Measuring Patient Satisfaction with Incontinence Treatment

Centre for Health Service Development

September, 2006
# Table of Contents

**ACKNOWLEDGEMENTS** ................................................................................................................................. 1

**GLOSSARY OF TERMS** .................................................................................................................................. 2

**LIST OF INSTRUMENT ABBREVIATIONS** ..................................................................................................... 5

1 EXECUTIVE SUMMARY .......................................................................................................................... 6

2 INTRODUCTION ....................................................................................................................................... 8

3 LITERATURE REVIEW ............................................................................................................................ 9

3.1 The role of and theories of patient satisfaction ......................................................................................... 9

3.2 Patient satisfaction instruments ............................................................................................................. 10

4 METHOD ......................................................................................................................................................... 12

4.1 Sample size .............................................................................................................................................. 12

4.2 Participants ............................................................................................................................................ 12

4.3 The study questionnaire ......................................................................................................................... 12

4.4 Procedures ........................................................................................................................................... 15

4.5 Statistical analysis ................................................................................................................................. 15

4.6 Assumptions behind the modelling ..................................................................................................... 16

5 DESCRIPTION OF THE SAMPLE CHARACTERISTICS AND OVERALL PATIENT SATISFACTION REPORTS ....................................................................................................................................................... 20

5.1 Response rate and participant details ..................................................................................................... 20

5.2 Incontinence status and treatment details ............................................................................................. 22

5.3 Patient satisfaction ............................................................................................................................... 24

6 PSYCHOMETRIC EVALUATION OF THE PATIENT SATISFACTION MEASURES .......................................................... 28

6.1 Revision of the original GUTSS ........................................................................................................... 28

6.2 Descriptive systems and item sensitivities ............................................................................................ 32

6.3 Scale properties .................................................................................................................................... 37

7 DERIVING A NEW SHORT PATIENT SATISFACTION SCALE ............................................................................. 42

7.1 Some issues arising from the analysis of instruments in Section 6 .......................................................... 42

7.1.1 A theoretical model of patient satisfaction and its measurement: issues in coverage .................. 42

7.1.2 Item response scales, data distribution and item stems .................................................................. 43

7.2 Constructing a short assessment of patient satisfaction scale ............................................................... 46

7.2.1 Data analysis procedure .................................................................................................................. 46

7.2.2 Findings .......................................................................................................................................... 47

7.2.3 Scoring the SAPS ........................................................................................................................... 48

7.3 A single measure for immediate assessment of patient satisfaction ................................................... 52
List of Tables

Table 1  Demographic details of participants ................................................................. 21
Table 2  Urinary incontinence status, treatment, pre- and post-treatment effect .................. 23
Table 3  Expectations, information and self-assessed change pre-post treatment .................. 24
Table 4  Patient satisfaction scores, percentage scores ................................................... 25
Table 5  Grouping of putative predictors of patient satisfaction ...................................... 26
Table 6  Predicting patient satisfaction ......................................................................... 27
Table 7  Modifications to the GUTSS items to remove the filter questions ......................... 28
Table 8  Mokken analysis of the GUTSS and revised GUTSS ........................................... 29
Table 9  Tests of convergent and divergent validity for the original and revised GUTSS .......... 32
Table 10  Patient satisfaction measures descriptive systems .......................................... 34
Table 11  Initial item and scale analysis ......................................................................... 36
Table 12  Scale properties ............................................................................................. 38
Table 13  Tests of convergent and divergent validity for the four incontinence measures, T-scores 41
Table 14  Patient satisfaction measures descriptive systems .......................................... 42
Table 15  Item response distributions (not recoded) ........................................................ 44
Table 16  Final model of a unidimensional Short Assessment of Patient Satisfaction scale (SAPS) 47
Table 17  Psychometric properties of SAPS items and the SAPS scale .............................. 48
Table 18  Patient satisfaction scores, percentage scores ................................................... 49
Table 19  Tests of convergent and divergent validity for the SAPS, T-scores ....................... 51
Table 20  Test of responsiveness to patient satisfaction, five measures, T-scores .................. 52
Table 21  Item response theory analysis of selected global items ..................................... 53
Table 22  Responsiveness of the single patient satisfaction item, PS#B ............................. 55
Table 23  Comparative sensitivity of patient satisfaction instruments, relative efficiency estimates 59
Table 24  GUTTS, SAPS and patient satisfaction items and their original source ............... 61

List of Figures

Figure 1  Monotone homogeneity and double monotonicity among a set of three dichotomous items 18
Figure 2  Distribution of the original and revised GUTSS T-scores by 0.5 standard deviations ........ 30
Figure 3  Scatterplot of original and revised GUTSS scores, T-scores .................................. 31
Figure 4  Standardized standard deviations by item number for the four instruments ............ 39
Figure 5  Score distribution for the Short Assessment of Patient Satisfaction scale .................. 49
Figure 6  Category probability curves for a single global patient satisfaction item: PS#B How satisfied are you with the outcome of your treatment? ................................. 54
Figure 7  Disordered category probability curves for a single global patient satisfaction item: PS#D How satisfied are you with the amount of help received? ................................. 54
Acknowledgements

We would like to thank the participating clinicians who contributed to this study. Our particular thanks go to Dr Marcus Carey and Ms Margaret Sherburn at the Royal Women’s Hospital, Melbourne, and to A/Professor Kate Moore at St George Hospital, Kogarah in NSW. Without their assistance this study could never have taken place. We would also like to thank all the patients who gave their time to complete the study questionnaire.

Our thanks also go to the various instrument developers who provided permission for us to use their measures in this study. We offer our apologies for any omissions.

This study was funded by the Commonwealth Department of Health and Ageing. Ethics approval was given by the Ethics Committees of the University of Wollongong, the Royal Women’s Hospital and St George Hospital.

There are no competing interests for the investigators.
Glossary of Terms

ANOVA  Analysis of variance.

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.

Cohen's effect size (d)  This quantifies the size of differences between groups or over time (Cohen, 1988). Cohen classified effect sizes into small effects (<0.20), moderate (~0.50) and large effect sizes (>0.80).

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 α  Measure of the reliability of a scale, based on examining the internal consistency of responses to items forming the scale (Cronbach, 1951). Cronbach alpha is based on both the correlations between items and the number of items within an instrument. However, where data distributions are highly skewed, α will represent the lower boundary of reliability rather than an accurate estimate.

Descriptive system  Refers to the actual items of an instrument and how these items are organized within an instrument.

DIF  Differential item functioning. DIF describes the extent to which two or more groups of respondents interpret an item differently (i.e. whether the item has significantly different meaning for the different groups).

Double-counting  Describes where the same issue is counting twice or more within an instrument. If there are redundant items in an instrument, then adding up their scores will produce double-counting.

End aversion  Describes where respondents avoid selecting an extreme option, e.g. a person may wish to avoid stating that they are ‘extremely dissatisfied’, so they will state that they are ‘dissatisfied’.

Guttman scale  Describes a response scale where the responses progressively increase (e.g. none, some, a lot, many).

Guttman scalogram  Describes an instrument comprising Guttman scales, where respondents order their responses such that a<b<c<d etc (Guttman, 1944).

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 alpha.
**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.

