.

Table 20   Summary table of ratings for faecal incontinence health related quality of life measures

	Tool	
Criteria	**FIQLS**	**GIQLI**
Availability of comparison data/usage	1	1
Length, ease and time to complete	1	1
Method of administration	2	2
Translations available	1	1
Ease of scoring	2	2
Sensitivity to incontinence	1	1
Reliability evidence available	2	1
Validity evidence available	1	1
Adherence to psychometric axioms	2	2
Cost of using the instrument	3	3
Weighted Total	**38**	**35**

FIQLS   Faecal Incontinence Quality of Life Scale
GIQLI   Gastro Intestinal Quality of Life Index

The available Faecal Incontinence Quality of Life measures are in an early stage of development. No tools reached the required 47-point cut-off score in this category and hence none are recommended.

Conclusions and recommendations for Faecal Incontinence Symptom Severity and Condition Specific Health-Related Quality of Life Measures

The development of measurement tools for the measurement of faecal incontinence symptoms is at an early stage in psychometric terms. Although measures to grade severity of symptoms associated with faecal incontinence such as the Wexner, Pescatori and St Marks scores are in reasonably wide use there remains a dearth of convincing studies in which the psychometric properties of the tools have been rigorously examined. The most convincing study to date has been the study performed by Vaizey et al. (1999) at St Marks Hospital. However, as noted in the review of this study the numbers of participants in the study was small, some 33 participants. While acknowledging the difficulty of performing studies with large-scale clinical groups the study size falls well short of the numbers required to provide a

credible psychometric analysis. Thus, while these tools enjoy a reasonable degree of use in clinical practice and some research studies, the research evidence for sound psychometric properties for these tools remains to be provided.

In terms of health-related quality of life measures for faecal incontinence, it is even earlier days in terms of progress in the field. The Rockwood et al. studies concerning the Faecal Incontinence Quality of Life Scale have attracted considerable attention and interest in the research literature and according to the WHO review show "promise". However, given the state of the published evidence in support of this tool, much further work remains if this prediction is to be achieved. The Byrne et al. studies (2002) in Sydney are also in the early stages of development of such tools and it might be useful to revisit these tools later, but at the present stage of development they cannot be recommended.

Therefore, at this point in the development of both symptom severity and quality of life specific to the condition of faecal incontinence, it is not possible to recommend a single measure or group measures over and above others on the basis of psychometric properties. From the viewpoint of simplicity of use and extent of use in symptom severity measurement, the Wexner scoring system for faecal incontinence probably has the advantage. However, once again, in the absence of proper psychometric data to support this choice, this would be a very tentative recommendation. The Wexner system would appear to have multiple conceptual and content dimensions combined together in the one index but in some contexts this has been nominated as an advantage. The consensus conference for treatment options for faecal incontinence held in October 2002 at Saint Vincent noted that 11 out of the 18 participants used the Wexner score with the next most frequent tool being used by only 2 participants.

The consensus group explained this situation of strong support by the Wexner by stating:

> "Because it is practical and easy to use and interpret. It was the first incontinence score system that takes into account the use of pads and lifestyle alteration as well as the frequency of incontinence".

Given the paucity of psychometric data in the faecal incontinence symptom and outcome measurement domain, it is not possible to apply a systematic tool selection framework such as employed in the review of multi-attribute utility measures reported in this study (see page 72). It is agreed with the recommendations made by the ICS/WHO Second Incontinence Consultation group that it is not possible to definitively recommend a condition specific quality of life measure for faecal incontinence as yet. The project team is of the view that it may be better to assess health-related quality of life by the use of a generic health-related quality of life measure which also allows comparison across conditions rather than using condition specific health-related quality of life measures. These measures tend to be all-in-one measures that combine symptoms, their severity, quality of life items and their impact. As a result, elements within these measures can be subject to a double counting of elements and to weighting systems that are not transparent. However, a provisional recommendation of the Wexner symptom scoring system for measurement of faecal incontinence symptoms for primary care practitioners, specialist practitioners and researchers has been made.

The recommendations for outcome measures in faecal incontinence based upon the literature analysis are as follows:

Table 21   Recommendations for outcome measures in faecal incontinence

User category	Recommended Symptom Measures	Recommended Condition Specific Health-Related Quality of Life Measures
Primary Care Practitioner	Wexner Symptom Scoring System	None
Specialist Practitioner	Wexner Symptom Scoring System	None
Researcher	Wexner Symptom Scoring System	None

Urinary Incontinence Outcome Measures

Urinary incontinence (as well as faecal incontinence) can occur in a variety of client groups: the aged and frail elderly, pregnant women, women post-partum, people undergoing cancer treatment, and others with dementia or other neurological problems. Yet it was not until recently, with the 2nd International Consultation on Incontinence in 2001, that uniform standards were developed for the measurement of incontinence conditions. For more information, readers are referred to the discussion about this important work on pages 28-30, as well as to the section on defining urinary incontinence on pages 33-37.

Urinary incontinence symptoms may be measured by a variety of means. These include symptom inventories and health-related quality of life measures (patient self-report, health status measures) as well as other means including pad tests, frequency and volume charts and bladder diaries. Symptom inventories are reviewed first, followed by the other methods.

Urinary Incontinence Symptom Measures

In their review of urinary incontinence symptom measure, the ICS 2nd International Consultation on Incontinence group (see Abrams et al. 2002a) charged with this task "highly recommended" the following symptom inventory measurement tools:

- Urogenital Distress Inventory (including the UDI 6 and the Urge UDI)
- Kings Health Questionnaire
- Incontinence Severity Index (for women only)
- DAN-PSS-1 (for men only)
- ICS Male (for men only)
- ICS Male SF (for men only)

The Bristol Lower Urinary Tract symptoms and the Symptom Severity Index were awarded a "recommended" grading.

The Kings Health Questionnaire

The Kings Health Questionnaire (KHQ) was developed by Kelleher and Cardozo's group at King's College Hospital in the UK (see Kelleher, Cardozo, Khullar & Salvatore, 1997). It has 21 items and was designed to measure symptoms and quality of life for males and females with urinary incontinence. The 21 items are allocated across three parts: Part 1 (general health perception, incontinence impact), Part 2 (role limitations, physical limitations, social limitations, personal relationships, emotions, and severity measures), Part 3 (frequency, nocturia, urgency, urge, stress, nocturnal enuresis, intercourse incontinence, infections, pain, and difficulty in voiding). A four point rating system is used and these yield 8 subscales ("domains") scored between 0 and 100, with 100 indicating greater impact upon health-related quality of life.

The KHQ has been used in a variety of settings in many countries including most European countries, the United States and Canada and in some Asian countries. For example, Kelleher, Reese, Pheil & Okano (2001), used the King's Health Questionnaire and the SF-36 to assess outcomes for patients with over-active bladder receiving immediate release from Tolterodine. The two questionnaires were administered at base line and also at the end of treatment twelve weeks later in an RCT intervention effectiveness study. The KHQ scores showed significant improvement. Interestingly, the SF-36 did not detect any changes in the study sample.

The KHQ has been subjected to psychometric study in a wide variety of settings. For example in Barcelona, Badia Llach et al. (2000) performed a psychometric study with women with urinary stress incontinence (n = 77), urge incontinence (n = 51) and mixed incontinence (n = 34). Cronbach's alpha measures of internal consistency ranged from 0.68 to 0.88 and test-retest reliability at 15 and 30 days ranged from 0.65 to 0.92. Other studies (Kelleher, Cardozo & Toozs-Hobson, 1995; and Kobelt, Kirchberger & Malone-Lee, 1999) reported similarly positive results for the psychometric robustness of the tool.

Several recent studies reported in the proceedings of the 2002 International Continence Society Congress held in Heidelberg describe studies that have assessed the validity and reliability of the KHQ. For example, Leung et al. (2002) report a validation study of Chinese women suffering from urinary incontinence. The KHQ and the Chinese version of the SF-36 were administered, along with a one-week urinary diary. As in the other studies of the psychometric properties of the KHQ, high internal consistency was found and moderate degrees of association between the SF-36 and the KHQ were reported. The researchers reported modest correlations between the KHQ and urinary diary findings.

The Urogenital Distress Inventory/Incontinence Impact Questionnaire and their derivatives

The Urogenital Distress Inventory (UDI) and the Incontinence Impact Questionnaire (IIQ) were developed by Shumaker, Wyman, Uebersax, McClish & Fantl (1994) to assess the impact of urinary incontinence symptoms upon quality of life for women. The original forms of the IIQ and the UDI had 30 and 19 items respectively but work by Uebersax, Wyman, Shumaker, McClish and Fantl (1995) created a 7 item version of the IIQ and a 6 item version of the UDI and these are now used widely in both clinical and research applications. The UDI is a symptom tool and the IIQ is a tool designed to measure the impact of the symptoms. The UDI long and short forms are reviewed below as it best fits within the Symptom Inventory focus of this section. However, it should be noted, that the IIQ provides an important additional dimension to the UDI as it assesses the impact of these symptoms.

Long Form of UDI

The UDI items ask the respondent "Do you experience and if so how much are you bothered by" which is followed by a list of symptoms:

- Frequent urination?
- A strong feeling of urgency to empty your bladder?
- Urine leakage related to the feeling of urgency?
- Urine leakage related to physical activity coughing or sneezing?
- General urine leakage not related to urgency or activity?
- Small amounts of urine leakage (drops)?
- Large amounts of urine leakage?
- Night-time urination?
- Bedwetting?
- Difficulty emptying your bladder?
- A feeling of incomplete bladder emptying?
- Lower abdominal pressure?
- Pain when urinating?
- Pain or discomfort in the lower abdominal or genital area?
- Heaviness or dullness in the pelvic area?
- A feeling of bulging or protrusion in the vaginal area?
- Bulging or protrusion you can see in the vaginal area?
- Pelvic discomfort when standing or physically exerting yourself?
- Having to push on the vaginal walls to have a bowel movement?

The response scale is:

- Not at all
- Slightly
- Moderately
- Greatly

A criticism of the UDI is that it appears to confound "bother" of the symptom with "presence or absence" of the symptom. The UDI focuses on the extent of bother of symptoms whereas tools such as the IIQ seem to focus on the experience or impact of the symptoms (for example, how urine leakage affects lifestyle and feelings). This focus would be an important aspect of the suitability of these tools for specific research questions.

Short Form of UDI

The six questions have the same items as reported above and cover:

- Frequent urination
- Urine leakage related to the feeling of urgency
- Urine leakage related to physical activity coughing or sneezing
- Small amounts of urine leakage (drops)
- Difficulty emptying your bladder
- Pain or discomfort in the lower abdominal or genital area

Although the short form of the UDI includes items on type of incontinence (urge/stress) this form only addresses the frequency of "a small amount of urine loss". It does not include the item from the UDI (long form) concerning the frequency of a large amount of urine loss. For this reason it may not be as good a measure of severity (frequency x amount) as other instruments.

Derivatives of the UDI

The UDI has recently been modified by Robinson and Shea (2002) into the Male Urogenital Distress Inventory (MUDI) and the Male Urinary Symptom Impact Questionnaire (MUSIQ). The MUDI has 27 items and the MUSIQ 32 items with a modal completion time of 20 minutes. Cronbach's alpha results in a sample of 153 male urology clinic patients were found to be 0.89 for the MUDI and 0.95 for the MUSIQ.

Brown, Posner and Stewart (1999) reported the re-development of the Incontinence Impact Questionnaire and Urogenital Distress Inventory. Cronbach's alpha for the new tools was found to be ranging between 0.74 and 0.95 and test-retest reliability using the ICC was found to be equal to 0.59, a result that is below the accepted threshold for reliability.

Wyman et al. (1997) used the Incontinence Impact Questionnaire along with a diary pad test and other clinical measures to assess changes in quality of life following a six week trial of bladder training in a study sample of 123 older women with urinary incontinence. The IIQ sub-scales and composite scale showed a significant improvement following training. However, the study reported minimal psychometric data.

Lubeck, Prebil, Peoples and Brown (1999) have modified the UDI and IIQ for urge incontinence patients, creating the U-UDI and the U-IIQ. The U-IIQ has 7 scales (Travel, Activities, Physical activities, Feelings, Relationships, Sexual function, Nighttime bladder control) that are scored seperately (a single index score was also calculated by the authors). The U-UDI consists of 9 items combined into a single score that measures the extent to which incontinence symptoms bother patients and an urge symptoms summary score. In their validation study, Lubeck et al. administered the tools at the start of baseline week, end of baseline week, 4 weeks and 12 weeks (final dose). Cronbach's alpha was between 0.82 and 0.96; and test-retest reliability, using the ICC, was between 0.68 and 0.83 for all scales. The UDI, IIQ and the associated family of modified derivatives show generally acceptable psychometric qualities and have been used in a variety of settings and hence warrant their highly recommended rating from the ICS/WHO consultation.

The Incontinence Severity Index

The Incontinence Severity Index is a rarity amongst measures of incontinence in that it is a very short measurement tool. It has two items, "How often do you experience urine leakage" and "How much urine do you lose". The index is calculated by multiplying the two item scores together. The structure of the index is shown below:

How often is urine leakage experienced?

Never = 0
Less than once a month = 1
1 to several times a month = 2
1 to several times a week = 3
Every day and/or night = 4

How much urine lost each time?

A few drops = 1
A little = 1
More = 2

Severity index = (points for frequency) x (points for amount)

The minimum score is 0 and the maximum score is 8. Higher scores denote more severe urinary incontinence.

The index was developed by Sandvik et al. (1993) for use in a Norwegian epidemiological survey of health problems. The index has been validated on several occasions. In the original 1993 study concurrent validity with a 48-hour pad weighing test was found to be $r = 0.59$. In a later study, Sandvik, Seim, Vanvik and Hunskaar (2000) reported a similar concurrent validity of 0.54 again with a 48-hour pad weighing test. The moderate correlation of the Index with "objective" measurement parameters is a strong feature of this tool, as many others show lower validation correlations.

The Incontinence Severity Index was originally intended for use with women and provides a useful short form diagnostic severity measure for urinary incontinence, although it has not been widely used as yet. Clearly the ISI does not measure either urgency or the type of urinary incontinence.

The DAN-PSS-1

The DAN-PSS-1 is a 12 item self-administered questionnaire developed for men with prostate problems by a Danish group of researchers (Hansen et al. 1997; Hansen et al. 1995). The developers report reasonable internal consistency of the scale with Cronbach's alpha = 0.73 and a median test-retest reliability of 0.84 across all questions over a week long test-retest interval.

This tool has not been used widely in clinical or research settings and is targeted at males with prostate problems. Its applicability to other contexts is yet to be demonstrated. For these reasons no in-depth review of this tool was undertaken. Given the low level of current use and evidence of applicability to a range of study populations, the strong endorsement of this tool by the ICS/WHO consultation group was somewhat surprising.

ICS Male and ICS Male SF

The International Continence Society (ICS) standard tool (ICS Male) has 22 items and the ICS Male Short Form has 11 questions concerning urinary incontinence symptoms for male respondents. This tool was developed to measure the symptomatology and "bothersomeness" of lower urinary tract problems for men with prostatic disease. The same team as the one that developed the Bristol Female Lower Urinary Tract Symptoms questionnaire worked on the ICS Male and there is considerable overlap in item content.

Donovan et al. (1996) reported the outcomes of a study of the psychometric properties of the ICS Male in a 12 country study of 1271 men with urinary tract problems and 471 British ambulant men. The study demonstrated sound psychometric properties and good discriminant validity for clinical and non-clinical groups.

The ICS Male SF consists of 2 sub-scales. The ICS Male VS: voiding sub-score comprising of five questions (hesitancy, straining, reduced stream, intermittency, incomplete emptying), and the ICS Male IS: incontinence sub-score comprising of six questions (urge, stress, unpredictable and nocturnal incontinence, urgency, postmicturition dribble). The scores are obtained by summing of the ratings for each item.

In a subsequent study reporting the development of the short form tool, Donovan et al. (2000) reported the psychometric properties of the tool components Voiding Sub-score (VS) and Incontinence Sub-score (IS). Cronbach's alpha for VS was 0.76 and 0.78 for IS. Responsiveness was also satisfactory. Thus the short form should be preferred to the full length form, as the longer form does not have major psychometric advantages.

Bristol Female Lower Urinary Tract Symptoms Questionnaire

The BFLUTS questionnaire was developed by the same team that developed the ICS Male questionnaire and its derivatives.

The developers Jackson, Donovan, Brookes, Eckford, Swithinbank and Abrams (1996), report sound psychometric properties for the BFLUTS tool based upon a robust psychometric trial. Cronbach's alpha was 0.78 and test retest reliability of 0.86 over a two week period reported in the initial study. This study also reported what the authors term 'criterion validity' (and what might be termed concurrent validity) by relating test results to frequency/volume charts and pad-test data, showing quite acceptable correlations between self report measures and "objective" tests.

The BFLUTS questionnaire has been used in a range of clinical studies. For example, Kulseng-Hansen and Berild (2002) studied outcomes of incontinence surgery using the BFLUTS, with patients who had undergone a Burch Colposuspension between 5 and 10 years after this procedure. The study found that leakage occurred with varying frequency, amounts and bother amongst the study population. Ward, Hilton, and the UK and Ireland TVT Trial Group (2004) have recently published a two-year follow-up prospective multicenter randomized trial of tension-free vaginal tape and colposuspension for primary urodynamic stress incontinence that used the BFLUTS. Temml, Haidinger, Schmidbauer, Schatzl, and Madersbacher (2000) also have used the BFLUTS in study of the prevalence rates of urinary incontinence and its impact upon quality of life and sexual behaviour. Van der Weide, Hilbrands,

Bemelmans, and Kiemeney (2004) have demonstrated satisfactory responsivity of the BFLUTS in their research.

It is interesting to note that the ICS Male questionnaire received a highly recommended rating from the ICS/WHO consultation groups, whereas the BFLUTS questionnaire received a recommended rating. When one considers that essentially this is the same tool applied in male and female populations, this was a surprising recommendation. If anything, the evidence base for the Bristol tool in terms of range of application is broader than that for the ICS male measure.

The Symptom Severity Index

This is a short tool designed to assess stress incontinence symptoms in women. The developers report the outcomes of a single study of the tool psychometrics and application, which shows satisfactory outcomes, (Black, Griffiths and Pope, 1996). However, this tool has, to date, been used in relatively few published studies of incontinence and for this reason no in-depth review was undertaken of this tool. This may change in time.

Other measures of symptom severity

Elser, Fantl & McClish (1995) conducted a study to compare subjective and objective measures of severity of urinary incontinence in 265 women participating in a clinical trial for urinary incontinence. The subjective measures of incontinence included patient recall of the number of incontinent episodes in one week, the number of perineal pads used during the week and the number of clothing changes required due to an incontinence episode. The objective measures included the number of incontinent episodes per week as recorded in a 7-day diary, as well as the number of perineal pads used during the week and the amount of fluid lost using a standardized pad test. Multi-variate correlation analyses revealed strong positive correlations between subjective and objective measures ranging between $r = 0.63$ to 0.81.

Groutz, Blaivas & Rosenthal (2000) conducted a study with the goal of developing a simplified urinary incontinence score for outcome measurement. Ninety-four per cent of women with incontinence problems underwent a clinical evaluation, 24-hour voiding diary and 24-hour pad test. The outcome measures were combined into a single score divided into 5 categories: cure, good response, fair response, poor response and failure. The patients were evaluated at baseline and one year post operatively. However, few psychometric data are reported about the properties of this score, and it has been used in few other studies to date.

Barry et al. (1992) reported the deliberations of a measurement committee of the American Urological Association in the development of a symptom index for benign prostatic hyperplasia. The validation study was conducted with 210 patients and 108 control participants. The AUA symptom index consists of seven questions concerning frequency, nocturia, urinary stream, hesitancy, intermittence, emptying and urgency of urine functions. The index was found to have a Cronbach's alpha of 0.86 and test-retest reliability was at the level of $r = 0.92$. The index was validated against participant ratings concerning the extent of the urinary incontinence with correlations ranging between $r = 0.65$ to 0.72. The Committee reported that satisfactory responsivity or sensitivity to change was achieved with the index.

Aydos, Memis, Yakupoglu, Ozdal and Oztekin (2001) reported the outcome of a study using the American Urological Association's Symptom Index in assessing outcomes of urethroplasty in a subject sample of 33 men. The symptom scores were correlated with maximum flow rates, but the correlations were modest, at $r = 0.40$.

Other studies have examined various aspects of the American Urological Association's Symptom Index. For example, Barnboym, Ahrens and Roehrborn (1999) studied the effects of scrambling of items on the short-term test-retest reliability. They found a high degree of reliability with item scrambling in a study sample of 111 participants. Similarly, MacDiarmid, Goodson, Holmes, Martin and Doyle (1998) studied the reading age level (n = 202 males) and concluded that a significant percentage of patients could not read the AUA Symptom Index and would require assistance from others to complete it. They argued that this assistance might introduce significant interviewer bias.

Sirls, Kirkemo and Jay (1996) reported a lack of correlation of the symptom index with urodynamic bladder outlet obstruction. They reported the sensitivity and specificity of the index for urodynamic obstruction as 42.5% and 54.3% respectively. This is not a good psychometric performance with respect to prediction of bladder outlet obstruction for the index.

Chancellor and Rivas (1993) administered the index to 38 women with voiding problems. The researchers concluded that the index was not suitable for this clinical population and given this an in-depth review of this tool has not been undertaken.

It would seem that the AUA Symptom Index would not be a preferred measure at this stage, as clarification is required regarding its application.

The ICIQ and ICIQ-SF

For future reference it is important to note that the second ICS/WHO Consultation on Incontinence report outlines the forthcoming developments of the ICIQ and ICIQ-SF tools. The domains covered by the tool include symptoms of incontinence, bothersomeness of symptoms, use of pads/protection, impact on aspects of everyday life including social and occupational factors, interference with sex life, emotional impact, follow-up after treatment, and diagnosis. However, the tool is in early stages of development and cannot currently be recommended.

The following table shows the outcomes of the expert panel deliberations using the evidence assembled from the literature review.

Table 22   Summary of ratings for urinary incontinence symptom measures

Criteria	Tool								
	Kings HQ	UDI Long	UDI Short	ISI	BFLUTS	AUASI	ICS-male	ICS-male short	Dan-PSS-1
Availability of comparison data/usage	3	2	2	1	3	2	1	1	1
Length, ease and time to complete	2	2	3	3	2	3	2	2	2
Method of administration	2	2	2	2	2	2	2	2	2
Translations available	3	2	2	3	2	1	2	2	2
Ease of scoring	2	2	2	3	2	2	2	2	2
Sensitivity to incontinence	3	3	3	2	2	1	2	2	1
Reliability evidence available	3	3	2	3	2	2	2	2	2
Validity evidence available	3	3	3	3	2	2	2	2	2
Adherence to psychometric axioms	3	3	3	2	2	2	2	2	2
Cost of using the instrument	3	3	3	3	3	3	3	3	3
Weighted Total	**66**	**62**	**61**	**58**	**53**	**48**	**47**	**47**	**44**

KHQ	King's Health Questionnaire
UDI Long	Urogenital Distress Inventory Long Form
UDI Short	Urogenital Distress Inventory Short Form
ISI	Incontinence Severity Index
BFLUTS	Bristol Female Lower Urinary Tract Symptom assessment
AUASI	American Urological Association Symptom Index
ICS-male	International Continence Society Male assessment
ICS-male short	International Continence Society Male short form assessment
Dan-PSS-1	No English Acronym

Of the tools studied in the Urinary Incontinence Symptom Measures category, the King's Health Questionnaire, the Urogenital Distress Inventory (UDI) Short Form, the Urogenital Distress Inventory (UDI) Long Form, the Incontinence Severity Index, the Bristol Female Lower Urinary Tract Symptom assessment, the American Urological Association Symptoim Index, the International Continence Society Male assessment, and the International Continence Society Male short form assessment, made the required 47-point score threshold. Given their high ratings, the first two measures, the King's Health Questionnaire and the UDI (in either form) are the recommended tools in this category and the ISI is recommended for use in primary care and public health settings.

The ICS/WHO consultation outcomes provide a sound starting point for consideration of recommendations for urinary incontinence symptom measures. However, the recommendations for some tools have been made on the basis of what is considered to be limited evidence and experience in terms of application for some of the tools. For example, there is a large difference between the uptake and current usage of tools such as the Kings Health Questionnaire and the Urogenital Distress Inventory and its derivatives, where on the one hand the tools have been used widely and are in translation in over 20 languages, and tools such as the DAN-PSS-1 where the application has been, to date, limited mainly to the country of development or a small number of similar language group countries in a small number of studies.

There is also the issue of applicability of the selected measurement tools across gender groups and conditions. Broadly it is considered that, not withstanding the importance of gender differences, it is inherently desirable that the one tool is applicable to as many groups as possible. This then enables benchmarking of clinical outcomes for groups against one another without the added distraction of having to take account of interpretation problems associated with the use of different tools.

Based on the criteria of reliability, validity, applicability and practicability, it is considered that the Kings Health Questionnaire and Urogenital Distress Inventory in either the short or long form derivatives are sensible choices for specialist practitioners and researchers. At this point, tools are probably impractical for primary care practitioners because of their length. The Incontinence Severity Index is probably worthy of consideration for primary care practitioners because of its brevity and the demonstrated validity with more "objective" measures of urinary incontinence. While the other tools may well be psychometrically sound and well conceptualised, their low usage in a wide variety of incontinence contexts and patient groups limits our ability to recommend them at this stage of the development of the evidence.

Table 23   Recommendations for symptom measures in urinary incontinence

User category	Recommended Urinary Incontinence Symptom Measures
Primary Care Practitioner	Incontinence Severity Index
Specialist Practitioner	1. King's Health Questionnaire 2. Urogenital Distress Inventory
Researcher	1. King's Health Questionnaire 2. Urogenital Distress Inventory

Urinary Incontinence Health-Related Quality of Life Measures

There has been some minimal work done in the development of condition specific, health-related quality of life measures. Patrick, Martin, Bushnell, Yalcin, Wagner and Buesching (1999) and Wagner, Patrick, Bavendam, Martin, and Buesching (1996) report the development of the I-QOL, but searches in PubMed revealed very scant empirical work had been done using these tools. This is a very similar situation to that encountered in the assessment of conditional specific faecal incontinence tools, i.e. there are almost no studies. It is likely that this situation reflects a reluctance to develop condition specific measures of health-related quality of life as a result of the obvious disadvantages associated with such tools.

No suitable tools were recommended in this category. The project team does not support the use of condition specific health-related quality of life measures in favour of generic measures. The use of condition specific measures means that a considerable test development and validation load is acquired, perhaps needlessly, and that the ability to benchmark the impact of specific conditions upon health-related quality of life with other conditions is needlessly constrained. This is a large price to pay for minimal if any advantage. It may be better to combine a symptom measurement tool with a generic health-related quality of life tool if it is desired to measure health-related quality of life in a particular condition domain.

(NB: For those interested, a review of the I-QOL has been provided because it does show some promise as a urinary incontinence condition specific, health-related quality of life measure.)

Table 24   Recommendations for health-related quality of life measures in urinary incontinence

User category	Recommended Condition Specific Health-Related Quality of Life Measures
Primary Care Practitioner	None
Specialist Practitioner	None
Researcher	None

Pad tests

The pad test is a means of quantitatively measuring the amount of urine leakage over a specified time interval. The big issue in the design of pad tests is the time over which the test should be administered and the conditions under which it should be administered. It is hardly surprising that longer tests have been found to have superior psychometric properties, as this is the case with the observation of any behavioural phenomenon. The longer the period of observation generally results in more reliability and validity. However, the length of the test may also impact upon adherence as the longer the test, the greater the degree of inconvenience to patients with the consequent threats to adherence.

Victor (1990) reviewed the literature on pad weighing tests for measurement of urinary incontinence. Victor argued that the short term tests, while having distinct practical advantages, actually have low sensitivity and require very close attention to procedure in order to reach acceptable test-retest reliability. He argued that the long term tests of duration 12 – 14 hours have much better sensitivity and test-retest reliability and hence validity.

Ryhammer, Djurhuus and Laurberg (1999) performed a comprehensive review of pad weighing tests for women with urinary incontinence. In their review they noted that there is a large degree of intra and individual variability and that it may not be possible to use such measures to formulate cut-off values to separate continence from incontinence.

The Second ICS/WHO consultation on incontinence concluded that:
- A 24-hour test has good reproducibility
- A test lasting longer than 24 hours has little advantage
- A pad test cannot distinguish between stress and urge incontinence

A review of some of the many studies of this issue is now presented.

Jorgensen, Lose and Andersen (1987) studied the reliability and validity of the one-hour pad weighing test in a study sample of 81 women. The test-retest reliability was found to be r = 0.68. The researchers described this is as relatively good. However, it is at the low end of expectation from a psychometric viewpoint.

Lose, Rosenkilde, Gammelgaard and Schroeder (1988) reported a further study of the one-hour pad weighing test in a study sample of 25 female patients with urinary incontinence. On this occasion, test-retest results were found to be highly correlated with r = 0.97. However, there were significant differences in measurement of the actual derived amounts. The researchers did not use the more appropriate ICC co-efficient to analyse their data. This would have revealed changes en masse in the absolute values and would have been sensitive to phase shifts.

Nygaard and Zmolek (1995) studied the psychometric properties of a modified pad test in a sample of 14 a-symptomatic participants. Using a Kendall co-efficient of concordance they found that the overall test-retest reliability was 0.96 for pad weight gain, 0.76 for area of stain and 0.60 for volume voided.

Rasmussen, Mouritsen, Dalgaard and Frimodt-Moller (1994) tested the test-retest reliability of the 24-hour pad weighing test and the impact of fluid intake and level of activity upon the test result. A study sample of 14 women referred to a urinary incontinence clinic performed 6 x 24-hour pad tests with different levels of activity and high and low fluid intake. The researchers reported that the 24-hour pad test gave a satisfactory level of test-retest reliability.

Ryhammer, Laurberg, Djurhuus and Hermann (1998) studied the association between urinary incontinence and pad tests using a 24-hour pad test in a sample of 144 healthy menopausal women. The researchers found little relationship between their subject assessment and urinary incontinence and pad test outcomes.

Simons, Yoong, Buckland and Moore (2001) reported an important study concerning the reproducibility of outcomes of the one-hour pad test. In a sample of 56 incontinent women, two one-hour pad tests were performed one week apart prior to treatment. Substantial variations between the tests were found and confidence intervals associated with these estimates were found to be large. The test-retest reliability of the one-hour pad test was found by these researchers to be clinically inadequate and they concluded its poor repeatability implies that it should not be used in the measurement of post treatment outcomes for urinary incontinence.

Persson, Bergqvist and Wolner-Hanssen (2001) reported the outcomes of an ultra-short perineal pad test for female stress urinary incontinence. In their study 34 women with stress incontinence, 13 with urge incontinence and 10 non-incontinent participants took part in the study. The researchers concluded that the pad test has acceptable reproducibility.

Wilson, Mason, Herbison and Sutherst (1989) reported the outcomes of an evaluation of the home pad test for quantifying urinary incontinence. They concluded that this test was psychometrically sound.

Versi et al. (1996) performed a study of the home pad test for women with urinary incontinence. Test-retest reliability analysis of 24-hour and 48-hour tests showed excellent correlation coefficients of 0.9 and 0.94. The researchers concluded that the home pad test, in combination with the frequency volume chart, is reliable, valid and practicable.

These studies confirm the reviews cited of the performance of pad tests, that the short tests of 1 or 2 hours or shorter do not provide sufficient psychometric performance to be used. If it is required to measure the volume of leaked urine then a 24-hour or 48-hour test is required whether this is for a practitioner or researcher. There is no point in collecting convenient but suspect data.

Recommendations Concerning Pad Tests

Pad tests are a means of quantitatively assessing the amount of urine leakage from a person with urinary incontinence. Primary care practitioners in general are not likely to require the outcomes of a pad test in routine practice although this is a useful tool for researchers and specialist incontinence clinicians.

Table 25   Recommendations for pad tests

User category	Recommended use of Pad Test Measures
Primary Care Practitioner	24-hour pad test
Specialist Practitioner	24-hour pad test
Researcher	24-hour or 48-hour pad test

Frequency Volume Charts and Bladder Diaries

The frequency/volume chart is a tool to record intake and urine output for each 24-hour period. Most of these charts are designed to chart the number of voidings and their timing over the daily period and each voided volume. In addition it can be used to record episodes of urgency and leakage and the number of incontinence pads used.

Many studies have investigated the properties of the charts and the diaries:

Locher, Goode, Roth, Worrell and Burgio (2001) report the outcomes of a major study of the reliability of bladder diaries for urinary incontinence in a study sample of 214 older women with a history of stress urge and/or mixed urinary incontinence. The participants maintained a 14 day bladder diary in which they kept records of the time and characteristics of each incontinence episode. It was found that 5 days of recording were necessary to obtain an internal consistency score of 0.90 for Cronbach's alpha. The researchers suggested that seven consecutive days of bladder diary provides a satisfactory provides a satisfactory, valid and reliable measure of frequency of incontinence episodes.

Nygaard and Holcomb (2000) performed a study of the reliability of the seven day voiding diary in a sample of 138 women with stress urinary incontinence. The study participants maintained two, seven day voiding diaries completed at four week intervals. The test-retest reliability for the number of weekly incontinence episodes was found to be equal to r = 0.83. The correlation between the first three days of the seven day diary with the last four days was found to be r = 0.89. The researchers concluded that their results supported the use of a three day diary as an appropriate outcome measure in stress urinary incontinence.

Rabin, McNett and Badlani (1996) reported the outcomes of an evaluation of an electronic voiding diary called "Compu-Void". However, they did not report reliability and validity outcomes that are superior to those of more conventional methods.

Robinson, McClish, Wyman, Bump and Fantl (1996) reported a study in which they investigated the impact of completion of seven day urinary diaries with, and without, detailed instructions. A sample of 278 women with urinary incontinence participated in the study. The test-retest reliability for the diaries was found to vary from 0.67 to 0.78. The researchers concluded that the seven day urinary diary is a valid and reliable tool to assess urinary incontinence symptoms.

Wyman, Choi, Harkins, Wilson and Fantl (1988) reported the outcomes of a study in which they investigated the use of a one week urinary diary in a sample of 50 community dwelling older women. The participants kept the diary for two consecutive weeks. Test-retest reliability was found to be satisfactory for the diary.

Chaikin, Blaivas, Rosenthal and Weiss (1999) reported the results of a study designed to assess the reliability and validity of alternative outcome measures for pubo-vaginal sling surgical intervention for stress incontinence. On the basis of their analysis, they concluded that the pad test and voiding diary were found to be psychometrically satisfactory and they recommended that these should be part of routine outcome measurement following surgical intervention.

The Second ICS/WHO Consultation on Incontinence has provided set of templates for micturition time charts, frequency volume charts, and bladder diaries. We consider that there is considerable advantage in the adoption of uniform recording standards.

It is recommended that the Committee's template be adopted by researchers and practitioners. The recommendations from the Committee are reproduced below in the form of recommended form templates from their report:

Three types of Bladder Charts and Diaries can be used to collect data:

1) Micturition Time Chart
   - Times of voiding
   - Incontinence episodes.

2) Frequency Volume Chart
   - Times of voiding with voided volumes measure
   - Incontinence episodes and number of changes or incontinence pads or clothing.

3) Bladder Diaries
   - The information above, but also
   - Assessments of urgency
   - Degree of leakage (slight, moderate or large) and descriptions of actions leading to symptoms such as stress leakage, e.g. running to catch a bus.

It is important to assess the individual's fluid intake, remembering that fluid intake includes fluids drunk plus the water content of foods eaten. It is often necessary to explain to a patient with LUTS that it may be important to change the timing of a meal and the type of food eaten, particularly in the evenings, in order to avoid troublesome nocturia.

The micturition time and frequency volumes charts can be collected on a single sheet of paper.

LUTS = Lower Urinary Tract Symptoms

Figure 1: Frequency/Volume Chart – Standard Version – 7 Days

Name: _____

Date	7:00 am	Midday	Midnight	6:00am	Pads used

No. of drinks per day: _____

In each chart/diary, the time the individual got out of bed in the morning and the time they went to bed at night should be clearly indicated.

Each chart/diary must be accompanied by clear instructions for the individual who will complete the chart/diary: the language used must be simple as in the suggestions given for patient instructions. There are a variety of designs of charts and diaries and examples of a detailed bladder diary are given. The number of days will vary from a single day up to one week.

INSTRUCTIONS FOR COMPLETING THE MICTURITION TIME CHART

This chart helps you and us to understand why you get trouble with your bladder. The diary is a very important part of the tests we do, so that we can try to improve your condition. On the chart you need to record:

1. When you get out of bed in the morning, show this on the diary by writing, "Got out of bed".
2. The time, e.g. 7.30 am when you pass your urine. Do this every time you pass urine throughout the day and also at night if you have to get up to pass urine.
3. If you leak urine, show this by writing a "W" (wet) on the diary at the time you leaked.
4. When you go to bed at the end of the day show it on the diary – write, "Went to bed".

INSTRUCTIONS FOR USING THE FREQUENCY VOLUME CHART

This chart helps you and us to understand why you get trouble with the bladder. The diary is a very important part of the tests we do, so that we can try to improve your condition. On the chart you need to record:

1. When you get out of bed in the morning, show this on the chart by writing "Got Out of Bed".
2. The time, e.g. 7.30 am when you pass your urine. Do this every time you pass urine throughout the day and also at night if you have to get up to pass urine.
3. Each time you pass urine, collect the urine in a measuring jug and record the amount (in mls or fluid ozs) next to the time you passed the urine, e.g. 1.30 pm – 320 mls.
4. If you leak urine, show this by writing W" (wet) on the diary at the time.
5. If you have a leak, please add "P" if you have to change a pad and "C" if you have to change your underclothes or even outer clothes. So, if you leak and need to change a pad, please write "WP" at the time you leaked.
6. At the end of each day please write in the column on the right the number of pads you have used, or the number of times you have changed clothes.
7. When you go to bed at the end of the day show it on the diary – write, "Went to Bed".

INSTRUCTIONS FOR USING THE BLADDER DIARY

This diary helps you and us to understand why you get trouble with your bladder. The diary is a very important part of the tests we do, so that we can try to improve your condition. On the chart you need to record:

1. When you get out of bed in the morning, show this on the diary by writing "Got out of bed".
2. During the day please enter at the correct time the drinks you have during the day, e.g. 8.00 am – two cups of coffee (total 400 ml).
3. The time you pass your urine, e.g. 7.30 am. Do this every time you pass urine throughout the day and night.
4. Each time you pass your urine, collect the urine in a measuring jug and record the amount (in mls or fluid ozs) next to the time you passed the urine, e.g. 1.30 pm/320 ml.
5. Each time you pass your urine, please write down how urgent was the need to pass urine:
   "O" means it was not urgent.
   + means I had to go within 10 minutes.
   ++ means I had to stop what I was going and go to the toilet.
6. If you leak urine, show this by writing a "W" on the diary at the time you leaked.
7. If you have a leak, please add "P" if you have to change a pad and "C" if you have to change your underclothes or even outer clothes. So if you leak and need to change a pad, please write "WP" at the time you leaked.
8. If you have a leakage please write in the column called "Comments" whether you leaked a small amount or a large amount and what you were doing when you leaked, e.g. "leaked small amount when I sneezed three times."
9. Each time you change a pad or change clothes, please write in the "Comments" column.
10. When you go to bed at the end of the day show it on the diary – write "Went to Bed."'

DAILY BLADDER CHART

Please fill in this chart before coming to the Unit; it gives your doctor important information and can help you to understand your condition.

Please fill in the <u>approximate time</u> and <u>amount</u> of how much fluid you drink, as illustrated in the example below.

Please fill in the <u>correct time</u> and <u>amount</u> of how much water (urine) you passed: either millilitres or ounces. Please mark with a star * when you leak/when pad is damp. Fill in the chart for 3 days (e.g. 3 x 24 hours = 72 hours) in a row.

Date/ Time	Fluids taken in	Amount of urine passed mls <u>or</u> ounces	*= wet	Date/ Time	Fluids taken in	Amount of urine passed mls <u>or</u> ounces	*= wet
EXAMPLE / ILLUSTRATION :							
7.6.99 2am	–	150 mls					
7am	–	250 mls	*				
8am	Mug coffee	–					
8.20	–	60 mls	*				
9.30	Cup orange	–					
10.00	–	100 mls					
12.00	2 mugs coffee	–					
14.00		360 mls	*				
AND	SO ON FOR 3	DAYS					

Recommendations for Frequency Volume Charts, Bladder Charts and Diaries

The research in this area, as with the research with pad tests shows that broadly the longer the period of the test the better the psychometric performance. However, Nygaard and Holcomb's study of the reliability of the seven day voiding diary found that the correlation between the first three days of the seven day diary with the last four days was found to be r = 0.89. This suggests that a 3-day diary may have sufficient psychometric performance for clinical evaluation.

It is unlikely that most primary care practitioners would require the degree of precision provided by diary/frequency volume and other charting methods but for those who do, a three-day period is required to achieve sufficient accuracy, Nygaard and Holcomb's study suggest that a three-day period is suitable. For researchers the longer period provides better psychometric performance but at a practicability cost.

Table 26   Recommendations for frequency volume charts, bladder charts and diaries

User category	Recommended use of Frequency Volume Charts Bladder Charts and Diaries
Primary Care Practitioner	ICS/WHO templates for 3 days
Specialist Practitioner	ICS/WHO templates for 3 days
Researcher	ICS/WHO templates for 3 to 7 days

Recommendations for Urinary Incontinence Outcome Measures

The following table summarises the recommendations made for urinary incontinence outcome measures.

Table 27   Recommendations for outcome measures in urinary incontinence

Tool Content Domain	User Category		
	Primary Care Practitioner	**Specialist Practitioner**	**Researcher**
Urinary Incontinence Symptom Measures	ISI	1. King's Health 2. UDI	1. King's Health 2. UDI
Urinary Incontinence Health-Related Quality of Life Measures	None	None	None
Pad Tests	24-hour pad test	24-hour pad test	24 or 48-hour pad test
Frequency Volume Charts and Bladder Diaries	ICS/WHO templates for 3 days	ICS/WHO templates for 3 days	ICS/WHO templates for 7 days

Multi Attribute Utility Instruments

This section reviews multi-attribute utility (MAU) instruments in the context of future Australian epidemiological research on incontinence, including population screening or surveillance as well as clinical treatment trials. It was authored by A/Professor Graeme Hawthorne.

The instruments reviewed are, in order of publication, the Rosser Index, QWB, HUI3, 15D, EQ5D, AQoL and SF6D.

The sources and publications used

The literature used in this report comes from searches of Medline, Psychlit and Econolit. Additionally, citations such as reports were sought out where they were deemed relevant. The following shows the number of such references for each of the instruments: Rosser Index: 18 (18 journal articles); QWB: 94 (90 journal articles); EQ6D: 308 (299 journal articles); 15D: 19 (10 journal articles); HUI3: 23 (21 journal articles); AQoL: 36 (13 journal articles); SF6D: 6 (5 journal articles).

The low number of journal articles for the AQoL as a proportion of all references may reflect the author's familiarity with non-journal publications for the AQoL. Subject to this bias, it would appear there is publication bias by instrument publication date: in general the earlier instruments (QWB, EQ5D) have greater publication; this of course, does not hold true for the Rosser Index which has a more specialized market. Regarding the literature, then, use in studies should not be accepted as indicative of instrument validity as this would imply that popular usage confers known properties!

Economic evaluation, cost-utility and the axioms of utility measurement

The growing interest in the measurement of health-related quality of life (HRQoL) can be attributed to four interrelated health and health care changes (Imhof, 1992). Health care technologies have reduced early mortality and prolonged the lives of those who would otherwise have died (Nordenfelt, 1994); there has been a shift in economically developed societies from exogenous to endogenous chronic diseases (Walker and Rosser, 1993); many health services are now designed to prevent deterioration in quality of life (Bowling, 1991); and there is increasing conflict between potentially useful interventions and the resources available to fund them (Drummond et al. 1998).

These changes suggest health resources should be allocated in ways that best benefit communities (Nordenfelt, 1994, Imhof, 1992). This can be achieved through providing services which lead to benefits in people's life-length and HRQoL. Evaluative research (Suchman, 1967) is needed to ensure that potential benefits are realised, including economic evaluation contributing to the decision-making processes associated with resource allocation.

HRQoL and economic evaluation

The World Health Organisation's definition of quality of life (QoL) is an individual's perception of their position in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns (WHOQoL Group, 1995).

Since it is the individual who experiences and values this state, the challenge is to find ways of measuring this individual perspective that are sensitive, valid and reliable and yields respondent information which can be considered alongside clinical and clinician data.

Health care evaluative research has traditionally been at the level of summative evaluation aimed at assessing the extent to which the intervention benefited people's health (Suchman, 1967, Shortell and Richardson, 1978, Ovretveit, 1998). Where comparative information across interventions is sought for resource allocation decisions, economic evaluation is used. Useful introductions to economic evaluation can be found in Drummond (2001), Singh et al. (Singh et al. 2001), and Drummond et al. (1998). Economic evaluation offers three important mechanisms: it can describe the cost of the burden of incontinence, it can predict the level of resources that will be needed in the future for treating incontinence, and it can provide information about the best use of resources available for incontinence interventions. Although there are four major kinds of economic analysis, the most comprehensive are cost-utility and cost-benefit (Drummond et al. 1998).

Multi-attribute utility (MAU) instruments are designed explicitly to be used in cost-utility evaluations. This review provides an overview of the leading MAU instruments and assesses them against issues relevant to incontinence.

The axioms of utility measurement

The basic axiom of cost-utility analysis is simple: life years are weighted by the value of a given health state in such a way that the values – referred to as "utilities" – act as an exchange rate between the quantity and quality of life. In this context, "utilities" are assumed to be preferences for a given health state. Regarding the measurement of utilities, Torrance (1986) provides the classic text.

To understand utilities, consider the following. Most people would prefer to be healthy over a given time rather than suffer constant urinary or faecal incontinence. Utility measurement refers to valuing these preferences on a life-death scale with endpoints of 1.00 and 0.00, where 0.00 is death equivalent and 1.00 is perfect (very good) HRQoL. For example, the measured utility for urinary incontinence may be 0.60. If treatment improves this to 0.70, then the value of the treatment is 0.70 – 0.60 = 0.10. If this utility gain is maintained over time, say for 10 years, then the gain is 0.10 x 10 = 1.00 Quality adjusted life year (QALY). Because utilities fall on the life-death scale, they are (in theory) common across all health states and therefore can be used to compare the effect of interventions in different health fields,

or different interventions within the same field. For example, the QALYs gained from Treatment A for incontinence could be compared with those gained from Treatment B for depression. Where treatment costs (including costs to the patient) are known, the treatment providing the lowest cost-per-QALY gained is preferred as this ensures society gains the greatest benefit from the health care dollar.

To allow for comparison, utility measures must be generic and must allow for respondents to report they have excellent HRQoL (full health equivalent state: 1.00); additionally they must allow those who have appalling HRQoL to report this (death equivalent state: 0.00). If an instrument does not permit this full range of responses, it cannot accurately measure the HRQoL of people who fall outside its range. For example, if an instrument only allows measurement between 0.50 to 1.00, then it is incapable of reporting the effect of treatment for people who are in a desperate health state (say, close to death). Under these circumstances, any claim to generalisability for the instrument is foregone.

The instrument must be applicable to HRQoL states deemed worse than death (i.e. the respondent indicates they would rather die now than continue living in their current HRQoL state). These negative health states are needed to allow for people who commit suicide or euthanasia; they have clearly made the decision that death is preferable to living in their current health state and any possible future health states. When determining negative utility boundaries, the developers of the EQ5D and HUI3 adopted Torrance's symmetry argument. This states that since a person can "lose" HRQoL value from 1.00 (full health) to 0.00 (death equivalent), they must be able to "gain" an equivalent amount from –1.00 to 0.00 (Torrance, 1986). However, since negative utility values do not possess the same interval properties as positive utility scores (Hawthorne et al. 2000a, Richardson and Hawthorne, 2000), there are difficulties. For example, improving the HRQoL of a person from –0.35 to –0.25 (ie. bringing them closer to a HRQoL death-equivalent state) does not have the same meaning as improving their HRQoL state from 0.25 to 0.35; yet both these would have a utility gain of +0.10. This is implausible. It seems likely, then, that negative values should have lower boundaries close to 0.00 (death equivalent) (Richardson and Hawthorne, 2000).

Implicit in axioms and mathematical modelling of utilities is that utility measurement must be at the interval level, where interval level refers to measurement scales that have equal-intervals between the measurement points. There are two forms of interval measurement that MAU-instruments must have if they are to do their job correctly. One is known as the "weak" interval property and the other the "strong" interval property (Richardson, 1994). The weak interval property is where a gain of 0.10 means the same thing across the range of instrument scores. For a person who has severe faecal incontinence, their utility score might be 0.15; as a result of treatment this rises to 0.25; i.e. the value of the treatment is 0.25 – 0.15 = 0.10. Similarly, the value of the treatment is also 0.10 for a person with urinary incontinence with an initial utility of 0.60, and who gains a utility of 0.70 after treatment (0.70 – 0.60 = 0.10). The strong interval property is where there is a direct relationship between gains in utility and gains in life-length. Since QALY calculation represents the time spent in a given state multiplied by the quality of that state, this implies that a 0.20 utility gain multiplied by 5 years in the improved health state equals 1.00 QALY (from 0.2 X 5). But a gain of 1 QALY could also be the product of a 0.40 utility gain over 2.5 years (or any other combination).

Measuring utilities using MAU-instruments

There are two steps to measuring utilities using MAU-instruments. First, the health state of interest is described. Second, the value or utility of the health state is assigned.

When a person completes a MAU-instrument, their numerical responses provide a description of their health. For example, consider two people completing an imaginary instrument with four dimensions each of which has four levels. This instrument's "descriptive system" would be: physical, mental, social and cognitive health dimensions and the response levels are: 1 = normal, 2 = some impairment, 3 = major impairment, 4 = gross impairment. Person A, who is in the best of health, selects the best response to each item (i.e. "1": normal). Her health state would be described as "1, 1, 1, 1". Person B who reported major incontinence (level 3: major impairment on the physical dimension), normal mental heath (level 1), some social impairment (level 2), and normal cognitive function (level 1). Her health state would be "3, 1, 2, and 1".

Valuing these health states is called "scaling". Five procedures have been used: time trade-off (TTO), standard gamble (SG), visual analog rating scale (VAS), magnitude estimation (ME) and person trade-off (PTO). Brief descriptions are given.

- Time trade-off (TTO). A person with severe incontinence can have a treatment which will restore her to full health; but a side effect is she will live a shorter life. She is asked to choose how many

years of her life she would be willing to "give up" in order to be in full health. If, in her untreated condition, her life expectancy was 10 years and after the treatment this was 5 years she may reject the treatment. If after the treatment it was 9 years, she may accept it; if her life expectancy was 6 years, she may not. Her choices would continue back-and-forth like this until she indicated that she was indifferent to whether she had the treatment or not. If the point of indifference was that 8 years of full health was the equivalent of 10 years with severe incontinence, then the quality of life value for her current health state is 8/10 or 0.80.

- Standard gamble (SG). A person with urinary incontinence is presented with a treatment option that has two possible outcomes: either full health for the remainder of his life, or death. He is free to choose either the treatment or to remain with lifelong urinary incontinence. If the probability of full health is 1.00 (i.e. his incontinence will be cured and there is no chance of death), then obviously he will choose to have the treatment. If the probability of full health is 0.90 and death 0.10, he may still choose the treatment. However there would be a point, for example at 0.80 for full health and 0.20 for death, where he is not clear as to whether he would want the treatment or would choose to remain in his current health state. This point of indifference is the "value" of his health state.

- Visual analog scale (VAS). The respondent is asked to consider an incontinent health state and then to rate this on a scale, where the endpoints are typically 0.00 (death equivalent) and 1.00 (full health equivalent). Unlike the TTO or SG, with the VAS there is no uncertainty: the respondent is not asked to "trade" anything. Consequently many consider that VAS scores do not represent utilities because they provide a simple ranking of health states. Where VAS scores are used, a transformation is generally required, based on TTO or SG (Brazier et al. 1999, Bennett and Torrance, 1996, Robinson et al. 1997).

- Magnitude estimation (ME). The respondent is asked to consider the distance of the health state of interest (e.g. incontinence) from 1.00 (full health). Once several of these rating exercises have been carried out, the respondent is then asked to rank these in order (Gudex et al. 1993). Because there is no uncertainty, it is uncertain if ME represents utility.

- Person trade off (PTO). The respondent is asked to estimate the number of people that would have to be treated to make an intervention worthwhile. For example, a respondent might be asked to choose between extending the life of 10,000 people who were in full health by 1 year against a treatment which extended the life of N people with incontinence, also for 1 year. The number of people with incontinence would be varied until the respondent indicated they were indifferent between the two choices (Gudex et al. 1993).

When these techniques are used to obtain the utility weights used in an MAU-instrument, in theory each health state described by the descriptive system can be scaled (as was done with the original Rosser Index (Rosser, 1993)), but this is impractical because MAU-instruments typically generate thousands of different health states. Instead, a limited number of health states are scaled and the values for other health states are then inferred using econometric or decision analytic techniques, typically either an additive or multiplicative model (Hawthorne et al. 2001b). During scoring, the health state descriptors (1, 2, 3, etc.) are replaced with the appropriate values. For example, if the value of suffering mild pain based on TTO is "0.70" and the response levels on an item measuring pain were "1" (no pain), "2" (mild pain), and "3" (severe pain), then a person who selected "2" would have this level replaced with the value "0.70" during scoring of the instrument.

Once item-level values have been assigned, these are combined into an index on a life-death scale. Three procedures have been used.

- Additive models. The substituted importance values are summed and the resulting score represents the utility index. The limitation is that for full health equivalent HRQoL states each instrument item or dimension must contribute a fixed amount. Under this model, a respondent can obtain a very poor utility score only if they report poor scores on all items or dimensions. Consider an instrument measuring two dimensions: physical and mental health. In an additive model, each may contribute 0.50 towards the utility score. In this model, appalling mental health (leading to suicide) could never, by itself, lead to a utility value lower than 0.50 because 0.50 (a person in good physical health) + 0.00 (mental health) = 0.50. Thus additive models cannot explain people who commit suicide if their physical health is good or euthanasia if their mental health is good.

- Econometric models. The items are treated as explanatory variables to derive a regression equation predicting utilities. This method, however, suffers the same limitation as the additive model.

- Multiplicative models. These involve multiplying items or dimension scores together. This overcomes the limitation of the additive model as it allows any dimension to carry a person to a death equivalent value. Consider the case above. Here the person's value for mental health would be 0.00, and 0.50 (physical health) x 0.00 (mental health) = 0.00.

Given these assumptions, preference independence is required to avoid double-counting, which is where the same underlying health condition contributes more than once to the MAU-instrument utility index. For example, if a person is incontinent this should be counted in their utility score once, although the effect of this health state may be measured in several different aspects of their life; i.e. on several different scales. Where these effects are measured using unidimensional scales that are orthogonal to each other there is no difficulty. Where the scales, however, are correlated the effect of incontinence will be counted several times over thereby biasing the utility measurement. It is for this reason that MAU-instruments are required to possess structural independence (i.e. where the scales are unidimensional and orthogonal) (von Winterfeldt and Edwards, 1986). For example, if incontinence is counted on dimensions measuring social, physical and psychological dimensions as well as its effects being directly measured, then there is loss of preference independence as the scores on the social dimension may be a function of physical scores.

Description of MAU-instruments

In order of their development, MAU-instruments are the Rosser Index (Rosser, 1993), the Quality of Well-Being (QWB) (Kaplan et al. 1993, Kaplan et al. 1996), the Health Utility Index 3 (HUI3) (Feeny et al. 1996a, Feeny et al. 1996b, Torrance et al. 1995), the 15D (Sintonen and Pekurinen, 1993, Sintonen, 1995, Sintonen, 1994), the EQ5D (formerly the EuroQol) (EuroQol Group, 1990, Kind, 1996), the Assessment of Quality of Life (AQoL) (Hawthorne et al. 1999, Hawthorne et al. 2000c, Hawthorne et al. 2000b) and the SF6D (Brazier et al. 2002, Brazier et al. 1998). Additionally, Fryback et al. (Fryback et al. 1997) have prepared an algorithm for mapping SF-36 scores onto the QWB.

This report considers the Rosser Index, QWB, HUI3, EQ5D, AQoL and SF6D. Although there are three HUI instruments, only the HUI3 is considered. The Fryback et al. SF-36 algorithm is not a MAU measure in its own right.

The descriptions presented in this section are largely based on those given by Hawthorne & Richardson (2001a, 2001).

Rosser Index

The British Rosser Index was designed for use in hospital settings. The original version had two dimensions measuring disability and distress, and measured 29 health states. Values were elicited using magnitude estimation from a convenience sample of 70 respondents (Rosser, 1993). A revised version was released in the early 1990s based on SG procedures and included an additional dimension of discomfort (Rosser, 1993). Administration requires a trained interviewer. The upper boundary is 1.00, and the lower boundary – 1.49; i.e. health states worse than death are permitted. The Rosser Index has given rise to two variants: the Health Measurement Questionnaire (HMG) (Kind and Gudex, 1994) and the Utility-based Quality of Life-Heart Questionnaire (UBQ-H) (Martin, 1996). Permission must be obtained for using the instrument; however there are no costs for its use. No website was identified for the Rosser Index.

Quality of Well-Being (QWB or IWB)

The American QWB was designed to bridge the gap between clinical measurement, functional status and health planning policy (McDowell and Newell, 1987) and was an adaptation of US health surveys (Cadet, 1994). It has three dimensions (Mobility, Physical Activity, and Social Activity) with 3–5 levels each. There are an additional 27 illness symptoms. Combined, these provide an index of "Well-life expectancy" of which there are 43 functioning levels (Kaplan et al. 1976, McDowell and Newell, 1987, Kaplan et al. 1993). This would seem to support Anderson et al.'s (Anderson et al. 1989) description of it as measuring dysfunction as mental and social health are not measured. The QWB was designed for interview administration (15–35 minutes), although a shorter version has been developed which takes about 15 minutes (Kaplan et al. 1996). Interviewer training is required (Bombardier and Raboud, 1991). The preference weights were elicited using VAS scores which were obtained from a sample of the San Diego population. A linear transformation was then used to place these on a 0.00-1.00 scale (Kaplan et al. 1976, Kaplan et al. 1996). An additive model is used to compute the index. Extensive efforts to validate that VAS provides interval properties led to the release of a revised version (Coons and Kaplan, 1993, Kaplan et al. 1976, Kaplan et al. 1993). The upper boundary is 1.00, and the lower boundary is 0.00 (death equivalent) and health states worse than death are not permitted. Permission must be obtained to use the QWB and there are no costs for its use. Further information on the QWB can be obtained at: http://medicine.ucsd.edu/fpm/hoap/instruments.html

Health Utilities Index, Mark 3 (HUI3)

The Canadian Health Utilities Index (HUI3), for general population use, is based on the HUI2 which was designed for survivors of childhood cancer. To render it generic and overcome reported difficulties, it was revised into the HUI3 (Feeny et al. 1996b). The HUI1 has been superseded. The HUI3 measures "within the skin" functional capacity (Feeny et al. 1996b), a perspective adopted to enhance its use in clinical studies (Furlong et al. 2001). Social aspects of HRQoL are not measured. Items have 4 – 6 response levels. Twelve of the 15 items form 8 attributes (Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition and Pain). Designed for self-completion, Nord (1997) reported it took 2 minutes to complete, although 5–10 minutes is more likely given it has 15 items. The utility weights were elicited using the VAS, and scores then transformed based on four "corner" health states valued with the SG where a 60 year timeframe was used. These results were based on stratified sampling (n = 256; response rate 22%) of the Hamilton, Ontario, population (Furlong et al. 1998). A multiplicative function combines the attributes into the utility score (Furlong et al. 1998, Torrance et al. 1995). The upper boundary is 1.00, and the lower boundary is –0.36, permitting health states worse than death. Users must be registered and the instrument is available at a cost of CAN $4,000 per trial. Copies of the HUI3 and application forms can be found at: http://www.healthutilities.com/hui3.htm/

15D

The Finnish 15D was defined by Finnish health concerns, the WHO definition of health and medical and patient feedback (Sintonen, 2001; Sintonen and Pekurinen, 1993). It is concerned with impairment and disability of "within the skin" functions. There are 15 items, each with 5 levels, measuring Mobility, Vision, Hearing, Breathing, Sleeping, Eating, Speech, Elimination, Usual Activities, Mental Function, Discomfort & Symptoms, Depression, Distress, Vitality and Sexual Function (Sintonen and Pekurinen, 1993). Nord (1997) reported it took 5–10 minutes for self-completion. The weights came from five random samples of the Finnish population (n = 1290 respondents; response rate 51%) using VAS questions; responses were combined using a simple additive model (Sintonen, 1994; Sintonen, 1995). The upper boundary is 1.00, and the lower boundary is +0.11: death-equivalent and worse than death health states are not allowed. Permission must be obtained to use the instrument; however there are no costs for its use. The 15D has been translated into a number of European languages. Although there is no website devoted to the 15D, details can be obtained from http://195.101.204.50:443/public/15D.html

EQ5D (formerly the EuroQol)

The EQ5D (formerly the EuroQoL), developed by a team from 7 European countries (Rabin and de Charro, 2001, EuroQol Group, 1990), was based on the QWB (Kaplan and Anderson, 1988), the Sickness Impact Profile (Bergner et al. 1981), the Nottingham Health Profile (Hunt et al. 1981), the Rosser Index (Rosser, 1993), and group members' opinions. Designed for use in cross-cultural comparisons it has 5 items, each with 3 response levels, measuring Mobility, Self-care, Usual Activities, Pain/Discomfort and Anxiety/Depression. It takes 1–2 minutes to self-complete (Nord, 1997). The utility weights are from a British population random sample (n = 3395 respondents, response rate 56%) based on the TTO for 42 marker health states using a 10 year timeframe (Dolan, 1997). Other utility values were regression modelled (MVHGroup, 1995; Dolan, 1997; Dolan et al. 1996). The index is computed using an econometric regression model. The upper boundary is 1.00, and the lower boundary is –0.59: it permits health state values worse than death. Although the EQ5D is in the public domain for public health research, the EQ5D management group ask that researchers register their use of it. There are no costs for its use, unless it is used by commercial organisations. The EQ5D has been translated in many languages. Further information on the EQ5D can be obtained from: http://www.eur.nl/bmg/imta/eq-net/EQ5d.htm

Assessment of Quality of Life (AQoL)

The Australian AQoL used the WHO's definition of health, and items describe "handicap" as distinct from impairment and disability (Hawthorne and Richardson, 1996). The descriptive system has 15 items and 12 are used in computing the index (Hawthorne et al. 2001b). Each item has 4 levels. There are five dimensions: Illness (not used in utility computation), Independent Living, Social Relationships, Physical Senses and Psychological Well-being (Hawthorne et al. 1999). Designed for self-completion, Nord (1997) reported the AQoL took 5–10 minutes. A stratified sample (n = 350 respondents; response rate 72%) representative of the Australian adult population completed TTOs based on a 10 year timeframe to provide the utility weights (Hawthorne et al. 2000c). A multiplicative model is used to compute the

utility index (Hawthorne et al. 2001b). The upper boundary is 1.00, and the lower boundary is –0.04: it permits health state values worse than death. Permission to use the AQoL must be obtained, but there is no cost for its use. Further information can be obtained at: http://chpe.buseco.monash.edu.au/aqol. html#1

Due to a concern that the AQoL is insensitive at the upper end (i.e. for well health states), the AQoL research team are developing AQoL II for use in health promotion. As part of this development, AQoL II has been designed to enable the addition of disease-specific modules. One is currently being developed for the visually impaired.

SF6D

Two different algorithms have been published by deriving preference-based values from the SF-36 (Brazier et al. 1998, Brazier et al. 2002). They are referred to as the SF6D-1 and SF6D-2. The SF-36 descriptive system is American and the SF6D weights are British. The advantage of the SF6D procedures is that wherever SF-36 raw scores are available, the SF6D preference measure can be used.

Brazier et al. (1998) SF6D-1 drew upon 20 of the 36 items; these were selected to avoid double-counting. During scoring, items are combined into composites and each composite has 2–6 response levels. There are six sub-scales; Physical Function, Role Limitation, Social Function, Bodily Pain, Mental Health and Vitality. Utility weights were computed from VAS scores and modelled using SG values for three "link" health states. These values were derived from a convenience sample of 165 British respondents. An additive model computes the utility index. The upper boundary is 1.00, the lower boundary is +0.46: it does not permit poor health states, death equivalent or worse than death health state values.

The SF6D-2 (2002) uses 10 items from the SF-36: three from the physical functioning scale, one from physical role limitation, one from emotional role limitation, one from social functioning, two bodily pain items, two mental health items and one vitality item. These form 6 dimensions: Physical Functioning (PF: 6 levels), Role Limitation (RL: 4 levels), Social Functioning (SF: 5 levels), Pain (PA: 6 levels), Mental Health (MH: 5 levels) and Vitality (VI: 5 levels). Utility weights were computed from VAS scores, which were modelled using SG values for two link health states. Values were obtained from a random sample (n = 611; response rate = 45%) of the British population. An additive econometric model is used to compute the utility index. The endpoints for the SF6D are 1.00, and 0.30 for the worst possible health state. No website for the SF6D was identified.

Comparison of instruments

Hawthorne and Richardson (2001) outlined the axioms of utility measurement which MAU-instruments should conform to in order to possess basic validity. These axioms can be used as a checklist in instrument selection. They are:

- The use of a preference measurement to weight instrument items.
- Instruments must measure the dimensions of HRQoL deemed to be important. These are usually defined as physical, mental, social and somatic sensations (e.g. pain).
- There must be coverage of the full spectrum of HRQoL values, from full health states to values representing states worse than death.
- The combination rule for the utility index must prevent double-counting.
- There must be evidence of both weak and strong interval measurement.
- Instruments must be sensitive to the health states of interest. This requirement is covered in the next section. For general sensitivity comparisons between the instruments in the three validation papers published by Hawthorne et al. should be consulted (Hawthorne et al. 2001a; Hawthorne and Richardson, 2001; Hawthorne et al. 2001b).

An additional requirement is that:

- There must be evidence of valid and reliable measurement.

Use of a preference measurement technique to weight instrument items

Instruments using the SG or TTO may be regarded as possessing preference weights since both involve decisions under uncertainty. In the SG, the life outcome is uncertain (the probability of full health versus death). In the TTO, life-length is uncertain (how many life-years a person is willing to sacrifice).

There are doubts over whether ME delivers preferences because the procedure requires the respondent to estimate the divergence of a given health state from the "full" health state (which is assigned a value of 1.00). Once several given health states have been so assigned, the respondent is then asked to rank these in order (Gudex et al. 1993).

As reported above, there is doubt whether the VAS delivers preference measurement. Consequently it has been argued that the VAS has no place in economic theory (Brazier et al. 1999) and that untransformed VAS scores should not be used (Robinson et al. 2001; Torrance et al. 2001). It is recommended that VAS data should always be transformed based on TTO or SG (Brazier et al. 1999; Bennett and Torrance, 1996; Robinson et al. 1997); the transformation function that has been used was developed by Torrance et al. (1982). The preference measurement of instruments weighted with VAS scores therefore rests upon the validity of this transformation. For the EQ5D, Dolan et al. (1995) reported that the explanatory power of the transformations used was $r^2 = 0.46$, which was considered to be very good. However Sintonen (1995) reported that when applied to the 15D VAS data it assigned 12–25% of the adult population to values worse than death, a result he stated was "implausible". Bleichrodt & Johannesson (1997) noted that individual transformations were unstable; Robinson et al. (2001) reported difficulties with the transformations; as did Torrance et al. (2001).

Instruments weighted with a preference measure are the EQ5D (which used the SG) and the AQoL (the TTO). The Rosser Index relies upon ME. The HUI3 and the SF6D both rely upon transformed VAS scores; the extent to which these can claim preference weighting is dependent upon the validity of the transformations. Nord (Nord, 1993) has questioned the validity of the linear transformations for the QWB, arguing that one of the primary reasons its use in Oregon was so heavily criticised was that it lacked cardinal values. Given that the 15D uses untransformed VAS ratings there are doubts that it meets this requirement, although Martin (1996) argued that this gave the opportunity to quickly establish new weights for different populations — a procedure which Sintonen argued should be followed for each population from which study participants were drawn (Sintonen and Pekurinen, 1993).

Instruments must measure the dimensions of HRQoL deemed to be important

Important areas of HRQoL are usually defined as physical, mental, social and somatic sensations (e.g. pain). Unless instruments measure all these they cannot claim to be "generic". It should be remembered that the measurement of utilities was explicitly developed to enable cross-condition, health state and health care comparisons; by definition MAU-instruments are supposed to be generic.

Generally there are no published formal tests of content validity (Hawthorne and Richardson, 2001). Where this is mentioned, instrument developers have reported "face" validation, i.e. that instrument content as judged by the instrument developers "looks about right". For example, it has been argued the very restricted Rosser Index descriptive system makes it insensitive and provides a narrow band of responses (Hollingworth et al. 1995; Nord et al. 1993; Mulkay et al. 1987; Elvik, 1995). In a study of the EQ5D descriptive system it was reported that it only covers 39% of the concepts regarded by the public as salient to health (Dept. Health, 1995). Feeny et al. (1996b) reported that the HUI3 was valid because all levels of scores had been assigned at least once in population surveys. These various assertions do not engender confidence that the universe of utilities is actually measured by any of the instruments, a point which has been noted in the literature.

In three recent review articles Hawthorne et al. (Hawthorne et al. 2001b, Hawthorne et al. 2001a; Hawthorne and Richardson, 2001) mapped the content of MAU-instruments against the dimensions of 14 HRQoL instruments published between 1971 and 1993. Table 28 summarizes their work. This shows that even in the better instruments, coverage of the universe is limited. Some instruments offer very narrow measurement (for example, the Rosser Index and EQ5D), others have in-depth or duplicated measurement in particular areas (for example, the QWB, 15D and HUI3), and some offer very broad but sketchy coverage (for example, the AQoL and SF6D). Duplicated measurement may bias the obtained utility values. Two examples illustrate the problems. Despite its broad coverage, the QWB primarily measures pain and physical disability (Kaplan et al. 1993) yet does not include either social or mental health (Anderson et al. 1989), and analysis of the HUI3 showed it was a measure of physical impairment which did not adequately measure physical, social or mental dimensions (Richardson and Zumbo, 2000).

Table 28   **Content of MAU instrumentsa**

Content of MAU-instruments (a)

HRQoL dimensions (b)	AQoL	EQ5D	HUI3	15D	QWB (c)	Rosser Kind (d)	SF6D
Relative to the body							
Anxiety/depression/distress	*	*		**	**	*	**
Bodily care	*	*					*
Cognitive ability			*	*	*		
General health							
Memory			*				
Mobility	*	*	*	*	*		**
Pain	*	*	*	*	********		*
Physical ability/vitality/disability			*	*	*****	*	*
Rest and fatigue	*			*	**		**
Sensory functions	**		****	*****			
Social expression							
Activities of daily living	*	*		*			*
Communication	*		**	*	*		
Emotional fullfilment			*				
Environment					*		
Family role	*						
Intimacy/Isolation	*						
Medical aids use					**		
Medical treatment							
Sexual relationships				*	*		
Social function	*				*		*
Work function							*

Note:
a = Table shows only those items used in calculation of utility scores. Each asterisk represents an item. Based on item content examination.
b = Dimensions of HRQoL defined by a review of 14 HRQoL instruments, 1971–1993.
c = Excludes intoxication.
d = Areas subsumed within the two items: mobility, employment, housework.
   Source: Adapted from Hawthorne and Richardson (2001).

There must be coverage of the full spectrum of HRQoL values

This refers to instruments providing utility values from full health states to values representing states worse than death. There are two issues here. First, instruments must have combination rules permitting very poor HRQoL, irrespective of how this is caused. Second, the range of utility scores must cover the full spectrum.

Regarding combination rules, multiplicative models are to be preferred for the reasons outlined above. Instruments with multiplicative models are the HUI3 and AQoL. The EQ5D and SF6D rely upon regression models which are essentially additive in nature, and the 15D is an additive instrument.

The Rosser Index, EQ5D and HUI3 allow large negative values. Given the difficulty with the symmetry argument, these values are problematic. Hawthorne & Richardson (Hawthorne and Richardson, 2001) calculated that the effect of restricting the lower boundary for the HUI3 and EQ5D to 0.00, in population studies, would raise mean utility values by 9% and 14% respectively. This suggests the net effect of the symmetry argument is to overstate the value of interventions where people are in very poor health states. This problem does not apply to the QWB and AQoL which have lower boundaries at or near to 0.00.

The lower endpoints for the 15D (+0.11) and SF6D (+0.46 and +0.30 for the SF6D-1 and SF6D-2 respectively) raise other questions. Hawthorne & Richardson (Hawthorne and Richardson, 2001) reported these boundaries resulted in very different QALY estimates: a 1 QALY gain from the AQoL, EQ5D or HUI3, where a person was returned from the lowest quartile to full health for 1 year, implied a 0.50 and 0.37 QALY gain on the 15D and SF6D respectively (Hawthorne and Richardson, 2001). These contradictory results suggest that at least one of these of instrument groups is wrong.

As they allow the full range of scores, the QWB or AQOL instruments would be preferred, as would the 15D.

The utility combination rule must prevent double-counting

During construction of the Rosser Index, care was taken to ensure orthogonality between the dimensions (Rosser, 1993). Brazier et al. (1999) reported that for the QWB there is multicollinearity between the scales and symptoms. In papers describing the EQ5D there is no mention of this issue (EuroQol Group, 1990; Kind, 1996). Based on clinicians' opinions, structural independence was claimed for the HUI3 (Furlong et al. 2001); the factor analysis of the HUI3 published by Richardson & Zumbo, which revealed a lack of independence between the attributes, challenges this claim. Sintonen (1995) claimed independence for the 15D, although no evidence was provided.

For the SF6D-1, Brazier et al. (1998) used correlation analysis: where items were highly correlated only one was included. Brazier et al. noted that since an econometric model was used for the SF6D-2, preference independence, structural independence and double counting were unimportant. Yet the form of the SF6D-2 for the prediction of SG scores is essentially an additive model. Therefore, this argument seems extraordinary given that orthogonality to prevent double-counting caused by multicollinearity has been axiomatic of both psychometric and decision-making theory for over 50 years (von Winterfeldt and Edwards, 1986; Cattell, 1952).

For the AQoL, exploratory factor analysis was used during construction to ensure orthogonality (Hawthorne et al. 1999); the structure has since been confirmed by structural equation modelling (Hawthorne et al. 2001a).

There must be evidence of both weak and strong interval measurement

For meeting this criterion, all MAU-instruments rely on the presumed interval properties of the TTO, SG, or VAS. No instrument construction or validation paper has reported any formal testing of these properties. It has not been convincingly demonstrated that these properties are embedded within the TTO, SG and VAS (Brazier et al. 1999). Rosser (1993) argued that the magnitude estimation procedure used with the Rosser Index produced cardinal values; thus, like the EQ5D and AQoL, the Rosser Index may meet the weak interval requirement.

The weak interval property

VAS responses may be functions of adaptation, context, endpoints or anchor points, end-aversion and rating effects. These imply VAS may produce ordinal rather than interval data (Cook et al. 2001; Robinson et al. 2001; Richardson, 1994; Torrance et al. 2001). For the TTO and SG even less is known as these issues do not appear to have ever been properly investigated. Although Cook et al. (2001) challenged the claim of interval data for all three techniques, this was refuted by Hawthorne et al. (2003b) on account of some major methodological difficulties.

Subject to these caveats, Hawthorne & Richardson (2001) asserted it was likely the SG and TTO possessed interval properties given they allowed incremental probabilities (SG) or time fractions (TTO). On this basis, those instruments weighted with the SG or TTO should be preferred.

The strong interval property

This means that any given incremental value in HRQoL utility was directly equivalent to the same incremental value in life-length or life-probability. There is no evidence available for any of the MAU-instruments that they meet this requirement.

Valid and reliable measurement

The validity and reliability of various MAU-instruments has been assessed through either test of concurrent validity where monotonic relationships are sought, or test-retest. Additionally, there are issues around the stability of the utility values used in the different instruments due to sample bias.

Monotonicity refers to a relationship in which the instruments of interest group or mean scores progressively increase in line with a criterion measure. For example, if a sample of people suffers incontinence from "a few drops" to "no bladder control at all", then an instrument measuring this underlying health condition should report manifest scores that systematically increase with the level of incontinence. This does not imply, of course, that there will always be a 1:1 relationship between the two measures, for there will be individual variation.

Hawthorne et al. (2001b) examined monotonicity for the EQ5D, 15D, HUI3, AQoL and SF6D-1 against health status as defined by their sample strata of community random sample, outpatients and inpatients; they also examined the same instruments by combined utility quartile (Hawthorne et al. 2001a) and by instrument predictive power (Hawthorne et al. 2001b). In general their findings support monotonicity for all the instruments, although they did observe that the instruments formed two groups: those which correctly classified >50% of cases (AQoL, 15D and SF6D) and those which predicted <50% (EQ5D and HUI3).

Data on the Rosser Index are mixed. Although Rosser Index scores have been shown to match empirical and population general health data quite well when predicting the healthy/unhealthy dichotomies (Kind and Gudex, 1994; Nord et al. 1993), in a replication study it was shown that there are several health states where monotonicity is violated leading to difficulties with assigning logical QALY values (Gudex et al. 1993).

For the QWB there is mixed evidence regarding monotonicity. Kaplan et al. (1978) reported very high correlations with a number of chronic conditions, where the average was r = 0.96. Based on the revised version, similar correlations with chronic conditions have been reported (Coons and Kaplan, 1993; Kaplan et al. 1993). For example, Kaplan et al. (1995) reported a monotonic relationship between QWB scores and HIV-status; similarly monotonicity has been reported for functional status of children suffering cancer (Bradlyn et al. 1993). Against this the QWB has been criticised for producing QALY values that are non-monotonic. Thus a person wearing glasses is worse off than someone confined to a wheelchair, or curing five people with pimples would equate with saving one life (O'Connor, 1993; Nord et al. 1993). In a study of heart disease, non-monotonicity was reported for half the QWB scales (Visser et al. 1994).

The Hawthorne et al. (2001a, 2001b) results for the EQ5D (see above) were particularly interesting as they indicated that the EQ5D assigned too many cases to a utility value of 1.00, a finding consistent with earlier work by Brazier et al. (1993). Both research groups suggested this may have been due to the insensitivity of the EQ5D at the healthy end of the range and the consequent limited capacity to discriminate between those with full health and some health problems. At the other end of the range (very poor health states) Nord et al. (1993), in a study comparing Norwegian and Australian populations, reported that the EQ5D assigned excessively low values for some health states; a finding supported by Hawthorne et al. (2000a) who found that the EQ5D assigned 4% of a population sample to health states worse than death. In a comparison with the SF-36, Brazier et al. (1993) pointed out that the EQ5D correlated poorly with physiological symptoms, and Andersen et al. (1995) reported that the EQ5D assigned non-monotonic values for people with fractures: a person with a fractured arm was assigned worse utility than someone with a fractured vertebra.

Sintonen (1995) tested the 15D for monotonicity in five population-based samples, reporting that up to 2.5% of respondents valued health states inconsistently, rising to 20% who valued "death" higher than being "unconscious".

For the AQoL, other than Hawthorne et al.'s work there is as yet insufficient published material examining its properties for any conclusive assessment to be made. Hawthorne et al.'s papers (Hawthorne et al. [2001b], Hawthorne et al. [2001a], Hawthorne and Richardson, [2001]) described above all report monotonicity. Monotonicity has also been reported for cochlear implants (Hogan et al. 2001), the health

status of those with long-term depression (Herrman et al. 2002), suicidal ideation (Goldney et al. 2001) and depression in a population sample (Hawthorne et al. 2003a), and stroke (Sturm et al. 2002).

Test-retest reliability estimates have been reported for the QWB, 15D, EQ5D, HUI3 and AQoL. For the QWB, Kaplan et al. (1978) reported test-retest reliability at r = 0.93–0.98. In a study of chronic obstructive pulmonary disease, at 14-day separation, Stavem (1999) reported that the EQ5D and 15D test-retest reliability using Spearman correlations were r = 0.73 and r = 0.90 respectively. This result for the 15D is more encouraging than that reported by Sintonen (1994), who did not give a statistical estimate but stated that the agreement was not very good. In a study of stroke patients, Dorman et al. (1998) reported test-retest reliability estimates for the EQ5D of ICC = 0.83; and in a Dutch population study of the EQ5D where test-retest was carried out at 10-month intervals the correlation was r = 0.90 (van Agt et al. 1994). Studies of the HUI3 (Boyle et al. 1995, Feeny et al. 1996b), based on a community random sample with telephone follow-up, reported test-retest reliability where r = 0.77. For the AQoL, Hawthorne (2003), using random population sampling and mail/telephone comparisons reported the test-retest ICC = 0.83. An earlier study reported test-retest reliability for the AQoL descriptive system where a = 0.80 (Hawthorne et al. 1996).

Finally, and importantly, there are issues concerning the stability of the utility weights used in the various instruments. This concern stems from that fact that utility weights for most of the instruments – with the notable exception of the EQ6D where the sample size was 3,395 – were obtained from either small (e.g. 70 cases for the Rosser Index) or convenience samples (e.g. the 1,290 respondents for the 15D). In most cases, this was because of the cost of data collection: face-to-face interviews where SG or TTO questions are asked are costly. Because the SG or TTO is extremely tedious, all the instrument designers eroded their sample sizes further by breaking their health states up into sub-interview routines and then administering each sub-interview to a strata within the sample. This is commonly referred to as a "sort" procedure. The extreme case where this occurred was with the SF6D-2 (Brazier et al. 2002). The weights for the revised SF6D-2 were obtained from a representative sample of 836 English persons of which 611 interviews were used. Based on a sort procedure, each respondent was asked to value 6 health states out of a possible 249 health states. Altogether 3,518 valuations were made; there was an average of 15 responses for each health state (the range was from 8 for health state 5,3,5,6,4,6 to 19 for health state 1,3,1,5,4,2). Similar procedures were followed for the HUI3 (Furlong et al. 1998), AQoL (Hawthorne et al. 1997), and 15D (Sintonen, 1995), although in each case the numbers were greater than for the SF6D-2. For example, for the HUI3 the numbers for each health state varied from 19 to 246; for the AQoL the range was 70 through 225. These difficulties for each instrument were compounded by the relatively low response rates (typically about 50% although the AQoL's was higher).

These wafer-thin estimates raise fundamental questions concerning the transparency of utility scores, their stability and the generalisability of the instruments. In no case have instrument developers reported validation of the obtained utility results or published an analysis of these data. Given this, it is highly likely the utility values for all instruments, other than the EQ5D, are biased and lack transparency. Because of the restricted response rates and small sample sizes, utility weights may be less than stable; a problem compounded by the fact that all instrument weights have been derived using means rather than medians. Clearly, under these circumstances, claims for generalisability to many health conditions, including incontinence, should be interpreted cautiously.

A review of MAU-instruments used in incontinence studies

The previous section examined the evidence for criterion validity where the criteria were the axioms of utility and psychometric measurement. This section reviews the performance of the MAU-instruments in incontinence studies where the criterion is the sensitivity of the instrument to detecting differences between those who are incontinent and continent.

To identify published studies a search of Medline and Econolit was undertaken using the terms "utility" and "incontinence", as well as the names of the instruments reviewed. Ninety articles were identified. Review of the abstracts revealed 16 articles using utility measures; all were retrieved. Reviews showed six did not contain any utility data so these were discarded.

No published papers were found for the AQoL or SF6D. For the AQoL and SF6D this was unsurprising given their recent development. Unpublished Australian data were available for both these instruments and these data have been reported here (for the SF6D the calculated values are from Brazier's second algorithm and the SF6D is described as the SF6D-2).

Of the studies reviewed, there were three population surveys, four trials (two non-randomised), and a modelling exercise.

Studies excluded

As part of a validation study of the uretal stent symptom questionnaire (USSQ), Joshi et al. (2003) reported values for the EQ5D. The treatment group were 85 patients, and the controls were 25 healthy volunteers. The EQ5D was administered to the patients 4 weeks after stent insertion and again at 4 weeks after stent removal. The results were reported as medians and inter-quartile ranges: for the stent group (with stent) the median scores was 0.76 (IQR: 0.62-0.90), after removal it was 1.00 (0.80-1.00), compared with the controls; 1.00 (0.76-1.00). A confounding factor was pain which was associated with the indwelling stents. The extent to which the differences in EQ5D scores were due to incontinence is therefore uncertain. Although the median EQ5D score at 4-weeks after stent removal was the same as that of the healthy controls, a proportion of the treatment group still reported incontinent, so, thus suggesting that the primary cause of the low EQ5D scores with stents may have been pain rather than incontinence. The implication is that the EQ5D may not be particularly sensitive to incontinence.

Krahn et al. (2003) elicited utility values from 141 older males (mean age = 72 years) who had treatment for prostate cancer. The utilities were elicited, on average, at 4 years post-diagnosis. The utility values were stratified by UCLA Prostate Cancer Index Scores. Urinary function was then reported for the HUI3 and the QWB. The values for the HUI3 were 1st quartile 0.85, 2nd 0.76, 3rd 0.80 and 4th 0.76; and for the QWB they were 1st 0.71, 2nd 0.64, 3rd 0.64, and 4th 0.57. If it is assumed that those in the 1st quartile had no urinary incontinence (something that is not stated in article), then the differences attributable to incontinence would be 0.09 for the HUI3 and 0.14 for the QWB. The HUI3 difference was reported as being non-significant, while the difference for the QWB was statistically significant ($p < 0.001$). In the case of the HUI3, the lack of significance could be attributable to either large variation in scores within quartiles or to the lack of a monotonic relationship with increasing severity across the quartiles. A limitation of the paper is that no standard errors, standard deviations or confidence intervals were reported for these estimates; therefore this paper was excluded from further analysis.

Manca et al. (2003) used the EQ5D in a randomized study comparing tension-free vaginal tape (n = 117) with colposuspension (n = 97). EQ5D data were collected at baseline (means 0.78, 0.79 for the tension-free and colposuspension groups respectively), 6-weeks (0.79, 0.75) and 6-months (0.81, 0.79) follow-up. Unfortunately the data were not reported as means, medians and interquartile ranges. The EQ5D values for women who were continent at the end of the study were not reported. Therefore this study was excluded from further analysis.

One Australian paper utilising the Rosser Index (Foote and Moore, 2001), was brought to the attention of the author although it was not identified in the literature search. This study compared five different treatments for female incontinence, reporting utility improvements between 1-2% across the five treatments, and the costs per QALY gained were between AUD\$28,000-\$134,000. Insufficient details of Rosser scores were included in the paper for it to be included in this review.

Several other papers were reviewed and excluded. One study which reported data for the EQ5D where scores were obtained on the EuroQoL VAS which is not a utility measure (Fuertes et al. 2000) was excluded. (On the EuroQol VAS a respondent is asked to rank their health state on a 100-point scale.) The study by Ogawa et al. (1988) which reported a global utility rating was also excluded. Kobelt's (1997) study of willingness-to-pay which included the EQ5D and correlated scores with micturition (Spearman $r = -0.25$) reported the mean EQ5D score was 0.68. Since no further information was given in the paper, it was also excluded from more detailed analysis.

Procedures

In the interests of comparability, Cohen's effect size d (1988) has been calculated from the data in the various studies. Given the variability in treatment and comparator groups and how data have been reported, the full formula was used:

$$d = \frac{m_A - m_B}{\sqrt{\dfrac{\sigma_A^2 + \sigma_B^2}{2}}}$$

where:

m_A = the mean score of the incontinence group expressed in raw (original) measurement units;

m_B = the mean of the comparator group or comparator expressed in raw (original) measurement units; and

$\_$ = the standard deviation of either population, on the assumption that the sample standard deviation equalled $\_$.

Cohen provided the following classification for interpreting d: 0.20 = a small effect, 0.50 = a moderate effect and 0.80 = a large effect.

Based on the effect sizes, a modified version of the relative efficiency statistic (Liang et al. 1985; Wright and Young, 1997; Fayers and Machin, 2000) was computed, thus allowing the relative sensitivities to be examined:

$$RE_{\,Ins1vs.\,In2} = \frac{d_{In1}^2}{d_{In2}^2}$$

where:

$d_{In\,2}$ = the instrument with the smallest effect size; and

$d_{In\,1}$ = the effect size for the instrument of interest.

When interpreting 'd' statistics, it should be borne in mind that the data were badly skewed in almost all the papers reviewed, which is normal for utility values (most people are healthy). Cohen argued for the robustness of the d measure and provided an example where the mean difference was 2.0 and the standard deviation 2.8 with respect to the number of trials rats required to learn a maze (Cohen, 1988, p41).

The studies are reviewed in chronological order and the effect sizes and relative efficiencies presented in Table 29.

Results

The HUI3 was used in the Canadian National Population Health Survey in 1994/5. Mittmann et al. (1999) broke down the data by chronic conditions reported by the 17,626 participants (54% were women). When computing the HUI3 utility score, the researchers used HUI2 weights. To identify chronic health conditions respondents were asked: "Do you have (chronic condition) that has been diagnosed by a health professional?" Twenty-two persons reported urinary incontinence only (i.e. no comorbid chronic conditions); the number with incontinence and comorbidities was not reported. Although the researchers stated that those with urinary incontinence obtained the lowest HUI3 utility scores of any chronic condition (HUI3 utility score = 0.70, sd = 0.23; there were no differences by gender), this assessment was based on all cases including those where multiple chronic conditions were reported. When they reported condition specific cases, for those with "incontinence only" the HUI3 utility was 0.85, sd = 0.12. That there was such a discrepancy in the HUI3 scores by "incontinence only" versus "incontinence only" and "incontinence and comorbidities" suggests that most of the difference in HUI3 scores can be attributed to the effect of the comorbidities. For the purposes of reporting the sensitivity of the HUI3, comparison in this report was made between those with incontinence only and those without any reported chronic condition. This latter group are those reported as the comparator group in Table 29.

Stach-Lempinen et al. (2001) reported data for the 15D on urinary incontinent women, where they were exploring the relationship between the Urinary Incontinence Severity Score, a visual analog scale describing the burden of incontinence ("how bothered are you by incontinence at this moment?") and the 15D. Treatment participants were 85 women consecutively presenting for symptomatic incontinence treatment; the comparators were 29 healthy women with urinary leakage who did not want any medical intervention. Two sets of results were reported: 15D scores comparing the two groups, and a comparison of pre-post scores for a sub-sample (n = 49) of women who improved as a result of treatment (d = 0.95) at follow-up (13 months); for women who did not statistically improve (n = 12), d = 55. Separately reported were the scores on the item within the 15D measuring elimination.

The EQ5D was used to examine the costs of incontinence in a study modelling the cost-effectiveness of tolterodine and oxybutynin (O'Brien et al. 2001). The authors defined four levels of incontinence, based on micturition (normal: $M \leq 8$; mild: $M = 9\text{-}12$; moderate: $M = 13\text{–}15$; severe: $M \geq 16$) and assigned women who participated in three 12-week trials comparing tolterodine with oxybutynin to these levels. The reported frequencies across all three trials were 1%, 38%, 33% and 28%. They then assigned EQ5D scores based on a sample of 455 women from a Swedish study examining the measurement of economic outcomes from incontinence (Kobelt, 1997). After assigning the Swedish women's data to the same four classification levels, the resulting EQ5D values were assigned to the women from tolterodine/oxybutynin trials in order to estimate costs.

In a randomised trial of colposuspension (laparoscopic versus abdominal incision) carried out in Melbourne, 45 women were administered at 6 month follow-up after surgery the AQoL and SF-36 as part of an investigation into patient satisfaction with medical care (Hawthorne and Harmer, 1999). Although the data have yet to be published, permission to use the data was given by the authors. The data were analysed by the incontinence status of the women at follow-up; of the 45 cases 16 were still incontinent at the time of data collection. The definition of incontinent was where the patient suffered both stress and urinary incontinence at follow-up as defined by self-report. The SF6D-2 scores were computed from the SF-36 data.

In the 1998 South Australian Health Omnibus Survey (HOS, n = 3010) both the AQoL and SF-36 were used. For this report the SF6D-2 utility scores were computed. There were three HOS questions on incontinence (faecal incontinence; loss of urine when coughing, sneezing or laughing; and urge incontinence involving urine loss prior to reaching a toilet). These data have not been previously published and are presented here with permission from Professor Alastair MacLennan, Department of Obstetrics and Gynacology, Adelaide University. The HOS is a user-pays survey for health organisations, covering people aged 15+ years, involving random sampling from the SA population (Wilson et al. 1992). For this report, urinary incontinence was defined as those cases meeting the two questions on urine loss (n = 194) and faecal incontinence referred to those who endorsed the faecal incontinence question (n = 87). There were 23 cases who reported both; these cases have been assigned to faecal incontinence. The comparator was all cases not assigned to incontinence. These data, of course, ignore the presence or absence of comorbidities.

Schultz and Kopec (2003) re-analysed HUI3 data from the Canadian National Population Health Survey, 1996/7 (n = 73,402) with respect to 21 chronic conditions. The definition of a chronic condition was where a condition had lasted or was expected to last for six months or more and had been diagnosed by a health professional. When HUI3 utility scores were assessed for those with no comorbid conditions, incontinence was the third most severe condition after Alzheimer's disease and stroke. For those with suffering only from incontinence the mean HUI3 utility was 0.82 (sd = 0.46), but for those with incontinence aswell as other comorbidities, it was 0.61 (sd = 0.48). As with the Mittmann et al. (1999) study, the fact that there was such a discrepancy in the HUI3 scores by "incontinence only" versus "incontinence and comorbidities" suggests that most of the difference in HUI3 scores can be attributed to the effect of the comorbidities. For the purposes of reporting the sensitivity of the HUI3, comparison in this report was made between those with incontinence only and those without any reported chronic condition. This latter group are those reported as the comparator group in Table 29.

Table 29   Summary of literature reporting use of MAU-instruments in incontinence

Summary of literature reporting use of MAU-instruments in incontinence

	Study	*Year*	*Groups*	*Utility (sd) (a)*	*N*	*Effect size*	*Relative efficiency*
AQoL	SA HOS (c)	2002	No incontinence	0.84 (0.19)	2729		
			Urinary incontinence	0.71 (0.26)	194	0.57	2.98
			Faecal incontinence	0.58 (0.29)	87	1.06	
AQoL	Hawthorne/Harmer (c)	1999	No incontinence	0.78 (0.23)	29		
			Incontinence	0.67 (0.23)	16	0.48	2.12
EQ5D	O'Brien et al.	2001	No incontinence	0.74 (0.11)	6		
			Mild	0.72 (0.22)	209	0.12	
			Moderate	0.69 (0.27)	182	0.24	
			Severe	0.61 (0.38)	154	0.46	1.94
HUI3	Mittmann et al.	1999	No incontinence	0.93 (0.08)	7509		
			Incontinence	0.85 (0.12)	22	0.78	5.59
HUI3	Schultz & Kopec	2003	No incontinence	0.95 (0.08)	71773		
			Incontinence	0.82 (0.46)	195	0.39	1.40
15D	Stach-Lempinen et al.	2001	Comparator group (d)	0.91 (0.08)	85		
			Incontinence (baseline)	0.80 (0.09)	29	1.29	15.28
SF6D-2	SA HOS (c)	2002	No incontinence#	0.76 (0.13)	2729		
			Urinary incontinence	0.71 (0.13)	194	0.38	1.33
			Faecal incontinence	0.63 (0.15)	87	0.93	
SF6D-2	Hawthorne/Harmer (c)	1999	No incontinence	0.70 (0.10)	29		
			Incontinence	0.67 (0.08)	16	0.33	1.00

Notes:
* = Median (IQR)
(a) = Where the study reported 95%CIs, these have been calculated.
(b) = Calculated from published study data. The comparator is each case is the No incontinence cohort.
(c) = Unpublished data. See the text for an explanation.
(d) = Women who were incontinent but who were not seeking treatment.