**Item response theory**

Referred to as IRT and modern test theory, this postulates that a person’s probability of selecting a particular response to an item is conditional upon their ability to select the correct response for him/her, and that abilities and probabilities can be separately described.

**IRTC**

Item rest of test correlation. Describes how well an item fits a scale based on correlations with the other items in the scale.

**Kappa (κ)**

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 indicates 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.

**Mokken analysis**

A form of item response theory analysis which assesses the unidimensionality of a scale based on the axioms of Guttman scalogram measurement.

**Mokken rho (ρ)**

Internal consistency reliability estimate for use with Guttman scales where the data are highly skewed.

**NCMS**

National Continence Management Strategy

**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 (Cronbach and Meehl, 1955).

**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.

**Relative efficiency (RE)**

Describes how responsive a scale is when compared with another scale (Fayers and Machin, 2000, Wright and Young, 1997).

**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 $\alpha$, or the correlation between half of a scale compared with the other half, administered at the same time (split-half reliability). It is difficult to be precise about the desired reliability levels. For example, longer scales will have higher reliability than shorter scales. The conventions are that for comparison of groups reliability 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 (Nunnally and Bernstein, 1994).
<table>
<thead>
<tr>
<th><strong>Response scale</strong></th>
<th>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 <em>Do you leak urine?</em> and the response scale might be <em>Not at all, a little, some, a moderate amount, a lot.</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responsiveness</strong></td>
<td>Describes the sensitivity of a scale to differences in the underlying condition.</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>Standard deviation.</td>
</tr>
<tr>
<td><strong>Then-test</strong></td>
<td>Describes the difference between a participant’s current health state and his/her previous health state, where the previous health state is assessed by asking the participant to recall his/her previous health state.</td>
</tr>
<tr>
<td><strong>T-score</strong></td>
<td>Standardized scores where the mean score = 50 and the standard deviation = 10 (McCall, 1922).</td>
</tr>
<tr>
<td><strong>Validity</strong></td>
<td>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.</td>
</tr>
</tbody>
</table>
List of Instrument Abbreviations

AQoL  Assessment of Quality of Life (Hawthorne et al., 1999; Hawthorne et al., 2001).

Consult SQ  Consultation Satisfaction Questionnaire (Baker, 1990). This is usually referred to as the CSQ, but in this report it is described as the Consult SQ to differentiate it from the CSQ-18.

CSQ-18  Client Satisfaction Questionnaire (Larsen et al., 1979).

FS  Friendship Scale (Hawthorne, 2006b).

GUTSS  Genito-Urinary Treatment Satisfaction Scale (Hawthorne and Harmer, 2000). Two versions of the GUTSS were administered to participants: the original version which consists of 10-items including 2 filter questions, and a revised version of 8 items without the filter questions. The two versions are directly compared in section 6 where it is shown that there is no substantive difference between the two versions. All references to the GUTSS throughout this report refer to the revised GUTSS, unless otherwise stated.

ISI  Incontinence Severity Index – 4 level version (Sandvik et al., 1993; Sandvik et al., 2000).

PSI  Patient Satisfaction Index (Guyatt et al., 1995).

SAPS  Short Assessment of Patient Satisfaction scale. Refer to Section 7 of this report.

UCA  Urinary Continence Assessment scale (Sansoni et al., In Press).

UDI-6  Urogenital Distress Inventory – short form (Shumaker et al., 1994; Uebersax et al., 1995).

WHO5  The World Health Organization’s wellbeing instrument, which can be used for depression screening (Bonsignore et al., 2001; Bech, 2004).
1 Executive Summary

As part of the National Continence Management Strategy, Hawthorne (2006c) reviewed patient satisfaction measures and recommended (a) investigation into a single patient satisfaction item for use by clinicians wishing to make an “on-the-spot” assessment, (b) that a short incontinence specific measure be developed based on revisions to the GUTSS instrument, and (c) that a short generic instrument be developed. This study implements these recommendations through the evaluation of various patient satisfaction instruments to determine their suitability for inclusion in the NCMS suite of measures.

A sample of women (N=178) who had treatment (physiotherapy or surgery) for urinary incontinence 6-12 months before recruitment participated in this study. Recruited through clinics in Melbourne and Sydney, they completed a patient satisfaction questionnaire and returned these to the researchers. This procedure ensured anonymity: the researchers had no knowledge of which women were sampled from the clinics and the clinics had no knowledge of which women returned the questionnaires. The response rate was 44% of those to whom questionnaires were sent. The questionnaire asked about type of incontinence, pre-treatment incontinence status, treatment, post-treatment incontinence status, expectations of treatment and four standard patient satisfaction questionnaires, the Client Satisfaction Questionnaire (CSQ-18) (Larsen et al., 1979), the Consultation Satisfaction Questionnaire (Consult SQ) (Baker, 1990), the Genito-Urinary Treatment Satisfaction Scale (GUTSS) (Hawthorne and Harmer, 2000) and the Patient Satisfaction Index (PSI) (Guyatt et al., 1995). The questionnaire also contained measures of social isolation, wellbeing and quality of life.

A model of patient satisfaction based on the work of Donabedian (1988) was used. This postulates that satisfaction is defined as the patient’s judgement on the quality of care, particularly the interpersonal relationships with clinicians and other care providers. The seven dimensions contributing to this model that were identified from the literature by Hawthorne (2006c) were used as the conceptual framework against which the patient satisfaction measures were reviewed.

Because of the nature of the data (most people are satisfied with their health care, thus the data were non-normally distributed), modern test psychometric methods were used; specifically Mokken analysis which investigates scale homogeneity and partial credit item response theory (IRT) which investigates the measurement properties of items.

Regarding revision of the GUTSS it was found that removal of two filter items and rewording of two other items shortened the GUTSS and simplified it’s scoring without any loss of validity or reliability. The revised GUTSS was used thereafter.

Selection of responsive items and their examination through iterative Mokken and partial credit IRT analyses against the model of patient satisfaction above, led to the selection of 7 items which formed a generic short assessment of patient satisfaction (the SAPS) scale. The internal psychometric properties of the SAPS were excellent.

The five resulting patient satisfaction instruments (the four original instruments and the SAPS) were examined by their descriptive systems, internal structures and responsiveness. Regarding the descriptive systems, the findings showed that the Consult SQ was mainly measuring the technical skill and relationship with the treating clinician, the CSQ-18 whether the treatment received was appropriate to the needs of the patient, the GUTSS measured satisfaction with treatment outcomes, the PSI the patient as a consumer of health services and the SAPS provided a broad patient satisfaction perspective.

The internal structures of the instruments suggested that all the items were responsive on the GUTSS and the SAPS, but that some items on the other measures were insensitive. All measures were shown to be unidimensional as assessed by the Mokken analysis and the reliability of all the
instruments was excellent. Tests of response bias suggested that this occurred in two of the instruments (the CSQ-18 and the PSI). In the PSI this may have been related to the overly difficult item stems and the use of 7-point response scales which, for many items, have convoluted descriptors. Redundancy was observed in the Consult SQ, CSQ-18 and PSI, typically because these instruments had several items measuring the same concept.