Discussion of the results presented in Table 29

Examination of the studies shows that they involved very different populations and samples, and the definitions of incontinence varied. For example, although the AQoL South Australian and HUI3 Canadian data (Mittmann et al. 1999; Schultz and Kopec, 2003) were both population surveys involving self-report, the AQoL estimates were derived from non-clinical questions whereas the HUI3 estimates were reports of clinical assessments. Additionally, the estimates for the HUI3 exclude those with comorbidities whilst those for the AQoL make no distinction between those with incontinence alone and those with comorbidities. This difference, however, does of itself invalidate the comparison since Mittmann et al. (1999) stated that those with incontinence alone reported the worst HUI3 utility value. The Schultz HUI3 study (Schultz and Kopec, 2003) reported the differences between those with incontinence alone and those with comorbidities, the effect of which was to reduce HUI3 values by 26%.

These differences must be kept in mind when interpreting the data in Table 29. Although caution should be exercised, the analysis may be useful as a guide to instrument selection. Generally, the obvious conclusion is that although incontinence has a significant effect on people's lives, it is not catastrophic. The table shows that all instruments report utility values for incontinence that are in the top half of the utility range. Indeed for urinary incontinence, the lowest utility value was 0.61 for those with severe incontinence on the EQ5D.

Regarding the effect of incontinence, under the axioms of utility theory it is possible to model the number of people who need to be treated to make a 1-QALY gain where it is assumed that the treatment effects last for one year

$$n = \frac{1.00}{U_{FH} - U_{IN}}$$

where U_{FH} is the mean utility value of non-incontinence and U_{IN} is the incontinence utility mean. The results show that a 1-QALY gain could be obtained by treating 7.69 people (EQ5D; 95%CI: 3.60-•), 7.69 (AQoL, SA HOS study; 95%CIs: 5.56-12.50); 7.69 (HUI3, 95%CIs:5.12-15.42 (Schultz and Kopec, 2003)), 9.09 (AQoL, colcosuspension study; 95%CI: 2.13-•), 9.09 (15D; 95%CI: 6.25–16.67), 12.50 (HUI3, 95%CI: 7.58-35.65 (Mittmann et al. 1999)), 20.00 (SF6D, HOS study; 95%CI: 14.29-50.00) and 33.33 (SF6D, colcosuspension study; 95%CI: 9.31-•). Although the very broad confidence intervals are due to the small numbers in some of the studies, these findings suggest that among those with incontinence there are large variations in how it affects people's lives.

There is also considerable variation in the effect sizes due to the large standard deviations. The range from was 0.33 (SF6D) to 1.29 (15D). When interpreted using Cohen's criteria (Cohen, 1988) the SF6D is capable of detecting a small effect, the EQ5D a moderate effect, and the HUI3, AQoL and 15D are capable of detecting large effects. That the 15D obtained the largest effect size is not surprising given its small standard deviations and its specific question on elimination.

The relative efficiencies also show marked differences. The data suggest that the least efficient instrument at detecting differences between incontinent and continent cases was the SF6D. Relative to the SF6D the efficiency of the other instruments was, in order, the EQ5D, the HUI3 and the AqoL, which were probably similar, and the 15D.

Finally, it should be noted that there are a number of inconsistencies between the utility values assigned across the different instruments for continent and incontinent cases. The mean incontinence score for the EQ5D (0.61) was approximately the same as faecal incontinence as reported by the SF6D-2 (0.63) when it might be expected that faecal incontinence would be worse than urinary incontinence (as shown on both the AQoL and SF6D-2). Another inconsistency, based on utility values that were derived from population estimates in Canada and Australia, is between the HUI3 (0.85 (Mittmann et al. 1999) and 0.82 (Schultz and Kopec, 2003) for incontinence)), the AQoL (0.84 for continent), the EQ5D (0.74 for continent) and the SF6D (0.76 (SA HOS) and 0.70 [colcosuspension study]).

There is also an inconsistency between the 15D (0.80 for incontinence), the SF6D-2 (0.76 and 0.70 for continent), the EQ5D (0.74 for continent) and the AQoL (0.78 for continent).

In addition to these findings from Table 29, the percentage gains reported in the Foote & Moore (2001) study using the Rosser Index, suggest that a 1-QALY gain could be obtained through the treatment of between 50-100 cases. This would suggest the Rosser Index is likely to be less sensitive in incontinence studies than the instruments above.

Although the inconsistencies reported above reflect the different descriptive systems, assigned weights, scoring mechanisms and study populations, they also suggest the different instruments deliver very different estimates: the results for the different instruments cannot all be right. When taken in conjunction with the differences in implied QALYs, effect sizes and relative efficiencies, they are suggestive that for similar studies the different instruments may give very different results when used in incontinence cost-utility analyses. The implications are extremely worrying as they suggest that study results may depend upon the instrument chosen rather than actual treatment benefits.

Recommendations

There has been so little work carried out in the area of incontinence and utility measurement that any recommendation is very speculative. The results of this review indicate that none of the existing MAU-

instruments meet al.l the requirements of utility theory. Additionally, available data show that there is a great deal of inconsistency among those instruments that have been used in incontinence studies. In short, no instrument can be recommended as the "gold standard" at this point.

The recommendations below have been framed by the fact that there are three levels at which utility instruments could be used in incontinence studies: (a) clinicians working in clinical practice, (b) specialists working in clinical practice, and (c) researchers or program evaluators.

At the clinical level, measurement is usually related to clinical management of individual patients and there are time and data collection issues which impact on recommended practice. Any instruments used at this level must possess sufficient nomological evidence to be used at the case level; i.e. for individual patient assessment. Additionally, at this level, data collection should be as brief and as possible and there should be few data analysis demands upon clinicians.

Under (b), data collected need to be sufficient to meet the needs of specialists. Whilst these include the requirements of clinical measurement, specialists need more information and are often involved in research or evaluation.

Researchers and program evaluators' needs centre around data that are useful for answering research questions where analyses are group-based; where data collection procedures may be remote; and where findings are aimed at demonstrating the effect of new treatments or at influencing policy decisions.

MAU-instruments, by definition, were designed for use by researchers undertaking economic evaluation. However, this does not necessarily imply that they have no role to play in clinical or specialist services. At the individual level, MAU-instruments may provide HRQoL profiles based on responses to individual questions or utility scores which may be compared to group or population norms. Additionally, in a health care system committed to evidence-based practice, basic data should be collected and held at the clinician level for local analysis as well as transference to research (e.g. for inclusion in incontinence monitoring or surveillance).

Summary comments

In general, conclusions drawn from this review should be placed in the following contexts which are germane to using MAU-instruments in incontinence studies.

- Instrument length. Given that the chosen MAU-instrument is likely to be used in an instrument battery and that longer batteries place higher cognitive demands on respondents who may be in frail health states, it would seem that short instruments should be considered. These would include the EQ5D, AQoL, HUI3 (both of which could be shortened to just the 12 items contributing to the utility score) and the 15D.
- Coverage. There is a clear difference between the utility instruments in relation to their coverage. If instruments providing "within the skin" coverage are to be preferred, the choices would be between the 15D and HUI3. On the other hand, if the "social expression" of HRQoL is desired, the AQoL or SF6D would be the instruments of choice.
- Administration. Recommended national instruments are likely to be used in a variety of settings, particularly in studies where data are collected through self-completion (whether in mail or interview situations) and where follow-up data are likely to be collected by mail or telephone. Instruments which require interviewer administration are therefore probably not recommended. Given the cognitive demand of telephone interviews, longer instruments with more complex item responses should also be avoided. This would preclude the use of the QWB and Rosser Index; and perhaps the 15D. Instruments that are more suitable for telephone administration are the EQ5D, AQoL and HUI3.
- Ease of use. Instruments which respondents find simple and easy to use are the EQ5D, HUI3, 15D, AQoL, and Rosser Index.
- Time to complete. To reduce the burden on participants and the costs associated with data collection this should be as short as possible. Those instruments taking less than ten minutes are the EQ5D, HUI3, 15D and AQoL.
- Translation. Translations will almost certainly be required for some sample sub-groups given the heterogenous Australian population. Because the SF6D relies upon the SF-36, the SF6D would be regarded as having been widely translated; i.e. SF6D scores can be obtained from any language into which the SF-36 has been translated. The only MAU-instrument, per se, that is readily available in a number of languages is the EQ5D. The 15D has been translated into several European languages.

- Ease of scoring. Although this does not directly impinge upon data collection, it does have some implications for data analysis where research groups may not have ready access to either a statistician or instrument technical support. The simpler the instrument the better. The simpler instruments regarding scoring are the 15D, EQ5D, AQoL and HUI3.
- Sensitivity. This is important given that in some situations the critical effect sizes for some incontinence interventions are likely to be small. Those instruments likely to be more sensitive are the 15D, AQoL and HUI3.
- Reliability. All the instruments reviewed are likely to possess similar reliability characteristics. However, this has not been fully investigated for all instruments.
- Validity. All the instruments reviewed have some questions about their validity. This has not been satisfactorily established and published for any of the instruments, particularly in relation to the generalisability of the utility weights – with the exception of the EQ5D – and the necessary strong interval property. Based on the available literature it would appear there are potential difficulties with the Rosser Index and QWB. There may also be some issues around sensitivity for the EQ5D, and some doubt as to whether the 15D actually measures utilities.
- Utility axioms. None of the instruments reviewed meet al.l the requirements for utility measurement at this time. However, the review suggests that those instruments with the better claims for meeting these axioms would be the HUI3 and AQoL, then the EQ5D and perhaps the 15D.

The following table provides a summary of the findings from this study. Each of the instruments reviewed was assessed against the descriptions and validity evidence presented in this report. For each of these criteria, the assessment was made on a 3-point scale where a low score indicated minimally meeting the criteria and a high score indicated mostly meeting the criteria. Additionally, following discussions with Jan Sansoni and A/Professor Shane Thomas, each of the criteria was weighted according to its perceived relevance to the Australian context (for example, it was decided that although having multiple language versions was advantageous, in the Australian context where English is almost universally spoken this was not an important criteria for instrument selection). The results suggest that the instruments of choice would be the AQoL, EQ5D and HUI3.

Table 30   Summary of ratings for multi-attribute utility instruments

	Tool						
Criteria	**EQ5D**	**AqoL**	**HUI3**	**15D**	**SF6D**	**QWB**	**Rosser**
Availability of comparison data/usage	3	2	2	1	1	2	2
Length, ease and time to complete	3	2	2	2	1	1	1
Method of administration	3	2	2	2	3	1	1
Translations available	3	1	2	1	3	1	1
Ease of scoring	3	3	3	3	3	2	1
Sensitivity to incontinence	1	2	2	3	1	1	1
Reliability evidence available	2	2	2	1	1	1	1
Validity evidence available	2	2	2	1	1	1	1
Adherence to psychometric axioms	2	3	3	1	1	1	1
Cost of using the instrument	2	3	1	3	3	3	3
Weighted Total	**55**	**54**	**51**	**42**	**38**	**33**	**31**

EQ5D	European Quality of Life Measure – 5D (formerly the EUROQOL)
AqoL	Assessment of Quality of Life
HUI3	Health Utilities Index – Version 3
15D	Fifteen-Dimensional measure of health-related quality of life
SF6D	No acronym available
QWB	Quality of Well Being
Rosser	Rosser Quality of Life Index

Of the tools studies in the Multi-Attribute Utility instruments category, three obtained the requisite 47-point score or higher. These were the Assessment of Quality of Life (AQoL), the European Quality of Life Measure-5D (EQ5D) and Health Utilities Index (HUI3). All three tools are recommended.

Recommendations

There are, therefore, a number of options which could be considered either individually or collectively.

1. A single MAU-instrument could be recommended as the preferred instrument of choice for routine use at the clinician- and specialist-levels. This instrument should be short, easy to administer and score and population norms could be made available for easy reference. If such a policy was adopted, it would be in light of the limitations outlined in this report and there would be no guarantee that results obtained would be comparable with results obtained elsewhere using another instrument. Indeed, where QALYs were computed as the result of a treatment, it is likely these would reflect instrument choice as much as treatment effect. Where two MAU-instruments were recommended as the preferred measures, these difficulties would be compounded if some studies included one of the instruments and other studies opted for the other instrument.
2. To overcome this uncertainty, it could be recommended that two MAU-instruments be included in any particular research or evaluation study, and that researchers be encouraged to provide both sets of results. One of the recommended instruments should be that recommended for clinician use. This strategy would have the benefit of reducing the bias inherent in a one-instrument strategy, and it would produce a range of estimated benefits from interventions, thus acknowledging the limitations of relying upon any particular existing MAU-instrument. Given that, inevitably, comparisons will be made with incontinence studies overseas, this strategy would have the further benefit of enabling cross-cultural comparisons.
3. Several instruments could be trialled in 3-4 large incontinence studies for the explicit purpose of identifying the instrument to be recommended for future use. Whilst this would impose an immediate burden for, say, 3 to 5 years, it would enable many of the questions raised in this report regarding the validity of MAU-instruments to be thoroughly investigated in an Australian context. This would place Australia in a position of world leadership in incontinence and utility research; it would enable a fully informed decision to be made regarding instrument selection; and it is likely the Australian model would become the world standard in the immediate future given the paucity of current research in the field. Should this latter scenario eventuate, it is likely this would enhance international cooperation in the field.
4. As an alternative to #3, the multi-attribute utility instruments that were considered above could be included in the 2004 South Australian Health Omnibus Survey (HOS), together with suitable questions on incontinence and incontinence-related health sequelae. This would enable the rapid collection of data and its analysis leading to instrument selection and recommendation. Since the HOS involves drawing a weighted population sample, the findings could be used to establish population norms against which future work could be interpreted. (NB: This work is currently in progress.)
5. A specific study could be funded to develop an incontinence module for attachment to a generic MAU-instrument descriptive system. This recommendation arises from the consideration that there are HRQoL areas of concern to those with incontinence that are not addressed with fully generic instruments. If an incontinence module for an existing instrument were constructed, researchers would be in a position to report both incontinence-specific HRQoL effects and generic utility scores. This model has been followed by the SF-36, for which there are now many disease-specific modules, and it is being followed by the World Health Organisation Quality of Life Group for the WHOQOL-OLD (being specifically developed for use with older adults) (see Murphy & Hawthorne, 2001), and also in Australia in the area of visual impairment and the AQoL. The chief difficulty lies in selecting the base instrument.

Recommended Multi-Attribute Utility Instruments

Based on this review the following Multi-Attribute Utility Instruments are recommended across the different practitioner and researcher categories.

Table 31   Recommendations for multi-attribute utility instruments

User category	Recommended Multi-Attribute Utility Instruments
Primary Care Practitioner	None recommended
Specialist Practitioner	1. EQ5D 2. AQoL 3. HUI3
Researcher	1. EQ5D 2. AQoL 3. HUI3

Generic Health-Related Quality of Life Measures

The two generic health-related quality of life/general health status measures highly recommended by the ICS/WHO working party were the SF-36 and the EQ-5D. The Nottingham Health Profile and the Sickness Impact Profile were also recommended. The full reviews for these tools are located in the technical instrument reviews appended to this report. The following text is a summary of the main elements of these reviews but the reviews should be referred to if additional detail is required.

Many studies have compared various health status and generic health-related quality of life measures through multiple administration of a battery of tests to the same patient population.

Andresen, Patrick, Carter and Malmgren (1995) compared the performance of the Sickness Impact Profile, the Quality of Well Being Scale and the SF-36 in a sample of 200 people aged 65 and older in New York. These scales were subjected to both internal consistency and test-retest checks with a one year test-retest interval. Using ICC coefficients between base line values and the one year results, the correlations for the tools were found to vary from r = 0.51 to 0.73. Satisfactory internal consistency outcomes were also obtained. The researchers, on the basis of the results obtained in their study, did not make a specific recommendation for the use of one tool over another.

Beaton, Hogg-Johnson and Bombardier (1997) studied the performance of five generic health status tools, including the SF-36, the Nottingham Health Profile, the Health Status section of the Ontario Health Survey, Duke Health Profile and the Sickness Impact Profile for 127 workers at varying stages after work related injury. Analysis of the performance of the tests led the researchers to recommend the SF-36 as the most appropriate tool to measure change in health status.

Edelman, Williams, Rothman and Samsa (1999) compared the Health Utilities Index Mark 11 with the SF-36 and the Sickness Impact Profile in a general medical outpatient population study. A sample of 160 patients from the Durham Veteran Affairs Medical Centre took part in the study. As this was a cross-sectional study no data concerning test-restest reliability was provided. The researchers made no specific recommendations as to the adoption of one instrument over the others.

Langfitt (1995) employed the SF-36 and the Sickness Impact Profile and the Washington Psycho-social Seizure Inventory in a study of patients with epilepsy. Based on an analysis of the psychometric performance of the three measures, Langfitt recommended the adoption of the SF-36 and the SIP in studies of the health status and quality life for people with epilepsy.

In Washington, Martin, Engelberg, Agel, Swionthowski (1997) studied the psychometric performance of the musculo-skeletal function assessment questionnaire, the SF-36, the Western Ontario and McMaster Universities' Osteo-arthritis Index and the Sickness Impact Profile. On the basis of the psychometric analysis of their study results, the researchers recommended the use of the musculo-skeletal function assessment in assessing health status of patients who have a musculo-skeletal disorder.

Smith, Taylor and Mitchell (2000) conducted a comparative study of the psychometric properties of four quality of life instruments in a group of 22 patients who had experienced myocardial infarction,

or coronary artery bypass graft. The SF-36, the Quality of Life Cardiac Version, the Quality of Life after Myocardiol Infarction Questionnaire and the schedule for the evaluation of the individual quality of life were used to evaluate quality of life at the beginning of rehabilitation and then six weeks' later. The researchers concluded that all of the measures lacked sensitivity to change but this study was based on a very small and specific sample. By contrast Dempster and Donnelly (2000) undertook a literature review to examine which generic and disease specific measures were most suitable to recommend for use with people with ischaemic heart disease. They recommended the SF-36 as it appeared to offer the most reliable, valid and sensitive assessment of quality of life in comparison with the Nottingham Health Profile and the Sickness Impact Profile.

Essink-Bot, Krabbe, Bonsel & Aaronson (1997) conducted a study involving four generic health status measures, including the Nottingham Health Profile, the SF-36, the COOP and EuroQol. Based on their psychometric analyses in the cross-sectional study, the researchers recommended the adoption of the SF-36, because of superior reliability and validity and responsivity.

In addition to the now rapidly growing set of studies in which multiple health status/health-related quality of life instruments have been administered to a range of patient populations, there is now a growing set of reviews concerning the relative merits of several of these measures. Coons, Rao, Keininger and Hays (2000) in a comparative review of generic quality of life instruments, note that hundreds of generic and specific health-related quality of life instruments have now been developed. Their review examined seven generic health-related quality of life instruments, including:

• The SF-36
• Nottingham Health Profile
• Sickness Impact Profile
• Dartmouth Primary Care Co-operative Information Project Charts (COOP)
• Quality of Well Being Scale
• The Health Utilities Index
• The EuroQol instrument (now the EQ5D)

The researchers did not recommend one particular instrument reliably over the others. However, most authors note that the SF-36 has sound psychometric properties and that it is the most commonly used health-related quality of life measure. They also note that the Dartmouth COOP charts were designed to be used in every day clinical practice but that they were not widely used (actually they are rarely used).

The SIP and NHP are also "count" or "Yes/No measures" rather than using a conventional rating scale format that permits a range of response levels in relation to the target attribute. This lessens the sensitivity of these instruments. The SIP is also a lengthy instrument and appears more complex to score.

Although there are many of these comparative studies and reviews, even taken as a body of work together, they are not particularly compelling from an evidence based viewpoint as to the selections that should be made between health-related quality of life measures. In any event, many of the reviews are focused on the psychometric properties of the tools and, in particular, upon reliability and validity issues. It is remarkable to note that many of these reviews do not include consideration of the issues of applicability and practicability, and when they do, these constructs are often narrowly defined.

While practicability may be defined as the ease of use of application of the particular tool, there are also other considerations. For example, practicability may be impacted by the ability of the clinician, or the researcher, to benchmark their results against known industry standards. Under these criteria, the most widely applied tools would be the most practicable.

The Short Form-36 Health Survey Questionnaire (SF-36)

The SF-36® Health Survey is a multi-dimensional profile measuring each of the following eight health concepts:
(1) physical functioning;
(2) role limitations because of physical health problems;
(3) bodily pain;
(4) social functioning;
(5) general mental health (psychological distress and psychological well-being);
(6) role limitations because of emotional problems;
(7) vitality (energy/fatigue); and
(8) general health perceptions.

The SF-36 also includes a single-item measure of health transition or change. The SF-36 can also be divided into two aggregate summary measures the Physical Component Summary (PCS) and the Mental Component Summary (MCS). (In the standard version of the SF-36 all scale questions refer to a 4 week time period and there is an acute one-week version also available.)

The SF-36® Health Survey items and scales were constructed using the Likert method of summated ratings. Answers to each question are scored (some items need to be recoded). These scores are then summed to produce raw scale scores for each health concept which are then transformed to a 0 – 100 scale. Scoring algorithms can then be applied to produce the PCS and MCS scores. (These two summary scores have the major advantage of being norm based. They also have reduced floor and ceiling effects.)

The SF-36® Health Survey developed out of work on the Medical Outcomes Study or RAND Health Insurance Experiment. It is a short-form derived from a larger 149-item instrument and is more precise than its predecessor the SF-20.

An Australian version of the SF-36 has been in use for many years (1997). Australian normative data for Version 1 of the SF-36® Health Survey is provided by Stevenson (1996) and from the Australian Bureau of Statistics National Health Survey, 1995 (ABS 1997). The ABS data is now the accepted normative data for use in Australia. Additional population health data using the SF-36 Version 1 can be found in the 1996 Australian Longitudinal Study on Women´s Health (Mishra and Schofield, 1998), the 1999-2000 Australian Diabetes, Obesity and Lifestyle Study (Chow et al. 2003), the 1998 National Drug Strategy Household Survey (Adhikari et al. 1998), the 1991 – 2002 SA Health Omnibus Survey (Behavioural Epidemiology Unit, 1995), the 1997 + 1998 NSW Health Survey (NSW Health Department, 2000a), the 1999 NSW Older People´s Health Survey (NSW Health Department, 2000b) and the 2002 Australian National Sexuality Survey (Purdie et al. 2002).

US Data for the SF-36 can be found in Ware, Kosinski & Keller (1994) and Ware, Kosinski, Bayliss, McHorney, Rogers & Raczek (1995). UK Data for the SF-36 can be found at Jenkinson, Coulter & Wright (1993) and Bowling, Bond, Jenkinson & Lamping (1999). International Data for the SF-36 in order to make cross country comparisons can be found in Ware, Gandek, Kosinski, Aaronson, Apolone, Brazier et al. (1998).

Recently Ware et al. (2000) have produced a revised and updated version of this instrument (Version 2/SF-36V2). They have refined the SF-36 Version 1 to develop a more "international" version given concerns over the wording of some items, the psychometric features of the role functioning scales and the survey's layout. Sansoni and Costi (2001) have indicated that the revisions encompassed in Version 2 of the SF-36 appear to have led to a greater precision of measurement for Version 2 as compared with Version 1. Thus as Jenkinson et al. (1999) indicated it is likely to be more responsive to change in health status which is an important feature for instruments used in outcomes research. The changes to wording and format should also make it a more appropriate instrument for use particularly with such groups as the elderly. The adoption of norm based scoring should also aid clinical interpretation (Sansoni & Costi 2001). An Australian edition of Version 2 is available from the Australian Health Outcomes Collaboration (AHOC). The only difference between the two editions however is the conversion from the imperial system to the metric system in a couple of items.

However, as Sansoni and Costi (2001) indicate the major problem with the use of Version 2 in Australia is that no Australian normative data had yet been collected and thus the instrument could not be scored using the norm based scoring methods that have been adopted for Version 1. Nor could the PCS and MCS summary scales be derived for this version of the instrument until such time as Australian normative data becomes available for the instrument. A study is currently underway in Australia (Harrison Health Research, 2004) to collect interim normative data for SF-36V2 and this data should be available by mid 2005. In view of this, and given that the revised role functioning scales are less prone to ceiling and floor effects (which are relevant domains for people with incontinence) the project team is recommending the use of Version 2 in preference to Version 1. However, either version of this instrument is suitable for use.

SF-12

The SF-12® Health Survey includes 12 questions from the SF-36® Health Survey (Version 1). These include: 2 questions concerning physical functioning; 2 questions on role limitations because of physical health problems; 1 question on bodily pain; 1 question on general health perceptions; 1 question on vitality (energy/fatigue); 1 question on social functioning; 2 questions on role limitations because of emotional problems; and 2 questions on general mental health (psychological distress and psychological well-being). This tool was developed for users who need an even shorter generic measure of perceived health status.

The scoring of individual items is identical to the SF-36® Health Survey. Scoring algorithms are then applied to produce the PCS and MCS scores. The SF-12® Health Survey was developed using normative data for the SF-36® Health Survey in the United States (see Ware, Kosinski & Keller, 1994 and Ware, Kosinski, Bayliss, McHorney, Rogers & Raczek, 1995). Wilson, Tucker & Chittleborough (2002) and Sanderson & Andrews (2002a) have conducted local equivalence studies and found the SF-12 suitable for use in Australia.

Population health data using the SF-12 can be found in the 1997 Australian National Survey of Mental Health and Well-being (Sanderson and Andrews, 2002a), the 2000 Mental Health Status of South Australian Population Study (Taylor et al. 2000), the 2002 Longitudinal Investigation of Depression Outcomes (LIDO) Study (Herrman et al. 2002) and the 2003 Australian Gulf War Veterans' Health Study (Sims et al. 2003).

In choosing between the SF-12® and the SF-36® Health Surveys users should consider the trade-off between test taker burden (i.e. number of questions, time to complete) and the precision of scores (i.e. how reliable does the obtained score need to be). Ware et al. (1996) reports that there is a 10% loss in the SF-12´s ability to distinguish between different disease groups as compared to the SF-36 and that the SF-12 less accurately reproduces the eight scale profile of the SF-36. Therefore it is recommended that the SF-36 be used for smaller studies (less than n = 500). A recent paper by Rubenach, Shadbolt, McCallum & Nakamura (2002) highlights this important distinction for clinical research studies.

Sanderson & Andrews (2002b) have done considerable work in utilising the SF-12 (MCS) as a disability measure for mental health disorders (especially anxiety and depression). Salyers et al. (2000) have utilised the SF-12 (MCS) for severe mental illness.

The SF-12 has been administered using interactive voice recognition technology and in computerised format, telephone vs. mail-out administration has also been compared. An acute (1 week) version of the SF-12® Health Survey is also available.

Like the SF-36® Health Survey, the SF-12® Health Survey has been recently updated by QualityMetric Incorporated. The new version is known as the SF-12V2™ Health Survey (Version 2). However, this update of the SF-12 has yet to be field tested in Australia for equivalence or new norms developed for the Australian population.

Conclusions Concerning the SF Family of Instruments

Licence fees are now payable for these tools (SF-36® and SF-12®), depending upon the proposed purpose. For individual clinical assessment, the SF-36 is recommended over the SF-12 because of its superior psychometric performance. Also see Sansoni and Costi (2001) for a more detailed discussion of related matters concerning the appropriate version of these tools to use. Chapter 2 of this report discusses the crucial importance of this performance in individual clinical studies especially those in which the desire is to measure or track change of individuals in health intervention programs. The SF-12's performance is not strong enough for this purpose and it should be reserved for population studies.

EQ5D

The EQ5D is reviewed in the MAU HRQoL section of this report. This material is not duplicated in this section. It is best considered to be a multi-attribute utility instrument.

Sickness Impact Profile

The Sickness Impact Profile (Deyo et al. 1982; 1983; Bergner, 1993), like the SF-36 was developed for use as a generic outcome measure (Bergner et al. 1976a; 1976b; 1981; Gilson et al. 1979).

The questionnaire has 136 items covering 7 dimensions and 12 categories – Physical Dimension (includes body care and movement, ambulation mobility); Psychosocial Dimension (emotional behaviour, alertness behaviour, social interaction, communication), and the independent dimensions of Sleep and Rest, Eating, Home Management, Work, Recreation and Pastimes. The focus of the items is on behaviours rather than feelings. It takes 20 – 30 minutes to complete.

A lower score indicates better quality of life. The 12 categories can be scored separately, 2 dimension scores can also be obtained (for physical and psychosocial dimensions). A SIP percent score can be calculated for all categories and for an overall SIP percent score. It takes approximately 10 minutes to score. As respondents only endorse applicable items it may be hard to differentiate a "no" response from missing data (Bowling, 1997).

It can be applied to populations with a wide range of levels of sickness; it seeks to be applicable across countries, age groups and clinical conditions. Healthy people or those with mild health problems score very low on this instrument – it is known that a normal population may only score 2-3/100 on this instrument and thus the instrument may be less sensitive to detecting health problems in those with minor ailments. Hall et al. (1987) reported that 18% of patients in a general practice setting scored 0 on the SIP total percent score. McKenzie et al. (1986) reported that it is more sensitive to deterioration rather than improvements in health status. Its length, also, may limit its use in some service settings.

Normative data is provided in the SIP User Manual. Only a few studies have used the SIP in Australia (Hall et al. 1987; Harper et al. 1995; Katsikitis et al. 1996; Gardner and Sibthorpe, 2002). However, the SIP has been used widely to assess health status for patients with many clinical conditions (see the review sheet).

The SIP requires patients to endorse applicable items. The endorsement can only be "yes" – there are no levels of response available for any item although item weights reflect the relative severity of limitation implied by each statement. This is unlike most other scales in which the user can provide ratings of the target attribute. This feature reduces the potential sensitivity of the scale, which is important for an outcome measure.

Hunskar has used the SIP in a number of studies of incontinence. For example, Hunskar & Vinsnes (1991) reported the outcomes of a small study of 36 women aged between 40 and 60 years and 40 women aged 70 years plus who were attending an incontinence clinic. The Sickness Impact Profile questionnaire was applied to the group. However, the psychometric performance of the profile was not reported in this study.

Nottingham Health Profile

The Nottingham Health Profile (NHP) was developed in the UK and was based on interviews with community samples concerning the effects of illness on behaviour. It has been much less frequently used in the study of incontinence than tools such as the SF-36 or the EQ-5D.

The NHP was designed to provide a brief indication of a patient's perceived emotional, social and physical health problems. Items ask about feelings and emotional states directly rather than via changes in behaviours – as contrasted with the SIP (McDowell and Newell, 1996). It is a self-administered survey but it can also be interview administered. It was developed for use in primary medical care settings but it has also been used in population surveys (although because of a floor effect the author no longer recommends it for this purpose) and in clinical trials.

Part 1 of the NHP contains 38 items grouped into six sections; physical abilities, pain, sleep, social isolation, emotional reactions and energy level. Part 2 contains a brief indicator of handicap and contains 7 items that record the effect of health problems on occupation, household jobs, personal relationships, social life, sex life, hobbies and holidays. Part 2 is optional, some items are not applicable to some groups such as the elderly, and the author has temporarily withdrawn it (Bowling, 1997). Yes/no responses are used throughout.

The NHP is scored from 0–100 with 0 reflecting no health problems/good health. The weighted scores for Part 1 are summed for each domain and can be presented as a health profile. A simpler scoring system is often used which counts the number of affirmative responses in each section and Jenkinson (1991) found that weighted and unweighted scores correlated so highly that he questioned the usefulness of the weights. Some users also present results as an overall score although the author does not recommend this. Summing the number of positive responses scores Part II of the NHP and no item weights are used.

Normative data is available for the UK and it has also been widely used in Europe. Reference scores are available for healthy people (by age, sex and social class) and for various categories of patient but these may need to be smoothed before using them as reference standards (McDowell and Newell, 1996). The NHP has been used in health surveys, medical care settings and in clinical trials. Current usage in Australia is limited but Australian data has been published by Baum & Cooke (1989), Kalucy & Baum (1992) and Crockett et al. (1996).

The NHP is short, inexpensive, and easy to use and overall has adequate psychometric properties. It has been criticised as it only provides a limited measure of function and some disabilities are not assessed – including sensory defects, incontinence and eating problems (Bowling, 1997). It has a very skewed distribution and a floor effect so it may not measure minor improvements in health nor be sensitive to people with mild conditions.

The weights in the scoring system have been criticised and scoring anomalies have been detected (Jenkinson, 1991). There has been some criticism that each section does not represent just one dimension and co-variation between items in different categories raises difficulties for interpretation (Bowling, 1997).

The WHOQOL-100

The WHOQOL-100 includes 100 items covering six domains: Physical, Psychological, Independence, Social relationships, Environment and Spiritual. Each domain is made up of a number of facets or issues. These facets are listed below:

Physical:
- Pain and discomfort
- Energy and fatigue
- Sleep and rest

Psychological:
- Positive affect
- Thinking, learning, memory and concentration
- Self-esteem
- Body image and appearance
- Negative affect

Independence:
- Mobility
- Activities of daily living
- Dependence on medication or treatments
- Working capacity

Social relationships:
- Personal relationships
- Social support
- Sexual activity

Environment:
- Physical safety and security
- Home environment
- Financial resources
- Health and social care: accessibility and quality
- Opportunities for acquiring new information and skills
- Participation in and opportunities for recreation/leisure activities
- Physical environment (pollution, noise, traffic, climate)
- Transportation

Spiritual:
- Spirituality/religion/personal beliefs

An additional four items are included on Overall Quality of life and General Health.

Each of the 24 facets is made up of 4 items. Each item is presented in a question format and is answered on a five-point response scale. Facet items are then summed to produce a Facet score (some items needing to be reversed scored). Domain scores are then calculated from the facet scores (some facet scores need to be reversed scored). Scoring algorithms are provided. All scores (facet and domain) range from 4 – 20. A higher score indicates better QOL.

The World Health Organisation (WHO) developed the WHOQOL-100 using a collaborative project involving 15 centres from around the world (Australia, Croatia, France, India (x2), Israel, Japan, The Netherlands, Panama, Russian Federation, Spain, Thailand, United Kingdom, United States of America, and Zimbabwe). Development of the items involved focus groups, forward and backward translations and linguistic review at each stage. A global item pool was produced and 236 common items were selected for pilot testing in each country (n = 300). These common items were further reduced to 100 using standard psychometric techniques. The five-point response scale descriptors for each question were also adapted for each language. This effort produced a set of QOL questions addressing common facets or issues across cultures with slightly different item wording for each language.

Australian normative data (n = 300) for the WHOQOL-100 is provided in the User's manual and interpretation guide (Murphy et al. 2000). International Data (n = 4802) is available for the WHOQOL-100 (WHOQOL Group, 1998a).

The WHO-QOL Bref

The WHOQOL-BREF includes 26 items (one item from each of the 24 facets contained in the WHOQOL-100, plus two items on overall Quality of life and satisfaction with health) covering four domains: Physical, Psychological, Social relationships and Environment (a reduction from the set of six produced by the WHOQOL-100). The 26 items are listed below:

Physical:
- Pain and discomfort
- Dependence on medical treatment
- Energy and fatigue
- Mobility
- Sleep and rest
- Activities of daily living
- Working capacity

Psychological:
- Positive affect
- Spirituality
- Thinking, learning, memory and concentration
- Body image and appearance
- Self-esteem
- Negative affect

Social relationships:
- Personal relationships
- Sexual activity
- Social support

Environment:
- Physical safety and security
- Physical environment (pollution, noise, traffic, climate)
- Financial resources
- Opportunities for acquiring new information and skills
- Participation in and opportunities for recreation/leisure activities
- Home environment
- Health and social care: accessibility and quality
- Transportation

Plus two items on Overall Quality of life and satisfaction with health.

Each of the 26 items is presented in a question format and is answered on a five-point response scale. They are then summed to produce the Domain scores (some items need to be reversed scored). Scoring algorithms are provided. Scores are then converted to a 0 – 100 scale. A higher score indicates better QOL. Facet scores cannot be derived from the single items used in the WHOQOL-BREF.

The WHOQOL Project attempted to derive the major issues or themes that are universally accepted as important in determining a person's Quality of Life (QOL). The WHOQOL-BREF is one result of this enterprise. It is a brief assessment of QOL extracted from its parent the WHOQOL-100.

Australian Data is provided by the Victorian Validation Study (n = 396) and the Longitudinal Investigation of Depression Outcomes (LIDO) Study (n = 115). International Data for the WHOQOL-BREF is provided in WHOQOL Group (1998b).

The following table shows the expert panel ratings for the generic health-related QoL measures reviewed in this paper.

Table 32   **Summary of ratings for generic health-related quality of life measures (general health status)**

Criteria	SF 36 (II)	SF 36 (I)	SIP	WHO-QOL Bref	SF 12	NHP	WHO-QOL 100
Availability of comparison data/usage	2	3	2	2	2	2	1
Length, ease and time to complete	2	2	1	2	2	1	1
Method of administration	2	2	2	2	2	2	2
Translations available	2	2	2	2	2	2	2
Ease of scoring	3	3	2	2	3	2	2
Sensitivity to incontinence	3	3	2	2	2	2	3
Reliability evidence available	3	2	2	2	2	2	2
Validity evidence available	3	2	2	1	1	1	1
Adherence to psychometric axioms	3	3	2	2	2	2	2
Cost of using the instrument	1	1	3	3	1	3	3
Weighted Total	**60**	**57**	**48**	**47**	**45**	**45**	**45**

SF 36 (II)   SF-36® Health Survey (Version 2.0)
SF 36 (I)   SF-36® Health Survey (Version 1.0)
SIP   Sickness Impact Profile
WHO-QOL Bref   World Health Organisation Quality of Life Assessment Bref (26 item) version
SF 12   SF-12® Health Survey (Version 1.0)
NHP   Nottingham Health Profile
WHO-QOL 100   World Health Organisation Quality of Life Assessment 100 item version

The SF-36® Health Survey (Version 1.0), the SF-36® Health Survey (Version 2.0), the Sickness Impact Profile, and the World Health Organisation Quality of Life Assessment Bref were the only tools in the Generic Health-Related Quality of Life Measures category to reach the required 47-point cut-off score. Given their high ratings, the first two measures, the SF-36® Health Survey (Version 1.0) and the SF-36® Health Survey (Version 2.0), are the recommended tools in this category. Both tools are recommended with a preference for the new Version 2.0 especially when Australian normative data becomes available. The SF-36 Version 2.0 is less prone to ceiling and floor effects on the role functioning scales, which may be an important consideration for patients with incontinence.

Recommended Generic Health-Related Quality of Life Measures

Based upon our review, the following Generic Health-Related Quality of Life Measures (General Health Status) are recommended.

Table 33   Recommended generic health-related quality of life measures (general health status)

User category	Recommended Generic Health-Related Quality of Life Measures (General Health Status)
Primary Care Practitioner	1. SF-36 V2 2. SF-36 V1
Specialist Practitioner	1. SF-36 V2 2. SF-36 V1
Researcher	1. SF-36 V2 2. SF-36 V1

Functional Measures

Rehabilitation outcomes have been assessed in the general rehabilitation clinical research literature for a considerable period. The tools to be reviewed in this section of the report include the Functional Independence Measure, the Katz Activities of Daily Living Scale and the Barthel Index. The Functional Independence Measure was developed in 1986, the Katz Activities of Daily Living Scale in 1959 and the Barthel Index in 1955. These tools have been used in hundreds of studies.

Functional Independence Measure

The FIM consists of 18 items with two sub-scales including consideration of independence in:

- Self care
- Sphincter control
- Transfers
- Locomotion
- Communication and
- Social cognition

The two sub-scales include physical (13 items) and cognitive (five items). The items are scored on a common 7-point scale ranging from 1 (total assistance required) to 7 (complete independence). The FIM was developed through a comprehensive process auspiced by the American Academy of Physical Medicine and Rehabilitation and the American Congress of Rehabilitation Medicine in the late 1980s (Granger et al. 1986: Hamilton et al. 1987). A trained observer administers the scale.

The FIM has been subjected to extensive psychometric study. Ottenbacher, Hsu, Granger & Fiedler (1996) reported a systematic review of 11 studies and reported median test-retest reliability of 0.95 and a median inter-rater reliability of 0.95.

Hsueh, Lin, Jent & Hsieh (2002) administered the FIM, Barthel 10 item form and the Barthel 5 item form to 118 inpatients with stroke at the National Taiwan University Hospital. The 10 item Barthel and the FIM demonstrated high internal consistency (Cronbach's alpha = 0.84), high concurrent validity (r = 0.92, ICC = 0.83) and high responsiveness. The 5 item Barthel had reasonable internal consistency (Cronbach's alpha = 0.71) but with lower concurrent validity with the other tools (r = 0.74, ICC = 0.55) at admission with better performance at discharge (r = 0.92, ICC = 0.74).

Hobart et al. (2000) compared the performance of the FIM and the 10 item Barthel index in a sample of 149 inpatients from two English rehabilitation units. They found high internal consistency, test-retest reliability and validity.

The FIM tool has been used in a wide range of incontinence studies. Foxx-Orestein et al. (2003) used the FIM in an outcomes study with 1013 rehabilitation inpatients from 17 centres in the United States. Eldar et al. (2001) used the FIM to study 37 patients with urinary incontinence in Israel. Gross (2000) have studied rehabilitation outcomes for 90 patients with stroke and urinary incontinence in the United States. Brock, Goldie and Greenwood (2002) in their study used a sample of 106 acute stroke patients to evaluate the discriminant validity of the FIM tool against others. They chose the FIM on the basis of its discriminant validity performance. The FIM has been used by the AN-SNAP study in Australia (see Lee, Eagar, Smith, 1998) and hence there are good clinical benchmarks. Professor Moore (Moore et al. 2004) is currently performing some work concerning the use of an appropriate cut-off on the FIM that is pertinent to the present discussion.

The Functional Independence Measure is a standard outcome measurement tool of long-standing and with demonstrated high levels of reliability and validity. It has been used in many studies of the effectiveness of rehabilitation interventions for patients with urinary incontinence. Because of its wide use, the FIM has excellent benchmark values across a wide variety of countries and patient groups. FIM is the central measure in the Uniform Data System for Medical Rehabilitation (UDS) in the US, and facilities in Australia, Canada, Japan and Europe also make use of this data management system. As a result the validation studies for this instrument are usually based on very large samples – which are unique in this field. In view of these considerations, it is recommended for use by specialist practitioners and researchers as a general functional outcome measure. As the FIM requires the rater to be trained, its use amongst non-specialist, primary care practitioners is likely to be less frequent.

The Barthel Index

The Barthel Index in its original form (Mahoney & Barthel, 1965) consists of 10 items relating to Activities of Daily Living. Minor modifications were made to the index by Collin and Wade et al. (1988) and McDowell and Newell (1996) recommend the use of this as the 10 item version. Eight of the items are self-care items (feeding, transfers, grooming, toileting, bathing, dressing, bowel and bladder continence) and two of the items are related to mobility (walking and ascending and descending stairs). A 15 item version has also been developed by Fortinsky, Granger and Seltzer (1981) which is also recommended for use by McDowell and Newell (1996).

Each item is scored on a three point scale and then these scores are weighted and summed to produce a score of between 0 and 100. The scale is administered by a trained observer and may take from 5 minutes to complete if based on reports but longer if based on direct clinical observations. It can also be self-administered.

An Australian research group, Shah, Vanclay and Cooper (1989) have attempted to modify the Barthel to improve its performance by changing the scoring system. However, Hocking, Williams, Broad and Baskett's study (1999) did not demonstrate any gains through this modification.

More recently a five-item version of the tool has been developed. Hobart and Thompson (2000) reported the development of the shorter tool in a sample of 844 patients admitted to a neurological rehabilitation unit in London. The researchers claimed that the tool they developed is "psychometrically equivalent" to the 10 item version with an ICC of 0.90 between the two measures. However, the work by Hsueh et al. (2002) and colleagues reviewed in the FIM section of this review do not entirely support this conclusion. Unfortunately in the 5-item version the continence items are omitted making this version inappropriate for use with patients with incontinence.

As is the case with the other functional outcome measures, the Barthel Index has been subjected to numerous reliability and validity studies. Test-retest reliability has been shown to be too high, ranging from r = 0.71 to 0.87 and inter-rater reliability is similarly high r =0.99 (see Cohen and Merino's, 2000 review for a full discussion of these studies).

The Barthel Index has been used in a wide range of studies of patients with incontinence problems. For example, Chua, Chuo and Kong (2003) have used the Barthel in a study of 84 patients with Traumatic Brain Injury and urinary incontinence in Singapore. Kalita, Shah, Kapoor and Misra (2002) used the Barthel in a study of 18 patients with urinary incontinence in India. Van Kujik, van der Linde and Van Limbeek also used the Barthel in the Netherlands in a study of stroke patients with urinary incontinence. The Barthel is a standard tool in these contexts and has excellent psychometric properties.

The scores on the FIM and Barthel can be mapped to each other using a table of equivalences (see Eagar et al. 1997).

The Katz ADL Scale

The Katz ADL Scale was developed in 1959 (Staff of the Benjamin Rose Hospital, 1959) and is designed to measure functional outcomes in older people and people with health problems. It consists of six items (bathing, dressing, toileting, transfers, continence and feeding). A clinical observer rates each of the 6 activities on a 3-point scale from independent to dependent, based on the degree of assistance required by the patient to complete each activity. Guttman scalogram analysis was used in the scale construction. Although it is a venerable scale and has been widely used in many studies, there are few studies of the psychometric properties of this tool. In their comprehensive review, Cohen and Marino (2001) give an equivocal recommendation for its continued use, preferring the Barthel and especially the FIM because of their extensive psychometric validation and demonstrated reliability. The use of the Katz is not recommended.

Table 34   Summary of ratings for functional measures

Criteria	Tool		
	FIM	Barthel	Katz
Availability of comparison data/usage	3	2	1
Length, ease and time to complete	1	2	2
Method of administration	1	1	1
Translations available	2	2	2
Ease of scoring	1	3	2
Sensitivity to incontinence	3	2	1
Reliability evidence available	3	2	1
Validity evidence available	3	2	1
Adherence to psychometric axioms	3	2	1
Cost of using the instrument	1	3	3
Weighted Total	**55**	**50**	**33**

FIM   Functional Independence Measure
Barthel   Barthel ADL Index
Katz   Katz ADL Scale

The functional measures category had two tools that reached the 47-point cut-off score. These were the Functional Independence Measure and the Barthel.

Recommendations concerning Functional Measures in continence studies

The FIM and the Barthel are the instruments of choice because of their demonstrated high levels of psychometric performance. Recent studies have attempted to shorten the 10 item Barthel to 5 items (Hobart and Thompson, 2000). However the shortened 5 item tool excludes the bladder and bowels items from the original 10 item tool and this is unlikely to be useful in studies and clinical assessment of incontinence.

The requirement of formal training limits the feasibility and suitability of the FIM in the primary care context. Therefore, we have recommended the Barthel as being suitable for primary care clinicians, but we recommend either the FIM or Barthel for specialist clinicians and researchers.

Based on our review the recommendations for functional outcome measures in incontinence studies are as follows:

Table 35   Recommendations for functional measures

User category	Recommended Functional Measures
Primary care practitioner	Barthel
Specialist practitioner	1. FIM 2. Barthel
Researcher	1. FIM 2. Barthel

Overall outcome measurement tool recommendations

The following table provides a summary of the overall recommendations for faecal and urinary incontinence outcome measures.

Table 36   Summary of recommended instruments

Tool Content Domain	User Category		
	Primary Care Practitioner	Specialist Practitioner	Researcher
Faecal Incontinence Symptom Measures	Wexner Symptom Scoring System	Wexner Symptom Scoring System	Wexner Symptom Scoring System
Faecal Incontinence Health-Related Quality of Life Measures	None	None	None
Urinary Incontinence Symptom Measures	ISI	1. King's Health 2. UDI	1. King's Health 2. UDI
Urinary Incontinence Health-Related Quality of Life Measures	None	None	None
Pad Tests	24-hour pad test	24-hour pad test	24 or 48-hour pad test
Frequency Volume Charts and Bladder Diaries	ICS/WHO templates for 3 days	ICS/WHO templates for 3 days	ICS/WHO templates for 7 days
Multi-Attribute Utility Instruments	None	1. EQ5D 2. AQoL 3. HUI3	1. EQ5D 2. AQoL 3. HUI3
General Health-Related Quality of Life Measures (General Health Status)	1. SF-36 V2 2. SF-36 V1	1. SF-36 V2 2. SF-36 V1	1. SF-36 V2 2. SF-36 V1
Functional Measures	Barthel	1. FIM 2. Barthel	1. FIM 2. Barthel

Chapter 4   Consultations with practitioners and researchers

Associated with this project a national meeting with incontinence clinicians and researchers was held in Melbourne on the 23rd of September, 2002. A second workshop was held in Adelaide on the 16th of October, 2002 associated with an incontinence scientific meeting. A structured interview was developed for the purpose of engaging with these key informants.

Issues covered in the consultations

In the consultations the following questions were asked:
- Do you advocate a specific definition of incontinence? If so, what is it?
- Do you advocate a specific definition of "improvement" or "partial response"? If so, what is it?
- Do you advocate a specific definition of "cure" with respect to incontinence? If so, what is it?

Current practice issues

In relation to current practice issues, the following questions were asked:
- How would you describe your involvement in incontinence practice and research?
- With your patients do you use any general screening question(s) concerning incontinence? If so, what is it (are they)?
- In your practice, do you use frequency volume charts?
- In your practice do you use bladder charts?
- In your practice do you conduct wet checks in elderly patients?
- In your practice do you use Urilos and/or other electronic devices?
- In your practice do you use pad tests?
- In your practice, do you use any other home tests?
- In your practice do you use the paper towel test?
- In assessing incontinence, do you use any of the following scoring systems?
    - Groutz Score?
    - Wexner Faecal Incontinence Score?
    - Lagro Janssen Score?
- Do you use any other scoring system?
- Do you have any comments on scoring systems for incontinence measures?
- In your practice do you use any tests of faecal incontinence?
- Do you ask your patients about outcomes in any other formal or informal way?
- Do you ever collect formal or informal health-related quality of life measures from your patients?
- Do you have any other comments you wish to make about the selection of incontinence outcome measures for this project?

Outcomes of the consultations

The following summary is derived from transcripts and notes of the two meetings held in Melbourne and Adelaide.

Issues related to Symptom Severity Measures

There was no current consensus on the tools used. Mostly the participants reported use of locally developed unstandardised tools and checklists. No consensus position was reached in the discussions but the following issues were canvassed in discussion:
- Incontinence needs to be measured objectively both in terms of the severity of symptoms and also subjectively in terms of the impact upon health-related quality of life.
- Practitioners need to establish quickly whether the person is potentially incontinent in the first place with some sort of general screening question and then to move on to something more specific.

- It is important to use standardised tests or parameters capturing objective and subjective information that can be repeated to demonstrate improvement by differences in ratings. This also enables the discussion of and benchmarking of outcomes across services because "everyone is talking the same language".

Issues related to Frequency Volume Charts/Bladder Diaries

The following points arose in discussion:
- The Frequency Volume Chart was seen as precise, recording input, measuring output and the occasions on which they "wet".
- It is necessary to record both the patient's input as well as output.
- It was suggested that for the purposes of reliability > 2 days of recording was required but that practicality made this difficult.
- Frequency volume charts and bladder diaries were seen as necessary measures in assessing incontinence and the impact of interventions.

Issues related to wet checks

The "wet check" procedure involves the clinician putting his or her hand under the patient and seeing whether they are wet). This was not endorsed because of concerns about lack of respect for patient dignity. This is also not a method for which there is a psychometric literature.

The following alternatives were mentioned in discussion:
- Number of linen changes,
- Changes of pads (pad checks),
- Pads with indicators that change colour,
- Changes of "kylies" or drawsheets as a wet check.

Issues concerning Electronic Device Measures

The following points were raised in discussion:
- They were considered to be potentially good in the future but not practical or affordable as yet.
- Limited evidence concerning validity and reliability is currently available.

Issues concerning paper towel measures

Some clinicians, especially physiotherapists, use the paper towel test. In this test a paper towel is put in the underpants with an incontinence pad behind it that can detect small amounts of urine. From the point of view of the non-specialist they are probably not practicable or applicable.

The view in the discussions was that the paper towel test may have a role for some practitioners but more as a supplementary measure rather than part of the basic outcomes suite.

Issues concerning pad tests

The following points were raised in discussion:
- The 48 hr pad test may be very good in terms of reliability and validity but it is not practical.
- The 1 & 2 hr pad test are not recommended because they are not reliable or valid and hence are considered to be inappropriate.
- The lower limit of normal for a 24 hr pad test about to be published in the British Journal of Obstetrics and Gynaecology is 0.3 gms. A major problem in the past has been that people have been using inaccurate kitchen scales that only have an accuracy of 1–2 gms.

Issues concerning faecal incontinence measures

The following points were raised in discussion:

- Something simple is needed, that has been validated and is applicable.

Suggestions:

- The Wexner 20 point Scale is very useful.
- Another measure could be how many faecal accidents a person has and whether this has changed.
- An incontinent episode diary – just describing those episodes – a frequency account over say 3 months.
- Need a bowel diary (with a number of variables) and then a longer term diary which is an episode count of accidents (frequency – a more simplified version of a frequency volume chart).
- Bowel accident charts need to be analysed.

Issues concerning health-related quality of life measures

The following points were raised in discussion:

- There was the impression that health-related quality of life instruments are mainly used in research not in clinical practice.
- There is a problem because there is no global QoL measure for all pelvic floor conditions that incorporates urinary or faecal incontinence or prolapse.
- Instruments need to be condition specific not generic.
- There has been some use of FIM, Kings' College Questionnaire, Bristol FLUTS, and International Prostate Scoring Scale (IPPS).
- IIQ & UDI are used together in research.
- The patient should be believed – if they say they're better – chances are that they are; if they say that they are worse – chances are that they are worse.
- There is a problem with no standardised scoring if you want to compare patients.
- In Melbourne most of the services use a questionnaire that has been locally developed (collaboratively) but it has not been validated. It includes QoL questions but no-one has analysed them – follow up is often interview based. (A copy of the form was inspected and found it to be useful but it has not been formally evaluated at this stage.)
- A questionnaire before the first consultation can be helpful as an ice-breaker.
- Should one of the recommendations of this project be whether a QoL tool is essential for assessment in this suite of measures or whether QoL measures should be restricted to a research settings or intervention assessment studies?
- The QoL measure that may be recommended by this project may not be based on validated evidence.

Summary of consultation outcomes

In order to facilitate comparison of the outcomes of the consultation with experts with the outcomes of the literature analysis, the consultation outcomes have been grouped under the following headings:

Symptom Severity Indexes

The participants did not recommend any index.

Quality of Life Measures

The participants made no specific recommendations except that a recommendation based upon the literature would be useful.

Pad Tests

For primary care practitioners 24 hours monitoring should be employed as the minimum duration in pad tests but this was sometimes difficult to implement.

For specialist practitioners and researchers, a minimum of 48 hours duration and preferably 72 hours should be employed in pad tests but that this was practically difficult.

Bladder Diary/Frequency Volume Chart

Frequency Volume Charts provide important insights but are difficult to implement. Bladder Diaries require the person to be cognitively intact.

Faecal Symptoms

The Wexner Faecal Incontinence Score ought to be considered but more basic tools would be useful.

Other tools

The participants did not endorse the wet check and paper towel test because they were considered not to provide sufficiently reliable and valid measures. These recommendations mainly accord with the ones derived from review of the published literature.

The main additional findings from the consultations were:

- There is currently considerable variability in practice in the use of incontinence outcome measurement tools.
- The practicability of most measurement tools in the clinical context is problematic because of the length of the time required to complete and score the tools.
- There is considerable interest in the development of realistic and practical recommendations for outcomes measurement.

Chapter 5   Concluding remarks and observations

This project has been a major undertaking not only because of its wide-ranging scope but also the lack of readiness of the outcome measurement field in the study of incontinence. The basic psychometric properties of many of the incontinence outcome measurement tools reviewed in this report are simply not known because they have not been studied or reported adequately in the published literature. Standard psychometric properties and outcomes are not reported consistently. These are significant barriers to the formulation of recommendations.

The inconsistency in reporting of outcome measures in clinical studies in the field of incontinence has itself been the subject of research by several groups of investigators. Lee, DeAntoni & Daneshgari (2002) reported a study of adherence to the Urodynamic Society's 1997 standards to assess the efficacy of therapy for urinary incontinence. Using the reporting criteria outlined in the standards, the researchers studied 39 articles that reported treatment for urinary incontinence. The overall mean compliance rate for the articles in the study was 29%. The authors concluded that:

> "There is far less than optimal reporting of outcomes for treatment of urinary incontinence per the recommendations of the Urodynamic Society. This low compliance makes standardized evaluation of treatment outcomes of urinary incontinence difficult, if not impossible."

Similarly, in other areas of the recording of clinical variables in outcomes research in incontinence, there is considerable variability in practice. Soroka, Drutz, Glazener, Hay-Smith & Ross (2002) reported wide variations in the application of clinical pad tests for studies of female urinary incontinence. The researchers reviewed 75 papers but found that only 25 had used pad tests that were consistent with International Continence Society guidelines.

Blaivas (1998) presented the outcomes of a major review sponsored by the Urodynamic Society in the areas of definition and classification of urinary incontinence and also standards of efficacy for evaluation of treatment outcomes in urinary incontinence. The recommended primary outcome variables nominated for routine collection within Blaivas's review included:

- The number of incontinent episodes;
- Volume of urinary loss;
- Type of incontinence.

The secondary measures that were proposed for selective collection included:

- Patient satisfaction;
- Health-related quality of life;
- Bladder systems;
- Uro flow;
- Post void;
- Residual urine;
- Other urodynamic variables.

Interestingly, despite the wide scope of Blaivas's review the conclusion regarding recommendations for specific collection was:

> "At the present time, there are no validated reproducible well accepted efficacy instruments for assessing treatment outcomes in urinary incontinence. Further work directed toward the development of such instruments is warranted."

Although there are several reviews of urinary and faecal incontinence outcome measures, there is limited consensus as to which instruments and tools should be employed in the measurement of outcomes of interventions for incontinence and limited agreement concerning, as well as adherence to, standards for the reporting of the outcomes of such studies. While bodies such as the ICS and the Urodynamic Society have attempted to define some of the basic characteristics of such standards, researchers and clinicians demonstrate low uptake of them. Studies such as those performed by Soroka's group and also Lee's group show that there is little adherence to these principles in practice. This remains a serious problem for the cause of adoption of some semblance of uniform practice in collection of outcome measurement data in incontinence research and practice.

The contribution of this work could be considered to be twofold. First, a rigorous methodology for tool evaluation has been developed. It includes criteria that are important to practitioners and researchers alike in the selection of measurement tools. This is important as it means that the process

of formulation of the recommendations has been made clear and transparent. Secondly, a practical set of recommendations for tools to be used by practitioners and researchers has been made.

Durability of the recommendations arising from the present study

As has been noted throughout this paper many areas in the measurement of outcomes for continence are in a state of flux. For example, it has been noted in the area of faecal incontinence health-related quality of life measures that it is only very recently that credible tools have been developed and that they await proper psychometric evaluation. However, in other areas such as the development of urinary incontinence symptom measures, it may well be the case that this period will see significant development of the evidence base.

In other areas, such as the generic/health-related quality of life measures and the multi-attribute utility measures, there is a much lengthier and sounder research base upon which to make recommendations. In these areas it is unlikely that within the next two to three years that there will be such significant developments so as to require the revision and overturning of current recommendations.

Thus, in formulating recommendations concerning the durability of the recommended suite of outcome measures, it is important to consider the differential states of development of the respective areas of the relevant literatures. For this reason, it is suggested that a review period of two to three years would be satisfactory after which the evidence base should be re-examined and the recommendations updated if necessary. For instance, during the finialisation of this publication, a number of important papers emerged in the literature. These include: Cockell et al. (2003) on the Postpartum Flatal and Faecal Incontinence Quality of Life Scale; the development of a new set of measures for urinary incontinence by Avery et al. (2004); a new paper by Rockwood (2004) on severity and quality of life scales for faecal incontinence; and two reviews of urinary incontinence quality of life meaures (Corcos et al. 2002 and Naughton et al. 2004). Also the recent South Australian Health Ominbus Survey (Harrison Health Research, 2004) will provide updated epidemological data about incontinence for the Australian population.

Recommendation: it is suggested that a review period of two to three years would be satisfactory after which the evidence base should be re-examined and the recommendations updated if necessary.

Recommendations concerning research priorities in continence outcome measurement research

Having made recommendations for the selection of incontinence outcome measurement tools for the various categories of practitioners in the various content domains, it is considered that there are three major research priorities that need to be addressed in subsequent studies:

- There is a need for some large-scale omnibus studies using the recommended tools in the Australian population to further explore validity, reliability, and applicability and practicability considerations within targeted populations with the full range of types of incontinence problems. The methodology employed by Vaizey, Garapeti, Cahill, & Kamm (1999) in their omnibus study provides a useful model for the proposed study or studies. The recommended tools should be trialled together in a study or linked studies to determine which are most suitable and also to develop appropriate Australian norms. Some tools such as the SF tools are in a better position in this regard than others, but most do not have convincing large-scale benchmark studies to date. There are nevertheless, opportunities to conduct them that ought to be pursued.

- There is a need for development of short form tools and common minimum data sets for use by primary care and specialist incontinence practitioners so that the needs of such practitioners can be better met. Although there are many tools in existence, many have poor or undemonstrated psychometric properties with lengthy formats that are unsuited to the generalist primary care practitioner. In various jurisdictions, some tools have been developed to capture the more general patient characteristics. Such tools include the Initial Needs Identification tool developed for the Victorian Primary Care Partnerships initiatives. There are also various coding dictionaries such as those developed by the Australian Institute for Health and Welfare, which are helpful in the design of common coding systems for minimum data sets. However, there is a dearth of usable tools in the area of incontinence, which are practical for primary care practitioners. As such practitioners provide the bulk of services to people with incontinence this is a significant deficiency in current capacity.

- Examination of the use of outcome measures for incontinence conditions across different cultural and lingustic groups is lacking in Australia. This is especially the case for the Aboriginal community. Here perhaps, the work of Dr. Kate A. Senior could be used as a guide (Senior, 2003).

References

Abrams, P., Blaivas, J., Stantion, S., & Andersen, J. (1990). The standardization of terminology of lower urinary tract function recommended by the International Continence Society. *International Urogynecological Journal, 1*, 45-58.

Abrams, P., Cardozo, L., Khoury, S. & Wein, A. [Editors]. (2002a). *Incontinence. 2nd International Consultation on Incontinence. Paris 1-3, July 2001*. (2nd Edition) Health Publications Limited: United Kingdom.

Abrams, P., Cardozo, L., Fall, M., Griffiths, D., Rosier, P., Ulmsten, U. et al. (2002b). The standardization of terminology of lower urinary tract function: Report of the Standardisation Sub-committee of the International Continence Society. *Neurourology and Urodynamics, 21*, 167-178.

Adhikari, P. & Summerill, A. (2000). 1998 National Drug Strategy Household Survey: Detailed findings. (AIHW Cat. No. PHE 27) Canberra: AIHW.

Andersen, L., Kristiansen, I., Falch, J., & Aursnes, I. (1995) *Cost-effectiveness of Alendronate for the prevention of osteoporotic fractures in Norwegian women*. Working Paper. Oslo.

Anderson, J., Kaplan, R., Berry, C., Bush, J., & Ruben, R. (1989), Interday reliability of function assessment for a health status measure: The Quality of Well-Being scale. *Medical Care, 27*(11), 1076-1084.

Andresen, E., Patrick, D., Carter, W., & Malmgren, J. (1995). Comparing the performance of health status measures for healthy older adults. *Journal of the American Geriatrics Society, 43*(9), 1030-1034.

Arya, L., Myers, D., & Jackson, N. (2001). Office screening test for intrinsic urethral sphincter deficiency: Pediatric Foley Catheter Test. *Obstetrics and Gynecology, 97*(6), 885-889.

Australian Bureau of Statistics. (1997) *1995 National Health Survey: SF-36 population norms,* Cat No 4399.0, Canberra.

Australian Health Ministers Advisory Council, February (1993). As quoted in: *Health Outcomes Bulletin*. No. 1, February 1994, 5.

Australian Health Ministers Advisory Council. (1996) The Final Report of the Taskforce on Quality in Australian Health Care AGPS, Canberra 1996.

Avery, K., Donovan, J., Peters, T.J., Shaw, C., Gotoh, M., & Abrams, P. (2004). ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourology & Urodynamics, 23*, 322-330.

Aydos, M., Memis, A., Yakupoglu, Y., Ozdal, O., & Oztekin, V. (2001). The use and efficacy of the American Urological Association Symptom Index in assessing the outcome of urethroplasty for post-traumatic complete posterior urethral strictures. *British Journal of Urology, 88*(4), 382-384.

Badia Llach, X., Castro, D., & Conejero, S. (2000). Validity of the King's Health questionnaire in the assessment of quality of life of patients with urinary incontinence. The King's Group. *Medicina Clinica, 114*(17), 647-652.

Baeten, C., Bailey, H., Bakka, A., Belliveau, P., Berg, E., Buie, W. et al. (2000). Safety and efficacy of dynamic graciloplasty for fecal incontinence: report of a prospective, multicenter trial. Dynamic Graciloplasty Therapy Study Group. *Diseases of the Colon & Rectum, 43*, 743-751.

Barber, M., Kuchibhatla, M., Pieper, C., & Bump, R. (2001). Psychometric evaluation of 2 comprehensive condition-specific quality of life instruments for women with pelvic floor disorders. *American Journal of Obstetrics and Gynecology, 185*(6), 1388-1395.

Barnboym, E., Ahrens, A., & Roehrborn, C. (1999). Effect of scramblilng on the short-term reliability of the American Urological Association Symptom Index. *Urology, 53*(3), 568-573.

Barry, M., Fowler, F. J., O'Leary, M., Bruskewitz, R., Holtgrewe, H., Mebust, W., et al. (1992). The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *Journal of Urology, 148*(5), 1549-1557.

Baum, F., & Cooke, R. (1989). Community-health needs assessment: use of the Nottingham health profile in an Australian study. *Medical Journal of Australia, 150*, 581-590.

Beaton, D., Hogg-Johnson, S., & Bombardier, C. (1997). Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *Journal of Clinical Epidemiology, 50*(1), 79-93.

Beddar, S., Holden-Bennett, L., & McCormick, A. (1997). Development and evaluation of a protocol to manage faecal incontinence in the patient with cancer. *Journal of Palliative Care, 13*(2), 27-38.

Behavioural Epidemiology Unit. (1995). *South Australian Population Norms for the Short Form 36 (SF-36) Health Status Questionnaire.* Adelaide: South Australian Health Commission.

Bennett, K., & Torrance, G. (1996) Measuring health state preferences and utilities: rating scale, time trade-off, and standard gamble techniques. In: B Spilker (ed.) *Quality of Life and Pharmacoeconomics in Clinical Trials* 253-265. Philadelphia: Lippincott-Raven Publishers.

Bergman, A., & Bhatia, N. (1987). Urodynamic appraisal of the Marshall-Marchetti test in women with stress urinary incontinence. *Urology, 29*(4), 458-462.

Bergman, A., McCarthy, T., Ballard, C., & Yanai, J. (1987). Role of the Q-Tip Test in evaluating stress urinary incontinence. *Journal of Reproductive Medicine, 32*(4), 273-275.

Bergner, M., Bobbitt, R., Kressel, S., Pollard, W., Gilson, B., & Morris, J. (1976a). The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services, 6*, 393-415.

Bergner, M., Bobbitt, R., Pollard, W., Martin, D. & Gilson, B. (1976b). The sickness impact profile: validation of a health status measure. *Medical Care, 14*, 57-67.

Bergner, M., Bobbitt, R., Carter, W., & Gilson, B. (1981). The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care, 19*, 787-805.

Bergner, M. (1993). Development, testing and use of the Sickness Impact Profile. In Walker, S. & Rosser, R. *Quality of Life Assessment: Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers.

Berlowitz, D., Brand, H., & Perkins, C. (1999). Geriatric syndromes as outcome measures of hospital care: can administrative data be used? *Journal of the American Geriatrics Society, 47*(6), 692-696.

Black, N., Griffiths, J., & Pope, C. (1996). Development of a symptom severity index and a symptom impact index for stress incontinence in women. *Neurourology and Urodynamics* (15), 630-640.

Blaivas, J. (1998). Outcome measures for urinary incontinence. *Urology, 51*(2A Suppl.), 11-19.

Bleichrodt, H., Johannesson, M. (1997). An experimental test of a theoretical foundation for rating- scale valuations. *Medical Decision Making, 17*(2), 208-16.

Bo, K. (1994). Reproducibility of instruments designed to measure subjective evaluation of female stress urinary incontinence. *Scandinavian Journal of Gastroenterology, 28*, 97-100.

Bo, K. (1998). Effect of electrical stimulation on stress and urge urinary incontinence. *Acta Obsetrica et Gynecologica Scandinavica, 77*(Supplement 168), 3-11.

Bo, K., & Finckenhagen, H. (2001). Vaginal palpation of pelvic floor muscle strength: inter-test reproducibility and comparison between palpation and vaginal squeeze pressure. *Acta Obsetrica et Gynecologica Scandinavica, 80*, 883-887.

Boci, R., Fall, M., Walden, M., Knutson, T., & Dahlstrand, C. (1999). Home uroflowmetry: Improved accuracy in outflow assessment. *Neurourology and Urodynamics, 18*, 25-32.

Bombardier, C., & Raboud, J. (1991). A comparison of health-related quality-of-life measures for rheumatoid arthritis research. The Auranofin Cooperating Group. *Control Clinical Trials, 12*(4 Suppl), 243S-256S.

Bowling, A. (1991). *Measuring Health: A Review of Quality of Life Measurement Scales.* Milton Keynes: Open University.

Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales*. 2nd Ed Milton Keynes: Open University.

Bowling, A., Bond, M., Jenkinson, C., & Lamping, D. L. (1999). Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by

the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *Journal of Public Health Medicine, 21*, 255-270.

Boyle, M. H., Furlong, W., Feeny, D., Torrance, G.W., & Hatcher, J. (1995). Reliability of the Health Utilities Index – Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. *Quality of Life Research, 4*, 249-257.

Bradlyn, A. S., Harris, C. V, Warner, J. E., Ritchey, A. K., & Zaboy, K. (1993). An investigation of the validity of the quality of Well-Being Scale with pediatric oncology patients. *Health Psychology, 12*(3), 246-50.

Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment, 3*(9), 1-165.

Brazier, J., Jones, N., & Kind, P. (1993). Testing the validity of the EuroQoL and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research, 2*, 169-180.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*, 271-292.

Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology, 51*(11), 1115-1128.

Brazier, J., Harper, R., & Jones, N. (1992). Validating the SF-36 health survey questionnaire: a new outcome measure for primary care. *British Medical Journal, 305*, 160-164.

Brink, C., Wells, T., Sampselle, C., Taillie, E., & Mayer, R. (1994). A digital test for pelvic muscle strength in women with urinary incontinence. *Nursing Research, 43*(6), 352-356.

Brock, K., Goldie, P., & Greenwood, K. (2002). Evaluating the effectiveness of stroke rehabilitation: choosing a discriminative measure. *Archives of Physical Medicine and Rehabilitation, 83(1),* 92-99.

Brocklehurst, J., Amess, M., Goldacre, M., Mason, A., Wilkinson, E., Eastwood, A., et al. (1999). *Health Outcome Indicators: Urinary Incontinence.* Report of a working group to the Department of Health. London, National Centre for Health Outcomes Development.

Brown, J., Posner, S., & Stewart, A. (1999). Urge incontinence: new health-related quality of life measures. *Journal of the American Geriatrics Society, 47*(8), 980-988.

Bug, G., Kiff, E., & Hosker, G. (2001). A new condition specific health related quality of life questionnaire for the assessment of women with anal incontinence. *British Journal of Gynaecology, 8*(10), 1057-1067.

Bump, R., Coates, K., Cundiff, G., Harris, R., & Weidner, A. (1997). Diagnosing intrinsic sphincter deficiency: Comparing urethral closure pressure, urethral axis, and Valsalva leak point pressures. *American Journal of Obstetrics and Gynecology, 177*(2), 303-310.

Bump, R., Elser, D., Theofrastous, J., & McClish, D. (1995). Valsalva leak point pressures in women with genuine stress incontinence: Reproducibility, effect of catheter caliber, and correlations with other measures of urethral resistance. *American Journal of Obstetrics and Gynecology, 173*(2), 551-557.

Bump, R., Hurt, W., Addison, W., Fantl, J., & McClish, D. (1998). Reliability and correlation of measurements during and after bladder neck surgery. *British Journal of Urology, 82*, 628-633.

Bump, R., Mattiasson, A., Bo, K., Brubaker, L., DeLancey, J., Klarskov, P., et al. (1996). The standardization of terminology of female pelvic organ prolapse and pelvic floor dysfunction. *American Journal of Obstetrics and Gynecology, 175*(1), 10-17.

Burgio, K., Matthews, K. & Engel, B. (1991). Prevalence, incidence and correlates of urinary incontinence in healthy, middle-aged women. *Journal of Urology, 146*, 1255-1259.

Button, D., Roe, B., Webb, C., Frith, T., Colin-Thome, D., & Gardner, L. (1998). Consensus guidelines for the promotion and management of continence by primary health care teams: development, implementation and evaluation. NHS Executive Nursing Directorate. *Journal of Advanced Nursing, 27*(1), 91-99.

Byrne, C., Pager, C., Rex, J., Roberts, R., & Solomon, M. (2002). Assessment of quality of life in the treatment of patients with neuropathic faecal incontinence. *Diseases of the Colon and Rectum, 45*(11), 1431-1436.

Cadet, B. (1994). History of the construction of a health indicator integrating social preference: the quality of well-being scale. In: *7th Meeting, International Network on Health Expectancy (REVES)*, 23-25 February; Canberra.

Campbell, S., Roland, M., & Buetow, S. (2000). Defining quality of care. *Social Science Medicine, 51*(11), 1611-1625.

Caputo, R. (1993). The Q-Tip test and urethrovesical junction mobility. *Obstetrics and Gynecology, 82*, 892-896.

Cattell, R. (1952). *Factor analysis: An Introduction and Manual for the Psychologist and Social Scientist*. New York: Harper & Row.

Chaikin, D., Blaivas, J., Rosenthal, J., & Weiss, J. (1999). Results of pubovaginal sling for stress incontinence: a prospective comparison of 4 instruments for outcome analysis. *Journal of Urology, 162*(5), 1670-1673.

Chambers, L. (1984). The McMaster Health Index Questionnaire, in Wenger, N. J., Mattson, M. E., Furberg, C. D. et al. (eds) *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*. New York: Le Jacq.

Chancellor, M., & Rivas, D. (1993). American Urological Association symptom index for women with voiding symptoms: lack of index specificity for benign prostate hyperplasia. *Journal of Urology, 150*(5), 1706-1709.

Chow, F., Briganti, E., Kerr, P., Chadban, S., Zimmet, P., Atkins, R. (2003). Health-related quality of life in Australian adults with renal insufficiency: a population-based study. *American Journal of Kidney Diseases, 41*, 596-604.

Chua, K., Chuo, A., & Kong, K. (2003). Urinary incontinence after traumatic brain injury: incidence, outcomes and correlates. *Brain Injury, 17(6),* 469-478.

Clayton, J., Smith, K., Qureshi, H., & Ferguson, B. (1998). Collecting patients' views and perceptions of continence services: the development of research instruments. *Journal of Advanced Nursing, 28*(2), 353-361.

Coates, K. (1998). Physiologic evaluation of the pelvic floor. *Obstetrics and Gynecology Clinics of North America, 25*(4), 805-823.

Cockell, S., Oates-Johnson, T., Gilmour, D., Vallis, T., Turnbull, G. (2003). Postpartum flatal and Fecal Incontinence Quality-of-Life Scale: a disease- and population-specific measure. *Qualitative Health Research, 13*, 1132-44.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. Hillsdale: Lawrence Erlbaum.

Cohen, M., & Marino, M. (2000). The tools of disability outcomes: research functional status measures. *Archives of Physical Medicine and Rehabilitation,* 81(Supplement 2), S21-S29.

Collin, C., Wade, D.T., Davies, S., & Horne, V. (1988). The Barthel ADL index: a reliability study. *International Disability Studies, 10*, 61-63.

Comiter, C., Sullivan, M., & Yalla, S. (1997). Retrograde leak point pressure for evaluating postradical prostatectomy incontinence. *Urology, 49*(2), 231-236.

Comiter, C., Sullivan, M., Schacterle, R., & Yalla, S. (1996). Prediction of prostatic obstruction with a combination of isometric detrusor contraction pressure and maximum urianry flow rate. *Urology, 48*(5), 723-730.

Cook, K., Ashton, C., Byrne, M., Brody, B., Geraci, J., Giesler, R., et al. (2001). A psychometric analysis of the measurement level of the rating scale, time trade-off, and standard gamble. *Social Science and Medicine, 53*, 1275-1285.

Coons, S., Kaplan, R. (1993). Quality of life assessment: understanding its use as an outcome measure. *Hospital Formul, 28*, 486-498.

Coons, S., Rao, S., Keininger, D., & Hays, R. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics, 17*(1), 13-35.

Corcos, J., Beaulieu, S., Donovan, J., Naughton, M., & Gotoh, M. (2002). Quality of life assessment in men and women with urinary incontinence. *The Journal of Urology, 168*, 896-905.

Cottenden, A., & Ledger, D. (1993). Predicting the leakage performance of bodyworn disposable incontinence pads using laboratory tests. *Journal of Biomedical Engineering, 15*, 213-220.

Cundiff, G., Harris, R., Theofrastous, J., & Bump, R. (1997). Pressure transmission ratio reproducibility in stress continent and stress incontinent women. *Neurourology and Urodynamics, 16*, 161-166.

Davis, G., McClure, G., Sherman, R., Hibbert, M., Wong, M., & Perez, R. (1998). Ambulatory urodynamics of female soldiers. *Military Medicine, 163*(12), 808-812.

De Bruin, A. F., Diederiks, J. P. M., de Witte, I. P., & J, S. F. C. (1993). *The first testing of the SIP-68, a short generic version of the Sickness Impact Profile*. Paper presented to the Fifth European Health Services Research Conference, Maastricht.

de Kruif, Y., & van Wegen, E. (1996). Pelvic floor muscle exercise therapy with myofeedback for women with stress urinary incontinence: A meta-analysis. *Physiotherapy, 82*(2), 107-113.

De Vellis, R. F. (2003) *Scale development: Theory and applications*. Newbury Park: Sage Publications.

Dempster, M., & Donnelly, M. (2000). Measuring the health related quality of life of people with ischaemic heart disease. *Heart*, *83*, 641–644.

Department of Health (1995). Research Group on the measurement and valuation of health. In, Methodology Workshop, *EUROQoL Conference*; 1995; London: Economics & Operational Research Division, Department of Health.

Devreese, A., Weerdt, W., Feys, H., Dewulf, M., Alaerts, R., Vereecken, R., et al. (1996). Functional assessment of urinary incontinence; the perineal pad test. *Clinical Rehabilitation, 10*, 210-215.

Deyo, R. A., Inui, T. S., & Leininger, J. D. (1982). Functions and psychological functions in rheumatoid arthritis: clinical use of a self-administered instrument. *Archives of Internal Medicine, 14442*, 879-882.

Deyo, R. A., Inui, T. S., & Leininger, J. D. (1983). Measuring functional outcomes in chronic disease: a comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. *Medical Care, 21*, 180-192.

Diokno, A., Normolle, D., Brown, M., & Herzog, A. (1990). Urodynamic tests for female geriatric urinary incontinence. *Urology, 36*(5), 431-439.

Dziki, A. & Bartos, M. Seton treatment of anal fistula: experience with a new modification. (1998). *European Journal of Surgery, 164*, 543-548.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1995). *Social Tariff for EUROQoL: Results from a UK General Population Survey*. York: Centre for Health Economics, University of York, Report No. Discussion Paper 138.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). Valuing health states: a comparison of methods. *Journal of Health Economics, 15*, 209-231.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35*(11), 1095-108.

Donovan, J., Abrams, P., Peters, T., Kay, H., Reynard, J., Chapples, C., et al. (1996). The ICS – 'BPH' Study: the psychometric validity and reliability of the ICSmale questionnaire. *British Journal of Urology, 77*, 554-562.

Donovan, J., Peters, T., Abrams, P., Brookes, S., de la Rosette, J., & Schafer, W. (2000). Scoring the short form ICS male SF questionnaire. International Continence Society. *Journal of Urology, 164*(6), 1948 – 1955.

Donovan, J. L., et al. (2002). Symptom and Quality of Life assessment. In Abrams, P., Cardozo, L., Khoury, S., & Wein, A. (Eds). *Incontinence: 2nd International Consultation on Incontinence*, (pp 267-316) Paris, 2nd Edition.

Dorman, P., Slattery, J., Farrell, B., Dennis, M., & Sandercock, P. (1998). Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. United Kingdom Collaborators in the International Stroke Trial. *Stroke, 29*(1), 63-68.

Dowell, C., Bryant, C., Moore, K., & Simons, A. (1999). Calculating the direct costs of urinary incontinence: a new test instrument. *British Journal of Urology, 83*, 596-606.

Drummond, M., O'Brien, B., Stoddart, G., & Torrance, G. (1998). *Methods for the Economic Evaluation of Health Care Programmes*. 2nd ed. Oxford, Oxford University Press.

Drummond, M. (2001). Introducing economic and quality of life measurements into clinical studies. *Annals of Medicine, 33*, 344-349.

DuBeau, C., Kiely, D., & Resnick, N. (1999). Quality of Life impact of urge incontinence in older persons: A new measure and conceptual structure. *Journal of the American Geriatrics Society, 47*(8), 989-994.

Eagar, K. et al. (1997). *The Australian National Sub-Acute and Non-Acute Patient Classification (AN-SNAP): Report of the National Sub-Acute and Non-Acute Casemix Classification Study.* Centre for Health Service Development, University of Wollongong.

Eagar, K. (2000). *An Overview of Health Outcomes Measurement in Australia.* Paper presented at the Mental Health Research & Development Strategy: Outcomes Conference, Wellington, New Zealand.

Edelman, D., Williams, G., Rothman, M., & Samsa, G. (1999). A comparison of three health status measures in primary care outpatients. *Journal of Intern Medicine, 14*(12), 759-762.

Eldar, R., Ring, H., Tshuwa, M., Dynia, A., & Revital, R. (2001). Quality of care for urinary incontinence in a rehabilitation setting for patients with stroke. Simultaneous monitoring of process and outcome. *International Journal for Quality in Health Care, 13(1),* 57-61.

Elser, D., Fantl, A., & McClish, D. (1995). Comparison of "subjective" and "objective" measures of severity of urinary incontinence in women. *Neurourology and Urodynamics, 14*, 311-316.

Elvik, R. (1995). The validity of using health state indexes in measuring the consequences of traffic injury for public health. *Social Science and Medicine, 40*(10), 1385-98.

Essink-Bot, M., Krabbe, P., Bonsel, G., & Aaronson, N. (1997). An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Medical Care, 35*(5), 522-537.

EuroQol Group. (1990). EuroQol: a new facility for measurement of health-related quality of life. *Health Policy, 16*, 199-208.

Eypasch, E., Williams, J., Wood-Dauphinee, S., Ure, B., Schmulling, C., Neugebauer, E., et al. (1995). Gastrointestinal quality of life index: development, validation and application of a new instrument. *British Journal of Surgery, 82(2)*, 216-222.

Eypasch, E., Wood-Dauphinee, S., Williams, J., Ure, B., Neugebauer, E., & Troidl, H. (1993). The Gastrointestinal Quality of Life Index. A clinical index for measuring patient status in gastroenterologic surgery. *Chirurgerie, 64*(4), 264-274.

Fantl, A., Harkins, S., Wyman, J., Choi, S., & Taylor, J. (1987). Fluid loss quantitation test in women with urinary incontinence: A test-retest analysis. *Obstetrics and Gynecology, 70*(5), 739-743.

Farrell, S. (1998). Application of the concept of 'Therapeutic Index' to surgery for stress incontinence. *International Urogynecology Journal, 9*, 299-302.

Fayers, P., & Machin D. (2000). *Quality of Life: Assessment, Analysis and Interpretation.* Chichester, Wiley.

Fedorkow, D., Sand, P., Retzky, S., & Johnson, D. (1995). The cotton swab test. Receiver-operating characteristic curves. *Journal of Reproductive Medicine, 40*, 42-46.

Feeny, D., Furlong, W., & Torrance, G. (1996). *Health Utilities Index Mark 2 and Mark 3 (HUI2/3) 15-item Questionnaire for Self-administered, Self-assessed Usual Health Status.* Hamilton: Centre for Health Economics and Policy Analysis, McMaster University.

Feeny, D., Torrance, G., & Furlong, W. (1996). Health Utilities Index. In: Spilker, B., (ed.) *Quality of Life and Pharmacoeconomics in Clinical Trials.* 2nd ed. Philadelpha: Lippincott-Raven Publishers.

Folkehelsa. (1995). *Statens Institutt for Folkehelsa.* Report No.: 11/1995.

Fonda, D., Brimage, P., & D'Astoli, M. (1993). Simple screening for urinary incontinence in the elderly: Comparison of Simple and Multichannel Cystometry. *Urology, 42*(5), 536 - 540.

Fonda, D., Resnick, N., Colling, J., Burgio, K., Ouslander, J., Norton, C., et al. (1998). Outcome measures for research of lower urinary tract dysfunction in frail older people. *Neurourology & Urodynamics, 17*, 273-281.

Fonda, D., Resnick, N., Colling, J., Burgio, K., Ouslander, J., Norton, C., et al. (1999). Outcome measures for research of lower urinary tract dysfunction in frail older people: Report from the International Continence Society Working Party on outcome standards. *Journal of Rehabilitation Outcomes and Measurement, 3*(3), 27-33.

Foote, A., & Moore, K. (2001). A comparative cost utility analysis of five treatments for female urinary incontinence. *Australian Continence Journal, 7*(4),108-109.

Fortinsky, R.H., Granger, C.V., Seltzer, G.B. (1981). The use of functional assessment in understanding home care needs. *Medical Care, 19*, 489-497.

Foxx-Orenstein, A., Kolakowsky-Hayner, S., Marwitz, J., Cifu, D., Dunbar, A., Englander, J., et al. (2003). Incidence, risk factors, and outcomes of faecal incontinence after acute brain injury: findings from the traumatic brain injury model system national database. *Archives of Physical Medicine and Rehabilitation, 84,* 231-237.

Fryback, D. G., Lawrence, W. F, Martin, P. A., Klein, R., Klein, B. E. (1997). Predicting quality of well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. *Medical Decision Making, 17*(1), 1-9.

Fuertes, M. E, Garcia Matres, M. J., Gonzalez Romojaro, V., de la Rosa, S., Anguera Vila, A., de la Pena, J., et al. (2000). Clinical trial to evaluate trospium chloride (Uraplex) effectiveness and tolerance in patients with detrusor instability incontinence and its impact on quality of life. *Archives Esp de Urology, 53*(2), 125-136.

Furlong, W., Feeny, D., Torrance, G., Goldsmith, C., DePauw, S., Zhu, Z., et al. (1998). *Multiplicative Multi-attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report. Working Paper*. Hamilton: McMaster University, Centre for Health Economics and Policy Analysis. Report No: 98-11.

Furlong, W. J., Feeny, D. H., Torrance, G. W., & Barr, R. D. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine, 33*, 375-384.

Gardner, A., & Sibthrope, B. (2002). Will he get back to normal? Survival and functional status after intensive care therapy. *Intensive & Critical Care Nursing, 18*, 138-145.

Gee, A. & Durdey, P. (1997). Preoperative increase in neuromuscular jitter and outcome following surgery for faecal incontinence. *British Journal of Surgery, 84*, 1265-1268.

Georgiou, A., Potter, J., Brocklehurst, J., Lowe, D., & Pearson, M. (2001). Measuring the quality of urinary continence care in long-term care facilities: an analysis of outcome indicators. *Age and Ageing, 30*, 63-66.

Gisolf, K., van Venrooij, G., Eckhardt, M., & Boon, T. (2000). Analysis and reliability of data from 24-hour frequency-volume charts in men with lower urinary tract symptoms due to benign prostatic hyperplasia. *European Urology, 38*, 45-52.

Gilson, B. S., Gilson, J. S., Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E. et al. (1979). The sickness impact profile: Development of an outcome measure of health care. *American Journal of Public Health, 65*, 1304-1310.

Glazer, H., Romanzi, L., & Polaneczky, M. (1999). Pelvic floor muscle surface electromyography: Reliability and clinical predictive validity. *Journal of Reproductive Medicine, 44*(9), 779-782.

Goldney, R. D, Fisher, L. J, Wilson, D. H, & Cheok, F. (2001). Suicidal ideation and health-related quality of life in the community. *Medical Journal of Australia, 175*(10), 546-549.

Gousse, A., Barbaric, Z., Safir, M., Madjar, S., Marumoto, A., & Raz, S. (2000). Dynamic half fourier acquisition, single shot turbo spin-echo magnetic resonance imaging for evaluating the female pelvis. *Journal of Urology, 164*, 1606-1613.

Graham, C., & Dmochowski, R. (2002). Questionnaires for women with urinary symptoms. *Neurourology and Urodynamics, 21*, 473-481.

Granger, C., Hamilton, B., Keith, R., Zielesky, M., & Sherwin, F. (1986). Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation, 1(3),* 59-74.

Griffiths, C., Murray, A., & Ramsden, P. (1986). Accuracy and repeatability of bladder volume measurement using ultrasonic imaging. *Journal of Urology, 136*, 808-812.

Griffiths, D., McCracken, P., & Harrison, G. (1991). Incontinence in the elderly: Objective demonstration and quantitative assessment. *British Journal of Urology, 67*, 467-471.

Gross, J. (2000). Urinary incontinence and stroke outcomes. *Archives of Physical Medicine and Rehabilitation*, 81, 22-27.

Groutz, A., Blaivas, J., & Rosenthal, J. (2000). A simplified urinary incontinence score for the evaluation of treatment outcomes. *Neurourology and Urodynamics, 19*(2), 127-135.

Groutz, A., Blaivas, J., Chaikin, D., Resnick, N., Engleman, K., Anzalone, D., et al. (2000). Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: A multicenter study of micturition diary and pad tests. *Journal of Urology, 164*, 698-701.

Gudex, C., Kind, P., van Dalen, H., Durand, M., Morris, J., & Williams, A. (1993). *Comparing Scaling Methods for Health State Valuations – Rosser Revisited. Discussion Paper*. York: Centre for Health Economics, University of York. Report No: 107.

Haeusler, G., Hanzal, E., Joura, E., Sam, C., & Koelbl, H. (1995). Differential diagnosis of detrusor instability and stress-incontinence by patient history: The Gaudenz-Incontinence-Questionnaire revisited. *Acta Obsetrica et Gynecologica Scandinavica, 74*, 635-637.

Hagen, S., Hanley, J., & Capewell, A. (2002). Test-retest reliability, validity and sensitivity to change of the urogenital distress inventory and the incontinence impact questionnaire. *Neurourology and Urodynamics, 21*(6), 534-530.

Hahn, I., & Fall, M. (1991). Objective quantification of stress urinary incontinence: A short, reproducible, provocative pad-test. *Neurourology and Urodynamics, 10*, 475-481.

Hall, J., Hall, N., Fisher, E. et al. (1987). Measurement of outcomes of general practice: comparison of three health status measures. *Family Practice, 4*, 117-23.

Hamilton, B., Granger, C., Sherwin, F., Zielesky, M., &Tashman, J. (1987). A uniform national data system for medical rehabilitation. In M. Fuhrer (Ed.), *Rehabilitation Outcomes: Analysis and Measurement* (pp. 137-147). Baltimore: Brooks.

Hanley, J., Capewell, A., & Hagen, S. (2001). Validity study of the severity index, a simple measure of urinary incontinence in women. *British Medical Journal (Clinical Research Ed.), 322*(7294), 1096-1097.

Hansen, B., Flyger, H., Brasso, K., Schou, J., Nordling, J., Anderson, J., et al. (1997). Validation of the patient-administered Danish Prostate Symptom Score Schedule. *Ugeskr Laeger, 1*(59), 591-597.

Hansen, B., Flyger, H., Brasso, K., Schou, J., Nordling, J.,Thorup Andersen, J., et al. (1995). Validation of the self-administered Danish Prostatic Symptom Score (DAN-PSS-1) system for use in benign prostatic hyperplasic. *British Journal of Urology, 76*(4), 451-458.

Hanzal, E., Berger, E., & Koelbl, H. (1991). Reliability of the urethral closure pressure profile during stress in the diagnosis of genuine stress incontinence. *British Journal of Urology, 68*, 369-371.

Harper, A. C., Harper, D. A., Lambert, L. J., de Klerk, N. H., et al. (1995). Development and validation of the Curtin Back Screening Questionnaire (CBSQ): A discriminative disability measure. *Pain, 60*, 73-81.

Harrison Health Research, (2004). *South Australian Health Omnibus Survey (HOS)*. (Report). Adelaide: South Australia Department of Health.

Harvey, M., Kristjansson, B., Griffith, D., & Versi, E. (2001). The Incontinence Impact Questionnaire and the Urogenital Distress Inventory: a revisit of their validity in women without a urodynamic diagnosis. *American Journal of Obstetrics and Gynecology, 185*(1), 25-31.

Hawthorne, G., Biddle, D., & Goulopoulos, J. (2002). *User's Manual: The ACPMH Assessment and Evaluation Protocols*. Version 2.2. In. Melbourne: Australian Centre for Posttraumatic Mental Health. p. 79.

Hawthorne, G., Cheok, F., Goldney, R., & Fisher, L. (2003). The excess cost of depression in South Australia: a comparative study of two methods of calculating burden. *Australian and New Zealand Journal of Psychiatry, 37*, 362-373.

Hawthorne, G., & Harmer, C. (1999). GUTSS: *The Genito-Urinary Treatment Satisfaction Scale Study. Melbourne*: Centre for Health Program Evaluation; Report No.: Working Paper 117.

Hawthorne, G., Osborne, R., McNeil, H., & Richardson, J. (1996). *The Australian Multi-Attribute Utility (AMAU): Construction and Initial Evaluation.* Melbourne: Centre for Health Program Evaluation, 1996. Report No: Working Paper 56.

Hawthorne, G., Osborne, R. H., & Elliott, P. (2003). Commentary on: a psychometric analysis of the measurement level of the rating scale, standard gamble and time-trade off, by Cook et al. *Social Science & Medicine, 56*, 895-897.

Hawthorne, G., Richardson, J., Day, N., & McNeil, H. (2000). *Life and Death: Theoretical and Practical Issues in Utility Measurement.* Melbourne: Centre for Health Program Evaluation. Report No: Working Paper 102.

Hawthorne, G., Richardson, J., Day, N., & McNeil, H. (2000). *Using the 'Assessment of Quality of Life' (AQoL) instrument.* Melbourne: CHPE; Report No.: Technical Report 12.

Hawthorne, G., Richardson, J., Day, N., Osborne, R., & McNeil, H. (2000). *Construction and Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument.* Melbourne: Centre for Health Program Evaluation. Report No: Working Paper 101.

Hawthorne, G., Richardson, J., & Day, N. (2001). A comparison of five multi-attribute utility instruments. *Australian Studies in Health Service Administration, 89*, 151-179.

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQol) with four other generic utility instruments. *Annals of Medicine, 33*, 358–370.

Hawthorne, G., Richardson, J., Osborne, R., & McNeil, H. (1997). *The Assessment of Quality of Life (AQoL) Instrument: Construction, Initial Validation and Utility Scaling.* Melbourne: Centre for Health Program Evaluation. Report No.: Working Paper 76.

Hawthorne, G., Richardson, J., & Osborne, R. (1999). The Assessment of Quality of Life (AQoL) Instrument: a psychometric measure of health related quality of life. *Quality of Life Research, 8*, 209-224.

Hawthorne, G., & Richardson, J. (2001). Measuring the value of program outcomes: a review of utility measures. *Expert Review of Pharmacoeconomics Outcomes Research, 1*(2), 215-228.

Hawthorne, G., & Richardson, J. (1995). *An Australian MAU/QALY Instrument: Rationale and Preliminary Results.* Melbourne: Centre for Health Program Evaluation. Report No.: Working Paper 49.

Hawthorne, G. (2002). The effect of different methods of collecting data: mail, telephone and filter data collection issues in utility measurement. *Quality of Life Research*, *12*, 1081-1088.