Responsiveness was assessed against various predictors of patient satisfaction and against a pooled patient satisfaction estimate derived from the four original patient satisfaction instruments. The results showed that although all five instruments were responsive, relative efficiency tests showed that the most responsive instrument was the GUTSS, then the SAPS, CSQ-18, and then the PSI. The Consult SQ was the least responsive instrument.

Six global items were extracted from the various instruments and examined using partial credit IRT. The results showed disordered thresholds or differential item functioning (i.e. where different groups of respondents interpret the item differently) for five of the items. The remaining item (PS#B) exhibited excellent psychometric properties, including responsiveness to known predictors.

From this study finding it is recommended that:

1. Where a single item is needed (e.g. in everyday clinical practice) the single item patient satisfaction (PS#B) is used (refer Appendix 3).

2. The revised GUTSS should be used where a specific incontinence patient satisfaction instrument is needed, such as in clinical trials.

3. For use in epidemiological or clinical studies that the newly developed SAPS is used, subject to further field testing as outlined in Recommendation #4.

4. Because the items contributing to the SAPS were edited to simplify the item stems, give them a common response scale and to ensure that the SAPS had a mixture of positive and negative items, that the SAPS be field tested prior to being recommended as the preferred patient satisfaction instrument.
2 Introduction

Incontinence is a common health problem that affects over 2 million Australians of all ages and backgrounds (Hawthorne, 2006a). 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.

For those suffering incontinence, treatment usually consists of either physiotherapy or, for more serious conditions, surgery. The primary outcome from treatment is, of course, relief from the symptoms of incontinence. Another outcome, however, is satisfaction with the health care process. As part of the NCMS, a report on 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 (Thomas et al., 2006). 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.

The first report on patient satisfaction arising from Thomas et al. (2006) was Hawthorne's review of patient satisfaction measures (Hawthorne, 2006c). This review covered the literature on patient satisfaction theories, the interpretation of patient satisfaction scores, the relationship between measurement theory and patient satisfaction instruments, and reviews of particular patient satisfaction instruments. The review found most instruments were not validated for use in incontinence treatment, the Australian healthcare system or its patients. It recommended that a study be undertaken to provide a validated instrument for use in the NCMS suite of outcomes measures (Thomas et al., 2006). The specific recommendations were that (a) a single patient satisfaction item is investigated for use by clinicians wishing to make an “on-the-spot” assessment, (b) a short incontinence specific measure be developed based on revisions to the GUTSS instrument, and (c) a short generic instrument be developed.

This study implements these recommendations through the evaluation of four patient satisfaction instruments to determine their suitability for inclusion in the NCMS suite of measures. It provides an overview of patient satisfaction within the context of urinary incontinence. Based on an analysis of data collected from female patients at 6 to 12 months after treatment for urinary incontinence, it makes recommendations relating to the inclusion of patient satisfaction instruments to be used as part of the NCMS.
3 Literature Review

A detailed review of the literature on patient satisfaction can be found in Hawthorne (2006c). This section summarizes the aspects of that work relevant to this study.

3.1 The role of and theories of patient satisfaction

The measurement of 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 rather than their role as health services consumers.

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 (Donabedian, 1988).

The major patient satisfaction theories were all published during the 1980s; almost all research since then is based on these theories. Ware et al (1983) argued that patient satisfaction was a function of patients' subjective responses to experienced care mediated by personal preferences and expectations. Linder-Pelz (1982) postulated it was mediated by personal beliefs and values about care as well as prior expectations of the care. Fox and Storms (1981) advocated that a person's orientation determined satisfaction; dissatisfaction, therefore, occurred where there was transgression of the relationship between expectation and experience. Fitzpatrick and Hopkins (1983) 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 (1988, 1980) 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:

.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. (Donabedian, 1988, p. 1746)

The implication is that the construct of patient satisfaction covers all aspects of care quality, particularly the interpersonal processes. Patient dissatisfaction will occur where there are a cluster of small transgressions of these dimensions or a major failure in service provision. Hawthorne (2006c) argued that the key dimensions of this construct were:

1. Appropriate access to health services, including the environment within which treatment takes place and the level of care coordination (Ware et al., 1983; Fox and Storms, 1981; Hardy et al., 1996);

2. The provision of health information (Suchman, 1965; Fox and Storms, 1981; Hardy et al., 1996; Hawthorne and Harmer, 2000; Donabedian, 1980);

3. The relationship between the patient and health care staff, specifically empathy with the patient (Ware et al., 1983; Hardy et al., 1996; Ben-Sira, 1976; Sitzia and Wood, 1997; Fitzpatrick, 1990; Donabedian, 1988; Kane et al., 1997);
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 (Hardy et al., 1996);

5. Satisfaction with the treatment provided, i.e. the technical quality of the care provided (Ware et al., 1983; Linder-Pelz and Struening, 1985; Fox and Storms, 1981; Hardy et al., 1996; Hawthorne and Harmer, 2000; Kane et al., 1997); and

6. The effectiveness of treatment, including the extent to which treatment helps the patient in their daily life (Ware et al., 1983; Hardy et al., 1996; Hawthorne and Harmer, 2000; Donabedian, 1988).

This model of patient satisfaction postulates that in a comprehensive assessment of patient satisfaction all seven dimensions will contribute and should be measured.

3.2 Patient satisfaction instruments

Hawthorne (2006c) comprehensively reviewed patient satisfaction and its measurement. 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. One of the most consistent findings was 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 was 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 suspect.

Although Hawthorne reviewed nine patient satisfaction instruments, he reported that there was no ‘stand out’ instrument and none could be considered truly satisfactory based on that no instrument had been sufficiently validated for its use in Australia to be automatically recommended. There were three key reasons for this finding:

A. There was 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 was 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.

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.

Four instruments, however, were recommended for further investigation. They were the:

- The CSQ-18 (Client Satisfaction Questionnaire) developed by Larsen et al. (1979). This instrument comprises 18 items measuring satisfaction with services.

- The Consult SQ (Consultation Satisfaction Questionnaire) which assesses satisfaction with a consultation with a general practitioner (Baker, 1990). This also consists of 18 items.
The GUTSS (Genito-Urinary Treatment Satisfaction Scale) which was designed for use in incontinence trials (Hawthorne and Harmer, 2000). It consists of 10 items assessing care and outcome satisfaction.

The PSI (Patient Satisfaction Inventory) which was designed to discriminate between patients with a life-threatening illness (Guyatt et al., 1995). It has 23 items. Detailed reviews of these measures can be found in Hawthorne (2006c). Descriptions of these are given in section 4.3 of this report.

This study reports on these four instruments through a direct head-to-head comparison in a sample of female patients between 6 to 12 months post incontinence treatment. It also reports on changes to the GUTSS to improve its scoring, the development of a short assessment of patient satisfaction scale and on the properties of several single global satisfaction items.
4 Method

This study investigated patient satisfaction in a sample of Australian female patients who had received treatment for their urinary incontinence. The sample was drawn from major public hospitals in Australia’s two largest cities, Melbourne and Sydney. In the interests of generalizability the sample was stratified by type of treatment (physiotherapy and surgery).