Hawthorne, G. (In press). *Review of Patient Satisfaction Measures for Continence Conditions.* (Report). Canberra: Department of Health and Ageing, 2005.

Haylen, B., Frazer, M., & Sutherst, J. R. (1988). Diuretic responses to fluid load in women with urinary incontinence: Optimum duration of pad test. *British Journal of Urology, 62*, 331-333.

Heavner, K. (1998). Urinary incontinence in extended care facilities: a literature review and proposal for continuous quality improvement. *Ostomy/Wound Management, 44*(12), 46-48, 50-43.

Herrman, H., Hawthorne, G., & Thomas, R. (2002). Quality of life assessment in people living with psychosis. *Social Psychiatry and Psychiatric Epidemiology; 37*(11), 510-518.

Herrman, H., Patrick, D., Diehr, P., Martin, M., Fleck, M., Simon, G., et al. (2002). Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. *Psychological Medicine, 32*, 889-902.

Hobart, J., & Thompson, A. (2000). The five item Barthel Index. *Journal Neurology, Neurosurgery & Psychiatry, 71*, 225-230.

Hocking, C., Williams, M., Borad, J., & Baskett, J. (1999). Sensitivity of Shah, Vanclay and Cooper's modified Barthel Index. *Clinical Rehabilitation, 13(2),* 141-147.

Hogan, A., Hawthorne, G., Kethel, L., Giles, E., White, K., Stewart. M., et al. (2001). Health-related quality-of-life outcomes from adult cochlear implantation: a cross-sectional study. *Cochlear Implants International, 2*(2), 115-128.

Hollingworth, W., Mackenzie, R., Todd, C. J., & Dixon A. K. (1995). Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? *Quality of Life Research*, *4*(4), 325-34.

Homma, Y., Kondo, Y., Takahashi, S., Kitamura, T., & Kawabe, K. (2000). Reproducibility of cystometry in overactive detrusor. *European Urology, 38*, 681-685.

Hool, G., Lieber, M., Church, J. (1999). Postoperative anal canal length predicts outcome in patients having sphincter repair for fecal incontinence. *Diseases of the Colon & Rectum, 42*(3), 313-318.

Hsu, T., Raymond, R., & Appell, R. (1999). The supine stress test: a simple method to detect intrinsic urethral sphincter dysfunction. *Journal of Urology, 162*, 460-463.

Hsueh, I., Lin, J., Jeng, J., & Hseih, C. (2002). Comparison of the psychometric characteristics of the functional characteristics of the FIM, 5 item Barthel Index and 10 item Barthel Index in patients with stroke. *Journal Neurology, Neurosurgery & Psychiatry, 73,* 188-190.

Hull, T., Floruta, C., & Piedmonte, M. (2001). Preliminary results of an outcome tool used for evaluation of surgical treatment for faecal incontinence. *Diseases of the Colon and Rectum, 44*(6), 799-805.

Hunskar, S., & Vinsnes, A. (1991). The quality of life in women with urinary incontinence as measured by the sickness impact profile. *Journal of the American Geriatrics Society, 39*(4), 378-382.

Hunt, S., McKenna, S., McEwen, J., Williams, J., & Papp, E. (1981). The Nottingham Health Profile: subjective health status and medical consultations. *Social Science and Medicine, 15A*, 221-229.

Hunt, S. M. (1986). Measuring health in clinical care and clinical trials. In Teeling-Smith, G. *Measuring health: a practical approach.* John Wiley, Chichester.

Hunt, S. M., McKenna, S. P., & Williams, J. (1981). Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteo-arthritis. *Journal of Epidemiology and Community Health, 35*, 297-300.

Imhof, A. (1992). The implications of increased life expectancy for family and social life. In: Wear, A., editor. *Medicine in Society: Historical Essays*. Cambridge: Cambridge University Press.

Jackson, S. (1997a). Female urinary incontinence-symptom evaluation and diagnosis. *European Urology,* 32 Suppl 2, 20-24.

Jackson, S. (1997b). The patient with an overactive bladder - symptoms and quality-of-life issues. *Urology, 50*(Supplement 6A), 18-22.

Jackson, S., Donovan, J., Brookes, S., Eckford, S., Swithinbank, L., & Abrams, P. (1996). The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *British Journal of Urology, 77*(6), 805-812.

Jakobsen, H., Vedel, P., & Andersen, J. (1987). Objective assessment of urinary incontinence: An evaluation of three different pad-weighing tests. *Neurourology and Urodynamics, 6*, 325-330.

Jenkinson, C., Fitzpatrick, R., & Argyle, M. (1988). The Nottingham Health Profile: an analysis of its sensitivity in differentiating illness groups. *Social Science and Medicine, 27*, 1411-1414.

Jenkinson, C. (1991). Why are we weighting? A critical examination of the use of item weights in a health status measure. *Social Science & Medicine, 32*, 1413-1416.

Jenkinson, C., Coulter, A., & Wright, L. (1993). Short form 36 (SF 36) health survey questionnaire: Normative data for adults of working age. *British Medical Journal, 306*, 1437-1440.

Jenkinson, C., Stewart-Brown, S., Petersen, S., & Paice, C. (1999). Assessment of the SF-36 Version 2 in the United Kingdom. *Journal of Epidemiology and Community Health, 53 (*1*), 46-50.*

Jenkinson, C., Coulter, A., & Bruster, S. (2002). The Picker Patient Experience Questionnaire: Development and validation using data from in-patient surveys in five countries. *International Journal of Quality in Health Care, 14(4)* 353-8.

Jeyaseelan, S., Haslam, E., Winstanley, J., Roe, B., & Oldham, J. (2000). An evaluation of a new pattern of electrical stimulation as a treatment for urinary stress incontinence: a randomised, double - blind, controlled trial. *Clinical Rehabilitation, 14*, 631 - 640.

Jorge, J., & Wexner, S. (1993). Etiology and management of faecal incontinence. *Dis Colon Rectum, Diseases of the Colon and Rectum, 36*(1), 77-97.

Jorgensen, J., Jacobsen, H., Bagi, P., Hvarnes, H., & Colstrup, H. (1998). Home uroflowmetry by means of the Da Capo Home Uroflowmeter. *European Urology, 33*, 64-68.

Jorgensen, L., Lose, G., & Andersen, J. (1987). One-hour pad-weighing test for objective assessment of female urinary incontinence. *Obstetrics and Gynecology, 69*(1), 39-42.

Joshi, H. B., Stainthorpe, A., Macdonagh R. P., Keeley, F. X., JR., Timoney, A. G. & Barry, M. J. (2003). Indwelling ureteral stents: evaluation of symptoms, quality of life and utility. *Journal of Urology, 169* (3), 1065-9.

Kalita, J., Shah, S., Kapoor, R., & Misra, U. (2002). Bladder dysfunction in acute transverse myelitis: magnetic resonance imaging and neurophysiological and urodynamic correlations. *Journal Neurology, Neurosurgery & Psychiatry, 73(2),* 154-159.

Kalucy, E., & Baum, F. (1992). The epidemiology of caring: The pattern in a South Australian suburban population. *Australian Journal on Ageing, 11*, 3-8.

Kaplan, R., Anderson, J., & Ganiats, T. (1993). The Quality of Well-Being Scale: rationale for a single quality of life index. In Walker S. and Rosser R. *Quality of Life Assessment: Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers.

Kaplan, R., Bush, J., & Berry, C. (1976). Health status: types of validity and the Index of Well-being. *Health Services Research, 11*(4): 478-507.

Kaplan, R., Bush, J., & Berry, C. (1978). The reliability, stability and generalizability of a health status index. In: *Proceedings of the Social Statistics Section*: American Statistical Association, p. 704-709.

Kaplan, R., Ganiats, T., Sieber, W., & Anderson, J. (1996). The Quality of Well-being Scale. *Medical Outcomes Trust Bulletin*: 2-3.

Kaplan, R. M., Anderson, J.P., Patterson, T.L., McCutchan, J.A., Weinrich, J.D., Heaton, R.K, et al. (1995). Validity of the Quality of Well-Being Scale for persons with human immunodeficiency virus infection. HNRC Group. HIV Neurobehavioral Research Center. *Psychosomatic Medicine, 57*(2), 138-47.

Kaplan, R. M, & Anderson, J. P. (1988). A general health policy model: update and applications. *Health Services Research, 23*(2): 203-35.

Karram, M., & Bhatia, N. (1988). The Q-Tip Test: Standardization of the technique and its interpretation in women with urinary incontinence. *Obstetrics and Gynecology, 71*(6), 807-810.

Katsikitis, M., Davis, A., & Burrow, D. (1996). An Australian community study of Parkinson's disease: 1. Disease severity and functional disability. *Australian Journal of Ageing, 15*, 22-26.

Katz, G., & Rodriguez, R. (2001). Use of a modified American Urological Association Symptom Score for the evaluation of the quality of life of patients with prostate cancer. *Urology, 57*(1), 112-116.

Katzman, McClish, D., Wyman, J., Sale, P., Camp, J., & Earle, B. (1999). Use and costs of incontinence pads in female study volunteers. *Journal of Wound, Ostomy and Continence Nurses Society, 26*(4), 207 – 213.

Kaufmann Rauen, K., Biefeld, T., & Balcom, A. (1998). New continence assessment tool: Development, validity, reliability. *Digital Urology Journal – Urology Nurses*.

Kelleher, C., Cardozo, L., & Toozs-Hobson, P. (1995). Quality of Life and urinary incontinence. *Current Opinion in Obstetric Gynecology, 7*, 404-408.

Kelleher, C., Cardozo, L., Khullar, V., & Salvatore, S. (1997). A new questionnaire to assess the quality of life of urinary incontinent women. *British Journal of Obstetrics and Gynaecology, 104*, 1374-1379.

Kelleher, C., Reese, P., Pheil, A., & Okano, A. (2002). Health-related quality of life of patients receiving extended-release totlterodine for overactive bladder. *American Journal of Managed Care, 8* (19 Suppl), S608 – S615.

Kind, P., & Gudex, C. (1994). Measuring health status in the community: a comparison of methods. *Journal of Epidemiology and Community Health, 48*, 86-91.

Kind, P. (1996). The EuroQoL instrument: an index of health-related quality of life. In: Spilker B, editor. *Quality of Life and Pharmacoeconomics in Clinical Trials*. 2nd ed. Philadelpha: Lippincott-Raven Publishers.

Kirschner, R., Scherr, P., Branch, L., Wetle, T., & Resnick, N. (1998). Accuracy of survey questions for geriatric urinary incontinence. *Journal of Urology, 159*, 1903-1908.

Klovning, A., Hunskaar, S., & Eriksen, B. (1996). Validity of a scored urological history in detecting detrusor instability in female urinary incontinence. *Acta Obsetrica et Gynecologica Scandinavica, 75*, 941-945.

Kobelt, G. (1997). Economic considerations and outcome measurement in urge incontinence. *Urology, 50*(6A Suppl): 100-107.

Kobelt, G. (1997). Economic considerations and outcome measurement in urge incontinence. *Urology, 50*(6A Suppl), 100-107; discussion 108-110.

Kobelt, G., Kirchberger, I., & Malone-Lee, J. (1999). Quality of life aspects of the overactive bladder and the effect of treatment with tolterodine. *British Journal of Urology, 83*, 583-590.

Korman, H. J., Sirls, L. T., & Kirkemo, A. K. (1994) Success rate of modified pereyra bladder neck suspension determined by outcomes analysis. *Journal of Urology, 152*, 1453-1457.

Korsgen, S., Deen, K. I., & Keighley, M. R. (1997). Long-term results of total pelvic floor repair for post-obstetric fecal incontinence. *Diseases of the colon and rectum*, *40*(7), 835-9.

Kromann-Andersen, B., Jakobsen, H., & Andersen, J. (1989). Pad-Weighing Tests: A literature survey on test accuracy and reproducibility. *Neurourology and Urodynamics, 8*, 237-242.

Krahn, M., Ritvo, P., Irvine, J., Tomlinson, G., Bremner, K., Bezjak, A. et al. (2003). Patient and community preferences for outcomes in prostate cancer: implications for clinical policy. *Medical Care, 41*, 153-164.

Kulseng-Hanssen, S. & Berild, G. H. (2002). Subjective and objective incontinence 5 to 10 years after Burch colposuspension. *Neurourology and Urodynamics, 21*(2), 1292-3.

Kulseng-Hanssen, S. (1997). Reliability and validity of stationary cystometry, stationary cysto-urethrometry and ambulatory cysto-urethro-vaginometry. *Acta Obsetrica et Gynecologica Scandinavica, 76*, 33-38.

Lagro-Janssen, A., Debruyne, F., & Van Weel, C. (1991). Value of the patient's case history in diagnosing urinary incontinence in General Practice. *British Journal of Urology, 67*, 569-572.

Langfitt, J. (1995). Comparison of the psychometric characteristics of three quality of life measures in intractable epilepsy. *Quality of Life Research, 4*(2), 101-114.

Lee, L., Eagar, K., & Smith, M. (1998). Subacute and non-acute casemix in Australia. *Medical Journal of Australia, 169*, S22-S25.

Lee, R. S., DeAntoni, E., & Daneshgari, F. (2002). Compliance with recommendations of the urodynamic society for standards of efficacy for evaluation of treatment outcomes in urinary incontinence. *Neurourology and Urodynamics, 21*(5), 482-485.

Leung, P. H. Y., Yip, S. K., Lee, D. T. S., Ho, S. C., & Chung, T. K. H. (2002). *Validity and reliability of the King's Health Questionnaire in a Chinese population.* Paper presented at the International Continence Society Congress, Heidelberg.

Liang, M., Larson, M., Cullen, K., & Schwartz, J. (1985). Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis and Rheumatism, 28*(5): 542-547.

Locher, J., Goode, P., Roth, D., Worrell, R., & Burgio, K. (2001). Reliability assessment of the bladder diary for urinary incontinence in older women. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences, 56*(1), M32-35.

Lose, G. (1997). Urethral pressure measurement. *Acta Obsetrica et Gynecologica Scandinavica, 166*, 39-42.

Lose, G. (1997). Uroflowmetry and pressure/flow study of voiding women. *Acta Obsetrica et Gynecologica Scandinavica, Supplement 166*(76), 43-47.

Lose, G., Fantl, A., Victor, A., Walter, S., Wells, T., Wyman, J., et al. (1998). Outcome measures for research in adult women with symptoms of lower urinary tract dysfunction. *Neurourology and Urodynamics, 17*, 255-262.

Lose, G., Gammelgaard, J., & Jorgensen, T. (1986). The one-hour pad-weighin test: reproducibility and the correlation between the test result, the start volume in the bladder, and the diuresis. *Neurourology and Urodynamics, 5*, 17-21.

Lose, G., Rosenkilde, P., Gammelgaard, J., & Schroeder, T. (1988). Pad-weighing test performed with standardised bladder volume. *Urology, 32*(1), 78-80.

Lubeck, D., Prebil, L., Peeples, P., & Brown, J. (1999). A health related quality of life measure for use in patients with urge urinary incontinence: a validation study. *Quality of Life Research, 8*(4), 337-344.

Lunniss, P., Kamm, M., & Phillips, R. (1994). Factors affecting continence after surgery for anal fistula. *British Journal of Surgery, 81*(9), 1382-1385.

McDowell, I., & Newell, C., (1987). *Measuring Health: A Guide to Rating Scales and Questionnaires.* New York: Oxford University Press.

McDowell, I., & Newell, C., (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires*. (2nd Ed.) New York: Oxford University Press.

McKenna, S., Whalley, D., Renck-Hooper, U., Carlin, S., & Doward, L. (1999). The development of a quality of life instrument for use with post-menopausal women with urogenital atrophy in the UK and Sweden. *Quality of Life Research, 8*(5), 393-398.

McKenna, S., Williamson, T., Renck-Hooper, U., & Whalley, D. (1998). The development of UK and Canadian English versions of the Incontinence Quality of Life Index (IQoLI). *Journal of Drug Assessment, 1*, 49 - 56.

McLennan, M., & Bent, A. (1998). Supine Empty Stress Test as a predictor of Low Valsalva Leak Point Pressure. *Neurourology and Urodynamics, 17*, 121-127.

MacDiarmid, S., Goodson, T., Holmes, T., Martin, P., & Doyle, R. (1998). An assessment of the comprehension of the American Urological Association Symptom Index. *Journal of Urology, 159*(3), 873-874.

MacKenzie, C. R., Charlson, M. E., DiGioia, D., & Kelley, K. (1986). Can the Sickness Impact Profile measure change? An example of scale assessment. *Journal of Chronic Diseases, 39*, 429-438.

Mahoney, F., & Barthel, D. (1965). Functional evaluation: the Barthel Index. *Maryland State Medical Journal, 14*, 61-65.

Makar, A., Thomas, P., Fletcher, M., & Harrison, N. (1998). Transballoon cystometry: a new technique to assess detrusor function after urinary retention. *British Journal of Urology, 82*, 21-24.

Manca, A., Sculpher, J., Ward, K. & Hilton, P. (2003). A cost-utility analysis of tension-free vaginal tape versus colposuspension for primary urodynamic stress incontinence. *Journal of Obstetrics & Gynaecology, 110*(3), 255-262.

Martin, A. (1996). Relative merits of the multi-dimensional health status indexes. AIHW workshop expert group evaluation of measures for the coordinated care trials. AIHW: Canberra.

Martin, D., Engelberg, R., Agel, J., & Swiontkowski, M. (1997). Comparison of the Musculoskeletal Function Assessment questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile health-status measures. *Journal of Bone Joint Surgery America, 79*(9), 1323-1335.

Mattiasson, A., Djurhuus, J., Fonda, D., Lose, G., Nordling, J., & Stohrer, M. (1998). Standardization of outcome studies in patients with lower urinary tract dysfunction: A report on general principles from the Standardisation Committee of the International Continence Society. *Neurourology and Urodynamics, 17*, 249-253.

Matzkin, H., Greenstein, A., Prager-Geller, T., Sofer, M., & Braf, Z. (1996). Do reported micturition symptoms on the American Urological Association questionnaire correlate with 24-hour home uroflowmetry recordings? *Journal of Urology, 155*, 197-199.

Medical Outcomes Trust (1997). *SF-36 Health Survey: Scoring Manual for English-Language Adaptations: Australia/New Zealand, Canada, United Kingdom.* Boston, Massachusetts; (Third printing).

Meyer, S., Francini, K., Achtari, C., & De Grandi, P. (2002). The King's Health Questionnaire: Its correlation with urogynecologic and urodynamic parameters. *Paper presented at International Continence Society Congress*, Heidelberg.

Miklos, J., Sze, E., & Karram, M. (1995). A critical appraisal of the methods of measuring leak-point pressures in women with stress incontinence. *Obstetrics and Gynecology, 86*(3), 349-352.

Mishra, G. & Schofield, M. (1998). Norms for the physical and mental health component summary scores for the SF-36 for young, middle-aged and older Australian women. *Quality of Life Research, 7*, 215-220.

Mittmann, N., Trakas, K., Risebrough, N., & Liu, B. A. (1999). Utility scores for chronic conditions in a community-dwelling population. *Pharmacoeconomics, 15*(4), 369-376.

Montella, J., Ewing, S., & Cater, J. (1997). Visual assessment of urethrovesical junction mobility. *International Urogynecology Journal, 8*, 13-17.

Montz, F., & Stanton, S. (1986). Q-Tip Test in female urinary incontinence. *Obstetrics and Gynecology, 67*(2), 258-260.

Moore, K., Ho, M.T., Lapsley, H., Brown, I. et al. (2004). *Development of a Framework for Economic and Cost Evaluation of Continence Conditions.* (Report). Canberra: Department of Health and Ageing.

Mouritsen, L., Berild, G., & Hertz, J. (1989). Comparison of different methods for quantification of urinary leakage in incontinent women. *Neurourology and Urodynamics, 8*, 579-587.

Mulkay, M., Ashmore, M., & Pinch, T. (1987). Measuring the quality of life: a sociological intervention concerning the application of economics to health care. *Sociology, 21*, 541-564.

Murphy, B., Herrman, H., Hawthorne, G., Pinzone, T., & Evert, H. (2000). *Australian WHOQOL instruments: User's Manual and Interpretation Guide.* Melbourne: Australian WHOQOL Field Study Centre.

Murphy, B. & Hawthorne, G. (2001). *Report of focus group research undertaken for the World Health Organisation Quality of Life for Older Persons (WHOQOL-OLD) Study*. Melbourne: Melbourne WHOQOL Field Study Centre, Australian Centre for Posttraumatic Mental Health, Department of Psychiatry, University of Melbourne.

Muylder, X., Claes, H., Neven, P., & Jaegher, K. (1992). Usefulness of urodynamic investigations in female incontinence. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 44*, 205-208.

MVHGroup. *The Measurement and Valuation of Health: Final Modelling of Valuation Tariffs.* York: Centre for Health Economics; 1995.

Naughton, M., Donovan, J., Badia, X., Corcos, J., Gotoh, M., Kelleher, C. et al. (2004). Symptom severity and QOL scales for urinary incontinence. *Gastroenterology, 126*(1 Suppl 1), S114-123.

Niecestro, R., Einhorn, C., Wheeler, J., Goggin, C., & Nanninga, J. (1992). Use of stresscath for diagnosing stress incontinence. *Urology, 39*(3), 266-269.

Nord, E., Richardson, J., & Macarounas-Kirchmann, K. (1993). Social evaluation of health care versus personal evaluation of health states. Evidence on the validity of four health-state scaling instruments using Norwegian and Australian surveys. International *Journal of Technology Assessment in Health Care, 9*(4): 463-78.

Nord, E. (1993). Unjustified use of the Quality of Well-Being Scale in priority setting in Oregon. *Health Policy, 24*(1): 45-53.

Nord, E. (1997). *A Review of Synthetic Health Indicators*. Oslo: National Institute of Public Health for the OECD Directorate for Education, Employment, Labour and Social Affairs.

Nordenfelt, L., (Ed.). (1994). *Concepts and Measurement of Quality of Life in Health Care*. Dordrecht: Kluwer Academic.

Nordling, J., Abrams, P., Ameda, K., Anderson, J., Donovan, J., Griffiths, D., et al. (1998). Outcome measures for research in treatment of adult males with symptoms of Lower Urinary Tract Dysfunction. *Neurourology and Urodynamics, 17*, 263-271.

Norton, C., & Christiansen, J. (2002). *Anal Incontinence: Report of Committee 15* (Report No. Report of Committee 15). Plymouth UK: Health Publication Ltd.

Norton, N. J. (2002). *Research Priorities for Faecal Incontinence: The Patient's Perspective* (Paper presented at IFFGD Symposium, Advancing the Treatment of Faecal and Urinary Incontinence Through Research). Milwaukee USA: IFFGD Symposium.

NSW Health Department. (2000a). *Report on the 1997 and 1998 NSW Health Surveys*. NSW Health Department, Sydney, 2000. Available at: http://www.health.nsw.gov.au/public-health/nswhs/hsindex.htm. Accessed (June 2003).

NSW Health Department. (2000b). New South Wales Older People's Health Survey 1999. *NSW Public Health Bulletin, 11*(S-2), 1-62.

Nygaard, I., & Holcomb, R. (2000). Reproducibility of the seven-day voiding diary in women with stress urinary incontinence. *International Urogynecology Journal, 11*(1), 15-17.

Nygaard, I., & Zomolek, G. (1995). Exercise Pad Testing in Continent Exercisers: Reproducibility and correlation with voided volume, pyridium staining and type of exercise. *Neurourology and Urodynamics, 14*, 125-129.

O'Brien, B. J, Goeree, R., Bernard, L, Rosner, A., & Williamson, T. (2001). Cost-Effectiveness of tolterodine for patients with urge incontinence who discontinue initial therapy with oxybutynin: a Canadian perspective. *Clinical Therapeutics, 23*(12): 2038-49.

O'Brien, B. (1988). Assessment of treatment in heart disease, in G. Teeling Smith (Ed). *Measuring Health: A Practical Approach. Chichester: John Wiley*.

O'Brien, B. J., Banner, N. R., & Gibson, S. (1988). The Nottingham Health Profile as a measure of quality of life following combined heart and lung transplantation. *Journal of Epidemiology and Community Health, 42*, 232-234.

O'Connor, R. (1993). *Issues in the Measurement of Health-Related Quality of Life*. Melbourne: National Centre for Health Program Evaluation; Working Paper 30.

Ogawa, A., Shimazaki, J., Mitsuya, H., Miyazaki, S., & Kurita, T. (1988). Clinical effects of terodiline hydrochloride on urinary frequency and sense of residual urine - a double blind clinical trial using flavoxate hydrochloride as a control. *Hinyokika Kiyo, 34*(4): 739-753.

Ortiz, O., & Nunez, F. (1996). Dynamic assessment of pelvic floor function in women using the intravaginal device test. *International Urogynecology Journal, 7*, 317-320.

Osterberg, A., Graf, W., Karlbom, U., & Pahlman, L. (1996). Evaluation of a questionnaire in the assessment of patients with faecal incontinence and constipation. *Scandinavian Journal of Gastroenterology, 31*(6), 575-580.

Ottenbacher, K., Hsu, Y., Granger, C., & Fiedler, R. (1996). The reliability of the Functional Independence Measure: a quantitative review. *Archives of Physical Medicine and Rehabilitation, 77,* 1226-1232.

Ouslander, J., Leach, G., Abelson, S., Staskin, D., Blaustein, J., & Raz, S. (1988). Simple versus multichannel cystometry in the evaluation of bladder function in an incontinent geriatric population. *Journal of Urology, 140*, 1482-1486.

Ouslander, J., Simmons, S., Tuico, E., Nigam, J., Fingold, S., Bates-Jensen, B., et al. (1994). Use of a portable ultrasound device to measure post-void residual volume among incontinent nursing home residents. *Journal American Geriatrics Society,* 42, 1189-1192.

Ouslander, J., Urman, H., & Uman, G. (1986). Development and testing of an incontinence monitoring record. *Journal American Geriatrics Society, 34*, 83-90.

Ovretveit, J. (1998). *Evaluating Health Interventions*. Buckingham: Open University Press.

Pager, C. K. (2002). Long-Term Outcomes of Pelvic Floor Exercise and Biofeedback Treatment for Patients with Faecal Incontinence. *Diseases of the Colon and Rectum, 45*(8), 997-1003.

Palmer, M., Czarapata, R., Wells, T., & Newman, D. (1997). Urinary outcomes in older adults: Research and clinical perspectives. *Urologic Nursing, 17*(1), 2-9.

Patrick, D., Martin, M., Bushnell, D., Marquis, P., Andrejasich, C., & Buesching, D. (1999). Cultural adaptation of a quality-of-life measure for urinary incontinence. *European Urology, 36*(5), 427-435.

Patrick, D., Martin, M., Bushnell, D., Yalcin, I., Wagner, T., & Buesching, D. (1999). Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). *Urology, 53*(1), 71-76.

Pearson, B., & Kelber, S. (1996). Urinary incontinence: treatments, interventions, and outcomes. *Clinical Nurse Specialist, 10*(4), 177-182.

Persson, J., Bergqvist, C., & Wolner-Hanssen, P. (2001). An ultra-short perineal pad-test for evaluation of female stress urinary incontinence treatment. *Neurourology and Urodynamics, 20*(3), 277-285.

Pescatori, M., Anastasion, G., Bottini, C., & A. M. (1992). New grading and scoring for anal incontinence. Evaluation of 335 patients. *Diseases of the Colon and Rectum, 35*(5), 482-487.

Pfister, S. (1999). Bladder Diaries and voiding patterns in older adults. *Journal of Gerontological Nursing,* March, 39-40.

Podnar, S., & Vodusek, D. (2001). Protocol for clinical neurophysiologic examination of the Pelvic Floor. *Neurourology and Urodynamics, 20*, 669-682.

Polgar, S. & Thomas, S. A. (2004). *Introduction to Research in the Health Sciences (5th edition)*. Edinburgh: Churchill Livingstone.

Poulsen, E., Kirkeby, H., & Djurhuus, J. (1989). Short and long-term reproducibility of cystometry. *Urology Research, 17*, 197-198.

Purdie, D., Dunne, M., Boyle, F., Cook, M., Najman, J. (2002). Health and demographic characteristics of respondents in an Australian national sexuality survey: comparison with population norms. *Journal of Epidemiology and Community Health, 56*, 748-753.

Rabin, R., & de Charro, F. (2001). EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine, 33*, 337-343.

Rabin, J., McNett, J., & Badlani, G. (1993). Computerized voiding diary. *Neurourology and Urodynamics, 12*, 541-554.

Rabin, J., McNett, J., & Badlani, G. (1996). "Compu-Void II": The Computerized Voiding Diary. *Journal of Medical Systems, 20*(1), 19-34.

Rao, S. & Patel, R. (1997). How useful are manometric tests of anorectal function in the management of defecation disorders? *American Journal of Gastroenterology, 92*, 469-475.

Rao, Y., Saxena, R., Sahni, P., Pande, G., & Chattopadhyay, T. (2002). Functional outcome and patient satisfaction after ileal pouch anal anastomosis for ulcerated colitis in a developing country. *Trop Gastroenterol, 23*(2), 66-69.

Rasmussen, A., Mouritsen, L., Dalgaard, A., & Frimodt-Moller, C. (1994). Twenty-four hour pad weighing teat: Reproducibility and dependency of activity level and fluid intake. *Neurourology and Urodynamics, 13*, 261-265.

Rieger, N., Wattchow, D., Sarre, R., Cooper, S., Rich, C., Saccone, G. (1997). Prospective trial of pelvic floor retraining in patients with fecal incontinence. *Diseases of the Colon & Rectum, 40*(7), 821-826.

Reilly, W., Talley, N., Pemberton, J., & Zinsmeister, A. (2000). Validation of a questionnaire to assess faecal incontinence and associated risk factors: Faecal Incontinence Questionnaire. *Diseases of the Colon and Rectum, 43*(2), 146-153.

Renck - Hooper, U., McKenna, S., & Whalley, D. (1998). Measuring quality of life in female urinary urge incontinence: development and psychometric properties of the IQoL. *Journal of Drug Assessment, 1*, 41 - 48.

Richardson, C. & Zumbo, B. (2000). A statistical examination of the Health Utility Index-Mark III as a summary measure of health status for a general Population Health Survey. *Social Indicators Research, 51*, 171-191.

Richardson, J., & Hawthorne, G. (2000). *Negative Utilities and the Evaluation of Complex Health States: Issues Arising from the Scaling of a Multiattribute Utility Instrument*. Melbourne: Centre for Health Program Evaluation; Report No.: Working Paper 113.

Richardson, J. (1994). Cost utility analysis: what should be measured? *Social Science and Medicine, 39*(1), 7-21.

Richmond, D., Sutherst, J., & Brown, M. (1987). Quantification of urine loss by weighing perineal pads. Observation on the exercise regimen. *British Journal of Urology, 59*, 224-227.

Robb, S. (1985). Urinary incontinence verification in elderly men. *Nursing Research, 34*, 278-282.

Robinson, A., Dolan, P., & Williams, A. (1997). Valuing health states using VAS and TTO: what lies behind the numbers? *Social Science and Medicine*, *45*(8), 1289-1297.

Robinson, D., Anders, K., Cardozo, L., & Bidmead, J. (2003). "Cure" do clinicians know best. *Neurourology & Urodynamics*, *22*, 406-407.

Robinson, D., McClish, D., Wyman, J., Bump, R., & Fantl, A. (1996). Comparison between urinary diaries completed with and without intensive patient instructions. *Neurourology and Urodynamics, 15*, 143-148.

Robinson, D., Pearce, K., Preisser, J., Dugan, E., Suggs, P., & Cohen, S. (1998). Relationship between patient reports of urinary incontinence symptoms and quality of life measures. *Obstetrics and Gynecology, 91*(2), 224-228.

Robinson, J., & Shea, J. (2002). Development and testing of a measure of health-related quality of life for men with urinary incontinence. *Journal of the American Geriatrics Society, 50*(5), 935-945.

Rockwood, T., Church, J., Fleshman, J., Kane, R., Mavrantonis, C., Thorson, A., et al. (1999). Patient and surgeon ranking of the severity of symptoms associated with faecal incontinence: the faecal incontinence severity index. *Diseases of the Colon and Rectum, 42*(12), 1525-1532.

Rockwood, T., Church, J., Fleshman, J., Kane, R., Mavrantonis, C., Thorson, A., et al. (2000). Faecal Incontinence Quality of Life Scale: quality of life instrument for patients with faecal incontinence. *Diseases of the Colon and Rectum, 43*(1), 9-16.

Rockwood, T. (2004). Incontinence severity and QOL scales for fecal incontinence. *Gastroenterology, 126*(1 Suppl 1), S106-13.

Roessner, D. (2002). *Outcome Measurement in the United States: State of the Art.* Paper presented at the annual meeting of the American Association for the Advancement of Science, Boston, MA, February 17, 2002.

Rogers, R., Kammerer-Doak, D., Villarreal, A., Coates, K., & Qualls, C. (2001). A new instrument to measure sexual function in women with urinary incontinence or pelvic organ prolapse. *American Journal of Obstetrics and Gynecology, 184*(4), 552-558.

Rosser, R. (1993). A health index and output measure. In Walker S. and Rosser R. *Quality of Life Assessment: Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers.

Royal College of Physicians. (1995). Incontinence. Causes, management and provision of services. A Working Party of the Royal College of Physicians. *Journal of the Royal College of Physicians of London 29*, 272-274.

Rubenach, S., Shadbolt, B., McCallum, J., & Nakamura, T. (2002). Assessing health-related quality of life following myocardial infarction: is the SF-12 useful? *Journal of Clinical Epidemiology, 55*, 306-309.

Ryhammer, A., Djurhuus, J., & Laurberg, S. (1999). Pad testing in incontinent women: a review. *International Urogynecology Journal and Pelvic Floor Dysfunction, 10*(2), 111-115.

Ryhammer, A., Laurberg, S., Djurhuus, J., & Hermann, A. (1998). No relationship between subjective assessment of urinary incontinence and pad test weight gain in a random population sample of menopausal women. *Journal of Urology, 159*(3), 800-803.

Salyers, M. P., Bosworth, H. B., Swanson, J. W., Lamb-Pagone, J., & Osher, F. C. (2000). Reliability and validity of the SF-12 health survey among people with severe mental illness. *Medical Care, 38*, 1141-1150.

Sand, P., Brubaker, L., & Novak, T. (1991). Simple screening incremental cystometry as a screening method for detrusor instability. *Obstetrics and Gynecology, 77*(3), 453-457.

Sand, P., Hill, R., & Ostergard, D. (1988). Incontinence history as a predictor of detrusor stability. *Obstetrics and Gynecology, 71*(2), 257-259.

Sanderson, K., & Andrews, G. (2002a). The SF-12 in the Australian population: cross-validation of item selection. *Australian & New Zealand Journal of Public Health, 26*, 343-345.

Sanderson, K., & Andrews, G. (2002b). Prevalence and severity of mental health related disability and relationship to diagnosis. *Psychiatric Services, 53*, 80-86.

Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993). Validation of a severity index in female urinary incontinence and its implementation in an epidemiological survey. *Journal of Epidemiology and Community Health, 47*, 497-499.

Sandvik, H., Hunskaar, S., Vanvik, A., Bratt, H., Seim, A., & Hermstad, R. (1995). Diagnostic classification of female urinary incontinence: an epidemiological survey corrected for validity. *Journal of Clinical Epidemiology, 48*, 339-343.

Sandvik, H., Seim, A., Vanvik, A., & Hunskaar, S. (2000). A severity index for epidemiological surveys of female urinary incontinence: Comparison with 48-hour Pad-Weighing Tests. *Neurourology and Urodynamics, 19*, 137-145.

Sansoni, J, and Costi J. (2001) SF-36 Version 1 or Version 2?: The need for Australian normative data. *Proceedings of Health Outcomes 2001: The Odyssey Advances*, 7th Annual National Health Outcomes Conference – 27/28 June 2001, Canberra; AHOC, Centre for Health Service Development, University of Wollongong.

Sansoni, J. (2002). *Health Outcomes: An Australian Perspective.* NZ Health Outcomes Conference, Vision Conferences, Auckland, March 2002.

Santiesteban, A. (1988). Electromyographic and dynamometric characteristics of female pelvic-floor musculature. *Physical Therapy, 68*(3), 344-350.

Schaer, G., Koechli, O., Schuessler, B., & Haller, U. (1995). Perineal Ultrasound for evaluating the bladder neck in urinary stress incontinence. *Obstetrics and Gynecology, 85*(2), 220-224.

Schultz, A., Dickey, G., & Skoner, M. (1997). Self-report of incontinence in acute care. *Urologic Nursing, 17*(1), 23-28.

Schultz, S., & Kopec, J. (2003). Impact of chronic conditions. *Health Reports, 14* (4), 41-53.

Scotti, R., & Myers, D. (1993). A comparison of the cough stress test and single-channel cystometry with multichannel urodynamic evaluation in genuine stress incontinence. *Obstetrics and Gynecology, 81*(3), 430-433.

Scotti, R., Flora, R., Greston, W., Budnivk, L., & Hutchinson-Colas, J. (2000). Characterizing and reporting pelvic floor defects: the Revised New York Classification System. *International Urogynecology Journal, 11*, 48-60.

Scotti, R., Ostergard, D., Guillaume, A., & Kohatsu, K. (1990). Predictive value of urethroscopy as compared to urodynamics in the diagnosis of genuine stress incontinence. *Journal of Reproductive Medicine, 35*(8), 772-776.

Senior, K. A. (2003). *A Gudbala Laif? Health and wellbeing in a remote Aboriginal community – What are the problems and where lies the responsibility?* Ph.D. Thesis, Australian National University.

Shah, S., Vanclay, F., & Cooper, B. (1989). Improving the sensitivity of the Barthel Index for stroke rehabilitation. *Journal of Clinical Epidemiology, 42(8),* 703-709.

Shaw, C., Matthews, R., Perry, S., Assassa, R., Williams, K., & McGrother, C., et al. (2002). Validity and reliability of an interviewer administered questionnaire to measure the severity of lower urinary tract symptoms of storage abnormality: the Leicester Urinary Symptom Questionnaire. *British Journal of Urology, 90*, 205-215.

Shinopulos, N., & Jacobson, J. (1999). Relationship between health promotion lifestyle profiles and patient outcomes of biofeedback therapy for urinary incontinence. *Urologic Nursing, 19*(4), 249-253.

Shone, S., Ryan, S., Rigby, P., & Jutai, J. (2002). Toward a comprehensive evaluation of the impact of electronic aids to daily living: evaluation of consumer satisfaction. *Disability Rehabilitation, 24*(1-3), 115-125.

Shortell, R., & Richardson, W. (1978). *Health Program Evaluation*. St Louis: Mosby.

Shumaker, S., Wyman, J., Uebersax, J., McClish, D., & Fantl, J. (1994). Health-related quality of life measures for women with urinary incontinence: the incontinence impact questionnaire and the urogenital distress inventory. Continence Program in Women (CPW) Research Group. *Quality of Life Research, 3*(5), 291-306.

Sielezneff, I., Malouf, A., Cesari, J., Brunet, C., Sarles, J., Sastre, B. (1999). Selection criteria for internal rectal prolapse repair by Delorme's transrectal excision. *Diseases of the Colon & Rectum, 42*, 367-373.

Siltber, H., Victor, A., & Larsson, G. (1997). Pad weighing tests: the best way to quantify urine loss in patients with incontinence. *Acta Obsetrica et Gynecologica Scandinavica, Supplement, 166*(76), 28-32.

Siltberg, H., Larsson, G., & Victor, A. (1996). Reproducibility of a new method to determine cough-induced leak-point pressure in women with stress urinary incontinence. *International Urogynecology Journal, 7*, 13-19.

Siltberg, H., Larsson, G., Hallen, B., Johansson, C., & Ulmsten, U. (1999). Validation of cough-induced leak point pressure measurement in the evaluation of pharmacological treatment of stress incontinence. *Neurourology and Urodynamics, 18*, 591-602.

Sims, M., Abramson, M., Forbes, A., Glass, D., Ikin, J., Ittak, P., et al. (2003). *Australian Gulf War Veterans' Health Study*. Canberra: Department of Veterans Affairs, 2003.

Simons, A., Dowell, C., Bryant, C., Prashar, S., & Moore, K. (2001). Use of the Dowell Bryant Incontinence Cost Index as a post-treatment outcome measure after non-surgical therapy. *Neurourology and Urodynamics, 20*(1), 85-93.

Simons, A., Yoong, W., Buckland, S., & Moore, K. (2001). Inadequate repeatability of the one-hour pad test: the need for a new incontinence outcome measure. *BJOG, 108*(3), 315-319.

Singh, B., Hawthorne, G., & Vos, T. (2001). The role of economic evaluation in mental health care. *Australian and New Zealand Journal of Psychiatry, 35*, 104-117.

Sintonen, H., & Pekurinen, M. (1993). A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In: Walker, S., Rosser, R., (Eds.). *Quality of Life Assessment.* Dordrecht: Kluwer Academic Publishers.

Sintonen, H. (1994) *The 15D Measure of Health-Related Quality of Life: Reliability, Validity and Sensitivity of its Health State Descriptive System.* Melbourne: National Centre for Health Program Evaluation. Working Paper 41.

Sintonen, H. (1995). *The 15D Measure of Health-Related Quality of Life: Feasibility, Reliability and Validity of its Valuation System.* Melbourne: National Centre for Health Program Evaluation. Working Paper 42.

Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine, 33*, 328-336.

Sirls, L., Kirkemo, A., & Jay, J. (1996). Lack of correlation of the American Urological Association Symptom 7 Index with urodynamic bladder outlet obstruction. *Neurourol Urodyn, 15*(5), 447-456.

Smith, C., & Witherow, R. (2000). The assessment of female pelvic floor dysfunction. *British Journal of Urology, 85*, 579-587.

Smith, H., Taylor, R., & Mitchell, A. (2000). A comparison of four quality of life instruments in cardiac patients: SF-36, QL1, QLM1 and SEIQoL. *Heart, 84*(4), 390-394.

Song, J., Rozanski, T., & Belville, W. (1995). Stress leak point pressure: a simple and reproducible method utilizing a fiberoptic microtransducer. *Urology, 46*(1), 81-84.

Sorensen, S. (1988). Urodynamic investigations and their reproducibility in healthy postmenopausal females. *Scandinavian Journal of Urology and Nephrology, Supplement 114*, 42-47.

Sorensen, S., Gregersen, H., & Sorensen, S. (1988). Long term reproducibility of urodynamic investigations in healthy fertile females. *Scandinavian Journal of Urology and Nephrology, Supplement, 114*, 35-41.

Soroka, D., Glazener, C. M., Hay-Smith, E. J., & Ross, S. (2002). Perineal pad test in evaluating outcome of treatments for female incontinence: a systematic review. *International Urogynecological Journal, 13*(3), 165-175.

Stach-Lempinen, B., Kujansuu, E., Laippala, P, & Metsanoja, R. (2001). Visual analogue scale, urinary incontinence severity score and 15 D-psychometric testing of three different health-related quality-of-life instruments for urinary incontinent women. *Scandinavian Journal of Urology and Nephrology, 35(*6), 476-483.

Staff of the Benjamin Rose Hospital. (1959). A new classification of functional status in activities of daily living. *Chronic Diseases*, *9*, 55-62.

Stavem, K. (1999). Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Quality of Life Research, 8*(1-2), 45-54.

Steven, I., Thomas, S., Eckerman, E., Browning, C., & Dickens, E. (1999). A patient determined general practice satisfaction questionnaire. *Australian Family Physician, 28*(4), 342-348.

Stevenson, C. E. (1996). *SF 36: Interim norms for Australian data*. Canberra: Australian Institute of Health and Welfare.

Stewart, A. L., & Ware, J. E. (1992). Measuring functioning and well-being: The medical outcomes and study approach. *Durham, N C, Duke University Press*.

Streiner, D. L, & Norman, G. R. (1995). *Health Measurement Scales: A Practical Guide to their Development and Use.* 2nd edition, New York: Oxford.

Sturm, J. W, Osborne, R. H, Dewey, H. M, Donnan, G. A, Macdonell, R. A, & Thrift, A. G. (2002). Brief comprehensive quality of life assessment after stroke: the assessment of quality of life instrument in the north East Melbourne stroke incidence study (NEMESIS). *Stroke, 33*(12), 2888-94.

Suchman, E. (1967). *Evaluative Research: Principle and Practice in Public Service and Social Action Programs*. New York: Russell Sage Foundation.

Sutherst, J., & Brown, M. (1984). Comparison of single and multichannel cystometry in diagnosing bladder instability. *British Medical Journal, 288*, 1720-1722.

Sutherst, J., Brown, J., & Richmond, D. (1986). Analysis of the pattern of urine loss in women with incontinence as measured by weighing perineal pads. *British Journal of Urology, 58*, 273-278.

Swift, S., & Utrie, J. (1996). The need for standardization of the Valsalva Leak-Point Pressure. *International Urogynecology Journal, 7*, 227-230.

Swithinbank, L., Donovan, J., du Heaume, J., Rogers, C., James, M., Yang, Q. et al. (1999). Urinary symptoms and incontinence in women: relationships between occurrence, age, and perceived impact. *British Journal of General Practice, 49*, 897-900.

Taylor, A., Wilson, D., Dal Grande, E., Ben-Tovim, D., Elzinga, R., Goldner, R., et al. (2000). Mental health status of the South Australian population. *Australian & New Zealand Journal of Public Health, 24*, 29-34.

Temml, C., Haidinger, G., Schmidbauer, J. Schatzl, G. & Madersbacher, S. (2000). Urinary incontinence in both sexes: prevalence rates and impact upon quality of life and sexual life. *Neurourology and Urodynamics, 19*(3), 259-71.

Theofrastous, J., & Swift, S. (1998). The clinical evaluation of pelvic floor dysfunction. *Obstetrics and Gynecology Clinics of North America, 25*(4), 783-903.

Theofrastous, J., Bump, R., Elser, D., Wyman, J., & McClish, D. (1995). Correlation of urodynamic measures of urethral resistance with clinical measures of incontinence severity in women with pure genuine stress incontinence. *American Journal of Obstetrics and Gynecology, 173*(2), 407-411.

Thorp, J., Jones, L., Wells, E., & Ananth, C. (1996). Assessment of pelvic floor function: A series of simple tests in nulliparous women. *International Urogynecology Journal, 7*, 94-97.

Thorpe, A., Williams, N., Badenoch, D., Blandy, J., & Grahn, M. (1993). Simultaneous dynamic electromyographic proctography and cystometrography. *British Journal of Surgery, 80*, 115-120.

Thyberg, M., Ertzgaard, P., Gylling, M., & Granerus. (1992). Blood pressure response to detrusor pressure elevation in patients with a reflex urinary bladder after a cervical or high thoracic spinal cord injury. *Scandinavian Journal of Rehabilitation Medicine, 24*, 187-193.

Tincello, D., & Richmond, D. (1998). The Larsson frequency/volume chart is not a substitute for cystometry in the investigation of women with urinary incontinence. *International Urogynecology Journal and Pelvic Floor Dysfunction, 9*(6), 391-396.

Torrance, G., Boyle, M., & Horwood, S. (1982). Application of multi-attribute theory to measure social preferences for health states. *Operations Research, 30*, 1043-1069.

Torrance, G., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions: health utilities index. *PharmacoEconomics, 7*(6), 503-520.

Torrance, G. (1986). Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics, 5*, 1-30.

Torrance, G. W, Feeny, D., & Furlong, W. (2001). Visual analog scales: do they have a role in the measurement of preferences for health states? *Medical Decision Making, 21*(4):329-34.

Uebersax, J., Wyman, J., Shumaker, S., McClish, D., & Fantl, J. (1995). Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program for Women Research Group. *Neurourology and Urodynamics, 14*(2), 1311-1319.

Vaizey, C., Garapeti, E., Cahill, J., & Kamm, M. (1999). Prospective comparison of faecal incontinence grading systems. *Gut, 44*(1), 77-80.

Valiquette, L., Duclos, A., & Lapointe, S. (2000). FPSUND: a new clinical classification of urinary incontinence. *Canadian Journal of Urology, 7*(3), 1038-1042.

van Agt, H. M, Essink-Bot, M. L, Krabbe, P. F, & Bonsel, G. J. (1994). Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Social Science and Medicine, 39*(11), 1537-1544.

van der Weide, M. J., Hilbrands, L. B., Bemelmans, B. L. & Kiemeney, L. A. (2004). Lower urinary tract symptoms after renal transplantation: are there changes over time? *Urology*, *63*(3), 442-6.

van Kuijk, A., van der Linde, H., & van Limbeek, J. (2001). Urinary incontinence in stroke patients after admission to a postacute inpatient rehabilitation program. *Archives of Physical Medicine and Rehabilitation, 82*(10), 1407-1411.

van Loenen, N., & Vierhout, M. (1997). Augmentation of urethral pressure profile by voluntary pelvic floor contraction. *International Urogynecology Journal, 8*, 284-287.

van Waalwijk, E., van Doorn, E., Ambergen, A., & Janknect, R. (1997). Detrusor Activity Index: Quantification of Detrusor overactivity by ambulatory monitoring. *Journal of Urology, 157*, 596-599.

Versi, E., Orrego, G., Hardy, E., Seddon, G., Smith, P., & Anand, D. (1996). Evaluation of the home pad test in the investigation of female urinary incontinence. *British Journal of Obstetrics and Gynaecology, 103*(2), 162-167.

Victor, A. (1990). Pad Weighing Test; a simple method to quantitate urinary incontinence. *Annals of Medicine, 22*, 443-447.

Violi, V., Boselli, A., De Bernardinis, M., Costi, R., Trivelli, M., & Roncoroni, L. (2002). A patient-rated, surgeon-corrected scale for functional assessment after total anorectal reconstruction. An adaption of the Working Party on Anal Sphincter Replacement scoring system. *International Journal of Colorectal Disorders, 17*(5), 327-337.

Visser, M., Fletcher, A., Parr, G., Simpson, A., & Bulpitt C. (1994). A comparison of three quality of life instruments in subjects with angina pectoris: the Sickness Impact Profile, the Nottingham Health Profile, and the Quality of Well Being Scale. *Journal of Clinical Epidemiology, 47*(2), 157-163.

von Winterfeldt, D., & Edwards, W. (1986). *Decision Analysis and Behavioural Research*. Cambridge: Cambridge University Press.

Wagner, T., Patrick, D., Bavendam, T., Martin, M., & Buesching, D. (1996). Quality of life of persons with urinary incontinence: development of a new measure. *Urology, 47*(1), 67-71; discussion 71-62.

Walker, S. & Rosser, R. (1993). *Quality of Life Assessment: Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers.

Walker, Q. J., Gebski, V. & Langlands, A. O. (1989). The misuse of mammography in the management of breast cancer revisited. *Medical Journal of Australia, Nov 6, 151(19)* 509-12.

Wan, J., McGuire, E., Bloom, D., & Ritchey, M. (1993). Stress leak point pressure: A diagnostic tool for incontinent children. *Journal of Urology, 150*, 700-702.

Ward, K. L., Hilton, P. & UK and Ireland TVT Trial Group. (2004). A prospective multicenter randomized trial of tension-free vaginal tape and colposuspension for primary urodynamic stress incontinence: two-year follow-up. *American Journal of Obstetrics and Gynecololology, 190*(2), 324-31.

Ware, J. E., Jr., Gandek, B., Kosinski, M., Aaronson, N., Apolone, G., Brazier, J. et al. (1998). The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology, 51*, 1167-1170.

Ware, J. E., Jr. & Kosinski, M. (1996). *SF-36 Health Survey (Version 2) Technical Note*. Boston, MA: Health Assessment Lab.

Ware, J. E., Jr., Kosinski, M., Bayliss, M., McHorney, C., Rogers, W. & Raczek, A. (1995). Comparison of methods for the scoring and statistical analysis of SF 36 Health Profile and Summary Measures: Summary of results from the Medical Outcomes Study. *Medical Care 33*, AS264-AS279.

Ware, J. E., Jr., Kosinski, M. & Dewey, J. E. (2000). *How to Score Version 2 of the SF-36 Health Survey*. Lincoln, RI: QualityMetric Incorporated.

Ware, J. E., Jr., Kosinski, M., & Keller, S. (1994). *SF 36 Physical and Mental Health Summary Scales: A User's Manual*. MA: The Health Institute, New England Medical Center.

Ware, J. E., Jr., Kosinski, M., & Keller, S. (1995). *How to Score the SF-12 Physical and Mental Health Summary Scales.* (2nd Ed.). Boston, MA: The Health Institute. New England Medical Centre.

Ware, J. E., Jr., Kosinski, M., & Keller, S. (1996). A 12-item short-form health survey construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*, 220-233.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1996). SF-12. An even shorter health survey. *Medical Outcomes Trust Bulletin, 4*(2).

Ware, J. E., Jr., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey: Manual and Interpretation Guide*. Boston, MA: The Health Institute. New England Medical Centre.

Weber, A., & Walters, M. (2000). Cost-effectiveness of urodynamic testing before surgery for women with pelvic organ prolapse and stress urinary incontinence. *American Journal of Obstetrics and Gynecology, 183*(6), 1338-1347.

Weber, A., Abrams, P., & Brubaker, L. (2001). The standardization of terminology for researchers in female pelvic floor disorders. *International Urogynaecological Journal of Pelvic Floor Dysfunction, 12*(3), 178-186.

Weidner, A., & Low, V. (1998). Imaging studies of the pelvic floor. *Obstetrics and Gynecology Clinics of North America, 25*(4), 825-848.

Whitehead, W., Wald, A., Diamant, N., Enck, P., Pemberton, J., & Rao, S. (1999). Functional disorders of the anus and rectum. *Gut 45*, 1155-1159.

WHOQoL Group. (1995). The World Health Organisation Quality of Life Assessment (WHOQOL): position paper from the World Health Organisation. *Social Science and Medicine, 41*(10): 1403-1409.

WHOQOL Group. (1998a). The World Health Organisation quality of life assessment (WHOQOL): Development and general psychometric properties. *Social Science & Medicine, 12*, 1569-1585.

WHOQOL Group. (1998b). Group development of the World Health Organisation WHOQOL-BREF quality of life assessment. Psychological Medicine, 28, 551-558.

Wilson, D., Wakefield, M., & Taylor, A. (1992). The South Australian Health Omnibus Survey. *Health Promotion Journal of Australia, 2*, 47-49.

Wilson, D., Tucker, G., & Chittleborough, C. (2002). Rethinking and rescoring the SF-12. *Sozial- und Praventivmedizin, 47*, 172-177.

Wilson, P., Mason, M., Herbison, G., & Sutherst, J. (1989). Evaluation of the home pad test for quantifying incontinence. *British Journal of Urology, 64*, 155-157.

Workman, D., Cassisi, J., & Dougherty, M. (1993). Validation of surface EMG as a measure of intravaginal and intra-abdominal activity: Implications for biofeedback-assisted Kegel exercises. *Psychophysiology, 30*, 120-125.

Wright, J., & Young, N. (1997). A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology, 50*(3): 239-246.

Wyman, J., Choi, S., Harkins, S., Wilson, M., & Fantl, A. (1988). The urinary diary in evaluation of incontinent women: A test-retest analysis. *Obstetrics and Gynecology, 71*(6), 812-817.

Wyman, J., Fantl, J., McClish, D., Harkins, S., Uebersax, J., & Ory, M. (1997). Quality of life following bladder training in older women with urinary incontinence. *International Urogynecology Journal and Pelvic Floor Dysfunction, 8*(4), 223-229.

Wyman, J., Harkins, S., Choi, S., Taylor, J., & Fantl, A. (1987). Psychosocial impact of urinary incontinence in women. *Obstetrics and Gynecology, 70*(3), 378-381.

Yoshioka, K., Ogunbiyi, O., Keighley, M. (1999). A pilot study of total pelvic floor repair or gluteus maximus transposition for postobstetric neuropathic fecal incontinence. *Diseases of the Colon & Rectum, 42,* 252-257.

Appendix

Summary review sheets for recommended tools

Summary review sheets are provided for each of the recommended tools. For the sake of completeness other reviews are also included in this section. The sheets include information concerning where the tools may be obtained and their clinical use and psychometric properties.

It should also be noted that there are several excellent books available that include wide ranging reviews of health related measurement tools. These include:

Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales*, 2nd Ed., Open University Press, Buckingham.

Bowling, A. (2001). *Measuring Disease*, 2nd Ed., Buckingham, Open University Press.

McDowell, I. & Newell, C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires*, 2nd Ed., New York, Oxford University Press.

In addition there are various web resources that contain detailed reviews. The best is probably the QOLID database at www.qolid.org. This site has a public and a subscriber area. The subscriber area provides substantial detail of many of the tools reviewed in this report.

The evidence available to support the psychometric performance of the tools reviewed varies very widely. For some tools such as the "SF" tools, there are hundreds of clinical studies. In others, such as the condition specific tools there may only be handful of studies. Age of a tool is no guide to the volume of literature available to support it. For example the Katz ADL Scale and Barthel Index are very widely used tools but as Bowling (1997) and McDowell & Newell (1996) both note, there are but a few studies that have systematically examined their validity and reliability.

Outcome measurement is a very active area, so as soon as one performs a review, it is outdated. Thus the reader is encouraged to perform their own searches using tools such PUBMED and the QOLID websites to supplement the information presented in his report and the following reviews.

Faecal Incontinence Symptom Severity Measures

Wexner Faecal Incontinence Symptom Severity Scoring System

Title:	Wexner Faecal Incontinence Symptom Severity Scoring System
	or the Cleveland Clinic Florida Faecal Incontinence Score CCF-FIS.
Abbreviations:	Wexner Score.
Author(s) Name:	Steven D Wexner
Author(s) Address:	Department of Colorectal Surgery
	Cleveland Clinic Florida
	Weston FL 33331
	USA
Supplied by:	The instrument is available from publications.
Cost:	None.
Administration Time:	1 minute.
Training Requirements:	None.
Purpose:	To assess symptom severity and QoL impact of Faecal Incontinence symptoms.
Administration:	Self-administered and can be clinician administered.
Instrument Type:	Patient self-report.
Structure:	5 items with 5 point rating scale (Never = 0, rarely = 1, sometimes = 2, usually = 3, always = 4)
Scoring:	Summed ratings for all five items. A score of 0 means no incontinence nor impact and a score of 20 means the worst possible incontinence and impact.
Developed for:	Patients of all ages with faecal incontinence symptoms.
Normative Data:	It has been included in the recent South Australian Health Omnibus Survey (Harrison Health Research, 2004) and a forthcoming report by Hawthorne will examine Australian prevalence data and the psychometric properties of this instrument. Also refer to clinical data below.
Clinical Data:	Extensive US and international data.
	A few clinical studies include:
	Electromyographic biofeedback for fecal incontinence: Beddy et al. (2004).
	Hyperbaric oxygen treatment of fecal incontinence: Cundall et al. (2003).
	Obstetric anal sphincter rupture: Kairaluoma et al. (2004).
	Sacral nerve stimulation: Rasmussen et al. (2004).
Rectal Surgery:	Bittorf et al. (2004); Lim & Ho (2001); Wiley et al. (2004).
Applications:	Individual faecal incontinence symptoms and impact, population studies.

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Adequate	See Vaizey et al. (1999).
Test – retest	Yes	Adequate	See Vaizey et al. (1999).
Inter – rater	NA		

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Adequate	See Vaizey et al. (1999).
Correlation with other measures	Yes	Adequate	See Vaizey et al. (1999); Rothbarth et al. (2001); Nazir et al. (2002).
Construct	Yes	Adequate	See Vaizey et al. (1999); Nazir et al. (2002).
Criterion	Yes	Adequate	See Vaizey et al. (1999).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Adequate	See Vaizey et al. (1999); Kairaluoma et al. (2004).

Cultural Applicability and Cultural Adaptations: Widely used in many countries.

Gender Appropriateness: Suitable for males and females.

Age Appropriateness: All ages.

Summary: The Wexner has been criticised by some for combining impact of faecal incontinence with severity by including the lifestyle question as part of the total score. It has also been noted that it does not address the issue of faecal urgency. However, the Wexner/Cleveland is a very widely used, industry standard, faecal incontinence symptom severity and impact measurement tool. It is suitable for primary care, specialist practitioners and clinical researchers. It is recommended.

References

Beddy, P., Neary, P., Eguare, E. I., McCollum, R., Crosbie, J., Conlon, K. C., & Keane, F. B. (2004). Electromyographic biofeedback can improve subjective and objective measures of fecal incontinence in the short term. *Journal of Gastrointestinal Surgery, 8*, 64-72.

Bittorf, B., Stadelmaier, U., Gohl, J., Hohen Berger, W., & Matzel, K.E. (2004). Functional outcome after intersphincteric resection of the rectum with colo-anal anastomosis in low rectal cancer. *European Journal of Surgical Oncology, 30*, 260-265.

Cundall, J. D., Gardiner, A., Chin, K., Laden, G., Grout, P., & Duthie, G. S. (2003). Hyperbaric oxygen in the treatment of fecal incontinence secondary to pudendal neuropathy. *Diseases of the Colon and Rectum, 46*, 1549-1554.

Jorge, J., & Wexner, S. (1993). Etiology and management of faecal incontinence. *Diseases of the Colon and Rectum, 36*(1), 77-97.

Kairaluoma, M. V., Raivio, P., Aarnio, M. T., & Kellokumpu, I. H. (2004). Immediate repair of obstetric anal sphinter rupture: medium-term outcome of the overlap technique. *Diseases of the Colon and Rectum, 47,* 1358-1363.

Lim, J. F., & Ho, Y. H. (2001). Total colectomy with ileorectal anastomosis leads to appreciable loss in quality of life irrespective of primary diagnosis. *Techniques in Coloproctology, 5*, 79-83.

Nazir, M., Carlsen, E., Jacobsen, A.F., & Nesheim, B.I. (2002). Is there any correlation between objective anal testing, rupture grade, and bowel symptoms after primary repair of obstetric anal sphincter rupture? An observational cohort study. *Diseases of the Colon and Rectum 45*, 1325-1331.

Rasmussen, O. O., Buntzen, S. Sorensen, M., Laurberg, S., & Christiansen, J. (2004). Sacral nerve stimulation in fecal incontinence. *Diseases of the Colon and Rectum 47*, 1158-1162.

Rothbarth, J., Bemelman, W. A., Meijerink, W. J., Stiggelbout, A. M., Zwinderman, A. H., Buyze-Westerweel, M. E., et al. (2001). What is the impact of fecal incontinence on quality of life? *Diseases of the Colon and Rectum, 44,* 67-71.

Vaizey, C., Garapeti, E., Cahill, J., & Kamm, M. (1999). Prospective comparison of faecal incontinence grading systems. *Gut, 44*(1), 77-80.

Wiley, M., Day, P., Rieger, N., Stephens, J., & Moore, J. (2004). Open vs. closed lateral internal sphinterotomy for idiopathic fissure-in-ano: a prospective, randomised controlled trial. *Diseases of the Colon and Rectum, 47,* 847-52.

Date of report:	May 2004
Reporter:	Shane Thomas

Urinary Incontinence Symptom Measures

King's Health Questionnaire

Title:	King's Health Questionnaire.
Abbreviations:	KHQ.
Author(s) Name:	Professor Linda Cardozo.
Author(s) Address:	8 Devonshire Place London W1G 6HP, UK
Email:	lcardozo@compuserve.com
Supplied by:	Available from publications.
Cost:	None but advise author of use.
Administration Time:	5 minutes.
Training Requirements:	None.
Purpose:	To assess impact upon quality of life for people with urinary incontinence symptoms.
Administration:	Self-administered.
Instrument Type:	Patient self report.
Structure:	21 items consisting of three parts. Part 1 contains general health perception and incontinence impact. Part 2 contains role limitations, physical limitations, social limitations, personal relationships, emotions and severity measures. Part 3 contains frequency, nocturia, urgency, urge, stress, nocturnal enuresis, intercourse incontinence, infections, pain, and difficulty in voiding. There is a four point rating system.
Scoring:	8 subscales ("domains") scored between 0 and 100.
Developed for:	Men and women with urinary incontinence.
Normative Data:	See clinical data.
Clinical Data:	Extensive international data. For example, see Reese et al. (2003). Recently, important work on this questionnaire has been published in Japan by Homma & Uemura (2004).
Applications:	QoL impact of urinary incontinence. The King's Health Questionnaire is a recommended tool in a recent European Clinical Practice Guideline. See Viktrup, Summers & Dennett (2004).

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Good	See Kelleher et al. (1997), Donovan et al. (2002) and Reese et al. (2003).
Test – retest	Yes	Good	See Kelleher et al. (1997) & Donovan et al. (2002).
Inter – rater	NA		

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Good	See Kelleher et al. (1997), Donovan et al. (2002) and Kelleher et al. (2002).
Correlation with other measures	Yes	Good	See Kelleher et al. (1997) and Donovan et al. (2002).
Construct	Yes	Good	See Kelleher et al. (1997), Donovan et al. (2002) and Reese et al. (2003).
Criterion	Yes	Good	See Kelleher et al. (1997) and Donovan et al. (2002).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Adequate	See Kelleher et al. (1997), Donovan et al. (2002) and Kelleher et al. (2004).

Cultural and Cultural Adaptations: The instrument is available in at least 26 Languages with Spanish and Portuguese versions having published data (refer QOLID reference below).

Gender Appropriateness: Suitable for males and females.

Age Appropriateness: 17-85 years.

Summary: This tool is in wide use across many countries to assess the impact of urinary incontinence on health related quality of life. It has excellent psychometric properties.

References

Donovan, J. L. (2002). *Symptom and Quality of Life Assessment*. In Abrams, P., Cardozo, L., Khoury, S., &.Wein, A. (Eds). Incontinence: 2nd International Consultation on Incontinence Paris, July 1-3, 2001 2nd Edition 267-316.

Homma, Y., & Uemura, S. (2004). Use of the short form of King's Health Questionnaire to measure quality of life in patients with an overactive bladder. *BJU International, 93,* 1009 – 1013.

Kelleher, C., Cardozo, L., Khullar, V., & Salvatore, S. (1997). A new questionnaire to assess the quality of life of urinary incontinent women. *British Journal of Obstetrics and Gynaecology, 104,* 1374-1379.

Kelleher, C. J., Reese, P. R., Pheil, A. M., & Okano, A. (2002). Health-related quality of life of patients receiving extended-release totlterodine for overactive bladder. *American Journal of Managed Care, 8* (19 Suppl), S608 – S615.

Kelleher, C. J., Pheil, A. M.., Reese, P. R., Burgess, S. M. & Brodish, P. H. (2004). How much is enough and who says so? The case of the King's Health Questionnaire and overactive bladder. *British Journal of Obstetrics and Gynaecology, 111*, 605 – 612.

Reese, P. R., Pheil, A. M., Okano, G. J., & Kelleher, C. J. (2003). Multinational study of reliability and validity of the King's Health Questionnaire in patients with overactive bladder. *Quality of Life Research, 12*, 427 – 442.

Uemura, S. & Homma, Y. (2004). Reliability and validity of King's Health Questionnaire in patients with symptoms of overactive bladder with urge incontinence in Japan. *Neurourology & Urodynamics, 23,* 94 – 100.

Viktrup, L., Summers, K. H., & Dennett, S. L. (2004). Clinical practice guidelines for the initial management of urinary incontinence in women: a European-focused review. *BJU International, 94* (Suppl. 1), 14-22.

Note: A bibliography is available for registered users on the QOLID database at www.qolid.org and a bibliographical search can also be undertaken at phi.uhce.ox.ac.uk/phidb.html

Date of report:	May 2004
Reporter:	Shane Thomas

Urogenital Distress Inventory/Incontinence Impact Questionnaire and their derivatives

Title:	Urogenital Distress Inventory/Incontinence Impact Questionnaire and their derivatives.
Abbreviations:	UDI/IIQ.
Author(s) Name:	Professor Jean F. Wyman Director, Centre for Nursing Research on Elders School of Nursing University of Minnesota 6-101 Weaver-Densford Hall 308 Harvard Street SE Minneapolis, MN 55455 USA Professor Sally A. Shumaker Department of Public Health Sciences Section on Social Sciences and Health Policy Wake Forest University School of Medicine Medicine Centre Boulevard Winston-Salem NC 27157 USA
Supplied by:	Women's Health Center of Excellance, Wake Forest University Baptist Medical Center (www.wfubmc.edu/women/whcoe_iiq_udi_instrument.htm)
Cost:	One-time administrative fee of US$250
Administration Time:	Depends upon the version chosen. The long forms of the UDI and IIQ take 10 minutes. The short forms take 2 to 3 minutes with most respondents.
Training Requirements:	None.
Purpose:	To assess the degree to which symptoms associated with incontinence are troubling to women.
Administration:	Self-administered.
Instrument Type:	Patient self report.
Structure:	Varies according to form chosen. See summary. Symptoms are rated on a 4 point scale: Not at all (0 points), slightly (1 point), moderately (2 points), greatly (3 points).
Scoring:	The ratings for each item are summed and the minimum score for both forms is 0, the maximum score for the long form is 57, the maximum score for the short form is 18.
Developed for:	These instruments were designed to assess the troublesomeness of urinary incontinence problems for adult women but are now being used with men as well.
Normative Data:	It has been included in the recent South Australian Health Omnibus Survey (Harrison Health Research, 2004) and a forthcoming report by Hawthorne will examine Australian prevalence data and the psychometric properties of this instrument. Also refer to clinical data below.
Clinical Data:	European and US studies, but no large scale Australian ones. See van der Vaart et al. (2002) for a major study form the Netherlands.
Applications:	The IIQ is a recommended tool in a recent European Clinical Practice Guideline. See Viktrup, Summers & Dennett (2004).

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Adequate	Brown, Posner & Stewart (1999); Moore & Jensen (2000).
Test – retest	Yes	Adequate	Brown, Posner and Stewart (1999); Hagen et al. (2002); van der Vaart et al. (2003).
Inter – rater	NA		

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Adequate	Brown, Posner and Stewart (1999); Hagen et al. (2002); Corcos et al. (2002); Handal & Massof (2004).
Correlation with other measures	Yes	Adequate	Lee et al. (1995); Brown, Posner and Stewart (1999); Moore & Jensen (2000); Simons et al. (2001); Barber et al. (2001); Harvey et al. (2001); Corcos et al. (2002).
Construct	Yes	Adequate	Brown, Posner and Stewart (1999); Moore & Jensen (2000); Harvey et al. (2001); Hullfish et al. (2004); van Brummen et al. (2004); Handal & Massof (2004).
Criterion	Yes	Adequate	Brown, Posner & Stewart (1999); Lemack & Zimmern (1999).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Adequate	Wyman et al. (1997); Woodman et al. (2001); Hagen et al. (2002); FitzGerald et al. (2001); Richter et al. (2003)

Cultural Applicability and Cultural Adaptations:	These instruments are widely used in Europe.
Gender Appropriateness:	Originally designed for women but now being used for men also.
Age Appropriateness:	Adult 14 years +.
Summary:	This review has assessed the group of instruments. The Urogenital Distress Inventory (UDI) and the Incontinence Impact Questionnaire (IIQ) were developed by Shumaker, Wyman, Uebersax, McClish & Fantl (1994) to assess the impact of urinary incontinence symptoms upon quality of life for women. The original forms of the IIQ and the UDI had 30 and 19 items respectively but work by Uebersax, Wyman, Shumaker, McClish and Fantl (1995) created a 7 item version of the IIQ and a 6 item version of the UDI

and these are now used widely in both clinical and research applications. As these are short instruments they are applicable for use in primary care settings.

The UDI's focus is on the experience of symptoms and the extent to which the patient is bothered by it whereas the focus of the IIQ is primarily on the impact of the symptoms in everyday life. Thus the choice of instrument will depend on the focus of the application. A criticism of the UDI has been that it appears to confound "bother" of the symptom with "presence or absence" of the symptom.

Although the short form of the UDI includes items on type of incontinence (urge/stress) this form only addresses the frequency of "a small amount of urine loss". It does not include the item from the UDI (long form) concerning the frequency of a large amount of urine loss. For this reason it may not be as good a measure of severity (frequency x amount) as other instruments.

The UDI and IIQ have recently been modified by Robinson and Shea (2002) into the Male Urogenital Distress Inventory (MUDI) and the Male Urinary Symptom Impact Questionnaire (MUSIQ).

References

Barber, M. D., Kuchibhatla, M. N., Pleper, C. F. & Bump, R. C. (2001). Psychometric evaluation of 2 comprehensive, condition specific quality of life instruments for women with pelvic floor disorders. *American Journal of Obstetrics and Gynecology, 185,* 1388 - 1395.

Brown, J., Posner, S., & Stewart, A. (1999). Urge incontinence: new health-related quality of life measures. *Journal of the American Geriatrics Society, 47*(8), 980-988.

Corcos, J., Behlouli, H. & Beaulieu, S. (2002). Identifying cut-off scores with neural networks for interpretation of the incontinence impact questionnaire. *Neurourology and Urodynamics, 21*, 198-203.

FitzGerald, M. P., Kenton, K., Shott, S. & Brubaker, L. (2001). Responsiveness of quality of life measurememts to change after reconstructive pelvic surgery. *American Journal of Obstetrics and Gynecology, 185*, 20-24.

Hagen, S., Hanley, J. & Capewell, A. (2002). Test–retest reliability, validity and sensitivity to change of the urogential distress inventory and the incontinence impact questionnaire. *Neurourology and Urodynamics, 21*, 534 – 539.

Handal, V. L. & Massof, R. W. (2004). Measuring the severity of stress urinary incontinence using the incontinence impact questionnaire. *Neurourology and Urodynamics, 23*, 27-32.

Harvey, M. A., Kristjansson, B., Griffith, D. & Versi, E. (2001). The Incontinence Impact Questionnaire and the Urogenital Distress Inventory: a revisit of their validity in women without a urodynamic diagnosis. *American Journal of Obstetrics and Gynecology, 185*, 25-31.

Hullfish, K. L., Bovbjerg, V. E. & Steers, W. D. (2004). Patient-centred goals for pelvic floor dysfunction surgery: long term follow-up. *American Journal of Obstetrics and Gynecology, 191,* 201 – 205.

Lee, P. S., Reid, D. W., Saltmarche, A. & Linton, L. (1995). Measuring the psychosocial impact of urinary incontinence: The York Incontinence Perceptions Scale (YIPS). *Journal of the American Geriatrics Society, 43,* 1275-1278.

Lemack, G. E. & Zimmern, P. E. (1999). Predictability of urodynamics findings based on the Urogenital Distress Inventory – 6 questionnaire. *Urology, 54,* 461-466.

Moore, K. N. & Jensen, L. (2000). Testing of the Incontinence Impact Questionnaire / IIQ-7 with men after radical prostatectomy. *Wound Ostomy and Continence Nursing Journal, 27*, 304-312.

Richter, H. E., Burgio, K. L. Holley, R. L., Goode, P. S., Locher, J. L., Wright, K. C., Varner, R. E. (2003). Cadaveric fascia lata sling for stress urinary incontinence: a prospective quality of life analysis. *American Journal of Obstetrics and Gynecology, 189*, 1590 – 1595.

Robinson, J. P., & Shea, J. A. (2002). Development and testing of a measure of health-related quality of life for men with urinary incontinence. *J Am Geriatr Soc, 50*, 935 – 945.

Simons, A. M., Dowell, C. J., Bryant, C. M., Prashar, S. & Moore, K. H. (2001). Use of the Bowell Bryant Incontinence Cost Index as a post-treatment outcome measure after non-surgical therapy. N*eurourology and Urodynamics, 20,* 85 – 93.

Shumaker, S., Wyman, J., Uebersax, J., McClish, D., & Fantl, J. (1994). Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. *Qual Life Res, 3*(5), 291-306.

Uebersax, J., Wyman, J., Shumaker, S., McClish, D., & Fantl, J. (1995). Short forms to assess life quality and symptom distress for urinary incontinence in women: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program for Women Research Group. *Neurourology and Urodynamics, 14*(2), 1311-1319.

van Brummen, H. J., Heintz, A. P., & van der Vaart, C. H. (2004). The association between overactive bladder symptoms and objective parameters from bladder diary and filling cystometry. *Neurourology and Urodynamics, 23*, 38 -42.

van der Vaart, C. H., de Leeuw, J. R.. J., Roovers, J. P. W. R., & Heintz, A. P. M. (2002). The effect of urinary incontinence and overactive bladder symptoms on quality of life in young women. *British Journal of Urology International, 90*, 544-549.

van der Vaart, C. H., de Leeuw, J. R. J., Roovers, J. P. W. R., & Heintz, A. P. M. (2003). Measuring health-related quality of life in women with urogenital dysfunction: the urogenital distress inventory and incontinence impact questionnaire revisited. *Neurourology and Urodynamics, 22*, 97-104.

Viktrup, L., Summers, K. H., & Dennett, S. L. (2004). Clinical practice guidelines for the initial management of urinary incontinence in women: a European-focused review. *British Journal of Urology International, 94* (Suppl. 1), 14-22.

Woodman, P. J., Misko, C. A. & Fischer, J. R. (2001). The use of short-form quality of life questionnaires to measure the impact of imipramine on women with urge incontinence. *International Urogynecology Journal, 12,* 312-315.

Wyman, J., Fantl, J., McClish, D., Harkins, S., Uebersax, J., & Ory, M. (1997). Quality of life following bladder training in older women with urinary incontinence. *International Urogynecology Journal and Pelvic Floor Dysfunction, 8*(4), 223-229.

Date of report: May 2004

Reporter: Shane Thomas

Incontinence Severity Index

Title:	Incontinence Severity Index.
Abbreviations:	ISI.
Author(s) Name:	Dr. Hogne Sandvik.
Author(s) Address:	Department of Public Health and General Practice University of Bergen Ulriksdal 8c 5009 Bergen Norway
Supplied by:	The instrument is available in the publications listed below.
Cost:	NA.
Administration Time:	<1 minute.
Training Requirements:	None.
Purpose:	To assess urinary incontinence severity in population surveys and for individual clinical assessment.
Administration:	Self-administered.
Instrument Type:	Patient self report.
Structure:	Two items scored as follows:

How often is urine leakage experienced?
Never = 0
Less than once a month = 1
1 to several times a month = 2
1 to several times a week = 3
Every day and/or night = 4

How much urine lost each time?
A few drops = 1
A little = 1
More = 2

Severity index = (points for frequency) x (points for amount).

Scoring:	The minimum score is 0 and the maximum score is 8. The higher the score the more severe is the urinary incontinence.
Developed for:	Public health population surveys of severity of urinary incontinence.
Normative Data:	European and North American (see for example the study by Hannestad et al. (2000) in Norway). This instrument has been included in the recent South Australian Health Omnibus Survey (Harrison Health Research, 2004) and a forthcoming report by Hawthorne will examine Australian prevalence data and the psychometric properties of this instrument.
Clinical Data:	Two clinical studies are listed below:
	Electrical Stimulation for incontinence: Indrekvam et al. (2001).
	Kidney Transplantation: Heit et al. (2004).
Applications:	Assessment of urinary incontinence severity in population surveys and for individual clinical assessment.
	Published clinical practice guidelines, have recommended the Sandvik Incontinence Severity Scale as a tool for monitioring improvement in symptoms for women with stress and urge urinary incontinence; as recorded by the National Guideline Clearinghouse in the USA (AHRQ) [University of Texas at Austin, School of Nursing, Family Nurse Practitioner Program].

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	NA		
Test – retest	Yes	Adequate	Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993).
Inter – rater	NA		

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power Discriminatory power (cont.)	Yes	Adequate	Sandvik, H., Seim, A., Vanvik, A., & Hunskaar, S. (2000). Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993).
Correlation with other measures Correlation with other measures	Yes	Adequate	Sandvik, H., Seim, A., Vanvik, A., & Hunskaar, S. (2000). Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993).
Construct	Yes	Adequate	Sandvik, H., Seim, A., Vanvik, A., & Hunskaar, S. (2000). Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993).
Criterion	Yes	Adequate	Sandvik, H., Seim, A., Vanvik, A., & Hunskaar, S. (2000). Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	No		

Cultural Applicability and Cultural Adaptations:	Widely used in Europe and has had some use in the US.
Gender Appropriateness:	Initially designed for women but suitable for both males and females (see Hunskaar & Sandvik, 1993).
Age Appropriateness:	Adult 14+.
Summary:	The ISI is a useful short form diagnostic severity measure for urinary incontinence, but it does not measure urgency or type of incontinence.

References

Hannestad, Y. S., Rortveit, G., Sandvik, H., & Hunskaar, S. (2000). A community base epidemiological survey of female urinary incontinence: the Norwegian EPINCONT study. Epidemiology of incontinence in the county of Nord-Trondelag. *Journal of Clinical Epidemiology, 53*(11), 1150-1157.

Heit, M., Blackwell, L., Thomas, S., & Ouseph, R. (2004). Prevalence and severity of urinary incontinence in kidney transplant recipients. *Obstetrics & Gynecology, 103*(2), 352-358.

Hunskaar, S., & Sandvik, H. (1993). One hundred and fifty men with urinary incontinence III. Psychosocial consequences. *Scandinavian Journal of Primary Health Care, 11*(3), 193-196.

Indrekvam, S., Sandvik, H., & Hunskaar, S. (2001). A Norwegian national cohort of 3198 women treated with home-managed electrical stimulation for urinary incontinence – effectiveness and treatment results. *Scandinavian Journal of Urology & Nephrology, 35*(1), 32-39.

Sandvik, H., Hunskaar, S., Seim, A., Hermstad, R., Vanvik, A., & Bratt, H. (1993), Validation of a severity index in female urinary incontinence and its implementation in an epidemiological survey. *Journal of Epidemiology and Community Health, 47*(6), 497-499.

Sandvik, H., Hunskaar, S., Vanvik, A., Bratt, H., Seim, A., & Hermstad, R. (1995). Diagnostic classification of female urinary incontinence: an epidemiological survey corrected for validity. *Journal of Clinical Epidemiology, 48*(3), 339-343.

Sandvik, H., Seim, A., Vanvik, A., & Hunskaar, S. (2000). A severity index for epidemiological surveys of female urinary incontinence: Comparison with 48-Hour Pad-Weighing Tests. *Neurourology and Urodynamics, 19*, 137-145.

University of Texas at Austin, School of Nursing, Family Nurse Practitioner Program, (2002). *Recommendations for the Management of Stress and Urge Urinary Incontinence in Women*. Austin (TX), Unversity of Texas at Austin, School of Nursing, 13.

Date of report:	May 2004
Reporter:	Shane Thomas

International Continence Society Male and International Continence Society Male Short Form

Title:	International Continence Society Male and International Continence Society Male (Short Form).
Abbreviations:	ICSmale and ICSmale SF.
Author(s) Name:	Professor Jenny Donovan.
Author(s) Address:	Department of Social Medicine University of Bristol Canynge Hall Whiteladies Road Bristol BS8 2PR United Kingdom
Supplied by:	NA. Available in publications.
Cost:	NA.
Administration Time:	Long form 5 minutes, short form 2 minutes.
Training Requirements:	None.
Purpose:	The tool was developed to measure the symptomatology and "bothersomeness" of lower urinary tract problems for men with prostatic disease.
Administration:	Self-Administered.
Instrument Type:	Patient self-ratings.
Structure:	The standard form has 22 items and the ICSmale Short Form has 11 questions concerning urinary incontinence symptoms for male respondents. Both forms are assessed in the review below.
Scoring:	The ICSmale SF consists of 2 sub-scales. The ICSmale VS (voiding sub-score) comprising five questions (hesitancy, straining, reduced stream, intermittency, incomplete emptying), and the ICSmale IS (incontinence sub-score) comprising six questions (urge, stress, unpredictable and nocturnal incontinence, urgency, postmicturition dribble. The scores are obtained by summing of the ratings for each item.

	For each item/symptom, the patient is asked to assess the degree of bother it is causing. Responses are:
	    No problem
	    A bit of a problem
	    Quite a problem
	    Severe problem

Developed for:	Men.
Normative Data:	See below.
Clinical Data:	Including the following studies:
	Laparoscopic and open prostatectomy for prostate cancer: Hara et al. (2003). Laser ablation of the prostate: Tan et al. (2003). Renal transplantation: van der Weide et al. (2004). Total prostatectomy: Bates et al. (1998); Gacci et al. (2003).
Applications:	To clinically assess and research men with urinary incontinence symptoms.
	An ICSsex questionnaire (4 items) is also available, see Tubaro et al. (2001); and Blanker et al. (2001a, 2001b, 2002).
	An ICSQoL questionnaire (5 multiple choice items plus an open ended question) is also available (see Donovan et al. 1997).

The ICS sex questionnaire, ICS Qol questionnaire and the ICS male questionnaire form a part of the ICS BPH study questionnaire containing over 50 items. See Corcos et al. (2002) for more details.

For key papers from the ICS BPH study see: Abrams et al. 1997; Witjes et al. 1997; Peters et al. 1997; Jolleys et al. 1994.

Finally, it should be noted that the ICSmale was used as the basis to develop the Bristol Female Lower Urinary Tract Symptoms Questionnaire (BFLUTS) for women.

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Adequate	Donovan et al. (2002). Bertaccini et al. (2001). Donovan et al. (1996). Donovan, Abrams, et al. (1996).
Test – retest	Yes	Adequate	Donovan et al. (2002). Donovan et al. (1996). Donovan, Abrams, et al. (1996).
Inter – rater	NA.		

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory Power	Yes	Adequate	Donovan et al. (2002). Donovan et al. (1996). Donovan, Abrams, et al. (1996).
Correlation with other measures	Yes	Adequate	Reynard et al. (1998). Sells et al. (2000). Porru et al. (2001). Bertaccini et al. (2001). Donovan et al. (2002). Donovan et al. (1996). Donovan, Abrams, et al. (1996).
Construct	Yes	Adequate	O'Leary (2000). Barry (2001). Donovan et al. (1996: 2002). Donovan, Abrams, et al. (1996).
Criterion	Yes	Adequate	Donovan et al. (2002). Donovan et al. (1996). Donovan, Abrams, et al. (1996).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Adequate	Gacci et al. (2003). Donovan et al. (2002). Donovan et al. (1999). Donovan et al. (1996). Donovan, Abrams, et al. (1996).

Cultural Applicability and Cultural Adaptations: The instrument is widely used in Europe.

Gender Appropriateness: Men only.

Age Appropriateness: Adult 14+.

Summary: The short form ICSmale should be considered in favour of the full length form, as the longer form does not have major psychometric advantages.

References

Abrams, P., Donovan, J. L., de la Rosette, J. J. & Schafer, W. (1997). International Continence Society "Benign Prostatic Hyperplasia" Study: background, aims, and methodology. *Neurourology & Urodynamics, 16*(2), 79-91.

Bates, T. S., Wright, M. P., & Gillatt, D. A. (1998). Prevalence and impact of incontinence and impotence following total prostatectomy assessed anonymously by the ICS-male questionnaire. *European Urology, 33*(2), 165-169.

Barry M. J. (2001). Evaluation of symptoms and quality of life in men with benign prostatic hyperplasia. *Urology, 58* (Suppl. 1), 25-32.

Bertaccini, A., Vassallo, F., Martino, F., Luzzi, L., Rocca Rossetti, S., Di Silverio, F. et al. (2001). Symptoms, bothersomeness and quality of life in patients with LUTS suggestive of BPH. *European Urology, 40* (Suppl. 1), 13-18.

Blanker, M. H., Bosch, J. L., Groeneveld, F. P., Bohnen, A. M., Prins, A., Thomas, S. et al. (2001a). Erectile and ejaculatory dysfunction in a community-based sample of men 50 to 78 years old: prevalence, concern, and relation to sexual activity. *Urology, 57*(4), 763-768.

Blanker, M. H., Bohnen, A. M., Groeneveld, F. P., Bernsen, R. M., Prins, A., Thomas, S. (2001b). Correlates for erectile and ejaculatory dysfunctionin older Dutch men: a community-based study. *Journal of the American Geriatrics Society, 49*(4), 436-442.

Blanker, M. H., Driessen, L. F., Bosch, J. L., Bohnen, A. M., Thomas, S., Prins, A. et al. (2002). Health status and its correlates among Dutch community-dwelling older men with and without lower urogenital tract dysfunction. *European Urology, 41*(6), 602-607.

Corcos, J., Beaulieu, S., Donovan, J., Naughton, M., & Gotoh, M. (2002). Quality of life assessment in men and women with urinary incontinence. *The Journal of Urology, 168*, 896-905.

de la Rosette, J. J., Witjes, W. P., Schafer, W., Abrams, P., Donovan, J. L., Peters, T. J. et al. (1998). Relationships between lower urinary tract symptoms and bladder outlet obstruction: results from the ICS-"BPH" study. *Neurourology & Urodynamics, 17*(2), 99-108.

Donovan, J., Abrams, P., Peters, T., Kay, H., Reynard, J., Chapples, C., et al. (1996). The ICS - 'BPH' Study: the psychometric validity and reliability of the ICSmale questionnaire. *British Journal of Urology, 77*, 554-562.

Donovan, J., Abrams, P., Kay, H., Reynard, J., Chapple, C., et al. (1996), The psychometric validity and reliability of the ICSmale questionnaire. *British Journal of Urology, 78*(6), 964.

Donovan, J. L, Kay, H. E., Peters, T. J., Abrams, P., Coast, J., Matos-Ferreira, A. et al. (1997). Using the ICSQol to measure the impact of lower urinary tract symptoms on quality of life: evidence from the ICS-'BPH' study. International Continence Society – Benign Prostatic Hyperplasia. *British Journal of Urology, 80*(5), 712-721.

Donovan, J. L., Brookes, S. T., de la Rosette, J. J., Peters, T. J., Porru, D., Kondo, A. et al. (1999). The responsiveness of the ICSmale questionnaire to outcome: evidence from the ICS-'BPH' study. *BJU International, 83*(3), 243-248.

Donovan, J., Peters, T., Abrams, P., Brookes, S., de la Rosette, J., & Schafer, W. (2000). Scoring the short form ICSmale SF questionnaire. *International Continence Society, Journal of Urology, 164*(6), 1948-1955.

Hara, I., Kawabata, G., Miyake, H., Nakamura, I., Hara, S., Okada, H. et al. (2003). Comparison of quality of life following laparoscopic and open prostatectomy for prostate cancer. *Journal of Urology, 169*(6), 2045-2048.

Gacci, M., Bartoletti, R., Figlioli, S., Sarti, E., Eisner, B., Boddi, V. et al. (2003), Urinary symptoms, quality of life and sexual function in patients with benign prostatic hypertrophy before and after prostatectomy: a prospective study. *BJU International, 91*(3), 196-200.

Jolleys, J. V., Donovan, J. L., Nanchahal, K., Peters, T. J., & Abrams, P. (1994). Urinary symptoms in the community: how bothersome are they? *British Journal of Urology, 74*(5), 551-555.

O'Leary, M. P. (2000). LUTS, ED, QOL: alphabet soup or real concerns to aging men? Urology, 56 (Suppl. 1), 7-11.

Peters, T. J., Donovan, J. L, Kay, H. E., Abrams, P., de la Rosette, J. J., Porru, D. et al. (1997). The International Continence Society 'Benign Prostatic Hyperplasia': the bothersomeness of urinary symptoms. *Journal of Urology, 157*(3), 885-889.

Porru, D., Bartoletti, R., Austoni, E., Carrino, M., Gianneo, E., & Melloni, D. (2001), Relationship of flow rate with symptoms, quality of life and other clinical parameters in patients with LUTS suggestive of BPH. *European Urology, 40* (Suppl. 1), 23-27.

Reynard, J. M., Yang, Q., Donovan, J. L., Peters, T. J., Schafer, W., de la Rosette, J. J. et al. (1998). The ICS – 'BPH' Study: uroflowmetry, lower urinary tract symptoms and bladder outlet obstruction. *British Journal of Urology, 82*(5), 619-623.

Sells, H., Donovan, J., Ewings, P., & MacDonagh, R. P. (2000). The development and validation of a quality of life measure to assess partner morbidity in benign prostatic enlargement. *British Journal of Urology International, 85*(4), 440-445.

Tan, A. H., Gilling, P. J., Kennett, K. M., Fletcher, H., & Fraundorfer, M. R. (2003). Long-term results of high-power holmium laser vaporization (ablation) of the prostate. *British Journal of Urology International, 92*(7), 707-709.

Tubaro, A., Polito, M., Giambroni, L., Famulari, C., Gange, E., & Ostardo, E. (2001). Sexual function in patients with LUTS suggestive of BPH. *European Urology, 40* (Suppl. 1), 19-22.

van der Weide, M. J., Hilbrands, L. B., Bemelmans, B. L., & Kiemeney, L. A. (2004). Lower urinary tract symptoms after renal transplantation: are there changes over time? *Urology, 63*(3), 442-446.

Witjes, W. P., de la Rosette, J. J., Donovan, J. L., Peters, T. J., Abrams, P. et al. (1997). The International Continence Society "Benign Prostatic Hyperplasia" Study: international differences in lower urinary tract symptoms and related bother. *Journal of Urology, 157*(4), 1295-1300.

Date of report: May 2004

Reporter: Shane Thomas

Note: In 2004, a new instrument, the International Consultation on Incontinence Questionnaire (ICIQ), was published by some members from the ICSmale team. This developing instrument is modular and is based on a systematic literature review of previous measures. It is designed to assess urinary and faecal incontinence symptoms and their impact on quality of life for both men and women.

Avery, K., Donovan, J., Peters, T. J., Shaw, C., Gotoh, M., & Abrams, P. (2004). ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourology & Urodynamics, 23,* 322-330.

Bristol Female Lower Urinary Tract Symptoms Questionnaire

Title:	Bristol Female Lower Urinary Tract Symptoms Questionnaire.
Abbreviations:	BFLUTS.
Author(s) Name:	S. Jackson, J. Donovan, S. Brookes, S. Eckford, L. Swithinbank & P. Abrams.
Author(s) Address:	Dr. Simon Jackson. Consultant Obstetrician and Gynaecologist John Radcliffe Hospital, Oxford OX3 9DU
Supplied by:	See Jackson, S., Donovan, J., Brookes, S., Eckford, S., Swithinbank, L., & Abrams, P. (1996). The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *British Journal of Urology, 77*(6), 805-812.
Cost:	No cost.
Administration Time:	10 minutes.
Training Requirements:	No formal training is required.
Purpose:	To assess symptomatology of the female lower urinary tract particularly symptom severity, impact upon quality of life and treatment outcomes.
Administration:	Self-administered tool.
Instrument Type:	Urinary incontinence symptoms questionnaire.
Structure:	Essentially the BFLUTS is a symptom checklist with 34 items. When asked about whether they have a particular symptom respondents can reply using a five point scale with the responses being:

Never
Occasionally
Sometimes
Most of the time
All of the time

When asked about the frequency of symptoms they can respond using a five point scale with the responses being:

Never
Once a week
2-3 times a week
Once a day
Several times a day

Also for each symptom, the patient is asked to assess the degree of bother it is causing. Responses are:

No problem
A bit of a problem
Quite a problem
Severe problem

Scoring:	The scores are summed to form an overall symptom severity and frequency index.
Developed for:	Women of all ages who may be experiencing urinary incontinence. NB: This questionnaire is derived from the ICSmale (Jackson et al. 1996).
Normative Data:	Normative data is available in the following papers: Chen, Lin, Hu, Chen & Lin (2003). Jackson, Donovan, Brookes, Eckford, Swithinbank & Abrams (1996). Schatzl, Temml, Schmidbauer, Dolezal, Haidinger & Madersbacher (2000). Swithinbank, Donovan, Rogers & Abrams (2000). Teleman, Lidfeldt, Nerbrand, Samsioe, Mattiasson (2004).

	Temml, Haidinger, Schmidbauer, Schatzl & Madersbacher (2000). Ward, Hilton & the UK and Ireland TVT Trial Group (2004).
Clinical Data:	The BFLUTS has been used in a variety of clinical studies. See: Bo, Talseth & Vinsnes (2000). Jackson, Shepherd, Brookes & Abrams (1999). Kulseng-Hanssen & Berild (2002). Van der Weide, Hilbrands, Bemelmans & Kiemeney (2004). Yoon, Jung, Lee & Lee (2002). Also refer to the previously cited papers in the normative data section.
Applications:	To assess the symptoms associated with urinary incontinence for women. The BFLUTS has also been used with men; see the papers by Schatzl et al. (2000) and Temml et al. (2000). See the recent paper by Brookes et al. (2004) for a scored form of the BFLUTS. Three domains were identified: Incontinence (5 items); Voiding (3 items); Filling (4 items). Also there are sexual function (2 items) and quality of life (5 items).

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Adequate	Cronbach's alpha of 0.78 (Jackson et al. 1996).
Test – retest	Yes	Adequate	The spearman rank order correlation coefficient for the overall symptom score was 0.86 across a 2 week interval (Jackson et al. 1996).
Inter – rater	NA	NA	The BFLUTS is a self-report measure.

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Good	Good discrimination between clinical and non-clinical cases was reported by Jackson et al. (1996).
Correlation with other measures	Yes	Adequate	Acceptable correlations with frequency/volume charts and pad-test data (see Jackson et al. 1996).
Construct	Yes	Adequate	See Jackson et al. (1996); Brookes et al. (2004).
Criterion	Yes	Adequate	See Kulseng-Hanssen et al. (2002); Ward et al. (2004).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Adequate	See Bo et al. (2000); Yoon et al. (2002); Van der Weide et al. (2004).

Cultural Applicability and Cultural Adaptations:

Developed for the English Language but has been used in Europe and Asia.

Gender Appropriateness:

Originally designed for women but now being used for men also. See the papers by Schatzl et al. (2000) and Temml et al. (2000).

Age Appropriateness:

Adults.

Summary:

The BFLUTS has acceptable psychometric properties and has been used in a number of independent clinical studies.

References

Bo, K., Talseth, T., & Vinsnes, A. (2000). Randomized controlled trial on the effect of pelvic floor muscle training on quality of life and sexual problems in genuine stress incontinent women. *Acta Obstetricia et Gynecologica Scandinavica, 79*(7), 598-603.

Brookes, S. T., Donovan, J. L., Wright, M., Jackson, S., & Abrams, P. (2004). A scored form of the Bristol Female Lower Urinary Tract Symptoms questionnaire: data from a randomized controlled trial of surgery for women with stress incontinence. *American Journal of Obstetrics & Gynecology, 191*(1), 73-82.

Chen, G. D., Lin, T. L., Hu, S. W., Chen, Y. C., & Lin, L. Y. (2003). Prevalence and correlation of urinary incontinence and overactive bladder in Taiwanese women. *Neurourology & Urodynamics, 22*(2), 109-117.

Jackson, S., Donovan, J., Brookes, S., Eckford, S., Swithinbank, L., & Abrams, P. (1996). The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *British Journal of Urology, 77*(6), 805-812.

Jackson, S., Shepherd, A., Brookes, S., & Abrams, P. (1999). The effect of oestrogen supplementation on post-menopausal urinary stress incontinence: a double-blind placebo-controlled trial. *British Journal of Obstetrics & Gynaecology, 106*(7), 711-718.

Kulseng-Hanssen, S. & Berild, G. H. (2002). Subjective and objective incontinence 5 to 10 years after Burch colposuspension. *Neurourology and Urodynamics, 21*(2), 1292-3.

Schatzl, G., Temml, C., Schmidbauer, J., Dolezal, B., Haidinger, G., & Madersbacher, S. (2000). Cross-sectional study of nocturia in both sexes: analysis of a voluntary health screening project. *Urology, 56*(1), 71-75.

Swithinbank, L. V., Donovan, J. L., Rogers, C. A., & Abrams, P. (2000). Nocturnal incontinence in women: a hidden problem. *Journal of Urology, 164*(3 Pt 1), 764-766.

Teleman, P. M., Lidfeldt, J., Nerbrand, C., Samsioe, G., & Mattiasson, A. (2004). Overactive bladder: prevalence, risk factors and relation to stress incontinence in middle-aged women. *BJOG: an International Journal of Obstetrics & Gynaecology, 111*(6), 600-604.

Temml, C., Haidinger, G., Schmidbauer, J. Schatzl, G. & Madersbacher, S. (2000). Urinary incontinence in both sexes: prevalence rates and impact upon quality of life and sexual life. *Neurourology and Urodynamics, 19*(3), 259-71.

Van der Weide, M. J., Hilbrands, L. B., Bemelmans, B. L. & Kiemeney, L. A. (2004). Lower urinary tract symptoms after renal transplantation: are there changes over time? *Urology, 63*(3), 442-6.

Ward, K. L., Hilton, P., & UK and Ireland TVT Trial Group. (2004). A prospective multicenter randomized trial of tension-free vaginal tape and colposuspension for primary urodynamic stress incontinence: two-year follow-up. *American Journal of Obstetrics and Gynecolology, 190*(2), 324-31.

Yoon, S. M., Jung, J. K., Lee, S. B. & Lee, T. (2002). Treatment of female urethral syndrome refactory to antibiotics. *Yonsei Medical Journal, 43*(5), 644-651.

Date of report: January, 2004

Reporter: Shane Thomas

American Urological Association (AUA) Symptom Index

Title:	American Urological Association (AUA) Symptom Index
	or the American Urological Association (AUA) Symptom Score or the American Urological Association – 7 BPH Questionnaire.
	The American Urological Association (AUA) Symptom Index is also known as the International Prostate Symptom Score (IPSS).
Abbreviations:	AUASI or IPSS.
Author(s) Name:	Michael J. Barry.
Author(s) Address:	Institute for Health Policy
	Massachusetts General Hospital
	50 Staniford Street (9th Floor)
	Boston MA, 02114
	USA
Supplied by:	A copy of the AUASI is provided in the article by Barry, Fowler, O'Leary, Bruskewitz, Holtgrewe, Mebust et al. (1992) published in the Journal of Urology.
Cost:	Nil.
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Those without these qualifications require basic training in survey administration and the characteristics of the AUASI.
Purpose:	The American Urological Association (AUA) Symptom Index is a symptom measure for lower urinary tract (LUTS) conditions like Benign Prostatic Hyperplasia (BPH).
Administration Time:	3 minutes.
Instrument Type:	Self-report questionnaire.
Structure:	The AUASI consists of 7 questions using a 6 point rating scales. Questions ask about urinary symptoms, for example:
	During the last month or so, how often have you had a weak urinary stream?
	During the last month or so, how many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?
	The AUA Symptom Index usually concludes with a single quality of life question using the delighted – terrible scale.
Scoring:	Item responses for the first 7 questions are added to give a total symptom severity score. The following table shows the relationship between total scores and the severity of symptoms.

Score	Severity
0 – 7	Mild
8 – 19	Moderate
20 – 35	Severe

The AUA Symptom Index can be used in conjunction with the Benign Prostatic Hyperplasia (BPH) Impact Index to get a better picture of the bothersomeness of symptoms.

Developed for:	The AUA Symptom Index was originally developed for Benign Prostatic Hyperplasia in men. However, studies have shown that the AUA Symptom Index is a "non-specific measure of LUTS possibly caused by BPH, and this is reflected by studies that have shown high symptom scores in men with lower urinary tract conditions and in women" (Pinnock & Marshall, 1997).
Normative Data:	Pinnock & Marshall (1997) published South Australian prevalence data on troublesome lower urinary tract symptoms using a modified version of the AUASI. In 2001, McCredie et al. published prevalence data using the IPSS in urban Australian men aged 40 – 69.
	Considerable community-based research has been conducted using the AUASI, especially from Olmsted County in Minnesota/Mayo Clinic. These include: papers by Chute et al. (1993), Girman et al. (1994, 1995), Jacobsen et al. (1993, 1996, 1997) and Roberts et al. (1994, 1997) examining a cohort of 2115 men aged 40 to 79. In another study from Olmsted County, Roberts et al. (1998) assessed community dwelling men and women with urinary incontinence. In 2004, Chung et al. reported data using the original Olmsted cohort examining the association between LUTS and sexual functioning.
	Two other important normative studies have been carried out in the United States by Moon et al. (1997) and Latini et al. (2004).
Clinical Data:	A few clinical studies are listed below:
	Bladder outlet obstruction in women: Groutz et al. (2000).
	Cystoscopy: Stav et al. (2004).
	Diabetes: Michel et al. (2000).
	Drug treatments for BPH: MacDonald & McNicholas (2003); O'Leary (2003).
	Early pelvic floor rehabilitation after transurethral resection of the prostate: Porru et al. (2001).
	Effect of hypertenson on LUTS: Sugaya et al. (2003).
	Erectile Dysfunction: Elliott et al. (2004).
	Genitourinary symptoms in younger men: Collins et al. (1998).
	Melatonin pharamacotherapy for nocturia in men with benign prostatic enlargement: Drake et al. (2004).
	Multiple Sclerosis: Araki et al. (2002).
	Parkinson's Disease: Lemack et al. (2000).
	Prostate Biopsy: Zisman et al. (2001).
	Prostate Cancer: Katz & Rodriguez (2001).
	Real life practice: Naderi et al. (2004).
	Relationship between the duration of the menapause and LUTS: Aygen et al. (2001).
	Spinal cord injury: Abrams et al. (2003).
	Transurethral ethanol injection therapy for BPH: Goya et al. (2004).
	Urethral stricture disease: Heyns & Marais (2002).
	Urethroplasty: Morey et al. (1998).
	Use of Alfuzosin for lower urinary tract symptoms: Lukacs et al. (2000).
	Use of laser ablation of the prostate: Norris et al. (1993); Cummings et al. (1995).
	Use of Zanoterone in men with Benign Prostatic Hyperplasia: Berger et al. (1995).
	Voiding symptoms in pregnancy: Asian et al. (2003).
	Water-Induced Thermotherapy: Muschter (2003).

Two further papers are noteworthy:

Su et al. (1996) found higher AUASI scores for men taking antidepressants and antihistamines, in the Olmsted County study.

Weiss et al. (1999) conducted a detailed investigation of Nocturia in men using the Nocturia question from the AUASI. A similar study was conducted by Fiske et al. (2004) with women.

Applications: The AUASI can be used with an interviewer, though with some care (see Barry et al. 1995, Netto Junior & de Lima, 1995, Rhodes et al. 1995, Plante et al. 1996, MacDiarmid et al. 1998, Cam et al. 2004).

Several noteworthy applications of the symptom index were reported in the recent literature. These included:

Van Schaik et al. (1999) used the AUASI/IPSS questionnaire in a multi-media education program for outpatients.

Wolters et al. (2002) used the AUASI/IPSS to examined effect of social influences on seeking medical care for LUTS conditions.

Wang et al. (2003) used the AUASI/IPSS with videourodynamics to investigate LUTS in young men.

The AUASI has also been used to survey whether discussion about Lower Urinary Tract Symptoms (LUTS) took place in primary care (Collins et al. 1996).

Meigs et al. (1999) reported a study, which defined incidence rates and risk factors for acute urinary retention amongst 8418 male, health professionals in the United States. Welch et al. (1998) and Welch et al. (2002) have also provided important validity data from this sample.

Welch et al. (1998) also found that the AUASI can be divided into filling and voiding symptom subscores. However, the clinical utility of this approach was not supported by Barry et al. (2000).

The evidence of a direct relationship between lower urinary tract symptoms, as measured by AUASI and the bothersomeness of the symptoms and its effect on the quality of life is mixed. Here is a summary of the recent literature:

Articles for:

Scarpero et al. (2003) found that the AUASI correlates well with symptom bother and affect on quality of life for women.

Welch et al. (2002) found that men with high-moderate and severe LUTS had poorer quality of life as measured by the SF-36.

Articles against:

Sarma et al. (2002) found that there is a great deal of variability in longitudinal changes in low urinary tract symptoms and bother for some men.

Cooperberg et al. (2003) found that differences in health related quality of life without any difference in AUASI scores for patients requiring more than 1 pad a day after radical prostatectomy.

A published clinical practice guideline, has recommended the AUASI as a tool for the assessment of symptoms of patients presenting with BPH; as recorded by the National Guideline Clearinghouse in the USA (AHRQ). See also Irani et al. (2003) for a review of published guidelines.

The AUASI/IPSS has also been used in economic evaluations (see Kok et al. 2002; Schulz et al. 2002).

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Barry et al. (1992) Barry et al. (1995)	Good	Cronbach's alpha = 0.86 (Barry et al. 1992).
Test – retest	Yes	Barry et al. (1992) Barry et al. (1995) Barnboym et al. (1999)	Good	Test-retest reliability = 0.92 (Barry et al. 1992).
Inter – rater	NA	NA	NA	The AUASI is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Ichiynagi et al. (1997) Morey et al. (1998)	Adequate	See also the references in the Construct Validity section.

Correlation with other measures	Yes	Barry et al. (1992)	Adequate	A few papers have found a poor correlation between the AUASI and physiological and anatomical measures of the severity of benign prostatic hyperplasia. These include: Barry et al. (1993); Ko et al. (1995); and Roehrborn (1996). (See also the papers in the comments section below.)
		Barry et al. (1993)		
		Jacobsen et al. (1993)		
		Roberts et al. (1994)		
		Ko et al. (1995)		
		Kaplan & Reis (1996)		
		Roehrborn (1996)		
		Roberts et al. (1997)		
		Jenkison et al. (1997)		
		Robers et al. (1998)		
		Morey et al. (1998)		
		Schwartz et al. (1998)		
		Weiss et al. (1999)		
		Wei et al. (2000)		
		Badia et al. (2001)		
		Blanker et al. (2001)		
		Heyns & Marais (2002)		
		Welch et al. (2002)		
		Okamura et al. (2002)		
		Boyle et al. (2003)		
		Ku et al. (2004)		
		Weiss et al. (2004)		

Construct	Yes	Barry et al. (1992)	Adequate	A number of papers have questioned the specificity of the AUASI for Benign Prostatic Hyperplasia. These include: Chancellor & Rivas (1993); Chancellor et al. (1994); Yalla et al. (1994); Ukimura et al. (1995); Roehrborn (1996); Kojima et al. (1997); and Roberts et al. (1998). To resolve this issue, an investigation into Symptom Index's sensitivity and specificity for BPH is required.
		Barry et al. (1993)		
		Chancellor & Rivas (1993)		
		Chancellor et al. (1994)		
		Roberts et al. (1994)		
		Yalla et al. (1995)		This issue is not helped by a lack of agreement about which physiological and anatomical measures to use, the confounding effects of age, co-existing conditions and the poor correlation of symptoms with pathophysiology (see: Barry et al. 1995; Bosch et al. 1995; Bosch et al. 1995; Jacobsen et al. 1995; van Venrooij et al. 1995; Madersbacher et al. 1996; Ezz el Din 1996; Gray 1998; Sciarra et al. 1998; Blanker et al. 2000; Barry, 2001; Eckhardt et al. 2001; Walker et al. 2001; Wadie et al. 2001; Wadie et al. 2001; Cohen et al. 2002; Joshi et al. 2002; Vesely et al. 2003 for more details).
		Ukimura et al. (1996)		
		Roehrborn (1996)		
		O'Leary (1997)		
		Kojima et al. (1997)		
		Gray (1998)		
		Teillac (1998)		
		Welch et al. (1998)		
		Roberts et al. (1998)		See also the above comments about quality of life in Applications section.
		Namasivayam et al. (1998)		
		O'Leary (2000)		
		Barry et al. (2000)		
		Badia et al. (2001)		
		O'Leary (2001)		
		Samra et al. (2002)		
		Cooperberg et al. (2003)		
		Rosen et al. (2003)		
		Scarpero et al. (2003)		

Criterion	Yes	Kaplan et al. (1996)	Adequate	
		Schacterle et al. (1996)		
		Jacobsen et al. (1997)		
		Bucci et al. (2002)		
		Kalra et al. (2003)		

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Barry et al. (1992) Barry et al. (1995) Barry et al. (1995) Rhodes et al. (1995) Jenkinson et al. (1997) Hakenberg et al. (1997) Quek et al. (2002) Masters & Rice (2003) Hakenberg et al. (2003)	Adequate	Ezz el Din et al. (1996, 1996) and O'Connor et al. (2003) report some evidence of clinical variablity. See also the comments in the Construct Validity section.

Cultural Applicability and Cultural Adaptations:	Data using the AUASI has been published for many countries and languages (including Spanish speakers in the United States, see Arocho et al. 1995 and Krongrad et al. 1997). Major studies have been conducted in the: United Kingdom (Trueman et al. 1999); Scotland (Guess et al. 1993); France (Sagnier et al. 1994, 1995); Sweden (Andersson et al. 2004); Norway (Overland et al. 2001) Spain (Badia et al. 1998; Vela-Navarrete et al. 2000); Austria (Madersbacher et al. 1998; Temml et al. 2003); Malaysia (Quek et al. 2001); Singapore (Tan et al. 1997); Korea (Lee et al. 1998); Japan (Ukimura et al. 1996; Masumori et al. 2003); Taiwan (Chen et al. 2003); India (Ganpule et al. 2004); Turkey (Cam et al. 2003; Tuncay et al. 2003), and Latin America (Rodrigues Netto et al. 1997).

Important cultural studies using the AUASI/IPSS have also been conducted in New Zealand (Gray et al. 2004; Nacey et al. 1995).

A number of international, cross-cultural studies have also taken place. These include: Sagnier et al. (1996); Homma et al. (1997); Badia et al. (1997); Boyle et al. (2003).

Gender Appropriateness:	The AUASI was originally designed for Benign Prostatic Hyperplasia (BPH) in men, but now it is being used generally for LUTS conditions in men and women. See studies by: Chancellor & Rivas (1993); Kaplan & Reis (1996); Pinnock & Marshall (1997); Roberts et al. (1998); Madersbacher et al. (1999); Groutz et al. (2000); Kakizaki et al. (2002); Scarpero et al. (2003); Boyle et al. (2003a, 2003b); and Terai et al. (2004).
Age Appropriateness:	18 – 65+ years.
Summary:	The American Urological Association Symptom Index (AUASI)/International Prostate Symptom Score (IPSS) is a very widely used symptom severity measure for patients with benign prostate hyperplasia.

Originally designed for benign prostate hyperplasia, the American Urological Association Symptom Index (AUASI) seems to be evolving into a clinical and epidemiological measure for all LUTS conditions. However, caution is required with regard to its construct validity; as some clarification is needed with regard to the Symptom Index's relationship to age, clinical symptoms, physiological and anatomical measures, the bothersomeness of symptoms, and health-related, quality of life in general. A few studies have also raised some concerns regarding the clinical variability of the AUASI (see Ezz & Din et al. 1996a, 1996b and O'Connor et al. 2003).

In terms of quality of life measurement, using a single item is limiting, as the AUA Guidelines for the Management of Benign Prostatic Hyperplasia state:

> "It is important to note, however, that the QOL question does not capture all quality-of-life issues related to BPH (e.g. Urinary retention and need for surgery)."

References

Abrams, P., Amarenco, G., Bakke, A., Buczynski, A., Castro-Diaz, D., Harrison, S. et al. (2003). Tamsulosin: efficacy and safety in patients with neurogenic lower urinary tract dysfunction due to suprasacral spinal cord injury. *Journal of Urology*, *170*, 1242-1251.

American Urological Association, Inc. 2003. *The Management of Benign Prostatic Hyperplasia*. Baltimore (MD), American Urological Association, Inc.

Andersson, S. O., Rashidkhani, B., Karlberg, L., Wolk, A., & Johansson, J. E. (2004). Prevalence of lower urinary tract symptoms in men aged 45-79 years: a population-based study of 40,000 Swedish men. *BJU International*, 94, 327-331.

Araki, I., Matsui, M., Ozawa, K., Nishimura, M., Kuno, S., & Sadia, T. (2002). Relationship between urinary symptoms and disease-related parameters in multiple sclerosis. *Journal of Neurology, 249*, 1010-1015.

Arocho, R., Kason, N. M., Colon, B., & McMillian, C. A. (1995). (1995). Translation and validation of the American Urological Association symptom index into Spanish. *Clinical Theraputics*, *17*, 777-785.

Asian, D., Asian, G., Yamazhan, M., Ispahi, C., & Tinar, S. (2003). Voiding symptoms in pregnancy: an assessment with international prostate symptom score. *Gynecological & Obstetric Investigation*, *55*, 46-49.

Aygen, E., Ekmekcioglu, O., & Serin, S. (2001). The relationship between the duration of menopause and lower urinary symptoms in women aged 40 – 59. *International Journal of Fertility & Womens Medicine, 46*, 16-22.

Badia, X., Garcia-Losa, M., & Dal-Re, R. (1997). Ten-language translation and harmonization of the International Prostate Symptom Score: developing a methodology for multinational clinical trials. *European Urology,* 129-140.

Badia, X., Garcia-Losa, M., Dal-Re, R., Carballido, J., & Serra, M. (1998). Validation of a harmonized Spanish version of the IPSS: evidence of equivalence with original American scale. International Prostate Symptom Score. *Urology, 52*, 614-620.

Badia, X., Rodiguez, F., Carballido, J., Garcia Losa M., Unda, M., Dal-Re R, et al. (2001). Influence of sociodemographic and health status variables on the American Urological Association symptom scores in patients with lower urinary tract symptoms. *Urology*, *57*, 71-77

Barnboym, E., Ahrens, A., & Roehrborn, C. G. (1998). Effect of scrambling on the short-term reliability of the American Urological Association Symptom Index. *Urology*, *53*, 568-573.

Barry, M. J, Fowler F. J, O'Leary M. P, Bruskewitz R. C, Holtgrewe H. L, Mebust W. K, et al. (1992). The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *Journal of Urology*, *148*, 1549-57.

Barry, M. J. (2001). Evaluation of symptoms and quality of life in men with benign prostatic hyperplasia. *Urology*, *58* (Suppl. 1), 25-32.

Barry, M. J., Cockett, A. T., Holtgrewe, H. L., McConnell, J. D., Sihelnik, S. A., & Winfield, H. N. (1993). Relationship of symptoms of prostatism to commonly used physiological and anatomical measures of the severity of benign prostatic hyperplasia. *Journal of Urology*, *150*, 351-358.

Barry, M. J., Fowler, F. J., Chang, Y. C., Liss, C. L, Wilson, H., & Stek, M. (1995a). The American Urological Association Symptom Index: Does mode of administration affect its psychometric properties? *Journal of Urology*, *154*, 1056-1059.

Barry, M. J., Fowler, F. J., O'Leary, M. P., Bruskewitz, R. C., Holtgrewe, H. L., & Mebust, W. K. (1992). Correlation of the American Urological Association symptom index with self-administered versions of the Madsen-Iversen, Boyarsky and Maine Medical Assessment Program symptom indexes. Measurement Committee of the American Urological Association. *Journal of Urology*, *148*, 1558-1563.

Barry, M. J., Girman, C. J., O'Leary, M. P., Walker-Corkery, E. S., Binkowitz, B. S., Cockett, A. T., et al. (1995c). Using repeated measures of symptom score, uroflowmetry and prostate specific antigen in the clinical management of prostate disease. *Journal of Urology*, *153*, 99-103.

Barry, M. J., Williford, W. O., Chang, Y., Machi, M., Jones, K. M., Walker-Corkery, E. et al. (1995b). Benign prostatic hyperplasia specific health status measures in clinical research: how much change in the American Urological Association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients? *Journal of Urology*, *154*, 1770-1774.

Barry, M. J., Williford, W. O., Fowler, F. J., Jones, K. M., & Lepor, H. (2000). Filling and voiding symptoms in the American Urological Association symptom index: the value of their distinction in a Veterans Affairs randomised trial of medical therapy in men with a clinical diagnosis of benign prostatic hyperplasia. *Journal of Urology*, *164*, 1559-1564.

Berger, B. M., Naadimuthu, A., Boddy, A., Fisher, H. A., McConnell, J. D., Milam, D., et al. (1995). The effect of zanoterone, a steroidal androgen receptor anatgonist, in men with benign prostatic hyperplasia. *Journal of Urology, 154*, 1060-1064.

Blanker, M. H., Groeneveld, F. P., Bohnen, A. M., Bernsen, R. M., Prins, A., Thomas, S., et al. (2001). Voided volumes: normal values and relation to lower urinary tract symptoms in elderly men, a community-based study. *Urology*, 57, 1093-1098.

Blanker, M. H., Groeneveld, F. P., Prins, A., Bersen, R. M., Bohnen, A. M., & Bosch, J. L. (2000). Strong effects of definition and nonresponse bias on prevalence rates of clinical benign prostatic hyperplasia: the Krimpen study of male urogenital tract problems and general health status. *British Journal of Uurology International*, *85*, 665-671.

Bosch, J. L., Hop, W. C., Kirkels, W. J., & Schroder, F. H. (1995). The International Prostate Symptom Score in a community-based sample of men between 55 and 74 years of age: prevalence and correlation of symptoms with age, prostate volume, flow rate and residual urine volume. *British Journal of Urology*, *75*, 622-630.

Bosch, J. L., Hop, W. C., Kirkels, W. J., & Schroder, F. H. (1995). Natural history of benign prostatic hyperplasia: appropriate case definition and estimation of its prevalence in the community. *Urology*, *46* (3 Suppl. A), 34-40.

Boyle, P., Roberston, C., Mazzetta, C., Keech, M., Hobbs, F. D., Fourcare, R., et al. (2003b). The relationship between lower urinary tract symptoms and health status: the UREPIK study. *British Journal of Urology International*, *92*, 575-580.

Boyle, P., Roberston, C., Mazzetta, C., Keech, M., Hobbs, F. D., Fourcare, R., et al. (2003a). The prevalence of lower urinary tract symptoms in men and women in four centres. The UrEpik study. *British Journal of Urology International, 92*, 409-414.

Bucci, J., Morris, W. J., Keyes, M., Spadinger, I., Sidhu, S., & Moravan, V. (2002). Predictive factors of urinary retention following prostate brachytherapy. *International Journal of Radiation Oncology, Biology, Physics*, *53*, 91-98.

Cam, K., Akman, Y., Cicekci, B., Senel, F., & Erol, A. (2004). Mode of administration of international prostate symptom score in patients with lower urinary tract symptoms: physician vs. self. *Prostate Cancer & Prostatic Diseases*, *7*, 41-44.

Cam, K., Senel, F., Akman, Y., & Erol, A. (2003). The efficacy of an abbreviated model of the International Prostate Symptom Score in evaluating benign prostatic hyperplasia. *British Journal of Urology International*, *91*, 186-189.

Chancellor, M. B & Rivas, D. A. American Urological Association symptom index for women with voiding symptoms: lack of index specificity for benign prostate hyperplasia. *Journal of Urology* 1993;150:1706-1709.

Chancellor, M. B., Rivas, D. A, Keeley, F. X., Lotfi, M. A., & Gomella, L. G. (1994). Similarity of the American Urological Association Symptom Index among men with benign prostate hyperplasia (BPH), urethral obstruction not due to BPH and detrusor hypereflexia without outlet obstruction. *British Journal of Urology*, *74*, 200-203.

Chen, T. I., Hsu, Y. S., & Wu, T. T. (2003). Lower urinary tract symptoms and uroflow in a community-based sample of Taiwanese men. *Journal of the Chinese Medical Association*, *66,* 84-88.

Chung, W. S., Nehra, A., Jacobson, D. J., Roberts, R. O., Rhodes, T., Girman, C. J. et al. (2004). Lower urinary tract symptoms and sexual dysfunction in community-dwelling men. *Mayo Clinic Proceedings, 79,* 745-749.

Chute, C. G., Panser, L. A., Girman, C. J., Oesterling, J. E., Guess, H. A., Jacobsen, S. J. et al. (1993). The prevalence of prostatism: a population-based survey of urinary symptoms, *Journal of Urology*, *150,* 85-89.

Cohen, D. D., Steinberg, J. R, Rossignol, M., Heaton, J., & Corcos, J. (2002). Normal variation and influence of stress, caffeine intake, and sexual activity on uroflowmetry parameters of a middle-aged asymptomatic cohort of volunteer male urologists. *Neurourology & Urodynamics*, *21*, 491-494.

Collins, M. F, Friedman, R. H, Ash, A., Hall, R., & Moskowitz, M. A. (1996). Underdetection of clinical benign prostatic hyperplasia in a general medical practice. *Journal of General Internal Medicine*, *11*, 513-518.

Collins, M. M., O'Leary, M. P., & Barry, M. J. (1998). Prevalence of bothersome genitourinary symptoms and diagnoses in younger men on routine primary care visits. *Urology, 52*, 422-427.

Cooperberg, M. R, Master, V. A, & Carroll, P. R. (2003). Health related quality of life significance of single pad urinary incontinence following radical prostatectomy. *Journal of Urology*, *170*, 512-515.

Cummings, J. M., Parra, R. O., & Boullier, J. A. (1995). Laser prostatectomy: initial experience and urodynamic follow-up. *Urology*, *45*, 414-418.

Drake, M. J., Mills, I. W., & Noble, J. G. (2004). Melatonin pharmacotherapy for nocturia in men with benign prostatic enlargement. *Journal of Urology*, *171*, 1199-1202.

Eckhardt, M. D., van Venrooij, G. E., & Boon, T. A. (2001). Symptoms and quality of life versus age, prostate volume, and urodynamic parameters in 565 strictly selected men with lower urinary tract symptoms suggestive of benign prostatic hyperplasia. *Urology, 57*, 695-700.

Elliott, S. P, Gulati, M., Pasta, D. J., Spitalny, G. M., Kane, C. J., Yee, R., et al. (2004). Obstructive lower urinary tract symptoms correlate with erectile dysfunction. *Urology, 63,* 1148-1152.

Ezz el Din, K. E., Kiemeney, L. A., de Wildt, M. J., Rosier, P. F, Debruyne, F. M, & de la Rosette, J. J. (1996b). The correlation between bladder outlet obstruction and lower urinary tract symptoms as measured by the international prostate symptom score. *Journal of Urology*, *156*, 1020-1025.

Ezz el Din, K. E., Kiemeney. L. A., de Wildt, M. J., Debruyne, F. M., & de la Rosette, J. J. (1996c). Correlation between uroflowmetry, prostate volume, postvoid residue, and lower urinary tract symptoms as measured by the International Prostate Symptom Score. *Urology, 48*, 393-397.

Ezz el Din, K. E., Koch, W. F., de Wildt, M. J., Kiemeney, L. A., Debruyne, F. M., & de al Rosette, J. J. (1996a). Reliability of the International Prostate Symptom Score in the assessment of patients with lower urinary tract symptoms and/or benign prostatic hyperplasia. *Journal of Urology*, *155*, 1959-1964.

Fiske, J., Scarpero, H. M., Xue, X., & Nitti, V. W. (2004). Degree of bother caused by nocturia in women. *Neurourology & Urodynamics, 23*, 130-133.

Ganpule, A. P., Desai, M. R., Desai, M. M., Wani. K. D., & Bapat, S. D. (2004). Natural history of lower urinary tract symptoms: preliminary report form a community based Indian study. *British Journal of Urology International*, *94*, 322-334.

Girman, C. J., Epstein, R. S., Jacobsen, S. J., Guess, H. A., Panser, L. A., Oesterling, J. E. et al. (1994). Natural history of prostatism: impact of urinary symptoms on quality of life in 2115 randomly selected community men, *Urology, 44,* 825-831.

Girman, C. J., Jacobsen, S. J., Guess, H. A., Oesterling, J. E., Chute, C. G., Panser, L. A. et al. (1995). Natural history of prostatism: relationship among symptoms, prostate volume and peak flow rate, *Journal of Urology, 153,* 1510-1515.

Goya, N., Ishikawa, N., Ito, F., Kobayashi, C., Tomizawa, Y., & Toma, H. (2004). Transurethral ethanol injection therapy for prostatic hyperplasia: 3 year results. *Journal of Urology 172*, 1017-1020.

Gray, M. (1998). Psychometric evaluation of the international prostate symptom score. *Urologic Nursing, 18*, 175-183.

Gray, M. A., Crampton, P., Weinstein, P., & Nacey, J. N. (2004). Differences in prostate disease symptoms and visits to the general practitioner among three ethnic groups in New Zealand. *British Journal of Urology International*, *94,* 96-100.

Groutz, A., Blaivas, J. G., Fait, G., Sassone, A. M., Chaikin, D. C., & Gordon, D. (2000). The signifiance of the American Urological Association symptom index score in the evaluation of women with bladder outlet obstruction. *Journal of Urology*, *163*, 207-211.

Guess, H. A., Chute, C. G., Garraway, W. M., Girman, C. J., Panser, L. A., Lee, R. J., et al. (1993). Similar levels of urological symptoms have similar impact on Scottish and American men – although Scots report less symptoms. *Journal of Urology*, *150*, 1701-1705.

Hakenberg, O. W., Pinnock, C. B., & Marshall, V. R. (1997). Does evaluation with the International Prostate Symptom Score predict the outcome of transurethral resection of the prostate? *Journal of Urology*, *158*, 94-99.

Hakenberg, O. W., Pinnock, C. B., & Marshall, V. R. (2003). Preoperative urodynamic and symptom evaluation of patients undergoing transurethral prostatectomy: analysis of variables relevant for outcome. *BJU International, 91*, 375-379.

Heyns, C. F., & Marais, D. C. (2002). Prospective evaluation of the American Urological Association symptom index and peak urinary flow rate for the follow-up of men with known urethral stricture disease. *Journal of Urology, 168*, 2051-2054.

Homma, Y., Kawabe, K., Tsukamoto, T., Yamanaka, H., Okada, K., & Okajima, E. (1997). Epidemiologic survey of lower urinary tract symptoms in Asia and Australia using the international prostate symptom score. *International Journal of Urology*, *4*, 40-46.

Ichiyanagi, O., Ishigooka, M., Hashimoto, T., Hayani, S., Suzuki, Y., & Nakada. T. (1997). The American Urological Association Symptom Index: early postoperative evaluation of irritative and obstructive symptoms due to benign prostatic hyperplasia. *International Urology & Nephrology*, *29*, 441-447.

Irani, J., Brown, C.T., van der Meulen, J., Emberton, M. (2003). A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? *British Journal of Urology International*, *92*, 937-942.

Jacobsen, S. J., Girman, C. J., Guess, H. A., Panser, L. A., Chute, C. G., Oesterling, J. E. et al. (1993). Natural history of prostatism: factors associated with discordance between frequency and bother of urinary symptoms, *Urology, 42,* 663-671.

Jacobsen, S. J., Girman, C. J., Guess, H. A., Panser, L. A., Chute, C. G., Oesterling, J. E., et al. (1995). Do prostate size and urinary flow rates predict health care-seeking behavior for urinary symptoms in men? *Urology*, *45,* 64-69.

Jacobsen, S. J., Girman, C. J., Guess, H. A., Rhodes, T., Oesterling, J. E., Lieber, & M. M. (1996). Natural history of prostatism: longitudinal changes in voiding symptoms in community dwelling men, *Journal of Urology, 155,* 595-600.

Jacobsen, S. J., Jacobson, D. J., Girman, C. J., Roberts, R. O., Rhodes. T., Guess, H. A., et al. (1997). Natural history of prostatism: risk factors for acute urinary retention, *Journal of Urology*, *158*, 481-487.

Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S., & Layte, R. (1997). Comparison of the Medical Outcomes Study 36 Item Short Form Health Survey, EuroQol, and disease specific measures. *Medical Care*, *35*, 1109-1118.

Joshi, H. B., Okeke, A., Newns, N., Keeley, F. X., & Timoney, A. G. (2002). Characterization of urinary symptoms in patients with ureteral stents. *Urology*, *59*, 511-516.

Kakizaki, H., Matsuura, S., Mitsui, T., Ameda, K., Tanaka, H., & Koyanagi, T. (2002). Questionnaire analysis on sex differences in lower urinary tract symptoms. *Urology*, *59*, 58-62.

Kalra, P., Togami, J., Bansal, B. S. G., Partin, A. W., Brawer, M. K., Babaian, R. J., et al. (2003). A neurocomputational model for prostate carcinoma detection. *Cancer*, *98*, 1849-1854.

Kaplan, S. A., & Reis, R. B. (1996). Significant correlation of the American Urological Association symptom score and a novel urodynamic parameter: detrusor contraction duration. *Journal of Urology*, *156*, 1668-1672.

Kaplan, S. A., Olssonm C. A., & Te, A. E. (1996). The American Urological symptom score in the evaluation of men with lower urinary tract symptoms: at 2 years of follow-up, does it work? *Journal of Urology*, *155*, 1971-1974.

Katz, G., & Rodriguez, R. (2001). Use of a modified American Urological Association Symptom Score for the evaluation of the quality of life of patients with prostate cancer. *Urology*, *57*, 112-116.

Ko, D. S., Fenster, H. N., Chambers, K., Sullivan, L. D., Jens, M., & Goldenberg, S. L. (1995). The correlation of multichannel urodynamic pressure-flow studies and American Urological Association symptom index in the evaluation of benign prostatic hyperplasia. *Journal of Urology*, *154*, 396-398.

Kojima, M., Naya, Y., Inoue, W., Ukimura, O., Watanabe, M., Saitoh, M., et al. (1997). The American Urological symptom index for benign prostatic hyperplasia as a function of age, volume and ultrasonic appearance of the prostate. *Journal of Urology*, *157*, 2160-2165.

Kok, E. T., McDonnell, J., Stolk, E. A., Stoevelaar, H. J., Busschbach, J. J., et al. The valuation of the International Prostate Symptom Score (IPSS) for use in economic evaluations. *European Urology*, *42*, 491-497.

Krongrad, A., Perczek, R. E., Burke, M. A., Granville, L. J., Lai, H., & Lai, S. (1997). Reliability of Spanish translations of select urological quality of life instruments. *Journal of Urology*, *158*, 493-496.

Ku, J. H., Jeong, I. G., Lim, D. J., Byun, S. S, Paick, J. S., & Oh, S. J. (2004). Voiding diary for the evaluation of urinary incontience and lower urinary tract symptoms: prospective assessment of patient compliance and burden. *Neurourology & Urodynamics*, *23*, 331-335.

Latini, J. M., Mueller, E., Lux, M. M., Fitzgerald, Kreder, K. J. (2004). Voiding frequency in a sample of asymptomatic American men. *Journal of Urology, 172,* 980-984.

Lee, E., Yoo, K. Y., Kim, Y., Shin, Y., & Lee, C. (1998). Prevalence of lower urinary tract symptoms in Korean men in a community-based study. *European Urology*, *33*, 17-21.

Lemack, G. E., Dewey, R. B., Roehrborn, C. G., O'Suilleabhain, P. E., & Zimmern, P. E. (2000). Questionnaire-based assessment of bladder dysfunction in patients with mild to moderate Parkinson's disease. *Urology, 56,* 250-254.

Lukacs, B., Grange, J. C., & Comet, D. (2000). One year follow-up of 2829 patients with moderate to severe lower urinary tract symptoms treated with alfuzosin in general practice according to IPSS and a health-related quality-of-life questionnaire. *Urology*, *55*, 540-546.

MacDiamid, S. A, Goodson, T. C, Holmes, T. M, Martin, P. R, & Doyle, R. B. (1998). An assessment of the comprehension of the American Urological Association Symptom Index. *Journal of Urology*, *159*, 873-874.

MacDonald, D., & McNicholas, T. A. (2003). Drug treatments for lower urinary tract symptoms secondary to bladder outflow obstruction: focus on quality of life. *Drugs, 63,* 1947-1962.

Madersbacher, S., Haidinger, G., Temml, C., & Schmidbauer, C. P. (1998). Prevalence of lower urinary tract symptoms in Austria as assessed by an open survey of 2,096 men. *European Urology*, *34*, 136-141

Madersbacher, S., Klinger, H. C., Schatzl, G., Stulnig, T., Schmidbauer, C. P., & Marberger, M. (1996). Age related urodynamic changes in patients with benign prostatic hyperplasia. *Journal of Urology*, *156*, 1662-1667.

Madersbacher, S., Pycha, A., Klinger, C. H., Schatzi, G., & Marberger, M. (1999). The International Prostate Symptom score in both sexes: a urodynamics-based comparison. *Neurourology & Urodynamics*, *18*, 173-182.

Masters, J. G., & Rice, M. L. (2003). Improvement in urinary symptoms after radical prostatectomy: a prospective evaluation of flow rates and symptom scores. *British Journal Urology International*, *91*, 795-797.

Masumori, N., Tanaka, Y., Takahashi, A., Itoh, N., Ogura, H., Furuya, S, et al. (2003). Lower urinary tract symptoms of men seeking medical care – comparison of symptoms found in the clinical setting and in a community study. *Urology*, *62*, 266-272.

McCredie, M., Staples, M., Johnson, W., English, D. R, & Giles, G. G. (2001). Prevalence of urinary symptoms in urban Australian men aged 40-69. *Journal of Epidemiology & Biostatistics, 6*, 211-218.

Meigs, J. B., Barry, M. J., Giovannucci, E., Rimm, E. B., Stampfer, M. J., & Kawachi, I. Incidence rates and risk factors for acute urinary retention: the health professionals' followup study. *Journal of Urology, 162*, 376-382.

Michel, M. C., Mehlburger, L., Schumacher, H., Bressel, H. U., & Goepel, M. (2000). Effect of diabetes on lower urinary tract symptoms in patients with benign prostatic hyperplasia. *Journal of Urology*, *163,* 1725-1729.

Moon, T. D., Hagen, L., & Heisey, D. M. (1997). Urinary symptomatology in younger men. *Urology, 50,* 700-703.

Morey, A. F., McAninch, J. W., Duckett, C. P., & Rogers, R. S. (1998). American Urological Association symptom index in the assessment of urethroplasty outcomes. *Journal of Urology, 159*, 1192-1194.

Muschter, R. (2003). Conductive heat: hot water-induced thermotherapy for ablation of prostatic tissue. *Journal of Endourology,* 17, 609-616.

Nacey, J. N., Morum, P., & Delahunt, B. (1995). Analysis of the prevalence of voiding symptoms in Maori, Pacific Island, and Caucasian New Zealand men. *Urology*, *46*, 506-511.

Naderi, N., Mochtar, C. A, & de la Rosette, J. J. (2004). Real life practice in the management of benign prostatic hyperplasia. *Current Opinion in Urology*, *14*, 41-44.

Namasivayam, S., Minhas, S., Brooke, J., Joyce, A. D., Prescott, S., & Eardley, I. (1998). The evaluation of sexual function in men presenting with symptomatic benign prostatic hyperplasia. *British Journal of Urology*, *82*, 842-846.

Netto Junior N. R., & de Lima, M. L. (1995). The influence of patient education level on the International Prostatic Symptom Score. *Journal of Urology*, *154*, 97-99.

Norris, J. P., Norris, D. M., Lee, R. D., & Rubenstein, M. A. (1993). Visual laser ablation of the prostate: clinical experience in 108 patients. *Journal of Urology, 150*, 1612-1614.

O'Connor, R. C., Bales, G. T., Avila, D., & Gerber, G. S. (2003). Variability of the International Prostate Symptom Score in men with lower urinary tract symptoms. *Scandinavian Journal of Urology & Nephrology*, *37*, 35-37.

O'Leary, M. (1997). The importance of standardisation and validation of symptom scores and quality of life: the urologist's point of view. *European Urology*, *32* (Suppl. 2), 48-49.

O'Leary, M. P. (2000). LUTS, ED, QOL: alphabet soup or real concerns to aging men? *Urology, 56* (Suppl. 1), 7-11.

O'Leary, M. P. (2001). Quality of life and sexuality: methodological aspects. *European Urology, 40* (Suppl. 3), 13-18.

O'Leary, M. P. (2003). Lower urinary tract symptoms/benign prostatic hyperplasia: maintaining symptom control and reducing complications. *Urology, 62*, 15-23.

Okamura, K., Usami, T., Nagahama, K., Maruyama, S., & Mizuta, E. (2002). "Quality of Life" assessment of urination in elderly Japanese men and women with some medical problems using International Prostate Symptom Score and King's Health Questionnaire. *European Urology*, *41*, 411-419.

Overland, G. B., Vatten, L., Rhodes, T., DeMuro, C., Jacobsen, G., Vada, K., et al. (2001). Lower urinary tract symptoms, prostate volume and uroflow in norwegian community men. *European Urology*, *39*, 36-41.

Pinnock, C. B, & Marshall, V. S. (1997). Troublesome lower urinary tract symptoms in the community: A prevalence study, *Medical Journal of Australia*, *167*, 72-75.

Plante, M., Corcos, J., Gregoire, I., Belanger, M. F., Brock, G., & Rossingol, M. (1996). The international prostate symptom score: physician versus self-administration in the quantification of symptomatology. *Urology*, *47*, 326-328.

Porru, D., Campus, G., Caria, A., Madeddu, G., Cucchi, A., Rovereto, B., et al. (2001). Impact of early pelvic floor rehabilitation after transurethral resection of the prostate. *Neurourology & Urodynamics*, *20*, 53-59.

Quek, K. F, Loh, C. S., Low, W. Y., Razack, A. H., & Chua, C. B. (2002). The effects of treating lower urinary tract symptoms on health-related quality of life: a short term outcome. *Singapore Medical Journal*, *43*, 391-398.

Quek, K. F, Low, W. Y., Razack, A. H., & Loh, C. S. (2001). Reliability and validity of the International Prostate Symptom Score in a Malaysian population. *British Journal of Urology International*, *88,* 21-25.

Rhodes, T., Girman, C. J., Jacobsen, S. J., Guess, H. A., Hanson, K. A., Oesterling, J. E. et al. (1995). Does the mode of questionnaire administration affect the reporting of urinary symptoms? *Urology, 46*, 341-345.

Roberts, R. O., Jacobsen, S. J., Jacobson, D. J., Reilly, W. T., Talley, N. J., & Lieber, M. M. (1998). Natural history of prostatism: high American Urological Association Symptom scores among community-dwelling men and women with urinary incontinence. *Urology, 51*, 213-219.

Roberts, R. O., Jacobsen, S. J., Rhodes, T., Girman, C. J., Guess, H. A., & Lieber, M. M. (1997). Natural history of prostatism: impaired health states in men with lower urinary tract symptoms. *Journal of Urology*, *157*, 1711-1717.

Roberts, R. O., Rhodes, T., Panser, L. A., Girman, C. J., Chute, C. G., Oesterling, J. E., et al. (1994). Natural history of prostatism: worry and embarrassment from urinary symptoms and health care-seeking behaviour. *Urology*, *43*, 621-628.

Rodrigues Netto, N., de Lima, M. L., de Andrade, E. F., Apuzzo, F., da Silva, M. B., Davidzon, I. M., et al. (1997). Latin American study on patient acceptance of the International Prostate Symptom Score (IPSS) in the evaluation of symptomatic benign prostatic hyperplasia. *Urology*, *49*, 46-49.

Roehrborn, C. G. (1996). The American Urological Association Symptom Index – Concerns and Confirmation. *Journal of Urology*, *155*, 1975-1976.

Rosen, R., Altwein, J., Boyle, P., Kirby, R. S., Lukacs, B., Meuleman, E., et al. (2003). Lower urinary tract symptoms and male sexual dysfunction: the multinational survey of the aging male (MSAM-7). *European Urology*, *44*, 637-649.

Sagnier, P. P., Girman, C. J., Garraway, M., Kumamoto, Y., Lieber, M. M., Richard, F., et al. (1996). International comparison of the community prevalence of symptoms of prostatism in four countries. *European Urology*, *29*, 15-20.

Sagnier, P. P., MacFarlane, G., Richard, F., Botto, H., Teillac, P., & Boyle, P. (1994). Results of an epidemiological survey using a modified American Urological Association symptom index for benign prostatic hyperplasia in France. *Journal of Urology*, *151*, 1266-1270.

Sagnier, P. P., MacFarlane, G., Teillac, P., Botto, H., Richard, F., & Boyle, P. (1995). Impact of symptoms of prostatism on level of bother and quality of life of men in the French community. *Journal of Urology*, *153,* 669-673.

Sarma, A. V., Jacobsen, S. J., Girman, C. J., Jacobson, D. J., Roberts, R. O., Rhodes, T. et al. (2002). Concomitant longitudinal changes in frequency of and bother from lower urinary tract symptoms in community dwelling men. *Journal of Urology*, *168*, 1446-1452.

Scarpero, H. M., Fiske, J., Xue, X., & Nitti, V. W. (2003). American Urological Association Symptom Index for lower urinary tract symptoms in women: correlation with degree of bother and impact on quality of life. *Urology, 61*, 1118-1122.

Schacterle, R. S., Sullivan, M. P., & Yalla, S. V. (1996). Combinations of maximum urinary flow rate and American Urological Association symptom index that are more specific for identifying obstructive and non-obstructive prostatism. *Neurourology & Urodynamics*, *15*, 459-470.

Schulz, M. W., Chen, J., Woo, H. H., Keech, M., Watson, M. E., Davey, P. J. (2002). A comparison of techniques for elicting patient preferences in patients with benign prostatic hyperplasia. *Journal of Urology*, *168*, 155-159.

Schwartz, B. F., Soderdahl, D. W., & Thrasher, J. B. (1998). Home flow rates in evaluation of lower urinary tract symptoms in men. *Techniques in Urology, 4*, 15-17.

Sciarra, A., D'Eramo, G., Casale, P., Loreto, A., Buscarini, M., Di Nicola, S., et al. (1998). Relationship among symptom score, prostate volume, and urinary flow rate in 543 patients with and without benign prostatic hyperplasia. *Prostate*, *34*, 121-128.

Stav, K., Leibovici, D., Goren, E., Livshitz, A., Siegel, Y. I., Lindner, A. et al. (2004). Adverse effects of cystoscopy and its impact on patients' quality of life and sexual performance. *Israel Medical Association Journal, 6,* 474-478.

Su, L., Guess, H. A., Girman, C. J., Jacobsen, S. J., Oesterling, J. E., Panser, L. A., et al. (1996). Adverse effects of medications on urinary symptoms and flow rate: a community-based study. *Journal of Clinical Epidemiology, 4*, 483-487.

Sugaya, K., Kadekawa, K., Ikehara, A., Nakayama, T., Gakiya. M., Nashiro, F., et al. (2003). Influence of hypertension on lower urinary tract symptoms in benign prostatic hyperplasia. International *Journal of Urology, 10,* 569-574.

Tan, H. Y, Choo, W. C., Archibald, C., & Esuvaranathan, K. (1997). A community based study of prostatic symptoms in Singapore. *Journal of Urology*, *157*, 890-893.

Teillac, P. (1998). Relief of BPO or improvement in quality of life? *European Urology, 8*, 34 Suppl 2:3-9.

Temml, C., Brossner, C., Schatzl, G., Ponholzer, A., Knoepp, L., & Madersbacher, S. (2003). The natural history of lower urinary tract symptoms over five years. *European Urology*, *43,* 374-380.

Terai, A., Matsui, Y., Ichioka, K., Ohara, H., Terada, N., & Yoshimura, K. (2004). Comparative analysis of lower urinary tract symptoms and bother in both sexes. *Urology*, *63*, 487-491.

Trueman, P., Hood, S. C., Nayak, U. S., & Mrazek. M. F. (1999). Prevalence of lower urinary tract symptoms and self-reported diagnosed 'benign prostatic hyperplasia', and their effect on quality of life in a community-based survey of men in the UK. *British Journal of Urology International, 83*, 410-415.

Tuncay, A. F, Aygun, C., Bilir, N., & Ozen, H. (2003). Prevalence of lower urinary tract symptoms in a community-based survey of men in Turkey. *International Journal of Urology*, *10*, 364-370.

Ukimura, O., Kojima, M., Inui, E., Ochiai, A., Hata, Y., Watanabe, M., et al. (1996). A statistical study of the American Unological Symptom index for benign prostatic hyperplasia in participants of mass screening program for prostatic diseases using transrectal sonography. *Journal of Urology*, *156*, 1673-1678.

van Schailk, P., Ahmed, T., Suvakovic, N., & Hindmarsh, J. R. (1999). Effect of an educational multimedia prostate program on the International Prostate Symptom Score. *European Urology*, *36*, 36-39.

van Venrooij, G. E., Boon, T. A., & de Gier, R. P. (1995). International prostate symptom score and quality of life assessment versus urodynamic parameters in men with benign prostatic hyperplasia symptoms. *Journal of Urology*, *153*, 1516-1519.

Vela-Navarrete, R., Alfaro, V., Badiella, L. L., & Fernandez-Hernanado, N. (2000). Age-stratified analysis of I-PSS and QoL values in Spanish patients with symptoms potentially related to BPH. *European Urology*, *38*, 199-207.

Vesely, S., Knutson, T., Damber, J. E., Dicuio, M., & Dahlstrand, C. (2003). Relationship between age, prostate volume, prostate-specific antigen, symptom score and uroflowmetry in men with lower urinary tract symptoms. *Scandinavian Journal of Urology & Nephrology*, *37*, 322-328.

Wadie, B. S., Badawi, A. M., & Ghoneim, M. A. (2001). The relationship of the International Prostate Symptom Score and objective parameters for diagnosing bladder outlet obstruction. Part II: the potential usefulness of artifical neural networks. *Journal of Urology*, *165*, 35-37.

Wadie, B. S., Ibrahim, E. H., de la Rosette, J. J., Gomha, M. A., & Ghoneim, M. A. (2001). The relationship of the International Prostate Symptom Score and objective parameters for diagnosing bladder outlet obstruction. Part 1: when statistics fail. *Journal of Urology*, *165*, 32-34.

Walker, R. M., Romano, G., Davies, A. H., Theodorou, N. A., Springhall, R. G., & Carter, S. S. (2001). Pressure flow study data in a group of asymptomatic male control patients 45 years old or older. *Journal of Urology*, *165*, 683-687.

Wang, C. C., Yang, S. S, Chen, Y. T., & Hsieh, J. H. (2003). Video-urodynamics identifies the causes of young men with lower urinary tract symptoms and low uroflow. *European Urology*, *43*, 386-390.

Wei, J. T, Dunn, R. L, Litwin, M. S, Sandler, H. M., & Sanda, M. G. (2000). Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. *Urology, 56*, 899-905.

Weiss, J. P., Blaivas, J. G., Stember, D. S., & Chaikin, D. C. (1999). Evaluation of the etiology of nocturia in men: the nocturia and nocturnal bladder capacity indices. *Neurourology & Urodynamics, 18*, 559-565.

Weiss, J. P., Blaivas, J. G., Tash, Anger, J. A., Di Blasio, C. J, Panagopoulos, G., & Gerboc, J. (2004). Development and validation of a new treatment outcome score for men with LUTS. *Neurourology & Urodynamics*, *23*, 88-93.

Welch, G., Kawachi, I., Barry, M. J., Giovannucci, E., Colditz, G. A. & Willet, W. C. (1998). Distinction between symptoms of voiding and filling in benign prostatic hyperplasia: findings from the Health Professions Follow-up Study. *Urology*, *51*, 422-427.

Welch, G., Weinger, K., & Barry, M. J. (2002). Quality-of-life impact of lower urinary tract symptoms severity: results from the Health Professionals Follow-up Study. *Urology, 59*, 245-250.

Wolters, R., Wensing, M., van Weel, C., van der Wilt, G. J., & Grol, R. P. (2002). Lower urinary tract symptoms: social influence is more important than symtoms in seeking medical care. *British Journal of Urology International*, *90*, 655-661.

Yalla, S. V., Sullivan, M. P, Lecamwasam, H. S., DuBeau, C. E., Vickers, M. A., & Cravalho, E. G. (1995). Correlation of American Urological Association symptom index with obstructive and non-obstructive prostatism. *Journal of Urology*, *153*, 674-679.

Zisman, A., Leibovici, D., Kleinmann, J., Cooper, A., Siegel. Y., & Lindner, A. (2001). The impact of prostate biopsy on patient well-being: a prospective study of voiding impairment. *Journal of Urology, 166*, 2242-2246.

Reporter: Nicholas Marosszeky, Research Psychologist

Date of report: 14 December 2004

Urinary Incontinence Health-Related Quality of Life Measures

Urinary Incontinence-Specific Quality of Life Instrument

Title:	Urinary Incontinence-Specific Quality of Life Instrument.
Abbreviations:	I-QOL.
Author(s) Name:	Donald L. Patrick, Ph. D.
Author(s) Address:	Box 357660 Department of Health Services School of Public Health and Community Medicine University of Washington H-689 Health Sciences Seattle, WA 98195-7660, USA
Supplied by:	Medical Outcomes Trust PMB #503 198 Tremont St Boston, MA 02116-4705 USA Phone: US 617-426-4046 Fax: US 617 587 4232 Email: info@outcomes-trust.org
Cost:	Free. However, one must seek permission to use the Urinary Incontinence-Specific Quality of Life Instrument (I-QOL). It costs $US200 to purchase the I-QOL Instrument Packet from the Medical Outcomes Trust.
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Those without these qualifications require basic training in survey administration and the characteristics of the I-QOL.
Purpose:	The Urinary Incontinence-Specific Quality of Life Instrument (I-QOL) is a condition specific quality of life measurement for urinary incontinence and urinary problems.
Administration Time:	5 minutes.
Instrument Type:	Self-report questionnaire.
Structure:	The I-QOL consists of 22 items with a 5-point Likert rating scale. The quality of life dimensions include: 8 items covering avoidance and limiting behaviour; 9 items assessing the psychosocial impacts of incontinence; and 5 items looking at social embarrassment.
Scoring:	Item responses are summed and averaged to create a total score which is transformed to a 0 – 100 scale; with higher scores indicating better urinary-incontinence specific quality of life.
	The three sub-scales (avoidance and limiting behaviour, psychosocial impacts, and social embarrassment) can also be calculated.
Developed for:	The I-QOL was developed by a team of researchers from the University of Washington, Seattle under the direction of Dr. Donald L. Patrick with sponsorship from Eli Lilly and Company.
	The questionnaire was developed from structured interviews with individuals with urinary incontinence (Wagner et al. 1996).
Normative Data:	No large scale, normative (population) studies using the I-QOL have been found in the literature to date.
Clinical Data:	A few clinical studies are listed below:
	Noninvasive tibial nerve stimulation treatment for overactive bladder: Svihra et al. (2002).

Perineal Magnetic Stimulation treatment for urinary incontinence: Almeida et al. (2004).

Posterior tibial nerve stimulation treatment for urge incontinence: Vandoninck et al. (2003).

Treatment with Duloxetine (serotonin-norepinephrine reuptake inhibitor) for stress and mixed urinary incontinence symptoms: Bump et al. (2003); van Kerrebroeck et al. (2004); Millard et al. (2004).

Applications: Published clinical practice guidelines have recommended the I-QOL as a tool for monitioring the quality of life of women with stress and urge urinary incontinence; as recorded by the National Guideline Clearinghouse in the USA (AHRQ).

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Wagner et al. (1996). Patrick et al. (1999). Patrick et al. (1999).	Good	Cronbach's alpha is in the range of 0.87 – 0.93 (Patrick et al. 1999).
Test – retest	Yes	Wagner et al. (1996). Patrick et al. (1999). Patrick et al. (1999).	Good	
Inter – rater	NA	NA	NA	The I-QOL is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Wagner et al. (1996). Patrick et al. (1999).	Adequate	
Correlation with other measures	Yes	Patrick et al. (1999). Finkelstein et al. (2002). Fialkow et al. (2003).	Weak - Adequate	Further study is required with other disease specific quality of life instruments.

Construct	Yes	Wagner et al. (1996). Patrick et al. (1999). Patrick et al. (1999).	Adequate	
Criterion	No			Studies in this area are required.

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	**Comment**
Sensitivity to change	Yes	Patrick et al. (1999).	Good	

Cultural Applicability and Cultural Adaptations:
The I-QOL has been translated into many languages, with validation data for the French, Spanish, Swedish and German languages.

Gender Appropriateness:
Most of the developmental work with The Urinary Incontinence-Specific Quality of Life Instrument has been conducted with women.

Age Appropriateness:
18 – 76+ years.

Summary:
The Urinary Incontinence-Specific Quality of Life Instrument (I-QOL) is a psychometrically promising condition specific, quality of life measure for women with urinary incontinence; which requires some further study and application.

References

Almeida, F. G., Bruschinin, H., & Srougi, M. (2004). Urodyamic and clinical evaluation of 91 female patients with urinary incontinence treated with perineal magnetic stimulation: 1-year followup. Journal of *Urology, 171,* 1571-1574.

Bump, R. C., Norton, P. A., Zinner, N. R., & Yalcin, I. (2003). Mixed urinary incontinence symptoms: urodynamics findings, incontinence severity, and treatment response. *Obstetrics & Gynecology, 102*, 76-83.

Fialkow, M. F., Melville, J. L, Lentz, G. M., Miller, E. A., Miller, J., & Fenner, D. E. (2003). The functional and psychosocial impact of fecal incontinence on women with urinary incontinence. *American Journal of Obstetrics & Gynecology, 189*, 127-129.

Finkelstein, M. M., Skelly, J., Kaczorowski, J., & Swanson, G. (2002). Incontinence Quality of Life Instrument in a survey of primary care physicians. *Journal of Family Practice, 51*, 952.

Millard, R. J., Moore, K., Rencken, R., Yalcin, I., & Bump, R. C. (2004). Duloxetine vs placebo in the treatment of stress urinary incontinence: a four continents randomised clinical trial. *British Journal of Urology International, 93*, 311-318.

Patrick, D. L., Martin, M. L., Bushnell, D. M., Marquis, P., Andrejasich, C. M., & Buesching, D. P. (1999). Cultural adaptation of a quality-of-life measure for urinary incontinence. *European Urology, 36*, 427-435.

Patrick, D. L., Martin, M. L., Bushnell, D. M., Valcin, I., Wagner, T. H., & Buesching, D. P. (1999). Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). *Urology, 53*, 71-76.

Svihra, J., Kurca, E., Luptak, J, & Kliment, J. (2002). Neuromodulative treatment of overactive bladder – noninvasive tibial nerve stimulation. *Bratisl Lek Listy, 103*, 480-483.

University of Texas at Austin, School of Nursing, Family Nurse Practitioner Program. (2002). *Recommendations for the Management of Stress and Urge Urinary Incontinence in Women*. Austin (TX): Unversity of Texas at Austin, School of Nursing, 13.

van Kerrebroeck, P., Abrams, P., Lange, R., Slack, M., Wyndaele, J. J, Yalcin, I., et al. (2004). Duloxetine versus placebo in the treatment of European and Canadian women with stress urinary incontinence. *British Journal of Gynaecology: an International Journal of Obstetrics & Gynaecology, 111*, 249-257.

Vandoninck, V., Van Balken, M. R., Finazzi Agro, E., Petta, F., Caltagirone, C., Heesakkers, J. P. et al. (2003). Posterior tibial nerve stimulation in the treatment of urge incontinence. Neurourology & *Urodynamics, 22,* 17-23.

Wagner, T. H., Patrick, D. L., Bavendam, T. G., Martin, M. L., & Buesching, D. P. (1996). Quality of life of persons with urinary incontinence: development of a new measure. *Urology, 47*, 67-72.

Note: A bibliography is available for registered users on the QOLID database at www.qolid.org and a bibliographical search for recent abstracts can be undertaken at phi.uhce.ox.ac.uk/phidb.html

Reporter: Nicholas Marosszeky, Research Psychologist

Date of report: 14 December 2004

Multi-Attribute Utility Instrument Reviews

AQoL

INSTRUMENT REVIEW SHEET	
Title:	Assessment of Quality of Life
Abbreviations:	AQoL
Author(s) name:	Hawthorne G. & Richardson J.
Author(s) address:	Hawthorne G, Australian Centre for Posttraumatic Mental Health, Department of Psychiatry, The University of Melbourne, PO Box 5444 Heidelberg West, Victoria, Australia, 3081. Richardson J, Centre for Health Economics, Building 75, Monash University, Clayton, Victoria, Australia 3800.
Supplied by:	A copy of the AQoL can be found in Hawthorne, Richardson et al. 1999.
Cost:	Free. Registration is required. To register contact A/Professor Graeme Hawthorne (graemeeh@unimelb.edu.au) or Ms Sharon Lakua (sharon.lakua@buseco.monash.edu.au)
Administration time:	5-10 minutes.
Training requirements:	None.
Purpose:	The AQoL is a generic health-related quality of life instrument. Although designed for use in economic evaluation, specifically cost-utility studies, it can also be used as a health-status profile instrument.
Administration:	Self-administered. Can also be administered by interviewer and over the telephone. Hawthorne (2003) showed that there was no significant effect by administration mode (mail versus telephone).
Languages available:	English, Danish.
Instrument type:	Multi-attribute utility instrument.
Structure:	The AQoL comprises 5 dimensions, each with 3 items. The dimensions and items are: Illness (medicine use, medical aids, and medical treatment), Independent Living (self-care, daily activities, and mobility), Social Relationships (intimacy, friends, and family role), Physical Senses (vision, hearing, communication) and Psychological Wellbeing (sleep, anxiety, pain). Each item has 4 responses levels, from a normal health state to a very poor health state. It should be noted that the different response levels are simply different levels of health; they do not possess interval or cardinal properties.

Scoring:	Only the last four dimensions of the AQoL are used in obtaining the utility scores. Endorsed responses are substituted with values obtained from a stratified sample of the general Australian community using time trade-off (TTO; n = 225). (Richardson and Hawthorne 1998; Hawthorne, Richardson et al. 1999).
	Within each dimension, the three items are combined using a multiplicative model. Dimension scores are then combined into a single index, again using a multiplicative model, and the resulting score is then transformed onto a life-death scale.
	The upper boundary, 1.00, represents the best HRQoL state, 0.00 represents HRQoL states that are death equivalent, and the lower boundary, -0.04, represents HRQoL states worse than death. HRQoL states worse than death are necessary to explain people who have either commited suicide or euthanasia.
Developed for:	The AQoL was designed for the economic evaluation of health interventions, specifically cost-utility analysis. The descriptive system was developed by the researchers based on the World Health Organisation's classification of disease, impairment, disability and handicap. A review of HRQoL instruments published since 1970 suggested the dimensions, and actual items were written in focus groups with clinicians. The pilot involved inpatients and a random sample of the general community. The utility weights used in scoring the AQoL were obtained from a representative sample of the Australian population, using the time trade-off (Richardson and Hawthorne 1998; Hawthorne, Richardson et al. 1999).
Normative data:	Normative data are reported in Hawthorne et al. (Hawthorne, Richardson et al. 2001; Hawthorne, Richardson et al. 2001). The general population norm for the AQoL utility, based on the 1998 South Australian Health Omnibus Survey (n = 3000 adults) is 0.83 (sd = 0.20). A smaller stratified sample of the Victorian population (n = 334) provided estimates from 0.71–0.84, depending upon age group (Hawthorne, Richardson et al. 2001).
Clinical data:	There are a limited number of published papers on the AQoL in clinical settings, mainly because it is a relatively new measure.
	Studies include Alzheimer's disease (Wlodarczyk, Brodaty et al. In press), cochlear implants (Hogan, Hawthorne et al. 2001; Hawthorne and Hogan 2002; Hawthorne, Hogan et al. 2004), pharmacist medication (Harris, Gospodarevskaya et al. 2001), psychosis (Herrman, Hawthorne et al. 2002), influenza (Osborne, Hawthorne et al. 2000), and stroke (Sturm, Osborne et al. 2002).
Public health data:	There are a limited number of published papers on the AQoL in public health settings, mainly because it is a relatively new measure.
	Studies include: aged care (Osborne, Hawthorne et al. 2003), depression (Hawthorne, Cheok et al. 2003), and suicidal ideation (Goldney, Fisher et al. 2001).

Applications:	There are three applications for the AQoL. As a generic instrument, it is applicable to all public health and clinical conditions and interventions.
	1. It can be used as a simple additive HRQoL measure, providing profile scores on independent living, social relationships, physical senses, and psychological wellbeing. Used like this, the AQoL can be used in program evaluation to assess the benefits associated with health interventions.
	2. When utilities are computed, the AQoL can used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions.
	3. When utilities are computed, the AQoL can used in economic evaluation, specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation.
Reliability:	Reliability has been assessed using internal consistency and test-retest.
	Based on community and hospital samples, the range of internal consistency estimates is Cronbach _ = 0.73 – 0.84 (Richardson and Hawthorne 1998; Osborne, Hawthorne et al. 2000; Hogan, Hawthorne et al. 2001; Osborne, Hawthorne et al. 2003; Hawthorne, Hogan et al. 2004).
	Based on a random community sample, two-week test-retest reliability was ICC = 0.83 (Hawthorne, 2003).
	Herrman et al. reported a Pearson r = 0.55 of proxy (case manager) and self-completion by those suffering long–term psychosis; and suggested that self-completion was to be preferred (Herrman, Hawthorne et al. 2002).
Studies* reported:	(Richardson and Hawthorne 1998; Hawthorne, Richardson et al. 1999; Richardson and Hawthorne 1999; Osborne, Hawthorne et al. 2000; Goldney, Fisher et al. 2001; Harris, Gospodarevskaya et al. 2001; Hawthorne and Richardson 2001; Hawthorne, Richardson et al. 2001; Hawthorne, Richardson et al. 2001; Hogan, Hawthorne et al. 2001; Hawthorne and Hogan 2002; Herrman, Hawthorne et al. 2002; Sturm, Osborne et al. 2002; Hawthorne 2003; Hawthorne, Cheok et al. 2003; Osborne, Hawthorne et al. 2003; Hawthorne, Hogan et al. 2004; Holland, Smith et al. In press; Segal, Dunt et al. In press; Wlodarczyk, Brodaty et al. In press)
* = academic papers only.	

References

Goldney, R. D., Fisher, L. J., Wilson, D. H. & Cheok, F. (2001). Suicidal ideation and health-related quality of life in the community. *Medical Journal of Australia*, *175* (10), 546-549.

Harris, A., Gospodarevskaya, E., Callaghan, J. & Story, I. (2001). The cost effectiveness of a pharmacist reviewing medication among the elderly in the community. *Australian Journal on Ageing*, *20* (4), 179-2001.

Hawthorne, G. & Hogan, A. (2002). Measuring disability-specific patient benefit in cochlear implant programs: developing a short form of the Glasgow Health Status Inventory, the Hearing Participation Scale. *International Journal of Audiology*, *41* (8), 535-44.

Hawthorne, G. & Richardson, N. J. (2001). Measuring the value of program outcomes: a review of utility measures. *Expert Review of Pharmacoeconomics Outcomes Research*, *1* (2), 215-228.

Hawthorne, G. (2003). The effect of different methods of collecting data: mail, telephone and filter data collection issues in utility measurement. *Quality of Life Research*, *12,* 1081-1088.

Hawthorne, G., Cheok, F., Goldney, R. & Fisher, L. (2003). The excess cost of depression in South Australia: a comparative study of two methods of calculating burden. *Australian and New Zealand Journal of Psychiatry*, *37* (3), 362-373.

Hawthorne, G., Hogan, A., Giles, E., Stewart, M., Kethel, L., White, K., Plaith, B., Pedley, K., Rushbrooke, E. & Taylor, A. (2004). Evaluating the health-related quality of life effects of cochlear implants: a prospective study of an adult cochlear implant program. *International Journal of Audiology*. *43*, 183-192.

Hawthorne, G., Richardson, J., & Day, N. (2001). A comparison of five multi-attribute utility instruments. *Australian Studies in Health Service Administration*, *89*, 151-179.

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQol) with four other generic utility instruments. *Annals of Medicine*, *33,* 358–370.

Hawthorne, G., Richardson, J., & Osborne, R. (1999). The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health related quality of life. *Quality of Life Research*, *8*, 209-224.

Herrman, H., Hawthorne, G. & Thomas, R. (2002). Quality of life assessment in people living with psychosis. *Social Psychiatry and Psychiatric Epidemiology*, *37* (11), 510-518.

Hogan, A., Hawthorne, G., Kethel, L., Giles, E., White, K., Stewart, M., Plath, B., & Code, C. (2001). Health-related quality-of-life outcomes from adult cochlear implantation: a cross-sectional study. *Cochlear Implants International*, *2* (2), 115-128.

Holland, R., Smith, R., Harvey, I., Swift, L. & Lenaghan, E. (In press). Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AqoL. *Health Economics*.

Osborne, R. H., Hawthorne, G., Lew, E. A. & Gray, L. C. (2003). Quality of Life assessments in the community-dwelling elderly: validation of the Assessment of Quality of Life (AQoL) instrument and comparison with the SF-36. *Journal of Clinical Epidemiology*, *56* (2), 138-147.

Osborne, R., Hawthorne, G., Papanicolaou, M. & Wegmuller, Y. (2000). Measurement of rapid changes in health outcomes in people with influenza symptoms. *Journal of Outcomes Research*, *4*, 15-30.

Richardson, J., & Hawthorne, G. (1998). The Australian quality of life (AQoL) instrument: psychometric properties of the descriptive system and initial validation. *Australian Studies in Health Service Administration*, *85*, 315-342.

Richardson, J., & Hawthorne, G. (1999). Difficulty with life and death: methodological issues and results from the utility scaling of the 'Assessment of Quality of Life' (AQoL) instrument. *Australian Studies in Health Service Administration*, *86*, 121-139.

Segal, L., Dunt, D., Day, S. E., Day, N. A., Robertson, I. & Hawthawne, G. (In press). Introducing coordinated care (1): a randomised trial assessing client and cost outcomes. *Health Policy*. Accepted January 2004.

Strum, J. W., Osborne, R. H., Dewey, H. M., Donnan, G. A., Macdonell, R. A. & Thrift, A. G. (2002). Brief comprehensive quality of life assessment after stroke: the assessment of quality of life instrument in the North East Melbourne stroke incidence study (NEMESIS). *Stroke*, *33* (12), 2888-94.

Wlodarczyk, J. H., Brodaty, H. & Hawthorne, G. (In press). The relationship between quality of life, Mini-Mental State Examination, and the Instrumental Activities of Daily Living in patients with Alzheimer's Disease. *Archives of Gerontology and Geriatrics*. Accepted: November 2003.

Reporter: Graeme Hawthorne, 2004

EQ5D

INSTRUMENT REVIEW SHEET	
Title:	EQ5D index (formerly the EuroQol).
Abbreviations:	EQ5D. There is some confusion in the literature about the name of the EQ5D index. There are three types of measurement, all of which are commonly referred to as the EQ5D: • The EQ5D health status measure. This describes where the 5 items of the EQ5D are completed and presented as a health profile measure. • The EQ5D VAS which is where the respondent rates their health state on a visual analog scale, where the endpoints are 0.00 (worst imaginable health) and 1.00 (best imaginable health). • The EQ5D index, which is where the responses to the EQ5D are weighted and combined to form a utility measure for use in economic evaluation. The EQ5D index is reviewed here. This review sheet does not review the EQ5D VAS. It only reviews the EQ5D health status measure where this is directly related to the EQ5D index. For a discussion of the differences between the EQ5D measurement types, the reader is referred to Krabbe & Weijnen, 2003.
Author(s) name:	The original EuroQol was developed by the EuroQol Group, although much of the work can be attributed to Professor Alan Williams, Centre for Health Economics, University of York. Membership of the EuroQol Group can be found in The EuroQol Group (EuroQol Group, 1990).
Author(s) address:	Enquiries should be directed to the EQ5D website http://www.euroqol.org/ Dr. Frank de Charro EuroQol Business Manager P.O. Box 4443 3006 AK Rotterdam The Netherlands Tel: +31 10 408 1545 Fax: +31 10 452 5303 E-mail: fdecharro@compuserve.com
Supplied by:	A copy of the EQ5D can be found in Rabin & De Charro, 2001.
Cost:	The EQ5D is available to public health researchers free of charge and is in the public domain, although the EQ5D Group requests that researchers register their use of the EQ5D. Where use is commercial, costs apply. Enquiries should be directed to the EQ5D website (http://www.euroqol.org/) where additional assistance can be found.
Administration time:	2-5 minutes.
Training requirements:	None.
Purpose:	The EQ5D was designed to provide a standardized, generic health-related quality of life measure for both describing and valuing health status. It was designed to be used alongside other measures rather than replace them, specifically in cross-cultural postal surveys.
Administration:	Self-administered. Can also be administered by interviewer and over the telephone.

Languages available:	Afrikaans, Armenian, Basque, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Filipino, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Malay, Norwegian, Polish, Portuguese, Romanian, Russian, Sesotho, Shona, Slovakian, Slovenian, Spanish, Swedish, Thai, Tongan, Turkish, Xhosa, Zulu.
Instrument type:	Multi-attribute utility instrument.
Structure:	The EQ5D comprises 5 dimensions, each with 1 item. The dimensions are: mobility, self-care, usual activities, pain/discomfort and anxiety/depression.
	Each item has 3 responses levels, from a normal health state to a very poor health state. It should be noted that the different response levels are are simply different levels of health; they do not possess interval or cardinal properties.
Scoring:	The EQ5D index is computed using an econometric regression model; essentially it is an additive model.
	The British utility weights are from a population random sample ($n = 3395$ respondents, response rate 56%) based on the TTO for 42 marker health states using a 10 year timeframe (Dolan, 1997). Other utility values were regression modelled (MVHGroup, 1995; Dolan, Gudex, et al. 1996; Dolan, 1997).
	Regarding the generalizability of the utility weights, a Swedish study showed similar values when compared with both British and Dutch values (Brooks, Jendteg, et al. 1991). A Spanish study (Badia, Roset, et al. 2001) however, showed there were significant differences between Spanish/British values on 35% of the EQ5D health states valued. It was also reported that there were significant differences in values between those who were ill and the healthy (Badia, Herdman, et al. 1998). Similar rankings of values were found for Zimbabweans when compared with the British weights (Jelsma, Hansen, et al. 2003).
	Johnson et al. used the EQ5D VAS to elicit US weights, using a transformation to convert the VAS into utilities; the $R^2 = 0.42$ (Johnson, Coons, et al. 1998). In a similar study, Polsky et al. compared the values of the general public with those of patients, reporting $R^2 = 0.57$ (Polsky, Willke, et al. 2001).
	The upper boundary is 1.00, and the lower boundary is –0.59: it permits health state values worse than death.
Developed for:	The EQ5D was developed to enable cross-cultural comparisons of health state valuations through providing a standardized, generic health-related quality of life measure for both describing and valuing health status. It was designed to be used alongside other measures rather than replace them.
Normative data:	British population norms were published by Kind et al. (Devlin & Hansen et al. 2003), broken down by gender and age group. The overall population norm was 0.86 ($n = 3392$; $sd = 0.23$). For males it was 0.86 ($n = 1467$; $sd = 0.24$) and for females it was 0.85 ($n = 1925$; $sd = 0.22$). The range, depending on population age group, was between 0.73 to 0.86.
	Preliminary Australian population values have been published ($n = 343$), giving a range of between 0.79–0.92 depending upon age group (Hawthorne, Richardson, et al. 2001).
	British normative data for the EQ5D items have been published by Kind, et al. 1998.

Clinical data:	The studies included here are representative of clinical studies using the EQ5D. Because of the number of published studies, no claim is made for comprehensiveness.
	AIDS/HIV (Delate & Coons, 2001; Miners, Sabin, et al. 2001; Wu, Jacobson, et al. 2002); alcohol and drug use (Foster, Peters, et al. 2002); allergies (Wasserfallen, Gold, et al. 1999); Alzheimer's disease/cognitive function/ dementia (Degl'Innocenti, Elmfeldt, et al. 2002; Thorgrimsen, Selwood, et al. 2003); anxiety and depression (Allsup & Gosney, 2002); arthritis/ osteoarthritis/rheumatoid arthritis/rheumatic disorders/rheumatology (Hurst, Jobanputra et al. 1994; Hurst, Kind, et al. 1997; Wolfe & Hawley, 1997; Hurst & McRorie, 1998; Brazier, Harper, et al. 1999; Kobelt, Eberhardt et al. 1999; Pipitone & Scott, 2001; Sokoll & Helliwell, 2001; Suarez-Almazor & Conner-Spady, 2001; Conner-Spady & Suarez-Almazor, 2003; Luo, Chew, et al. 2003); cancer (Norum, Angelsen, et al. 1996; Jenkinson, Gray, et al. 1997; MacDonagh, Cliff, et al. 1997; Anderson & Palmer, 1998; McMillan, Wigmore, et al. 1999; Conner-Spady, Cumming, et al. 2001; Schulz, Chen, et al. 2002; Ananth, Jones, et al. 2003; van den Hout, van der Linden, et al. 2003; Brennan, Bansback, et al. 2004); cardiovascular disease (Bosch, van der Graaf, et al. 1999; Bosch & Hunink, 2000; Beck, Joseph, et al. 2001; Campbell, Tait, et al. 2001; Asadi-Lari, Packham, et al. 2003); de Vries, 1998; Durkin, 2001; Goodacre, 2001; Oostenbrink, 2001; Serruys, 1999; Sollano, 1998; Tangelder, 1999; cosmetic surgery (Kerrigan, Collins, et al. 2000; Kerrigan, Collins, et al. 2001); Collins, Kerrigan, et al. 2002); deafness (Joore, Van Der Stel, et al. 2003); cystic fibrosis (Johnson, Connolly, et al. 2000); dermatitis (Klassen, Newton, et al. 2000; Bergstrom, Arambula, et al. 2003); dialysis (de Wit, Merkus, et al. 2002; Manns, Johnson, et al. 2002; Sennfalt, Magnusson, et al. 2002); diabetes (Ragnarson, Tennvall, & Apelqvist, 2000; Clarke, Gray, et al. 2002; Redekop, Koopmanschap, et al. 2002, Bech, Moses, et al. 2003); digestive system (Richards, Hughes, et al. 2001; Akehurst, Brazier, et al. 2002; Casellas, Lopez Vivancos, et al. 2003); neck fractures (Tidermark, Zethraeus, et al. 2002; Tidermark, Zethraeus, et al. 2002; Tidermark, Bergstrom, et al. 2003); gynaecology (Davies & Doyle, 2002); hepatitis (Chong, Gulamhussein, et al. 2003); hernia (Poley, Stolk, et al. 2002); hypertension (Roca-Cusachs, Dalfo, et al. 2001); immune system (McCrone, Chisholm, et al. 2003); incontinence (Kobelt, 1997; Fuertes, Garcia Matres, et al. 2000; Joshi, Adams, et al. 2001; O'Brien, Goeree, et al. 2001; Joshi, Stainthorpe, et al. 2003); intensive care (Sznajder, Aegerter, et al. 2001); migraine (Essink-Bot, van Royen et al. 1995; Essink-Bot, Krabbe, et al. 1997); mobility (Hollingworth, Mackenzie, et al. 1995; Chetter, Dolan, et al. 1997; Chetter, Spark, et al. 1997; Soderman, 2000; Ekert, Brewin, et al. 2001; Tidermark, Blomfeldt, et al. 2003; Tidermark, Ponzer, et al. 2003); multiple sclerosis (Rothwell, McDowell, et al. 1997; Forbes, Lees, et al. 1999); muscular dystrophy (van Dieten, Perez, et al. 2003); neonatal surgery (Poley, Stolk, et al. 2001); neurology (Jarman, Hurwitz, et al. 2002; Siderowf, Ravina, et al. 2002; Brilstra, Rinkel, et al. 2004); orthopaedics (James, St Leger, et al. 1996; Glick, Polsky, et al. 1999; Brown, Kennedy, et al. 2001; Green, Kiebert, et al. 2003); pain (Suarez-Almazor, Kendall, et al. 2000; Garratt, Klaber Moffett, et al. 2001; Miller, Kendrick, et al. 2002; Forouzanfar, Kemler, et al. 2004); respiratory diseases (Andersson, Strom, et al. 1998; Allin, James, et al. 2001); sleep apnea (Jenkinson, Stradling, et al. 1997; Jenkinson, Stradling, et al. 1998; Chakravorty, Cayton, et al. 2002); spondylitis (Boonen, van der Heijde, et al. 2003); stroke (Dorman, Slattery, et al. 1997; Dorman, Waddell, et al. 1997; Dorman, Slattery, et al. 1998; Dorman, Dennis, et al. 1999; Myers, McPherson, et al. 2003; Pickard, Johnson, et al. 2004); transplants (Bryan, Ratcliffe, et al. 1998; Anyanwu, McGuire, et al. 2001; Ratcliffe, Longworth, et al. 2002; Longworth & Bryan, 2003).

Public health data:	The studies included here are representative of clinical studies using the EQ5D. Because of the number of published studies, no claim is made for comprehensiveness.
	Exercise (Van Tubergen, Boonen, et al. 2002); gypsy travellers (Van Cleemput & Parry, 2001); inner city low-socioeconomic status (Lubetkin & Gold, 2002); older adults (Brazier, Walters, et al. 1996; Coast, Peters, et al. 1998); population samples (Devlin, Hansen, et al. 2000); smoking (Tillmann & Silcock, 1997); trauma injury (Meerding, Looman, et al. 2004).
Applications:	There are three applications for the EQ5D. As a generic instrument, it is applicable to all public health and clinical conditions and interventions.
	1. It can be used as a simple HRQoL measure, providing profile scores on all five items. Used like this, the EQ5D can be used in program evaluation to assess the benefits associated with health interventions.
	2. When utilities are computed, the EQ5D can used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions.
	3. When utilities are computed, the EQ5D can used in economic evaluation, specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation.
Reliability:	Reliability of the EQ5D items at 10-month test-retest was assessed using G-study, which showed that 82% of the variance was expained by respondents' health state; when expressed as test-retest reliabiility correlation coefficient this was reported as 0.36 at the group level and 0.90 at the individual level (van Agt, Essink-Bot, et al. 1994).
	Brazier (Brazier, Deverill et al. 1999) cites two earlier studies of the EQ5D involving older women and those with chronic obstructive pulmonary disease. Test-retest at 6-months among those who stated their health had not changed was 0.67 and 0.83. Using test-retest among stroke patients, Dorman et al. reported $Kappa$ = 0.63 to 0.80 for EQ5D items (Dorman, Slattery et al. 1998). Stavem reported Spearman correlations at 2-week test-retest of r_s = 0.73 for those with chronic obstructive pulmonary disease (Stavem 1999). The ICC was reported to be 0.70 for those with osteoarthritis (Fransen & Edmonds, 1999).

References

Akehurst, R. L., Brazier, J. E., Mathers, N., O'keefe, C., Kaltenthaler, E., Morgan, A., Platts, M. & Walters, S. J. (2002). Health-related quality of life and cost impact of irritable bowel syndrome in a UK primary care setting. *PharmacoEconomics*, *20* (7), 455-62.

Allin, D., James, I., Zachariah, J., Carr, W., Cullen, S., Middleton, A., Newson, P., Lytle, T. & Coles, S. (2001). Comparison of once- and twice-daily clarithromycin in the treatment of adults with severe acute lower respiratory tract infections. *Clinical Therapeutics*, *23* (12), 1958-68.

Allsup, S. J. & Gosney, M. A. (2002). Anxiety and depression in an older research population and their impact on clinical outcomes in a randomised controlled trial. *Postgraduate Medical Journal*, *78* (925), 674-7.

Ananth, H., Jones, L., King, M. & Tookman, A. (2003). The impact of cancer on sexual function: a controlled study. *Palliative Medicine*, *17* (2), 202-5.

Anderson, H., & Palmer, M. K. (1998). Measuring quality of life: impact of chemotherapy for advanced colorectal cancer. Experience from two recent large phase III trials. *British Journal of Cancer*, *77* Suppl 2, 9-14.

Andersson, A., Strom, K., Brodin, H., Alton, M., Boman, G., Jakobsson, P., Lindberg, A., Uddenfeldt, M., Walter, H., & Levin, L. A. (1998). Domiciliary liquid oxygen versus concentrator treatment in chronic hypoxaemia: a cost-utility analysis. *European Respiratory Journal*, *12* (6), 1284-9.

Anyanwu, A. C., Mcguire, A., Rogers, C. A., & Murday, A. J. (2001). Assessment of quality of life in lung transplantation using a simple generic tool. *Thorax*, *56* (3), 218-22.

Asadi-Lari, M., Packham, C. & Gray, D. (2003). Unmet health needs in patients with coronary heart disease: implications and potential for improvement in caring services. *Health Qual Life Outcomes*. *1* (1), 26.

Badia, X., Herdman, M., & Kind, P. (1998). The influence of ill-health experience on the valuation of health. *PharmacoEconomics*, *13* (6), 687-96.

Badia, X., Roset, M., Herdman, M., & Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making*, *21* (1), 7-16.

Bech, P., Moses, R. & Gomis, R. (2003). The effect of prandial glucose regulation with repaglinide on treatment satisfaction, wellbeing and health status in patients with pharmacotherapy naive Type 2 diabetes: a placebo-controlled, multicentre study. *Quality of Life Research*, *12* (4), 413-25.

Beck, C. A., Joseph, L., Belisle, P. & Pilote, L. (2001). Predictors of quality of life 6 months and 1 year after acute myocardial infarction. *American Heart Journal*, *142* (2), 271-9.

Bergstrom, K. G., Arambula, K. & Kimball, A. B. (2003). Medication formulation affects quality of life: a randomized single-blind study of clobetasol propionate foam 0.05% compared with a combined program of clobetasol cream 0.05% and solution 0.05% for the treatment of psoriasis, *Cutis*, *72* (5), 407-11.

Boonen, A., van der Heijded, D., Landewe, R., Guillemin, F., Spoorenberg, A., Schouten, H., Rutten-Van Molken, M., Dougados, M., Mielants, H., De Vlam, K., van der Tempel, H. & van der Linden, S. (2003). Costs of analysing? Spondylitis in three European countries: the patient's perspective. *Annals of the Rheumatic Diseases*, *62* (8), 741-7.

Bosch, J. L., & Hunink, M. G. (2000). Comparison of the Health Utilities Index Mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. *Quality of Life Research*, *9* (6), 591-601.

Bosch, J. L., van der Graaf, Y., & Hunink, M. G. (1999). Health-related quality of life after angioplasty and stent placement in patients with iliac artery occlusive disease: results of a randomized controlled clinical trial. The Dutch Iliac Stent Trial Study Group. *Circulation*, *99* (24), 3155-60.

Brazier, J. E., Harper, R., Munro, J., Walters, S. J., & Snaith, M. L. (1999). Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology*, *38* (9), 870-7.

Brazier, J. E., Walters, S. J., Nicholl, J. P. & Kohler, B. (1996). Using the SF-36 and Euroqol on an elderly population. *Quality of Life Research*. *5* (2), 195-204.

Brazier, J., Deverill, M., Green, C., Harper, R. & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *Health Technology Assessment*, *3* (9), 1-165.

Brennan, A., Bansback, N., Reynolds, A., & Conway, P. (2004). Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. *Rheumatology*, *43,* (1), 62-72.

Brilstra, E. H., Rinkel, G. J., Van Der Graaf, Y., Sluzewski, M., Groen, R. J., Lo, R. T. & Tulleken, C. A. (2004). Quality of life after treatment of unruptured intracranial aneurysms by neurosurgical clipping or by embolisation with coils. A prospective, observational study. *Cerebrovascular Diseases*, *17* (1), 44-52.

Brooks, R. G., Jendteg, S., Lindgren, B., Persson, U., & Bjork, S. (1991). EuroQol: health-related quality of life measurement. Results of the Swedish questionnaire exercise. *Health Policy*, *18* (1), 37-48.

Brown, A. P., Kennedy, A. D., Torgerson, D. J., Campbell, J., Webb, J. A., & Grant, A. M. (2001). The OMENS trial: opportunistic evaluation of musculo-skeletal physician care among orthopaedic outpatients unlikely to require surgery. *Health Bulletin*, *59* (3), 198-210.

Bryan, S., Ratcliffe, J., Neuberger, J. M., Burroughs, A. K., Gunson, B. K., & Buxton, M. J. (1998). Health-related quality of life following liver transplantation. *Quality of Life Research*, *7* (2), 115-20.

Campbell, H. E., Tait, S., Buxton M. J., Sharples, L. D., Caine, N., Schofield, P. M. & Wallwork, J. (2001). A UK trial-based cost-utility analysis of transmyocardial laser revascularization compared to continued medical therapy for treatment of refractory angina pectoris. *European Journal of Cardio-Thoracic Surgery*, *20* (2), 312-8.

Casellas, F., Lopez Vivancos, J. & Malagela, J. R. (2003). Previous experience and quiality of life in patients with inflammatory bowel disease during relapse. *Revista Espanola de Enfermedades Digestivas*, *95* (7), 476-9, 471-5.

Chakravorty, I., Cayton, R. M., & Szczepura, A. (2002). Health utilities in evaluating intervention in the sleep apnoea/hypopnoea syndrome. *European Respiratory Journal*, *20* (5), 1233-8.

Chetter, I. C., Dolan, P., Spark, J. I., Scott, D. J., & Kester, R. C. (1997). Correlating clinical indicators of lower-limb ischaemia with quality of life. *Cardiovascular Surgery*, *5* (4), 361-6.

Chetter, I. C., Spark, J. I., Dolan, P., Scott, D. J., & Kester, R. C. (1997). Quality of life analysis in patients with lower limb ischaemia: suggestions for European standardisation. *European Journal of Vascular and Endovascular Surgery*, *13* (6), 597-604.

Chong, C. A., Gulamhussein, A., Heathcote, E. J., Lilly, L., Sherman, M., Naglie, G, & Krahn, M. (2003). Health-state utilities and quality of life in hepatitis C patients. *American Journal of Gastroenterology*, *98* (3), 630-8.

Clarke, P., Gray, A., & Holman, R. (2002). Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). *Medical Decision Making*. *22* (4), 340-9.

Coast, J., Peters, T. J., Richards, S. H., & Gunnell, D. J. (1998). Use of the EuroQoL among elderly acute care patients. *Quality of Life Research*, *7* (1), 1-10.

Collins, E. D., Kerrigan, C. L., Kim, M., Lowery, J. C., Striplin, D. T., Cunningham, B., & Wilkins, E. G. (2002). The effectiveness of surgical and nonsurgical interventions in relieving the symptoms of macromastia. *Plastic and Reconstructive Surgery*, *109* (5), 1556-66.

Conner-Spady, B. & Suarez-Almazor Me. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care*, *41* (7), 791-801.

Conner-Spady, B., Cumming, C., Nabholtz, J. M., Jacobs, P., & Stewart, D. (2001). Responsiveness of the EuroQol in breast cancer patients undergoing high dose chemotherapy. *Quality of Life Research*, *10* (6), 479-86.

Davies, J. E. & Doyle, P. M. (2002). Quality of life studies in unselected gynaecological outpatients and inpatients before and after hysterectomy. *Journal of Obstetrics and Gynaecology*, *22* (5), 523-6.

de Wit, G. A., Merkus, M. P., Krediet, R. T., & de Charro, F. T. (2002). Health profiles and health preferences of dialysis patients. *Nephrology, Dialysis, Transplantation*, *17* (1), 86-92.

Degl'innocenti, A., Elmfeldt, D., Hansson, L., Breteler, M., James, O., Lithell, H., Olofsson, B., Skoog, I., Trenkwalder, P., Zanchetti, A. & Wiklund I. (2002). Cognitive function and health-related quality of life in elderly patients with hypertension - baseline data from the study on cognition and prognosis in the elderly (SCOPE). *Blood Pressure*, *11* (3), 157-65.

Delate, T., & Coons, S. J. (2001). The use of 2 health-related quality-of-life measures in a sample of persons infected with human immunodeficiency virus. *Clinical Infectious Diseases*, *32* (3), E47-52.

Devlin, N. J., Hansen, P., Kind, P. & Williams, A. (2003). Logical inconsistencies in survey respondents' health state valuations - a methodological challenge for estimating social tariffs. *Health Economics*, *12* (7), 529-44.

Devlin, N., Hansen, P., & Herbison, P. (2000). Variations in self-reported health status: results from a New Zealand survey. *NZ Medical Journal*, *113* (1123), 517-20.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, *35* (11), 1095-108.

Dolan, P., Gudex, C., Kind, P. & Williams, A. (1996). Valuing health states: a comparison of methods. *Journal of Health Economics*, *15*, 209-231.

Dorman, P. J., Dennis, M., & Sandercock, P. (1999). How do scores on the EuroQol relate to scores on the SF-36 after stroke? *Stroke*, *30* (10), 2146-51.

Dorman, P. J., Slattery, J., Farrell, B., Dennis, M. S., & Sandercock, P. A. (1997). A randomised comparison of the EuroQol and Short Form-36 after stroke. United Kingdom collaborators in the International Stroke Trial. *British Medical Journal*, *315* (7106), 461.

Dorman, P. J., Waddell, F, Slattery, J., Dennis, M. & Sandercock, P. (1997). Is the EuroQol a valid measure of health-related quality of life after stroke? *Stroke*, *28* (10), 1876-82.

Dorman, P., Slattery, J., Farrell, B., Dennis, M., & Sandercock, P. (1998). Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. United Kingdom Collaborators in the International Stroke Trial. *Stroke*, *29* (1), 63-68.

Ekert, H., Brewin, T., Boey, W., Davey, P. & Tilden, D. (2001). Cost-utility analysis of recombinant factor VIIa (NovoSeven) in six children with long-standing inhibitors to factor VIII or IX. *Haemophilia*, *7* (3), 279-85.