4.1 Sample size

The sample size was calculated on two parameters. First, it was assumed that there would be differences in patient satisfaction by treatment cohort (physiotherapy/surgery). It was hypothesized that Cohen’s d effect size (Cohen, 1988) between the two cohorts would be 0.25, i.e. a small effect. Translated to an estimate of sample size for the comparison of group means using a standard sample size formula for the difference in means (N = (2(Zα + Z1-β))/d2) (Davison et al., 1986), this indicated at least 85 cases in each study cohort – a total of 170 cases. Second, this initial estimate was increased to the requirements of sample sizes for stable factor analysis (N=~150 overall) (Guadagnoli and Velicer, 1988) and item response theory analysis (N=~125 in each of the study cohorts (physiotherapy/surgery), i.e. a total of 250 participants (Holman et al., 2003).

Based on the literature, a participation rate of 60% was expected (Hill et al., 1997; Day et al., 1995; Hawthorne et al., 2006a; Brogger et al., 2002). The sample size was therefore increased to over-sample by 170 cases (250/0.6 = 420), i.e. the initial sample sizes were estimated to be 210 in each of the study cohorts (physiotherapy/surgery).

4.2 Participants

A post-treatment cross-sectional survey of patients who received urinary incontinence treatment was conducted. Although it had been planned to include both urinary and faecal incontinence patients in both males and females, due to the timing constraints on the study only patients who had received urinary incontinence treatment were identified from the clinical records.

Eligibility criteria were having received treatment three to twelve months prior to study participation and having sufficient English to complete a self-report questionnaire. Patient lists at participating clinics were scanned by participating clinicians and a random sample of eligible cases drawn (N = 420). The participating clinics were the St George Hospital, Sydney, and the Royal Women’s Hospital, Melbourne, and its associated private practices in surgery and physiotherapy.

The initial physiotherapy patient numbers were 30 from St George Hospital, 81 from the Royal Women’s Hospital and 99 from the private clinics. The initial patient numbers who had undergone surgery were 40 from St George Hospital, 49 from the Royal Women’s Hospital and 121 from the private clinics.

4.3 The study questionnaire

The questionnaire consisted of four parts. The first, Part A, included items concerning the nature and severity of the urinary incontinence and its treatment. Part B elicited demographic information. The largest section, Part C, included four patient satisfaction measures and Part D reviewed participants’ quality of life.
Part A  Participants’ treatment

This was a short questionnaire about treatment, including questions on type of incontinence, date of treatment start and finish, expectations of treatment, main treatment modality and self-report success/failure.

The incontinence severity measures recommended by Thomas et al. (2006) were included. Two versions of each instrument were included, once for before treatment severity (retrospective) and again for after treatment severity (current). The difference between the retrospective and current incontinence estimates is described as the ‘then-test’. The instruments were:

- The Incontinence Severity Index – 4 level version (ISI) (Sandvik et al., 2000, Sandvik et al., 1993). This comprises 2 items measuring how often a person experiences urine leakage and the amount of urine lost. The possible scoring range is 1 to 12, with a high score indicating more severe incontinence symptoms.

- The Urogenital Distress Inventory – short form (UDI-6) (Shumaker et al., 1994; Uebersax et al., 1995) comprises 6 items measuring the experience of and being bothered by frequent urination, leakage related to urgency, leakage related to physical activity, coughing or sneezing, the amount of urine leaked, difficulty with emptying the bladder and pain in the lower abdominal or genital area. The possible scoring range is 0 to 18, with a high score indicating more severe incontinence symptoms.

- From these two measures the Urinary Continence Assessment (UCA) scale score was computed. This new composite measure of 3 items was developed by Hawthorne from the ISI and UDI and is reported in Sansoni et al. (In Press). In this study it is used as a composite measure of incontinence. The score range is 0 to 9, with a higher score indicating greater incontinence.

Part B  Participants’ backgrounds

Typical demographic items were included such as gender, age, country of birth, relationship status, education level and work status.

Part C  Satisfaction with health care received for incontinence

This section was the patient satisfaction item bank, with the four scales recommended in the Hawthorne report (2006c). For detailed reviews of the instruments, the reader should consult that report. The scales were:

- The Client Satisfaction Questionnaire (CSQ-18; n=18 items) was developed to measure satisfaction with health care services (Larsen et al., 1979). It 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 Guttman-type scale, where the responses cover a poor service through to an excellent service. Scoring of the CSQ-18 is by simple summation. For the CSQ-18 the score range is 18-72. In this study the scoring range has been adjusted to 0 to 54, with a high score indication high satisfaction with treatment.

- The Consultation Satisfaction Questionnaire (Consult SQ) (Baker, 1990) 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). The items are Likert scales with 5 response levels. The possible scoring range is 0 to 76, with a high score indicating high satisfaction with treatment.

- The Genito-Urinary Treatment Satisfaction Scale (GUTSS) was developed to measure women’s satisfaction with the outcome of treatment for urinary stress incontinence (Hawthorne and Harmer, 2000). The original GUTSS comprises 10 items, 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). The overall GUTSS score is obtained by summing the two sub-scale scores. The score range is from 0 to 32, with a high score indicating high satisfaction with treatment.

Hawthorne (2006c) recommended removal of the two filter questions. Two additional items were included with the GUTSS to test the effect of removal of the filter questions. The analysis reported in section 6.1 showed that the two versions of the GUTSS could be regarded as equivalent scales; consequently throughout this report the revised GUTSS has been used and is referred to as the GUTSS. Where the original GUTSS is referred to it is described as being the ‘original GUTSS’.

- The Patient Satisfaction Index (PSI) was designed to discriminate between patients with a life-threatening illness who were satisfied with their medical care and those who were not (Guyatt et al., 1995). 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, adjusted in this study to 0 to 138, with a high score indicating high satisfaction with treatment.

To minimise the differences between the four instruments, item stems were modified (where this was indicated permissible by the instrument developers) to reflect incontinence (e.g. the PSI item stems had the phrase ‘the treatment you received for your incontinence’ inserted), and a similar format was used for all items so that they had a similar look for the respondents.

**Part D Questions about other aspects of the participant’s life (well being and quality of life).**

This section consisted of three instruments covering screening for depression, social isolation and quality of life.

- Screening for depression was assessed by the World Health Organization’s WHO5 measure (Heun et al., 1999; Bonsignore et al., 2001; Henkel et al., 2003; Primack, 2003). This comprises 5 items assessing being in good spirits, being calm, feeling active, refreshed and experiencing interest in life. The response scales are 6-point Guttman scales. Scoring is through summation; the range is 0 to 25, where 25 represents the highest wellbeing. For depression screening, scores <13 suggest probable depression as does endorsement of any item response in the range 0/1 (i.e. where a respondent endorses the worst possible levels). Using these criteria, scores can be dichotomized into no signs of depression and probable depression.
Social isolation was measured with the Friendship Scale (Hawthorne, 2006b). This 6-item instrument assesses relationships with others, being isolated, sharing, getting in touch, feelings of being separate and being alone. Response scales are 5-point Guttman-type assessing item stem frequency. The range of scores is 0-24 with lower scores indicating greater social isolation.

Quality of life was assessed with the Assessment of Quality of Life (AQOL) instrument (Hawthorne et al., 1999; Hawthorne and Osborne, 2005). This is a multi-attribute utility instrument comprising 15 items in 5 scales measuring Illness, Independent Living, Social Relationships, Physical Senses and Psychological Wellbeing. An overall utility index is computed from the last four dimensions using a weighted multiplicative model. The score range is from -0.04 (HRQoL worse than death) through 0.00 (HRQoL equivalent to death) to 1.00 (full HRQoL). Norms are available for Australian populations (Hawthorne and Osborne, 2005).