Essink-Bot, M. L., Krabbe, P. F., Bonsel, G. J., & Aaronson, N. K. (1997). An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Medical Care*, *35* (5), 522-37.

Essink-Bot, M. L., van Royen, L., Krabbe, P., Bonselm G. J., & Rutten, F. F. (1995). The impact of migraine on health status. *Headache*, *35* (4), 200-6.

EUROQOL Group. (1990). EuroQol: a new facility for measurement of health-related quality of life. *Health Policy*, *16*, 199-208.

Forbes, R. B., Lees, A., Waugh, N. & Swingler, R. J. (1999). Population based cost utility study of interferon beta-1b in secondary progressive multiple sclerosis. *British Medical Journal*, *319* (7224), 1529-33.

Forouzanfar, T., Kemler, M. A., Weber, W. E., Kessels, A. G. & Van Kleef, M. (2004). Spinal cord stimulation in complex regional pain syndrome: cervical and lumbar devices are comparably effective. *British Journal of Anaesthesia*, *92* (3), 348-53.

Foster, J. H., Peters, T. J. & Kind, P. (2002). Quality of life, sleep, mood and alcohol consumption: a complex interaction. *Addict Biol. 7* (1), 55-65.

Fransen, M., & Edmonds, J. (1999). Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology*, *38* (9), 807-13.

Fuertes, M. E., Garcia Matres, M. J., Gonzalez Romojaro, V., de la Rosa, S., Anguera Vila, A., de la Pena, J., Leiva Galvis, O. & Sanchez Chapado, M. (2000). Clinical trial to evaluate trospium chloride (Uraplex) effectiveness and tolerance in patients with detrusor instability incontinence and its impact on quality of life. *Arch Esp de Urol*, *53* (2), 125-136.

Garratt, A. M., Klaber Moffett, J. & Farrin, A. J. (2001). Responsiveness of generic and specific measures of health outcome in low back pain. *Spine*. 26 (1): 71-7; discussion 77.

Glick, H. A., Polsky, D., Willke, R. J., & Schulman, K. A. (1999). A comparison of preference assessment instruments used in a clinical trial: responses to the visual analog scale from the EuroQol EQ-5D and the Health Utilities Index. *Medical Decision Making*, *19* (3), 265-75.

Green, C., Kiebert, G., Murphy, C., Mitchell, J. D, O'Brien, M., Burrell, A. & Leigh P. N. (2003). Patients' health-related quality-of-life and health state values for motor neurone disease/amyotrophic lateral sclerosis. *Quality of Life Research*, *12*, (5), 565-74.

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQol) with four other generic utility instruments. *Annals of Medicine*, *33,* 358–370.

Hollingworth, W., Mackenzie, R., Todd, C. J. & Dixon, A. K. (1995). Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? *Quality of Life Research*, *4* (4), 325-34.

Hurst, N. P., & Mcrorie, E. R. (1998). The short-term health outcome of out-patient rheumatology consultations in relation to rationing: a pilot study. *British Journal of Rheumatology*, *37* (5), 509-13.

Hurst, N. P., Kind, P., Ruta, D., Hunter, M., & Stubbings, A. (1997). Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *British Journal of Rheumatology*, *36* (5), 551-9.

Hurst, N., Jobanputra, P., Hunter, M., Lambert, M., Lochhead, A. & Brown, H. (1994). Validity of Euroqol – a generic health status instrument – in patients with rheumatoid arthritis. Economic and Health Outcomes Research Group. *British Journal of Rheumatology*, *33* (7), 655-662.

James, M., Leger, S., & Rowsell, K. V. (1996). Prioritising elective care: a cost utility analysis of orthopaedics in the north west of England. *Journal of Epidemiology and Community Health*, *50* (2), 182-9.

Jarman, B., Hurwitz, B., Cook, A., Bajekal, M. & Lee, A. (2002). Effects of community based nurses specialising in Parkinson's disease on health outcome and costs: randomised controlled trial. *British Medical Journal*, *324* (7345), 1072-5.

Jelsma, J., Hansen, K., de Weerdt, W., de Cock, P. & Kind, P. (2003). How do Zimbabweans value health states? *Population Health Metric*, *1* (1), 11.

Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S. & Layte, R. (1997). Evaluation of index and profile measures of health status in a randomized controlled trial. Comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and disease specific measures. *Medical Care*, *35* (11), 1109-18.

Jenkinson, C., Stradling, J., & Petersen, S. (1997). Comparison of three measures of quality of life outcome in the evaluation of continuous positive airways pressure therapy for sleep apnoea. *Journal of Sleep Research*, *6* (3), 199-204.

Jenkinson, C., Stradling, J., & Petersen, S. (1998). How should we evaluate health status? A comparison of three methods in patients presenting with obstructive sleep apnoea. *Quality of Life Research*, *7* (2), 95-100.

Johnson, J. A., Connolly, M., Zuberbuhler, P., & Brown, N. E. (2000). Health-related quality of life for adults with cystic fibrosis: a regression approach to assessing the impact of recombinant human DNase. *Pharmacotherapy*, *20* (10), 1167-74.

Johnson, J. A., Coons, S. J., Ergo, A., & Szava-Kovats, G. (1998). Valuation of EuroQOL (EQ-5D) health states in an adult US sample. *PharmacoEconomics*, *13* (4), 421-33.

Joore, M. A., van der Stel, H., Peters, H. J., Boas, G. M. & Anteunis L. J. (2003). The cost-effectiveness of hearing-aid fitting in the Netherlands. *Archives of Otolaryngology – Head and Neck Surgery*, *129* (3), 297-304.

Joshi, H. B., Adams, S., Obadeyi, O. O., & Rao, P. N. (2001). Nephrostomy tube or 'JJ' ureteric stent in ureteric obstruction: assessment of patient perspectives using quality-of-life survey and utility analysis. *European Urology*, *39* (6), 695-701.

Joshi, H. B., Stainthorpe, A., Macdonagh R. P., Keeley, F. X., JR., Timoney, A. G. & Barry, M. J. (2003). Indwelling ureteral stents: evaluation of symptoms, quality of life and utility. *Journal of Urology*, *169* (3), 1065-9, discussion 1069.

Kerrigan, C. L., Collins, E. D., Kneeland, T. S., Voigtlaender, D., Moncur, M. M., Matheney, T. H., Grove, M. R., & Tosteson, A. N. (2000). Measuring health state preferences in women with breast hypertrophy. *Plastic and Reconstructive Surgery, 106* (2), 280-8.

Kerrigan, C. L., Collins, E. D., Striplin, D., Kim, H. M., Wilkins, E., Cunningham, B. & Lowery, J. (2001). The health burden of breast hypertrophy. *Plastic and Reconstructive Surgery*, *108* (6), 1591-9.

Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: results from a United Kingdom national questionnaire survey. *British Medical Journal, 316* (7133), 736-41.

Klassen, A. F., Newton, J. N., & Mallon, E. (2000). Measuring quality of life in people referred for specialist care of acne: comparing generic and disease-specific measures. *Journal of the American Academy of Dermatology*, *43* (2 Pt 1), 229-33.

Kobelt, G. (1997). Economic considerations and outcome measurement in urge incontinence. *Urology*, *50* (6A Suppl), 100-107.

Kobelt, G., Eberhardt, K., Jonsson, L., & Jonsson, B. (1999). Economic consequences of the progression of rheumatoid arthritis in Sweden. *Arthritis and Rheumatism*, *42,* (2), 347-56.

Krabbe, P. & Weijnen, T. (2003). Guidelines for analysing and reporting EQ-5D outcomes. In Brooks R, Rabin R, and de Charro F. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective.* Dordrecht: Kluwer Academic.

Longworth, L. & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics*, *12* (12), 1061-7.

Lubetkin, E. I. & Gold, M. R. (2002). Comprehensibility of measures of health-related quality of life in minority and low-income patients. *Journal of the National Medical Association*, *94* (5), 327-35.

Luo, N., Chew, L. H., Fong, K. Y., Koh, D. R., Ng, S. C., Yoon, K. H., Vasoo, S., Li, S. C., & Thumboo, J. (2003). A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease. *Journal of Rheumatology*, *30* (10), 2268-74.

Macdonagh, R. P., Cliff, A. M., Speakman, M. J., O'boyle, P. J., Ewings, P. & Gudex, C. (1997). The use of generic measures of health-related quality of life in the assessment of outcome from transurethral resection of the prostate. *British Journal of Urology*, *79* (3), 401-8.

Manns, B. J., Johnson, J. A., Taub, K., Mortis, G, Ghali, W. A. & Donaldson, C. (2002). Dialysis adequacy and health related quality of life in hemodialysis patients. *ASAIO Journal*. *48* (5), 565-9.

McCrone, P., Chisholm, D., Knapp, M., Hughes, R., Comi, G., Dalakas, M. C., Illa, I., Kilindireas, C., Nobile,-Orazio, E., Swan, A., van den Bergh, P. & Willison, H. J. (2003). Cost-utility analysis of intravenous immunoglobulin and prednisolone for chronic inflammatory demyelinating polyradiculoneuropathy, *European Journal of Neurology*, *10* (6), 687-94.

McMillan, D. C., Wigmore, S. J., Fearon, K. C., O'gorman, P., Wright, C. E. & Mcardle, C. S. (1999). A prospective randomized study of megestrol acetate and ibuprofen in gastrointestinal cancer patients with weight loss. *British Journal of Cancer*, *79* (3-4), 495-500.

Meerding, W. J., Looman, C. W., Essink-Bot, M. L., Toeth Mulder, S. & Van Beeck, E. F. (2004). Distribution and determinants of health and work status in a comprehensive population of injury patients. *Journal of Trauma*, *56* (1), 150-61.

Miller, P., Kendrick, D., Bentley, E., & Fielding, K. (2002). Cost-effectiveness of lumbar spine radiography in primary care patients with low back pain. *Spine*, *27* (20), 2291-7.

Miners, A. H., Sabin, C. A., Mocroft, A., Youle, M., Fisher, M., & Johnson, M. (2001). Health-related quality of life in individuals infected with HIV in the era of HAART. *HIV Clin Trials*, *2* (6), 484-92.

MVH Group. (1995). *The Measurement and Valuation of Health: Final Modelling of Valuation Tariffs.* York, Centre for Health Economics.

Myers, J. A., Mcpherson, K. M., Taylor, W. J., Weatherall, M. & Mcnaughton, H. K. (2003). Duration of condition is unrelated to health-state valuation on the EuroQoL. *Clinical Rehabilitation*, *17* (2), 209-15.

Norum, J., Angelsen, V., Wist, E. & Olsen, J. A. (1996). Treatment costs in Hodgkin's disease: a cost-utility analysis. *Eur J Cancer*, *32A* (9), 1510-7.

O'Brien, B. J., Goeree, R., Bernard, L., Rosner. A., & Williamson, T. (2001). Cost-Effectiveness of tolterodine for patients with urge incontinence who discontinue initial therapy with oxybutynin: a Canadian perspective. *Clinical Therapeutics*, *23* (12), 2038-49.

Pickard, A. S., Johnson, J. A., Feeny, D. H., Shuaib, A., Carriere, K. C. & Nasser, A. M. (2004). Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. *Stroke*, *35* (2), 607-12.

Pipitone, N., & Scott, D. L. (2001). Magnetic pulse treatment for knee osteoarthritis: a randomised, double-blind, placebo-controlled study. *Current Medical Research and Opinion*, *17* (3), 190-6.

Poley, M. J., Stolk, E. A., Langemeijer, R. A., Molenaar, J. C., & Busschbach, J. J. (2001). The cost-effectiveness of neonatal surgery and subsequent treatment for congenital anorectal malformations. *Journal of Pediatric Surgery*, 36 (10), 1471-8.

Poley, M. J., Stolk, E. A., Tibboel, D., Molenaar, J. C., & Busschbach, J. J. (2002). The cost-effectiveness of treatment for congenital diaphragmatic hernia. *Journal of Pediatric Surgery*. *37* (9), 1245-52.

Polsky, D., Willke, R. J., Scott, K., Schulman, K. A., & Glick, H. A. (2001). A comparison of scoring weights for the EuroQol derived from patients and the general public. *Health Economics*, *10* (1), 27-37.

Rabin, R., & de Charro, F. (2001). EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine*, *33*, 337-343.

Ragnarson Tennvall, G., & Apelqvist, J. (2000). Health-related quality of life in patients with diabetes mellitus and foot ulcers. *Journal of Diabetes and Its Complications*, *14* (5), 235-41.

Ratcliffe, J., Longworth, L., Young, T., Bryan, S., Burroughs, A., & Buxton, M. (2002). Assessing health-related quality of life pre- and post-liver transplantation: a prospective multicenter study. *Liver Transpl*. *8* (3), 263-70.

Redekop, W. K., Koopmanschap, M. A., Stolk, R. P., Rutten, G. E., Wolffenbuttel, B. H. & Niessen, L. W. (2002). Health-related quality of life and treatment satisfaction in Dutch patients with type 2 diabetes. *Diabetes Care*, *25* (3), 458-63.

Richards, D. M., Hughes, S. A., Irving, M. H. & Scott, N. A. (2001). Patient quality of life after successful restorative proctocolectomy is normal. *Colorectal Disease*, *3* (4), 223-6.

Roca-Cusachs, A., Dalfo, A., Badia, X., Aristegui, I., & Roset, M. (2001). Relation between clinical and therapeutic variables and quality of life in hypertension. *Journal of Hypertension*, *19* (10), 1913-9.

Rothwell, P., Mcdowell, Z., Wong, C. & Dorman, P. (1997). Doctors and patients don't agree: cross-sectional study of patients' and doctors' perceptions and assessment of disability in multiple sclerosis. *British Medical Journal*, *314*, 1580-1583.

Schulz, M. W., Chen, J., Woo, H. H., Keech, M., Watson, M. E. & Davey, P. J. (2002). A comparison of techniques for eliciting patient preferences in patients with benign prostatic hyperplasia. *Journal of Urology*, *168* (1), 155-9.

Sennfalt, K., Magnusson, M. & Carlsson, P. (2002). Comparison of hemodialysis and peritoneal dialysis-
-a cost-utility analysis. *Peritoneal Dialysis International*. *22* (1), 39-47.

Siderowf, A., Ravina, B., & Glick, H. A. (2002). Preference-based quality-of-life in patients with Parkinson's disease. *Neurology*, *59* (1), 103-8.

Soderman, P. (2000). On the validity of the results from the Swedish National Total Hip Arthroplasty register. *Acta Orthopaedica Scandinavica. Supplementum*, *71* (296), 1-33.

Sokoll, K. B., & Helliwell, P. S. (2001). Comparison of disability and quality of life in rheumatoid and psoriatic arthritis. *Journal of Rheumatology*, *28* (8), 1842-6.

Stavem, K. (1999). Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Quality of Life Research*, *8* (1-2), 45-54.

Suarez-Almazor, M. E., & Conner-Spady, B. (2001). Rating of arthritis health states by patients, physicians, and the general public. Implications for cost-utility analyses. *Journal of Rheumatology*, *28* (3), 648-56.

Suarez-Almazor, M. E., Kendall, C., Johnson, J. A., Skeith, K., & Vincent, D. (2000). Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology*, *39* (7), 783-90.

Sznajder, M., Aegerter, P., Launois, R., Merliere, Y., Guidet, B., & Cubrea. (2001). A cost-effectiveness analysis of stays in intensive care units. *Intensive Care Medicine*, 27 (1), 146-53.

Thorgrimsen, L., Selwood, A., Spector, A., Royan, L., De Madariaga Lopez, M., Woods R. T. & Orrell, M. (2003). Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. *Alzheimer Dis Assoc Disord*, *17* (4), 201-8.

Tidermark J., Ponzer, S., Svensson, O., Soderqvist, A. & Tornkvist, H. (2003). Internal fixation compared with total hip replacement for displaced femoral neck fractures in the elderly. A randomised, controlled trial. *Journal of Bone and Joint Surgery. British Volume, 85* (3), 380-8.

Tidermark, J., Bergstrom, G., Svensson, O., Tornkvist, H. & Ponzer, S. (2003). Responsiveness of the EuroQol (EQ 5-D) and the SF-36 in elderly patients with displaced femoral neck fractures. *Quality of Life Research*, *12* (8), 1069-79.

Tidermark, J., Blomfeld, R., Ponzer, S, Soderqvist, A. & Tornkvist, H. (2003). Primary total hip arthroplasty with a Burch-Schneider antiprotusion cage and autologous bone grafting for acetabular fractures in elderly patients. *Journal of Orthopaedic Trauma*. 17 (3): 193-7.

Tidermark, J., Zethraeus, N., Svensson, O., Tornkvist, H. & Ponzer, S. (2002). Femoral neck fractures in the elderly: functional outcome and quality of life according to EuroQol. *Quality of Life Research*, *11* (5), 473-81.

Tidermark, J., Zethraeus, N., Svensson, O., Tornkvist, H., & Ponzer, S. (2002). Quality of life related to fracture displacement among elderly patients with femoral neck fractures treated with internal fixation. *Journal of Orthopaedic Trauma*, *16* (1), 34-8.

Tillmann, M., & Silcock, J. (1997). A comparison of smokers' and ex-smokers' health-related quality of life. *Journal of Public Health Medicine*, *19* (3), 268-73.

van Agt H. M., Essink-Bot, M. L., Krabbe, P. F. & Bonsel, G. J. (1994). Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Social Science and Medicine*, *39* (11), 1537-1544.

van Cleemput, P., & Parry, G. (2001). Health status of Gypsy Travellers. *Journal of Public Health Medicine*, *23* (2), 129-34.

van den Hout, W. B., van der Linden, Y. M., Steenland, E., Wiggenraad, R. G., Kievit, J., de Haes, H. & Leer, J. W. (2003). Single- versus multiple-fraction radiotherapy in patients with painful bone metastases: cost-utility analysis based on a randomized trial. *Journal of the National Cancer Institute*, *95* (3), 222-9.

van Dieten, H. E., Perez, R. S., van Tulder, M. W., de Lange, J. J., Zuurmond, W. W., Ader, H. J., Vondeling, H. & Boers, M. (2003). Cost effectiveness and cost utility of acetylcysteine versus dimethyl sulfoxide for reflex sympathetic dystrophy. *PharmacoEconomics*, *21* (2), 139-48.

van Tubergen, A., Boonen, A., Landewe, R., Rutten-van Molken, M., van der Heijde, D., Hidding, A. & van der Linden, S. (2002). Cost effectiveness of combined spa-exercise therapy in ankylosing spondylitis: a randomized controlled trial. *Arthritis and Rheumatism*, *47* (5), 459-67.

Wasserfallen, J. B., Gold, K., Schulman, K. A., & Baraniuk, J. N. (1999). Item responsiveness of a rhinitis and asthma symptom score during a pollen season. *Journal of Asthma*. *36* (5), 459-65.

Wolfe, F., & Hawley, D. J. (1997). Measurement of the quality of life in rheumatic disorders using the EuroQol. *British Journal of Rheumatology*, *36* (7), 786-793.

Wu, A. W., Jacobson, K. L., Frick, K. D., Clark, R., Revicki, D. A., Freedberg, K. A., Scott-Lennox, J., & Feinberg, J. (2002). Validity and responsiveness of the EuroQol as a measure of health-related quality of life in people enrolled in an AIDS clinical trial. *Quality of Life Research*, *11* (3), 273-82.

Reporter: Graeme Hawthorne, 2004

HUI3

INSTRUMENT REVIEW SHEET	
Title:	Health Utilities Index – Version 3.
Abbreviations:	HUI3.
Author(s) name:	Feeny D, Torrance G and Furlong W.
Author(s) address:	Professor David Feeny Institute of Health Economics University of Alberta #1200, 10405 Jasper Avenue Edmonton, AB T5J 3N4 Canada Tel: +7804484881 Fax: +7804480018 Email: dfeeny@pharmacy.ualberta.ca
Supplied by:	Copies of the HUI3 and application forms can be found at: http://www.healthutilities.com/hui3.htm/. Although the actual HUI3 questionnaire is not freely available, descriptions of the 8 attributes and response levels forming the HUI3 descriptive system can be found in Feeny, Furlong et al. (1995), Feeny, Torrance et al. (1996) or Furlong, Feeny et al. (2001).
Cost:	Users must be registered and the instrument is available at a cost of CAN$4,000 per trial.
Administration time:	5-10 minutes.
Training requirements:	None.
Purpose:	The HUI3 is a generic health-related quality of life instrument, designed for use in economic evaluation, specifically cost-utility studies.
Administration:	Self-administered. Can also be administered by interviewer and over the telephone. Special interviewer and telephone administration versions are available. For the interview and telephone versions, the standard items are replaced with dichotomous responses for each response level; i.e. there are a total of 42 questions instead of the standard 15. Verrips et al. evaluated the effect of interview versus mail data collection on adolescents, and reported that on the physical attributes there was no significant difference, but that on the psychological attributes of the HUI3 there was poor agreement. The researchers concluded that self-completion is to be preferred (Verrips, Stuifbergen et al. 2001).
Languages available:	Chinese, Dutch, English, French, German, Italian, Japanese, Portuguese, Spanish, Swedish.
Instrument type:	Multi-attribute utility instrument.

Structure:	The HUI3 comprises 15 items measuring: vision (2 items), hearing (2 items), communication (2 items), happiness, pain (2 items), walking, and use of hands, forgetfulness, thinking, personal care, and worry.
Structure (cont.)	The HUI3 measures "within the skin" functional capacity (Feeny, Torrance et al. 1996), a perspective adopted to enhance its use in clinical studies (Furlong, Feeny et al. 2001). Social aspects of HRQoL are not measured. Consistent with this, Richardson & Zumbo (2000) reported that the HUI3 primarily measured physical health attributes.
	Items have 4–6 response levels. Twelve of the 15 items form the 8 attributes (Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition and Pain) that are used in the utility scoring algorithm. It should be noted that the different response levels are are simply different levels of health; they do not possess interval or cardinal properties.
Scoring:	Item responses are substituted with population weights. The utility weights were elicited using the visual analog scale, and scores then transformed based on four "corner" health states valued with the standard gamble where a 60 year timeframe was used. These results were based on stratified sampling ($n = 256$; response rate 22%) of the Hamilton, Ontario population (Furlong, Feeny et al. 1998).
	A multiplicative function combines the attributes into the utility score (Boyle, Furlong et al. 1995; Furlong, Feeny et al. 1998). The upper boundary is 1.00, and the lower boundary is –0.36, permitting health states worse than death.
Developed for:	The Canadian Health Utilities Index (HUI3), for general population use, is based on the HUI2 which was designed for survivors of childhood cancer. To render it generic and overcome reported difficulties with the HUI2, it was revised into the HUI3 (Feeny, Torrance et al. 1996).
Normative data:	Although the HUI3 has been widely used in Canadian population surveys, normative data have been rarely published. Furlong et al. provide population norms from the 1996/7 Canadian National Population Health Survey at 0.90 (Furlong, Feeny et al. 2001). Mittman et al. provided a breakdown of the 1995/6 Canadian National Population Health Survey ($n = 17,000$) and reported that for those without any health conditions the norm was 0.93 (sd = 0.08) (Mittmann, Trakas et al. 1999).
	Macran et al. reported British population norms for the HUI3, broken down by age group. The range was 0.65–0.97 (Macran, Weatherly et al. 2003). Wang & Chen provide normative data for Singaporeans by each of the HUI3 attributes, but not by HUI3 utility scores (Wang and Chen, 1999).
	Preliminary Australian population values have been published ($n = 343$), giving a range of between 0.77–0.89 depending upon age group (Hawthorne, Richardson et al. 2001).

Clinical data: **Clinical data (cont.)**	The HUI3 has been mostly used with both acute and chronic conditions. Areas include: arthritis/osteoarthritis (Grootendorst, Feeny et al. 2000; Raynauld, Torrance, et al. 2002; Torrance, Raynauld, et al. 2002); brain tumour (Whitton, Rhydderch, et al. 1997); childhood cancer survivors (Barr, Simpson, et al. 1999; Glaser, Furlong, et al. 1999; Le Gales, Costet, et al. 1999; Nixon Speechley, Maunsell, et al. 1999; Szecket, Medin, et al. 1999; Van Schaik, Barr, et al. 1999; Barr, Chalmers, et al. 2000; Felder-Puig, Frey, et al. 2000; Sung, Greenberg, et al. 2003); chronic obstructive pulmonary disease (Schunemann, Griffith et al. 2003); diabetes (Maddigan, Feeny, et al. 2003; Maddigan, Majumdar, et al. 2003); heart disease (Bosch & Hunink, 2000; Bosch, Halpern. et al. 2002; O'Brien, Spath, et al. 2003); hemophilia (Barr, Saleh et al. 2002); hemostasis (Barr, Sek et al. 2003); hip replacement (Blanchard, Feeny et al. 2003; Feeny, Blanchard et al. 2003); rheumatic disease (Luo, Chew et al. 2003); rheumatology (Conner-Spady and Suarez-Almazor 2003); and stroke (Grootendorst, Feeny et al. 2000; Stiell, Nichol et al. 2003; Pickard, Johnson et al. 2004).
Public health data:	Alzheimer's disease and care-giving (Neumann, Sandberg et al. 2000); chronic health conditions (Mittmann, Trakas et al. 1999); drug regime adherence (Cote, Farris et al. 2003); different language speakers in Canada (English vs. French) (Kopec, Williams et al. 2000); inner-city (Macran, Weatherly et al. 2003) and low-income patients (Lubetkin and Gold 2002; Lubetkin & Gold, 2003); low-birth weight children (Verrips, Stuifbergen, et al. 2001); obesity (Trakas, Oh, et al. 2001); and home or residential care (Wodchis, Hirdes, et al. 2003).
Applications:	There are three applications for the HUI3. As a generic instrument, it is applicable to all public health and clinical conditions and interventions. 1. It can be used as a simple additive HRQoL measure, providing profile scores on independent living, social relationships, physical senses, and psychological wellbeing. Used like this, the HUI3 can be used in program evaluation to assess the benefits associated with health interventions. 2. When utilities are computed, the HUI3 can used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions. 3. When utilities are computed, the HUI3 can used in economic evaluation, specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation.

Reliability:	Reliability has been assessed using internal consistency and test-retest.
	Internal consistency was reported for a French cross-cultural adaptation study at Cronbach _ = 81 (Costet, Le Gales et al. 1998). Cronbach _ = 0.79 was reported in the Spanish validation study (Ruiz, Rejas et al. 2003).
	Furlong et al. reported an ICC = 0.91 for the contruction sample of the HUI3 (Furlong, Feeny et al. 1998), which was identical to that reported by Ruiz et al. (Ruiz, Rejas et al. 2003). Test-retest reliability was assessed using telephone administration at a one month interval was ICC = 0.77 (Boyle, Furlong et al. 1995). In a study involving children (>9.5 years) the ICC was reported to be >0.50 (Barr, Simpson et al. 1999). An ICC = 0.75 at 7-day interval was reported for those with rheumatic disease (Luo, Chew et al. 2003).
	The reliability of case completion versus proxy completion has been reported to be ICC = 0.60–0.70 (Pickard, Johnson et al. 2004), suggesting that cases and proxies complete the HUI3 somewhat differently. Regarding different modes of administration, a Dutch study indicated that self-completion is to be preferred (Verrips, Stuifbergen, et al. 2001).
Studies* reported:	(Feeny, Furlong, et al. 1992; Boyle, Furlong, et al. 1995; Feeny, Furlong, et al. 1995; Torrance, Furlong, et al. 1995; Juniper, Guyatt, et al. 1997; Whitton, Rhydderch, et al. 1997; Costet, Le Gales, et al. 1998; Barr, Simpson, et al. 1999; Glaser, Furlong, et al. 1999; Le Gales, Costet, et al. 1999; Mittmann, Trakas, et al. 1999; Nixon, Speechley, Maunsell, et al. 1999; Szecket, Medin, et al. 1999; Van Schaik, Barr, et al. 1999; Wang, & Chen, 1999; Weintraub, Culler, et al. 1999; Barr, Chalmers, et al. 2000; Bosch, & Hunink, 2000; Felder-Puig, Frey, et al. 2000; Grootendorst, Feeny, et al. 2000; Kopec, Williams, et al. 2000; Neumann, Sandberg, et al. 2000; Richardson, & Zumbo, 2000; Furlong, Feeny, et al. 2001; Hawthorne, & Richardson, 2001; Hawthorne, Richardson, et al. 2001; Trakas, Oh, et al. 2001; Verrips, Stuifbergen, et al. 2001; Barr, Saleh, et al. 2002; Bosch, Halpern, et al. 2002; Feeny, Furlong, et al. 2002; Le, Buron, et al. 2002; Raynauld, Torrance, et al. 2002; Torrance, Raynauld, et al. 2002; Barr, Sek, et al. 2003; Blanchard, Feeny, et al. 2003; Conner-Spady & Suarez-Almazor, 2003; Cote, Farris, et al. 2003; Feeny, Blanchard, et al. 2003; Franks, Lubetkin, et al. 2003; Horsman, Furlong, et al. 2003; Lubetkin & Gold, 2003; Luo, Chew, et al. 2003; Macran, Weatherly, et al. 2003; Maddigan, Feeny, et al. 2003; Maddigan, Majumdar, et al. 2003; O'Brien, Spath, et al. 2003; Ruiz, Rejas, et al. 2003; Schultz & Kopec, 2003; Schunemann, Griffith, et al. 2003; Stiell, Nichol, et al. 2003; Sun, 2003; Sung, Greenberg, et al. 2003; Wodchis, Hirdes, et al. 2003; Feeny, Furlong, et al. 2004; Pickard, Johnson, et al. 2004; Raat, Bonsel, et al. 2004).
* = academic papers only.	

References

Barr, R. D., Chalmers, D., de Pauw, S., Furlong, W., Weitzman, S., & Feeny, D. (2000). Health-related quality of life in survivors of Wilms' tumor and advanced neuroblastoma: A cross-sectional study. *Journal of Clinical Oncology*, *18* (18), 3280-7.

Barr, R. D., Saleh, M., Furlong, W., Horsman, J., Sek, J., Pai, M., & Walker, I. (2002). Health status and health-related quality of life associated with hemophilia. *American Journal of Hematology*, *71* (3), 152-60.

Barr, R. D., Sek, J., Horsman, J., Furlong, W., Saleh, M., Pai, M., & Walker, I. (2003). Health status and health-related quality of life associated with von Willebrand disease. *American Journal of Hematology*, *73* (2), 108-14.

Barr, R. D., Simpson, T., Whitton, A., Rush, B., Furlong, W. & Feeny, D. H. (1999). Health-related quality of life in survivors of tumours of the central nervous system in childhood – a preference-based approach to measurement in a cross-sectional study. *European Journal of Cancer*, *35* (2), 248-55.

Blanchard, C., Feeny, D., Mahon, J. L., Bourne, R., Rorabeck, C., Stitt, L., & Webster-Bogaert, S. (2003). Is the Health Utilities Index responsive in total hip arthroplasty patients? *Journal of Clinical Epidemiology, 56* (11), 1046-54.

Bosch, J. L., & Hunink, M. G. (2000). Comparison of the Health Utilities Index Mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. *Quality of Life Research*, *9* (6), 591-601.

Bosch, J. L., Halpern, E. F., & Gazelle, G. S. (2002). Comparison of preference-based utilities of the Short-Form 36 Health Survey and Health Utilities Index before and after treatment of patients with intermittent claudication. *Medical Decision Making*, *22* (5), 403-9.

Boyle, M. H., Furlong, W., Feeny, D., Torrance, G. W. & Hatcher, J. (1995). Reliability of the Health Utilities Index – Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. *Quality of Life Research*, *4*, 249-257.

Conner-Spady, B. & Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care, 41* (7), 791-801.

Costet, N., Le Gales, C., Buron, C., Kinkor, F., Mesbah, M., Chwalow, J. & Slama, G. (1998). French cross-cultural adaptation of the Health Utilities Indexes Mark 2 (HUI2) and 3 (HUI3) classification systems. Clinical and Economic Working Groups. *Quality of Life Research*, *7* (3), 245-56.

Cote, I., Farris, K., & Feeny, D. (2003). Is adherence to drug treatment correlated with health-related quality of life? *Quality of Life Research*, *12* (6), 621-33.

Feeny, D., Blanchard, C., Mahon, J. L., Bourne, R., Rorabeck, C., Stitt, L. & Webster-Bogaert, S. (2003). Comparing community-preference-based and direct standard gamble utility scores: evidence from elective total hip arthroplasty. *International Journal of Technology Assessment in Health Care*, *19* (2), 362-72.

Feeny, D., Furlong, W., Barr, R. D., Torrance, G. W., Rosenbaum, P. & Weitzman, S. (1992). A comprehensiuve multi-attribute system for classifying the health status of survivors of childhood cancer. *Journal of Clinical Oncology*, 923-928.

Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems. Health Utilities Index. *PharmacoEconomics*, *7* (6), 490-502.

Feeny, D., Furlong W, Torrance, G., Goldsmith, C., Zhu, Z., Depauw, S., Denton, M., & Boyle, M. (2002). Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care*, *40* (2), 113-128.

Feeny, D., Torrance, G., & Furlong, W. (1996). Health Utilities Index. B. Spilker *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelpha: Lippincott-Raven Publishers.

Feeny, D., Furlong, W., Saigal, S., & Sun, J. (2004). Comparing directly measured standard gamble scores to HUI2 and HUI3 utility scores: group – and individual – level comparisons. *Social Science & Medicine*, *58* (4), 799-809.

Felder-Puig, R., Frey, E., Sonnleithner, G., Feeny, D., Gadner, H., Barr, R. D., Furlong, W., & Topf, R. (2000). German cross-cultural adaptation of the Health Utilities Index and its application to a sample of childhood cancer survivors. *European Journal of Pediatrics*, *159* (4), 283-8.

Franks, P., Lubetkin, E. I., Gold, M. R., & Tancredi, D. J. (2003). Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population. *Medical Care*, *41* (11), 1277-83.

Furlong, W., Feeny, D., Torrance, G., Goldsmith, C., Depauw, S., Zhu, Z., Denton, M., & Boyle, M. (1998). *Multiplicative Multi-attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report.* Hamilton: McMaster University, Centre for Health Economics and Policy Analysis. Working Paper 98.

Furlong, W. J., Feeny, D. H., Torrance, G. W. & Barr. R. D. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*, *33*, 375-384.

Furlong, W. J., Feeny, D. H., Torrance, G. W., & Barr, R. D. (2001). *The Health Utilities Index (HUI) System for Assessing Health-Related Quality of Life in Clinical Studies.* Hamilton: McMaster University Centre for Health Economics and Policy Analysis.

Glaser, A. W., Furlong, W., Walker, D. A., Fielding, K., Davies, K., Feeny, D. H., & Barr, R. D. (1999). Applicability of the Health Utilities Index to a population of childhood survivors of central nervous system tumours in the U.K. *Eur J Cancer*, *35* (2), 256-61.

Grootendorst, P., Feeny, D., & Furlong, W. (2000). Health Utilities Index Mark 3: evidence of construct validity for stroke and arthritis in a population health survey. *Medical Care*, *38* (3), 290-299.

Hawthorne, G., & Richardson, J. (2001). Measuring the value of program outcomes: a review of utility measures. *Expert Review of Pharmacoeconomics Outcomes Research*, *1* (2), 215-228.

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Annals of Medicine*, *33*, 358–370.

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI(R)): concepts, measurement properties and applications. *Health Quality of Life Outcomes*, *1* (1), 54.

Juniper, E. F., Guyatt, G. H., Feeny, D. H, Griffith, L. E. & Ferrie, P. J. (1997). Minimum skills required by children to complete health-related quality of life instruments for asthma: comparison of measurement properties. *European Respiratory Journal*, *10* (10), 2285-94.

Kopec, J. A., Williams, J. I., To, T., & Austin, P. C. (2000). Measuring population health: correlates of the Health Utilities Index among English and French Canadians. *Canadian Journal of Public Health*, *91* (6), 465-70.

Le Gales, C., Costet, N., Gentet, J. C., Kalifa, C., Frappaz, D., Edan, C., Sariban, E., Plantaz, D., & Doz, F. (1999). Cross-cultural adaptation of a health status classification system in children with cancer. First results of the French adaptation of the Health Utilities Index Marks 2 and 3. *International Journal of Cancer. Supplement*, *12*, 112-8.

Le, G. C., Buron, C., Costet, N., Rosmanm S., & Slama, P. R. (2002). Development of a preference-weighted health status classification system in France: the Health Utilities Index 3. *Health Care Management Science*, *5* (1), 41-51.

Lubetkin, E. I., & Gold, M. R. (2002). Comprehensibility of measures of health-related quality of life in minority and low-income patients. *Journal of the National Medical Association*, *94* (5), 327-35.

Lubetkin, E. I., & Gold, M. R. (2003). Areas of decrement in health-related quality of life (HRQL): comparing the SF-12, EQ-5D, and HUI 3. *Quality of Life Research*, *12* (8), 1059-67.

Luo, N., Chew, L. H., Fong, K. Y., Koh, D. R., Ng, S. C., Yoon, K. H., Vasoo, S., Li, S. C., & Thumboo, J. (2003). A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease. *The Journal of Rheumatology*, *30* (10), 2268-74.

Macran, S., Weatherly, H. & Kind, P. (2003). Measuring population health: a comparison of three generic health status measures. *Medical Care*, *41* (2), 218-31.

Maddigan, S. L., Feeny, D. H., & Johnson, J. A. (2003). A comparison of the health utilities indices Mark 2 and Mark 3 in type 2 diabetes. *Medical Decision Making*, *23* (6), 489-501.

Maddigan, S. L., Majumdar, S. R., Toth, E. L., Feeny, D. H. & Johnson, J. A. (2003). Health-related quality of life deficits associated with varying degrees of disease severity in type 2 diabetes. *Health Quality of Life Outcomes*, *1* (1), 78.

Mittmann, N., Trakas, K., Risebrough, N., & Liu, B. A. (1999). Utility scores for chronic conditions in a community-dwelling population. *PharmacoEconomics*, *15* (4), 369-376.

Neumann, P. J., Sandberg, E. A., Araki, S. S., Kuntz, K. M., Feeny, D., & Weinsteinm, M. C. (2000). A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease. *Medical Decision Making*, *20* (4), 413-22.

Nixon Speechley, K., Maunsell, E., Desmeules, M., Schanzer, D., Landgraf, J. M., Feeny, D. H., & Barrera, M. E. (1999). Mutual concurrent validity of the child health questionnaire and the health utilities index: an exploratory analysis using survivors of childhood cancer. *International Journal of Cancer, 12* (Supp), 95-105.

O'Brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P., & Brazier, J. (2003). A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics*, *12* (11), 975-81.

Pickard, A. S., Johnson, J. A., Feeny, D. H., Shuaib, A., Carriere, K. C., & Nasser, A. M. (2004). Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. *Stroke*, *35* (2), 607-12.

Raat, H., Bonsel, G. J., Hoogeveen, W. C., & Essink-Bot, M. L. (2004). Feasibility and reliability of a mailed questionnaire to obtain visual analogue scale valuations for health states defined by the Health Utilities Index Mark 3. *Medical Care*, *42* (1), 13-8.

Raynauld, J. P., Torrance, G. W., Band, P. A., Goldsmith, C. H., Tugwell, P., Walker, V., Schultz, M., & Bellamy, N. (2002). A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results. *Osteoarthritis and Cartilage*, *10* (7), 506-17.

Richardson, C., & Zumbo, B. (2000). A statistical examination of the Health Utility Index-Mark III as a summary measure of health status for a general population survey. *Social Indicators Research*, *51* (2), 171-191.

Ruiz, M., Rejas, J., Soto, J., Pardo, A., & Rebollo, I. (2003). Adaptation and validation of the Health Utilities Index Mark 3 into Spanish and correction norms for Spanish population. *Medicina Clinica*, *120* (3), 89-96.

Schultz, S. E., & Kopec, J. A. (2003). Impact of chronic conditions. *Health Reports*, *14* (4), 41-53.

Schunemann, H. J., Griffith, L., Stubbing, D., Goldstein, R., & Guyatt, G. H. (2003). A clinical trial to evaluate the measurement properties of 2 direct preference instruments administered with and without hypothetical marker states. *Medical Decision Making*, *23* (2), 140-9.

Stiell, I., Nichol, G., Wells, G., de Maio, V., Nesbitt, L., Blackburn, J., & Spaite, D. (2003). Health-related quality of life is better for cardiac arrest survivors who received citizen cardiopulmonary resuscitation. *Circulation*, *108* (16), 1939-44.

Sun, J. (2003). Adjusting distributions of the Health Utilities Index Mark 3 utility scores of health-related quality of life. *Quality of Life Research*, *12* (1), 11-20.

Sung, L., Greenberg, M. L., Doyle, J. J., Young, N. L., Ingber, S., Rubenstein, J., Wong, J., Samanta, T., Mclimont, M., & Feldman, B. M. (2003). Construct validation of the Health Utilities Index and the Child Health Questionnaire in children undergoing cancer chemotherapy. *British Journal of Cancer*, *88* (8), 1185-90.

Szecket, N., Medin, G., Furlong, W. J., Feeny, D. H., Barr, R. D., & Depauw, S. (1999). Preliminary translation and cultural adaptation of Health Utilities Index questionnaires for application in Argentina. *International Journal of Cancer. Supplement*, *12*, 119-24.

Torrance, G. W., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions. Health Utilities Index. *PharmacoEconomics*, *7* (6), 503-20.

Torrance, G. W., Raynauld, J. P., Walker, V., Goldsmith, C. H., Bellamy, N., Band, P. A., Schultz, M., & Tugwell, P. (2002). A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 2 of 2): economic results. *Osteoarthritis and Cartilage*, *10* (7), 518-27.

Trakas, K., Oh, P. I., Singh, S., Risebrough, N., & Shear, N. H. (2001). The health status of obese individuals in Canada. *International Journal of Obesity and Related Metabolic Disorders*, *25* (5), 662-8.

van Schaik, C. S., Barr, R. D., Depauw, S., Furlong, W. & Feeny, D. (1999). Assessment of health status and health-related quality of life in survivors of Hodgkin's disease in childhood. *International Journal of Cancer. Supplement*, *12*, 32-8.

Verrips, G. H., Stuifbergen, M. C., den Ouden, A. L., Bonsel, G. J., Gemke, R. J., Paneth, N., & Verloove-Vanhorick, S. P. (2001). Measuring health status using the Health Utilities Index: agreement between raters and between modalities of administration. *Journal of Clinical Epidemiology*, *54* (5), 475-81.

Wang, Q., & Chen, G. (1999). The health status of the Singaporean population as measured by a multi-attribute health status system. *Singapore Medical Journal*, *40* (6), 389-96.

Weintraub, W. S., Culler, S. D., Kosinski, A., Becker, E. R., Mahoney, E., Burnette, J., Spertus, J. A., Feeny, D., Cohen, D. J., Krumholz, H., Ellis, S. G., Demopoulos, L., Robertson, D., Boccuzzil, S. J., Barr, E., & Cannon, C. P. (1999). Economics, health-related quality of life, and cost-effectiveness methods for the TACTICS (Treat Angina with Aggrastat [tirofiban] and Determine Cost of Therapy with Invasive or Conservative Strategy)-TIMI 18 trial. *American Journal of Cardiology*, *83* (3), 317-22.

Whitton, A. C., Rhydderch, H., Furlong, W., Feeny, D., & Barr, R. D. (1997). Self-reported comprehensive health status of adult brain tumor patients using the Health Utilities Index. *Cancer*, *80* (2), 258-65.

Wodchis, W. P., Hirdes, J. P., & Feeny, D. H. (2003). Health-related quality of life measure based on the minimum data set. *International Journal of Technology Assessment in Health Care*, *19* (3), 490-506.

Reporter: Graeme Hawthorne, 2004

15D

INSTRUMENT REVIEW SHEET	
Title:	15D.
Abbreviations:	15D.
Author(s) name:	Sintonen H.
Author(s) address:	Harri Sintonen, Professor of Health Economics Department of Public Health P.O. Box 41 00014 University of Helsinki, Finland Tel. +358 (0)9 19127562 Fax +358 (0)9 19127540 Mobile +358 40 5464298
Supplied by:	A copy of the 15D can be found in Sintonen, 2001.
Cost:	Free. Registration is required. To register contact Professor Harri Sintonen (Harri.Sintonen@helsinki.fi).
Administration time:	5-10 minutes.
Training requirements:	None.
Purpose:	The 15D is a generic health-related quality of life instrument, designed for use in economic evaluation, specifically cost-utility studies.
Administration:	Self-administered. Can also be administered by interviewer and over the telephone.
Languages available:	Arabic, Czech, Danish, English, Finnish, German, Greek, Hebrew, Japanese, Norwegian, Russian, Serbian, Swedish, Turkish.
Instrument type:	Multi-attribute utility instrument.
Structure:	The 15D comprises 15 items measuring: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort & symptoms, depression, distress, vitality and sexual activity. Each item is measured on a 5-point scale, where the endpoints are 1 = "no problems/feeling fine" to 5 = "extremely" difficult/compromised. It should be noted that the different response levels are are simply different levels of health; they do not possess interval or cardinal properties.
Scoring:	There are two ways the 15D can be scored. The published population weights may be used, or specific importance study weights can be derived and then used. Most users do not collect their own weights. Only two studies using specific study importance weights appear to have been published see Kannisto & Sintonen, or Vaara, Sintonen, et al. 1999. Usually, item responses are substituted with population weights. The weights came from five random samples of the Finnish population ($n = 1290$ respondents; response rate 51%) using visual analog scale questions; responses were combined using a simple additive model (Sintonen 1994; Sintonen 1995). The upper boundary is 1.00, and the lower boundary is +0.11: death-equivalent and worse than death health states are not allowed.
Developed for:	The 15D was defined by Finnish health concerns, the WHO definition of health and medical and patient feedback (Sintonen and Pekurinen 1993; Sintonen 2001). It is concerned with impairment and disability of "within the skin" functions.

Normative data:	No Australian population norms have been published, although Hawthorne et al. (2001) reported values for community samples (n = 334; the range was 0.85–0.92 depending on age group).
	Finnish population norms for individual items have been published (Sintonen 2001), but population norms do not appear to have been published yet.
Clinical data:	The 15D has been used in a range of medical conditions, including asthma (Kauppinen, Sintonen, et al. 1998; Kauppinen, Sintonen, et al. 1998; Ritva, Pekka, et al. 2000; Kauppinen, Vilkka, et al. 2001); brain tumour (Salo, Niemela, et al. 2002); chronic obstructive pulmonary disease (Stavem, 1999); depression (Lonnqvist, Sintonen, et al. 1994; Lonnqvist, Sihvo, et al. 1995); diabetes and ageing (Hahl, Hamalainen, et al. 2002), epilepsy (Stavem, 1998); hip and knee replacements (Rissanen, Aro, et al. 1996; Rissanen, Aro, et al. 1997; Rissanen, Sogaard, et al. 2000); osteoporosis (Andersen, Kristiansen, et al. 1995); prostate cancer (Nygard, Norum, et al. 2001); spinal cord injury or deformity (Kannisto, & Sintonen, 1997; Kannisto, Merikanto, et al. 1998; Vaara, Sintonen, et al. 1999) and urinary incontinence (Stach-Lempinen, Kujansuu, et al. 2001).
Public health data:	The 15D has been used more in clinical settings that population settings. Studies include general population quality of life status (Hawthorne & Richardson, 2001; Hawthorne, Richardson, et al. 2001).
Applications:	There are three applications for the 15D. As a generic instrument, it is applicable to all public health and clinical conditions and interventions.
	1. It can be used as a simple additive HRQoL measure, providing profile scores on each of the 15 items in the instrument. Used like this, the 15D can be used in program evaluation to assess the benefits associated with health interventions.
	2. When utilities are computed through simple unweighted addition of item scores, the 15D can be used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions.
	3. When utilities are computed through use weighting item responses, the 15D can be used in economic evaluation, specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation. Because the weights used in scoring the 15D were obtained using VAS values, it may be that the 15D is not a "true" utility measure (it is an axiom of utility theory that "value" is obtained under uncertainty; VAS provide data that is "rated").
Reliability:	Stavem (1999) reported that among patients with chronic pulmonary disease the test-retest Spearman correlation with 14 days between administrations was r_s = 0.90. Based on the construction samples for the weights used, Sintonen reported that the groups means between samples provided a reliability estimate of Spearman r_s = 0.94 (Sintonen 1995).

Studies* reported:	(Sintonen 1981; Sung, Olukotun, et al. 1987; Sintonen, 1993; Lonnqvist, Sintonen, et al. 1994; Lonnqvist, Sihvo, et al. 1995; Rissanen, Aro, et al. 1996; Kannisto & Sintonen, 1997; Rissanen, Aro, et al. 1997; Kannisto, Merikanto, et al. 1998; Kauppinen, Sintonen, et al. 1998; Kauppinen, Sintonen, et al. 1998; Stavem, 1998; Stavem, 1999; Vaara, Sintonen, et al. 1999; Galli, Crabb, et al. 2000; Rissanen, Sogaard, et al. 2000; Ritva, Pekka, et al. 2000; Stavem, Bjornaes, et al. 2000; Sullivan, Kempen, et al. 2000; Hawthorne & Richardson, 2001; Hawthorne, Richardson, et al. 2001; Hawthorne, Richardson, et al. 2001; Kauppinen, Vilkka, et al. 2001; Nygard, Norum, et al. 2001; Sintonen, 2001; Stach-Lempinen, Kujansuu, et al. 2001; Hahl, Hamalainen, et al. 2002; Penttinen, Nevala-Puranen, et al. 2002; Salo, Niemela, et al. 2002; Bowden & Fox-Rushby, 2003).
* = academic papers only.	

References

Andersen, L., Kristiansen, I., Falch, J., & Aursnes, I. (1995). *Cost-effectiveness of Alendronate for the Prevention of Osteoporotic Fractures in Norwegian Women.* Oslo: Folkehelsa; Statens Institutt for Folkehelsa. Working Paper 11.

Bowden, A., & Fox-Rushby, J. A. (2003). A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Social Science & Medicine*, *57* (7), 1289-306.

Galli, A., Crabb, D., Price, D., Ceni, E., Salzano, R, Surrenti, C. & Casini, A. (2000). Peroxisome proliferator-activated receptor gamma transcriptional regulation is involved in platelet-derived growth factor-induced proliferation of human hepatic stellate cells. *Hepatology*, *31* (1), 101-8.

Hahl, J., Hamalainen, H., Sintonen, H., Simell, T., Arinen, S., & Simell, O. (2002). Health-related quality of life in type 1 diabetes without or with symptoms of long-term complications. *Quality of Life Research*, *11* (5), 427-36.

Hawthorne, G. & Richardson, J. (2001). Measuring the value of program outcomes: a review of utility measures. *Expert Review of Pharmacoeconomics Outcomes Research*, *1* (2), 215-228.

Hawthorne, G., Richardson, J., & Day, N. (2001). A comparison of five multi-attribute utility instruments. *Australian Studies in Health Service Administration*, *89*, 151-179.

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQol) with four other generic utility instruments. *Annals of Medicine*, *33*, 358–370.

Kannisto, M., & Sintonen, H. (1997). Later health-related quality of life in adults who have sustained spinal cord injury in childhood. *Spinal Cord*, *35* (11), 747-51.

Kannisto, M., Merikanto, J., Alaranta, H., Hokkanen, H., & Sintonen, H. (1998). Comparison of health-related quality of life in three subgroups of spinal cord injury patients. *Spinal Cord*, *36* (3), 193-9.

Kauppinen, R., Sintonen, H., & Tukiainen, H. (1998). One-year economic evaluation of intensive vs conventional patient education and supervision for self-management of new asthmatic patients. *Respiratory Medicine*, *92* (2), 300-7.

Kauppinen, R., Sintonen, H., Vilkka, V., Pekurinen, M., & Tukiainen, H. (1998). Quality-of-life measures and clinical parameters in asthmatics during three year follow-up. *Monaldi Archives for Chest Disease*, *53* (4), 400-4.

Kauppinen, R., Vilkka, V., Sintonen, H., Klaukka, T., & Tukiainen, H. (2001). Long-term economic evaluation of intensive patient education during the first treatment year in newly diagnosed adult asthma. *Respiratory Medicine*, *95* (1), 56-63.

Lonnqvist, J., Sihvo, S., Syvalahti, E., Sintonen, H., Kiviruusu, O., & Pitkanen, H. (1995). Moclobemide and fluoxetine in the prevention of relapses following acute treatment of depression. *Acta Psychiatrica Scandinavica*, *91* (3), 189-94.

Lonnqvist, J., Sintonen, H., Syvalahti, E., Appelberg, B., Koskinen, T., Mannikko, T., Mehtonen, O., Naarala, M., Sihvo, S., Auvinen, J., et al. (1994). Antidepressant efficacy and quality of life in depression: a double-blind study with moclobemide and fluoxetine. *Acta Psychiatrica Scandinavica*, *89* (6), 363-369.

Nygard, R., Norum, J., & Due, J. (2001). Goserelin (Zoladex) or orchiectomy in metastatic prostate cancer? A quality of life and cost-effectiveness analysis. *Anticancer Research*, *21* (1B), 781-8.

Penttinen, J., Nevala-Puranen, N., Airaksinen, O., Jaaskelainen, M., Sintonen, H. & Takala, J. (2002). Randomized controlled trial of back school with and without peer support. *Journal of Occupational Rehabilitation, 12* (1), 21-9.

Rissanen, P., Aro, S., Sintonen, H., Asikainen, K., Slatis, P., & Paavolainen, P. (1997). Costs and cost-effectiveness in hip and knee replacements. A prospective study. *International Journal of Technology Assessment in Health Care*, *13* (4), 575-88.

Rissanen, P., Aro, S., Sintonen, H., Slatis, P., & Paavolainen, P. (1996). Quality of life and functional ability in hip and knee replacements: a prospective study. *Quality of Life Research*, *5* (1), 56-64.

Rissanen, P., Sogaard, J., & Sintonen, H. (2000). Do QOL instruments agree? A comparison of the 15D (Health-Related Quality of Life) and NHP (Nottingham Health Profile) in hip and knee replacements. *International Journal of Technology Assessment in Health Care*, *16* (2), 696-705.

Ritva, K., Pekka, R., & Harri, S. (2000). Agreement between a generic and disease-specific quality-of-life instrument: the 15D and the SGRQ in asthmatic patients. *Quality of Life Research*, *9* (9), 997-1003.

Salo, J., Niemela, A., Joukamaa, M., & Koivukangas, J. (2002). Effect of brain tumour laterality on patients' perceived quality of life. *Journal of Neurology, Neurosurgery and Psychiatry*, *72* (3), 373-7.

Sintonen, H., & Pekurinen, M. (1993). A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In Walker S. and Rosser R. *Quality of Life Assessment: Key issues in the 1990s*. Dordrecht: Kluwer Academic Publishers.

Sintonen, H. (1981). An approach to measuring and valuing health states. *Social Science & Medicin, 15*C, 55-65.

Sintonen, H. (1993). Health-related quality of life measures. *Sairaanhoitaja, 4*, 17-9.

Sintonen, H. (1994). *The 15D Measure of Health-Related Quality of Life: Reliability, Validity and Sensitivity of its Health State Descriptive System.* Melbourne: National Centre for Health Program Evaluation. Working Paper 41.

Sintonen, H. (1995). *The 15D Measure of Health-Related Quality of Life: Feasibility, Reliability and Validity of its Valuation System.* Melbourne: National Centre for Health Program Evaluation. Working Paper 42.

Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine*, *33*, 328-336.

Stach-Lempinen, B., Kujansuu, E., Laippala, P., & Metsanoja, R. (2001). Visual analogue scale, urinary incontinence severity score and 15 D - psychometric testing of three different health-related quality-of-life instruments for urinary incontinent women. *Scandinavian Journal of Urology and Nephrology*, *35* (6), 476-83.

Stavem, K., Bjornaes, H., & Lossius, M. I. (2000). Reliability and validity of a Norwegian version of the quality of life in epilepsy inventory (QOLIE-89). *Epilepsia*, *41* (1), 91-7.

Stavem, K. (1998). Quality of life in epilepsy: comparison of four preference measures. *Epilepsy Research*, *29* (3), 201-9.

Stavem, K. (1999). Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Quality of Life Research*, *8* (1-2), 45-54.

Sullivan, M.D., Kempen, G. I., van Sonderen, E., & Ormel, J. (2000). Models of health-related quality of life in a population of community-dwelling Dutch elderly. *Quality of Life Research*, *9* (7), 801-10.

Sung, R. J., Olukotun, A. Y., Baird, C. L., & Huycke, E. C. (1987). Efficacy and safety of oral nadolol for exercise-induced ventricular arrhythmias. *American Journal of Cardiology*, *60* (6), 15D-20D.

Vaara, P., Sintonen, H., Peltonen, J., Hokkanen, H., Poussa, M., & Ryoppy, S. (1999). Health-related quality of life in patients with diastrophic dysplasia. *Scandinavian Journal of Public Health*, *27* (1), 38-42.

Reporter: Graeme Hawthorne, 2004

SF6D

INSTRUMENT REVIEW SHEET	
Title:	SF-6D.
Abbreviations:	SF6D.
Author(s) name:	Brazier J, Roberts J & Deverill M.
Author(s) address:	Profesor John Brazier Sheffield Health Economics Group School of Health and Related Research University of Sheffield Regent Court, 30 Regent Street Sheffield S1 4DA UK T: 0114 2220715 F: 0114 2724095 E: <j.e.brazier@sheffield.ac.uk>
Supplied by:	The SF6D algorithm is based on the SF-36 items and data. For copies of the SF-36 see the SF-36 Review Sheet. The descriptive systems for the two different versions of the SF6D algorithm are given in Brazier, Usherwood et al. (1998) and Brazier, Roberts et al. (2002). It is the latter version that is described here.
Cost:	Details of costs for use of the SF-36 can be found on the SF-36 Review Sheet. There are no costs for using the SF6D algorithm.
Administration time:	10 minutes.
Training requirements:	None.
Purpose:	The SF6D is a generic health-related quality of life instrument, designed for use in economic evaluation, specifically cost-utility studies where SF-36 data have been collected.
Administration:	Details of SF-36 administration are given on the SF-36 Review Sheet.
Languages available:	See the SF-36 Review Sheet.
Instrument type:	Multi-attribute utility instrument algorithm.
Structure:	Two different algorithms have been published for deriving preference-based values from the SF-36 (Brazier, Usherwood et al. 1998; Brazier, Roberts et al. 2002). The latter algorithm replaces the earlier version, which was a preliminary algorithm. It is this latter algorithm which is described here. The SF6D (Brazier, Roberts et al. 2002) uses 10 items from the SF-36: three from the physical functioning scale, one from physical role limitation, one from emotional role limitation, one from social functioning, two bodily pain items, two mental health items and one vitality item. These form 6 dimensions: Physical Functioning (PF: 6 levels), Role Limitation (RL: 4 levels), Social Functioning (SF: 5 levels), Pain (PA: 6 levels), Mental Health (MH: 5 levels) and Vitality (VI: 5 levels).

Scoring:	Item responses are substituted with population weights. These weights were computed from visual analog scale scores, which were modelled using standard gamble values for two link health states. Values were obtained from a random sample ($n = 611$) of the British population. An additive econometric model is used to compute the utility index.
	The endpoints for the SF6D are 1.00 (best health state) and 0.30 for the worst possible health state.
Developed for:	The SF6D was developed to enable the conversion of SF-36 data to utilities for use in economic evaluation.
Normative data:	No normative population data have been published, almost certainly due to the SF6D's recent publication.
	Hawthone et al. reported a range of SF6D original algorithm community sample ($n = 343$) scores from 0.89 to 0.93 (Hawthorne, Richardson, et al. 2001).
Clinical data:	The SF6D has been used in studies of arthritis (Russell, Conner-Spady et al. 2003); heart disease (O'Brien, Spath et al. 2003); HIV/AIDS (Schackman, Goldie et al. 2002); liver transplant (Longworth & Bryan, 2003); and musculo-skeletal disease (Conner-Spady, & Suarez-Almazor, 2003).
Public health data:	No public health data have been published.
Applications:	There are two applications for the SF6D. As a generic instrument, it is applicable to all public health and clinical conditions and interventions.
	1. When utilities are computed, the SF6D can be used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions.
	2. When utilities are computed, the HUI3 can be used in economic evaluation, specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation.
Reliability:	No estimates of reliability for the SF6D have been published.
Studies* reported:	(Brazier, Usherwood, et al. 1998; Hawthorne, & Richardson, 2001; Hawthorne, Richardson, et al. 2001; Hawthorne, Richardson, et al. 2001; Brazier, Roberts, et al. 2002; Schackman, Goldie, et al. 2002; Conner-Spady & Suarez-Almazor, 2003; Kopec, & Willison, 2003; Longworth, & Bryan, 2003; O'Brien, Spath, et al. 2003; Russell, Conner-Spady, et al. 2003; Walters & Brazier, 2003)
* = academic papers only.	

References

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21,* 271-292.

Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology*, *51* (11), 1115-1128.

Conner-Spady, B., & Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care, 41* (7), 791-801.

Hawthorne, G., & Richardson, J. (2001). Measuring the value of program outcomes: a review of utility measures. *Expert Review of Pharmacoeconomics Outcomes Research*, *1* (2), 215-228.

Hawthorne, G., Richardson, J., & Day, N. (2001). A comparison of five multi-attribute utility instruments. *Australian Studies in Health Service Administration*, *89*, 151-179.

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQol) with four other generic utility instruments. *Annals of Medicine*, *33*, 358–370.

Kopec, J. A. & Willison, K. D. (2003). A comparative review of four preference-weighted measures of health-related quality of life. *Journal Clinical Epidemiology*, *56* (4), 317-25.

Longworth, L., & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics*, *12* (12), 1061-7.

O'brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P. & Brazier, J. (2003). A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics*, *12* (11), 975-81.

Russell, A. S., Conner-Spady, B., Mintz, A., & Maksymowych, W. P. (2003). The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *Journal of Rheumatology*, *30* (5), 941-7.

Schackman, B. R., Goldie, S. J., Freedberg, K. A., Losina, E., Brazier, J., & Weinstein, M. C. (2002). Comparison of health state utilities using community and patient preference weights derived from a survey of patients with HIV/AIDS. *Medical Decision Making*, *22* (1), 27-38.

Walters, S. J., & Brazier, J. E. (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes*, *1* (1), 4.

Reporter: Graeme Hawthorne, 2004

Generic Health-Related Quality of Life Measures

SF-12® Health Survey (Version 1.0)

Title:	SF-12® Health Survey (Version 1.0) for use in Australia (*also known as the Short-Form 12-Item Health Survey*).
Abbreviations:	SF-12.
Author(s) Name:	John E. Ware, Jr.
Author(s) Address:	QualityMetric Incorporated 640 George Washington Highway Lincoln, RI 02865 USA www.qualitymetric.com
Supplied by:	QualityMetric Incorporated 640 George Washington Highway Lincoln, RI 02865 USA
	In Australia, SF-12® Health Survey manuals can be obtained from the:
	Australian Health Outcomes Collaboration (AHOC) c/- Centre for Health Service Development University of Wollongong NSW 2522 Phone: 02 4221-4411
Cost:	An annual license fee applies for the use of the SF-12® Health Survey.
	Survey users are required to register with QualityMetric Incorporated and obtain a quote for the annual license fee that applies to their project.
	The license charge will depend upon whether users require a commercial or research license.
	Register online at www.qualitymetric.com Information on the SF group of instruments can also be found at http://www.sf-36.com/
	SF-12® manuals can be purchased in Australia from AHOC by contacting Laura Willmott at willmott@uow.edu.au or by telephone on 02 4221-4411.
	For technical questions about using the SF-12® Health Survey in Australia (including latest developments and research advice) contact Jan Sansoni at jansan@netspeed.com.au or by telephone on 02 6291-7271 or 02 6205-0869.
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. For those professionals without these qualifications basic training is required in survey administration and the characteristics of the SF-12® Health Survey. The AHOC provides training workshops for the SF-12 and other instruments.
Purpose:	A shorter version of the SF-36® Health Survey designed to reproduce the Physical Component Summary (PCS) and the Mental Component Summary (MCS) scores.
Administration Time:	2 minutes.
Instrument Type:	Self-report Questionnaire.
Structure:	The SF-12® Health Survey includes 12 questions from the SF-36® Health Survey (Version 1). These include: 2 questions concerning physical functioning; 2 questions on role limitations because of physical health problems; 1 question on bodily pain; 1 question on general health perceptions; 1 question on

vitality (energy/fatigue); 1 question on social functioning; 2 questions on role limitations because of emotional problems; and 2 questions on general mental health (psychological distress and psychological well-being).

Scoring:	Scoring of individual items is identical to the SF-36® Health Survey. Scoring algorithms are then applied to produce the PCS and MCS scores.
Developed for:	Those who need an even shorter generic measure of perceived health status.
Normative Data:	The SF-12® Health Survey was developed using normative data for the SF-36® Health Survey in the United States. [See Ware, Kosinski & Keller (1994) and Ware, Kosinski, Bayliss, McHorney, Rogers & Raczek (1995)] Wilson, Tucker & Chittleborough (2002) and Sanderson & Andrews (2002) have conducted local equivalence studies and found the SF-12 suitable for use in Australia.
	Population health data using the SF-12 can be found in the 1997 Australian National Survey of Mental Health and Well-being, the 2000 Mental Health Status of South Australian Population Study, the 2002 Longitudinal Investigation of Depression Outcomes (LIDO) Study and the 2003 Australian Gulf War Veteran´s Health Study.
Clinical Data:	A few clinical studies are listed below:
	Arthritis: Gandhi, Salmon, Zhao, Lambert, Gore & Conrad (2001).
	Diabetes: Siddique, Ricci, Stewart, Sloan & Farup (2002).
	Elective Surgery: Derrett, Devlin, Hansen & Herbison (2003).
	Heart and Stroke Patients: Lim & Fisher (1999).
	Homeless Persons: Larson (2002).
	Myocardial Infarction: McBurney, Eagle, Kline-Rogers, Cooper, Mani, Smith et al. (2002).
	Older Adults in a retirement community: Resnick & Nahm (2001).
	Retinal Diseases: Globe, Levin, Chang, Mackenzie & Azen (2002).
Applications:	In choosing between the SF-12® and the SF-36® Health Surveys users should consider the trade-off between test taker burden (i.e. number of questions, time to complete) and the precision of scores (i.e. how reliable does the obtained score need to be). Ware et al. (1996) reports that there is a 10% loss in the SF-12´s ability to distinguish between different disease groups as compared to the SF-36 and that the SF-12 less accurately reproduces the eight scale profile of the SF-36. Therefore it is recommended that the SF-36 be used for smaller studies (less than n = 500). A recent paper by Rubenach, Shadbolt, McCallum & Nakamura (2002)[16] highlights this important distinction for clinical research studies.
	Sanderson & Andrews have done considerable work in utilising the SF-12 (MCS) as a disability measure for mental health disorders (especially anxiety and depression). Salyers et al. (2000) have utilised the SF-12 (MCS) for severe mental illness.
	The SF-12 has been administered using interactive voice recognition technology and in computerised format. Telephone vs. mail-out administration has also been compared.
	An acute (1 week) version of the SF-12® Health Survey is also available.
	Like the SF-36® Health Survey, the SF-12® Health Survey has been recently updated by QualityMetric Incorporated. The new version is known as the SF-12v2™ Health Survey (Version 2). However, this update of the SF-12 has yet to be field tested in Australia for equivalence or new norms developed for the Australian population.
	See also the Instrument Review on the SF-36® Health Survey.

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	NA	NA	NA	The important issue here is how well the SF-12 reproduces the PCS and MCS scores of the SF-36.
Test – retest	Yes	Ware et al. (1996). Salyers et al. (2000). Lenert (2000). Resnick & Parker (2001).	Adequate	Test-Retest Reliability – PCS = 0.89; MCS = 0.76.
Inter – rater	NA	NA	NA	The SF-12 is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Ware et al. (1996). Sugar et al. (1998). Sanderson et al. (2001). Sanderson & Andrews (2002).	Adequate	See also the references in the Construct Validity section.
Correlation with other measures	Yes	Ware et al. (1996). Johnson & Coons (1998). Lundberg et al. (1999). Burdine et al. (2000). Marcan et al. (2003).	Good	
Construct	Yes	Ware et al. (1996). Jenkinson & Layte (1997). Gandek et al. (1998).	Good	The SF-12 PCS and MCS scores correlate 0.95 and 0.96 with their SF-36 counterparts.
Criterion	Yes	Ware et al. (1996). Jenkinson & Layte (1997). Gandek et al. (1998).	Good	The criterion is how well the SF-12 reproduces the PCS and MCS scores of the SF-36 (see above).

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Jenkinson et al. (1997). Sugar et al. (1998). Lenert et al. (2000). Riddle et al. (2001).	Adequate	

Cultural Applicability and Cultural Adaptations:	Jenkinson, Chandola, Coulter & Bruster (2001) in the United Kingdom have made a useful contribution in this area. However, in Australia, little research has been reported on the use of SF-12 with people from a non-English speaking background and Aboriginal and Torres Strait Islanders.
Gender Appropriateness:	Normative data is available for males and females.
Age Appropriateness:	14 years and over.
Summary:	The SF-12® Health Survey is a suitable measure for large group epidemiological studies (greater than $n = 500$) where information on the SF-36® Health Survey Summary Scores (PCS + MCS) is required.

References

Andrews, G., (2002). A brief integer scorer for the SF-12 validity of the brief scorer in Australian community and clinic settings. *Australian & New Zealand Journal of Public Health*, *26*, 508-510.

Burdine, J. N., Felix, M. R., Abel, A. L., Wiltraut, C. J., & Musselman, Y. J. (2000). The SF-12 as a population health measure: an exploratory examination of potential for application. *Health Services Research*, *35*, 885-904.

Derrett, S., Devlin, N., Hansen, P., & Herbison, P. (2003). Prioritizing patients for elective surgery: a prospective study of clinical priority assessment criteria in New Zealand. *International Journal of Technology Assessment in Health Care*, *19*, 91-105.

Gandek, B., Ware, J.E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, et al. (1998). Cross validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51*, 1171-1180.

Gandhi, S. K., Salmon, J. W., Zhao, S. Z., Lambert, B. L., Gore, P. R., & Conrad, K. (2001). Psychometric evaluation of the 12-item short-form survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clinical Therapeutics*, *23*, 1080-1098.

Globe, D. R., Levin, S., Chang, T. S., Mackenzie, P. J., & Azen, S., (2002). Validity of the SF-12 quality of life instrument in patients with retinal diseases. *Ophthalmology*, *109*, 1793-1798.

Herrman, H., Patrick, D. L., Diehr, P., Martin, M. L., Fleck, M., Simon, G. E., et al. (2002). Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. *Psychological Medicine*, *32*, 889-902.

Jenkinson, C., & Layte, R. (1997). Development and testing of the UK SF-12 (short form health survey). *Journal of Health Services & Research Policy*, *2*, 14-18.

Jenkinson, C., Chandola, T., Coulter, A., & Bruster, S. (2001). An assessment of the construct validity of the SF-12 summary scores across ethnic groups. *Journal of Public Health Medicine*, *23*, 187-194.

Jenkinson, C., Layte, R., Jenkinson, D., Lawrence, K., Petersen, S. Paice, et al. (1997). A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *Journal of Public Health Medicine*, *19*, 179-186.

Johnson, J. A., & Coons, S. J. (1998). Comparison of the EQ-5D and SF-12 in an adult US sample, *Quality of Life Research*, *7*, 155-166.

Jones, D., Kazis, L., Lee, A., Rogers, W., Skinner, K., Cassar, L. et al. (2001). Health status assessments using the Veterans SF-12 and SF-36: methods for evaluating outcomes in the Veterans Health Administration. *Journal of Ambulatory Care Management*, *24*, 68-86.

Larson, C. O. (2002). Use of the SF-12 instrument for measuring the health of homeless persons. *Health Services Research*, 37, 733-750.

Lenert, L. A. (2000). The reliability and internal consistency of an internet-capable computer program for measuring utilities. *Quality of Life Research*, *9*, 811-817.

Lenert, L. A., Sherbourne, C. D., Sugar, C., & Wells, K. B. (2000). Estimation of utilities for the effects of depression from the SF-12. *Medical Care*, *38*, 763-770.

Lim, L. L., & Fisher, J. D. (1999). Use of the 12-item short-form (SF-12) Health Survey in an Australian heart and stroke population, *Quality of Life Research*, *8*, 1-8.

Lundberg, L., Johannesson, M., Isacson, D. G., & Borgquist, L. (1999). The relationship between health-state utilities and the SF-12 in a general population. *Medical Decision Making*, *19*, 128-140.

Macran, S., Weatherly, H., & Kind, P. (2003). Measuring population health: a comparison of three generic health status measures. *Medical Care*, *41*, 218-231.

McBurney, C. R., Eagle, K. A., Kline-Rogers, E. M., Cooper, J. V., Mani, O. C., Smith, D. E. et al. (2002). Health-related quality of life in patients 7 months after a myocardial infarction: factors affecting the Short Form-12. *Pharmacotherapy*, *22,* 1616-1622.

Millard, R. W., & Carver, J. R. (1999). Cross-sectional comparison of live and interactive voice recognition administration of the SF-12 health status survey. *American Journal of Managed Care*, *5*, 153-159.

Resnick, B., & Nahm, E. S. (2001). Reliability and validity testing of the revised 12-item Short-Form Health Survey in older adults. *Journal of Nursing Management*, *9,* 151-161.

Resnick, B., & Parker, R. (2001). Simplified scoring and psychometrics of the revised 12-item Short-Form Health Survey. *Outcomes Management for Nursing Practice*, *5*, 161-166.

Riddle, D. L., Lee, K. T., & Stratford, P. W. (2001). Use of SF-36 and SF-12 health status measures: A quantitative comparison for groups versus individual patients. *Medical Care*, *39*, 867-878.

Rubenach, S., Shadbolt, B., McCallum, J., & Nakamura, T. (2002). Assessing health-related quality of life following myocardial infarction: is the SF-12 useful? *Journal of Clinical Epidemiology, 55*, 306-309.

Salyers, M. P., Bosworth, H. B., Swanson, J. W., Lamb-Pagone, J., & Osher, F. C. (2000). Reliability and validity of the SF-12 health survey among people with severe mental illness. *Medical Care, 38*, 1141-1150.

Sanderson, K., & Andrews, G. (2002). Prevalence and severity of mental health related disability and relationship to diagnosis. *Psychiatric Services*, *53*, 80-86.

Sanderson, K., & Andrews, G. (2002). The SF-12 in the Australian population: cross-validation of item selection. *Australian & New Zealand Journal of Public Health*, *26*, 343-345.

Sanderson, K., Andrews, G., & Jelsma, W. (2001). Disability measurement in the anxiety disorders: comparison of three brief measures. *Journal of Anxiety Disorders*, *15*, 333-344.

Siddique, R., Ricci, J. A., Stewart, W. F., Sloan, S., & Farup, C. E. (2002). Quality of life in a US national sample of adults with diabetes and motility-related upper gastrointestinal symptoms. *Digestive Diseases & Sciences*, *47*, 683-689.

Sugar, C. A., Sturm, R., Lee, T. T., Sherbourne, C. D., Olshen, R. A., & Wells, K. B. et al. (2001). Empirically defined health states for depression from the SF-12. *Health Services Research*, *33*, 911-928.

Taylor, A. W., Wilson, D. H., Dal Grande, E., Ben-Tovim, D., Elzinga, R. H., Goldner, R. D. et al. (2000). Mental health status of the South Australian population. *Australian & New Zealand Journal of Public Health*, *24*, 29-34.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1996). A 12 Item Short Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34,* 220-233.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. *SF 36 Physical and Mental Health Summary Scales: A User's Manual.* MA: The Health Institute, New England Medical Center, 1994.

Ware, J. E., Jr., Kosinski, M., Bayliss, M. S., McHorney, C. A., Rogers, W. H., & Raczek, A. (1995). Comparison of methods for the scoring and statistical analysis of SF 36 Health Profile and Summary Measures: Summary of results from the Medical Outcomes Study. *Medical Care* 1995, *33*, AS264-AS279.

Wilson, D., Tucker, G., & Chittleborough, C. (2002). Rethinking and rescoring the SF-12. *Sozial- und Praventivmedizin*, *47*, 172-177.

Reporter: Nicholas Marosszeky, Research Psychologist

Date of report: 30 June 2003

SF-36® Health Survey (Version 1.0)

Title:	SF-36® Health Survey (Version 1.0)
	for use in Australia
	(also known as the Medical Outcomes Study (MOS) 36-Item Short Form Health Survey (SF-36)).
Abbreviations:	SF-36.
Author(s) Name:	John E. Ware, Jr.
Author(s) Address:	QualityMetric Incorporated 640 George Washington Highway Lincoln, RI 02865 USA www.qualitymetric.com
Supplied by:	QualityMetric Incorporated 640 George Washington Highway Lincoln, RI 02865 USA
	In Australia, SF-36® Health Survey manuals can be obtained from the: Australian Health Outcomes Collaboration (AHOC) c/- Centre for Health Service Development University of Wollongong NSW 2522 Phone: 02 4221-4411.
Cost:	An annual license fee applies for the use of the SF-36® Health Survey.
	Survey users are required to register with QualityMetric Incorporated and obtain a quote for the annual license fee that applies to their project.
	The license charge will depend upon whether users require a commercial or research license.
	Register online at http://www.qualitymetric.com/. Information of the SF group of instruments can also be found at http://www.sf-36.com/.
	SF-36 manuals can be purchased in Australia from AHOC by contacting Laura Willmott at willmott@uow.edu.au or by telephone on 02 4221-4411.
	For technical questions about using the SF-36® Health Survey in Australia (including latest developments and research advice) contact Jan Sansoni at jansan@netspeed.com.au or by telephone on 02 6291-7271 or 02 6205-0869.
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Basic training in survey administration and the characteristics of the SF-36® Health Survey is required by those without these qualifications. The AHOC provides training workshops for the SF-36 and other instruments.
Purpose:	The SF-36® Health Survey is a generic outcome measure designed to examine a person's perceived health status.
Administration Time:	5 – 10 minutes.
Instrument Type:	Self-report Questionnaire.
Structure:	The SF-36® Health Survey includes one multi-item scale measuring each of the following eight health concepts: (1) physical functioning; (2) role limitations because of physical health problems; (3) bodily pain; (4) social functioning;

(5) general mental health (psychological distress and psychological well-being);

(6) role limitations because of emotional problems;

(7) vitality (energy/fatigue); and

(8) general health perceptions.

The SF-36 also includes a single-item measure of health transition or change. The SF-36 can also be divided into two aggregate summary measures the Physical Component Summary (PCS) and the Mental Component Summary (MCS). (In the standard version of the SF-36 all scale questions refer to a 4 week time period.)

Scoring:

The SF-36® Health Survey items and scales were constructed using the Likert method of summated ratings. Answers to each question are scored (some items need to be recoded). These scores are then summed to produce raw scale scores for each health concept which are then transformed to a 0 – 100 scale. Scoring algorithms can then be applied to produce the PCS and MCS scores. (These two summary scores have the major advantage of being norm based. They also have reduced floor and ceiling effects.)

Developed for:

The SF-36® Health Survey developed out of work on the Medical Outcomes Study or RAND Health Insurance Experiment. It is a short-form derived from a larger 149-item instrument and is more precise than its predecessor the SF-20.

Normative Data:

Australian data for the SF-36® Version 1 Health Survey is provided by Stevenson (1996) and from the Australian Bureau of Statistics (1997), National Health Survey (1995). (These are the accepted norms for use in Australia.) Additional population health data using the SF-36 can be found in the 1996 Australian Longitudinal Study on Women´s Health, the 1999–2000 Australian Diabetes, Obesity and Lifestyle Study, the 1998 National Drug Strategy Household Survey, the 1991–2002 SA Health Omnibus Survey, the 1997 & 1998 NSW Health Survey, the 1999 NSW Older People´s Health Survey and the 2002 Australian National Sexuality Survey. Version 2 of the SF-36® has been included in the recent South Australian Health Omnibus Survey (Harrison Health Research, 2004) and a forthcoming report by Hawthorne will examine Australian data and the psychometric properties of this instrument.

US Data for the SF-36 can be found in Ware, Kosinski & Keller (1994) and Ware, Kosinski, Bayliss, McHorney, Rogers & Raczek (1995).

UK Data for the SF-36 can be found at Jenkinson, Coulter & Wright (1993) and Bowling, Bond, Jenkinson & Lamping (1999).

World Data for the SF-36 in order to make cross country comparisons can be found at Ware, Gandek, Kosinski, Aaronson, Apolone, Brazier et al. (1998).

Clinical Data:

A few clinical studies are listed below:

Arthritis: Hill, Parsons, Taylor & Leach (1999).

Asthma: Adams, Wakefield, Wilson, Parsons, Campbell, Smith et al. (2001).

Cardiac Rehabilitation: Jette & Downing (1994).

Chronic Fatigue Syndrome: Komaroff, Fagioli, Doolittle, Gandek, Gleit, Guerriero, et al. (1996).

Geriatrics: Weinberger, Nagle, Hanlon, Samsa, Schmader, Landsman et al. (1994) and Cohen, Feussner, Weinberger, Carnes, Hamdy, Hsieh et al. (2002).

Hip or Knee Replacement for osteoarthritis: March, Gross, Lapsley, Brnabic, Tribe, Bachmeier et al. (1999).

Kidney Diseases: Kurtin, Davies, Meyer, DeGiacomo & Kantz (1992) and Chow, Briganti, Kerr, Chadban, Zimmet & Atkins (2003).

Knee Replacement: Bombardier, Melfi, Paul, Green, Hawker, Wright et al. (1995).

Mental Health: Sherbourne, Wells & Ludd (1996) and Goldney, Fisher, Wilson & Cheok (2001).

Methadone Maintenance Treatment: Ryan & White (1996).

Sexual Health: Patel, Boselli, Cairo, Barnett, Price & Wulf (2001).

Sleep Problems: Manocchia, Keller & Ware (2001).

Stroke: Anderson, Laubscher & Burns (1996), Anderson, Rubenach, Mhurchu, Clark, Spencer & Winsor (2000) and Middleton, Donnelly, Harris, Lusby & Ward (2002).

The ACT Care Continuum and Health Outcomes Project is a useful source of Australian clinical data for hospitalised patients.

Applications:	Outcome studies using the SF-36® Health Survey are not only restricted to the Doctor´s waiting room, but can be also administered via mail-out survey or telephone interview. The SF-36 can also be used in a computerised format.

Interpretation guidelines and cautions are also available. One clear recommendation from the literature is that SF-36 Summary Scores (PCS + MCS) should be compared with the eight SF-36 Scale Scores before interpretation.

An acute (1 week) version of the SF-36® Health Survey is also available.

Rasch Analysis, a form of Item Response Theory, has also been applied to the SF-36 10-item Physical Functioning Scale (PF-10) with good result and future applications (especially for the use of computerised adaptive testing with patients).

Recently QualityMetric Incorporated has developed an improved version of the SF-36® Health Survey known as the SF-36v2™ Health Survey (Version 2). This new version of the SF-36 has refinements to layout, item wording and response categories, as well as norm based scoring for all of the eight SF-36 health concept scales (not just for the summary scores: PCS + MCS). The SF-36 Version 2 also uses new norms – 1998 general US population. Interim norms will shortly be available for this instrument from the 2004 South Australian Health Omnibus Survey and those interested should contact Dr Graeme Hawthorne at graemeeh@unimelb.edu.au or by telephone on 03 9496-4031.

See also the Instrument Review of the SF-12® Health Survey.

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	McHorney et al. (1994)\n\nWare, Kosinski & Keller (1994)\n\nMcCallum (1995)\n\nStevenson (1996)\n\nGandek et al. (1998)\n\nSanson-Fisher & Perkins (1998)	Good	
Test – retest	Yes	Ware, Kosinski & Keller (1994)\n\nBowling (1995)\n\nSanson-Fisher & Perkins (1998)\n\nKagee (2001)	Adequate	More information could be published on this aspect of the SF-36´s reliability.\n\n(Cronbach´s Alpha is used to construct the SEM for the SF-36 Summary scores. Cronbach´s Alpha: PCS = 0.92; MCS = 0.91)
Inter – rater	NA	NA	NA	The SF-36 is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Komaroff et al. (1996) Shadbolt, McCallum & Singh (1997) Kagee (2001)	Adequate	See also the references in the Construct Validity section.
Correlation with other measures	Yes	Beaton, Hogg-Johnson & Bombardier (1997) Prieto et al. (1997) Essink-Bot et al. (1997) Kagee (2001)	Good	
Construct	Yes	Tarlov et al. (1989) McHorney et al. (1992) McHorney, Ware & Raczek (1993) McHorney et al. (1994) Ware et al. (1995) Keller et al. (1998) Ware et al. (1998) Jenkinson (1999)	Good	
Criterion	Yes	Jenkinson, Wright & Coulter (1994) Kagee (2001)	Adequate	

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Jenkinson, Peto & Coulter (1994) Jenkinson et al. (1995) Jenkinson et al. (1997) Sharples et al. (2000) Ferguson, Robinson & Splaine (2002)	Adequate	

Cultural Applicability and Cultural Adaptations:	The SF-36® Health Survey has been translated into many languages and its content examined cross culturally. In Australia, the SF-36 has been utilised for people from a non-English speaking background in Western Sydney and a large group of new Vietnamese migrants. However, limited research has been reported with Aboriginal and Torres Strait Islanders. (A recent paper by Scott, Sarfali, Tobias & Haslett [2000] may provide a useful template for future work in this area.)
Gender Appropriateness:	Normative data is available for males and females.
Age Appropriateness:	14 years and over.
Summary:	The SF-36 is a highly recommended measure with superior psychometric properties. It has been used extensively in Australia for both population health and clinical research.

References

Adams, R., Wakefield, M., Wilson, D., Parsons, J., Campbell, D., Smith, B., & Ruffin, R. (2001). Quality of life in asthma: a comparison of community and hospital asthma patients. *Journal of Asthma*, *38*, 205-214.

Anderson, C., Laubscher, S., & Burns, R. (1996). Validation of the Short Form 36 (SF 36) Health Survey Questionnaire among stroke patients. *Stroke*, *27*, 1812-1816.

Anderson, C., Rubenach, S., Mhurchu, C. N., Clark, M., Spencer, C., & Winsor, A. (2000). Home or hospital for stroke rehabilitation? Results of a randomized controlled trial: 1. Health outcomes at 6 months. *Stroke*, 31, 1024-1031.

Beaton, D. E., Hogg-Johnson, S., & Bombardier, C. (1997). Evaluating changes in health status: Reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *Journal of Clinical Epidemiology*, 50, 73-93.

Bombardier, C., Melfi, C. A., Paul, J., Green, R., Hawker, G., Wright, J., et al. (1995). Comparison of a generic and a disease specific measure of pain and physical function after knee replacement surgery. *Medical Care*, *4*, AS131-AS144.

Bowling, A., Bond, M., Jenkinson, C., & Lamping, D. L. (1999). Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *Journal of Public Health Medicine*, *21*, 255-270.

Bowling, A. (1995). *Measuring Disease: A Review of Disease Specific Quality of Life Measurement Scales*. Buckingham: Open University Press.

Bullinger, M., Alonso, J., Apolone, G., Leplege, A., Sullivan, M., Wood-Dauphinee, S. et al. (1998). Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51*, 913-923.

Cardona, M., Jorm, L., Williamson, M., & Chey, T. (1995). *The Blacktown Health Survey of People from Non English Speaking Background.* Western Sector Public Health Unit.

Chow, F. Y., Briganti, E. M., Kerr, P. G., Chadban, S. J., Zimmet, P. Z., & Atkins, R. C. (2003). Health-related quality of life in Australian adults with renal insufficiency: a population-based study. *American Journal of Kidney Diseases*, *41,* 596-604.

Cohen, H. J., Feussner, J. R., Weinberger, M., Carnes, M., Hamdy, R. C., Hsieh, F. et al. (2002). A controlled trial of inpatient and outpatient geriatric evaluation and management. *New England Journal of Medicine*, *346*, 905-912.

Diehr, P., Patrick, D. L., Spertus, J., Kiefe, C. I., McDonell, M., & Fihn, S. D. (2001). Transforming self-rated health and the SF-36 scales to include death and improve interpretability. *Medical Care*, *39*, 670-680.

Essink-Bot, M. L., Krabbe, P. F. M., Bonsel, G. J., & Aaronson, N. K. (1997). An empirical comparison of four generic health status measures: The Nottingham Health Profile, the Medical Outcomes Study 36 Item Short Form, the COOP/WONCA Charts, and the EuroQol Instrument. *Medical Care*, *35*, 522-537.

Ferguson, R. J., Robinson, A. B., & Splaine, M. (2002). Use of the Reliable Change Index to evaluate clinical significance in SF-36 outcomes. *Quality of Life Research*, *11*, 509-516.

Gandek, B., Ware, J. E., Jr., Aaronson, N. K., Alonso, J., Apolone, G., & Bjorner, J. et al. (1998). Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51,* 1149-1158.

Goldney, R. D., Fisher, L. J., Wilson, D. H., & Cheok, F. (2001). Suicidal ideation and health related quality of life in the community. *Medical Journal of Australia*, *175*, 546-549.

Hill, C. L., Parsons, J., Taylor, A., & Leach, G. (1999). Health related quality of life in a population sample with arthritis. *Journal of Rheumatology*, *26*, 2029-2035.

Jenkinson, C., Coulter, A., & Wright, L. (1993). Short form 36 (SF 36) health survey questionnaire: Normative data for adults of working age. *British Medical Journal, 306*, 1437-1440.

Jenkinson, C., Fitzpatrick, R., Garratt, A., Peto, V., & Stewart-Brown, S. (2001). Can item response theory reduce patient burden when measuring health status in neurological disorders? Results from Rasch analysis of the SF-36 physical functioning scale (PF-10). *Journal of Neurology, Neurosurgery and Psychiatry*, *71*, 220-224.

Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S., & Layte, R. (1997). Evaluation of index and profile measures of health status in a randomized controlled trial. Comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and disease specific measures. *Medical Care*, *35*, 1109-1118.

Jenkinson, C., Lawrence, K., McWhinnie, D., & Gordon, J. (1995). Sensitivity to change of health status measures in a randomized controlled trial: Comparison of the COOP charts and the SF-36. *Quality of Life Research*, *4*, 47-52.

Jenkinson, C., Peto, V., & Coulter, A. (1994). Measuring change over time: a comparison of results from a global single item of health status and the multi dimensional SF 36 health status survey questionnaire in patients presenting with menorrhagia. *Quality of Life Research*, *3*, 317-321.

Jenkinson, C., Stewart-Brown, S., Petersen, S., & Paice, C. (1999). Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiology and Community Health*, *53*, 46-50.

Jenkinson, C., Wright, L., Coulter, A. 1994. Criterion validity and reliability of the SF 36 in a population sample. *Quality of Life Research*, *3*, 7-12.

Jenkinson, C. (1999). Comparison of UK and US methods for weighting and scoring the SF-36 summary measures. *Journal of Public Health Medicine. 21*, 372-376.

Jenkinson, C. (1998). The SF-36 physical and mental summary measures: an example of how to interpret scores. *Journal of Health Services & Research Policy*, *3*, 92-96.

Jette, D. U., & Downing, J. (1994). Health status of individuals entering a cardiac rehabilitation program as measured by the Medical Outcomes Study 36 item Short Form survey (SF 36). *Physical Therapy*, *74*, 521-527.

Kagee, A. (2001). Review of the SF-36 Health Survey, in Plake BS & Impara, JC. (Eds). *The Fourteenth Mental Measurements Yearbook.* Lincoln NE, Buros Institute of Mental Measurements.

Keller, S. D., Ware, J. E., Jr., Bentler, P. M., Aaronson, N. K., Alonso, J., Apolone, G. et al. (1998). Use of structural equation modelling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51*, 1179-1188.

Komaroff, A. L., Fagioli, L. R., Doolittle, T. H., Gandek, B., Gleit, M. A., Guerriero, R. T., et al. (1996). Health status in patients with chronic fatigue syndrome and in general population and disease comparison groups. *American Journal of Medicine*, *101*, 281-290.

Kurtin, P. S., Davies, A. R., Meyer, K. B., DeGiacomo, J. M., & Kantz, M. E. (1992). Patient based health status measures in outpatient dialysis: Early experiences in developing an outcomes assessment program. *Medical Care*, *30*, MS136-MS149.

Manocchia, M., Keller, S., & Ware, J. E. (2001). Sleep problems, health-related quality of life, work functioning and health care utilization among the chronically ill. *Quality of Life Research*, *10*, 331-345.

March, L. M., Cross, M. J., Lapsley, H., Brnabic, A. J., Tribe, K. L., Bachmeirer, C. J. et al. (1999). Outcomes after hip or knee surgery for osteoarthritis. A prospective cohort study comparing patients´ quality of life before and after surgery with age-related population norms. *Medical Journal of Australia*, *171*, 235-238.

McCallum, J. (1995). The SF 36 in an Australian sample: validating a new, generic health status measure. *Australian Journal of Public Health*, *19*, 160-166.

McHorney, C. A., Haley, S. M., & Ware, J. E., Jr., (1997). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, *50*, 451-461.

McHorney, C. A., Kosinski, M., & Ware, J. E., Jr. (1994). Comparisons of the costs and quality of norms for the SF 36 Health Survey collected by mail versus telephone interview: Results from a national survey. *Medical Care*, *32*, 551-567.

McHorney, C. A., Ware, J. E., Jr., Lu, J. F. R., & Sherbourne, C. D. (1994). The MOS 36 Item Short Form Health Survey (SF 36): 3. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*, *32*, 40-66.

McHorney, C. A., Ware, J. E., Jr., & Raczek, A. E. The MOS 36 Item Short Form Health Survey (SF36): 2. Psychometric and clinical tests of validity measuring physical and mental health constructs. *Medical Care*, *31*, 247-263.

McHorney, C. A., Ware, J. E., Jr., Rogers, W., Raczek, A. E., Lu, J. F. R. (1992). The validity and relative precision of MOS Short and Long Form Health Status Scales and Dartmouth COOP Charts. *Medical Care*, *30*, MS253-MS265.

Middleton, S., Donnelly, N., Harris, J., Lusby, R., & Ward, J. (2002). Audit of long-term mortality and morbidity outcomes for carotid endarterectomy. *Australian Health Review*, *25*, 81-91.

Patel, R., Boselli, F., Cairo, I., Barnett, G., Price, M., & Wulf, H. C. (2001). Patient's perspectives on the burden of recurrent genital herpes. *International Journal of STD and AIDS*, *12*, 640-645.

Perkins, J. J., & Sanson-Fisher, R. W. (1998). An examination of self- and telephone-administered modes of administration for the Australian SF-36. *Journal of Clinical Epidemiology*, 51, 969-973.

Perneger, T. V., Lepledge, A., & Etter, J. F. (1999). Cross-cultural adaptation of a psychometric instrument: two methods compared. *Journal of Clinical Epidemiology*, *52*, 1037-1046.

Prieto, L., Alonson, J., Ferrer, M., & Antò, J. M. (1997). Are results of the SF 36 Health Survey and the Nottingham Health Profile similar?: A comparison in COPD patients. *Journal of Clinical Epidemiology*, *50*, 463-473.

Raczek, A. E., Ware, J. E., Bjorner, J. B., Gandek, B., Haley, S. M., Aaronson, N. K. et al. (1998). Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51*, 1203-1214.

Ryan, C. F., & White, J. M. (1996). Health status at entry to methadone maintenance treatment using the SF-36 health survey questionnaire. *Addiction*, *91*, 39-45.

Ryan, J. M., Corry, J. R., Attewell, R., & Smithson, M. J. (2002). A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Quality of Life Research*, *11*, 19-26.

Sanson-Fisher, R. W., & Perkins, J. J. (1998). Adaptation and validation of the SF-36 Health Survey for use in Australia. *Journal of Clinical Epidemiology*, 51, 961-967.

Scott, K. M., Sarfali, D., Tobias, M. I., Martin, I., Haslett, S. J. (2000). A challenge to the cross-cultural validity of the SF-36 health survey: Factor structure in Maori, Pacific and New Zealand ethnic groups. *Social Science & Medicine*, *51*, 655-1664.

Shadbolt, B., McCallum, J., & Singh, M. (1997). Health outcomes by self-report: validity of the SF-36 among Australian hospital patients. *Quality of Life Research*, 6, 343-352.