4.4 Procedures

After the sample had been drawn, four hundred and twenty questionnaires were mailed to putative participants on incontinence treatment lists at St George Hospital, Sydney, and the Royal Women’s Hospital, Melbourne, and its associated private physiotherapy clinics. The questionnaire package included an introductory letter from researchers, a letter of explanation from the clinicians at the treatment centre, a participant information form, the questionnaire and a consent form. A reply paid addressed envelope was also included.

The researchers prepared the questionnaire packages which were then forwarded to participating clinics. The clinics inserted the personalised letter of explanation and then mailed the questionnaire on behalf of the researchers. This procedure ensured that no personal or identifying information was revealed to the researchers and that the participating clinics were unaware of which patients participated. Questionnaires were returned via mail directly to researchers using the reply paid envelope.

A reminder letter was posted at 2-week follow-up.

4.5 Statistical analysis

The analysis strategy

Study participants were stratified by incontinence type and severity, treatment success or failure, and relevant socio-economic variables (e.g. age group). These known groups were then used as stratifying variables for testing scales and items.

Four analyses of the patient satisfaction instruments and items were carried out:
- The first analysis examined the GUTSS to assess the effect of removing the two filter questions.
- The second analysis was of the item and scale properties of each of the four measures, using a mixture of classic test theory (CTT, e.g. item data distributions, tests of convergent and divergent validity by known groups) and IRT (Mokken analysis) to assess the unidimensionality of scales.
- The third analysis was to assess the possibility of whether it was possible to build a short, user-friendly and valid measure of patient satisfaction. The steps were
  - A preliminary analysis to explore the properties of items on psychometric and logical criteria (such as ceiling or floor effects, and failure to discriminate). Items failing these basic tests were removed;
Mokken analysis and IRT was used to identify the properties of remaining items and to build a parsimonious model of patient satisfaction;

Tests of validity were then conducted (e.g. by known groups, such as treatment success/failure, gender)

The fourth analysis was to select items which had the potential to be single-item measures of patient satisfaction with a view to recommending an item for everyday use.

Statistical procedures used

Following data entry, the raw data was assessed for missing data, outliers and other irregularities. Missing data were imputed using horizontal imputation for scale items (Hawthorne and Elliott, 2005). No attempt was made to impute categorical missing data (e.g. gender).

Scale reliabilities (internal consistencies) are reported using Cronbach alpha. All the measures used, however, contained non-normally distributed item-level data. Although the use of alpha under these circumstances violates its axioms, studies which have examined alpha under these circumstances have reported that it is a robust measure (Bandalos and Enders, 1996; Enders and Bandalos, 1999; Zimmerman et al., 1993; Bollen and Lennox, 1991). In addition to alpha, for double-monotone scales Mokken’s ρ (rho) is reported because under nonparametric item response theory scale models, alpha underestimates reliability (Sijtsma and Molenaar, 1987).

Categorical data are reported as frequencies with percentages and were analysed with $\chi^2$. Where the distributional requirement of the $\chi^2$ test was violated, the Fisher Exact Test was used instead. Effect sizes were calculated with Cohen’s $d$ with scores of $d=0.20$ indicating a small effect, $d=0.50$ indicating a moderate effect and $d=0.80$ indicating a large effect (Cohen, 1988). Pearson and Spearman correlations were performed between scales, determined by tests of normality. Principal component and oblimin factor analyses were used to group variables or examine latent constructs.

Regression was used to examine the predictors of patient satisfaction, and analysis of variance (ANOVA) to examine differences between known groups. For direct comparison between patient satisfaction scales, raw scores were computed as percentage scores and T-scores (mean = 50, standard deviation = 10) (McCall, 1922). T-tests were used to examine differences between groups or differences over time, and Welch’s approximate t was used to compare between groups where the variances were dissimilar. Skew was assessed and non-normal data transformed using reciprocal and log transformations prior to analysis. For comparing the relative responsiveness of scales the relative efficiency (RE) statistic was used (Wright and Young, 1997; Fayers and Machin, 2000), based on F-values, thus: $RE = \frac{F_i}{F_D}$, where $F_i$ was the instrument of interest and $F_D$ the instrument with the smallest F-value.

Scale homogeneity was examined with Mokken analysis (Mokken, 1982), which overcomes the restrictions associated with parametric methods such as factor analysis. For a detailed description see section 4.5. Partial credit item response theory (IRT) for polytomous items was used to examine the properties of specific items. Percentages were rounded up to the nearest integer. Statistical analyses were performed using SPSS Version 14.0. (Salahuddin et al., 2005), MSP5 Version 5.0 (Molenaar et al., 2000), Instat Version 3.06 (GraphPad, 2003) and RUMM2020 Version 4.0 (Andrich et al., 2004).

Assumptions behind the modelling

The data analyses used in this report are based on several measurement axioms derived from representational, classical and modern test theories (Michell, 1986; Streiner and Norman, 2006; Pedhazur and Schmelkin, 1991; Nunnally and Bernstein, 1994).
It was assumed that patient satisfaction is a latent construct (i.e. a virtual concept) which can be represented by manifest measurement models (i.e. an instrument) comprising items (i.e. formal questions) representative of all possible items. All the items in a model form a scale, measure or instrument which provides for graded measurement (Gutman, 1944). Graded measurement is where the response scales of the individual items possess values that are ordinal in nature, i.e. an assigned value of ‘2’ is bigger than ‘1’ (Stevens, 1951). Respondents to a questionnaire have personal abilities (knowledge, experiences or traits) that determine how they respond to items. Modern test theory postulates that these abilities are independent of a person’s actual item response on a particular test (e.g. think of an experienced driver who misjudges parking a car and scratches the paintwork on a wall) (Rasch, 1960). A further assumption of test theory is that there is a symbolic interaction between the purpose of the researcher (who is using the instrument for some purpose) and a respondent’s responses (Foddy, 1993). Finally, it was assumed that the participants in this study were representative of incontinence patients in general. The implication is that the findings are relevant to other incontinence patients.

From these axioms it follows that the construct ‘patient satisfaction’ is unidimensional in the sense that there are statistically observable relationships between both different patient satisfaction measures and their various items. Where items form a unidimensional scale, the scale is described as being an ‘homogenous’ scale. For example, this was claimed for at least two of the patient satisfaction measures reported in this study. Hawthorne & Harmer (2000) reported that all GUTSS items fell on a single principal component, while Guyatt et al. (1995) argued that the PSI was unidimensional based on the Cronbach alpha of 0.94.

Homogeneity implies that the items will order respondents similarly. For example, think of a 2-item instrument, where one item asks “Are you satisfied with your care overall?” and the other asks “Are you satisfied with the pre-treatment care provided in the clinic?” Respondents who are satisfied overall with their care will endorse the first item, but only those respondents who considered they had excellent care in the clinic before their treatment will endorse both the first and second items. It should be noted that a respondent’s endorsement of each item is assumed to be independent of his/her response to the other item; this condition is known as ‘local independence’. Thus the probabilities (P(θ)) of selecting responses to items are functions of the underlying ability of respondents (e.g. from no incontinence symptoms to extreme incontinence). Where these probabilities increase monotonically within an item response scale, the item is described as having ‘monotone homogeneity’. This is a desirable property of an item since it minimises endorsement errors. Where all items within an instrument meet this requirement, the instrument is described as possessing a set of ‘monotonely homogenous items’, or rather more elaborately a ‘monotone locally independent unidimensional model’ (Stout, 2002).

‘Doubly monotone’ describes a further subset of items; viz., those items where the P(θ) curves do not intersect. This property is illustrated in Figure 1. The x-axis line (0-θ) represents different ability levels of respondents, from those with little ability (θ0) to those with high ability (θ3), and the P(θ) y-axis represents the probability of endorsing a dichotomous response (in this case No/Yes). As shown, those with low ability have a low probability of endorsing a ‘Yes’ response (<0.50). The figure shows that for each of the three items the item characteristic curves (ICCs) suggest that the probabilities of endorsing a response level are monotone – hence all three items meet the requirements for monotone homogeneity. Because the ICCs for each of the items do not cross each other, this implies that there is a graded order of probability of response between the items, hence these three items meet the requirements for being considered a set of doubly monotone items: there is agreement both on the endorsement order within items and on the endorsement order between items. This is a highly desirable property of a set of instrument items; where this property does not hold the items may be repetitious of each other leading to redundancy and double-counting.
Figure 1  Monotone homogeneity and double monotonicity among a set of three dichotomous items

Source: Adapted from Mokken (1982, p. 419).

Tests in which items meet the axiom of monotone homogeneity are considered good, and tests which are doubly monotone are considered excellent. It should be noted, however, that very few tests will meet these requirements perfectly. Where there are few violations of these axioms, an instrument will be considered robust, whereas where there are a high number of violations the validity of the instrument is in doubt (the greater the number of violations the greater the randomness of endorsement errors). A poor instrument is one where neither of these properties can be demonstrated.

Under the assumptions outlined at the beginning of this section, valid patient satisfaction measures will be those meeting the axioms outlined above. They imply that a nonparametric item response theory and scale analysis investigating monotonicity is appropriate. Mokken analysis meets these requirements because it employs the axioms of item response theory (IRT) to examine the unidimensionality of instruments (including the fit of nonparametric items) and the extent to which monotone homogeneity and double monotonicity are met (Mokken, 1982).

Item monotone homogeneity is assessed through computation of Loewinger’s $H_i$ coefficient of homogeneity (Loevinger, 1948). Simply, the coefficient of probability is the probability of obtaining a score (or endorsement level) on an item where the score on another item is already known, thus:

$$H_{ij} = \frac{p_{i|j} - p_i}{1 - p_i}$$

Where $H_{ij}$ is the coefficient of homogeneity for the two items being compared, $p$ is the probability of a given score, $i$ is the item of interest and $j$ is another item for which the endorsement level is already known, where $i$ and $j$ are Guttman scalogram ordered items (Guttman, 1944), i.e. $i \leq j$. $H_{ij}$ is the weighted average of the probabilities for the item of interest. It is adjusted so that a value of 0.00 indicates perfect heterogeneity (i.e. complete disagreement between the items) and 1.00 indicates perfect homogeneity (i.e. perfect agreement between the items). For inclusion in a homogenous scale, $H_i$ should exceed 0.30 if the item is to be acceptable, and >0.40 for an item to be definitely included in the scale.

Scale homogeneity is assessed through computation of Loewinger’s $H$ coefficient of homogeneity of a test (Loevinger, 1948). This is the weighted average of the $H_i$ coefficients. The score range is from 0.00 (where no item pairs within an instrument are monotonic) through to 1.00 (where all item

1 Guttman scalograms assume that items $a<b<c<d<e........n$. A perfect scalogram thus describes a set of monotonic items.
pairs are perfectly monotonic). The accepted values for H are <0.4 which indicates a weak scale, 0.40 to 0.50 a medium scale and >0.50 a strong scale (Mokken, 1982).

Item monotonic violations in the computer program MSP5 are described as the Crit-value; this is the weighted number of violations (weighted by the number of tests made, etc.) 2. Molenaar et al. (2000) proffers the advice that where the Crit-value <40 the violations may be due to sampling variation, and that values >80 imply that the item does not fit the homogenous scale and should be deleted.

---

2 Where items in a Guttman scalogram are non-monotonic (e.g. a<b<c>d<e), this is described as a violation of monotonicity.
5 Description of the Sample Characteristics and Overall Patient Satisfaction Reports

5.1 Response rate and participant details

One hundred and eighty-four completed questionnaires were returned to the researchers, giving a simple response rate of 44% of those patients initially drawn. Because of the methodology the number of patients out of scope is unknown: no questionnaires were returned to the researchers because of unknown addresses. Although lower than expected, this response rate is consistent with the literature for postal surveys in health studies where there are no follow-up telephone calls eliciting participation (Fowler et al., 2002). (Follow-up telephone calls could not be made since the researchers had no knowledge of patient’s names, addresses or other contact details.)

The percentage of females in the sample was 97%; there were just 6 males in the study population; therefore these males were excluded from all the data analyses leaving 178 cases for data analysis. This gender bias reflects the type of clinics that participated in the research: these were female-focused services, such as the Royal Women’s Hospital and women’s health physiotherapy clinics.

Eighty-one percent of participants were born in Australia, 7% born in Europe, 3% in Asia and the Pacific region, 7% in Africa or the Middle East, and 2% from North America. For analysis purposes, participants were coded into Australian- and Overseas-born cases. The details are given in Table 1. The mean age of participants was 58 years (SD = 12 years). There was, however, a good distribution by age as shown by the 15-year age cohorts in Table 1. Regarding education attainment, 9% had finished primary school only, 43% high school, 21% held a trade or TAFE certificate and 27% had a college or university degree.

Seventy-eight percent of the sample were partnered (married or in a de facto relationship), 12% were widowed, and 10% had never married, were separated or divorced. Finally, 48% were participating in the workforce or were seeking work, 25% were homemakers and 28% were either retired or in receipt of sickness benefits. The details are given in Table 1.

Regarding participants’ current mental health status, the WHO5 standard classification scores indicated that 24% (N = 42) were probably suffering depression. For social isolation, 84% were socially connected, 8% reported some social isolation and 8% were socially isolated based on their Friendship Scale scores. In terms of health related quality of life, the mean score on the AQoL was 0.78 (SD = 0.19, N = 173), which was only just below the published population norm (0.83, SD = 0.20) (Hawthorne and Osborne, 2005).