Sharples, L. D., Todd, C. J., Caine, N., & Tait, S. (2000). Measurement properties of the Nottingham Health Profile and Short Form 36 Health status measures in a population sample of elderly people living at home. Results from ELPHS. *British Journal of Health Psychology*, *5*, 217-233.

Sherbourne, C. D., Wells, K. B., & Ludd, L. L. (1996). Functioning and well being of patients with panic disorder. *American Journal of Psychiatry*, *153*, 213-218.

Stevenson, C. E. (1996). SF 36: *Interim norms for Australian data.* Canberra: Australian Institute of Health and Welfare.

Tarlov, A. R., Ware, J. E., Jr., Greenfield, S., Nelson, E. C., Perrin, E., & Zubkoff, M. (1989). The Medical Outcomes Study: An application of methods for monitoring the results of medical care. *Journal of the American Medical Association (JAMA)*, *262*, 925-930.

Wagner, A. K., Gandek, B., Aaronson, N. K., Acquadro, C., Alonso, J., Apolone, G. et al. (1998). Cross-cultural comparisons of the content of SF-36 translations across 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51*, 925-932.

Ware, J. E., Jr., Bjorner, J. B., & Kosinski, M. (2000). Response to Hays et al. and McHorney and Cohen: Practical implications of Item Response Theory and computerized Adaptive Testing: A brief summary of ongoing studies of widely used Headache Impact Scales. *Medical Care*, *38* (Supp 2), 73-82.

Ware, J. E., Jr., Gandek, B., Kosinski, M., Aaronson, N. K., Apolone, G., Brazier, J. et al. (1998). The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, *51*, 1167-1170.

Ware, J. E., Jr., & Hays, R. D. (1988). Methods for measuring patient satisfaction with specific medical encounters. *Medical Care*, *26*, 393-402.

Ware, J. E., Jr., Kosinski, M., Bayliss, M. S., McHorney, C. A., Rogers, W. H., & Raczek, A. (1995). Comparison of methods for the scoring and statistical analysis of SF 36 Health Profile and Summary Measures: Summary of results from the Medical Outcomes Study. *Medical Care*, *33*, AS264-AS279.

Ware, J. E., Jr., Kosinski, M., Gandek, B., Aaronson, N. K., Apolone, G., Bech, P. et al. (1998). The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology*, 51, 159-1165.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1994). *SF 36 Physical and Mental Health Summary Scales: A User's Manual*. MA: The Health Institute, New England Medical Center.

Ware, J. E., & Kosinski, M. (2001). Interpreting SF-36 summary health measures: a response. *Quality of Life Research*, *10*, 405-413.

Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36 Item Short Form Health Survey (SF 36). Conceptual framework and item selection. *Medical Care*, *30*, 473-483.

Watkins, R. E., Plant, A. J., Sang, D., O´Rourke, T., & Gushulak, B. (2000). Development of a Vietnamese version of the Short form-36 Health Survey. Asia-Pacific *Journal of Public Health*, *12*, 118-123.

Watson, E. K., Firman, D. W, Baade, P. D., & Ring, I. (1996). Telephone administration of the SF-36 Health Survey: validation studies and population norms for adults in Queensland. *Australian & New Zealand Journal of Public Health*, *20*, 359-363.

Weinberger, M., Nagle, B., Hanlon, J. T., Samsa, G. P., Schmader, K., Landsman, P. B., et al. Assessing health related quality of life in elderly outpatients: Telephone versus face to face administration. *Journal of the American Geriatrics Society*, *42*, 1295-1299.

Wetzler, H. P, Lum, D. L., & Bush, D. M. (2000). Using the SF-36 health survey in primary care. In Maruish ME. (Editor) *Handbook of Psychological Assessment in Primary Care Settings.* Mahwah NJ: Lawrence Erlbaum Associates.

Wilson, D., Parsons, J., & Tucker, G. (2000). The SF-36 summary scales: problems and solutions. *Sozial- und Praventivmedizin*, *45*, 239-246.

Reporter: Nicholas Marosszeky, Research Psychologist

Date of report: 30 June 2003

World Health Organisation Quality of Life-100

Title:	World Health Organisation Quality of Life-100. *(Australian English Language Version)*
Abbreviations:	WHOQOL-100
Author(s) Name:	The WHOQOL Group, Division of Mental Health, World Health Organisation (WHO).
Author(s) Address:	WHOQOL Group Division of Mental Health World Health Organisation CH-1211 Geneva 27 SWITZERLAND
Supplied by:	Australian WHOQOL Field Centre Australian Centre for Posttraumatic Mental Health Austin and Repatriation Medical Centre PO Box 5444 Heidelberg West VIC 3081 www.acpmh.unimelb.edu.au/whoqol/ *(For more information about the WHOQOL-100 contact A/Professor Graeme Hawthorne at graemeeh@unimelb.edu.au or by telephone on 03 9496-4031).*
Cost:	Free.
	Users must register with the Melbourne WHOQOL Field Study Centre. (Users are not authorised to make any modifications to the instrument without permission.)
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Basic training in survey administration and the characteristics of the WHOQoL-100 is required by those without these qualifications.
Purpose:	The WHOQOL-100 is a quality of life (QOL) assessment instrument, which is comparable across cultures and languages.
Administration Time:	20 – 30 minutes.
Instrument Type:	Self-report Questionnaire.
Structure:	The WHOQOL-100 includes 100 items covering six domains: Physical, Psychological, Independence, Social relationships, Environment and Spiritual (e.g. Spirituality/religion/personal beliefs). Each domain is made up of a number of facets or issues. These facets are listed below:

Physical
  Pain and discomfort
  Energy and fatigue
  Sleep and rest

Psychological
  Positive affect
  Thinking, learning, memory and concentration
  Self-esteem
  Body image and appearance
  Negative affect

Independence
  Mobility
  Activities of daily living
  Dependence on medication or treatments
  Working capacity

Social relationships
    Personal relationships
    Social support
    Sexual activity

Environment
    Physical safety and security
    Home environment
    Financial resources
    Health and social care: accessibility and quality
    Opportunities for acquiring new information and skills
    Participation in and opportunities for recreation/leisure activities
    Physical environment (pollution, noise, traffic, climate)
    Transportation

Spiritual
    Spirituality/religion/personal beliefs

An additional four items are included on Overall Quality of life and General Health.

Scoring:	Each of the 24 facets is made up of 4 items. Each item is presented in a question format and is answered on a five-point response scale. Facet items are then summed to produce a Facet score (some items needing to be reversed scored). Domain scores are then calculated from the facet scores (some facet scores need to be reversed scored). Scoring algorithms are provided. All scores (facet and domain) range from 4 - 20. A higher score indicates better QOL.
Developed for:	The World Health Organisation (WHO) developed the WHOQOL-100 using a collaborative project involving 15 centres from around the world (Australia, Croatia, France, India (x2), Israel, Japan, The Netherlands, Panama, Russian Federation, Spain, Thailand, United Kingdom, United States of America, and Zimbabwe). Development of the items involved focus groups, forward and backward translations and linguistic review at each stage. A global item pool was produced and 236 common items were selected for pilot testing in each country ($n = 300$). These common items were further reduced to 100 using standard psychometric techniques. The five-point response scale descriptors for each question were also adapted for each language. This effort produced a set of QOL questions addressing common facets or issues across cultures with slightly different item wording for each language.
Normative Data:	Australian Data ($n = 300$) for the WHOQOL-100 is provided in the User's Manual and Interpretation Guide.
	International Data ($n = 4802$) for the WHOQOL-100 is provided in WHOQOL Group (1998).
Clinical Data:	A few clinical studies are listed below:
	Bone Marrow Transplantation: De Souza, Duraes, Vigorito, Penteado Aranha, Oliveira, De Brito Eid et al. (2002).
	Cancer: Struttman, Fabro, Romieu, de Roquefeuil, Touchon, Dandekar et al. (1999).
	Chronic Fatigue Syndrome: van Heck & de Vries (2002).
	Chronic Pain: Skevington (1998) and Skevington, Carse & de C. Williams (2001).
	Depression: Skevington & Wright (2001) and Angermeyer, Holzinger, Matschinger & Stenger-Wenzke (2002).
	Diabetes: Pibernik-Okanovic, Szabo & Metelko (1998).
	Epilepsy: Amir, Roziner, Knoll & Neufeld (1999).
	Liver Transplantation: O'Carroll, Cossar, Couston & Hayes (2000).

Lung Disease: Wirnsberger, de Vries, Breteler, van Heck, Wouters, Drent (1998).

Senile Dementia: Struttman, Fabro, Romieu, de Roquefeuil, Touchon, Dandekar et al. (1999).

Applications: It should be noted that the WHOQOL instruments have been designed for population research or clinical trials. They are not to be used for individual assessment.

The WHOQOL-100 can also be administered by an interviewer.

A set of 31 optional "importance questions" is included in the User's manual and interpretation guide. These questions were designed to weight a person's responses to the WHOQOL-100. However a standardised method for manipulating these weightings with WHOQOL-100 scores has yet to be devised.

Skevington, Bradshaw & Saxena (1999) researching the WHOQOL have outlined a method to develop national quality of life items in addition to the core international set.

See also the Instrument Review on the WHOQOL-BREF.

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	WHOQOL Group (1998) Skevington (1999) Bonomi et al. (2000)	Adequate	
Test – retest	Yes	Bonomi et al. (2000)	Adequate	
Inter – rater	NA	NA	NA	The WHOQOL-100 is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	WHOQOL Group (1998) Skevington (1999)	Adequate	
Correlation with other measures	Yes	Bonomi et al. (2000) Skevington et al. (2001)	Adequate	At this time, there is no generally accepted measure (i.e. gold standard) for the construct of Quality of Life in the scientific literature.
Construct	Yes	Power et al. (1999) Bonomi et al. (2000)	Adequate	
Criterion	Yes	Bonomi et al. (2000)	Adequate	

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Bonomi et al. (2000) O'Carroll et al. (2000) Skevington & Wright (2001)	Adequate	

Cultural Applicability and Cultural Adaptations: The WHOQOL-100 is currently available in over 20 language versions A Chinese-Australian version of the WHOQOL-100 has been developed. Research using the WHOQOL with people from the Aboriginal and Torres Strait Islanders population has yet to be undertaken.

Gender Appropriateness: Normative data for males and females is not available at this time.

Age Appropriateness: Adults.

Summary: The WHOQOL-100 is a recommended measure of quality of life (QOL). It has undergone an extensive process of translation and development in order to make it comparable across languages and cultures. However further field-testing and normative data is required for the Australian population.

References

Amir, M., Roziner, I., Knoll, A., & Neufeld, M. Y. (1999). Self-efficacy and social support as mediators in the relationship between disease severity and quality of life in patients with epilepsy. *Epilepsia*, *40*, 216-224.

Angermeyer, M. C., Holzinger, A., Matschinger, H., & Stenger-Wenzke, K. (2002). Depression and quality of life: Results of a follow-up study. *International Journal of Social Psychiatry*, *48*, 189-199.

Bonomi, A. E., Patrick, D. L., Bushnell, D. M., & Martin, M. (2000). Validation of the United States' version of the World Health Organisation Quality of Life (WHOQOL) Instrument. *Journal of Clinical Epidemiology*, *53*, 1-12.

De Souza, C. A., Duraes, M. I., Vigorito, A. C., Penteado Aranha, F. J, Oliveira, G. B., De Brito, Eid, K. A. et al. (2002). Quality of Life in patients randomized to receive a bone marrow or a peripheral blood allograft. *Haematologica*, *87*, 1281-1285.

Murphy, B., Herrman, H., Hawthorne, G., Pinzone, T., & Evert, H. (2000). *Australian WHOQOL Instruments: User's Manual and Interpretation Guide.* Melbourne: Australian WHOQOL Field Study Centre.

O'Carroll, R. E., Cossar, J. A., Couston, M. C., Hayes, & P. C. (2000). Sensitivity to change following liver transplantation: A comparison of three instruments that measure quality of life. *Journal of Health Psychology*, *5*, 69-74.

Pibernik-Okanovic, M., Szabo, S., & Metelko, Z. (1998). Quality of life following a change in therapy for diabetes mellitus. *Pharmacoeconomics*, *14*, 201-207.

Power, M., Harper, A., & Bullinger, M. (1999). The World Health Organisation WHOQOL-100: tests of the universality of Quality of Life in 15 different cultural groups worldwide. *Health Psychology*, *18*, 495-505.

Skevington, S. M., Bradshaw, J., & Saxena, S. (1999). Selecting national items for the WHOQOL: conceptual and psychometric considerations. *Social Science & Medicine*, *48*, 473-487.

Skevington, S. M., Carse, M. S., de C. Williams, A. C. (2001). Validation of the WHOQOL-100: Pain management improves quality of life for chronic pain patients. *Clinical Journal of Chronic Pain*, *17*, 264-275.

Skevington, S. M., & Wright, A. (2001). Changes in the quality of life of patients receiving antidepressant medication in primary care: validation of the WHOQOL-100. *British Journal of Psychiatry*, *178*, 261-267.

Skevington, S. M. (1998). Investigating the relationship between pain and discomfort and quality of life, using the WHOQOL. *Pain*, *76*, 395-406.

Skevington, S. M. (1999). Measuring quality of life in Britain: Introducing the WHOQOL-100. *Journal of Psychosomatic Research*, *47*, 449-459.

Struttman, T., Fabro, M., Romieu, G., de Roquefeuil, G., Touchon, J., Dandekar, T. et al. (1999). Quality-of-Life assessment in the old using the WHOQOL-100: Differences between patients with senile dementia and patients with cancer. *International Psychogeriatrics*, *11*, 273-279.

Szabo, S., Orley, J., & Saxena, S. (1997). An approach to response scale development for cross-cultural questionnaires. *European Psychologist*, *2*, 270-276.

van Heck, G. L., & de Vries, J. (2002). Quality of life of patients with chronic fatigue syndrome. *Journal of Chronic Fatigue Syndrome*, *10*, 17-35.

WHOQOL Group. (1993). Study protocol for the World Health Organisation project to develop a Quality of Life assessment instrument (WHOQOL). *Quality of Life Research*, *2*, 153-159.

WHOQOL Group. (1998). The World Health Organisation quality of life assessment (WHOQOL): Development and general psychometric properties. *Social Science & Medicine*, *12*, 1569-1585.

Wirnsberger, R. M., de Vries, J., Breteler, M. H., van Heck, G. L., Woutres, E. F., & Drent, M. (1998). Evaluation of quality of life in sarcoidosis patients. *Respiratory Medicine*, 92, 750-756.

Reporter: Nicholas Marosszeky, Research Psychologist.

Date of report: 12 August 2003.

World Health Organisation Quality of Life-BREF

Title:	World Health Organisation Quality of Life-BREF *(Australian English Language Version).*
Abbreviations:	WHOQOL-BREF.
Author(s) Name:	The WHOQOL Group, Division of Mental Health, World Health Organisation (WHO).
Author(s) Address:	WHOQOL Group Division of Mental Health World Health Organisation CH-1211 Geneva 27 SWITZERLAND
Supplied by:	Australian WHOQOL Field Centre Australian Centre for Posttraumatic Mental Health Austin and Repatriation Medical Centre PO Box 5444 Heidelberg West VIC 3081 www.acpmh.unimelb.edu.au/whoqol/ (*For more information about the WHOQOL-BREF contact A/Professor Graeme Hawthorne at graemeeh@unimelb.edu.au or by telephone on 03 9496-4031*).
Cost:	Free. Users must register with the Melbourne WHOQOL Field Study Centre. (Users are not authorised to make any modifications to the instrument without permission.)
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Basic training in survey administration and the characteristics of the WHOQOL-BREF is required by those without these qualifications.
Purpose:	The WHOQOL-BREF is a brief quality of life (QOL) assessment instrument, which is comparable across cultures and languages.
Administration Time:	5 – 10 minutes.
Instrument Type:	Self-report Questionnaire.
Structure:	The WHOQOL-BREF includes 26 items (one item from each of the 24 facets contained in the WHOQOL-100) covering four domains: Physical, Psychological, Social relationships, Environment (a reduction from the set of six produced by the WHOQOL-100). The 26 items are listed below:

Physical

    Pain and discomfort
    Dependence on medical treatment
    Energy and fatigue
    Mobility
    Sleep and rest
    Activities of daily living
    Working capacity

Psychological

    Positive affect
    Spirituality
    Thinking, learning, memory and concentration
    Body image and appearance
    Self-esteem
    Negative affect

Social relationships
> Personal relationships
> Sexual activity
> Social support

Environment
> Physical safety and security
> Physical environment (pollution, noise, traffic, climate)
> Financial resources
> Opportunities for acquiring new information and skills
> Participation in and opportunities for recreation/leisure activities
> Home environment
> Health and social care: accessibility and quality
> Transportation

Plus two items on Overall Quality of life and satisfaction with health.

Scoring:	Each of the 26 items is presented in a question format and is answered on a five-point response scale. They are then summed to produce the Domain scores (some items need to be reversed scored). Scoring algorithms are provided. Scores are then converted to a 0 – 100 scale. A higher score indicates better QOL.
	NB: Facet scores can not be derived from the single items used in the WHOQOL-BREF.
Developed for:	The WHOQOL Project attempted to derive the major issues or themes that are universally accepted as important in determining a person's Quality of Life (QOL). The WHOQOL-BREF is one result of this enterprise. It is a brief assessment of QOL extracted from its' parent the WHOQOL-100.
Normative Data:	Australian Data is provided by the Victorian Validation Study (n = 396) and the Longitudinal Investigation of Depression Outcomes (LIDO) Study (n = 115).
	International Data for the WHOQOL-BREF is provided in WHOQOL Group (1998b).
Clinical Data:	A few clinical studies are listed below:
	Depression: the Longitudinal Investigation of Depression Outcomes (LIDO) Study.
	HIV: Fang, Hsiung, Yu, Chen & Wang (2002) and Chandra, Deepthivarma, Jairam & Thomas (2003).
	Liver Transplantation: O'Carroll, Smith, Couston, Cossar & Hayes (2000).
Schizophrenia:	Herrman, Hawthorne & Thomas (2002) and Ritchie, Chiu, Harrigan, Hall, Hassett, Macfarlane et al. (2003).
Applications:	It should be noted that the WHOQOL instruments have been designed for population research or clinical trials. They are not to be used for individual assessment.
	The WHOQOL-BREF can also be administered by an interviewer.
	Leplege et al. (2002) have attempted to derive a single index score for QOL from the WHOQOL-BREF using Rasch Analysis.
	Saxena, Carlson, Billington & Orley (2001) have examined the importance of the WHOQOL-BREF items across cultures.
	See also the Instrument Review on the WHOQOL-100.

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	WHOQOL Group (1998b) Murphy et al. (2000)	Adequate	Cronbach's α is in the range of 0.60 – 0.90 for the four domains. It is low for the Social Relationships domain.
Test – retest	Yes	WHOQOL Group (1998b) Murphy et al. (2000)	Adequate	Pearson correlations were above 0.80 for the four domains.[1]
Inter – rater	NA	NA	NA	The WHOQOL-BREF is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	WHOQOL Group (1998b) Murphy et al. (2000)	Adequate	
Correlation with other measures	Yes	Murphy et al. (2000)	Adequate	At this time, there is no generally accepted measure (i.e. gold standard) for the construct of Quality of Life in the scientific literature.
Construct	Yes	WHOQOL Group (1998b) Murphy et al. (2000)	Adequate	The domain scores of the WHOQOL-BREF correlate highly (0.89 or above) with WHOQOL-100 domain scores (calculated on a 4-domain structure).
Criterion	Yes	WHOQOL Group (1998b)	Adequate	The criterion is how well the WHOQOL-BREF reproduces the domain scores of the WHOQOL-100 (see above).

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Murphy et al. (2000) O'Carroll et al. (2000) Ritchie et al. (2003)	Adequate	

Cultural Applicability and Cultural Adaptations:

The WHOQOL-BREF is currently available in over 20 language versions However, further research is required to examine the application of the WHOQOL-BREF within the context of a multi-cultural society.

Gender Appropriateness:

Normative data for males and females is not available at this time.

Age Appropriateness:

Adults, with "adult" being culturally defined.

Summary:

The WHOQOL-BREF is a recommended brief assessment of quality of life (QOL). Derived from the WHOQOL-100, the WHOQOL-BREF has undergone an extensive process of translation and development in order to make it comparable across languages and cultures.

References

Chandra, P. S., Deepthivarma, S., Jairam, K. R., & Thomas, T. (2003). Relationship of psychological morbidity and quality of life to illness-related disclosure among HIV-infected persons. *Journal of Psychosomatic Research*, *54*, 199-203.

Fang, C. T., Hsiung, P. C., Yu, C. F., Chen, M. Y., & Wang, J. D. (2002). Validation of the World Health Organisation quality of life instrument in patients with HIV infection. *Quality of Life Research*, *11*, 753-762.

Herrman, H., Hawthorne, G., & Thomas, R. (2002). Quality of life assessment in people living with psychosis. *Social Psychiatry & Psychiatric Epidemiology*, *37*, 510-518.

Herrman, H., Patrick, D. L., Diehr, P., Martin, M. L., Fleck, M., Simon, G. E, et al. (2002). Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. *Psychological Medicine*, *32*, 889-902.

Leplege, A., Ecosse, E., & WHOQOL Rasch Project Scientific Committee. (2000). Methodological issues in using the Rasch model to select cross culturally equivalent items in order to develop a quality of life index: The analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong, United Kingdom). *Journal of Applied Measurement*, *1*, 372-392.

Murphy, B., Herrman, H., Hawthorne, G., Pinzone, T., & Evert, H. (2000). *Australian WHOQOL instruments: User's Manual and Interpretation Guide*. Melbourne: Australian WHOQOL Field Study Centre.

O'Carroll, R. E., Smith, K., Couston, M., Cossar, J. A., & Hayes, P. C. (2000). A comparison of the WHOQOL-100 and the WHOQOL-BREF in detecting change in quality of life following liver transplantation. *Quality of Life Research*, *9*, 121-124.

Ritchie, C. W., Chiu, E., Harrigan, S., Hall, K., Hassett, A., Macfarlane, S. et al. (2003). The impact upon extra-pyramidal side effects, clinical symptoms and quality of life of a switch from conventional to atypical antipsychotics (risperidone or olanzapine) in elderly patients with schizophrenia. *International Journal of Geriatric Psychiatry*, *18*, 432-440.

Saxena, S., Carlson, D., Billington, R., & Orley, J. (2001). The WHO quality of life assessment instrument (WHOQOL-Bref): The importance of its items for cross-cultural research. *Quality of Life Research*, *10*, 711-721.

WHOQOL Group. (1998b). Group development of the World Health Organisation WHOQOL-BREF quality of life assessment. *Psychological Medicine*, *28*, 551-558.

WHOQOL Group. (1993). Study protocol for the World Health Organisation project to develop a Quality of Life assessment instrument (WHOQOL). *Quality of Life Research*, *2*, 153-159.

WHOQOL Group. (1998a). The World Health Organisation quality of life assessment (WHOQOL): Development and general psychometric properties. *Social Science & Medicine, 12*, 1569-1585.

Reporter: Nicholas Marosszeky, Research Psychologist.

Date of report: 12 August 2003.

Sickness Impact Profile

Title:	Sickness Impact Profile (Final Revision) *(Also known as the Functional Limitation Profile (UK Version)).*
Abbreviations:	SIP™ *or FLP.*
Author(s) Name:	Marilyn Bergner.
Author(s) Address:	Late of the Johns Hopkins School of Public Health 615 N. Wolfe St. Baltimore MD 21205-2179 USA
Supplied by:	Medical Outcomes Trust PMB #503 198 Tremont St Boston, MA 02116-4705 USA Phone: US 617-426-4046 Fax: US 617 587 4232 Email: info@outcomes-trust.org or Elizabeth Skinner Health Policy and Management Bloomberg School of Public Health John Hopkins University (USA) 624 N. Broadway, Baltimore, Maryland 21205, USA Phone: (410) 614-4022 Fax: (410) 955-0470 Email: askinner@jhsph.edu
Cost:	Free. However, one must seek permission to use the Sickness Impact Profile. It costs $US160 to purchase the Sickness Impact Profile User Manual from the Medical Outcomes Trust.
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Those without these qualifications require basic training in survey administration and the characteristics of the SIP™.
Purpose:	The Sickness Impact Profile is a generic outcome measure designed to provide a descriptive profile of changes in a person's health status due to sickness.
Administration Time:	20 – 30 minutes. Respondents check or endorse only the items that describe them on a given day. It can be self-administered or administered by interview.
Instrument Type:	Self-report questionnaire, generic health status measure.
Structure:	Patients are asked to describe themselves today and their state of health on a number of behaviourally based statements e.g. "I am not doing heavy work around the house"; "I am going out less to visit people". The conceptual model of the SIP™ is based on a continuum of sickness and behaviour. At one end of the continuum individuals are healthy and behave without limitation. At the other end, sick individuals experience limitations and exhibit dysfunction.
	The final revision of the SIP™ consists of 136 behaviourally based items covering 7 dimensions with 12 categories. There is the Physical Dimension (includes Ambulation, Mobility, Body Care & Movement); the Psychosocial Dimension (includes the categories of Communication, Alertness Behavior, Emotional Behaviour, Social Interaction) and the independent dimensions

of Sleep & Rest, Eating, Work, Home Management, and Recreation & Pastimes.

Scoring:	Items endorsed (with a check mark) by patients are counted and the weighted total of each category is calculated. The sum is then divided by the maximum possible score of the SIP™ and multiplied by 100 to give a total score (0 – 100). Physical and Psychosocial dimensions are calculated from the appropriate categories Higher scores indicate greater dysfunction.
	It takes approximately 10 minutes to score. As respondents only endorse applicable items it may be hard to differentiate between a no response and missing data (Bowling, 1997).
Developed for:	The SIP™ can be used for all diseases and for all adult populations as it is a generic instrument. It is broadly applicable and was intended for use in measuring the outcomes of care, in health surveys, in program planning and policy formulation, and in monitoring patient progress (McDowell and Newell, 1996).
	The final revision of the SIP™ published by Bergner, Bobbitt, Carter & Gilson (1981) was a result of an extensive process of psychometric testing.
	SIP™ behaviour statements were elicited from over 1000 health care professionals, patients, family caregivers and healthy individuals. During extensive field testing, the pool of candidate behaviours was assigned a weight reflecting its degree of dysfunction. During the course of its development 312 health behaviour statements were reduced to 136.
Normative Data:	Normative data is provided in the Sickness Impact Profile User Manual, "Published mean overall SIP™ scores range from 3 in a general population sample to 44 in a frail elderly sample."
	Only a few studies using the SIP™ have been conducted in Australia. They include:

    Hall et al. (1987).
    Harper, Harper, Lambert, de Klerk, et al. (1995).
    Katsikitis, Davis & Burrow (1996).
    Gardner & Sibthorpe (2002).

Clinical Data:	It has been used to assess health status for patients with many clinical conditions, which include myocardial infarction, renal disease, rheumatoid arthritis, hyperthyroidism, hip replacement, heart transplant, in-patients, outpatients, and homecare patients with chronic diseases.

A few clinical studies are listed below:

Angina Pectoris: Visser, Fletcher, Parr, Simpson & Bulpitt (1994).
Cancer: Bergman, Sullivan & Sorenson (1992).
Cardiac Rehabilitation: Christensen, Edwards, Moran, Burke, Lounsbury & Gordon (1999).
Cataract Surgery: Damiano, Steinberg, Cassard, Bass et al. (1995).
Chronic Hepatitis C: Davis, Balart, Schiff, Lindsay, Bodenheimer, Perrillo et al. (1994).
Chronic Low Back Pain: Follick, Smith & Ahern (1985); and Stratford, Solomon, Binkley, Finch Gill (1993); and Slater, Doctor, Pruitt, Atkinson (1997).
Chronic Pain: Watt-Watson & Graydon (1989).
Chronic Airflow Limitation: Jones, Baveystock & Littlejohns (1989).
Elderly: Yeto, Ingram, Skurnick, Crapo (1987); and Mulrow, Gerety, Cornell, Lawrence, Kanten (1994); and Weinberger, Samsa, Hanlon, Schmader et al. (1991).
End-Stage Renal Disease: Hart & Evans (1987).
Guillain-Barre Syndrome: Bernsen, Jacobs, de Jager & van der Meche (1997).
Head Injury: Temkin, McLean, Dikmen, Gale, Bergner & Almes (1988);

and Temkin, Dikmen, Machamer & McLean (1989); and Temkin, Dikmen, Machamer & McLean (1989).

Hearing Impairment: Bess, Lichtensen, Logan, Burger, et al. (1989).

Heart Transplantation: Konstam, Surman, Hijazi, Konstam, Fierstein, Turbett et al.(1997).

Hip Replacement: Katz, Larson, Phillips, Fossel et al. (1992); and Katz, Phillips, Fossel, Liang et al. (1994).

HIV & AIDS: Weinstein (1990).

Knee or Hip Osteoarthritis: Weinberger, Samsa, Tierney, Belyea, Hiner (1992).

Multiple Sclerosis: Zeldow & Pavlou (1988).

Nursing Home Residents: Rothman, Hedrick & Inui (1989).

Parkinson's Disease: Longstreth, Nelson, Linde & Munoz (1992); and Katsikitis, Davis & Burrow (1996); and Straits-Troester, Fields, Wilkinson, Pahwa, Lyons, Koller & Troester (2000).

Pneumonia: Hasley, Brancati, Rogers, Hanusa et al. (1993).

Polio Survivors: Schanke (1997).

Post-Traumatic Stress: Friedland & Dawson (2001).

Renal Dialysis: Essink-Bot, Krabbe, van Agt & Bonsel (1996).

Rheumatoid Arthritis: Sullivan, Ahlmen & Bjelle (1990); and Sullivan, Ahlmen, Bjelle & Karlsson (1993).

Schizophrenia: Awad, Lapierre, Angus & Rylander (1997).

Stroke: Mayer, Kreiter, Copeland, Bernardini, Bates, Peery et al. (2002); and van Straten, de Haan, Limburg, Schuling, Bossuyt & van den Bos (1997).

Trauma – Lower Extremity Fractures: Jurkovich, Mock, MacKenzie, Burgess, Cushing, deLateur et al. (1995).

Trauma – General: Temkin, McLean, Dikmen, Gale, Bergner & Almes (1988); and Temkin, Dikmen, Machamer & McLean (1989); and Temkin, Dikmen, Machamer & McLean (1989).

Urinary Incontinence in women: Hunskaar & Vinsnes (1991).

Applications:	The SIP™ can also be administered by an interviewer. Hulsebos, Beltan, dos Reis Miranda & Spangenberg (1991) provide information comparing the mode of administration for the SIP™ – interview, mail, and telephone and mail.

It can be applied to populations with a wide range in levels of sickness. However, healthy people or those with mild health problems get very low scores on this instrument. It is known that a normal population may only score 2-3/100 on this instrument and thus the instrument may be less sensitive to detecting health problems in those with minor ailments. Hall et al. (1987) reported that 18% of patients in a general practice setting scored 0 on the SIP total percent score. McKenzie et al. (1986) reported that it is more sensitive to deterioration rather than improvements in health status. Its length, also, may limit its use in some service settings.

The work of de Bruin & colleagues has produced a shorter 68 item version of Sickness Impact Profile – the SIP-68.

A number of reduced item versions of the SIP™ for specific disease groups or purposes have also been developed:

The Reduced SIP™ for Nursing Homes.

The Roland Scale for pain patients.

The 64 item SIP for rheumatoid arthritis.

The 30 item SIP for stroke patients (SA-SIP30).

The Mobility Control sub-scale of the SIP-68 has been used to screen for mobility disorders in the elderly.

However, the work of Temkin and colleagues makes the case for the use of general versions of the SIP™ rather than disease specific ones (at least for head injured patients).

Research into the use of the caregivers or proxies to generate SIP™ scores for patients has been undertaken by McCusker & Stoddard (1984) and Rothman, Hedrick, Bulcroft, Hickam et al. (1991).

Pollard & Johnston (2001) have suggested a new method of scoring the SIP™ improving the summated total scores.

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Bergner et al. (1981) Rothman et al. (1989) de Bruin et al. (1994)	Good	
Test – retest	Yes	Pollard et al. (1976) Bergner et al. (1981) de Bruin et al. (1994) Damiano (1996) Beckerman et al. (2001)	Adequate	
Inter – rater	NA	NA	NA	The SIP™ is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Bergner et al. (1981) Deyo et al. (1983) Visser et al. (1994) Damiano (1996)	Adequate	
Correlation with other measures		Bergner et al. (1981) Follick et al. (1985) Zeldow & Pavlou (1988) Jones et al. (1989) Rothman et al. (1989) Watt-Watson & Graydon (1989) Brooks et al. (1990) Sullivan et al. (1990) Weinberger et al. (1991) de Bruin et al. (1992) Bergman et al. (1992) Katz et al. (1992) Weinberger et al. (1992) Visser et al. (1994) Beaton et al. (1996) Lipsett et al. (2000)		Beaton et al. (1996) reports a ceiling effect with the SIP™ for healthy workers

Construct	Yes	Bergner et al. (1981) Rothman et al. (1989) Brooks et al. (1990) de Bruin et al. (1992) de Bruin et al. (1994) Tian & Miranda (1995) Jette et al. (1998)	Adequate	Pollard & Johnston (2001) criticise the current method of summated scoring for the SIP™
Criterion	Yes	Bergner et al. (1981) Post et al. (2001)	Adequate	

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Follick et al. (1985) MacKenzie et al. (1986) Sullivan et al. (1990) Katz et al. (1992) Stratford et al. (1993) Hasley et al. (1993) Katz et al. (1994) Davis et al. (1994) Visser et al. (1994) de Bruin et al. (1997) Konstam et al. (1997) Beckerman et al. (2001)	Adequate	Two papers by Davis et al. (1994) and MacKenzie et al. (1986) criticise the SIP's ability to measure change over time and limit it to describing health states. Refer McDowell and Newell (1996) and de Bruin et al. (1992) concerning its sensitivity to change compared with other measures.

Cultural Applicability and Cultural Adaptations:	Some cross cultural work using the SIP™ has been conducted in the USA and the UK.1,75,76. The SIP™ has been translated into Chinese.77 There are a large number of translations available and information on this aspect is available through the MOT at http://www.outcomes-trust.org/ or the QOLID database at http://www.qolid.org/
Gender Appropriateness:	"The SIP™ is a behaviourally based instrument that is broadly applicable across types and severity of illness, as well as across demographically and culturally diverse groups."
Age Appropriateness:	Adults.
Summary:	The Sickness Impact Profile is a recommended generic outcome measure of perceived health status. The SIP is a well-constructed measure with adequate psychometric properties. It has been used extensively in a wide range of settings. However, its length may preclude its use for some applications and questions have been raised concerning its responsiveness to change, and thus its use as an outcome measure.

References

Awad, A., Lapierre, Y. D., Angus, C., & Rylander, A. (1997). Quality of life and response of negative symptoms in schizophrenia to haloperidol and the atypical antipsychotic remoxipride. *Journal of Psychiatry & Neuroscience*, *22*, 244-248.

Beaton, D. E., Bombardier, C., & Hogg-Johnson, S. A. (1996). Measuring health in injured workers: a cross-sectional comparison of five generic health status instruments in workers with musculoskeletal injuries. *American Journal of Industrial Medicine*, *29*, 618-631.

Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. M. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research*, *10*, 571-578.

Bergman, B., Sullivan, M., & Sorenson, S. (1992). Quality of life during chemotherapy for small cell lung cancer. II. A lognitudinal study of the EORTC Core Quality of Life Questionnaire and comparison with the Sickness Impact Profile. *Acta Oncologica*, *31*, 19-28.

Bergner, M., Bobbitt, R. A., Carter, W. B., & Gilson, B. S. (1981). The Sickness Impact Profile: development and final revision of a health status measure. *Medical Care*, *19*, 787-805.

Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E., Gilson, B. S., & Morris, J. R. (1976). The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services*, *6*, 393-415.

Bergner, M., Bobbitt, R. A., Pollard, W. E., Martin, D. P., & Gilson, B. S. (1976). The sickness impact profile: validation of a health status measure. *Medical Care*, *14*, 57-67.

Bernsen, R., Jacobs, H., de Jager, A., & van der Meche, F. (1997). Residual health status after Guillain-Barre syndrome. *Journal of Neurology, Neurosurgery & Psychiatry*, *62*, 637-640.

Bess, F. H., Lichtenstein, M. J., Logan, S. A., Burger, M. C., et al. (1989). Hearing impairment as a determinant of function in the elderly. *Journal of the American Geriatrics Society*, *37*, 123-128.

Brooks, W. B., Jordan, J. S., Divine, G. W., Smith, K. S., & Neelon, F. A. (1990). The impact of psychologic factors on measurement of functional status. Assessment of the sickness impact profile. *Medical Care*, *28*, 793-804.

Carter, W. B., Bobbitt, R. A., Bergner, M., & Gilson, B. S. (1976). Validation of an interval scaling: the sickness impact profile. *Health Services Research*, *11*, 516-528.

Christensen, A. J., Edwards, D. L., Moran, P. J., Burke, R., Lounsbury, P., & Gordon, E. E. I. (1999). Cognitive distortion and functional impairment in patients undergoing cardiac rehabilitation. *Cognitive Therapy & Research*, *23*, 159-168.

Damiano, A. M., Steinberg, E. P., Cassard, S. D., Bass, E. B., Diener-West, M., Legro, M. et al. (1995). Comparison of generic versus disease-specific measures of functional impairment in patients with cataract. *Medical Care*, *33* (Suppl 4), AS120-AS130.

Damiano, A. M. (1996). The Sickness Impact Profile: Part 1. *Medical Outcomes Trust Bulletin*, *4*, 2.

Davis, G. L., Balart, L. A., Schiff, E. R., Lindsay, K., Bodenheimer, H. C. Jr., Perrillo, R. P. et al. (1994). Assessing health-related quality of life in chronic hepatitis C using the Sickness Impact Profile. *Clinical Therapeutics*, *16*, 334-343.

de Bruin, A. F., Buys, M., de Witte, L. P., & Diederiks, J. P. (1994). The Sickness Impact Profile: SIP68, a short generic version. First evaluation of the reliability and reproducibility. *Journal of Clinical Epidemiology*, *47*, 863-871.

de Bruin, A. F., Diederiks, J. P., de Witte, L. P., Stevens, F. C., & Philipsen, H. (1994). The development of a short generic version of the Sickness Impact Profile. *Journal of Clinical Epidemiology*, *47*, 407-418.

de Bruin, A. F., Diederiks, J. P., de Witte, L. P., Stevens, F. C., & Philipsen, H. (1997). Assessing the responsiveness of a functional status measure: the Sickness Impact Profile versus the SIP68. *Journal of Clinical Epidemiology*, *50*, 529-540.

de Bruin, A. F., de Witte, L. P., & Diedericks, J.P. (1992). Sickness Impact Profile: the state of the art of a generic functional status measure. *Social Science and Medicine*, I*35*, 1003-14.

Deyo, R. A., Inui, T. S., Leininger, J. D., & Overman, S. S. (1983). Measuring functional outcomes in chronic disease: A comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. *Medical Care*, *21*, 180-192.

Dhanda, R., Mulrow, C. D., Gerety, M. B., Lee, S., & Cornell, J. E. (1995). Classifying change with the Sickness Impact Profile for Nursing Homes (SIP-NH). *Aging-Clinical & Experimental Research*, *7*, 228-233.

Dittmar, S. S., & Gresham, G. E. (1997). *Functional Assessment and Outcome Measures for the Rehabilitation Health Professional*. MA: Aspen Publishers Inc.

Essink-Bot, M. L., Krabbe, P. F. M., van Agt, H. M. E, Bonsel, G. J. NHP or SIP: (1996). A comparative study in renal insufficiency associated anemia. *Quality of Life Research*, *5*, 91-100.

Follick, M. J., Smith, T. W., Ahern, D. K. (1985). The sickness impact profile: a global measure of disability in chronic low back pain. *Pain*, *21*, 67-76.

Friedland, J. F., & Dawson, D. R. (2001). Function after motor vehicle accidents: A prospective study of mild head injury and posttraumatic stress. *Journal of Nervous & Mental Disease*, *189*, 426-434.

Gardner, A., & Sibthorpe, B. (2002). Will he get back to normal? Survival and functional status after intensive care therapy. *Intensive & Critical Care Nursing*, *18*, 138-145.

Gerety, M. B., Cornell, J. E., Mulrow, C. D., Tuley, M., Hazuda, H. P., Lichtenstein, M. et al. (1994). The Sickness Impact Profile for nursing homes (SIP-NH). *Journal of Gerontology*, *49*, M2-M8.

Gilson, B. S., Erickson, D., Chavez, C. T., Bobbitt, R. A., Bergner, M., & Carter, W. B. (1980). A Chicano version of the Sickness Impact Profile. *Culture Medicine and Psychiatry*, *4*, 137-150.

Gilson, B. S., Gilson, J. S., Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E. et al. (1979). The sickness impact profile: Development of an outcome measure of health care. *American Journal of Public Health*, *65*, 1304-1310.

Hall, J., Hall, N., Fisher, E. et al. (1987). Measurement of outcomes of general practice: comparison of three health status measures. *Family Practice*, *4*, 117-23.

Harper, A. C., Harper, D. A., Lambert, L. J., de Klerk, N. H., et al. (1995). Development and validation of the Curtin Back Screening Questionnaire (CBSQ): A discriminative disability measure. *Pain*, *60*, 73-81.

Hart, L. G., & Evans, R. W. (1987). The functional status of ESRD patients as measured by the Sickness Impact Profile. *Journal of Chronic Diseases*, *40* (Suppl 1), 117S-136S.

Hasley, P. B., Brancati, F. L., Rogers, J., Hanusa, B. H., et al. (2003). Measuring functional change in community-acquired pneumonia: A preliminary study using the Sickness Impact Profile. *Medical Care*, *31*, 649-657.

Hulsebos, R. G., Beltman, F. W., dos Reis Miranda, D., (1991). Spangenberg, J. F. Measuring quality of life with the sickness impact profile: a pilot study. *Intensive Care Medicine*, *17*, 285-288.

Hunskaar, S., & Vinsnes, A. The quality of life in women with urinary incontinence as measured by the sickness impact profile. *Journal of the American Geriatrics Society*, *39*, 378-382.

Jannink-Nijlant, J. M., Diederiks, J. P., Brouwers, M. A., Metsemakers, J. F. (1999). Screening for mobility disorders by the Mobility Control subscales of the short version of the Sickness Impact Profile. *Clinical Rehabilitation*, *13*, 492-497.

Jensen, M. P., Strom, S. E., Turner, J. A., & Romano, J. M. (1992). Validity of the Sickness Impact Profile Roland scale as a measure of dysfunction in chronic pain patients. *Pain*, *50*, 157-162.

Jette, A. M., Assmann, S. F., Rooks, D., Harris, B. A., & Crawford, S. (1998). Interrelationships among disablement concepts. *Journals of Gerontology Series A - Biological Sciences & Medical Sciences*, *53A*, M395-M404.

Jones, P. W., Baveystock, C. M., & Littlejohns, P. (1989). Relationships between general health measured with the sickness impact profile and respiratory symptoms, physiological measures, and mood in patients with chronic airflow limitation. *American Review of Respiratory Disease*, *140*, 1538-1543.

Jurkovich, G., Mock, C., MacKenzie, E., Burgess, A., Cushing, B., deLateur, B. et al. (1995). The Sickness Impact Profile as a tool to evaluate functional outcome in trauma patients. *Journal of Trauma-Injury Infection & Critical Care*, *39*, 625-631.

Katsikitis, M., Davis, A., & Burrow, D. (1996). An Australian community study of Parkinson's disease: 1. Disease severity and functional disability. *Australian Journal of Ageing*, *15*, 22-26.

Katz, J. N., Larson, M. G., Phillips, C. B., Fossel, A. H., et al. (1992). Comparative measurement sensitivity of short and longer health status instruments. *Medical Care*, *30*, 917-925.

Katz, J. N., Phillips, C. B., Fossel, A. H., & Liang, M. H. (1994). Stability and responsiveness of utility measures. *Medical Care*, *32*, 183-188.

Konstam, V., Surman, O., Hijazi, K. H., Konstam, M. A., Fierstein, J., Turbett, A. et al. (1997). A longitudinal study of health-related quality of life in recipients of heart transplantations. *Journal of Applied Rehabilitation Counseling*, *28*, 51-56.

Lipsett, P. A., Swoboda, S. M., Campbell, K. A., Cornewell, E. I., Dorman, T., & Pronovost, P. J. (2000). Sickness Impact Profile Score versus a Modified Short-Form survey for functional outcome assessment: acceptability, reliability, and validity in critically ill patients with prolonged intensive care unit stays. *Journal of Trauma-Injury Infection & Critical Care*, *49*, 737-743.

Longstreth, W. T. J., Nelson, L., Linde, M., & Munoz, D. (1992). Utility of the sickness impact profile in Parkinson's disease. *Journal of Geriatric Psychiatry & Neurology*, *5*, 142-148.

MacKenzie, C. R., Charlson, M. E., DiGioia, D., & Kelley, K. (1986). Can the Sickness Impact Profile measure change? *An example of scale assessment. Journal of Chronic Diseases*, *39*, 429-438.

Mayer, S. A., Kreiter, K. T., Copeland, D., Bernardini, G. L., Bates, J. E., Peery, S. et al. (2002). Global and domain-specific cognitive impairment and outcome after subarachnoid hemorrhage. *Neurology*, *59*, 1750-1758.

McCusker, J., & Stoddard, A. M. Use of a surrogate for the Sickness Impact Profile. *Medical Care*, *22*, 789-795.

McDowell, I. & Newell, C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd Ed. Oxford University Press, New York.

McKenzie, C. R., Charlson, M. E., DiGioia, D. et al. (1986). Can the Sickness Impact Profile measure change? An example of scale assessment. *Journal of Chronic Disease*, *39*, 429-438.

Mulrow, C. D., Gerety, M. B., Cornell, J. E, Lawrence, V. A., & Kanten, D. N. (1994). The relationship between disease and function and perceived health in frail elders. *Journal of the American Geriatrics Society*, *42*, 374-380.

Patrick, D. L., Sittampalam, Y., Sommerville, S. M., Carter, W. B., & Bergner, M. (1985). A cross-cultural comparison of health status values. *American Journal of Public Health*, *95*, 1402-1407.

Pollard, B., & Johnston, M. (2001). Problems with the sickness impact profile: a theoretically based analysis and a proposal for a new method of scoring. *Social Science & Medicine*, *52*, 921-934.

Pollard, W. E., Bobbitt, R. A., Bergner, M., Martin, D. P., & Gilson, B. S. (1976). The Sickness Impact Profile: reliability of a health status measure. *Medical Care*, *14*, 146-155.

Post, M. W., Gerritsen, J., Diederiks, J. P., & DeWittet, L. P. (2001). Measuring health status of people who are wheelchair-dependent: validity of the Sickness Impact Profile and the Nottingham Health Profile. *Disability & Rehabilitation*, *23*, 245-253.

Rothman, M. L., Hedrick, S., & Inui, T. (1989). The Sickness Impact Profile as a measure of health status of noncognitively impaired nursing home residents. *Medical Care*, *27* (Suppl 3), S157-S167.

Rothman, M. L., Hedrick, S. C., Bulcroft, K. A., Hickam, D. H., et al. (1991). The validity of proxy-generated scores as measures of patient health status. *Medical Care*, *29*, 115-124.

Schanke, A. K. (1997). Psychological distress, social support and coping behaviour among polio survivors: A 5-year perspective on 63 polio patients. *Disability & Rehabilitation*, *19*, 108-116.

Short, T. G., Rowbottom, M. Y., Lau, J. P., Lai, G. W., Buckley, T. A., & Oh, T. E. (1998). Translation and calibration of a Chinese version of the Sickness Impact Profile for use in Hong Kong. *Hong Kong Medical Journal*, *4*, 375-381.

Slater, M. A., Doctor, J. N., Pruitt, S. D., & Atkinson, J. H. (1997). The clinical significance of behavioral treatment for chronic low back pain: An evaluation of effectiveness. *Pain*, *71*, 257-263.

Straits-Troester, K., Fields, J. A., Wilkinson, S. B., Pahwa, R., Lyons, K. E., Koller, W. C. et al. (2000). Health-related quality of life in Parkinson's disease after pallidotomy and deep brain stimulation. Brain & Cognition, *42*, 399-416.

Stratford, P., Solomon, P., Binkley, J., Finch, E., & Gill, C. (1993). Sensitivity of Sickness Impact Profile items to measure change over time in a low-back pain patient group. *Spine*, *18*, 1723-1727.

Sullivan, M., Ahimen, M., Bjelle, A., & Karlsson, J. (1993). Health status measurement in rheumatoid arthritis. II. Evaluation of a modified Shorter Sickness Impact Profile. *Journal of Rheumatology*, *20*, 1500-1507.

Sullivan, M., Ahimen, M., & Bjelle, A. (1990). Health status measurement in rheumatoid arthritis. I. Further work on the validity of the sickness impact profile. *Journal of Rheumatology*, *17*, 439-447.

Temkin, N., McLean, A. J., Dikmen, S., Gale, J., Bergner, M., & Almes, M. J. (1988). Development and evaluation of modifications to the Sickness Impact Profile for head injury. Journal of Clinical Epidemiology, *41*, 47-57.

Temkin, N. R., Dikmen, S., Machamer, J., & McLean, A. (1989). General versus disease-specific measures. Further work on the Sickness Impact Profile for head injury. *Medical Care*, *27* (Suppl 3), S44-S53.

Temkin, N. R., Dikmen, S., Machamer, J., & McLean, A. (1989). Modification of the Sickness Impact Profile. Medical Care, *27* (Suppl 3), S44-S53.

Tian, Z. M., & Miranda, D. R. (1995). Quality of life after intensive care with sickness impact profile. *Intensive Care Medicine*, *21*, 422-428.

van Straten, A., de Haan, R. J., Limburg, M., Schuling, J., Bossuyt, P. M., & van den Bos, G. A. (1997). A stroke-adapted 30-item version of the Sickness Impact Profile to assess quality of life (SA-SIP30). *Stroke*, *28*, 2155-2161.

Visser, M. C., Flecther, A. E., Parr, G., Simpson, A., & Bulpitt, C. J. (1994). A comparison of three quality of life instruments in subjects with angina pectoris: the Sickness Impact Profile, the Nottingham Health Profile, and the Quality of Well Being Scale. *Journal of Clinical Epidemiology*, *47*, 157-163.

Watt-Watson, J. H., & Graydon, J. E. (1989). Sickness Impact Profile: a measure of dysfunction with chronic pain patients. Journal of Pain & Symptom Management, *4*, 152-156.

Weinberger, M., Samsa, G. P., Hanlon, J. T., Schmader, K., et al. (1991). An evaluation of a brief health status measure in elderly veterans. *Journal of the American Geriatrics Society*, *39*, 691-694.

Weinberger, M., Samsa, G. P., Tierney, W. M., Belyea, M. J., & Hiner, S. L. (1992). Generic versus disease specific health status measures: comparing the sickness impact profile and the arthritis impact measurement scales. *Journal of Rheumatology*, *19*, 543-546.

Weinstein, B. D. (1990). Assessing the impact of HIV disease. *American Journal of Occupational Therapy*, *44*, 220-226.

Yeto, G., Ingram, L., Shurnick, J., & Crapo, L. (1987). (1987). Effects of a geriatric clinic on functional health and well-being of elders. *Journal of Gerontology*, *42*, 252-258.

Zeldow, P. B., & Pavlou, M. (1988). Physical and psychosocial functioning in multiple sclerosis: Descriptions, correlations, and a tentative typology. *British Journal of Medical Psychology*, *61*, 185-195.

Note: An extensive bibliography is available for registered users on the QOLID database at www.qolid.org and a bibliographical search for recent abstracts can undertaken at phi.uhce.ox.ac.uk/phidb.html

Reporter: Nicholas Marosszeky, Research Psychologist.

Date of report: 12 August 2003.

Nottingham Health Profile

Title:	Nottingham Health Profile.
Abbreviations:	NHP.
Author(s) Name:	Sonja M. Hunt, James McEwen, and Stephen P. McKenna.
Author(s) Address:	Dr. Stephen P. McKenna Galen Research Enterprise House Manchester Science Park Lloyd Street North Manchester, M15 6SE UK www.galen-research.com
Supplied by:	A copy of the Nottingham Health Profile (NHP) is provided in the article by Baum & Cooke (1989) published in the Medical Journal of Australia (MJA).
Cost:	Nil. Need to seek permission to use the NHP from one of the authors.
Training Requirements:	Nil training is required for those professionals with qualifications and experience in psychometrics and statistics. Those without these qualifications require basic training in survey administration and the characteristics of the NHP.
Purpose:	The Nottingham Health Profile is a generic outcome measure designed to examine a person's perceived health status. Items ask about feelings and emotional status directly rather than via changes in behaviours – as contrasted with the SIP (McDowell and Newell, 1996).
Administration Time:	5 – 10 minutes.
Instrument Type:	Self-report questionnaire; generic health status measure.
Structure:	The Nottingham Health Profile is in two parts. Part I: Consists of 38 statements about health problems, covering six domains: Physical Mobility (8 items); Pain (8 items); Energy (3 items); Social Isolation (5 items); Sleep (5 items); Emotional Reactions (9 items). Part II: Consists of 7 statements about aspects of daily life most often affected by health: paid employment, jobs around the house, social life, personal relationships, sex life, hobbies & interests, and holidays. *Part 2 is optional, and some items are not applicable to some groups such as the elderly (Bowling, 1997).*
	All statements require a Yes or No response.
Scoring:	In Part I of the Nottingham Health Profile: Each Yes response is weighted (based on results from the original sample). Weights sum to 100 for each of the six domains. This represents the situation where someone answers "yes" to all questions for the particular domain.
	The scores range from 0 – 100. Thus, the higher the score obtained the greater the health problem. Scores are to be presented for each domain in profile form. The authors state that domain scores are not to be combined to produce an overall score although some users have reported this.
	A simpler scoring system is often used which counts the number of affirmative responses in each section and Jenkinson (1991) found that weighted and unweighted scores correlated so highly that he questioned the usefulness of the weights.
	In Part II of the Nottingham Health Profile: Yes responses are not weighted; they are just tabulated (0 or 1) for each question.
	Mean scores for each of the six domains in Part I and the seven question in Part II can be produced for groups of 20 people or more.

NB: Hunt, McEwen & McKenna (1986) make the point that responses to the seven questions are related to a person's perceived health problems.

> *"It should not be taken to mean that the individual has that area affected in the absence of a health problem" (Page 280).*

Developed for:

The Nottingham Health Profile (NHP) was designed for clinical and epidemiological research. It was developed to reflect lay rather than professional perceptions of health. In 1975, work began on collecting statements about the typical effects of ill-health. 2200 statements were collected from 700 members of the general public. These statements were then refined to reduce redundancy, colloquialisms and ambiguity to produce a list of 138. After testing with a number of diverse patient groups they NHP to produce a population survey tool of 38 statements. These statements were then weighted for severity using Thurstone's Method of Paired Comparisons ($n = 215$).

This extensive process of development culminated with the publication of an article by Hunt, McEwen & McKenna (1985) and the book *Measuring Health Status* by Hunt, McEwen & McKenna in 1986.

Normative Data:

Normative Data is provided in the Nottingham Health Profile Manual reproduced in the book by Hunt, McEwen and McKenna (1986).

Baum & Cooke (1989) conducted a major community health survey using the Nottingham Health Profile in South Australia ($n = 1528$). Kalucy & Baum (1992) and Crockett, Cranston, Moss & Alpers (1996) have also published Australian data on the NHP.

Clinical Data:

The NHP has been used with patients with a wide variety of conditions. These include arthritis, brain tumours, stroke, cardiac and orthopaedic conditions, and it has also been used in rehabilitation settings and with the elderly.

Some clinical studies are listed below:

Brain Tumors: Salo, Niemelae, Joukamaa & Koivukangas (2002).

Chronic Obstructive Pulmonary Disease (COPD): Tsukino, Nishimura, McKenna, Ikeda, Hajiro, Zhang et al. (2002).

Chronic Pain: Alamo, Moral & de Torres (2002).

Coronary Artery Bypass Graft Surgery: Lukkarinen & Hentinen (1997); Caine, Sharples & Wallwork (1999); and Skinner, Farrer, Albers, Neil & Adams (1999).

Coronary Artery Disease: Skinner, Albers, Hall & Adams (1995).

Cystic Fibrosis: Congleton, Hodson & Duncan-Skingle (1998).

Day Rehabilitation: Burch, Longbottom, McKay, Borland & Prevost (2000).

Diabetes: Keinanen-Kiukaanniemi, Ohinmaa, Pajunpaa & Koivukangas (1996).

General Practice: Hopton, Porter & Howie (1991).

Heart and Lung Transplantation: O'Brien, Banner, Gibson & Yacoub (1998).

Hip Replacement: Nilsson, Frazen, Carlsson & Onnerfalt (1994); Hayes, Leary, Gillespie, Pinder & Sher (2000); and Rissanen, Sogaard & Sintonen (2000).

Knee Replacement: Rissanen, Sogaard & Sintonen (2000).

Lung Cancer: Montazeri, Milroy, Hole, McEwen & Gillis (2001).

Lung Transplantation: TenVergert, Vermeulen, Geertsma, van Enckevort, De Boer, van der Bij et al. (2001).

Methadone Maintenance Treatment: Torrens, San, Martinez, Castillo, Domingo-Salvany & Alonso (1997).

Migrane: Jenkinson (1990).

Multiple Sclerosis: Sitzia, Haddrell, Rice-Oxley (1998).

Myositis: Chung, Mitchell, Houssien, Al-Mahrouki, Carr & Scott (2001).

Open Heart Surgery: Chocron, Tatou, Schjoth, Naja, Clement, Viel et al. (2000).

Parkinson's Disease: Sitzia, Haddrell, Rice-Oxley (1998); and Karlsen, Tandberg, Arsland & Larsen (2000).

Rheumatoid Arthritis: Houssien, McKenna & Scott (1997).

Snorers: Loth, Petruson, Wiren & Wihelmsen (1998).

Stroke: Ebrahim, Barer & Nouri (1986).

Zoster: Mauskopf, Austin, Dix & Berzon (1994).

Applications:

The Nottingham Health Profile can be administered by an interviewer or used in a postal survey.

Hunt, McEwen & McKenna (1985) outline an important design limitation of the Nottingham Health Profile. The statements in Part I "represent rather severe situations. It was found necessary to have such items in preference to less severe statements in order to avoid picking up large quantities of 'false positives' i.e. people who are feeling temporarily 'under the weather'. However, the severity of the items does mean that some individuals who are suffering discomfort may not show up on the NHP. 'Normal' populations or those with minor aliments may affirm few statements on some sections. This makes it difficult to compare their scores, or to be able to demonstrate change. 'Zero scorers' can not be shown to improve on the NHP sections, although in actuality they may be feeling better than on a previous occasion." (Page 280)

Kind & Carr-Hill (1987) found that almost half of their population sample (n = 1672) scored 0 on all 38 statements.

This issue of "zero scorers" affects the performance of the NHP in two ways:

(1) it produces score distributions which are highly skewed toward zero (this outcome can also be described as a floor effect); and

(2) it reduces the instrument's sensitivity to change.

This limits the applicability of the NHP to specified groups (e.g. the elderly and chronically ill).

The Nottingham Health Profile has been adapted for wheel-chair bound patients by Post, Gerritsen, van Leusen, Paping & Prevo (2001).

Spanish researchers have produced a shorter 22 item NHP by using Rasch Analysis.

In the SENECA study researcher's calculated the percentage of Yes or No responses to the NHP statements for each domain.

Geerlings, Twisk, Beekman, Deeg & van Tiburg (2002) focused on using the NHP Pain statements only.

RELIABILITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Essink-Bot et al. (1997) Jans et al. (1999) Franks & Moffatt (2001) Klevsgard et al. (2002)	Adequate	
Test – retest	Yes	Hunt, McEwen & McKenna (1985) Hunt, McEwen & McKenna (1986) Beaton et al. (1997) Bouchet et al. (2000) Bureau-Chalot et al. (2002) Demet et al. (2002)	Adequate	
Inter – rater	NA	NA	NA	The NHP is a self-report measure.

VALIDITY	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Hunt, McEwen & McKenna (1985) Hunt, McEwen & McKenna (1986) Kind & Gudex (1994) Mauskopf et al. (1994) Prieto et al. (1997) Beaton et al. (1996) Jenkinson et al. (1988) Klevsgard et al. (2002)	Adequate	

Correlation with other measures	Yes	Kind & Gudex (1994)	Adequate	
		Crockett et al. (1996)		
		Meyer-Rosberg et al. (2001)		
		Post et al. (2001)		
Construct	Yes	Hunt, McEwen & McKenna (1985)	Adequate	The NHP's weighting method for the 38 statements has been criticised by Jenkinson (1991) suggesting the use of dichotomous scoring or Likert methods.
		Hunt, McEwen & McKenna (1986)		
		Kind & Carr-Hill (1987)		
		Jenkinson et al. (1988)		Kind & Carr-Hill (1987) Plant et al. (1996) and Jans et al. (1999) have reported that the NHP produces highly skewed distributions. (This is not surprising when you consider the NHP's design)
		Jenkinson (1991)		
		Plant et al. (1996)		
		Jans et al. (1999)		
		Prieto & Alonso (2000)		Kind & Carr-Hill (1987) and Jenkinson, Fitzpatrick & Argyle (1988) have criticised the six domain structure of the NHP.
Criterion	Yes	Hunt, McEwen & McKenna (1985)	Adequate	
		Hunt, McEwen & McKenna (1986)		
		Doll et al. (1993)		

RESPONSIVENESS	Studies reported Yes/No	References	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Hopton et al. (1991)	Adequate	Hopton et al. (1991); VanderZee et al. (1996); Congleton et al. (1998); Brown et al. (2000); and Bouchet et al. (2000) report limitations with the NHP as a measure of change in health status.
		VanderZee et al. (1996)		
		Congleton et al. (1998)		
		Brown et al. (2000)		
		Bouchet et al. (2000)		
		Franks & Moffatt (2001)		
		Klevsgard et al. (2002)		
		Tsukino et al. (2002)		

Cultural Applicability and Cultural Adaptations:	The NHP has been used extensively in the United Kingdom and Europe (translated into Danish, Dutch, Finnish, French, German, Hungarian, Norwegian, Spanish and Swedish) but has had limited use in Australia.
Gender Appropriateness:	Age and sex norms have been produced in the original manual ($n = 2173$, random sample from the community). As a rule "women usually have higher scores than men and scores tend to rise with age." Baum & Cooke (1989) also provide age and sex data for their Australian community sample.
Age Appropriateness:	16 years and over (requiring a minimum reading age of 10 years).
Summary:	The NHP is short, inexpensive and easy to use. It has been criticised as it only provides a limited measure of function and some disabilities are not assessed including sensory defects, incontinence and eating problems (Bowling, 1997). It has a very skewed distribution and a floor effect so it may not measure minor improvements in health nor be sensitive to people with mild conditions.

The weights in the scoring system have been criticised and scoring anomalies have been detected (Jenkinson, 1991).

There has been some criticism that each section does not represent just one dimension and co-variation between items in different categories raises difficulties for interpretation (Bowling, 1997).

The Nottingham Health Profile is a recommended generic outcome measure of perceived health status. However, in light of a number of findings made in the scientific literature it is in need of a revision to further improve its applicability and psychometric performance.

References

Alamo, M. M., Moral, R. R., de Torres, L. A. P. (2002). Evaluation of a patient-centred approach in generalized musculoskeletal chronic pain/fibromyalgia patients in primary care. *Patient Education & Counseling*, *48*, 23-31.

Baum, F. E., & Cooke, R. D. (1989). Community-health needs assessment: use of the Nottingham health profile in an Australian study. *Medical Journal of Australia*, *150*, 581-590.

Beaton, D. E., Bombardier, C., & Hogg-Johnson, S. A. (1996). Measuring health in injured workers: a cross-sectional comparison of five generic health status instruments in workers with musculoskeletal injuries. *American Journal of Industrial Medicine*, *29*, 618-631.

Beaton, D. E., Hogg-Johnson, S., & Bombardier, C. (1997). Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *Journal of Clinical Epidemiology*, 50, 79-93.

Bouchet, C., Guillemin, F., Paul-Dauphin, A., & Briancon, S. (2000). Selection of quality-of-life measures for a prevention trial: a psychometric analysis. *Controlled Clinical Trials*, *21*, 30-43.

Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales*. 2nd Ed. Buckingham, UK: Open University Press.

Brown, N., Melville, M., Gray, D., Young, T., Skene, A. M., & Hampton, J. R. (2000). Comparison of the SF-36 health survey questionnaire with the Nottingham Health Profile in long-term survivors of myocardial infarction. *Journal of Public Health Medicine*, *22*, 167-175.

Burch, S., Longbottom, J., McKay, M., Borland, C., & Prevost, T. (2000). The Huntingdon Day Hospital Trial: secondary outcome measures. *Clinical Rehabilitation*, *14*, 447-453.

Bureau-Chalot, F., Novella, J. L., Jolly, D., Ankri, J., Guillemin, F., & Blanchard, F. (2002). Feasibility, acceptability and internal consistency reliability of the Nottingham Health Profile in dementia patients. *Gerontology*, *48*, 220-225.

Caine, N., Sharples, L. D., & Wallwork, J. (1999). Prospective study of health-related quality of life before and after coronary bypass grafting: outcome at five years. *Heart*, *81*, 347-351.

Chocron, S., Tatou Schjoth, B., Naja, G., Clement, F., Viel, J. F. et al. (2000). Perceived health status in

patients over 70 before and after open-heart operations. *Age & Aging*, *29*, 329-334.

Chung, Y. L., Mitchell, H. L., Houssien, D. A., Al-Mahrouki, H., Carr, A. J., & Scott, D. L. (2001). A comparative study of outcome in myositis and other musculokeletal disorders assessed using the Nottingham Health Profile. *Clinical & Experimental Rheumatology*, *19*, 447-450.

Congleton, J., Hodson, M. E., & Duncan-Skingle, F. (1998). Do Nottingham Health Profile scores change over time in cystic fibrosis? *Respiratory Medicine*, *92*, 268-272.

Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *Phamacoeconomics*, *17*, 13-35.

Crockett, A. J., Cranston, J. M., Moss, J. R., & Alpers, J. H. (1996). The MOS SF-36 health survey questionnaire in severe chronic airflow limitation: comparison with the Nottingham Health Profile. *Quality of Life Research*, *5*, 330-338.

Demet, K., Guillemin, F., Martinet, N., & Andre, J. M. (2002). Nottingham Health Profile: reliability in a sample of 542 subjects with major amputation of one or several limbs. *Prosthetics & Orthotics International*, *26*, 120-123.

Doll, H. A., Black, N. A., Flood, A. B., McPherson, K. (1993). Criterion validation of the Nottingham Health Profile: patient views of surgery for benign prostatic hypertrophy. *Social Science & Medicine*, *37*, 115-122.

Ebrahim, S., Barer, D., & Nouri, F. (1986). Use of the Nottingham Health Profile with patients after a stroke. *Journal of Epidemiology & Community Health*, *40*, 166-169.

Essink-Bot, M. L., Krabble, P. F., Bonsel, G. L., & Aaronson, N. K. (1997). An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Medical Care*, *35*, 522-537.

Franks, P. J., & Moffatt, C. J. (2001). Health related quality of life in patients with venous ulceration: use of the Nottingham Health Profile. *Quality of Life Research*, *10*, 693-700.

Geerlings, S. W., Twisk, J. W., Beekman, A. T., Deeg, D. J., & van Tiburg, W. (2002). Longitudinal relationship between pain and depression in older adults: sex, age and physical disability. *Social Psychiatry & Psychiatric Epidemiology*, *37*, 23-30.

Hayes, J. H., Leary, R., Gillespie, W. J., Pinder, I. M., & Sher, J. L. (2000). Are clinical and patient assessed outcomes affected by reducing length of hospital stay for total hip arthroplasty? *Journal of Arthroplasty*, *15*, 448-452.

Hopton, J. L., Porter, A. M., & Howie, J. G. (1991). A measure of perceived health in evaluating general practice: the Nottingham Health Profile. *Family Practice*, *8*, 253-260.

Houssien, D. A., McKenna, S. P., & Scott, D. L. (1997). The Nottingham Health Profile as measure of disease activity and outcome in rheumatoid arthritis. *British Journal of Rheumatology*, 36, 69-73.

Hunt, S. M., McEwen, J., McKenna, S. P. (1986). *Measuring Health Status*. London: Croom Helm Ltd.

Hunt, S. M., McEwen, J., McKenna, S. P. (1985). Measuring health status: a new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners*, *35*, 185-188.

Jans, M. P., Schellevis, F. G., & van Eijk, J. T. (1999). The Nottingham Health Profile: score distribution, internal consistency and validity in asthma and COPD patients. *Quality of Life Research*, *8*, 501-507.

Jenkinson, C. Fitzpatrick, R. & Argyle, M. (1988). The Nottingham Health Profile: An analysis of its sensitivity in differentiating illness groups. *Social Science & Medicine*, *27*, 1411-1414.

Jenkinson, C. Health status and mood state in a migraine sample. *International Journal of Social Psychiatry*, *36*, 42-48.

Jenkinson, C. (1991). Why are we weighting? A critical examination of the use of item weights in a health status measure. *Social Science & Medicine*, *32*, 1413-1416.

Kalucy, E., & Baum, F. (1992). The epidemiology of caring: The pattern in a South Australian suburban population. *Australian Journal on Ageing*, *11*, 3-8.

Karlsen, K. H., Tandberg, E., Arsland, D., & Larsen, J. P. (2000). Health related quality of life in Parkinson's disease: A prospective longitudinal study. *Journal of Neurology, Neurosurgery & Psychiatry*, *69*, 584-589.

Keinanen-Kiukaanniemi, S., Ohinmaa, A., Pajunpaa, H., & Koivukangas, P. (1996). Health related quality of life in diabetic patients measured by the Nottingham Health Profile. *Diabetic Medicine*, *13*, 382-388.

Kind, P., & Carr-Hill, R. (1987). The Nottingham Health Profile: a useful tool for epidemiologists? *Social Science & Medicine*, *25*, 905-910.

Kind, P., & Gudex, C. M. (1994). Measuring health status in the community: A comparison of methods. *Journal of Epidemiology & Community Health*, *48*, 86-91.

Klevsgard, R., Froberg, B. L., Risberg, B., & Hallberg, I. R. (2002). Nottingham Health Profile and the Short-Form 36 Health Survey questionnaires in patients with chronic lower limb ischemia: before and after revascularization. *Journal of Vascular Surgery*, *36*, 310-317.

Loth, S., Petruson, B., Wiren, L., & Wihelmsen, L. (1998). Evaluation of the quality of life of male snorers using the Nottingham Health Profile. *Acta Oto-Laryngolgica*, *118*, 723-727.

Lukkarinen, H., & Hentinen, M. (1997). Assessment of quality of life with the Nottingham Health Profile among patients with coronary heart disease. *Journal of Advanced Nursing*, *26*, 73-84.

Mauskopf, J., Austin, R., Dix, L., & Berzon, R. (1994). The Nottingham Health Profile as a measure of quality of life in zoster patients: convergent and discriminant validity. *Quality of Life Research*, *3*, 431-435.

McDowell, I. & Newell, C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd Ed. New York, Oxford University Press.

Meyer-Rosberg, K., Burckhardt, C. S., Huizar, K., Kvarnstroem, A., Nordfors, L. O, & Kristofferson, A. (2001). A comparison of the SF-36 and the Nottingham Health Profile in patients with chronic neuropathic pain. *European Journal of Pain*, *5*, 391-403.

Montazeri, A., Milroy, R., Hole, D., McEwen, J., & Gillis, C. R. (2001). Quality of life in lung cancer patients: as an important prognostic factor. *Lung Cancer*, *31*, 233-240.

Nilsson, L. T., Frazen, H., Carlsson, A. S., Onnerfalt, R. (1994). Early radiographic loosening impairs the function of a total hip replacement. The Nottingham Health Profile of 49 patients at five years. *Journal of Bone & Joint Surgery*, *76*, 235-239.

O'Brien, B. J., Banner, N. R., Gibson, S., & Yacoub, M. H. (1998). The Nottingham Health Profile as a measure of quality of life following combined heart and lung transplantation. *Journal of Epidemiology & Community Health,* 42, 232-234.

Plant, P., McEwan, J., & Prescott, K. (1996). Use of the Nottingham Health Profile to test the validity of census variables to proxy the need for health care. *Journal of Public Health Medicine*, *18*, 313-320.

Post, M. W., Gerritsen, J., Diederikst, J. P., (2001). DeWittet, L. P. Measuring health status of people who are wheel-chair dependent: validity of the Sickness Impact Profile 68 and the Nottingham Health Profile. *Disability & Rehabilitation*, *23*, 245-253.

Post, M. W., Gerritsen, J., van Leusen, N. D., Paping, M. A., & Prevo, A. J. (2001). Adapting the Nottingham Health Profile for use in people with severe physical disabilities. *Clinical Rehabilitation*, *15*, 103-110.

Prieto, L., Alonso, J., Ferrer, M., & Anto, J. M. (1997). Are results of the SF-36 health survey and the Nottingham Health Profile similar? A comparison in COPD patients. Quality of Life in COPD Study Group. *Journal of Clinical Epidemiology*, *50*, 463-473.

Prieto, L., Alonso, J., Lamarca, R., & Wright, B. D. (1998). Rasch measurement for reducing the items of the Nottingham Health Profile. *Journal of Outcome Measurement*, *2*, 285-301.

Prieto, L., & Alonso, J. (2000). Exploring health preferences in socio-demographic and health related groups through the paired comparison of the items of the Nottingham Health Profile. *Journal of Epidemiology & Community Health*, *54*, 537-543.

Rissanen, P., Sogaard, J., & Sintonen, H. (2000). Do QOL instruments agree? A comparison of the 15D (Health-Related Quality of Life) and NHP (Nottingham Health Profile) in hip and knee replacements. *International Journal of Technology Assessment in Health Care*, *16*, 696-705.

Salo, J., Niemelae, A., Joukamaa, M., & Koivukangas, J. (2002). Effect of brain tumour laterality on patients' quality of life. *Journal of Neurology, Neurosurgery & Psychiatry*, *72*, 373-377.

Schroll, M., Schlettwein, D., van Staveren, W., & Schlienger, J. L. (2002). Health related quality of life and physical performance. SENECA 1999. *Journal of Nutrition, Health & Aging*, *6*, 15-19.

Sitzia, J., Haddrell, V., & Rice-Oxley, M. (1998). Evaluation of a nurse-led multidisciplinary neurological rehabilitation programme using the Nottingham Health Profile. *Clinical Rehabilitation*, *12*, 389-394.

Skinner, J. S., Albers, C. J., Hall, R. J., & Adams, P. C. (1995). Comparison of Nottingham Health Profile (NHP) scores with exercise duration and measures of ischaemia during treadmill exercise testing in patients with coronary artery disease. *European Heart Journal*, *16*, 1561-1565.

Skinner, J. S., Farrer, M., Albers, C. J., Neil, H. A., & Adams, P. C. (1999). Patient-related outcomes five years after coronary artery bypass graft surgery. *QJM*, *92*, 87-96.

TenVergert, E. M., Vermeulen, K. M., Geertsma, A., van Enckevort, P. J., De Boer, W. J., van der Bij, W. et al. (2001). Quality of life before and after lung transplantation in patients with emphysema versus other indications. *Psychological Reports*, *89*, 707-717.

Torrens, M., San, L., Martinez, A., Castillo, C., Domingo-Salvany, A., & Alonso, J. (1997). Use of the Nottingham Health Profile for measuring health status of patients in methadone maintenance treatment. *Addiction*, *92*, 707-716.

Tsukino, M., Nishimura, K., McKenna, S. P., Ikeda, A., Hajiro, T., Zhang, M. et al. (2002). Change in generic and disease-specific health-related quality of life during a one-year period in patients with newly detected chronic obstructive pulmonary disease. *Respiration*, *69*, 513-520.

VanderZee, K. I., Sanderman, R., & Heyink, J. (1996). A comparison of two multidimensional measures of health status: the Nottingham Health Profile and the RAND 36-Item Health Survey 1.0. *Quality of Life Research*, *5*, 165-174.

Note: Some recent relevant abstracts can be found at phi.uhce.ox.ac.uk/phidb.html

Reporter: Nicholas Marosszeky, Research Psychologist.

Date of report: 12 August 2003.

Functional Measures

Barthel Index

Title:	Barthel Index (10 items).
Abbreviations:	None.
Author(s) Name:	Mahoney and Barthel (1965) created the Barthel Index nearly 40 years ago. It was formerly known as the Maryland Disability Index and was first published in 1958. A modification of the 10 item version and its scoring system (20 point) was produced by Collin et al. (1988). A 15 item version has also been produced by Granger et al. (1981) and there are also many other variants of this index. Use of the Collin et al. (1988) version of the 10 item scale, or the Fortinsky et al. (1981) 15-item Modified Barthel Index is recommended and the use of other variants is discouraged (McDowell and Newell, 1996).
Author(s) Address:	As Mahoney and Barthel created the index nearly 40 years ago, it is unlikely that either author is still available.
Supplied by:	Available in publications.
Cost:	None.
Administration Time:	2-5 minutes for a health professional, up to 10 minutes for self-administration. It can be used with direct observation, interview, and telephone and from medical records.
Training Requirements:	None.
Purpose:	To assess activities of daily living (ADL) functions of adult respondents. It was designed to monitor performance in chronic patients before and after treatment and to indicate the level of nursing care required.
Administration:	Clinician administered, can also be self-administered
Instrument Type:	Clinician/nursing administered functional activities rating scale and it can be self-administered.
Structure:	The Barthel Index in its original form (Mahoney & Barthel, 1965) and the Collin et al. (1988) modification consists of 10 items relating to Activities of Daily Living. Eight of the items are self-care items (feeding, transfers, personal toilet/grooming, toileting, bathing, dressing, bowel and bladder continence) and two of the items are related to mobility (walking on a level surface and ascending and descending stairs).
Scoring:	Each item is scored on a three-point scale and then these scores are weighted and summed to produce a score of between 0 and 100 in the original version or 0-20 in the Collin et al. (1988) version. The scale is administered by a trained observer and may take from 5 minutes to complete if based on reports but longer if based on direct clinical observations. Shah et al. (1989) and Sansoni (2004, private communication) have produced versions that have each item scored on a five-point scale and these scores are summed to produce a total score between 10 and 50. Such versions might be considered if finer calibration for the level of functional limitation is required.
Developed for:	Adults.
Normative Data:	Extensive clinical data including from Australia. It forms part of the HACC minimum dataset (Eagar et al. 2001).
Clinical Data:	Amputees: O'Toole, Goldberg & Ryan (1985).
	Geriatrics: Stone, Herbert, Chrisostomou, Vessey & Horwood (1993).
	Hip Fracture: Cameron, Lyle & Quine (1993).
	Spinal Cord Injury: Roth, Lawler & Yarkony (1990).

Stroke: Wood-Dauphinee, Williams & Shaprio (1990).

Traumatic Brain Injury: Tuel, Presty, Meythaler, Heinemann & Katz (1992).

Applications: Rehabilitation outcome measure and therapy planning tool.

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Adequate	Cohen & Marino (2000); Sherwood, Morris, Mor & Gutkin (1977).
Test – retest	Yes	Adequate to good	Wade (1992); Collin, Wade, Davies & Horne (1988); Granger, Albrecht & Hamilton (1979).
Inter – rater	Yes	Good	Wade (1992); Granger et al. (1979).

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Adequate	Wade (1992); Cohen and Marino (2000).
Correlation with other measures	Yes	Adequate	Gresham, Phillips & Labi (1980); Gresham & Labi (1984); Basmajian (1994); Dittmar & Gresham (1997); Cohen and Marino (2000).
Construct	Yes	Adequate	Gresham & Labi (1984); Wade (1992); Basmajian (1994); Dittmar & Gresham (1997); Cohen and Marino (2000).
Criterion	Yes	Adequate	Wade & Collin (1988); Wade (1992); Cohen and Marino (2000).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Adequate	Granger et al. (1979); Cohen & Marino (2000).

Cultural Applicability and Cultural Adaptations: The index has been used in many countries and cultures, including French, German, Dutch, Japanese and Chinese (Dittmar & Gresham, 1997).

Gender Appropriateness: Suitable for males and females.

Age Appropriateness: Adults 14+.

Summary: The Barthel Index has been used in a wide range of studies of patients with incontinence problems. The Barthel Index is a standard tool in rehabilitation contexts and has excellent psychometric properties. It is simple to use. The original developers of the Barthel now advocate the use of the Functional Independence Measure, but the Barthel continues to be widely used.

References

Basmajian, J. (Ed.) (1994). *Physical Rehabilitation Outcome Measures*. Toronto, Ontario, Canada: Canadian Physiotherapy Association in Co-operation with Health and Welfare Canada and the Canadian Communications Group – Publishing, Supply & Services Canada.

Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales.* 2nd Ed. Buckingham, Open University Press.

Cameron, I.D., Lyle, D.M. & Quine, S. (1993). Accelerated rehabilitation after proximal femoral fracture: A randomised controlled trial. *Disability and Rehabilitation, 15*, 29 – 34.

Cohen, M., & Marino, M. (2000). The tools of disability outcomes: research functional status measures. *Archives of Physical Medicine and Rehabilitation, 81*(Supp. 2), S21-S29.

Collin, C., Wade, D.T., Davies, S., & Horne, V. (1988). The Barthel ADL index: a reliability study. *International Disability Studies*, *10*, 61-63.

Dittmar, S. S. & Gresham, G. E. (1997). *Functional Assessment and Outcome Measures for the Rehabilitation Health Professional*. Gaithersburg, Maryland, Aspen Publishers Inc.

Eagar, K et al. (2001). *Stage 1: Report of the HACC dependency data items project.* Wollongong, *CHSD*, University of Wollongong.

Fortinsky, R.H., Granger, C.V., Seltzer, G.B. (1981). The use of functional assessment in understanding home care needs. *Medical Care*. 19, 489-497.

Granger, C.V., Albrecht, G. L., & Hamilton, B. B. (1979). Outcome of comprehensive medical rehabilitation: measurement by PULSES profile and the Barthel Index. *Archives of Physical Medicine and Rehabilitation, 60,* 145-154.

Gresham, G. E. & Labi, M. L. C. (1984). Functional assessment instruments currently available for documenting outcomes in rehabilitation medicine, in C.V. Granger & G.E. Gresham (Eds). *Functional Assessment in Rehabilitation Medicine*. Baltimore: Williams & Wilkins.

Gresham, G. E., Phillips, T. F. & Labi, M. L. (1980). ADL status in stroke: Relative merits of three standard indexes. *Archives of Physical Medicine and Rehabilitation. 61*, 355 – 358.

Mahoney, F., & Barthel, D. (1965). Functional evaluation: the Barthel Index. *Maryland State Medical Journal, 14*, 61-65.

McDowell, I. & Newell, C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires.* 2nd Ed. New York, Oxford University Press.

O'Toole, D. M., Goldberg, R. T. & Ryan, B. (1985). Functional changes in vascular amputee patients: Evaluation by Barthel Index, PULSES Profile and ESCROW Scale. *Archives of Physical Medicine and Rehabilitation*, *66*, 508 – 511.

Roth, E. J., Lawler, M. H., & Yarkony, G. M. (1990). Traumatic Cord Syndrome: Clinical features and functional outcomes. *Archives of Physical Medicine and Rehabilitation*, *71*, 15 – 23.

Sherwood, S. J., Morris, J., Mor, V. & Gutkin, C. (1977). *Compendium of Measures for Describing and Assessing Long Term Care Populations.* Boston, MA: Hebrew Rehabilitation Center for the Aged.

Stone, S. P., Herbert P., Chrisostomou, J., Vessey, C. & Horwood, C. (1993). The assessment of disability in patients on an acute medical ward for the elderly. *Disability and Rehabilitation*, *15*, 35 – 37.

Tuel, S. M., Presty, S. K. Meythaler, J. M., Heinemann, A. W. & Katz, R.T. (1992). Functional improvement in severe head injury after readmission for rehabilitation. *Brain Injury*, *6*, 363 – 372.

Wade, D. T. (1992). *Measurement in Neurological Rehabilitation Medicine*. Oxford, England: Oxford University Press.

Wade, D. T. & Collin, C. (1988). The Barthel Index: A standard measure of physical disability. *International Disability Studies*, *10*, 64 – 67.

Wood-Dauphinee, S. L., Williams, J. J. & Shapiro, S. H. (1990). Examining outcome measures in a clinical study of stroke, *Stroke, 21*, 731 – 739.

Reporter: Shane Thomas and Jan Sansoni.

Date of report: October 2003.

Functional Independence Measure

Title:	Functional Independence Measure.
Abbreviations:	FIM.
Author(s) Name:	Carl V. Granger and Byron B. Hamilton, 1987.
Author(s) Address:	See the supplier and http://www.udsmr.org/adultfim.htm
Supplied by:	Australian Rehabilitation Outcomes Centre. 02 4221 4411
Cost:	Agencies must participate in FIM licensed network. The licensing arrangements for Australia are administered by Australian Rehabilitation Outcomes Centre (AROC). There are costs to the agency and also to the individual. AROC can provide a schedule of costs.
Administration Time:	1 hour for observations, 30 minutes to administer and score the scale for each patient.
Training Requirements:	Must be trained by FIM accredited trainer. Australia has 6 Master trainers.
Purpose:	Functional assessment of adults undergoing rehabilitation.
Administration:	By clinician, based upon observation of patient.
Instrument Type:	Clinician completed rating scale. (The FIM can also be completed by Multi-Disciplinary Team).
Structure:	The FIM includes 18 items covering including consideration independence in self-care, sphincter control, mobility, locomotion, communication and social cognition.
	The two sub-scales include physical (13 items) and cognitive (five items). These scales have been converted into interval measures by Rasch Analysis (Linacre, Heinemann, Wright, Granger & Hamilton, 1994).
	The FIM is now sometimes used in conjunction with a complementary tool, the Functional Assessment Measure that was developed by Hall et al. (1993).
Scoring:	The items are scored on a common 7 point scale ranging from 1 (total assistance required) to 7 (complete independence). The possible total score ranges from 18 (lowest) to 126 (highest level of performance).
Developed for:	Functional assessment of adults undergoing rehabilitation.
Normative Data:	The FIM forms part of the Uniform Data System for Medical Rehabilitation (UDS). Extensive clinical data is available including from Australia.
Clinical Data:	Australian Data – Marosszeky, Granger & Marosszeky (1995).
	U.S. Data – Granger & Hamilton (1992); Granger & Hamilton (1993); Granger & Hamilton (1994); Granger, Ottenbacher & Fiedler (1995) and Fiedler, Granger & Ottenbacher (1996).
Applications:	Data gathering in the United States has put the UDSMR SM in the position to develop "quality benchmarks" for health service providers (Fiedler & Granger, 1997).
	Margaret G. Stineman and colleagues have used the FIM to develop Functional Related Groups (FRGs) for major rehabilitation impairment categories (Stineman, 1995; Stineman, Hamilton, Goin, Granger, Fiedler, 1996).
	A WeeFIM for children and a PhoneFIM for telephone administration are also available.

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Good	Stineman, Shea, Jette, Tassoni, Ottenbacher , Fiedler et al. (1996).
Test – retest	Yes	Good	Ottenbacher, Hsu, Granger & Fiedler (1996); Kidd et al. (1995). ** Median Test – Retest Reliability of FIM Total = 0.95 * *
Inter – rater	Yes	Good	Ottenbacher, Hsu, Granger & Fiedler (1996); Hamilton, Laughlin, Fiedler, & Granger (1994).

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Good	Dodds, Martin, Stolov & Deyo (1993) (see also the construct validity section).
Correlation with other measures	Yes	Good	Kidd, Sewart, Baldry, Johnson, Rossiter, Petruckevitch et al. (1995) and Hall, Hamilton, Gordon & Zasler (1993).
Construct	Yes	Good	Heinemann, Linacre, Wright, Hamilton & Granger (1993).
Criterion	Yes	Good	Heinemann, Linacre, Wright, Hamilton & Granger (1994).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Good	Linacre et al. (1994); Heinemann, Hamilton, Linacre, Wright, Granger (1995); Cook, Smith & Truman (1994); Bohannon & Leary (1995). see also Ottenbacher, Hsu, Granger & Fiedler (1996).

Cultural Applicability and Cultural Adaptations: The FIM has been translated for use in many countries and languages.

Gender Appropriateness: Suitable for males and females.

Age Appropriateness: Adults 14+.

Summary: The Functional Independence Measure is a standard outcome measurement tool of long-standing and with demonstrated high levels of reliability and validity. It has been used in many studies of the effectiveness of rehabilitation interventions for patients with urinary incontinence. Because of its wide use, the FIM has excellent benchmark values across a wide variety of countries and patient groups. In view of these considerations, it is recommended for use by specialist practitioners and researchers as a general functional outcome measure. It is unlikely that primary care practitioners would require such sophisticated measurement in routine practice.

References

Bohannon, R. W., Leary, K. M. (1995). Standing balance and function over the course of acute rehabilitation. *Archives of Physical Medicine and Rehabilitation, 76*, 994 – 996.

Cook, L., Smith, D. S., Truman, G. (1994). Using Functional Independence Measure profiles as an index of outcome in the rehabilitation of brain-injured patients. *Archives of Physical Medicine and Rehabilitation, 75*, 390 – 393.

Dodds, R. A., Martin, D. P., Stolov, W. C., & Deyo, R. A. (1993). A validation of the Functional Independence Measure: A comparative validity and reliability study. *Disability and Rehabilitation, 74*, 531 – 536.

Fiedler, R. C., Granger, C. V. & Ottenbacher, K. J. (1996). The Uniform Data System for Medical Rehabilitation – report of first admissions for 1994. *American Journal of Physical Medicine and Rehabilitation*, *75*, 125 – 129.

Granger, C., Hamilton, B., Keith, R., Zielesky, M., & Sherwin, F. (1986). Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation, 1*(3), 59 – 74.

Granger, C. V., & Hamilton, B. B. (1992). UDS Report – The Uniform Data System for Medical Rehabilitation report of first admissions for 1990. *American Journal of Physical Medicine and Rehabilitation*, *71*, 108 – 113.

Granger, C. V., & Hamilton, B. B. (1993). The Uniform Data System for Medical Rehabilitation report of first admissions for 1991. *American Journal of Physical Medicine and Rehabilitation*, *72*, 33 – 38.

Granger, C. V., & Hamilton, B. B. (1994). The Uniform Data System for Medical Rehabilitation report of first admissions for 1992. *American Journal of Physical Medicine and Rehabilitation*, *73*, 51 – 55.

Granger, C. V., Ottenbacher, K. J. & Fiedler, R. C. (1995). The Uniform Data System for Medical Rehabilitation: Report of first admissions for 1993. *American Journal of Physical Medicine and Rehabilitation*, *74*, 62 – 66.

Hall, K. M., Hamilton, B. B., Gordon, W. A. & Zasler, N. D. (1993). Characteristics and comparisons of functional assessment indices: Disability Rating Scale, Functional Independence Measure and the Functional Assessment Measure. *Journal of Head Trauma Rehabilitation*, *75*, 390 – 393.

Hamilton, B., Granger, C., Sherwin, F., Zielesky, M., &Tashman, J. (1987). A uniform national data system for medical rehabilitation. In M. Fuhrer (Ed.), *Rehabilitation Outcomes: Analysis and Measurement* (pp. 137-147). Baltimore: Brooks.

Hamilton, B. B., Laughlin, J. A., Fiedler, R. C., Granger, C. V. (1994). Inter-rater reliability of the 7-Level Functional Independence Measure (FIM). *Scandanvian Journal of Rehabilitation Medicine*, *77*, 1226 – 1232.

Heinemann, A. W., Linacre, J. M., Wright, B. D., Hamilton, B. B., & Granger C. (1993). Relationships between impairment and physical disability as measured by the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation, 74*, 566 – 573.

Heinemann, A. W., Linacre, J. M., Wright, B. D., Hamilton, B. B., & Granger C. (1994). Prediction of rehabilitation outcomes with disability measures. *Archives of Physical Medicine and Rehabilitation, 75*, 133 – 143.

Heinemann, A. W., Hamilton, B., Linacre, J. M., Wright, B. D., & Granger C. (1995). Functional status and therapeutic intensity during inpatient rehabilitation. *American Journal of Physical Medicine and Rehabilitation*, *74*, 315 – 326.

Kidd, D., Stewart, G, Baldry, J., Johnson, J., Possiter, D., Petruckevitch, A. et al. (1995). The Functional Independence Measure: A comparative validity and reliability study. *Disability and Rehabilitation, 17*, 10 – 14.

Linacre, J. M., Heinemann, A. W., Wright, B. D., Granger, C. V., Hamilton, B. B. (1994). The structure and stability of the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation, 75*, 127 – 132.

Marosszeky, J. E., Granger, C. V., & Marosszeky, N. E. V. (1995). The uniform data system for medical rehabilitation: First admissions in Australia for 1992 – 93. *Health Outcomes Bulletin*, *6*, 29 – 32.

Ottenbacher, K., Hsu, Y., Granger, C., & Fiedler, R. (1996). The reliability of the functional independence measure: a quantitative review. *Archives of Physical Medicine and Rehabilitation, 77*, 1226 – 1232.

Ottenbacher, K. S., Mann, W. G., & Granger, C. V. (1994). Inter-rater agreement and stability of functional

assessment in the community based elderly. *Archives of Physical Medicine and Rehabilitation*, *75*, 1297-301.

Stineman, M. G. (1995). Case-mix measurement in medical rehabilitation. *Archives of Physical Medicine and Rehabilitation*, *76*, 1163 – 1170.

Stineman, M. G., Hamilton, B. B., Goin, J. E., Granger, C. V., & Fiedler, R. (1996). Functional gain and length of stay for major rehabilitation categories: Patterns revealed by Functional Related Groups. *American Journal of Physical Medicine and Rehabilitation*, *75*, 68 – 78.

Stineman, M. G., Shea, J. J., Jette, A., Tassoni, C. J., Ottenbacher, K. J., Fiedler, R. et al. (1996). The Functional Independence Measure: Tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Archives of Physical Medicine and Rehabilitation*, *77*, 1101 – 1108

Note: Relevant web sites include http://www.udsmr.org/adultfim.htm and tbims.org/combi/FIM/index.html

Reporter: Shane Thomas.

Date of report: December 2003.

Katz ADL Scale

Title:	Katz Index of Independence in Activities of Daily Living.
Author(s) Name:	Sidney Katz.
Author(s) Address:	Dr Katz is deceased.
Supplied by:	The tool is available through publications including McDowell, I. & Newell, C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires.* Oxford: Oxford University Press.
Cost:	Nil.
Administration Time:	Estimated 5-10 minutes.
Training Requirements:	Minimal.
Purpose:	To assess the physical functioning of elderly and chronically ill patients.
Administration:	Clinician administered rating scale of functional skills.
Instrument Type:	It is a rating scale based on observation and interview. The most dependent degree of performance during a 2-week period is recorded.
Structure:	The instrument assesses independence in six activities: bathing, dressing, toileting, transferring from bed to chair, continence and feeding. Each is rated on a 3-point scale of independence and descriptors are provided for each level reflecting the degree of dependence/independence.
Scoring:	Scoring involves translating the 3-point scales into a 'dependent/independent' classification. The patients overall performance is summarized on an eight-point scale that considers the numbers of areas of dependency and their relative importance. An alternative scoring scheme is also available.
Developed for:	To assess those with severe chronic illness and the elderly and to evaluate the effectiveness of their treatment.
Normative Data:	This is surprisngly limited given its wide clinical use (refer to Chen and Bryant (1975), McDowell and Newell (1996).
Clinical Data:	It has been used in studies of many conditions including cerebral palsy, strokes, multiple sclerosis, paraplegia, quadriplegia, and rheumatoid arthritis.
Applications:	It has also been used with children and with adults, with people with intellectual and physical disabilities, in the community and in institutions. The index is only appropriate with severely sick respondents as minor ailments or disabilities often do not translate into the limitations assessed by this scale. As with many such scales it has a floor effect and is insensitive to variations in low levels of disability (Eagar et al. 2001).

RELIABILITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Internal consistency	Yes	Adequate	Sonn & Asberg (1991); Brorsson & Asberg (1984).
Test – retest	Yes	Adequate	Katz, Ford, Moskowitz, Jackson & Jaffee (1963).
Inter – rater	Yes	Adequate	Katz, Ford, Moskowitz, Jackson & Jaffee (1963); Asberg & Sonn (1989); Dittmar & Gresham (1997).

VALIDITY	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Discriminatory power	Yes	Weak to adequate	Katz & Akpom (1976); Ashberg (1986).
Correlation with other measures	Yes	Adequate	Katz, Ford, Moskowitz, Jackson & Jaffee (1963); Dencker & Gottfries (1995); Dittmar & Gresham (1997).
Construct	Yes	Weak	Chen & Bryant (1975); Spector, Katz, Murphy & Fulton (1987).
Criterion	Yes	Weak	Chen & Bryant (1975); Asberg & Nydevik (1991).

RESPONSIVENESS	Studies reported Yes/No	Adequacy Weak/ Adequate/ Good	Comment
Sensitivity to change	Yes	Weak	Donaldson, Wagner & Gresham (1973); Gresham, Phillips & Labi (1980); Dittmar & Gresham (1997).

Cultural Applicability and Cultural Adaptations: This scale has been used in the Scandanivaian countries and German research, hence there are some translations available but not from a central source.

Gender Appropriateness: It has been used widely in the past with both male and female patients.

Age Appropriateness: The scale has been used with children, adults and older people.

Summary: Katz's Index was one of the first such scales published and was used widely in the past but now has been superseded by more recently developed instruments such as the Barthel and the FIM. Although it has been widely used in clinical settings there is surpisingly little information available on its psychometric properties and for this reason McDowell and Newell (1996) do not recommend its use. They prefer other scales because of their greater availability of published psychometric data.

References

Asberg, K.H. (1986). Change in ADL and use of short-term hospital care. *Scandinavian Journal of Social Medicine,* 14, 105 – 111.

Asberg, K. H., & Sonn, U. (1989). The cumulative structure of personal and instrumental ADL. A study of elderly people in a health service district. *Scandinavian Journal of Rehabilitation Medicine, 21*, 171 – 177.

Asberg, K. H. & Nydevik, I. (1991). Early prognosis of stroke outcome by means of Katz Index of Activities of Daily Living. *Scandinavian Journal of Rehabilitation Medicine, 23*, 187 – 191.

Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales.* 2nd Ed. Buckingham, Open University Press.

Brorsson, B. & Asberg, K. H. (1984). Katz index of activities in daily living: realibility and validity in short-term care, *Scandinavian Journal of Rehabilitation Medicine, 16*, 125-132.

Chen, M. K. & Bryant, B. E. (1975). The measurement of health: a critical and selective overview. *International Journal of Epidemiology, 4*, 257-264.

Dencker, K. & Gottfries, C. G. (1995). Activities of daily living ratings of older people using KATz's ADL Index and the GBS-M scale. *Scandinavian Journal of Caring Sciences, 9*(1), 35-40.

Dittmar, S. S. & Gresham, G. E. (1997). *Functional Assessment and Outcome Measures for the Rehabilitation Health Professional.* Gaithersburg, Maryland, Aspen Publishers Inc.

Donaldson, S. W., Wagner, C. C. & Gresham, G. E. (1973). A unified ADL evaluation form. *Archives of Physical Medicine and Rehabilitation*, *54*, 175 – 179.

Gresham, G. E., Phillips, T. F., & Labi, M. L. (1980). ADL status in stroke: Relative merits of three standard indexes. *Archives of Physical Medicine and Rehabilitation. 61*, 355 – 358.

Katz, S. & Akpom, C. A. (1976). A measure of primary sociobiological functions. *International Journal of Health Services*, *6*(6), 493-507.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffee, M. W. (1963). Studies of illness in the aged. The Index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association, 185*, 914-9.

McDowell, I. & Newell, C. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires.* 2nd Ed. New York, Oxford University Press.

Sonn, U. & Asberg, K. H. (1991). Assessment of activities of daily living: A study of a populationof 76 year olds living in Gothenburg, Sweden. *Scandinavian Journal of Rehabilitation Medicine*, *23*(4), 193-202.

Spector, W. D., Katz, S., Murphy, J. B. & Fulton, J. P. (1987). The hierarchical relationship between activities of daily living and instrumental activities of daily living. *Journal of Chronic Diseases*, *40*, 481 – 489.

Reporter: Shane Thomas.

Date of report: November, 2003.

Review of Patient Satisfaction Measures

2006

By
A/Professor Graeme Hawthorne

Background

Incontinence is a common health problem that affects over 2 million Australians of all ages and backgrounds. To address this important issue, the Commonwealth Government has resourced the National Continence Management Strategy (NCMS). Through the NCMS the Government aims to improve continence treatment and management so that more Australians can live and participate in their communities with dignity and confidence.

As part of the NCMS, a report on the possibility of a national suite of outcome measures to be used by Australian clinicians and researchers working in the continence field was commissioned, the Continence Outcomes Measurement Suite Project (1). One of the recommendations of this report was that there should be a separate report on patient satisfaction measures that could be considered for inclusion in the national suite of outcome measures.

Table of Contents

List of Tables

List of Figures

228

Glossary of Terms

ARSB	Acquiescent response set bias. This refers to the situation where a respondent provides biased answers to questions because he/she wishes to please the researchers.
Ceiling effect	Refers to scores on an instrument being "bunched" up at the top end of the scoring range.
Correlation	Describes the linear relationship between two variables, and is used in psychometrics as a test of validity. The conventional interpretation is that a correlation of <0.60 between two variables would indicate that they measuring different things; between 0.60-0.80 indicates they share something in common, but are not measuring the same thing; and correlations >0.80 imply the two measures are probably measuring the same thing. Correlations of >0.90 are needed before it can be asserted that the two measures are equivalent.
Coverage	Describes how well the descriptive system of a manifest instrument covers the latent construct of interest.
Cronbach _	Measure of the reliability of a scale, based on examining the internal consistency of responses to items forming the scale. Cronbach _ is based on both the correlations between items and the number of items within an instrument. Thus it is difficult to be precise about the desired _-value. For example, longer scales will have higher values than shorter scales. The conventions are that for comparison of groups _ should be within the range 0.70 to 0.90. For individual assessment (e.g. clinical diagnosis) the literature has suggested values in the range of 0.70 to 0.95. A discussion of the usefulness of _ in clinical decision-making can be found in Charter and Feldt (16).
Descriptive system	Refers to the actual items of an instrument and how these items are organized within an instrument.
Ecological validity	An instrument is described as possessing ecological validity when it was developed in consultation with a sample of the population with which the final instrument will be used. This is usually done through focus groups where the sample of respondents help with all phases of instrument construction.
Guttman scale	Describes a response scale where the responses progressively increase (e.g. *none, some, a lot, many*).
Homogenous scale	Describes a scale where all the items in the scale are measuring the same latent construct. Ideally, all scales should be homogenous as this minimizes measurement error. Homogeneity is usually tested using factor analysis, which groups items according to how well they are correlated.
Internal consistency	Describes the extent to which a scale is reliable. The most common method of testing for internal consistency is Cronbach _.
Instrument	An instrument is the formal language used to describe the descriptive system of a measure. It usually comprises several scales, each of which contains several items.
Item	Is the term used to describe a single question, where the psychometric properties of the question are known. A question has no formally known measurement properties. Items consist of two parts: the item stem is the question part, and the item response is the response part.
Kappa	A measure of the level of agreement between two observers.
Latent construct	Describes an object that doesn't exist but that is presumed to exist, such as love. In this report the latent construct of interest is patient satisfaction. A latent construct is defined by a theoretical model postulated by the researchers.
Likert scale	Describes a scale where the distance of responses from a mid-point indicate the strength of agreement or disagreement with a statement (e.g. the responses to the question: *You are satisfied with your treatment* might be: *strongly disagree/disagree/neither/agree/strongly agree*).

Manifest instrument	The descriptive system of an instrument that is used to represent a latent construct. It is the instrument that is administered to respondents.
Nomological net	Because validity is never established, researchers collect a variety of different types of validity evidence relating to an instrument. Where sufficient evidence is collected this is referred to as a nomological net of evidence (17).
Psychometrics	This is the discipline of measurement, where psychometric refers to the formal measurement properties of an item, scale or instrument.
Redundancy	Refers to items that are not needed in a scale, i.e. their presence does not contribute to the scale, and the scale is as reliable and valid with these items removed.
Reliability	Describes the stability of scale scores over time. A person who scores *X* on a scale should also score *X* on the scale if they complete the scale a second time. Reliability is usually assessed through correlation at test-retest, Cronbach _, or the correlation between half of a scale compared with the other half, administered at the same time (split-half reliability).
Response scale	Items often use a response scale on which the respondent selects the response that best describes his/her position. e.g. An item may ask *Do you leak urine?* and the response scale might be *Not at all, a little, some, a moderate amount, a lot.*
Responsiveness	Describes the sensitivity of a scale to differences in the underlying condition.
Scale	Refers to a collection of items that, between them, measure a construct. It is accepted that the items within a scale should be homogenous. Several scales may be included in an instrument.
SM	Scale maximum score.
VAS	Visual analog scale. Visual analog scales are usually presented as a line, a percentage scale (0-100) or a ladder (where the intervals are marked). In each of these, the respondent is asked to select a position on the VAS that best describes him/her.
Validity	Refers to evidence that suggests an instrument (or scale) measures what it is claimed to measure. Since validity is made up of two components – the properties of the descriptive system and the ability of the respondents – validity varies from sample to sample. Researchers therefore collect different types of validation evidence about an instrument; hence the "nomological net of evidence". Because respondents vary in their ability to answer questions (e.g. consider those who are continent compared with those who are incontinent), an instrument that has validity in one population sample may not be valid in another sample. Therefore validation exercises should be undertaken each time an instrument is used in a new study.

List of Abbreviations

AQoL   Assessment of Quality of Life (18, 19).

ConsultSQ   Consultation Satisfaction Questionnaire (10). Usually referred to as the CSQ, but in this report referred to as the Consult SQ to differentiate it from the CSQ-18 and CSQ-8.

CS   Convenience Scale (4).

CSQ-18   Client Satisfaction Questionnaire (9).

CSQ-8   Short version of the Client Satisfaction Questionnaire (9).

DCS   Doctor Conduct Scale (4).

GPAS   General Practice Assessment Survey (20).

GSS   General Satisfaction Scale (4).

GUTSS   Genito-Urinary Treatment Satisfaction Scale (8).

HMO   Health maintenance organisation.

LOPPS   La Monica-Oberst Patient Satisfaction Scale (11).

MISS   Medical Interview Satisfaction Scale (12, 13).

MSG   Multi-speciality practice.

PSI   Patient Satisfaction Index (14).

PSQ   Patient Satisfaction Questionnaire (3).

PVRQ   Patient Visit Rating Questionnaire (15), also referred to as the Medical Outcomes Trust patient satisfaction scale and the RAND 9-item patient satisfaction survey .

RAND-9   See PVRQ.

SOLO   Solo clinician practice.

TOP   Treatment Outcome Profile (21).

WHO   World Health Organisation.

Executive Summary

This review provides an overview of patient satisfaction within the context of incontinence. It's purpose is to recommend a patient satisfaction instrument to be used as part of the national dataset being developed under the aegis of the Australian National Continence Management Strategy (NCMS). Through the NCMS the Government aims to improve continence treatment and management so that more Australians can live and participate in their communities with dignity and confidence.

The role of and theories of patient satisfaction

Patient satisfaction has increased in popularity due to three changes in health care. First, the role of clinicians has changed from one of helping patients through their illness to where the clinician is expected to either cure the patient or alleviate chronic symptoms. Second, the rise of the patients' rights movement, which presents patients as consumers of health care, has led to patient views being taken into account during medical decision-making. Third, patient perspectives are increasingly sought for inclusion in the monitoring of health care and the legitimizing of health policy. This paper takes the position that health care recipients are patients rather than consumers because (a) most patients in Australia are not fully informed consumers and (b) this review is concerned with their personal health care satisfaction.

Despite this rise in research, there are conflicting definitions of patient satisfaction. It is defined here as the patient's judgement on the quality of care, particularly the interpersonal relationships with clinicians and other care providers (2).

The major patient satisfaction theories were all published during the 1980s; almost all research since then is based on these theories. Ware et al. (3) argued that patient satisfaction was a function of patients' subjective responses to experienced care mediated by personal preferences and expectations. Linder-Pelz (4) postulated it was mediated by personal beliefs and values about care as well as prior expectations of the care. Fox and Sturms (5) advocated that a person's orientation determined satisfaction; dissatisfaction, therefore, occurred where there was transgression of the relationship between expectation and experience. Fitzpatrick (6) argued that expectations were socially mediated, reflecting the health goals of the patient and the extent to which illness and health care violated the patient's personal sense of self. Finally, Donabedian (2, 7) postulated it was based on personal relationships within health care systems and health care outcomes from treatment, where these were mediated by the values of the patient.

Reviewing patient satisfaction instruments

Electronic searches of databases revealed 858 unique articles reporting patient satisfaction studies, of which 130 formed the key papers reviewed for this study. The Genito-Urinary Treatment Satisfaction Scale (GUTSS) (8) was the only incontinence specific patient satisfaction instrument.

Nine prominent generic patient satisfaction instruments met the study criteria for inclusion: two versions of the Client Satisfaction Questionnaire (the CSQ-18 and CSQ-8) (9); the Consultation Satisfaction Questionnaire (abbreviated to ConsultSQ) (10); the La Monica-Oberst patient satisfaction scale (LOPSS) (11); the Linder-Pelz Doctor Conduct, General Satisfaction and Convenience scales (4); the Medical Interview Satisfaction Scale (MISS) (12, 13); the Patient Satisfaction Index (PSI) (14); the Patient Satisfaction Questionnaire (PSQ) (3); and the Patient Visit Rating Questionnaire (PVRQ, also known as the RAND-9 survey) (15). Following review of each instrument, each was assessed against the study criteria (see below).

The first key finding was that most papers do not adequately report patient satisfaction. It is usually reported in a single sentence, where it is offered as evidence complementing treatment success. Few papers report either the instruments used, their psychometric properties in the study samples, or the actual results. This situation is unsatisfactory, and every effort should be made to rectify it if patients' views are to be incorporated alongside clinical indicators into health care decision-making and policy.

One of the most consistent findings across the literature reviewed is that most people are satisfied with their health care. Typically, between 70-90% of patients report satisfaction, even when there is evidence of continuing health problems. The reasons for this are primarily to do with health literacy, the unequal relationship between patients and their clinicians, instrument administration and bias, and that most people are satisfied with their lives generally.

The third key finding is that most studies report patient satisfaction based on a single item, such as *How satisfied are you with your health care?*. These kinds of items are short, quick and easy to administer.

They are widely used in clinical settings because they are easy to understand and interpret immediately, and they are frequently used by clinicians as discussion starters with patients. Almost no research, however, has been undertaken regarding their psychometric properties, and, since patients are usually in a dependent relationship with their clinician when responding to such questions, the value of the responses is extremely suspect.

Regarding the instruments reviewed, the criteria were: evidence of a theoretical model of patient satisfaction (a latent construct); parsimony (a surrogate for ease of use); adequate coverage of the latent construct; data distribution and ceiling effects; validity evidence; reliability and responsiveness evidence; and relevance in an Australian context. Each instrument was rated on each criterion on a scale of 0, 1 or 2, where 0 indicated the instrument did not meet the criterion, 1 indicated there was some evidence the instrument partly met the criterion, and 2 indicated the instrument met the criterion. Where insufficient information was available, no rating was given.

The criteria were weighted to reflect their relative importance. For example, validity evidence was deemed more important than being based on an explicit theory of patient satisfaction. The scale of weights used was 1, 2 and 3, where 1 indicated the criterion was important, 2 that it was very important and 3 that it was extremely important. Weighted scores on each criterion were then summed and divided by the number of criteria.

Results

A summary of the results is presented in the table. Although there were differences between the instruments, there was no "stand out" instrument and none of the instruments reviewed could be considered truly satisfactory. The main finding is that no instrument has been sufficiently validated for its use in Australia to be automatically recommended. There are three key reasons for this finding:

A. There is evidence throughout the literature that patient satisfaction is culturally specific. It cannot be assumed that an instrument that is relevant, valid and reliable in one culture retains those properties in another culture.

B. There is no agreed theoretical model of patient satisfaction or of its constituent parts. The consequence is that instrument designers have proceeded on an *ad hoc* basis with the result that there are thousands of patient satisfaction measures available.

C. Among recognised generic patient satisfaction instruments there is insufficient evidence of their psychometric properties for any instrument to be fully accepted as possessing a nomological net of validity evidence.

Summary of instruments

Criteria	Weight	CSQ-18	CSQ-8	Consult SQ	LOPPS-18	Linder-Pelz	MISS-21	PSI	PSQ-III	PVRQ	GUTSS
Theory & purpose (a)	1	1	1	2	1	2	2	1	2	2	1
Parsimony (b)	2	1	2	1	1	1	1	1	0	2	2
Coverage (c)	2	2	0	1	1	1	1	1	2	1	1
%SM (d)	1	1	1	1	0	—	0	1	—	—	1
Validity evidence (e)	3	1	1	2	1	—	0	1	2	1	1
Reliability evidence (f)	3	1	1	1	0	1	1	2	1	0	1
Aust. relevance (g)	3	1	1	1	0	0	1	1	0	0	1
Weighted total (h)		2.43	2.14	2.71	1.14	1.29	1.71	2.57	2.14	1.57	2.43

Notes:
a = Clinical = 2; Other = 1; None = 0.
b = <10 items = 2, 10 to 24 items = 1, 25 or more items = 0.
c = Seven dimensions measured =2, 4 to 6 dimensions = 1; 0-3 dimensions = 0.
d = Scale maximum score, a surrogate for data distribution and ceiling effects, <75%SM = 1; >75%SM = 0.
e = "Yes" = 2, 2-3 "Yes" = 1; 0-1 "Yes" = 0.
f = Reliability estimate in the range of 0.70 to 0.90 = 2; >0.90 = 1; < 0.70 = 0. If evidence of unresponsiveness, then 1 has been subtracted from the score.
g = Estimated for relevance to Australian clinicians, specialists and researchers.
h = Based on (Score X Weight)/N. categories.

Review of individual instruments

CSQ-18   Although its development followed good practice and it has good psychometric properties, too many items measure general satisfaction (30%). The distribution of item responses is poor (less than 10% of cases endorse the lower half of the scales). This lack of response distribution combined with the general satisfaction emphasis may explain the high Cronbach _ (0.91), which suggests redundancy. There are questions about its cultural relevance in an Australian context.

CSQ-8   The CSQ-8, while attractive because of its brevity, is restricted in its measurement. This is because 6/8 items are general satisfaction items. The removal of items from the CSQ-18 resulted in retaining items that were very similar to each other, hence explaining the very high Cronbach ⍺ of 0.93. The distribution of item responses is a concern.

ConsultSQ   The developmental procedures were excellent. It has good coverage measuring general satisfaction, professional care, relationship with the clinician and consultation time. There is validity and reliability evidence (the range of Cronbach _s = 0.73 to 0.94). However, the items are repetitious, and there is no item measuring treatment outcomes. The repetition implies that it gains its reliability (and perhaps some validity) through replication.

LOPPS   Although the LOPPS has been re-written several times, it is of limited value. There is evidence of acquiescent response bias, poor data distribution and ceiling effects, an inability to discriminate between groups known to vary by satisfaction, and a failure to be sensitive to dissatisfaction. The very high Cronbach ⍺s (0.93 to 0.98) reflect poor data distribution. Because it is primarily concerned with nursing care in an institution, it may have limited application in Australian general practice settings.

Linder-Pelz   These three scales were designed to test the expectancy theory of patient satisfaction, and there is insufficient psychometric information available. The reliability of the scales varies from 0.49 to 0.81. There are problems of negativity with the doctor conduct scale (all 10 items are negative in tone), and the scales may be less relevant in an Australian context because of the emphasis on entitlements and access.

MISS   The MISS has fair psychometric properties, including reasonable coverage. Its internal structure is poor, because two different research teams reported this was not congruent with the original descriptive system. There are redundant items and they present patients with a paternalistic perspective. There is some evidence that the MISS may be culturally specific. Cronbach ⍺ has been reported between 0.78-0.96.

PSI   The conceptual base and instrument construction steps were exemplary. It is a reliable instruments (the test-retest intraclass correlation coefficient was 0.86), although there is insufficient validity evidence. Because it was designed for patients with life-threatening conditions, the item content reflects experiences rather than satisfaction. Unlike the other instruments reviewed, the PSI was designed for interviewer administration.

PSQ-III   This was the most rigorously developed instrument reviewed. Its psychometric properties are very good, although some scales have poor reliability (the test-retest and Cronbach _ range is 0.23 to 0.93). The PSQ-III may reflect life satisfaction. It has 51 items: most clinicians and researchers would not have the time or resources to use it. Finally, because it emphasizes access to health care services and the capacity of the patient to afford health care, its validity in an Australian context is suspect.

PVRQ   This appears to be an excellent instrument, which was rigorously developed and tested. However, there are difficulties with the distribution of item responses and also with the responsiveness of the measure. Additionally, 4/9 items are concerned with access issues that may not be relevant in an Australian context.

GUTSS   This was based on a sound theory of patient satisfaction, and has good psychometric properties (e.g. the Cronbach _ = 0.83), although its coverage is less than ideal. Because it was developed to assess surgery satisfaction, whether it is applicable to all incontinence interventions is open to question. The structure of the GUTSS includes two filter questions, thus making it unnecessarily complex to use and score.

Recommendations

The recommendations are:

1. That a single item patient satisfaction measure should be adopted for use in Australian settings by clinicians wishing to assess the satisfaction of their patients "on the spot". Strategies should be put in place to encourage clinicians to adopt this measure as a common metric across Australia. Encouragement should be given to specialists and researchers to also include this common metric in their work. In this way a bank of shared understanding will be progressively established.

2. That if a continent-specific measure is required, a short, valid and reliable incontinence-specific patient satisfaction measure be developed for use by Australian specialists. Strategies should be put in place to encourage Australian researchers to adopt this measure as well. A suitable starting point would be revision of the GUTSS through removal of the two filter questions. This could be done quickly using the available GUTSS development database to model the effect of this change.

3. The better instruments identified in this report are the ConsultSQ, PSI and CSQ-18. For the reasons outlined above, however, none of these can be recommended in their current form. Therefore it is recommended that a short, valid and reliable generic patient satisfaction measure be developed for use by Australian researchers, or the revision of one or more of these instruments be undertaken. In the interim period it is recommended that the ConsultSQ be used.

1. Papers, theories and instruments reviewed in this study

To identify papers discussing or reporting patient satisfaction (as defined in section 2), searches of Medline/Pub Med and the internet were undertaken using the terms patient, client, or consumer and crossing these with any of the terms satisfaction, questionnaires, instrument, measurement or theory. In addition separate searches were made of the terms patient/client/consumer, satisfaction and theory/instrument.

Using the term "patient satisfaction", 38,193 articles were identified through Medline/Pub Med and over 10,000 websites in January 2004. Refinement of the search terms as described above led to the identification of 858 unique articles or reports and a further 126 websites. Abstracts (or, in the case of internet sites, first paragraphs) were reviewed. Based on an assessment of the contribution of the paper to the literature in a way not made elsewhere or providing a particularly good illustration of an issue of interest, unique articles and reports were obtained for close reading or critique. The reference lists were scanned for additional papers of interest. These were then extracted in turn. Altogether 130 unique articles were extracted, and the review is primarily based on these articles or reports.

In selecting instruments and papers for review, the following criteria were used.

1. Patient satisfaction theories had to be original and developed as a generic theory;
2. Patient satisfaction instruments had to be generic, that is designed for use in all health conditions, with all patients and across studies or research settings, other than incontinence specific measures;
3. Instruments had to be concerned with patient satisfaction assessment at the intervention level rather than the health care system level;
4. Papers must have been published in English;
5. They must have been accessible through the academic press or internet; and
6. Basic psychometric data must have been reported.

Theories, instruments and papers were excluded on the following grounds:

- Theories were excluded if they were elaborations of earlier theories, or if they had been developed for use with a specific condition. The review revealed that most of the modern theories or models of patient satisfaction are either restatements of earlier theories or have been developed for specific health conditions. For example, Hudak's (22) embodiment theory of patient satisfaction for those with hand surgery was excluded, as was Aragon's (23) primary provider theory of patient satisfaction for use in hospital emergency departments.

- Instruments measuring constructs other than patient satisfaction or that measured just one aspect of patient satisfaction were excluded. For example, the EUROPEP (24, 25) was excluded because it was designed to assess patients' views of their medical care, not satisfaction with that care. This exclusion draws the distinction between patients' cognitive awareness of care and their satisfaction with that care. Likewise the Patient Perception of Hospital Experience with Nursing (26) was excluded because it specifically focussed on nursing care within hospital settings.

- Instruments designed for use in single studies or for specific medical procedures. Thus the Surgery Satisfaction Questionnaire (27) was excluded. Likewise, Kane et al.'s (28) patient satisfaction measure was excluded because it was developed for a single study and a single type of medical procedure.

- Instruments designed for use with specific groups of patients were excluded. For example, the Older Patients Satisfaction Scale (29) was excluded because it was developed for use with a particular cohort, namely older adults.

- Instruments designed for use with patients suffering specific conditions (the exception, of course, being incontinence) were excluded. Thus the Verona Service Satisfaction Scale (30), while enjoying widespread international support (particularly across Europe), was excluded because it is concerned with the assessment of health care for those with mental health conditions. Margolis et al.'s (31) patient satisfaction scale was excluded because it was thought to be culturally limited (e.g. that males and females are separated in a medical clinic). Similarly, Westaway et al.'s (32) diabetes patient satisfaction scale was also excluded.

- Scales that were embedded within instruments were also excluded. For example, the Treatment Outcome Profile (TOP 21) is a 27-item measure designed for use with psychiatric patients to assess changes in quality of life, symptomatology and functioning level. A fourth scale of 9 items,

embedded within the instrument, measures patient satisfaction with services (the effectiveness of treatment, perceived competence of staff and the treatment environment). Another example would be the HIV/AIDS satisfaction questionnaire developed by Beck et al. (33).

A common feature of most of the articles reviewed was that although researchers reported patient satisfaction estimates, the actual measures used are very poorly reported, if at all. This situation applied regardless of whether the paper was written in the 1980s or since 2000; it applied to single-item measures as well as to formal instruments.

The psychometric properties of instruments used were particularly poorly reported — a finding consistent with that of Sitzia (34) who found, based on a review of 195 patient satisfaction papers published in 1994, that less than half reported any psychometric data, yet that 81% reported using a new patient satisfaction instrument and a further 10% reported modifying a previously existing instrument. He reported that most of the study instruments reviewed had little evidence of reliability or validity, and that of those papers reporting a new instrument, 60% reported no psychometric data whatsoever. For example, Johannsson et al. (35) in a long-term follow-up study of haemorrhoidectomy reported patient satisfaction and examined the relationship between continuing faecal symptoms and patient satisfaction. Yet nowhere in their paper was the patient satisfaction measure described, nor was there any reference to it so that it could be tracked down.

As reported by Sitzia (34), this situation is unacceptable research practice.

Note: The author was the principal researcher in the development of the Genito-Urinary Treatment Satisfaction Scale (GUTSS 8), which was the only incontinence specific measure identified. It is possible this affected the review of this measure.

2. Introduction to health care and patient satisfaction

Between about 1850 and 1950 there was a fundamental shift in the role of clinicians (36). Their role changed from being one of helping patients through their sickness (where the determinants of the outcomes were largely a function of the natural course of the condition) to one where the clinician was expected to either cure the patient or alleviate the symptoms of a chronic condition (where the determinants of the outcomes were perceived to be largely a function of the efficacy of the medical intervention, or the clinician's expertise).

Consequent upon this change in perceived role was accountability, first defined around 1900 as assessing the value of the care provided. Two broad areas of value assessment developed. First was the search to find ever better clinical outcomes through improved interventions (for a recent example see Campanella et al. (37)). This in turn gave rise, from the 1960s onwards, to the patient rights movement (38) – a movement that led directly to management concerns with service quality, and the assessment of that quality by those using the services; hence patient or consumer satisfaction. By the late 1960s, then, the debate over the relationship between patient satisfaction as an assessment of the value of the technical care versus the process of care was well established (39, 40).

At its most basic, whether at the technical or process level, patient satisfaction is, as observed by Carr-Hill (41), an outcome of the health care system specifically justified by health outcomes, the need for quality assurance, treatment evaluation and consumerism (42).

Patient satisfaction was initially perceived as being related to issues around access to medical infrastructure (43-47) and nursing care (48). Donabedian, with a focus on quality assessment, also saw it as arising from the medical infrastructure (the quality of amenities), but extended it to include the technical health outcomes from the treatment process and process quality which focussed on personal relationships within health care systems (2, 7). Importantly, he explicitly stated that patient satisfaction was not an indicator of the quality of technical care (7). This position was reiterated by Wilson & Goldschmidt (49) who separated patient satisfaction from patient outcomes on the basis that patient satisfaction was concerned only with the interpersonal aspects of hospital care. The corollary of these arguments was that patient satisfaction could be used as a surrogate indicator enabling the incorporation of the patient perspective into the process of monitoring and improving health care services or, even more broadly, as an evaluative tool in the formulation of social and health policy (38, 50) — even though patient satisfaction *per se* is not an evaluation of medical care (51).
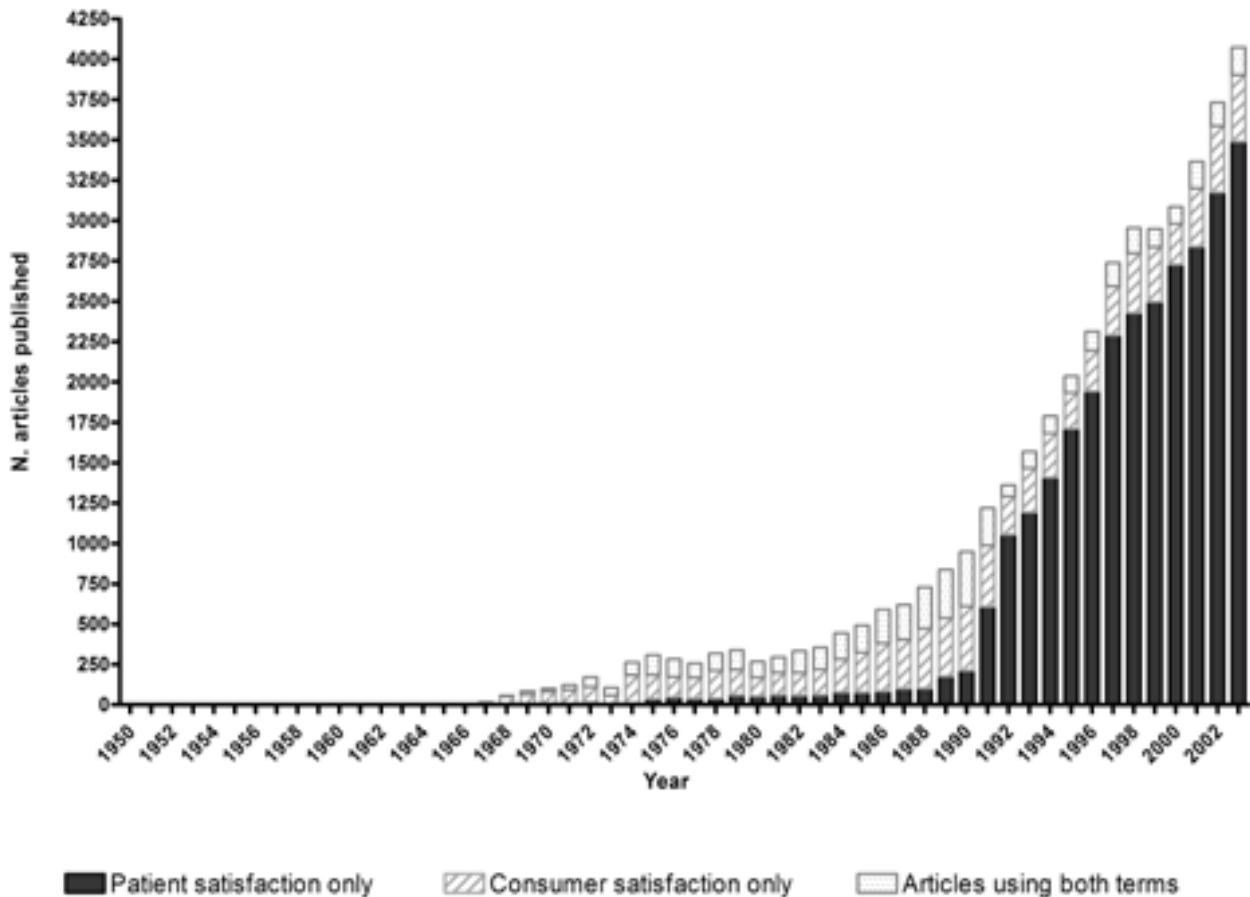
The rise in the use of patient satisfaction measures has also been justified on health outcome grounds, despite the lack of evidence showing that these measures are widely used in routine clinical practice or that they actually influence day-to-day clinical practice (52). These reasons for inclusion include that interventions with higher patient satisfaction outcomes are to be preferred, that satisfied patients are more likely to seek medical care, have greater compliance with treatment, continuing relationships with the clinician and have better health outcomes (10, 53-58).

Finally, patient satisfaction is also justified on the grounds that it enables patients to select health care clinicians, facilities or insurance plans: less satisfied patients are more likely to seek health care elsewhere (15, 59).

Figure 1 shows that since the early papers referred to above, there has been a continuous increase in the annual number of published academic papers cited in Medline; in 2003 there were over 4,000 papers cited. This increase in academic publication is more than matched by the number of internet websites. In January 2004, using the search term "patient satisfaction questionnaire", links to over 10,000 websites were identified. Most were links to hospitals, clinics or treatment centres, and many were primarily concerned with satisfaction with the environment within which health care was provided rather than with either medical care *per se* or the relationship between the clinician and the patient.

The rise in academic publication reflects the use of patient satisfaction as an indicator variable in monitoring and evaluating the delivery of care at the system level as well as an increasing interest in using it to assess the relationship between medical staff, treatment outcomes and patients. As noted by Valentine et al. (60) this rise is largely attributable to the rise of patient satisfaction measurement in the USA and the UK, often as either information for the marketing of a particular health service or in response to the consumer movement (42, 61, 62).

Figure 1: Journal articles by patient/consumer satisfaction, 1950-2003.

Patient satisfaction only · Consumer satisfaction only · Articles using both terms

Regarding the use of patient satisfaction at the system level, based on the tenets of quality assurance by the 1980s patient satisfaction had been widely incorporated into hospital and other health care services quality monitoring processes in the UK and the USA (50, 54, 63, 64). Within Australia, the systematized collection of patient satisfaction data can be traced to national evaluation policy. Since 1984 the objectives of Australian federal evaluation policy have emphasized efficiency, effectiveness and accountability. Patient satisfaction has been used as an indicator of accountability (65).

Importantly, the development and ongoing monitoring of the evaluation policy was located within the Department of Finance, thus giving rise to the importance of economic efficiency. Patient satisfaction was seen, therefore, as a surrogate indicator justifying and validating health care initiatives. Additionally, as the emphasis on the "new managerialism" increased in the 1990s patient satisfaction came to be seen as an accountability tool ensuring responsiveness to the patients' needs (66, 67). Simply, patient satisfaction has been a means of legitimising health policy and as a means of identifying for improvement those areas where dissatisfaction is expressed (54).

The result of the above socio-political context ensured that patient satisfaction has been increasingly seen as a key indicator assessing the quality of care within the Australian health care system (68, 69). Within this policy framework, it has become *de rigueur* for patient satisfaction data to be collected at both the system and treatment levels. For example, the routine collection of patient satisfaction data was introduced into Victorian hospitals along with casemix funding in 1993. In 2000 the Victorian Patient Satisfaction Monitor annual surveys were introduced with the express purpose of assessing patient satisfaction with key aspects of service delivery, the perceived strengths and weaknesses of health care services, to identify areas for quality improvement initiatives and to set appropriate benchmarks allowing hospitals to undertake comparative performance assessment (70-72).

Because these and other system-level measures are routinely administered to patients as part of hospital management programs, this level of patient satisfaction measurement is not specifically considered in this report even though many of the same issues regarding the definition of patient satisfaction and the operationalization of construct measurement are referred to.

Rather, this report is concerned with patient satisfaction assessment at the intervention level; i.e. at the level of a response to the relationship between the health care provider and the patient, and the

effect of an intervention for a particular condition. It is pertinent to note that research has shown that it is this relationship which is the key to understanding patient satisfaction. As pointed out by Locker & Dunt (51) over 20 years ago, interest in patient satisfaction at this level reflects, among other things, the continuing shift from acute to chronic conditions. This shift causes many people to live their lives in states requiring continual medical care.

The result has been the surge in measuring patient interest shown in Figure 1. As well as the system quality of health care issue discussed above, other factors contributing to this have been concerns with patient expectations, the face-to-face interaction between health care provider and patient, the provision of appropriate condition knowledge, and treatment compliance and treatment outcomes (53, 73-75).

3. Defining patient satisfaction

Like other constructs, defining patient satisfaction and measuring it is problematic. A key reason is that people vary in what they perceive to be a health condition worthy of health care and in their health literacy (76, 77). There is also variation in what people want from the health care system, and what they expect from any particular encounter with health care providers (5, 51). These differences operate at both the individual as well as at the societal level with the consequence that there may be differences of such magnitude as to render what is important in one context relatively unimportant in another. For example, issues of patient satisfaction with health care costs will be important in systems where health care insurance (in whatever form) is purchased by the individual. Where purchasing is at the state level satisfaction with health care costs may be deemed irrelevant by patients. Against this, much of the research on patient satisfaction is from the USA. Given that the structure, operation and financing of the health care system in the USA is systematically different to that of many other countries, caution needs to be exercised in assuming generalizability of theories, instruments and studies. At least six recent studies have shown that it cannot be assumed that a patient satisfaction scale developed in one country is directly transferable to another country without modification (78-82), and Baker et al. (83) reported that patient satisfaction was lower among US patients when compared with UK patients. This caveat should be borne in mind when interpreting this report.

Although there are many general definitions of patient satisfaction, there is widespread agreement that the literature suffers from inadequate conceptualization of the construct; a situation that has not changed much since the 1970s (41, 50, 51). One consequence of this situation is that there are many competing models of patient satisfaction (42), the key ones of which are reviewed below.

3.1 Whose satisfaction: patients or consumers?

As shown in Figure 1, both the terms "patient" and "consumer" satisfaction are used in the literature. It also shows, however, that until the 1990s, most papers used the term "consumer" or used both "consumer" and "patient". Since the early 1990s the term "patient satisfaction" has come to dominate the literature, possibly because there has been a relative shift in the number of papers concerned with health service quality at the systemic level when compared with the number of papers reporting satisfaction with treatment at the clinical level.

It is important to realise that although there has been this shift, in terms of actual numbers of published papers there has been no decline in the use of the term "consumer"; indeed in terms of raw numbers between 1980/1 and 2002/3 the annual number of papers published using the term "consumer" doubled. Many of the early papers which used the term "consumer" were studies of access to health services (in terms of the location of services), the role of health services in communities, or the organisation of health service provision (e.g. see 45, 46, 47). A second wave of papers using the term "consumer" originated in the consumer movement of the 1960s which vigorously championed the rights of consumers and the accountability of professionals (51). More recently, there has been a shift to include health system responsiveness based on the World Health Organisation's (WHO 84) report that the patient experience included the relationship between patients (as a population) and health systems. Valentine et al. (60) have defined this responsiveness as how people are treated within the health care system, and the environment within which they are treated. An axiom of responsiveness is that it can have direct effects on health outcomes through making access to the health care system easier, through providing greater openness and through enabling consumers to better assimilate health information. An important feature of responsiveness is that it is independent of health treatment outcomes (which can be measured by the capture of direct health changes).

This situation can be contrasted with the bulk of papers using the term "patient". In the early literature, the concerns of these papers were primarily with the patient-doctor relationship, patient views on hospital infrastructure, information provision, and service provision satisfaction (e.g. see 74, 85, 86-88). As noted by Hudak et al. (89), the emphasis in patient-oriented papers has recently become one where patient satisfaction is perceived to be a desirable outcome from the patient-clinician encounter, primarily defined as personal satisfaction with the process of care and the outcomes of that care.

In line with these broad differences in the published literature, Carr-Hill noted that the term "consumer" originated in the private sector under the aegis of consumer rights, and that it was largely at the system level (e.g. by accountants) that the term was first applied to medicine (41, 51). In contrasting the principles of consumer rights with the situation in health care, Carr-Hill (41) reported that there were clear mismatches relating to issues of access, choice, information, redress, safety, value for money and equity — mismatches that led health economists to simply refer to these from the 1960s onwards

as "market failure" (90-92). Owens and Batchelor (93) examined the extent to which elderly patients could be considered consumers and concluded that they couldn't in the normal sense of the word, since they were often uninformed, had few expectations and did not behave like consumers. Sitzia and Wood (50) provide a concise review of the terms, noting that the consumer rights and quality management movements use "consumer" and "customer", whereas "patient" is preferred by the medical establishment (possibly because the term respects the knowledge and power advantage of the clinician). Use of "consumer", then, may carry with it a clear signal of the changing relationship between the clinician and patient, where patients are perceived as collaborators with clinicians in the production of health (50, 94). Hudak et al. (89), however, in examining the assumptions behind the terms "consumer" and "patient", pointed out that consumers were in a business relationship with salespersons where the relationship was governed by the monetary value of objects. They contrasted this with "patient", where the relationship was based on the endeavours of the clinician to improve the patient's health in a collaborative relationship.

Consistent with these arguments, this paper describes health care recipients as patients because this review is concerned with their personal satisfaction with their health care (3).

3.2   Defining patient satisfaction

As foreshadowed above, there is no agreed definition of patient satisfaction in the literature; a situation that has existed since the beginning of its measurement (50, 51). Generally, rather than pursue adequate definitions of patient satisfaction, researchers have reported the measurement of patient satisfaction itself and its correlates (e.g. see 41, 95-97). For example, in one of the early seminal review papers, Locker and Dunt (51, p283) avoided defining patient satisfaction; the closest they got was that ... *the quality of care can become synonymous with the quality of life and satisfaction with care an important component of life satisfaction.* Equally vague was the definition of Carr-Hill (41, 241) that satisfaction was *the extent to which aspirations are met given self-perceived status.* Rather more pointedly Beck et al. (33, 332) quote Fitzpatrick's 1993 definition that patient satisfaction is a *cognitive evaluation of an emotional reaction to healthcare.*

Rather than develop a definition of the construct itself, many have defined patient satisfaction in terms of treatment success (e.g. see 98, 99, 100) or by the areas measured in patient satisfaction instruments (50). For example, Linder-Pelz (4), based on a review of the literature, defined the determinants of patient satisfaction as comprising expectations, value, entitlement, occurrences and interpersonal comparisons.

Almost certainly this failure to define the construct is one of the key reasons patient satisfaction has been roundly criticized for being ambiguous and unsatisfactory (4, 60, 101). As pointed out by Donabedian, *...to proceed to measurement without a firm foundation of prior agreement on what quality consists in is to court disaster* (2, p1743). Williams (38) was equally scathing when he noted that researchers wishing to measure patient satisfaction often simply assumed that "it" existed in the population and was awaiting measurement; he argued that because "it" was not properly defined and ecologically verified many patient satisfaction surveys did not measure the true experiences of patients.

In 1982 Linder-Pelz (4, p578) defined patient satisfaction as the patient's *...positive evaluations of distinct dimensions of their health care.* This definition explicitly argues that patient satisfaction must capture the patient's perspective. Ware et al. (3) advanced a definition that although similar, was uncompromising about this perspective in that patient satisfaction was

> *...a personal evaluation of health care services and providers... Satisfaction ratings are intentionally more subjective [than care reports]; they attempt to capture a personal evaluation that cannot be known by observing care directly... this is their unique strength. (3, p247)*

Apart from the emphasis on the patient perspective, this is an unformed definition; essentially Ware offered a generalized statement that provided little guidance regarding its precise meaning. A more refined definition was advanced by Donabedian:

> *Patient satisfaction may be considered to be one of the desired outcomes of care, even an element in health status itself. An expression of satisfaction or dissatisfaction is also the patient's judgement on the quality of care in all its aspects, but particularly as concerns the interpersonal process. (2, p1746)*

Like the definitions above, this states that it is the patient's subjective perspective that is central to patient satisfaction. Unlike Linder-Pelz and Ware, however, Donabedian was quite precise in defining

the principal target of the patient's satisfaction — the interpersonal processes of health care. Few others, however, have been this precise; indeed few others have attempted to define the construct at all. Generally, where definitions have been advanced, they explicitly include the patient's perspective but have referred to the various aspects of care rather than emphasizing the interpersonal process. For example, Goldstein et al. (96, p854) stated that patient satisfaction was *...a health care recipient's reaction to aspects of the service delivered and satisfaction over time which result in overall perceptions of quality of service.*

Many researchers have simply accepted Donabedian's or Linder-Pelz's definitions (e.g. see 32, 102). As shown above, however, there are distinct differences between Linder-Pelz's, Ware's and Donabedian's definitions, with Donabedian's being both the most clearly elucidated and (implicitly, since it has such a strong emphasis on interpersonal relationships, which are the most important aspects of care quality according to patients (103)) the most widely accepted.

# 4.	Theories of patient satisfaction

The key theories of generic patient satisfaction were all published in the 1980s. Five key models were advanced; these share several similarities which can contribute towards an overall theory of patient satisfaction.

Although there are more recent theories of patient satisfaction, for the most part these are restatements of the principles laid out in the 1980s and have therefore been excluded as outlined in section 1.1.

## 4.1	Ware et al. (1982): Determinants and components of patient satisfaction

Drawing on their work carried out in the 1970s, Ware et al. (3) proposed a subjective model of patient satisfaction which was the conceptual basis of the Patient Satisfaction Questionnaire. Their model was primarily derived through a content analysis of previous literature. At the heart of this model lay the distinction between patient satisfaction ratings and reports. Ratings were the subjective responses of patients whereas reports were perceived to be more objective and factual accounts of care. The reason for drawing this distinction was that the subjective component was based on personal preferences, expectations and the experienced care. The first two of these (personal preferences and expectations) formed a patient's attitude towards their experienced care. The argument was that patient satisfaction provided information both about the experienced care and also about the patient who provided that rating.

It was argued that where dissatisfaction occurred it was possible that this implied the need for a change in either the medical care or in the patient (e. g. through an increase in health literacy). Based on their empirical findings, however, the within-patient characteristics were reported to be small relative to the within-care component. Sitzia and Wood (50) referred to the within-patient component of Ware et al.'s work as the *determinants of satisfaction* and the medical care aspect as the *components of satisfaction.*

Regarding the components of care, Ware et al. (3) provided a taxonomy for classifying the content (the dimensions of care) which should be measured, and they asserted that patients would develop attitudes towards each dimension. The dimensions were:

- Interpersonal manner; i.e. how clinicians interacted with their patients;
- Technical quality; i.e. the competence and care standards of the clinician;
- Accessibility/convenience; i.e. issues in arranging to receive medical care;
- Finances; i.e. payments for medical care;
- Efficacy/outcomes; i.e. the helpfulness of clinicians in improving or maintaining health;
- Continuity; i.e. provision of care through the same clinician or at the same location;
- Physical environment; i.e. the physical setting in which care is delivered; and
- Availability; i.e. the presence of medical resources within the community.

Further work suggested that these factors could be collapsed into four key vectors:

- Conduct of the physician;
- The availability of services;
- Continuity/Convenience; and
- The cost of services.

## 4.2	Linder-Pelz (1982): Expectancy-value theory of patient satisfaction

Consistent with Ware et al. (3), Linder-Pelz (4) also defined patient satisfaction as patients' attitudes towards their health care. But, in doing so she drew on a variety of psychological theories.

At the core of her work was Fishbein and Ajzen's *theory of reasoned action* (104) for her conceptual model of attitudes, specifically their expectancy-value theory. Essentially, for Fishbein & Ajzen the term attitude was conceptualized as the amount of affect for or against some object whereas beliefs represented the information a person has about an object. Under expectancy-value theory, expectations were defined as the beliefs that an action will have consequences which, in turn, will have positive or negative valence (affect). From this relationship, Fishbein & Ajzen derived their claim that *A person's attitude toward an object is related to his beliefs that the object possesses certain attributes and his evaluation of those attributes* (104, p59). Linder-Pelz, drawing on Vroom's work in job satisfaction (105), argued that expectancy (the evaluation of object attributes) was determined by the relationship between

the importance of an outcome and the perceived chances of that outcome being achieved. Linder-Pelz reported that, based on the empirical data she collected and analysed, there was no support for the Fishbein and Ajzen model that attitudes are determined by the interaction of beliefs (expectations) and values. She did, however, report that both expectations and values were independent predictors of satisfaction (95).

This fitted very neatly with Lawler's discrepancy, fulfilment and equity models of pay satisfaction (106). Under discrepancy theory, satisfaction is determined by the relationship between what a person desires and what occurs. Fulfilment theory, on the other hand, defines satisfaction as the difference between the rewards desired and those obtained. Equity theory posits that satisfaction results from evaluation of one's own position against that of others. A critical issue arising from this exposition was in relation to explaining discrepancies between social groups. For this, Linder-Pelz drew on the work of the relative deprivation theorists (107, 108), and argued that dissatisfaction would arise where there was perceived discrepancy with the relevant social group. For explaining individual intrapersonal comparisons, Linder-Pelz drew on Festinger's theory of social comparison (109), whereby satisfaction is obtained through comparison with others mediated by the cultural setting. Thus she cited Thibault and Kelley's proposal that people evaluate circumstances in relation to those they believe others achieve or in relation to those they have themselves experienced in the past (110).

Drawing these theories together, Linder-Pelz proposed that patient satisfaction was related to the sum of the products of beliefs and valuations regarding various aspects of care; that this was subject to the extent to which perceived occurrences concurred with prior expectations and that the relationship between these could be defined as the perceived occurrence score minus the expectation score, divided by the expectation score. She further proposed that this was subject to the valuing of the object (i.e. satisfaction would only occur where an object was valued).

Regarding actual satisfaction, Linder-Pelz postulated that there were two conditions under which high satisfaction would be reported: (a) where positive expectations and positive experiences coincided, and (b) where experiences were perceived to be as good as or better than those of others. Dissatisfaction would occur where there were positive expectations and negative experiences.

From this theory, Linder-Pelz argued that the determinants of patient satisfaction were:
- Expectations; i. e. beliefs about an object;
- Value; i.e. a person's attitude towards an object;
- Entitlement; i.e. the belief held by an individual that he/she has proper and accepted grounds for claiming a particular outcome;
- Occurrences; i.e. the perception of what actually occurred during an encounter with the health care system, at whatever level; and
- Interpersonal comparisons; i.e. the comparison with others or with other encounters.

Under the expectancy model of satisfaction, dissatisfaction could occur where a patient's expectations are not met (for whatever reason). A limitation, of course, is the assumption that patients have formed expectations; it may be that where there are unformed expectations the experience of health care forms expectations.

Since publication, the expectancy-value theory of patient satisfaction has been widely cited and used in the construction or review of patient satisfaction measures, e.g. see Crow et al. (42).

4.3   Fox and Storms (1981): Discrepancy and transgression theories

The genesis of the discrepancy theory of patient satisfaction lay in the observation that different studies reported contradictory findings for the relationship between socio-demographic variables and patient satisfaction. To explain this Fox and Storms (5) drew on the work of Suchman (76) who had argued that cultural differences would predispose individuals' beliefs and attitudes concerning their reactions to illness, their health care knowledge, and their scepticism of health care. Fox and Storms described this as a person's orientation. They illustrated this with an example of dysentery prevention; a person who believes in allopathic medicine would boil water before drinking it, whereas a person with a humoral disease model of health would argue adding heat is counterproductive in treating a hot disease. The second assumption of Fox and Storms was that there were different medical treatments available for the same condition and that which treatment was provided and how it was provided was a function of different health providers' approach to care.

They postulated that where a person's orientation matched that of the health care provider there would be high satisfaction, but that where there was mismatch dissatisfaction would follow. Regarding the components of the orientation model, Fox & Storms reported they were:

- Knowledge about the health condition of interest (based on Suchman, they reported this concerned disease etiology, symptoms and prognosis);
- Beliefs about care (theories of the condition etiology and how it should be treated);
- The conditions of care as defined above.

Others have developed discrepancy theory further because patient satisfaction surveys consistently show that most people are satisfied with their health care (see below, section 5). Carr-Hill (41) argued that because patient satisfaction was a derived concept (derived from receiving a health service), there was a power relationship between the clinician and patient which confounded attempts to measure satisfaction:

> *Crudely, in situations where patients have, or perceive themselves to have, more control, they are more likely to pursue their own goals; where patients see themselves as powerless, then expectations will be redefined to match the probable outcome. Goals (expectations, aspirations) cannot, therefore, be measured in a vacuum; they have to be situated in the context of the structural relationship between the patient and the health care agents. These additional complexities mean that those who set out to "measure satisfaction" are probably on a hopeless quest... (41, p242)*

He recommended that researchers should search for sources of dissatisfaction: primarily the mismatch between expectation and experience. For example, Linder-Pelz (95) reported that discrepancy was strongest where there was a mismatch between expectations and occurrences which resulted in a negative correlation with satisfaction (for an example see Bramadat and Driedger (111)). However, others have shown that patient satisfaction is more a reflection of current health than expectations; dissatisfaction is particularly likely where there are major current health complications (112) or treatment failure (113-119); indeed Kane et al. (28, 120) showed that satisfaction was a function of a clutch of factors. Based on factor analysis, Hardy et al. (55) identified the key components:

- The proximal components of patient satisfaction, referring to the process of care (general view of the hospital, quality of care, and overall satisfaction with care), physical health (understanding of the illness of interest, illness and general health management and health improvement) and psychological well-being (feelings of anxiety, fearfulness, loneliness, loss of control, and the empathy of the clinical and nursing staff);
- Organisation components of satisfaction, including socialisation (admission procedures, quality of provided information, staff helpfulness), participation (information provision, participation in decision-making, interaction with care staff) and the hospital facilities.

Regarding the relative effects of these factors, Hardy et al. reported that based on regression analysis, of the three proximal components, patient satisfaction was most strongly associated with the process of care (r^2_{adj} = 0.50), then health (0.33) and psychological wellbeing (0.21). Consistent with this, Dozier et al. (26) argued that patient satisfaction with nursing care reflected nothing more than meeting the immediate needs of patients.

The obverse is that dissatisfaction results from a cluster of failures or where a critical event occurs (38, 42, 101, 121). This research has given rise to theories of patient satisfaction where dissatisfaction is reported only where multiple roles are transgressed (e.g. expectations of health care, the role of the clinician, the provision of information, treatment outcomes, and the patient's own understanding of his/role as a consumer of health services). Williams (38), for instance, argued that dissatisfaction was a function of patients applying their social values and norms to encounters with clinicians; dissatisfaction occurred where these norms were transgressed. Hawthorne and Harmer (8) postulated a dual transgression model of patient satisfaction, which was where patient satisfaction was a function of health status following an intervention and the role perception of the patient (defined as the health care worker both providing information and treatment, and the role of the patient as a health care consumer). For dissatisfaction to be expressed, they postulated that transgression of both these factors was necessary.

4.4 Fitzpatrick (1984): Multiple models as determinants of patient satisfaction

Generally, the models of patient satisfaction above assume there is a single underlying construct of patient satisfaction, albeit with multiple dimensions. Underlying this argument is an awareness that the measurement models of patient satisfaction assume a continuum (satisfaction—dissatisfaction);

however, other research suggests that patients do not think of satisfaction/dissatisfaction as the endpoints of a continuum (59, 101).

Fitzpatrick argued rather differently to this assumption, presenting the case that combining the different dimensions of patient satisfaction into a single construct was most likely leading to confounded measurement, based on the observation that patients appeared to have multiple concerns with their health care (6). Consequently he argued that there were three models and that each should be measured separately (122).

The first model, the need for the familiar, reflected that expectations were socially mediated, as described above (5, 76). The second related to the reason the patient was seeking health care, i.e. the goals of help-seeking. Under this model, patient satisfaction is a function of the extent to which the patient's health goals are met; similar to fulfilment theory as described above (4, 106). The third related to the importance of meeting emotional needs. The argument was that since illness strikes at the very core of our being, illness and its treatment has a major emotional impact. Since patients do not possess medical knowledge, they place their health care need in the hands of the clinician — leading to the relationship issues described above (3, 41, 123). The corollary of this argument was that patient satisfaction was therefore a function of the interpersonal affective behaviour and communication skills of the clinician (since the patient was in no position to evaluate the quality of technical skill). Sitzia and Wood (50) speculated that this third construct was derived from the work of Ben-Sira (124) which explored the patient-clinician relationship.

4.5   Donabedian (1988): Health care quality assessment

Donabedian's writings on the quality of health care contain references to patient satisfaction, rather than patient satisfaction being the focus of his work (7). This is important, because it places patient satisfaction within overall care assessment. As already presented, Donabedian saw quality care as comprising the medical infrastructure, process quality which focussed on personal relationships within health care systems, and the technical health outcomes from the treatment process (2, 7). In general, he argued that care quality could only be attained where service providers aligned these three aspects of care to the needs and values of the patient population (125). Quality care, then, was to be based on a series of broadening concentric rings: at very the centre was the interpersonal relationship between the patient and the clinician and the level of technical care provided, further out was the patient's contribution to their care (including that of their family), and finally the level of community care (e.g. care access) (2). Generally, it is the interpersonal relationships and technical care that have been incorporated by others into their work (e.g. see Kayne et al. (28)).

Figure 2:   **Diagrammatic representation of Donabedian's levels of care quality**
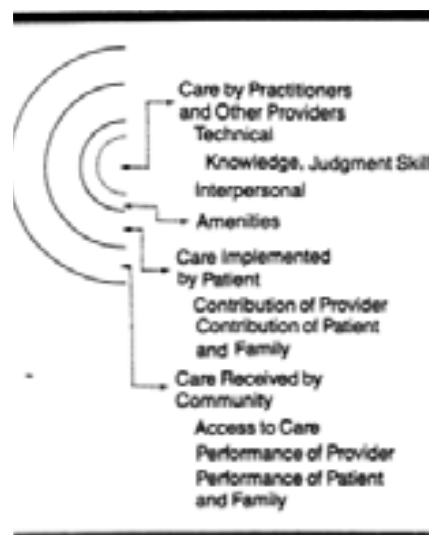
Source: Adapted from Donabedian (1988), Figure 1, p1744

Regarding the assessment of the level of technical care provided, Donabedian argued this was a function of comparing the actual care with the best practice, where the best practice was defined by its ability to produce the greatest improvement in health (2). He noted, however, that patients were often not in a position to fully appreciate either what technical care could achieve or understand certain aspects of health care (e.g. dissatisfaction could occur where a patient had unreasonable expectations of the

efficacy of treatment). From this, he argued that clinicians had a role in promoting patients' health literacy (7), which gave rise to the view that technical care satisfaction was a function of technological expectations which were formed by the current standard of medicine (2). This argument has since been supported by other research and extensively used as the underlying model in patient satisfaction instrument construction (28, 32, 33, 93, 126, 127).

5. Issues in interpreting patient satisfaction scores

One of the most consistent findings across the literature is that people are satisfied with their health care. In a very early study, Cartwright (73) reported that over 80% of patients were satisfied — a finding that has been replicated time and time again regardless of whether single item, generic or specific patient satisfaction measures or qualitative research techniques (open-ended interviews) have been used, although qualitative research tends to report more reasons for dissatisfaction. A possible reason for the high level of patient satisfaction, as argued by Crow et al. (42), is to do with the term "satisfaction" itself which means "enough". The implication is that as long as a service satisfices it will provide satisfaction. In addition, high rates of satisfaction are reported even when there is evidence of continuing health problems. For example, in a study of internal sphincterotomy for chronic anal fissure it was reported that 28% of respondents continued to experience difficulty with wind control and 22% with underwear soiling yet 90% reported overall general satisfaction (113). In another study of abdominoperineal resection and iliac colostomy, where approximately 40% of patients were incontinent at 12-month follow-up, 85% reported being satisfied (128).

High satisfaction levels are found in short-term assessments, as well as in long-term follow-up. For example, at between 1 to 7-year longitudinal follow-up of closed haemorrhoidectomy with local anesthesia 93% reported being satisfied (129); likewise in a study of ileoanal pouch procedure, at five year follow-up 94% of cases reported being satisfied (130). However, it is important to qualify any assumption that patient satisfaction is a constant over time. In a cross-sectional mailed survey of women with a history of continence surgery it was reported that following surgery 67% reported being satisfied compared with 45% at 5 year follow-up, suggesting that satisfaction changes over time (131). As a result of this confused situation, at least one research team has suggested that researchers should focus on sources of dissatisfaction rather than satisfaction (101). A possible reason for these changes over time is that patients tend to focus on their current state of health rather than consider the extent of health improvement over time (28).

Several possible reasons for this high level of satisfaction have been advanced, including reasons relating to individual differences, differences in perception and in communication between patients and doctors, instrument administration and patient dependence, instrument bias, and that most people are satisfied with their lives generally. Additionally, there are socio-demographic correlates of patient satisfaction which many have been confused with the construct itself; likewise patient satisfaction has also been shown to be a predictor of various outcomes from treatment.

5.1 Changes in health consumer role and health literacy

There has been a major change in the role that patients perceive themselves to play in relation to the health care system (38). This change is directly attributable to the growth in the consumer rights movement since the 1950s, as outlined in section 1. As part of this change there has been a corresponding change in the role of both patients and clinicians. These changes are predicated on the assumption that patients would become informed consumers of health care who took an active part in the decision-making process along with their clinician. Obviously, this change in patient role demands a higher level of health literacy than situations where the doctor was seen to be in charge of health care decision-making.

Meredith (132) argued that because this change in role was externally imposed on patients and doctors it should have led to an increase in dissatisfaction. No study examining this issue appears to have been reported, although Elkadry et al. (133) reported that women who were unprepared for pelvic reconstructive surgery were more likely to report dissatisfaction. The evidence for the argument would be that there is both a consistent age and education effect across patient satisfaction reports because older less well educated patients are more likely to be less well informed consumers, hence they are more likely to accept offered treatment. There is some support for this theory; older adults (generally defined as those aged >60 years) generally, although not in all studies (e.g. see Tannenbaum et al. (134)), report higher levels of satisfaction when compared with younger adults. Likewise there is some evidence that those with lower education levels report higher satisfaction (61, 135-137). The interpretation would be that younger patients are more likely to be more health literate, and hence more likely to express dissatisfaction. An alternative explanation is that older adults may be confronting multiple chronic health conditions, and that they are grateful for care which prevents further deterioration through providing relief.

5.2   General life satisfaction and patient satisfaction

A second general explanation lies in Cummins' life satisfaction work. This shows that on scales measuring life satisfaction respondents select responses that provide satisfaction estimates at approximately 75% of the scale maximum score (75%SM 138). The implication of the 75%SM is that most people are predisposed to indicate satisfaction generally regardless of the actual circumstances. Although there is some literature reporting that changes in wait-time change patient satisfaction scores (139), Meredith (132), in a study of wait times for general surgery, found that excessive times were interpreted by patients as meaning that the doctors were indifferent to them as people. She also reported that this dissatisfaction was often not communicated because patients were not predisposed to complain. This may explain why some researchers have found that an important predictor of patient satisfaction is general life satisfaction (140).

5.3   Unformed expectations and opinions on medical care

Although it is commonly assumed that patients will have expectations of health care and opinions about quality health care, this is not necessarily the case. Patients may not have well formed opinions at all, thus when patients are asked to evaluate the satisfactoriness of their care and the health intervention they respond as uninformed consumers who are limited in their capacity to pass legitimate judgement on their care (101).

Regarding expectations, there is equivocal evidence. Williams (141) interviewed patients and reported that their expectations were based on beliefs about what a service should offer, and the duty of care of the service towards its patients, whereas Kravtiz et al. (142) reported that many expectations were formed from anecdotal vicarious evidence. Jackson and Kroenke (143) in a study of unmet need reported that 98% of cases had at least one pre-visit expectation, the most common being diagnosis (80%), anticipated recovery time (62%), receiving a prescription (66%), diagnostic test (56%), and referral (47%). The most common unmet expectations were in relation to prognostic (51%) or diagnostic information (33%). They also reported that there was a relationship between expectations and satisfaction: those whose expectations were met reported higher satisfaction levels; a finding supported by Zebiene et al. (82) who reported a monotonic relationship between the number of unmet expectations and satisfaction. This finding was consistent with that reported by Lin et al. (144) in a study of consultation time. They reported that where the expected consultation time was met or exceeded, patient satisfaction was higher.

On the other hand, the argument has been challenged by Coyle and Williams (101) on the grounds that patients may have a cluster of both positive and negative expectations. In a study of older patients receiving district nursing care, Owens and Batchelor (93) reported that patients fell into three groups: those who were informed and had clear expectations, those who had no prior conceived expectations, and those who had misconceived expectations. The overall finding was that patients had few prior expectations of the service. From this they argued that expectations were formed from years of contact with health services, and that therefore patient satisfaction would always be congruent with these expectations since it was based on the patient's experiences of what services were delivered. O'Connell et al. (145), for example, in a study of satisfaction with nursing care administered both a standardized measure, the La Monica-Oberst Patient Satisfaction Scale (LOPPS), and interviewed the patients. The findings showed that the mean score was at 82%SM — indicating a very high degree of satisfaction — but that there were so many negative comments made during interviews that these called the validity of the LOPSS into question. One reason was that the patients were unable to discriminate care from the rest of the hospital experience, with the result that the patients completed the LOPSS in terms of social desirability.

Based on consumer satisfaction theories, Kane et al. (28) argued that patients may have intertemporal expectations that were driven by past experiences. However, they also argued that many patients do not have sufficient past experience on which to form expectations, with the consequence that expectations were formed by the opinion of others, including the clinician. Under these circumstances, patient satisfaction became post-hoc rationalization. They argued that patient satisfaction was a relative function of outcomes (before-after treatment), severity (before-after treatment), demographics and procedure. At 6-month follow-up they reported that this model explained about 8% of the variance; and that the strongest predictor of satisfaction was current health state (not change in health state) rather than expectations.

5.4   Miscommunication and dependence between patients and clinicians

There may be mismatches in communication between patients and clinicians. Generally, data collection requires an interaction between the researcher (often the clinician) and the respondent (the patient). This involves four distinct steps: (a) the researcher must encode the question in a way that takes into account the purpose of the question, as well as the respondent's ability to answer the question; (b) the respondent has to decode the question, taking into account his/her own purposes during which assumptions are made about the purposes of the researcher; (c) the respondent has to encode a response which draws upon his/her experiences and which reflects the responder's understanding of the researcher's needs; and (d) the researcher has to decode the response in light of the researcher's purpose and what the researcher thinks was the purpose of the respondent (146).

At each of these stages there may be miscommunication between the researcher and respondent. For example, if a clinician assumes that successful treatment (e.g. no further major episodes of urinary leakage) is synonymous with satisfaction, then the researcher may neglect to ask about satisfaction *per se*. Likewise, if a patient feels he/she is in a dependent relationship with the clinician and perceives that the clinician is not interested in occasional small leakage or satisfaction then the patient may fail to bring these issues to the attention of the clinician. The clinician, having received an answer that conforms with his/her expectations, will report that the patient was satisfied. For example, Flocke (147) reported that interpersonal communication and patient satisfaction were only moderately correlated ($r = 0.46$). In a study of how patients and clinicians evaluated total hip arthroplasty outcomes, Leiberman et al. (148) reported that where patients were satisfied there was congruence between patient and clinician ratings, but that where patients reported dissatisfaction there was disagreement; indeed the clinicians' satisfaction estimates for these patients were almost double what the patients reported themselves. This discrepancy between patient and clinician perspective can also be seen in the Black et al. (149) study of surgery for stress incontinence where at 1-year follow-up 66% of women reported that the operation had met or exceeded their expectations compared with clinicians reporting they were satisfied with the case outcomes in 85% of cases. Other studies reporting similar findings include that of McCammon et al. (150) and Wei (151).

Regarding dependence, many patient satisfaction measures are administered during or at the end of a consultation (e.g. see 116). In the case of single-item measures, this is often done as a stimulus to further discussion of issues arising from the treatment. There are two difficulties here. Because of the dependent relationship the patient may feel obligated to report satisfaction leading to acquiescent response set bias (ARSB) or yea-saying. Coyle and Williams (101) provide a review of dependence, arguing that dependence prevents patients reporting dissatisfaction because of perceptions of powerlessness, fear of retribution, the emotional cost of complaining, the perception that complaints would have no impact on clinicians' behaviours or organisation, and feelings of gratitude.

It may also be the case that because of the immediate care, the patient is likely to respond favourably, i.e. there may be a Hawthorne effect or, in evaluation terms, a happy sheet response. Williams (38) argued that these difficulties reflect patient passivity where there may be no formed satisfaction opinion coupled with medical paternalism. Both of these possibilities may systematically bias patient responses upwards resulting in higher satisfaction levels than is warranted. During the development of the Patient Satisfaction Questionnaire (PSQ), for example, Ware et al. (3) reported that between 40-60% of respondents exhibited some form of ARSB, and that for 2-10% this was substantial, inducing significant bias particularly where items were positively worded. The solution Ware et al. adopted in the PSQ was to ensure a balance between positive and negative items.

5.5   Caring and patient satisfaction

Some patients will indicate satisfaction where they believe that they have received the best available treatment and care. It is difficult to formally express dissatisfaction when one has received significant caring at a time of personal distress, regardless of whether the treatment has proved effective or not (152). Williams (121) interviewed patients and reported that where a patient was dissatisfied, but believed that the clinician had not breached his/her duty of care this dissatisfaction was not reported since the patient assigned responsibility elsewhere (often within themselves). Kinnersley et al. (153) compared patient satisfaction with a measure of patient-centeredness based on analysis of recorded transcripts of the consultation. The findings showed that 8% of patient satisfaction could be explained by patient-centeredness during the consultation; adjustment for potential confounders had almost no effect on this relationship. In a study exploring the effect of leaflets aimed at empowering patients, Little et al. (154) reported that where patients were encouraged to raise their concerns with the clinician, there was an improvement with communication and with subsequent patient satisfaction.

Regarding caring, Hyrkas and Paunonen (155) reported that the key attributes sought by patients were in relation to nurses being objective, friendly and pleasant; and that negative experiences were related to a breakdown in these relationships. Different perceptions applied to clinicians, who were more valued for their professional skills, although the patient-clinician relationship was still raised as an important issue. Coyle et al. (156) reported that 44% of patients reported that nursing staff were unavailable to patients even though they were physically present because the patients saw that the nurses were so busy.

These findings suggest the central importance of the empathy, courtesy and communication skills of the clinician or nurse; certainly Meredith (132), Calnan et al. (157) and O'Connell et al. (145) reported that this aspect of caring was critically important in determining high levels of patient satisfaction. Drain (59) defined patient satisfaction in terms of perceived communication – sensitivity to patient needs, concern for privacy, meeting patients concerns or worries, including patients in decisions, friendliness towards patients, instructions for follow-up, time spent with patients and explanations of conditions – on the grounds that clinics which scored high on these characteristics were more likely to retain their patients in a competitive environment.

5.6   Bias in patient satisfaction instruments

Where patient satisfaction measures have been designed from the researchers' or clinicians' perspective (thus only reflecting the concerns of these professionals) and no ecological validation has been undertaken to verify that this is congruent with patients' perspectives, there is considerable opportunity for miscommunication or the inadvertent omission of important information. This review shows this situation for many patient satisfaction measures, particularly those involving single-item measures which seek global assessments. The effect is likely to have washed out particular concerns thus systematically biasing responses upwards so that higher levels of satisfaction than might be warranted are reported. A case in point is Rahmqvist's Patient Satisfaction Index (136) which consists of just two global questions which ask the patient how they felt about their hospital admission as a whole, and how close their experience was to being perfect in every respect. Each question was evaluated on a scale of 1-7, and then summed together.

A second important consideration is that many patient satisfaction measures are embedded within care process or treatment outcome measures, particularly where single item measures are used. A good example is the General Practice Assessment Survey (20, 158) which measures the process of care. It comprises 13 scales/items measuring access, receptionists, continuity of care, technical care, communication, inter-personal care, trust, knowledge of patient, nursing care, referral, coordination, recommend to others and overall satisfaction (1 item).

The consequence of this type of arrangement is that there is often confusion between treatment success and satisfaction with treatment. For example, Bomalaski et al. (159) in a study of collagen injections for structural urinary incontinence among children assessed patient satisfaction through questionnaires covering self-esteem, activity level and treatment benefit. Similarly Burgio et al. (160) used items covering satisfaction with progress (completely/somewhat/not at all) and whether patients wished to receive another form of treatment (yes/no) as measures of patient satisfaction. In another case, patient satisfaction was assumed from items such as the helpfulness of the person at the registration desk, the courtesy of medical technicians and how well blood was taken (quickly, little pain etc.) where the response categories were on a 5-point scale from 1 = poor to 5 = excellent (37). Whether these items and response categories actually measured satisfaction as opposed to health care processes is uncertain.

Two extreme examples of the mismatch between what was measured and the inference of patient satisfaction were uncovered in this review. The review paper by Demirci and Petri (161, 174) assessed 60 papers on perioperative complications of Burch colposuspension. At no point did they review patient satisfaction measures or outcomes in any of the studies, yet they concluded that *Knowledge of the possible risks and complications of Burch colposuspension ... increases postoperative surgical success and patient satisfaction!* The other example is the paper by Williams et al. (162) who evaluated a nurse-led incontinence intervention. No patient satisfaction measures are reported in the paper, yet it is stated that 99% of patients were satisfied.

Calnan et al. (157) observed that these biases were major reasons for the high satisfaction levels reported in the literature. Baker and Whitfield (163) argued that this implied measuring patient satisfaction was often a fairly meaningless activity.

5.7   Instrument administration effects

Equally important are issues of administration mode. Patients may be asked to self-complete questionnaires (typically through mail-back), or respond to telephone interview or face-to-face interview at follow-up. There are two important effects of administration mode. The first is in relation to patients who drop out of studies and fail to complete questionnaires, and the second is in relation to the reported assessments.

Regarding completion rates, Ware et al. (3) reported that with mail-back self-completion the response rate was 69% compared with 95% for self-completion under interviewer supervision; non-completers were more likely to be younger, non-white and from a low-income background. Few studies have investigated or reported these issues in depth, although some papers have reported them. An important corollary relates to the effect of missing data on the analyses. Although there is an extensive literature on missing data, including advice on appropriate imputation methods, over 95% of the papers reviewed in this study used listwise deletion, which is where a case is simply omitted from the analysis if there is a missing datum in his/her dataset. This situation is disappointing because there are known biasing effects where listwise deletion is used; generally people remain in studies where their treatment is effective and they are more satisfied.

For the effect of administration mode on satisfaction assessments, there are good psychometric reasons why researchers should not mix-and-match administration modes (164). There is evidence suggesting acquiescent response set bias (ARSB) is present when patients are interviewed; i.e. that during interview patients are more likely to give socially acceptable answers. Nguyen et al. (152) reported on oral versus written administration of a single satisfaction item and reported that in oral (interviewer) mode scores were 15% higher when compared with written mode. Consistent with this, a recent study of clinic interview compared with mail self-completion of the General Practice Assessment Survey (GPAS) showed that on average mail completers obtained scores that were 6% lower than those given by clinic interviewees (165). In a study of mail versus telephone interview, Hawthorne (166) reported that when filter questions were used in telephone interviews they increased quality of life scores by 7% because respondents assumed the researchers were only interested in major health conditions. Finally, there are question order effects. Ware et al. (3) reported that where health questions were asked prior to satisfaction items, when compared with asking the satisfaction items first, lower levels of patient satisfaction were reported on 78% of PSQ sub-scales, suggesting a sensitizing effect due to contextual factors.

5.8   Correlates of patient satisfaction

A constant theme throughout the literature is that there are correlates of patient satisfaction which explain scores. Generally, these predictors are reported as explaining <20% of the variance in satisfaction scores, suggesting that there are both important predictors outside health care and that there are multiple predictors of patient satisfaction (167).

Linder-Pelz (168) reported that the key correlates of patient satisfaction were the nature of the encounter in general, the conduct of the clinician, and access. Hardy (55) used regression analysis to identify the key predictors of patient satisfaction, finding that patient satisfaction was most strongly associated with the process of care (r^2_{adj} = 0.50), then health (0.33) and psychological wellbeing (0.21). Additionally, the literature suggests that the environment within which care is provided is important. Accordingly the correlates of patient satisfaction are reviewed under four headings: communication, health and treatment outcome, the clinical setting, and social determinants.

5.8.1   Clinician-patient communication and relationship

Tucker (102, p72, 169) argued that *…patient satisfaction is predicted by evaluations of access, communication, outcomes and quality.* He examined both patient and health system characteristics in a military sample, using factor analysis and regression modelling. The findings showed that between these two vectors 49% of variance in patient satisfaction could be explained. The specific predictors for health system characteristics were communication, outcomes and care quality, access and telephone access/communication. The significant patient predictors were demographics (rank, service and marital status), health status and service utilization, and geographic location/education. Although some of these predictors can be discounted as being unique to military organisations (e.g. rank and service organisation), the findings are consistent with other studies (e.g. see Drain (59)) and provide a useful overview of patient satisfaction correlates.

A word of caution, however, is in order. Although clinician-patient communication is widely reported to be the most important of these predictors, Jackson et al. (167) reported that while this was true in the short-term, by 2 to 3 weeks after the consultation, satisfaction was most strongly related to the course and impact of the patient's symptoms.

Subject to this, communication with patients seems to be the most important patient satisfaction predictor, particularly clinician attention to the patient, clinician courtesy and concern for the patient (59, 61, 123, 168, 170). Baker et al. (83) reported that satisfaction was higher among patients who had high levels of trust and care continuity with their clinician. Consistent with this, in an earlier paper Feletti et al. (171) reported that the most important predictors of satisfaction were: communication, care and reassurance (which explained 24% of the variance); professional attitude and behaviours (5%); the clinician being a personal confidant of the patient (5%); technical competence (4%); and generating trust in the physician (3%). Williams et al. (172) reported patients wanted explanations of their problems and where expectations were met patients reported significantly higher levels of satisfaction.

Campanella et al. (37) reported that the strongest significant predictors were showing concern for the patient's comfort, the seriousness with which the patient was treated, and the clinicians' behaviour. Process variables (e.g. access such as waiting time) were negligibly important. Linder-Pelz (123) in a study of expectations, reported that the highest correlations were those between expectations and doctor conduct ($r^2 = 0.08$) and between values and general satisfaction ($r^2 = 0.05$). Jackson et al. (143) showed that the most satisfied patients were those with no unmet expectations, who received a explanation of how long their symptoms were likely to last, and who were informed of their symptom etiology. These were all more important predictors than older patients (>65 years), those with higher functional status, those whose symptoms at follow-up had been ameliorated and those who did not require treatment from another doctor.

5.8.2   Health status and treatment outcome

Regarding health status, patients with fewer symptoms, those who are healthier, those with higher functioning and those with higher quality of life generally report higher satisfaction (21, 150, 173), with the exception of some chronically ill groups.

Although many studies have directly equated treatment outcomes with patient satisfaction, Fontana et al. point out that, *ceteris paribus*, this is subject to bias because researchers, generally, have not reported pre-post treatment change outcomes (174). In this review, almost all incontinence papers described the proportion of cases who were continent/incontinent at follow-up; the assumption clearly being that all cases were incontinent at pre-treatment. The assumption of outcomes equaling satisfaction are particularly strong where there has been treatment failure resulting in dissatisfaction (113-117), but this is not always the case. Thus Hazard et al. (175) in a study of the long-term outcomes of chronic low back pain reported that despite ongoing pain and disability, the correlations with patient satisfaction were in the range 0.01 to 0.15, suggesting that as little as 2% of satisfaction scores in patients with chronic conditions could be explained by treatment success. Studies in incontinence reporting similar findings (although not formally quantified in this way) include Jonler et al. (176) and Litwiller et al. (177). Jonler et al.'s study of radical prostatectomy at 2 year follow-up showed that, despite significant erectile and urine leakage problems (58% of respondents), 74% reported being satisfied with the surgery. Litwiller et al. (177) reported that among patients with post-prostatectomy incontinence who received an artificial urinary sphincter, at 2-year follow-up, although 90% reported continuous urinary leakage, 90% stated that they were satisfied and 96% that they would have the procedure again. Other studies include Korman et al. (178) in a follow-up study of stress urinary incontinence after modified Pereyra bladder neck suspension who reported that although 38% of respondents reported their urinary incontinence had not improved or had deteriorated, 77% were satisfied with the treatment.

5.8.3   The treatment environment

Although the hospital or clinic environment is often measured, there is less convincing evidence these environmental aspects of care play a critical role over and above the interpersonal care issues discussed. For example, Linder-Pelz (168) reported that although access (convenience) was important, it ranked behind the care encounter itself and the conduct of the clinician. Maitra and Chikhani (179) reported that although waiting time, information communicated by the doctor and the total time spent in the emergency department were all significant predictors of patient satisfaction, they played almost no part in affecting actual scores. This finding was consistent with that of Young et al. (180) who reported that institutional factors accounted for 1-2% of variation in patient satisfaction.

5.8.4   Social determinants

Although many researchers postulated that social factors, such as age, education level, marital status, and ethnicity affect patient satisfaction there is little agreement in the literature. Linder–Pelz (95) reported that social factors actually explained <10% of the variation in satisfaction scores. Almost 20 years later, Young et al. (180) in a study of institutional and social predictors reported that patient characteristics (age, health status, and ethnicity) explained between 9–15% of patient satisfaction, with age the dominant predictor. Age was also reported as the dominant predictor by Baker et al. (181). These findings are consistent with Lewis' earlier review (61).

Being able to work in the immediate years following treatment is a predictor of satisfaction with health care, but in the long term (e.g. over a 5 year period) the predictive power of working status dissipates (175).

6.   Measurement theory and patient satisfaction

Psychometric theory postulates that the valid and reliable measurement of a latent construct requires the construction of a manifest instrument that delivers an observed model which is isomorphic with the construct. To achieve this, the following axioms are widely accepted:

1. There should be a latent model of the construct, including an adequate description of its dimensions;
2. For each dimension, there should be measurement items, such that the item content covers the dimension adequately. All items combined form the descriptive system of an instrument from which the manifest model is derived;
3. The resulting instrument should possess a nomological net of evidence suggesting validity (17);
4. It should also be reliable and responsive.

Where there is a nomological net of evidence (17) relating to each of these four criteria, it may be inferred that an instrument is valid. Since validity is a function of both the instrument itself and the respondents who complete it, validity is never a fixed property but varies from sample to sample. The important corollary from this is that although there may be validity evidence for an instrument developed in, say, the USA, that same instrument may be invalid in Australia due to cultural differences (see section 3). It is accepted among psychometricians that this implies basic tests of validity and reliability need to be applied each time an instrument is used with a different population. As this review shows, this principle has not been well applied in the patient satisfaction field.

These criteria can be used as a checklist against which instruments can be assessed. Each criterion is briefly described.

6.1   An underlying latent model

For the measurement of constructs, there is no "gold" standard against which instruments can be assessed; a situation that can be contrasted with, for example, the presence of breast cancer cells where the results of a mammogram can be histopathologically confirmed. Defining constructs, then, is usually carried through by developing a theoretical model which often draws upon previous research, the views of experts and those who will complete the measure.

The models of patient satisfaction reviewed above suggest that patient satisfaction is multi-dimensional and that comprehensive instruments should measure the key dimensions. As pointed out by Kane et al. (28), however, most instruments only measure some of the key dimensions. In a review of the literature, based on examination of 221 articles, Lewis (61) showed that the most commonly measured dimensions were clinician humaneness (65% of articles), informativeness (50%), overall quality (45%), competence and satisfaction (43% each), bureaucracy (28%), access (27%), cost (18%), facilities (16%), outcomes (6%) and continuity (4%). Whilst a popularity list like this does not define what should be measured, it does provide insight into how patient satisfaction is most commonly perceived by researchers and clinicians. Based on the literature reviewed above the key dimensions are:

1. Appropriate access to health services, including the environment within which treatment takes place and the level of care coordination (3, 5, 55);
2. The provision of health information (5, 7, 8, 55, 76);
3. The relationship between the patient and health care staff, specifically empathy with the patient (2, 3, 28, 50, 55, 124, 182);
4. Participation in making choices regarding health treatment, including the associated fears and sense of loss of control as well as the appropriate use of treatment therapies and medications (55);
5. Satisfaction with the treatment provided, i.e. the technical quality of the care provided (3, 5, 8, 28, 55, 123); and
6. The effectiveness of treatment, including the extent to which treatment helps the patient in their daily life (2, 3, 8, 55).

Many instruments have incorporated a global patient satisfaction assessment to supplement these dimensions.

The dimensions above exclude social demographics, patient beliefs, expectations and interpersonal comparisons (5, 41, 95). Although these omissions may be thought surprising, they are excluded since they are predictors of patient satisfaction rather than dimensions of the construct itself (28, 183).

Because they mediate patient satisfaction, however, there is a strong case for their inclusion in surveys of patient satisfaction so that patient satisfaction scores can be adjusted for these independent variables during data analysis.

6.2   Adequate coverage of the latent construct

This requirement stems from the axiom that the manifest items forming an instrument are only ever a sample of the total number of all possible items. The assumption is made that these items are "sampled" from the underlying pool of items that measure the construct. If items are selected in such a way that they measure only part of the construct, then the measurement will be biased. For example, if patient satisfaction comprises the dimensions described above, but an instrument only measures the first dimension, then the content validity of the instrument is open to serious question.

Where examination of an instrument reveals adequate coverage of the latent construct, then content validity of the descriptive system is inferred.

6.2.1   Response scale categories

The choice of response scale categories can influence how respondents assess the question that is being asked of them. In most instances, researchers use one of two response scale formats; these are usually referred to as Likert scales or Guttman-type scales.

Likert scales

Likert scales are designed to elicit attitudes towards some object. In the present study, these may be attitudes towards nursing care, for example. They are usually presented with 5 options:

> *Very dissatisfied*
>
> *Dissatisfied*
>
> *Neither (neutral or don't know)*
>
> *Satisfied*
>
> *Very satisfied*

The assumption is that the mid-scale is the neutral position, and that each of the anchorpoints on either side of this represents equi-distant intervals progressively marking further deviance from this neutral position. Many researchers often add an extra choice to cater for those to whom the item is not applicable.

Where the mid-point (neutral) is omitted, so that the scale has only four options, it is described as a forced choice scale because respondents are not given the option of indicating that they may have no position on the question of interest. For example, Miller et al. (184) in the Australian Longitudinal Study of Women's Health used a single item measure of satisfaction and a forced choice 4-point scale (1 = very satisfied and 4 = very dissatisfied); i.e. the respondents were not given the opportunity of indicating they had no opinion on the item of interest. A 4-point forced choice response scale was also used in the French inpatient experience questionnaire; to overcome the forced choice difficulty the authors added a "Not applicable" category and imputed these scores using vertical mean substitution (80). Sometimes researchers have included both a neutral position and a not applicable category (156).

That this matters may be deduced from a study reporting that on average, across 17 items measuring patient satisfaction, 11% of respondents chose the neutral position (no opinion) (185); the implication is that forcing respondents into a choice through use of a forced choice scale could seriously distort study findings since most people will select the category above the neutral position because it is both socially acceptable and most people are predisposed to indicating that "things are all right".

Guttman scales

Guttman scales assess the amount of something. For example, this might be the amount of time a patient felt their clinician took an interest in him/her. They are usually presented with 5 options, such as this example of the frequency of events:

> *Never*
>
> *Occasionally*
>
> *Sometimes*
>
> *Often*
>
> *Always*

The assumptions here are that the neutral position is "never", and the anchorpoints represent increasing distances away from this neutral position. Unlike Likert scales, Guttman scales do not make the assumption that the anchorpoints are equi-distant. Nor is there any particular difficulty with a Guttman scale of, say, 4-points other than that the range of choices is restricted.

Both forms of response scale are widely used in the patient satisfaction literature. The two examples above assume, for convenience, a scoring system of 1,2,3,4,5. But not all scales are scored in this direction, they may be scored the opposite way. Where they are scored the opposite way, the responses are usually reversed prior to scale summation.

There is some evidence that patients interpret Likert and Guttman scales differently and that they do not produce equivalent results. For example, the Patient Satisfaction Questionnaire (PSQ 3) uses Likert scale responses. However, Ware and Hays (186) tested the PSQ with Guttman scales and reported that the Guttman response rating scales (poor/fair/good/very good/excellent) worked just as well, if not better. Hendriks et al. (187) tested both Likert and Guttman formats in their construction of a Dutch patient satisfaction scale (Assessing Inpatient Satisfaction) and reported that Guttman-type response scales provided lower satisfaction scores than those using Likert scales.

Regarding other forms of response scales, Meric (188) investigated differences on the PSQ using two forms of Likert scaling (agreement, where 1 = strongly agree, 5 = strongly disagree; and satisfaction, where 1 = very satisfied, 5 = very dissatisfied) and expectations (1 = more than expected, 5 = less than expected). It was reported that the different forms produced comparable results.

6.3   A nomological net of validity evidence

Validity may be defined as the extent to which an instrument measures the latent construct (17, 189-191). Under perfect validity, there would be an isomorphic relationship between the observed or manifest model, which is derived from the descriptive system, and the latent construct model. Although different terms and emphases have been fashionable at different times, conventionally three types of validity evidence have been sought:

- Content validity: describes the extent to which a respondent's responses to the items of an instrument may be considered a representative sample of his/her responses to the hypothetical universe of situations which constitute the latent construct (192). This is usually assessed by mapping manifest items against the latent model of the construct and reporting the coverage. Ecological validity is where a sample of those who have to respond to an instrument, i.e. the patients in this case, have participated in the development of the instrument model and items (193). Many commentators have confused content validity with "face" validity, which is where expert opinion asserts an instrument "looks right". While there are formal procedures for assessing content validity, there are no formal procedures for assessing face validity, which, therefore, has little psychometric merit.

- Construct validity: describes the extent to which an instrument measures an underlying latent construct. This implies that the observed model, constructed from the manifest instrument, enables inferences about the construct to be made. There are two popular ways of assessing construct validity. Factor analysis can be used to examine the relationships between items; items that are measuring a similar construct will be assigned to a "vector"; the content of vectors should be congruent with the previously hypothesized model. A second way of assessing construct validity evidence is through comparing the scores of respondents with the construct with scores from respondents without the construct.

- Criterion validity: describes the relationship between instrument scores and either other independent measures (criteria) or other specific measure (predictors). This includes:
  - Concurrent validity which is where the criterion data are collected "at about the same time" as the instrument data are collected;
  - Convergent validation which is where an instrument is explicity tested against similar measures of the same construct (convergent evidence) and against measures of different constructs (divergent evidence); and
  - Predictive validation is where a score now can be used to predict some future state.

6.4 Reliability and responsiveness evidence

Reliability describes the stability of measurement. The scores obtained on an instrument may be thought of comprising two components:

$$X = T_s + e_s$$

where T_s represents the "true" situation and e_s random "error". Reliable measurement is where the true score is maximized and error score minimized. Generally, reliability is assessed through test-retest, internal consistency analysis or split-half analyses.

Responsiveness describes the sensitivity of measurement. This is usually assessed by verifying that scores from an instrument systematically vary between respondent groups, or systematically vary within a group over time when there has been an intervention between measurement points.

Finally, instruments must be practical to use in field settings.

# 7.   Review of single-item patient satisfaction measures

Given the literature above, it is beyond the scope of this study to comprehensively review all patient satisfaction measures, particularly those developed for specific studies, patient groups or institutions. Based on the inclusion criteria outlined in section 1.1, three groups of patient satisfaction measures were identified and are reviewed:

1. Single item patient satisfaction measures;
2. Multi-item generic patient satisfaction measures (see section 8); and
3. Multi-item incontinence specific patient satisfaction measures (see section 9).

Most patient satisfaction reports are based on single item measures which are short, quick and easy to administer. These are widely used in clinical settings because they are easy to understand and interpret immediately, and they are frequently used by clinicians as discussion starters with patients. Generally, single item measures take one of three forms.

Patients may be asked to indicate on a continuum their level of satisfaction where the endpoints are defined, thus:

How satisfied are you with your health care?

0	1	2	3	4	5

Very
dissatisfied   Very satisfied
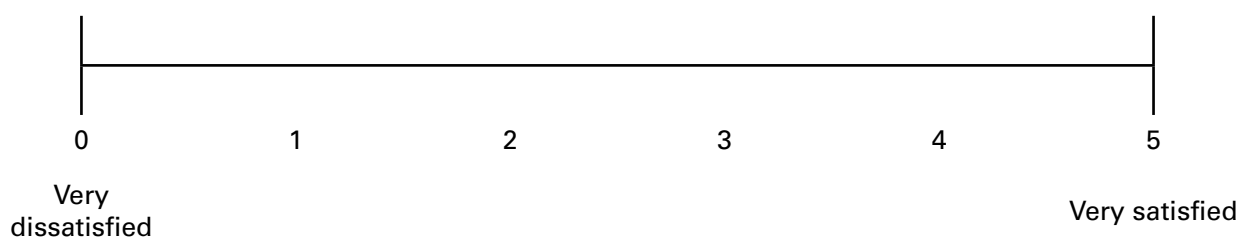
The number of points provided on the continuum may vary from none to 10 or more, or use a visual analogue scale (VAS). For example, Peters et al. (194) in a study of transrectal biopsy taken under sedation used a VAS where the endpoints were 0 and 10, with lower scores indicating higher satisfaction. For those where the biopsy was taken under sedation the mean level of satisfaction was 0.9, indicating an extremely high satisfaction level. Maher et al. (195) provides another example of a VAS, and Hendriks et al. (187) report that a 10-point scale is the standard used in The Netherlands for school reports, therefore, due to familiarity, it was argued that 10-point scales are appropriate for assessing hospital care quality.

Fox and Storms (5, 560) assessed patient satisfaction on a scale of 10-points with a single question:

> *If a score of ten represents the best possible medical care available and one represents a very poor quality of medical care, how would you rate the medical care you have received in the past year?*

Two problems are evident. Regarding what is being measured, the question asks about the quality of medical care. Whether this is an adequate representation of patient satisfaction was not assessed by Fox and Storms; they assumed this measured patient satisfaction. Regarding how this should be interpreted, the question refers to all medical care over a defined time period, thus there is no way of knowing what medical care was received and what care was being assessed. A generous interpretation might be that this was a question about respondents' beliefs regarding the quality of health care generally, rather than it measuring patient satisfaction with a particular health care intervention. It is not inconceivable that a respondent may endorse this question yet be dissatisfied.

Although few studies have replicated this question, the literature is replete with studies reporting similar questions (175, 196-199).

Alternately, the patient may be asked to respond to a question with categorical answers, like this:

How satisfied are you with your health care?

> *Very satisfied*
>
> *Satisfied*
>
> *Not sure*
>
> *Dissatisfied*
>
> *Very dissatisfied*

As discussed in section 6.2, different response scales can be used with this formulation, including satisfaction (as in this example), agreement, expectations and events. Ruiz-Deya et al. (120) provide a good example. In their follow-up study of radical prostatectomy, they asked patients how satisfied they were with the surgery, and used a 5-point Guttman-type agreement scale of 0 = Not at all to 4 = Very much. This question was interpreted as measuring patient satisfaction; not surprisingly 95% of respondents reported being satisfied. Rather than being general, this question is probably too specific to be used as an estimate of patient satisfaction since it focuses on the surgery alone and ignores all other aspects of care. Another recent example of a very simple categorical patient satisfaction measure is that reported by Grumbach et al. (200) who used a global satisfaction question with the five response categories listed above in a study of patient satisfaction with their primary clinician (the GP). The results showed that 82% were satisfied.