---

3 Although the researchers approached clinics for inclusion of males, no clinics were able to help within the study period.
### Table 1 Demographic details of participants

<table>
<thead>
<tr>
<th>Demographic details of participants</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St George Hospital (Sydney)</td>
<td>31</td>
<td>17%</td>
</tr>
<tr>
<td>Royal Women’s Hospital (Melbourne)</td>
<td>117</td>
<td>66%</td>
</tr>
<tr>
<td>Private clinic (Melbourne)</td>
<td>30</td>
<td>17%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>178</td>
<td>97%</td>
</tr>
<tr>
<td>Male (a)</td>
<td>6</td>
<td>3%</td>
</tr>
<tr>
<td><strong>Country of birth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>142</td>
<td>80%</td>
</tr>
<tr>
<td>Other</td>
<td>36</td>
<td>20%</td>
</tr>
<tr>
<td><strong>Age group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-45 years</td>
<td>31</td>
<td>17%</td>
</tr>
<tr>
<td>46-60 years</td>
<td>73</td>
<td>41%</td>
</tr>
<tr>
<td>61-75 years</td>
<td>57</td>
<td>32%</td>
</tr>
<tr>
<td>75+ years</td>
<td>17</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary only</td>
<td>16</td>
<td>9%</td>
</tr>
<tr>
<td>High school</td>
<td>76</td>
<td>43%</td>
</tr>
<tr>
<td>Trade certificate/TAFE</td>
<td>37</td>
<td>21%</td>
</tr>
<tr>
<td>College/University</td>
<td>48</td>
<td>27%</td>
</tr>
<tr>
<td><strong>Relationship status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partnered</td>
<td>138</td>
<td>78%</td>
</tr>
<tr>
<td>Never married/Divorced/Separated</td>
<td>18</td>
<td>10%</td>
</tr>
<tr>
<td>Widowed</td>
<td>22</td>
<td>12%</td>
</tr>
<tr>
<td><strong>Labour force status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour force (i.e. working, student, unemployed)</td>
<td>85</td>
<td>48%</td>
</tr>
<tr>
<td>Homemaker</td>
<td>44</td>
<td>25%</td>
</tr>
<tr>
<td>Retired/Sickness benefit</td>
<td>49</td>
<td>28%</td>
</tr>
</tbody>
</table>

**Notes:**
All percentages are rounded up to the nearest integer, so percentages may not round up to 100%.
Excludes all missing cases. The base number was 178 cases (excluding males).
a = Males were excluded from all the data analyses.
5.2 Incontinence status and treatment details

Regarding incontinence status for which participants reported they had sought treatment, although the clinical records showed that all cases had sought treatment for urinary incontinence, on the self-report questionnaire 82% reported urinary incontinence, 11% did not report their incontinence type, 6% reported suffering both urinary and faecal incontinence and 2% reported suffering faecal incontinence only (the most likely source of this discrepancy is that these cases may have suffered both urinary and faecal symptoms and sought treatment for their urinary incontinence which left them with some faecal incontinence symptoms). The details are given in Table 2. The table also shows the type of treatment undergone for incontinence and the proportions receiving each treatment type: 27% had participated in physiotherapy alone, 40% had experienced surgery alone and 33% had received both physiotherapy and surgery. Retrospective (i.e. before treatment) and current urinary incontinence self-report status are given in Table 2 for the ISI, UDI and UCA, along with then-test change scores indicating treatment effectiveness. The data suggest that between 78% to 88% of participants improved following treatment; that for 7-15% there was no change in their condition and that 5-7% of participants experienced deterioration in their condition following treatment.
### Table 2  Urinary incontinence status, treatment, pre- and post-treatment effect

<table>
<thead>
<tr>
<th>Urinary incontinence status, treatment, pre- and post-treatment effect</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report incontinence status (a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urinary only</td>
<td>145</td>
<td>82%</td>
</tr>
<tr>
<td>Faecal only</td>
<td>4</td>
<td>2%</td>
</tr>
<tr>
<td>Both</td>
<td>10</td>
<td>6%</td>
</tr>
<tr>
<td>Not stated</td>
<td>19</td>
<td>11%</td>
</tr>
<tr>
<td>Incontinence treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiotherapy</td>
<td>47</td>
<td>27%</td>
</tr>
<tr>
<td>Physiotherapy and surgery</td>
<td>57</td>
<td>33%</td>
</tr>
<tr>
<td>Surgery</td>
<td>69</td>
<td>40%</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>2%</td>
</tr>
<tr>
<td>ISI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre (retrospective) (Mean, SD)</td>
<td>13.49</td>
<td>(5.75)</td>
</tr>
<tr>
<td>Post (current) (Mean, SD)</td>
<td>6.05</td>
<td>(4.97)</td>
</tr>
<tr>
<td>Then-test mean change score (Mean, SD)</td>
<td>7.47</td>
<td>(6.57)</td>
</tr>
<tr>
<td>Percentage of cases reporting improvement</td>
<td>135</td>
<td>78%</td>
</tr>
<tr>
<td>Percentage of cases with no change</td>
<td>26</td>
<td>15%</td>
</tr>
<tr>
<td>Percentage of cases with worse outcomes</td>
<td>13</td>
<td>7%</td>
</tr>
<tr>
<td>UDI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre (retrospective) (Mean, SD)</td>
<td>9.20</td>
<td>(3.72)</td>
</tr>
<tr>
<td>Post (current) (Mean, SD)</td>
<td>3.72</td>
<td>(3.17)</td>
</tr>
<tr>
<td>Then-test mean change score (Mean, SD)</td>
<td>5.25</td>
<td>(4.28)</td>
</tr>
<tr>
<td>Percentage of cases reporting improvement</td>
<td>147</td>
<td>88%</td>
</tr>
<tr>
<td>Percentage of cases with no change</td>
<td>12</td>
<td>7%</td>
</tr>
<tr>
<td>Percentage of cases with worse outcomes</td>
<td>9</td>
<td>5%</td>
</tr>
<tr>
<td>UCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre (retrospective) (Mean, SD)</td>
<td>5.86</td>
<td>(2.28)</td>
</tr>
<tr>
<td>Post (current) (Mean, SD)</td>
<td>2.78</td>
<td>(2.44)</td>
</tr>
<tr>
<td>Then-test mean change score (Mean, SD)</td>
<td>3.08</td>
<td>(2.68)</td>
</tr>
<tr>
<td>Percentage of cases reporting improvement</td>
<td>139</td>
<td>82%</td>
</tr>
<tr>
<td>Percentage of cases with no change</td>
<td>20</td>
<td>12%</td>
</tr>
<tr>
<td>Percentage of cases with worse outcomes</td>
<td>10</td>
<td>6%</td>
</tr>
</tbody>
</table>

**Notes**

a = All cases clinical records indicated urinary incontinence.

Statistical analysis of each of the three incontinence scales showed that the mean change scores in Table 2 were significant with very little difference between the measures (dependent t-test, \( t_{\text{ISI}} = 14.95, \text{df} = 172, p < 0.01; t_{\text{UDI}} = 15.88, \text{df} = 167, p < 0.01; \) and \( t_{\text{UCA}} = 14.93, \text{df} = 168, p < 0.01 \).

In addition to these ‘objective’ indicators of treatment success or failure, participants were asked their pre-treatment retrospective expectations, their satisfaction with pre-treatment information,
and their subjective rating of the success of their treatment. These data are presented in Table 3 and show that 69% of participants rated the pre-treatment information as being excellent or very good. The data also show that 53% of participants expected their incontinence would be cured compared with 36% who reported that their incontinence was cured. As shown, these figures suggest that the outcomes were as expected for 45% of participants and that for 39% of cases the outcomes were worse than their original expectations. After excluding the 3 “Other” cases whose status was unknown, there was no significant difference in these outcomes by treatment modality ($\chi^2 = 3.74, df = 4, p = 0.44$).