A third form of single question is where patients are asked would they either (a) have the procedure again, or (b) would they recommend the procedure to their friends. Positive responses are interpreted as indicators of satisfaction (150, 160, 201-206). However, invariably most patients state that they would have the procedure again even when the medical procedure involves considerable health losses including complications (116). A case in point is that of Lim et al. (207) who reported loss of sexual function combined with ongoing incontinence among males who had radical prostatectomy or external beam radiotherapy; yet over 90% reported that they would have the procedure again. For a recent example see Latini et al. (208) who asked patients undergoing vaginal vault prolapse both these questions. They stated that these measures were sufficient to measure patient satisfaction and reported that 100% of participants responded positively.

8.   Review of generic patient satisfaction measures

During the past 50 years, many generic patient satisfaction questionnaires have been developed. In most cases, these have been developed in the USA for *ad hoc* hospital use (55). Although many are available through the internet, few are used outside the developing hospital and almost none have published psychometric data in the academic literature. Hardy (55) pointed out that many of these measures were not developed using sound psychometric principles with the consequence that patient satisfaction measures proved to be inadequate and brought patient satisfaction measurement into disrepute. In recent years there has been greater attention paid to the measurement properties of instruments, but most studies still do not report on the properties of the measures used. Additionally, most instruments are concerned with satisfaction with the service facilities rather than satisfaction with the patient-clinician interface. For these reasons these measures are not reviewed here.

8.1   Client Satisfaction Questionnaire (CSQ-18 and CSQ-8)

The CSQ-18 (18 items) developed by Larsen et al. (9) was intended to measure satisfaction with services. From the literature, 9 service dimensions were identified and 9 items written for each dimension. These were then assessed by health professionals and the resulting pool of 31 items administered to 248 mental health patients. Factor analysis revealed a single factor accounting for 75% of the common variance (152). A shorter version, the CSQ-8 (8 items), is also available; it was developed through removal of items from the CSQ-18 where the criteria for removal was those items with the lowest internal consistency properties; thus the CSQ-8 is a more homogenous scale. The CSQ-18 (CSQ-8 items are marked with an asterisk) consists of 18 items measuring the promptness of being seen, the comfort and attractiveness of the facility and building, the amount of help received (*), the appropriateness of the help given, the helpfulness of the services (*), how well the patient was listened to, whether the patient received the service(s) he/she wanted (*), whether there were other services the patient wanted but did not receive, how clearly the patient was understood, the competence of the clinician, rating the quality of service received (*), overall satisfaction with services received (*), recommending the service to a friend (*), being understood, having needs met (*), having rights respected and returning to the service (*). Each item is scored on a 4-point scale, where the responses cover a poor service through to an excellent service. Scoring of the CSQ-18 and CSQ-8 is by simple summation. For the CSQ-18 the score range is 18-72 and for the CSQ-8 it is 8-32.

The psychometric properties of the CSQ-18 and CSQ-8 were reported by Attkisson and Zwick (209) and Nguyen et al. (152). The Attkisson and Zwick sample was a sub-set of 45 cases who completed the CSQ-18 as part of a larger trial (n = 62) at an urban community mental health centre where participants were randomly assigned to treatment or control group. The treatment group viewed a videotape on pretherapy orientation whereas the controls were admitted normally. Follow-up was at one month. Data for the CSQ-8 was extracted and properties of both the CSQ-18 and CSQ-8 reported. Cronbach _ for the CSQ-18 was 0.91 and 0.93 for the CSQ-8. Regarding predictive validity, this was assessed through correlation with three service use variables during the month following administration. For remaining in treatment (Yes/No) the CSQ-18 correlation was 0.61 (CSQ-8: 0.57), for the number of sessions it was 0.54 (0.56) and for the proportion of sessions missed it was 0.06 (0.01). Correlations between various outcome measures (various symptom measures as assessed by both the patient and the clinician) ranged from –0.01 to –0.35 (0.01 to –0.40). In general, the findings showed that greater CSQ satisfaction ratings were associated with more sessions attended and with greater symptom reduction, but not with current symptomatology.

The Nguyen et al. split-half reliability study sample equivalent forms analysis was based on 34 cases, where the CSQ-18 was randomly split into two scales; the correlation was $r = 0.82$. In a further analysis, based on 44 cases, where the CSQ-18 was administered in written and oral modes, the oral mode produced scores that were 10% higher (more satisfied) (152). Also reported in the Nguyen et al. paper was a study of the CSQ-8 involving 49 cases with 4-week follow-up; the Cronbach _ was 0.92 and scores were correlated with self-reported clinical improvement scores ($r = 0.53$), and that in a further study (n = 3,120) the CSQ-8 mean score was 27.09 (sd = 4.01) with an _ = 0.87. Although these reports suggest good psychometric properties, they also suggest that the CSQ is subject to differences in administration. The implication is that in a study where, say, the CSQ was administered, post-treatment, orally to the treatment group and self-completed by the control group, any differences favouring the treatment group could be due to the difference in administration mode and not the new treatment.

Pang et al. (210) in a study of concomitant tension-free vaginal tape insertion during pelvic floor reconstruction surgery follow-up at 1-year post-operation reported on the Chinese version of the CSQ-8. Regarding data distribution across the items, only 10% of all responses involved the third and fourth

level (poor service) of the response scale, across all items. The implication is that half the response scale was redundant and the responses on the CSQ-8 items were essentially dichotomous. In a Costa Rican study of diabetes, Firestone et al. (81) reported psychometric limitations (the original article was unable to be extracted, so no further details are available). Hilton et al. (211) used a shortened version of the CSQ-18 (through removal of 9 items, leaving just 9 items in the version used) and reported a Cronbach _ = 0.78. They also commented that there was positive skew on item responses and recommended a different method of assessing dissatisfaction was needed.

Roberts et al. (212) directly compared the CSQ-18 with the Patient Satisfaction Questionnaire (PSQ, see below in section 8.7) in a study of 148 public health patients. The two measures were shown to assess different aspects of patient satisfaction. While the CSQ-18 provided information that was orientated towards service planning and monitoring, the data from the PSQ was more highly correlated with global life satisfaction and well-being rather than with the specific health care services used. Generally, the CSQ-18 scores were significantly higher than those obtained on the PSQ.

8.2 Consultation Satisfaction Questionnaire (CSQ, described here as the ConsultSQ)

Based on a literature review and iterative consultation with clinicians and patients, the ConsultSQ assesses patient's satisfaction with a consultation with a general practitioner (10). From the review and consultation an item bank was developed and administered to patients in a surgery following a consultation. After further modification, it was re-issued. This procedure was iteratively followed and the bank progressively modified as more data about the performance of items within the bank became available. Following iteration, factor and correlation analyses were used to discard further items and refine the final form of the ConsultSQ.

The ConsultSQ comprises 18 items located in four scales: general satisfaction (3 items); professional care (7 items describing the patient's concerns, the provision of information, treatment by the doctor, agreement with the doctor's advice, and the doctor treating the person as a whole); depth of relationship (5 items measuring the doctor's intimate knowledge of the patient and the transmission of personal information to the doctor); and perceived time (3 items measuring the length of the consultation in relation to the patient's perceived needs). A limitation of the ConsultSQ is that there are no items assessing treatment effects.

Items were written as attitude statements, such as *I am totally satisfied with my visit to this doctor*, and comprised both positive and negative statements. Responses were 5-point Likert scales. Scoring is by simple summation following reversal of negative items. Following administration to 239 patients, the psychometric properties were examined. Internal reliability of the ConsultSQ was Cronbach _ = 0.91, and for the scale professional care it was 0.87, for depth of relationship 0.83, for perceived time 0.82 and for general satisfaction 0.67. Spearman correlations between the general satisfaction scale and other scales were 0.50 for depth of relationship and 0.64 for professional care, suggesting these were measuring different, but related, constructs.

The ConsultSQ was assessed in a trial comparing those who changed doctor (n = 272) with those who did not (n = 711) (163). The results showed ConsultSQ scores systematically varied as predicted, from which the authors concluded that the ConsultSQ possessed validity because changing doctors is a strong statement of dissatisfaction.

Kinnersley et al. (213) compared the ConsultSQ with the Medical Interview Satisfaction Scale (MISS, see section 8.5) in a sample of 198 patients attending GP surgeries. The findings were that there was very little difference in the psychometric properties of the two measures. The correlation between the ConsultSQ and MISS was 0.82, suggesting they were measuring the same latent concept. For the ConsultSQ the mean score was 72% (sd = 12.6) of the scale range. Correlations between sub-scales ranged from 0.40 to 0.79 for the ConsultSQ. The reliability of the ConsultSQ scales were Cronbach _ = 0.73 to 0.94. In a study examining the competence of medical students, McKinley et al. (214) reported that the correlation between consultation assessment and the ConsultSQ scales ranged from 0.16 to 0.44; they suggested it should not be used for assessing medical students. The mean scores on the ConsultSQ scales ranged from 37-69% of the scale ranges.

8.3 La Monica-Oberst patient satisfaction scale (LOPSS)

Developed using factor analysis by La Monica et al. (11) for measuring satisfaction in oncology, the LOPSS originally consisted of 48 items. It was revised by Munro et al. (215) through the removal of

redundant items, making it more suitable for general health care satisfaction assessment. The standard version has 28 items, although Vahey et al. (216) reported using a 21-item version and O'Connell et al. (145) an 18-item version.

The LOPSS measures interpersonal support (9 items), good impressions (5 items) and dissatisfaction with nursing care (14 items). A typical item is *In general, the nurse seems more interested in completing tasks than in listening to concerns.* Responses are on a 5-point Likert scale, from strongly agree to strongly disagree, although Vahey et al. (216) used a forced choice 4-point response scale. Scoring is by summation after reversal, giving a range of 28 through 140 where the highest scores reflect the greatest satisfaction with nursing.

Regarding the LOPSS's psychometric properties, the factor structure (3 sub-scales) reported by La Monica et al. was confirmed by Munro et al. (215). Likewise, reliability was reported by La Monica et al. (11) at Cronbach _ = 0.98 and the revised version at 0.97 (215). O'Connell et al. (145) reported that Cronbach _ = 0.96 and Vahey et al. (216) reported 0.93. Munro et al. reported that the mean score was 118.7 (sd = 17.3).

O'Connell et al. (145) investigated the psychometric properties of the LOPSS in a sample of 105 surgical patients who were questioned about their nursing care. The mean score was 115.7 (sd = 17.4). When LOPSS scores were examined by presumed correlates (age, gender, length of stay) of satisfaction, no significant differences were observed. Telephone interviews revealed that dissatisfaction with several aspects of care did not appear to be reflected in instrument scores, leading to the conclusion that the LOPSS items were too insensitive and that the measure may be prone to acquiescent response bias.

8.4   Linder-Pelz satisfaction scales

Linder-Pelz (4) developed scales to test the expectancy hypothesis arising from her work on the theory of patient satisfaction (section 4.2). Three scales were developed: the Doctor Conduct (DCS), General Satisfaction (GSS) and Convenience scales (CS). The 10 items in DCS were all negative in tone, e.g. *Doctor should have told me how to care for condition*, and measured condition care, being thorough, showing interest, doing the patient a favour, explaining the medical problem, having better clinical equipment, ordering tests, making the patient feel foolish, ignoring previous medical problems and the patient liking more time with the doctor. Six of the 7 items in the GSS were all positive, e.g. *My questions were answered to my complete satisfaction*. It measured answering questions, the doctor understanding the patient, the patient being satisfied with the visit, understanding the medical condition better, receiving better medical care than most people, the doctor being one of the best and not wanting to see the same doctor again. The CS had 4 items measuring easy getting to the clinic, the waiting area being comfortable, how the staff treated the patient and having to wait too long. Scoring of the scales was through simple summation.

Validation of the scales was through administration to all first-time patients in a medical centre in Upper Manhattan (n = 125) following a session with the doctor. Cronbach _ for each scale was 0.81 (DCS), 0.77 (GSS) and 0.49 (CS). No other psychometric properties were reported.

Other than Linder-Pelz's own work, there appears to have been no further psychometric work on her scales.

8.5   Medical Interview Satisfaction Scale (MISS)

The MISS was developed in the USA to measure patient satisfaction with a clinical consultation (12, 13). Originally the MISS consisted of 29 items, however recent work in the UK (78) has suggested a more coherent structure with 21 items nominally organised in four scales (the MISS-21). The scales measure distress relief (told what the trouble is, how serious the illness is, how long before getting better, worries relieved, and that the clinician knew what to do); communication comfort (uncertain, embarrassed, not allowed to say what the patient wanted, and clinician did not understand); rapport (clinician interested in patient, clinician warm and friendly, clinician treated problems seriously, patient felt free to talk about private matters, patient given chance to say what was on his/her mind, being understood by the clinician, feeling trust in the clinician, and the clinician knew what he/she was doing); and compliance intent (easy to follow clinician's advice, difficult to follow clinician's advice, and not sure if worth the trouble of following clinician's advice). A typical item is *the physician told me the name of my disease in words I could understand.*

In addition to revision of the instrument length, a further difficulty is that different research teams have used different response scales. The original item response scales were 7-point Likert scales, but 5-point Likert scales have been used (82). Scoring is through summation.

Several studies have reported its psychometric properties. Wolf et al. (12, 13) reported that the internal consistency of the construction sample was _ = 0.93 and that the four original scales explained 40% of the variance. In the Kinnersley et al. (213) study reported above (section 8.2) the MISS mean score was 76.6% (sd = 11.4) of the scale range; and the reliability of the MISS scales was Cronbach _ = 0.78 to 0.96.

Zebiene et al. (82) administered the MISS to 460 cases, and examined the internal properties. They reported that factor analysis revealed eleven factors explaining 62% of the variance. Of these there were four substantive factors explaining 42% of the variance. They labelled these emotional support, understanding and explanation, information and diagnosis, and treatment. The Cronbach _ for each of the four substantive factors was 0.88 (emotion support), 0.85 (understanding), 0.70 (information) and 0.66 (diagnosis and treatment). Although Zebiene et al. reported similar psychometric properties to those obtained by Wolf et al. (12, 13), they also reported that 5 items did not contribute to the scale, perhaps because of differences in expectations in Lithuania.

Meakin and Weinman (78) administered the MISS to 150 patients and examined its internal structure. They obtained a five factor solution, which they then constrained to four factors through the removal of 8 items which failed to load on any factor >0.40, or which cross-loaded. The correlations between the four scales were from 0.46 to 0.65, perhaps suggesting the presence of an underlying single construct. Internal consistency was examined for each of the four scales, and the Cronbach _ range was 0.67 to 0.92.

Finally, several studies have suggested that the MISS is culturally specific and that cultural adaptation is needed prior to it being used in other countries than the US (78, 82).

8.6   Patient Satisfaction Index (PSI)

The PSI was designed to discriminate between patients with a life-threatening illness who were satisfied with their medical care and those who were not (14). An initial item bank was assembled from the literature, patient interviews, and interviews with family members and health care providers. Following item review, three parallel questionnaires were constructed and interviewer-administered to 102 patients and 153 relatives. Preliminary scales were developed based on the most frequent and important items. These were then re-administered at two week interval along with other items measuring satisfaction. Logical criteria were used to sort items into dimensions, and 8 domains were identified. Health care providers were asked to verify the domains and sort items into the domains. The pattern of responses suggested that three different scales were needed: one for patients (the PSI); one for relatives of competent patients; and one for relatives of incompetent patients. Only the PSI is reviewed here.

The PSI comprises 23 items measuring gone through a lot, decisions made without involving patients, went through more than expected, felt helpless in decision-making, felt out of control of the situation, wanted decisions made by clinicians, feeling overwhelmed, involved in decisions too late, didn't understand what was happening, problems not clearly explained, not firm enough about wishes, options explained, co-operation from clinician, understood by clinician, understood clinician, family involved, respected by clinician, received appropriate care level, decision choices available, comfortable with decision-making, sharing same goals as clinician, clinicians clarify wishes and feel clinicians care. Item response scales are Guttman-type 7-point scales where 7 indicates the highest level of satisfaction and 1 the lowest. Scoring is by summation, providing a range of scores from 23 to 161. The PSI is designed for interviewer administration, and takes 20-30 minutes to administer.

The PSI properties were assessed through administration to a sample of 105 patients, and re-test was carried out with 97 patients. All data were collected through interview. The intraclass correlation coefficient for the test-retest was 0.86. Correlations with other patient satisfaction measures were in the range 0.67 to 0.75, correlations with health care provider (mainly nurses) were 0.19, and with relatives' estimates were 0.28.

8.7   Patient Satisfaction Questionnaire (PSQ)

The PSQ (3) was based on the model of patient satisfaction outlined above in section 4.1. This model construct was derived from a review of the literature and the responses of convenience samples of patients. All the statements written for the PSQ were based on classic attitude measurement theory such that they expressed an opinion, e.g. *I'm very satisfied with the medical care I receive*. Altogether 2,300 statements were prepared and submitted to a panel of judges, who reduced the item pool to about 500 items. The item pool was then administered to patients in four different patient groups and various

psychometric tests, including factor analysis, used to eliminate most items. A reduced item pool of 87 items was then administered to fresh samples and, the PSQ-II constructed based on both logical and statistical criteria. The PSQ-II had 68 items; further revision led to the PSQ-III with 51 items. The items are presented 7 dimensions of satisfaction covering: Access to care (emergency care (3 items), convenience of services (2) and access (2 items)); Financial aspects (cost of care (4 items), payment mechanisms (4) and insurance coverage (3)); Availability of resources (family doctors (2 items), specialists (2) and hospitals (2)); Continuity of care (family (2 items) and self (2)); Technical quality (quality competence (9 items), prudence-risks (2) and doctor's facilities (2)); Interpersonal manner (explanations (3 items), consideration (5), prudence-expenses (2)); and Overall satisfaction (4 items). Scoring of the PSQ-III scales is by simple addition of items within scales.

Tests of reliability (Cronbach _ and test-retest) on the original construction sample ranged from 0.77 to 0.88 (217). However, subsequent reliability tests produced more varied results in that the reliability of PSQ sub-scales ranged from 0.23 (Prudence – risks) to 0.93 (Availability – hospitals) (3). Overall, 78% of the PSQ sub-scales obtained reliability estimates >0.50, which was the standard adopted from Helmstadter (3, 218). An abbreviated version of the PSQ was used by Ross et al. (62) in a study of patient preferences; they reported that the internal consistency of six scales (Access to care, Availability of services, Technical quality of care, Inter-personal care, Communication and Financing of care) ranged from 0.79 to 0.91.

Regarding the validity of the PSQ, Ware et al. (3) argued that because the internal structure of the PSQ was replicated across their many field trials, validity was implied. They also assessed PSQ scales against respondents' concerns, and reported that the scales performed as expected (e.g. those who complained of technical deficiency obtained lower scores on the Technical Quality sub-scale). When assessed against single-item measures of satisfaction, the PSQ behaved as expected and discriminated between groups; although Ware et al. did report that the PSQ provided lower scores than did single-item measures.

8.8 Patient Visit Rating Questionnaire (PVRQ)

The PVRQ is also referred to as the Medical Outcomes Trust patient satisfaction scale and the RAND 9-item patient satisfaction survey. Developed as part of the Medical Outcomes Study (MOS), the PVRQ was designed to provide comprehensive measurement of all aspects of patient satisfaction with a medical consultation for the purpose of comparing patients' views of the quality of care in different systems of care (15). It consists of 9-items measuring the visit overall, the technical skills of the clinician, the personal manner of the clinician, the time to get an appointment, the convenience of the medical rooms, contacting the medical rooms by telephone, the time spent waiting at the medical rooms, the time spent with the clinician and the explanation of the treatment. The response categories are *poor/fair/good/very good/excellent*. Two scoring methods were reported by Rubin et al. (15). Scores were summed and transformed into percentile scores (scale 0 to 100, where 0 = poor and 100 = excellent). The second scoring system was where item responses were dichotomized at excellent/not excellent.

Regarding the psychometric properties of the PVRQ, Rubin et al. (15) reported data from the MOS survey (n = 17,671) which was carried out in three US cities (Boston, Chicago and Los Angeles) where within each city three different types of medical practice were sampled (a prepaid medical insurance practice (health maintenance organisation, HMO), a multi-speciality practice (MSG) and with solo clinician practices (SOLO). The PVRQ was administered after consultation, but before the patient left the office; thus there may well have been a Hawthorne effect. In terms of data distribution, on average less than 5% of respondents endorsed the lowest two categories (reporting that the service was either fair or poor), implying that the response scales were essentially 3-point scales. Based on dichotomization, 55% of respondents reported excellent satisfaction. Responsiveness was assessed by comparing across the different types of medical practice; this revealed that the PVRQ showed SOLO practices obtained higher satisfaction levels, followed by HMO and then MSG practices. Furthermore, differences were also obtained by prepaid and fee-for-service practices and whether patients changed clinicians.

These findings stand in marked contrast with those of Kikano et al. (219), who used the PVRQ in a study comparing self-employed clinicians (n = 2,185 patients) compared with those who were employed (n = 1,351 patients). The findings showed there was no significant difference in PVRQ satisfaction scores, even though there were significant differences on other aspects of health care (e.g. history taking, planning treatment, doing physical examinations).

9. Condition-specific instruments for incontinence

The initial literature search identified hundreds of condition-specific patient satisfaction scales, many of which were embedded within other assessment instruments.

Only one incontinence-specific patient instrument was identified.

9.1 The Genito-Urinary Treatment Satisfaction Scale (GUTSS)

The GUTSS was developed to measure women's satisfaction with the outcome of treatment for urinary stress incontinence (8, 220). Based on the theories of Donabedian (7) and Ware et al. (3), a literature review and focus groups with patients and clinicians, an item bank was prepared and telephone administered to 45 women at 6-months post-treatment. Standard psychometric procedures, including factor and reliability analyses, were used to finalise the GUTSS. The GUTSS comprises 10 questions, two of which are filter questions, organised into two subscales, measuring outcome satisfaction (happy with effect of operation, satisfaction with operation, continuing problems, and disappointment with outcomes) and care satisfaction (doctor's explanations of outcomes, happy with care in hospital, attitudes/behaviour of doctors/nurses and prior information). Following reversal of the two negatively-worded items, scoring of each sub-scale is through summation of raw scores minus 4 (to anchor the scale at 0), an overall GUTSS score is obtained by summing the two sub-scale scores. The score range is from 0 to 32, and the mean score reported in the construction sample was 25 (sd = 6.1).

Regarding the psychometric properties of the GUTSS in the construction sample, principal components analysis showed that all items loaded >0.30, and that each item loaded >0.60 on its vector. Overall the GUTSS explained 72% of the variance. Internal consistency was good (Cronbach _ = 0.83 for the GUTSS, 0.93 for the outcomes scale and 0.76 for the care scale).

Convergent-discriminant validity was assessed by correlation with the Urinary Distress Inventory (r –0.70) and the Incontinence Impact Questionnaire (–0.72), the Patient Satisfaction Scale (0.65), the SF-36 (physical health scale $r = 0.33$, and mental health scale $r = 0.59$) and the Assessment of Quality of Life (AQoL; $r = 0.32$) measure. There were no significant differences in GUTSS scores by demographic variables, whereas scores did vary by absolute incontinence status.

Replication of the psychometric tests in a larger sample (n = 152) confirmed the internal structure, reliability was Cronbach _ = 0.84, and significant differences by post-operation continence status were observed (8).

In a subsequent paper by Karantanis et al. (221) evaluating tension-free vaginal tape in younger and older women, at 6-month follow-up there were significant differences between the two cohorts. The older women were less satisfied (median GUTSS score = 87% of the range) when compared with younger women (95%), perhaps because they reported less satisfactory clinical outcomes.

9.2 Some issues for incontinence specific patient satisfaction measures

The most difficult issue is the paucity of scales: other than the GUTSS none were identified.

Regarding the GUTSS, the construction sample was a very small sample (n = 45) although its structure was confirmed in a larger sample. Its internal structure is possibly clumsy in light of the literature showing the effect of filter questions and administration mode (see section 5.7). Probably the GUTSS would be simpler if the two filter questions were removed (*Since the operation do you still have problems...* and *Over the past 4 weeks have you been disappointed...*) and the response categories of subsequent questions modified accordingly.

9.3. Summary assessment of incontinence specific patient satisfaction measures

When the GUTSS was assessed against the criteria presented in section 6, it appeared that:

- There was evidence of an underlying latent model of patient satisfaction, overlaid by the pragmatic experiences of both patients and clinicians;
- The coverage of the GUTSS was just adequate;
- The psychometric properties appeared to be good, although these need to be replicated in further studies; and
- Based on the preliminary evidence of three studies, the GUTSS appears to be responsive to both continence and satisfaction status.

10. Discussion and recommendations

10.1 Discussion of single-item patient satisfaction measures

Often single item assessments are asked in the context of health care, perhaps at the end of a consultation when the patient is in a dependent relationship with the clinician (e.g. see 116). Many clinicians use this as lead-in material for more detailed discussions of issues arising from treatment. Not surprisingly, given this relationship, most patients report very high levels of satisfaction (e.g. see 196, 222).

Quite apart from this administrative issue, there are substantive psychometric grounds for rejecting this model of patient satisfaction measurement.

It assumes that patient satisfaction is a single holistic dimension, which is adequately captured by a single item. As shown in section 6.1, however, patient satisfaction is a construct with at least 6 substantive dimensions (other than general satisfaction) implying that the level of satisfaction will vary depending upon which aspect of medical care is being assessed by the patient. Where different dimensions of care are assessed globally, there may be no way of determining which aspects are in need of improvement (51).

It is also assumed that a given response to a single item is reliable. However, none of the studies reviewed reported on the reliability of a single item, with the exception of Ware et al. (3) who reported that, based on Helmstadter's recommendation of a reliability estimate of 0.50 or greater for group comparisons (218), 75% of single items from the PSQ (Patient Satisfaction Questionnaire) failed this criteria, compared with 18% of the PSQ sub-scales. Where single item reliability has been systematically investigated elsewhere, the results suggest that single items are of doubtful reliability. Wyrwich (223), for example, reported 1-4 day test-retest kappa agreement of 0.64 to 0.73 for single item patient change scores; estimates which fall outside the normally accepted psychometric standard for reliability (generally reported to be >0.80).

An additional objection is that single-assessment items often incorporate bias because it is harder for a patient to admit to dissatisfaction if they are in a dependent situation when they are asked to respond to a closed item with a limited number of response options. For example, Tocchi et al. (222) used a single categorical item with three response levels (*very successful, moderately successful, failure*) to assess satisfaction with surgery for anal fissure repair; unsurprisingly 96% of patients reported they were satisfied. A similar situation arises with the scale used by Burgio et al. (224) which assessed patient satisfaction on a 3-point scale (*completely, somewhat, not at all*) in a study of behavioural training for urge incontinence; just 2% of participants endorsed the *not at all* category.

Finally, there are issues in the interpretation of single item measures and the cut off points used to determine satisfaction. A careful reading of the literature indicates that there is no equivalence in the interpretation of satisfaction scores between different studies. To illustrate this, consider three of the studies described above, the Fox and Storms (5), Grumbach et al. (200) and Miller et al. (184) studies. Fox and Storms dichotomized their 10-point scale at 10/9, where those scoring a "10" were classified as the "most satisfied" versus all others (1-9); similarly Jackson et al. (167) on a 5-point scale dichotomized respondents into those with "excellent" satisfaction (5) versus the rest (1-4). Grumbach et al. on the other hand, used a 5-point scale and dichotomized responses into those who were satisfied, defined as those indicating *very satisfied/satisfied*, versus those who were dissatisfied, defined as those indicating *uncertain/dissatisfied/very dissatisfied.* Although the reader is informed that 82% fell within the satisfied category, the numbers reporting uncertainty were not reported. In the Miller et al. (184) study where a forced choice 4-point scale (1 = very satisfied and 4 = very dissatisfied) was used respondents were dichotomized at the mid-point, thus forcing respondents to be either satisfied or unsatisfied. Perhaps this forced choice explains the high rates of dissatisfaction reported by these researchers (the proportion classified as being dissatisfied was between 26-100% for different interventions).

Given these psychometric difficulties with single item assessments, it should be expected that few cases will report dissatisfaction. Indeed, where patient satisfaction has been collected through qualitative research and open-ended questions more negative experiences have been reported (41, 51, 141).

10.2 Discussion of patient satisfaction instruments

Over 20 years ago Nguyen et al. (152) noted that it is almost impossible to make any meaningful comparisons between different patient satisfaction scale scores for two key reasons: first, that satisfaction scores across studies are so high that comparative interpretation is almost impossible,

and second that because there are almost no standard instruments that are widely used or reported it is difficult to equate scores from one study to another. They pointed out that most patient satisfaction questionnaires have been developed based on the researchers' views, i.e. that at best most patient satisfaction measures have face validity only. Hardy et al. (55) observed that most patient satisfaction measures were developed in hospitals in the USA, the implication being that they may have little applicability elsewhere.

Furthermore, Sitzia (34) in a review paper found that 81% of studies reported using a new patient satisfaction instrument and a further 10% reported modifying a previously existing instrument, yet 60% of studies examined failed to report any psychometric data.

The findings from this study are consistent with these earlier reports, although some improvements are evident, such as the reporting of basic psychometric tests (e.g. scale means, standard deviations and estimates of internal reliability). In general, however, the literature is still characterized by measures developed for particular studies where almost no psychometric data are available. Surprisingly, very few of the recognized patient satisfaction measures were identified as being used in studies other than by the original authors, and in many cases instruments were modified without appropriate psychometric testing. An additional difficulty uncovered in this review is that even where patient satisfaction measures have been available over time and are widely cited, there is almost no further psychometric work reported in the literature beyond that of the instrument developers. As a consequence, very few patient satisfaction instruments that met the study criteria were identified for inclusion.

Concerning the instruments reviewed, when these were examined against the psychometric criteria outlined in section 6, none met al.l the criteria.

10.2.1 Evidence of a latent construct of patient satisfaction

The definition of patient satisfaction varies by the purpose of the researchers. Regarding the instruments reviewed, the stated purposes are shown in Table 1. This reveals that five instruments were primarily developed to assess satisfaction with a clinical or medical consultation (ConsultQS, Linder-Pelz, MISS, PSQ-III and PVRQ).

Table 1   Purpose of instruments reviewed

		CSQ-18	CSQ-8	Consult SQ	LOPPS-18	Linder-Pelz	MISS-21	PSI	PSQ-III	PVRQ	GUTSS
Purpose	Clinical consultation	—	—	Yes	—	Yes	Yes	—	Yes	Yes	—
	Health care generally	Yes	Yes	—	—	—	—	—	—	—	—
	Incontinence surgery	—	—	—	—	—	—	—	—	—	Yes
	Life-threat care	—	—	—	—	—	—	Yes	—	—	—
	Nursing care	—	—	—	Yes	—	—	—	—	—	—
Admin.	Self-completed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Length (a)		M	S	M	M	M	M	M	L	S	S

Notes:
a = S: short instrument; M: medium length; L: long instrument.; details are given in Table 2.

Other than the ConsultSQ, all instrument developers referred to a theory of patient satisfaction in general. However, none invoked a theoretical model and then tested their manifest instrument against the model, with the exception of the Linder-Pelz scales and the PSQ-III. All instrument developers stated that the model of patient satisfaction used was created by reading the literature and consulting with clinicians. The more thorough instrument developers also consulted with patients; these instruments were the ConsultSQ, LOPSS, PSI, PSQ-III and GUTSS. Generally though, the impression was that instrument developers defined a theoretical model in accordance with their particular concerns, and then created the instrument around those concerns. This judgement rests on the fact that of the latent dimensions contributing to patient satisfaction presented in section 6.1, no instrument covered all these;

i.e. instrument developers chose to measure some of the theoretical parts of the patient satisfaction construct. Under these circumstances, the accepted interpretation of what is being measured in psychometric terms is to examine the content of the instrument.

- The preferred instruments are those where the stated purpose is consistent with the theories of satisfaction given in section 4, i.e. that patient satisfaction is primarily around the interaction relationship between the patient and the clinician. Based on this criteria the better instruments are the ConsultSQ, Linder-Pelz, MISS, PSQ-III and PVRQ.

Regarding the number of items, i.e. instrument length, in clinical work and epidemiological studies parsimony is important. Clinicians do not have the time to administer long instruments or the resources to score them, and in most research studies instrument batteries are administered where there are competing demands for the available space.

- Based on the need for parsimony, the shorter instruments are the CSQ-8, GUTSS and PVRQ.

10.2.2 Adequate coverage of the latent construct

Table 2 shows the coverage of the reviewed instruments, where the instrument items, based on the item content, have been mapped against the dimensions of patient satisfaction outlined in section 6.1. No instrument provided complete coverage; the best instruments were the Linder-Pelz scales and the PSQ-III. The PSQ-III, however, had very strong measurement of access to health care, other areas (most of which were measuring the patient's ability to pay for health care), and the technical skill of the clinician. Other instruments with particular emphases were the PSI for measuring patient participation, the PVRQ for measuring access, the ConsultSQ for technical skill and the CSQ-18/8 for general satisfaction. The LOPSS primarily measures information and patient-clinician relationships.

Table 2: Content validity (coverage)

	CSQ -18	CSQ -8	Consult SQ	LOPPS -18	Linder -Pelz	MISS -21	PSI	PSQ -III	PVRQ	GUTSS
Access & facilities	3	—	3	—	4	—	—	12	5	—
Information	—	—	2	5	4	3	3	3	1	2
Relationship	2	—	4	5	5	6	6	8	1	2
Participation	2	—	1	5	1	6	9	3	—	—
Technical skill	2	1	5	2	5	3	3	9	1	1
Effectiveness	1	1	—	—	—	3	—	—	—	3
Satisfaction general	5	6	3	1	1	—	—	6	1	—
Other	3	—	—	—	1	—	2	10	—	—
Total items	18	8	18	18	20	21	23	51	9	8 (a)

Notes:
a = Excludes two filter questions.

Areas that were poorly measured were treatment effectiveness or outcome (this was measured by the MISS, GUTSS and CSQ-18), and satisfaction generally (not measured by the MISS, PSI or GUTSS).

The PSQ-III, although offering the most comprehensive coverage of any instrument, is excessively weighted towards issues around access and payment — indeed these constitute 20/51 of its items, and this emphasis is likely to be misplaced in an Australian context (see section 10.3.1).

Although the CSQ-18 has good coverage, the CSQ-8's coverage is poor because 6/8 items are about satisfaction in general (the content is primarily about help being given, and needs being met).

Although the coverage of the ConsultSQ is very good in that it has items about 6 of the 7 satisfaction dimensions, this coverage is subject to issues around item repetition. For example, *I am totally satisfied with my visit to this doctor* and *I am not completely satisfied with my visit to the doctor*. Essentially these are the same item, one expressed positively and the other negatively. Altogether there are 6 pairs of such items in the instrument; thus 6/18 items are repetitive.

Although the coverage of the MISS is good, examination of the actual items reveals that most are

about the doctor being fully in charge of the health of the patient, particularly with respect to decision-making. For example, *The doctor told me just what my trouble is*, or *The doctor gave me a chance to say what was really on my mind*. As such the tone of the MISS is out of step with one of the key reasons for the rise of patient satisfaction measurement: to give patients a voice. The fundamental issue this tone raises is whether the MISS actually measures patient satisfaction at all rather than measuring the behaviour and attitude of the clinician towards the patient.

The items in the PSI pose a different problem because many are concerned with the inner feeling of the patient in coping with the life-threatening condition. For example, *Gone through a lot*, or *Felt out of control of situation*. Because these kinds of items comprise most items in the PSI it is difficult to know whether the PSI is measuring patient satisfaction or patient internalization of their experiences with their health care.

- Based on the coverage of content criteria, the better instruments are the Linder-Pelz and CSQ-18.

10.2.3 Data distribution and ceiling effects

Regarding the response scale used in the instruments, Likert scales are used by the ConsultSQ, LOPPS, MISS, and PSQ-III. The other instruments use Guttman scales.

All instruments had different score ranges; but all (with the exception, perhaps, of the GUTSS) suffered from assigning high levels of satisfaction. To examine whether these scores were so high that ceiling effects were likely Cummins' %SM (138) was computed for all instruments, where the standard was 75%SM of the theoretical score range. Instruments with %SM scores above this standard are more likely to suffer ceiling effects, whereas instruments below this standard are less likely to. As shown in Table 3, score ranges were not reported for the Linder-Pelz, MISS, PSQ-III or PVRQ instruments.

Table 3: Scoring of the instruments

		CSQ-18	CSQ-8	Consult SQ	LOPPS-28	Linder-Pelz	MISS-21	PSI	PSQ-III	PVRQ	GUTSS
Responses	Likert-type	—	—	Yes	Yes	—	Yes	—	Yes	—	—
	Guttman-type	Yes	Yes	—	—	—	—	Yes	—	Yes	Yes
	N. points	4	4	5	5	—	7	7	5	5	5
Scale range	Minimum	18	8	18	28	—	—	23	—	—	0
	Maximum	72	32	90	140	—	—	161	—	—	32
	Mean scores (range)	49-55	24-27	52-77	115-120	—	—	122	—	—	21
Cummins %SM range (a)		56-68%	65-78%	46-82%	77-82%	—	77% (b)	72%	—	—	65-93%

Notes:
a = mean score as percentage of potential scale range.
b = not computed from data, but reported in the papers.

The LOPSS instrument was that with the highest %SM, suggesting ceiling effects were more likely; a conclusion consistent with O'Connell et al. who reported high levels of acquiescent response bias through item insensitivity (145). The same situation has been reported for the PVRQ where less than 5% of respondents endorsed the lower two response categories.

An important point to note is that the %SM values presented in Table 3 show that there is a divide in %SM scores by response scale type: those instruments using Likert scales (ConsultSQ, LOPPS, and MISS) all obtained higher %SM scores at the upper end of the scale than did the instruments using Guttman scales. This finding is consistent with the literature (186, 187). Since high patient satisfaction scores are both endemic and problematic in the measurement of patient satisfaction, this finding would, *prima facie*, suggest that instruments with Guttman-type scales might be preferred.

- Based on scoring ranges and the likelihood of ceiling effects, the better instruments are the GUTSS, CSQ-8 and CSQ-18.

As the individual instrument reviews show, there is very little sustained evidence of validity for any of the instruments reviewed. Generally, the available evidence is from the instrument developers and perhaps one or two other research teams. This evidence is summarized in Table 4. When reading the table, it should be remembered that "Yes" means that some evidence is available, and a null entry (—) that no evidence was reported in the papers reviewed (this does not mean that the instruments have not been assessed against the validation criteria, just that no evidence was uncovered).

10.2.4 Validity evidence

Table 4:   Validity evidence

Validation	CSQ -18	CSQ -8	Consult SQ	LOPPS -18	Linder -Pelz	MISS -21	PSI	PSQ -III	PVRQ	GUTSS
Ecological (a)	—	—	Yes	Yes	—	—	Yes	Yes	—	Yes
Factor analysis (b)	Yes	Yes	Yes	Yes	—	Yes	—	Yes	—	Yes
Concurrent (c)	Yes	Yes	Yes	Yes	—	—	Yes	Yes	Yes	Yes
Convergent (d)	Yes	Yes	Yes	No	—	—	Yes	Yes	Yes	Yes
Predictive (e)	—	—	—	—	—	—	—	—	Yes	—

Notes:
a = involving patients during instrument construction.
b = a measure of construct validity because it examines the relationships between items.
c = correlation with other measures of patient satisfaction.
d = assessed against patients groups known to be satisfied/dissatisfied or treatment success/failure.
e = assessed by power to predict future outcomes.

Regarding the validity evidence for the LOPPS, it must be recognised that although there is some evidence available, this does not fully support the LOPPS in that the LOPPS did not show significant variation in scores by groups known to differ in satisfaction (based on age), and that it did not register patient dissatisfactions.

- The instruments that have the most validity evidence are the PSQ-III and ConsultSQ. Although there is validity evidence for the GUTSS, it comes from the one study.

Table 5:   Reliability and responsiveness evidence

Estimates	CSQ -18	CSQ -8	Consult SQ	LOPPS -18	Linder -Pelz	MISS -21	PSI	PSQ -III	PVRQ	GUTSS
Cronbach _ (a)	0.78 -0.91	0.92- 0.93	0.91	0.93- 0.98	—	0.93	—	0.88	—	0.83- 0.84
Cronbach _ (b)	—	—	0.67- 0.94	—	0.49- 0.81	0.66- 0.96	—	0.23- 0.93	—	0.76- 0.93
Split – half correlation	0.82	—	—	—	—	—	—	—	—	—
Test – retest (c)	—	—	—	—	—	—	0.86	—	—	—
Responsiveness (d)	Yes	Yes	Yes	?	—	Yes	Yes	Yes	?	Yes

Notes:
a = for summated instrument scores. Where different estimates are available, the lower and upper estimates are given.
b = for scales within the instrument. Where different estimates are available, the lower and upper estimates are given.
c = intra-class correlation.
d = ability to detect differences between groups of patients.

10.2.5 Reliability and responsiveness evidence

Three forms of reliability were reported, Cronbach _, split-half and test-retest. The data are given in Table 5. The conventional range for internal consistency as assessed by Cronbach _ is from 0.70 to 0.90 where cohorts are to be compared (191). However, _ is a function of both the correlation between items within a scale and of the scale length (225). The implication is that where _ exceeds 0.90 there is likely to be redundancy in the scale because the same concept is being asked twice or more often.

High _ values can also be brought about where there is little variance in responses; typically where the range of responses is truncated. In this study this was the case with many instruments because respondents mainly utilized the first two or three categories (i.e. where most respondents ticked a 1 or 2 in a scale of 5). Equally where _ < 0.70 the scale is likely to be made up of items that are too disparate and that do not form a homogenous scale.

Table 5 shows that most of the instruments were within the acceptable range, albeit at the upper end. No instrument was reported as being unreliable, although several instruments' scales fell outside the conventional range for reliability (scales on the ConsultSQ, Linder-Pelz, MISS and PSQ). It is likely, however, that the LOPPS contained redundant items; and there may be redundancy in the MISS, and the CSQ-18/8.

As reported in Table 2, in the case of the CSQ-18/8 this may be a function of the lack of breadth of measurement because a high proportion of the items measure the same dimension (e.g. in the case of the CSQ-8 6/8 items are concerned with overall or general satisfaction, thus there is likely to be repetitive measurement). In the case of the LOPPS, although the items are spread out over different dimensions, this situation is almost certainly caused by insensitive items leading to the endorsement of high end categories indicating satisfaction (as shown in Table 3).

Regarding responsiveness, the evidence suggests that all instruments were responsive, although there was mixed evidence for the LOPPS and for the PVRQ, and insufficient for the Linder-Pelz scales.

- Based on reliability criteria, the better instruments are the ConsultSQ (although there are some questions over its sub-scales), PSI, and GUTSS.

10.3 Additional criteria

In addition to the psychometric criteria discussed in the previous section, there are two contextual issues that are relevant in assessing patient satisfaction instruments for use in Australian settings.

10.3.1 Relevance to the Australian health care system

As shown in the literature review section (section 3), patient satisfaction measures may be culturally specific (78-83). This implies that different instruments are not equally relevant in different settings.

The Australian health care system is characterised by multi-level funding: the Commonwealth and State Governments provide about 70% of all health costs, primarily through Medicare and the Pharmaceutical Benefits Scheme (both of which provide subsidized services) and the funding of public hospitals where emergency and outpatient services are provided free of charge (226). Safety nets for high consumption users who have limited resources also apply (e.g. for those on unemployment or pension benefits). For most Australians the first contact for health care is a general practitioner (GP). Access to GP services, from the patient perspective, is uncapped. Although GPs are located across Australia, thus ensuring ready access to primary health care, there are some distributional issues that affect access to and quality of services, mostly in country areas (227, 228). Additionally, about 49% of the population has private health insurance for hospital and ancillary health care (226).

The implication is that patient satisfaction measures which focus on costs borne by the patient, access to health care or emphasis on the buildings within which care is provided are likely to be less relevant in the Australian context. Additionally, the literature suggests that these issues play little part in determining patient satisfaction (see section 5).

The instruments which have scales or items measuring these aspects of care are the CSQ-18, the Linder-Pelz scales, the PSQ-III and the PVRQ. The CSQ-18 has two items measuring the promptness of being seen, and the comfort and attractiveness of the facility and building. The Linder-Pelz scales have 3 items assessing a patient's entitlements (rights to see a clinician immediately, not wait for an appointment and the right to tell the clinician everything) and 3 questions on access (ease of getting to the clinic, the comfort of the waiting room, and having to wait). The PSQ-III has 7 items assessing the effect of health costs (e.g. going without services due to the cost), and 9 items assessing access to care (e.g. that health care facilities should have longer opening hours). The PVRQ has items covering the time to get an appointment, the convenience of the medical rooms, contacting the medical rooms by telephone, the time spent waiting at the medical rooms. Additionally, the LOPSS has an item on responsiveness to the patient ringing the nurse call bell; which is a reflection of its primary use inside hospitals.

Although some of these items (e.g. the convenience of the medical rooms) may be appropriate for Australians living in locations where access to services is compromised, it is doubtful these items are relevant beyond this.

- This criteria would suggest that the better instruments are the CSQ-8, the ConsultSQ, the MISS-21, the PSI and the GUTSS.

10.3.2 Instrument users

There are three main users of patient satisfaction instruments: (a) clinicians, (b) specialists, and (c) researchers or program evaluators.

At the clinical level, patient satisfaction is likely to be related to the clinical management of individual patients. Assessment may be near the end of a consultation, and the patient's response may be used by the clinician as a discussion starter. At the clinician level there are time and data collection issues: busy clinicians may not have the time and expertise required to use long, multi-scaled instruments. Additionally, they may need a measure where the scoring is instant so that they can discuss the results with the patient immediately. Data collection should be as brief as possible and there should be few data analysis demands upon clinicians.

Table 6: Additional criteria

Criterion	Single item	CSQ-18	CSQ-8	Consult SQ	LOPPS-18	Linder-Pelz	MISS-21	PSI	PSQ-III	PVRQ	GUTSS
Australian relevance	Yes	—	Yes	Yes	—	—	Yes	Yes	—	—	Yes
Best for:											
Clinicians	Yes	—	—	—	—	?	—	—	—	—	—
Specialists	?	Yes	—	?	—	?	—	?	—	Yes	Yes
Researchers	No	Yes	—	?	—	?	—	?	—	—	Yes

Specialists working with patients may have somewhat different needs. Although when working with individual patients their needs may be similar to those of clinicians, many specialists also need more information and are often involved in research or evaluation. Instruments used at this level may need to possess sufficient nomological evidence to be used at the case level; i.e. for individual patient assessment.

Researchers and program evaluators' needs centre round data that are useful for answering research questions where analyses are group-based; where data collection procedures may be remote; and where findings are aimed at demonstrating the effect of new treatments or at influencing policy decisions.

Finally, in a health care system committed to evidence-based practice, basic data which is collected and held at the clinician level should be suitable for transfer to local level analysis and also to research settings (e.g. for inclusion in incontinence monitoring or surveillance).

These different needs imply that at each level different patient satisfaction measures may be needed. Based on the reviews of instruments in sections 7, 8 and 9, when assessed against this criterion, the rankings of instruments presented in Table 6 was made. This suggests that:
- For clinicians working with individual patients a single global question may be sufficient.
- For specialists involved in research studies, in addition to a single global question, short instruments assessing satisfaction with incontinence care and treatment outcomes may be needed. The preferred instruments would be the CSQ-18, ConsultSQ and GUTSS.
- For researchers, in addition to a single global question and short patient satisfaction instruments, generic patient satisfaction measures may be needed where study results are to be compared with those obtained from other studies in other fields of medicine. The preferred instrument would be the CSQ-18.

10.4 Summary assessment

A brief review of each instrument is given, followed by some summary assessment comments.

Single item measures

Single item patient satisfaction measures may have a place in clinical practice where the item serves as an introduction to a broader discussion between the patient and the clinician. Where this is the case,

however, the response to this item should not be interpreted as a measure of patient satisfaction due to its psychometric limitations. It is possible that such an item may also have a place in other settings as a global estimate of satisfaction. Beyond this, however, single items should be avoided because the available evidence suggests that they do not meet the criteria for good psychometric measurement:

- There is almost no evidence of an underlying latent model of patient satisfaction, unless it is assumed that an holistic manifest model forms such a construct;
- Because single items represent holistic measurement, they cannot provide adequate coverage of a multi-dimensional construct;
- None of the studies consulted in this review reported on the psychometric properties of single item measures; and
- No study examined the responsiveness of a single item measure.[1]

Generic patient satisfaction instruments

No generic patient satisfaction instrument satisfied all the criteria of the study described in sections 4, 5 and 6. Brief summaries on each instrument are given, followed by rating the instruments.

CSQ-18	Although the development of the CSQ-18 followed good practice and it has good psychometric properties, there are questions over the number of items measuring general satisfaction (30%), and the distribution of item responses (less than 10% of cases endorse the lower half of the scales). It is likely this lack of response distribution combined with the emphasis on general satisfaction explains the high Cronbach _, which instead of indicating good psychometrics suggests redundancy. Additionally, there are questions about its cultural relevance in an Australian context.
CSQ-8	The CSQ-8, while very attractive because of its brevity, is restricted in its measurement. This is because 6/8 items are general satisfaction items. The implication is that the removal of items from the CSQ-18 resulted in retaining items that were very similar to each other. As with the CSQ-18, the distribution of item responses is a concern, leading to the same conclusion regarding the measure's reliability.
ConsultSQ	The developmental procedures followed with the ConsultSQ were excellent. It has good coverage of patient satisfaction dimensions measuring general satisfaction, professional care, relationship with the clinician and consultation time, and there is evidence of validity and reliability. However, the instrument is disappointing because of the repetition of items as described above and that there is no item measuring treatment outcomes. The repetition implies that the ConsultSQ gains its reliability (and perhaps some validity evidence) through replication.
LOPPS	Although the LOPPS has been re-written several times to improve its psychometric properties, the instrument appears to be of limited value. There are several reasons for this, including difficulties with acquiescent response bias, poor distribution of data on response scales, evidence of a ceiling effect on total scores, an apparent inability to discriminate between groups known to vary by satisfaction, and a failure to be sensitive to patient dissatisfaction. Because it is primarily concerned with nursing care in an institution, it may have limited application in Australian general practice settings.
Linder-Pelz	These scales were designed to test the expectancy theory of patient satisfaction, rather than to measure patient satisfaction *per se*. Subject to this caveat, there is insufficient psychometric information available to make a good assessment of the scales. However, there are problems of negativity with the doctor conduct scale (all 10 items are negative in tone), and the scales may be less relevant in an Australian context because of the emphasis on entitlements and access.

---

1. Although it might be thought that Fox and Storms reported on the responsiveness of their item, because they assessed responses against known correlates of patient satisfaction, their correlates did not include any variables related to a health condition or to a particular health care intervention. The closest variable to these criteria for responsiveness was a dichotomous question (Yes/No) relating to chronic conditions.

MISS	The MISS has fair psychometric properties, including reasonable coverage. There are some serious questions about its internal structure, because two different research teams reported that the factorial structure was not congruent with that of the instrument developers, and that there were non-contributing items. Additionally, there are issues of the tone of the items which present patients with a paternalistic perspective, and questions regarding what the MISS is actually measuring. There is some evidence that the MISS may be culturally specific.
PSI	The conceptual base and methodological steps taken during instrument construction are exemplary. It is a well constructed, reliable and (apparently) valid instrument, although on this latter point there is insufficient evidence. However, almost certainly because it was designed to be used with patients suffering life-threatening conditions, the item content reflects the patient experiences rather than their satisfaction with care. Additionally, unlike the other instruments reviewed here, the PSI was designed for interviewer administration.
PSQ-III	This was the most rigorously developed instrument of any reviewed. Generally, its psychometric properties are very good in almost all respects, although there are concerns with some of the scales where poor reliability estimates have been reported. There is also some evidence that the PSQ-III reflects life satisfaction rather than satisfaction with health care. The other difficulty with the PSQ-III is its length: most clinicians and researchers would not have the time or resources to use it. Finally, its validity (as an instrument) is heavily constrained by the emphasis on those areas of care which may be less relevant in an Australian context, particularly the emphasis measuring access to health care services and the capacity of the patient to afford the health care needed.
PVRQ	At face value, this appears to be an excellent instrument, which was rigorously developed and tested. However, there are reported difficulties with the distribution of item responses and also with the responsiveness of the measure. Additionally, 4/9 items are concerned with access issues that may not be relevant in an Australian context.

Incontinence specific patient satisfaction instruments, and the GUTSS

The argument that there should be incontinence-specific patient satisfaction scales should be carefully scrutinized. In general, unless there is reason to believe that the effect of incontinence treatment is systematically different to that of other interventions for other health conditions, then in the interests of comparability generic patient satisfaction measures should be used. This is because where generic instruments have widespread use, their psychometric properties become well known and norms are also available to assist researchers interpret their work.

In addition to this general argument, the only incontinence-specific patient satisfaction measure identified, the GUTSS, is affected by three important issues. Because it was developed to assess surgery satisfaction, whether it is applicable to all incontinence interventions is open to question. The structure of the GUTSS includes two filter questions, thus making it unnecessarily complex to use and score. The coverage of the items is less than ideal. Apart from these issues, it appears to have good psychometric properties.

Before it could be recommended, the GUTSS needs to have further developmental work carried out in terms of its internal structure.

11. Recommendations

Based on the criteria for measuring patient satisfaction outlined in sections 4, 5 and 6, and the reviews of instruments in sections 8, 9 and 10, it was possible to compare the multi-item instruments reviewed. This is done in Table 7.

For each of the study criteria described in section 10, each instrument was rated on a scale of 0, 1 or 2, where 0 indicated the instrument did not meet the criterion, 1 indicated there was some evidence the instrument partly met the criterion, and 2 indicated the instrument met the criterion. Where insufficient information was available, no rating was given.

The criteria were weighted to reflect their relative importance. For example, validity evidence was deemed more important than being based on an explicit theory of patient satisfaction. The scale of weights used was 1, 2 and 3, where 1 indicated the criterion was important, 2 that it was very important and 3 that it was extremely important. The details are given in the notes to Table 7.

Scores on each criterion were weighted and then summed. The weighted values were then divided by the number of criteria. The results are presented in Table 7. Although there were differences between the instruments, there was no "stand out" instrument and none of the instruments reviewed could be considered truly satisfactory.

*

The key finding from this review of patient satisfaction instruments is that no instrument has been sufficiently validated for its use in Australia to be automatically recommended. There are three key reasons for this finding:

A.   There is evidence throughout the literature that patient satisfaction is culturally specific. It cannot be assumed that an instrument that is relevant, valid and reliable in one culture retains those properties in another culture. Thus instruments developed overseas may not be appropriate in Australian settings.

B.   There is no agreed theoretical model of patient satisfaction or of its constituent parts. As this review has shown, the consequence is that instrument designers have proceeded on an *ad hoc* basis with the result that there are thousands of patient satisfaction measures available.

C.   Among recognised generic patient satisfaction instruments there is insufficient evidence of their psychometric properties for any instrument to be fully accepted as possessing a nomological net of validity evidence.

Table 7:   Summary of instruments

Criteria	Weight	CSQ-18	CSQ-8	Consult SQ	LOPPS-18	Linder-Pelz	MISS-21	PSI	PSQ-III	PVRQ	GUTSS
Theory & purpose (a)	1	1	1	2	1	2	2	1	2	2	1
Parsimony (b)	2	1	2	1	1	1	1	1	0	2	2
Coverage (c)	2	2	0	1	1	1	1	1	2	1	1
%SM (d)	1	1	1	1	0	—	0	1	—	—	1
Validity evidence (e)	3	1	1	2	1	—	0	1	2	1	1
Reliability evidence (f)	3	1	1	1	0	1	1	2	1	0	1
Aust. relevance (g)	3	1	1	1	0	0	1	1	0	0	1
Weighted total (h)		2.43	2.14	2.71	1.14	1.29	1.71	2.57	2.14	1.57	2.43

Notes:

a = Based on Table 1. Clinical = 2; Other = 1; None = 0.

b = Based on Table 2. <10 items = 2, 10 to 24 items = 1, 25 or more items = 0.

c = Based on Table 2. Seven dimensions measured = 2, 4 to 6 dimensions = 1; 0-3 dimensions = 0.

d = Based on Table 3. <75%SM = 1; >75%SM = 0.

e = Based on Table 4. 4 "Yes" = 2, 2-3 "Yes" = 1; 0-1 "Yes" = 0.

f = Based on Table 5. Reliability estimate in the range of 0.70 to 0.90 = 2; >0.90 = 1; < 0.70 = 0. If evidence of unresponsiveness, then 1 has been subtracted from the score.

g = Based on Table 6.

h = Based on (Score X Weight)/N. categories.

The recommendations below should be read with these caveats in mind. They are:

1. That a single item patient satisfaction measure should be adopted for use in Australian settings by clinicians wishing to assess the satisfaction of their patients "on the spot". Strategies should be put in place to encourage clinicians to adopt this measure as a common metric across Australia. Encouragement should be given to specialists and researchers to also include this common metric in their work. In this way a bank of shared understanding will be progressively established. It may be possible that a single item measure could be drawn from the continent-specific or generic instruments recommended in #2 or #3.

2. That if a continent-specific measure is required, a short, valid and reliable incontinence-specific patient satisfaction measure be developed for use by Australian specialists. Strategies should be put in place to encourage Australian researchers to adopt this measure as well. A suitable starting point would be revision of the GUTSS through removal of the two filter questions. This could be done quickly using the available construction database to model the effect of this change.

3. The better generic patient satisfaction instruments identified in this report are the ConsultSQ, PSI and CSQ-18. For the reasons outlined in the report, however, none of these can be recommended in their current form. Therefore it is recommended that a short, valid and reliable generic patient satisfaction measure be developed for use by Australian researchers, or the revision of one or more of these instruments be undertaken. In the interim period it is recommended that the ConsultSQ be used.

12 References

1. Thomas S, Moore K, Nay R, Fonda D, Marosszeky N, Sansoni J, et al. Continence Outcomes Measurement Suite Project Final Report. Melbourne: La Trobe University; 2004 April.

2. Donabedian A. The quality of care: how can it be assessed? Journal of the American Medical Association 1988;260(12):1743-48.

3. Ware JE, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. Evaluation and Program Planning 1983;6:247-263.

4. Linder-Pelz S. Toward a theory of patient satisfaction. Social Science and Medicine 1982;16:577-582.

5. Fox JG, Storms DM. A different approach to sociodemographic predictors of satisfaction with health care. Social Science and Medicine 1981;15(5):557-64.

6. Fitzpatrick R, Hopkins A. Problems in the conceptual framework of patient satisfaction research: an empirical exploration. Sociology of Health and Illness 1983;5(3):297-311.

7. Donabedian A. Explorations in Quality Assessment and Monitoring: Vol 1 The Definition of Quality and the Approaches to its Assessment. MI: Health Administration Press; 1980.

8. Hawthorne G, Harmer C. GUTSS: The Genito-Urinary Treatment Satisfaction Scale Study. In. Melbourne: Centre for Health Program Evaluation; 2000. p. 50.

9. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Evaluation and Program Planning 1979;2(3):197-207.

10. Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. British Journal of General Practice 1990;40(December):487-490.

11. La Monica EL, Oberst MT, Madea AR, Wolf RM. Development of a patient satisfaction scale. Research in Nursing and Health 1986;9(1):43-50.

12. Wolf MH, Putnam SM, James SA, Stiles WB. The Medical Interview Satisfaction Scale: development of a scale to measure patient perceptions of physician behavior. Journal of Behavioral Medicine 1978;1(4):391-401.

13. Wolf MH, Stiles WB. Further development of the Medical Interview Satisfaction Scale. In: American Psychological Association Convention; 1981; Los Angeles; 1981.

14. Guyatt GH, Mitchell A, Molloy DW, Capretta R, Horsman J, Griffith L. Measuring patient and relative satisfaction with level or aggressiveness of care and involvement in care decisions in the context of life threatening illness. Journal of Clinical Epidemiology 1995;48(10):1215-24.

15. Rubin HR, Gandek B, Rogers WH, Kosinski M, McHorney CA, Ware JE, Jr. Patients' ratings of outpatient visits in different practice settings. Results from the Medical Outcomes Study. Journal of the American Medical Association 1993;270(7):835-40.

16. Charter RA, Feldt LS. Meaning of reliability in terms of correct and incorrect clinical decisions: the art of decision making is still alive. Journal of Clinical and Experimental Neuropsychology 2001;23(4):530-7.

17. Cronbach J, Meehl P. Construct validity in psychological tests. Psychological Bulletin 1955;52(4):281-302.

18. Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health related quality of life. Quality of Life Research 1999;8:209-224.

19. Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Annals of Medicine 2001;33:358–370.

20. National Primary Care research and Development Centre. General Practice Assessment Survey (GPAS) Manual. Manchester: University of Manchester; 1999.

21. Holcomb WR, Parker JC, Leong GB, Thiele J, Higdon J. Customer satisfaction and self-reported treatment outcomes among psychiatric inpatients. Psychiatric Services 1998;49(7):929-34.

22. Hudak PL, Hogg-Johnson S, Bombardier C, McKeever PD, Wright JG. Testing a new theory of patient satisfaction with treatment outcome. Medical Care 2004;42(8):726-39.

23. Aragon SJ, Gesell SB. A patient satisfaction theory and its robustness across gender in emergency departments: a multigroup structural equation modeling investigation. American Journal of Medical Quality 2003;18(6):229-41.

24. Grol R, Wensing M, Mainz J, Ferreira P, Hearnshaw H, Hjortdahl P, et al. Patients' priorities with respect to general practice care: an international comparison. European Task Force on Patient Evaluations of General Practice (EUROPEP). Family Practice 1999;16(1):4-11.

25. Grol R, Wensing M. Patients Evaluate General/Family Practice: The EUROPEP Instrument: The Task Force on Patient Evaluations of General Practice Care; 2000.

26. Dozier AM, Kitzman HJ, Ingersoll GL, Holmberg S, Schultz AW. Development of an instrument to measure patient perception of the quality of nursing care. Research in Nursing and Health 2001;24(6):506-17.

27. Baker R. The reliability and criterion validity of a measure of patients' satisfaction with their general practice. Family Practice 1991;8(2):171-177.

28. Kane RL, Maciejewski M, Finch M. The relationship of patient satisfaction with care and clinical outcomes. Medical Care 1997;35(7):714-30.

29. Cryns AG, Nichols RC, Katz LA, Calkins E. The hierarchical structure of geriatric patient satisfaction. An Older Patient Satisfaction Scale designed for HMOs. Medical Care 1989;27(8):802-16.

30. Ruggeri M, Dall'Agnola R. The development and use of the Verona Expectations for Care Scale (VECS) and the Verona Service Satisfaction Scale (VSSS) for measuring expectations and satisfaction with community-based psychiatric services in patients, relatives and professionals. Psychological Medicine 1993;23:511-523.

31. Margolis SA, Al-Marzouq S, Revel T, Reed RL. Patient satisfaction with primary health care services in the United Arab Emirates. International Journal for Quality in Health Care 2003;15(3):241-9.

32. Westaway MS, Rheeder P, Van Zyl DG, Seager JR. Interpersonal and organisational dimensions of patient satisfaction: the moderating effects of health status. International Journal for Quality in Health Care 2003;15(4):337-44.

33. Beck EJ, Griffith R, Fitzpatrick R, Mandalia S, Carrier J, Conlon C, et al. Patient satisfaction with HIV service provision in NPMS hospitals: the development of a standard satisfaction questionnaire. NPMS Steering Group. AIDS Care 1999;11(3):331-43.

34. Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. International Journal for Quality in Health Care 1999;11(4):319-28.

35. Johannsson HO, Graf W, Pahlman L. Long-term results of haemorrhoidectomy. European Journal of Surgery 2002;168(8-9):485-9.

36. Porter R. The Greatest Benefit to Mankind: A Medical History of Mankind. New York: Norton & Co; 1997.

37. Campanella HC, Campanella PM, Grayson K. Factors affecting Department of Defense patient satisfaction in a military emergency department. Military Medicine 2000;165(5):396-402.

38. Williams B. Patient satisfaction: a valid concept? Social Science and Medicine 1994;38(4):509-516.

39. Donabedian A. Evaluating the quality of medical care. Milbank Memorial Fund Quarterly 1966;44(3):Suppl:166-206.

40. Donabedian A. The quality of medical care: a concept in search of a definition. Journal of Family Practice 1979;9(2):277-84.

41. Carr-Hill R. The measurement of patient satisfaction. Journal of Public Health Medicine 1992;14(3):236-249.

42. Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. Health Technology Assessment 2002;6(32):1-244.

43. Rowbotham MC. Centralized or decentralized service; Faulkner Hospital combiness them for greater efficiency and patient satisfaction. Modern Hospital 1953;80(6):114-6.

44. Caplan EK, Sussman MB. Rank order of important variables for patient and staff satisfaction with outpatient service. Journal of Health and Human Behavior 1966;7(2):133-7.

45. Bashshur RL, Metzner CA, Worden C. Consumer satisfaction with group practice, the CHA case. American Journal of Public Health and the Nation's Health 1967;57(11):1991-9.

46. James G. The medical school of the future and its role in the community. Journal of the American Medical Association 1967;202(5):415-8.

47. Rouse MO. Organizing and delivering health care. 3. Journal of the American Medical Association 1967;202(4):305.

48. Abdellah FG, Levine E. Developing a measure of patient and personnel satisfaction with nursing care. Nursing Research 1957;5(3):100-8.

49. Wilson W, Goldschmidt P. Quality Management in Healthcare. Sydney: McGraw-Hill Book Company Australia; 1995.

50. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. Social Science and Medicine 1997;45(12):1829-1843.

51. Locker D, Dunt D. Theoretical and methodological issues in sociological studies of consumer satisfaction with medical care. Social Science and Medicine 1978;12:283-292.

52. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. Journal of Evaluation in Clinical Practice 1999;5(4):401-16.

53. Larsen D, Rootman I. Physician role performance and patient satisfaction. Social Science and Medicine 1976;10:29-32.

54. Fitzpatrick R. Surveys of patient satisfaction: I – important general considerations. British Medical Journal 1991;302:887-889.

55. Hardy G, West M, Hill F. Components and predictors of patient satisfaction. British Journal of Health Psychology 1996;1:65-85.

56. Chung CC, Ha JP, Tai YP, Tsang WW, Li MK. Double-blind, randomized trial comparing Harmonic Scalpel hemorrhoidectomy, bipolar scissors hemorrhoidectomy, and scissors excision: ligation technique. Diseases of the Colon and Rectum 2002;45(6):789-94.

57. Lindsey I, Smilgin-Humphreys MM, Cunningham C, Mortensen NJ, George BD. A randomized, controlled trial of fibrin glue vs. conventional treatment for anal fistula. Diseases of the Colon and Rectum 2002;45(12):1608-15.

58. Hara I, Kawabata G, Miyake H, Nakamura I, Hara S, Okada H, et al. Comparison of quality of life following laparoscopic and open prostatectomy for prostate cancer. Journal of Urology 2003;169(6):2045-8.

59. Drain M. Quality improvement in primary care and the importance of patient perceptions. Journal of Ambulatory Care Management 2001;24(2):30-46.

60. Valentine NB, de Silva A, Kawabata K, Darby C, Murray CJL, Evans DB. Health system responsiveness: concepts, domains and operationalization. In: Murray CJL, Evans DB, editors. Health Systems Performance Assessment: Dabates, Methods and Empiricism. Geneva: World Health Organisation; 2003. p. 573-596.

61. Lewis J. Patient views on quality care in general practice: literature review. Social Science and Medicine 1994;39(5):655-670.

62. Ross C, Steward C, Sinacore J. The importance of patient preferences in the measurement of health care satisfaction. Medical Care 1993;31(12):1138-1149.

63. Cleary P, Edgeman-Levitan D, Roberts M, Maloney T, McMullan W, Walker J, et al. Patients evaluate their hospital care: a national survey. Health Affairs 1991;10(4):254-67.

64. McIvor S. Obtaining the Views of Inpatients and Users of Casualty Departments. London: Kings Fund Centre; 1992.

65. Department of Finance and Australian Public Service Board. Evaluating government programs: Financial management improvement program. Canberra: Commonwealth of Australia, AGPS; 1987.

66. Shafritz J, Russell EW. Introducing public administration. United States: Addison-Wesley Educational Publishers Inc.; 1997.

67. Corbett D. Australian Public Sector Management. Second Edition ed. St Leonards: Allen & Unwin; 1998.

68. Draper M, Hill S. The Role of Patient Satisfaction Surveys in a National Approach to Hospital Quality Management: Commonweatlh of Australia, AGPS.; 1996.

69. Draper M. Involving Consumers in Improving Hospital Care: Lessons from Australian Hospitals: Commonwealth of Australia; 1997.

70. TQA Research Consulting. Victorian Patient Satisfaction Monitor Annual Survey Report, Year One, 2000-2001. Melbourne: Department of Human Services; 2002.

71. TQA Research Consulting. Victorian Patient Satisfaction Monitoring Annual Survey Report, Year Two, 2001-2002. Melbourne: Department of Health Services; 2003.

72. Quint T. Patient Satisfaction Survey: Executive Summary. Melbourne: TQA Research for the Dept. Health Victoria; 1995 14th November.

73. Cartwright A. Patients and Their Doctors. London: Routledge & Kegan Paul; 1967.

74. Korsch BM, Gozzi EK, Francis V. Gaps in doctor-patient communication. 1. Doctor-patient interaction and patient satisfaction. Pediatrics 1968;42(5):855-71.

75. Kincey J, Bradshaw P, Ley P. Patients' satisfaction and reported acceptance of advice in general practice. Journal of the Royal College of General Practitioners 1975;25(157):558-66.

76. Suchman EA. Social patterns of illness and medical care. Journal of Health and Human Behavior 1965;106:2-16.

77. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Medical Journal of Australia 1997;166:182-186.

78. Meakin R, Weinman J. The "Medical Interview Satisfaction Scale" (MISS-21) adapted for British general practice. Family Practice 2002;19(3):257-63.

79. Henderson C, Hales H, Ruggeri M. Cross-cultural differences in the conceptualisation of patients' satisfaction with psychiatric services – content validity of the English version of the Verona Service Satisfaction Scale. Social Psychiatry and Psychiatric Epidemiology 2003;38(3):142-8.

80. Labarere J, Fourny M, Jean-Phillippe V, Marin-Pache S, Patrice F. Refinement and validation of a French in-patient experience questionnaire. International Journal of Health Care Quality Assurance 2004;17(1):17-25.

81. Firestone DN, Jimenez-Briceno L, Reimann JO, Talavera GA, Polonsky WH, Edelman SV. Predictors of diabetes-specific knowledge and treatment satisfaction among Costa Ricans. Diabetes Education 2004;30(2):281-292.

82. Zebiene E, Razgauskas E, Basys V, Baubiniene A, Gurevicius R, Padaiga Z, et al. Meeting patient's expectations in primary care consultations in Lithuania. International Journal of Quality in Health Care 2004;16(1):83-9.

83. Baker R, Mainous AG, 3rd, Gray DP, Love MM. Exploration of the relationship between continuity, trust in regular doctors and patient satisfaction with consultations with family doctors. Scandinavian Journal of Primary Health Care 2003;21(1):27-32.

84. WHO. The World Health Report 2000. Health Systems: Improving Performance. Geneva: World Health Organisation; 2000.

85. Harding WG, 2nd, Silvers G. What the patient really wants in hospital accommodations. Ohio State Medical Journal 1968;64(12):1337-9.

86. Rassi J, Harford E. An analysis of patient attitudes and reactions to a clinical hearing aid selection program. Asha 1968;10(7):283-90.

87.   Roy AL, Prasad BG. A socio-medical study of adult male patients attending the general medical out-patient department of a medical college hospital. Indian Journal of Public Health 1967;11(2):73-81.

88.   Henley B, Davis M. Satisfaction and dissatisfaction: a study of the chronically ill aged patient. Journal of Health and Social Behaviour 1967;8:65.

89.   Hudak PL, McKeever P, Wright JG. The metaphor of patients as customers: implications for measuring satisfaction. Journal of Clinical Epidemiology 2003;56(2):103-8.

90.   Drummond M, O'Brien B, Stoddart G, Torrance G. Methods for the Economic Evaluation of Health Care Programmes. 2nd ed. Oxford: Oxford University Press; 1998.

91.   Cutler DM. The cost and financing of health care. American Economic Review 1995;85:32-37.

92.   Arrow K. Uncertainty and the welfare economics of medical care. American Economic Review 1963;53:941-973.

93.   Owens D, Batchelor C. Patient satisfaction and the elderly. Social Science and Medicine 1996;42(11):1483-1491.

94.   Draper M, Cohen P, Buchan H. Seeking consumer views: what use are results of hospital patient satisfaction surveys? International Journal for Quality in Health Care 2001;13(6):463-8.

95.   Linder-Pelz S. Social psychological determinants of patient satisfaction: a test of five hypotheses. Social Science and Medicine 1982;16:583-589.

96.   Goldstein MS, Elliott SD, Guccione AA. The development of an instrument to measure satisfaction with physical therapy. Physical Therapy 2000;80(9):853-63.

97.   Zimlichman E, Kreiss Y, Mandel D, Wartenfeld R, Ohana N, Cohen-Rosenberg G, et al. Evaluating delivery of primary health care to military personnel during low-intensity conflict using a patient satisfaction survey. Military Medicine 2003;168(6):471-4.

98.   Tjandra JJ, Fazio VW, Church JM, Milsom JW, Oakley JR, Lavery IC. Ripstein procedure is an effective treatment for rectal prolapse without constipation. Diseases of the Colon and Rectum 1993;36(5):501-7.

99.   Engel AF, van Baal SJ, Brummelkamp WH. Late results of postanal repair for idiopathic faecal incontinence. The European Journal of Surgery 1994;160(11):637-40.

100.   Engel AF, van Baal SJ, Brummelkamp WH. Late results of anterior sphincter application for traumatic faecal incontinence. The European Journal of Surgery 1994;160(11):633-6.

101.   Coyle J, Williams B. Seeing the wood for the trees: defining the forgotten concept of patient dissatisfaction in the light of patient satisfaction research. International Journal of Health Care Quality Assurance Incorporating Leadership in Health Services 1999;12(6-7):i-ix.

102.   Tucker JL, 3rd, Kelley VA. The influence of patient sociodemographic characteristics on patient satisfaction. Military Medicine 2000;165(1):72-6.

103.   Hall JA, Dornan MC. What patients like about their medical care and how often they are asked: a meta analysis of the satisfaction literature. Social Science and Medicine 1988;27(9):935-939.

104.   Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behaviour: An Introduction to Theory and Research. Reading, MA: Addison-Wesley; 1975.

105.   Vroom VH. Work and Motivation. New York: Wiley; 1964.

106.   Lawler EE. Pay and Organisational Effectiveness: A Psychological View. New York: McGraw-Hill; 1971.

107.   Merton RK. Continuities in the theory of reference groups. In: Merton RK, editor. Social Theory and Social Structure. New York: Free Press; 1968.

108.   Williams R. Relative deprivation. In: Coser LA, editor. The Idea of Social Structure: Papers in Honor of Robert K Merton. New York: Harcourt Brace Jovanovich; 1975. p. 547.

109.   Festinger L. A theory of social comparison processes. Human Relations 1954;7:114-140.

110.   Thibault JW, Kelley HH. The Social Psychology of Groups. New York: Wiley; 1959.

111.   Bramadat IJ, Driedger M. Satisfaction with childbirth: theories and methods of measurement. Birth 1993;20(1):22-9.

112. Mikkola K, Luukkonen P, Jarvinen HJ. Long-term results of restorative proctocolectomy for ulcerative colitis. International Journal of Colorectal Disease 1995;10(1):10-4.

113. Garcia-Aguilar J, Belmonte C, Wong WD, Lowry AC, Madoff RD. Open vs. closed sphincterotomy for chronic anal fissure: long term results. Diseases of the Colon and Rectum 1996;39:440-443.

114. Kuczyk MA, Klein S, Grunewald V, Machtens S, Denil J, Hofner K, et al. A questionnaire-based outcome analyis of the Stamey bladder neck suspension procedure for the treatment of urinary stress incontinence: the Hannover experience. British Journal of Urology 1998;82(2):174-80.

115. N'Dow J, Leung HY, Marshall C, Neal DE. Bowel dysfunction after bladder reconstruction. Journal of Urology 1998;159(5):1470-4; discussion 1474-5.

116. Everaert K, De Ridder D, Baert L, Oosterlinck W, Wyndaele JJ. Patient satisfaction and complications following sacral nerve stimulation for urinary retention, urge incontinence and perineal pain: a multicenter evaluation. International Urogynecology Journal and Pelvic Floor Dysfunction 2000;11(4):231-5; discussion 236.

117. Garcia-Aguilar J, Davey CS, Le CT, Lowry AC, Rothenberger DA. Patient satisfaction after surgical treatment for fistula-in-ano. Diseases of the Colon and Rectum 2000;43(9):1206-12.

118. Montague DK, Angermeier KW, Paolone DR. Long-term continence and patient satisfaction after artificial sphincter implantation for urinary incontinence after prostatectomy. Journal of Urology 2001;166(2):547-9.

119. Deval B, Jeffry L, Al Najjar F, Soriano D, Darai E. Determinants of patient dissatisfaction after a tension-free vaginal tape procedure for urinary incontinence. Journal of Urology 2002;167(5):2093-7.

120. Ruiz-Deya G, Davis R, Srivastav SK, A MW, Thomas R. Outpatient radical prostatectomy: impact of standard perineal approach on patient outcome. Journal of Urology 2001;166(2):581-6.

121. Williams S, Weinman J, Dale J. Doctor-patient communication and patient satisfaction: a review. Family Practice 1998;15(5):480-92.

122. Fitzpatrick R. Satisfaction with health care. In: Fitzpatrick R, editor. The Experience of Illness. London: Tavistock; 1984.

123. Linder-Pelz S, Struening EL. The multidimensionality of patient satisfaction with a clinic visit. Journal of Community Health 1985;10(1):42-54.

124. Ben-Sira Z. The function of the professional's affective behaviour in client satisfaction: a revised approach to social interaction theory. Journal of Health and Social Behavior 1976;17:3-11.

125. Donabedian A. The quality of care in a health maintenance organisation: a personal view. Inquiry 1983;20(3):218-22.

126. Krupat E, Fancey M, Cleary PD. Information and its impact on satisfaction among surgical patients. Social Science and Medicine 2000;51(12):1817-25.

127. Woodring S, Polomano RC, Haagen BF, Haack MM, Nunn RR, Miller GL, et al. Development and testing of patient satisfaction measure for inpatient psychiatry care. Journal of Nursing Care Quality 2004;19(2):137-48.

128. Gamagami RA, Chiotasso P, Lazorthes F. Continent perineal colostomy after abdominoperineal resection: outcome after 63 cases. Diseases of the Colon and Rectum 1999;42(5):626-30; discussion 630-1.

129. McConnell JC, Khubchandani IT. Long-term follow-up of closed hemorrhoidectomy. Diseases of the Colon and Rectum 1983;26(12):797-9.

130. Marcello PW, Roberts PL, Schoetz DJ, Jr., Coller JA, Murray JJ, Veidenheimer MC. Long-term results of the ileoanal pouch procedure. Archives of Surgery 1993;128(5):500-3; discussion 503-4.

131. Diokno AC, Burgio K, Fultz H, Kinchen KH, Obenchain R, Bump RC. Prevalence and outcomes of continence surgery in community dwelling women. Journal of Urology 2003;170(2 Pt 1):507-11.

132. Meredith P. Patient satisfaction with communication in general surgery: problems of measurement and improvement. Social Science and Medicine 1993;37(5):591-602.

133. Elkadry EA, Kenton KS, FitzGerald MP, Shott S, Brubaker L. Patient-selected goals: a new perspective on surgical outcome. American Journal of Obstetrics and Gynecology 2003;189(6):1551-7; discussion 1557-8.

134. Tannenbaum C, Bachand G, Dubeau CE, Kuchel GA. Experience of an incontinence clinic for older women: no apparent age limit for potential physical and psychological benefits. Journal of Women's Health & Gender-based Medicine 2001;10(8):751-6.

135. Laitinen P. Participation of informal caregivers in the hospital care of elderly patients and their evaluations of the care given: pilot study in three different hospitals. Journal of Advanced Nursing 1992;17(10):1233-7.

136. Rahmqvist M. Patient satisfaction in relation to age, health status and other background factors: a model for comparisons of care units. International Journal of Quality in Health Care 2001;13(5):385-90.

137. Hall J, Dornan M. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. Social Science and Medicine 1990;30(7):811-818.

138. Cummins R. On the trail of the gold standard for subjective wellbeing. Social Indicators Research 1995;35:179-200.

139. Bar-dayan Y, Leiba A, Weiss Y, Carroll JS, Benedek P. Waiting time is a major predictor of patient satisfaction in a primary military clinic. Military Medicine 2002;167(10):842-5.

140. Chesterman J, Bauld L, Judge K. Satisfaction with the care-managed support of older people: an empirical analysis. Health and Social Care in the Community 2001;9(1):31-42.

141. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Social Science and Medicine 1998;47(9):1351-1359.

142. Kravitz RL, Callahan EJ, Paterniti D, Antonius D, Dunham M, Lewis CE. Prevalence and sources of patients' unmet expectations for care. Annals of Internal Medicine 1996;125(9):730-737.

143. Jackson JL, Kroenke K. The effect of unmet expectations among adults presenting with physical symptoms. Annals of Internal Medicine 2001;134(9 Pt 2):889-97.

144. Lin CT, Albertson GA, Schilling LM, Cyran EM, Anderson SN, Ware L, et al. Is patients' perception of time spent with the physician a determinant of ambulatory patient satisfaction? Archives of Internal Medicine 2001;161(11):1437-42.

145. O'Connell B, Young J, Twigg D. Patient satisfaction with nursing care: a measurement conundrum. International Journal of Nursing Practice 1999;5(2):72-7.

146. Foddy W. Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research. Cambridge: Cambridge University Press; 1993.

147. Flocke SA. Measuring attributes of primary care: development of a new instrument. Journal of Family Practice 1997;45(1):64-74.

148. Lieberman JR, Dorey F, Shekelle P, Schumacher L, Thomas BJ, Kilgus DJ, et al. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. The Journal of Bone and Joint Surgery. American volume 1996;78(6):835-8.

149. Black N, Griffiths J, Pope C, Bowling A, P A. Impact of surgery for stress incontinence on morbidity: cohort study. British Medical Journal 1997;315:1493-1498.

150. McCammon KA, Kolm P, Main B, Schellhammer PF. Comparative quality-of-life analysis after radical prostatectomy or external beam radiation for localized prostate cancer. Urology 1999;54(3):509-16.

151. Wei JT, Montie JE. Comparison of patients' and physicians' rating of urinary incontinence following radical prostatectomy. Seminars in Urologic Oncology 2000;18(1):76-80.

152. Nguyen TD, Attkisson CC, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. Evaluation and Program Planning 1983;6(3-4):299-313.

153. Kinnersley P, Stott N, Peters TJ, Harvey I. The patient-centredness of consultations and outcome in primary care. British Journal of General Practice 1999;49(446):711-6.

154. Little P, Dorward M, Warner G, Moore M, Stephens K, Senior J, et al. Randomised controlled trial of effect of leaflets to empower patients in consultations in primary care. British Medical Journal 2004;328(7437):441.

155. Hyrkas K, Paunonen M. Patient satisfaction and research-related problems (Part 2). Is triangulation the answer? Journal of Nursing Management 2000;8(4):237-45.

156. Coyle J, Williams B. Valuing people as individuals: development of an instrument through a survey of person-centredness in secondary care. Journal of Advanced Nursing 2001;36(3):450-9.

157. Calnan M, Katsouyiannopoulos V, Ovcharov VK, Prokhorskas R, Ramic H, Williams S. Major determinants of consumer satisfaction with primary care in different health systems. Family Practice 1994;11(4):468-78.

158. Ramsay J, Campbell JL, Schroter S, Green J, Roland M. The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties. Family Practice 2000;17(5):372-9.

159. Bomalaski MD, Bloom DA, McGuire EJ, Panzl A. Glutaraldehyde cross-linked collagen in the treatment of urinary incontinence in children. Journal of Urology 1996;155(2):699-702.

160. Burgio KL, Locher JL, Goode PS, Hardin JM, McDowell BJ, Dombrowski M, et al. Behavioral vs drug treatment for urge urinary incontinence in older women: a randomized controlled trial. Journal of the American Medical Association 1998;280(23):1995-2000.

161. Demirci F, Petri E. Perioperative complications of Burch colposuspension. International Urogynecology Journal and Pelvic Floor Dysfunction 2000;11(3):170-5.

162. Williams KS, Assassa RP, Smith NK, Jagger C, Perry S, Shaw C, et al. Development, implementation and evaluation of a new nurse-led continence service: a pilot study. Journal of Clinical Nursing 2000;9(4):566-73.

163. Baker R, Whitfield M. Measuring patient satisfaction: a test of construct validity. Quality in Health Care 1992;1:104-109.

164. Shah DK, Paul EM, Amukele S, Eisenberg ER, Badlani GH. Broad based tension-free synthetic sling for stress urinary incontinence: 5-year outcome. Journal of Urology 2003;170(3):849-51.

165. Bower P, Roland MO. Bias in patient assessments of general practice: general practice assessment survey scores in surgery and postal responders. British Journal of General Practice 2003;53(487):126-8.

166. Hawthorne G. The effect of different methods of collecting data: mail, telephone and filter data collection issues in utility measurement. Quality of Life Research 2003;12:1081-1088.

167. Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. Social Science and Medicine 2001;52(4):609-20.

168. Linder-Pelz S, Stewart MM. Patient satisfaction with outpatient primary health care in a metropolitan medical center. American Journal of Preventive Medicine 1986;2(2):89-96.

169. Tucker JL. The moderators of patient satisfaction. Journal of Management in Medicine 2002;16(1):48-66.

170. Tucker JL, 3rd. The importance of caring as a determinant of patient satisfaction among active duty beneficiaries. Military Medicine 1998;163(11):758-60.

171. Feletti G, Firman D, Sanson-Fisher RW. Patient satisfaction with primary care consultations. Journal of Behavioural Medicine 1986;9:389-399.

172. Williams B, Wilkinson G. Patient satisfaction in mental health care. Evaluating an evaluative method. British Journal of Psychiatry 1995;166(5):559-62.

173. Hall JA, Milburn MA, Epstein AM. A causal model of health status and satisfaction with medical care. Medical Care 1993;31(1):84-94.

174. Fontana A, Rosenheck R. A model of patients' satisfaction with treatment for posttraumatic stress disorder. Administration and Policy in Mental Health 2001;28(6):475-89.

175. Hazard R, Haugh L, Green P, Jones P. Chronic low back pain: the relationship between patient satisfaction and pain, impairment, and disability outcomes. Spine 1994;19(8):881-887.

176. Jonler M, Messing EM, Rhodes PR, Bruskewitz RC. Sequelae of radical prostatectomy. British Journal of Urology 1994;74(3):352-8.

177. Litwiller SE, Kim KB, Fone PD, White RW, Stone AR. Post-prostatectomy incontinence and the artificial urinary sphincter: a long-term study of patient satisfaction and criteria for success. Journal of Urology 1996;156(6):1975-80.

178. Korman HJ, Sirls LT, Kirkemo AK. Success rate of modified Pereyra bladder neck suspension determined by outcomes analysis. Journal of Urology 1994;152(5 Pt 1):1453-7.

179. Maitra A, Chikhani C. Patient satisfaction in an urban accident and emergency department. British Journal of Clinical Practice 1992;46(3):182-4.

180. Young GJ, Meterko M, Kamal R. Patient satisfaction with hospital care: effects of demographic and institutional characteristics. Medical Care 2000;38(3):325-334.

181. Baker R. Characteristics of practices, general practitioners and patients related to levels of patients' satisfaction with consultations. British Journal of General Practice 1996;46(411):601-5.

182. Fitzpatrick R. Measurement of patient satisfaction. In: Hopkins D, Costain D, editors. Measuring the Outcomes of Medical Care. London: Royal College of Physicians and King's Fund Centre; 1990.

183. Coulter A. Evaluating the outcomes of health care. In: Gabe J, Calnan M, Bury M, editors. The Sociology of the Heatlh Service. London: Routledge; 1991. p. 115-139.

184. Miller YD, Brown WJ, Smith N, Chiarelli P. Managing urinary incontinence across the lifespan. International Journal of Behavioral Medicine 2003;10(2):143-61.

185. Moumtzoglou A, Dafogianni C, Karra V, Michailidou D, Lazarou P, Bartsocas C. Development and application of a questionnaire for assessing parent satisfaction with care. International Journal for Quality in Health Care 2000;12(4):331-7.

186. Ware JE, Jr., Hays RD. Methods for measuring patient satisfaction with specific medical encounters. Medical Care 1988;26(4):393-402.

187. Hendriks AA, Vrielink MR, Smets EM, van Es SQ, De Haes JC. Improving the assessment of (in)patients' satisfaction with hospital care. Medical Care 2001;39(3):270-83.

188. Meric HJ. The effect of scale form choice on psychometric properties of patient satisfaction measurement. Health Mark Quarterly 1994;11(3-4):27-39.

189. Pedhazur E, Schmelkin L. Measurement, Design and Analysis: An Integrated Approach. Hillsdale: Lawrence Erlbaum; 1991.

190. Anastasi A. Evolving concepts of test validation. Annual Review of Psychology 1986;37:1-15.

191. Nunally J. Psychometric Theory. 2nd ed. New York: McGraw Hill; 1978.

192. Lennon R. Assumptions underlying the use of content validity. Educational & Psychological Measurement 1965;16:294-304.

193. Streiner D, Norman G. Health Measurement Scales: A Practical Guide to their Development and Use. 2nd ed. Oxford: Oxford Medical Publications; 1995.

194. Peters JL, Thompson AC, McNicholas TA, Hines JE, Hanbury DC, Boustead GB. Increased patient satisfaction from transrectal ultrasonography and biopsy under sedation. BJU International 2001;87(9):827-30.

195. Maher CF, Murray CJ, Carey MP, Dwyer PL, Ugoni AM. Iliococcygeus or sacrospinous fixation for vaginal vault prolapse. Obstetrics and Gynecology 2001;98(1):40-4.

196. Taylor BM, Beart RW, Jr., Dozois RR, Kelly KA, Phillips SF. Straight ileoanal anastomosis v ileal pouch – anal anastomosis after colectomy and mucosal proctectomy. Archives of Surgery 1983;118(6):696-701.

197. Church JM, Fazio VW, Lavery IC, Oakley JR, Milsom J, McGannon E. Quality of life after prophylactic colectomy and ileorectal anastomosis in patients with familial adenomatous polyposis. Diseases of the Colon and Rectum 1996;39(12):1404-8.

198. Pinedo G, Vaizey CJ, Nicholls RJ, Roach R, Halligan S, Kamm MA. Results of repeat anal sphincter repair. British Journal of Surgery 1999;86(1):66-9.

199.   Choe JM, Ogan K, Battino BS. Antimicrobial mesh versus vaginal wall sling: a comparative outcomes analysis. Journal of Urology 2000;163(6):1829-34.

200.   Grumbach K, Selby JV, Damberg C, Bindman AB, Quesenberry C, Jr., Truman A, et al. Resolving the gatekeeper conundrum: what patients value in primary care and referrals to specialists. Journal of the American Medical Association 1999;282(3):261-6.

201.   Haab F, Trockman BA, Zimmern PE, Leach GE. Quality of life and continence assessment of the artificial urinary sphincter in men with minimum 3.5 years of followup. Journal of Urology 1997;158(2):435-9.

202.   Czaplicki M, Dobronski P, Torz C, Borkowski A. Long-term subjective results of Marshall-Marchetti-Krantz procedure. European Urology 1998;34(2):118-23.

203.   Badalament RA, Bahn DK, Kim H, Kumar A, Bahn JM, Lee F. Patient-reported complications after cryoablation therapy for prostate cancer. Urology 1999;54(2):295-300.

204.   Brown SL, Govier FE. Cadaveric versus autologous fascia lata for the pubovaginal sling: surgical outcome and patient satisfaction. Journal of Urology 2000;164(5):1633-7.

205.   Winters JC, Chiverton A, Scarpero HM, Prats LJ, Jr. Collagen injection therapy in elderly women: long-term results and patient satisfaction. Urology 2000;55(6):856-61.

206.   Walsh IK, Nambirajan T, Donellan SM, Mahendra V, Stone AR. Cadaveric fascia lata pubovaginal slings: early results on safety, efficacy and patient satisfaction. BJU International 2002;90(4):415-9.

207.   Lim AJ, Brandon AH, Fiedler J, Brickman AL, Boyer CI, Raub WA, Jr., et al. Quality of life: radical prostatectomy versus radiation therapy for prostate cancer. Journal of Urology 1995;154(4):1420-5.

208.   Latini JM, Brown JA, Kreder KJ. Abdominal sacral colpopexy using autologous fascia lata. Journal of Urology 2004;171(3):1176-9.

209.   Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Evaluation and Program Planning 1982;5(3):233-7.

210.   Pang MW, Chan LW, Yip SK. One-year urodynamic outcome and quality of life in patients with concomitant tension-free vaginal tape during pelvic floor reconstruction surgery for genitourinary prolapse and urodynamic stress incontinence. International Urogynecology Journal and Pelvic Floor Dysfunction 2003;14(4):256-60; discussion 259-60.

211.   Hilton BA, Budgen C, Molzahn AE, Attridge CB. Developing and testing instruments to measure client outcomes at the Comox Valley Nursing Center. Public Health Nursing 2001;18(5):327-39.

212.   Roberts RE, Pascoe GC, Attkisson CC. Relationship of service satisfaction to life satisfaction and perceived well-being. Evaluation and Program Planning 1983;6(3-4):373-83.

213.   Kinnersley P, Stott N, Peters T, Harvey I, Hackett P. A comparison of methods for measuring patient satisfaction with consultations in primary care. Family Practice 1996;13(1):41-51.

214.   McKinley RK, Fraser RC, Baker RH, Riley RD. The relationship between measures of patient satisfaction and enablement and professional assessments of consultation competence. Medical Teacher 2004;26(3):223-8.

215.   Munro BH, Jacobsen BS, Brooten DA. Re-examination of the psychometric characteristics of the La Monica-Oberst Patient Satisfaction Scale. Research in Nursing and Health 1994;17(2):119-25.

216.   Vahey DC, Aiken LH, Sloane DM, Clarke SP, Vargas D. Nurse burnout and patient satisfaction. Medical Care 2004;42(2 Suppl):II57-66.

217.   Hays RD, Davies AR, Ware JE. Scoring the Medical Outcomes Study Patient Satisfaction Questionnaire PSQ-III. Boston: Medical Outcomes Trust; 1987.

218.   Helmstadter GC. Principles of Pyschological Measurement. New York: Appleton-Century-Crofts; 1964.

219.   Kikano GE, Goodwin MA, Stange KC. Physician employment status and practice patterns. Journal of Family Practice 1998;46(6):499-505.

220. Cornish A, Fynes M, Harmer C, Hawthorne G, Rosamillia A, Carey M, et al. The Genitourinary Treatment Satisfaction Score for continence surgery. In: Neurourology and Urodynamics; 2001; 2001. p. 506-507.

221. Karantanis E, Fynes MM, Stanton SL. The tension-free vaginal tape in older women. British Journal of Obstetrics and Gynaecology 2004;111(8):837-41.

222. Tocchi A, Mazzoni G, Miccini M, Cassini D, Bettelli E, Brozzetti S. Total lateral sphincterotomy for anal fissure. International Journal of Colorectal Disease 2004;19(3):245-9.

223. Wyrwich KW, Metz SM, Babu AN, Kroenke K, Tierney WM, Wolinsky FD. The reliability of retrospective change assessments. Quality of Life Research 2002;11(7):636.

224. Burgio KL, Goode PS, Locher JL, Umlauf MG, Roth DL, Richter HE, et al. Behavioral training with and without biofeedback in the treatment of urge incontinence in older women: a randomized controlled trial. Journal of the American Medical Association 2002;288(18):2293-9.

225. Cortina J. What is coefficient alpha? Examination of theory and applications. Journal of Applied Psychology 1993;78(1):98-104.

226. AIHW. Australia's Health 2004: The Ninth Biennial Health Report of the Australian Institute of Health and Welfare. Canberra: Australian Institute of Health and Welfare; 2004. Report No.: AIHW Cat AUS44.

227. Birrell R, Hawthorne L. Doctor dilemmas: how GP medicine is practised in Australia. People and Place 2001;9(3):41-53.

228. Hawthorne L, Birrell B, Young D. The Retention of Overseas Trained Doctors in General Practice in Regional Victoria. Melbourne: Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne; 2003.