### Table 3 Expectations, information and self-assessed change pre-post treatment

<table>
<thead>
<tr>
<th>Expected outcomes</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incontinence cured</td>
<td>91</td>
<td>53%</td>
</tr>
<tr>
<td>Partly cured</td>
<td>52</td>
<td>30%</td>
</tr>
<tr>
<td>Some improvement</td>
<td>27</td>
<td>16%</td>
</tr>
<tr>
<td>That treatment would not help at all</td>
<td>2</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pre-treatment information</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>57</td>
<td>33%</td>
</tr>
<tr>
<td>Very good</td>
<td>63</td>
<td>36%</td>
</tr>
<tr>
<td>Good</td>
<td>36</td>
<td>21%</td>
</tr>
<tr>
<td>Fair</td>
<td>16</td>
<td>9%</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
<td>1%</td>
</tr>
<tr>
<td>Very poor</td>
<td>1</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceived treatment success</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incontinence cured</td>
<td>62</td>
<td>36%</td>
</tr>
<tr>
<td>Partly cured</td>
<td>57</td>
<td>33%</td>
</tr>
<tr>
<td>Some improvement</td>
<td>42</td>
<td>24%</td>
</tr>
<tr>
<td>Treatment did not help at all</td>
<td>13</td>
<td>8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did treatment meet expectations? (a)</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes better than expected</td>
<td>28</td>
<td>17%</td>
</tr>
<tr>
<td>Outcomes as expected</td>
<td>76</td>
<td>45%</td>
</tr>
<tr>
<td>Outcomes worse than expected</td>
<td>66</td>
<td>39%</td>
</tr>
</tbody>
</table>

**Notes**

a = Composite measure derived from expected outcomes – perceived outcomes.

### 5.3 Patient satisfaction

Each of the patient satisfaction measures was scored using the developer's instructions. For comparison, the raw scores were converted to percentages and T-scores. The percentage data are presented in Table 4. They show that the original GUTTS, CSQ-18 and PSI all performed similarly in assessing satisfaction based on mean scores and the percentage of cases at scale ceilings. The Consult SQ assessed patient satisfaction quite differently. The mean score was ~10% lower than that on the other measures and the percentage of cases with ceiling scores was 3%. (There were no floor effects for any of the instruments.) It is noteworthy that the Consult SQ was the only patient satisfaction instrument in the study using Likert scales for all items.
Tables 5 and 6 show the predictors of patient satisfaction based on a three-step process. Step 1 was to ensure all variables were at least ordinal. Where predictors were non-ordinal they were recoded to form ordinal scales. For example, treatment was coded in order of presumed severity into 3 levels, thus ‘physiotherapy’, ‘physiotherapy & surgery’ and ‘surgery’ (other treatments’ (3 cases) were excluded from the analysis). Similarly, variables were logically recoded so that all response categories contained >10 cases. For example, expected treatment outcomes were recoded into 3 levels ‘cured’, ‘partly cured’, ‘some improvement/no help’. Step 2 was to conduct an oblimin factor analysis on the available predictors in order to identify co-varying variables which, where used in a regression model, could confound the model through co linearity. The oblimin procedure allows the factors to be correlated – which was the case in this study. Step 3 was to construct a multivariate regression model, entering into the model the pivotal (i.e. highest loading) and lowest loading items from each of the factors (excluding cross-loading variables) up to a maximum of 8 predictors (to meet stable model requirements based on the available sample). The purpose in selecting pivotal and lowest loading items was to try, as far as possible, to avoid co linearity. Some co linearity is, of course, acknowledged. The selected items, therefore, should be regarded as being representative of all items on the factor. Four models were constructed with the dependent variable in each model one of the four patient satisfaction measures.

Table 5 shows the grouping of variables based on an unconstrained oblimin model. Four factors were identified where the eigen values were >1.00. The model explained 68% of the variance. The first factor comprised the different measures of current incontinence (ISI, UDI & UCA) along with perceived treatment success; it could be described as current status. The second factor was pre-treatment incontinence status (ISI, UDI, UCA), the change scores (pre-now) on these measures and treatment (physiotherapy, physiotherapy & surgery, surgery); based on the item content this factor could be labelled incontinence severity improvement. Factor 3 comprised age group, labour force status, education attainment, and relationship status; this can be thought of as demographic background. Factor 4 was pre-treatment expected outcomes, whether treatment met expectations and pre-treatment information provided; clearly this was an expectations factor.
### Table 5  Grouping of putative predictors of patient satisfaction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factors &amp; loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>UCA - now</td>
<td>-.95</td>
</tr>
<tr>
<td>ISI - now</td>
<td>-.89</td>
</tr>
<tr>
<td>UDI - now</td>
<td>-.88</td>
</tr>
<tr>
<td>How successful was your treatment?</td>
<td>-.85</td>
</tr>
<tr>
<td>UCA – pre-treatment</td>
<td>.94</td>
</tr>
<tr>
<td>UDI – pre-treatment</td>
<td>.89</td>
</tr>
<tr>
<td>ISI – pre-treatment</td>
<td>.88</td>
</tr>
<tr>
<td>UCA change scores</td>
<td>.53</td>
</tr>
<tr>
<td>UDI change score, pre - post</td>
<td>.48</td>
</tr>
<tr>
<td>ISI change score, pre - post</td>
<td>.58</td>
</tr>
<tr>
<td>Treatment received (excluding Other) (a)</td>
<td>.47</td>
</tr>
<tr>
<td>Age group (b)</td>
<td>.83</td>
</tr>
<tr>
<td>Labour force participation (c)</td>
<td>.80</td>
</tr>
<tr>
<td>Education attainment (d)</td>
<td>-.68</td>
</tr>
<tr>
<td>Relationship status (e)</td>
<td>.49</td>
</tr>
<tr>
<td>Before treatment expected outcomes (f)</td>
<td>.84</td>
</tr>
<tr>
<td>Were expectations met? (g)</td>
<td>-.79</td>
</tr>
<tr>
<td>Information given (h)</td>
<td>-.31</td>
</tr>
</tbody>
</table>

**Notes:**

- Unconstrained oblimin factor analysis
- N. eigenvalues >1.00 = 4
- % explained variance = 68% overall.
  - 30%  19%  11%  8%
- Cross-loadings <0.30 not shown
- Item not loading on any factor: Country of birth (Australia/Other)
- a = physiotherapy/physiotherapy & surgery/surgery;
- b = 31-45/46-60/61-75/75+ years;
- c = in labour force/homemaker/retired or on sickness benefits
- d = primary only/high school/trade certificate/college or university degree
- e = partnered/never partnered, divorced or separated/widowed
- f = cured/partly cured/improved & not helped
- g = better outcome than expected/expectations met/worse outcome than expected
- h = excellent/very good/good/fair

The regression results are shown in Table 6. Across all four patient satisfaction measures, the strongest predictors were the representative predictors from factor 1 followed by those from factor 4 (Table 5). The two strongest predictors explaining the highest proportion of variance were the level of information given by the clinician and the self-assessed treatment effect – these were only two significant predictors across all four measures. The other significant predictors were expected...
outcomes from treatment for the original GUTSS and CSQ-18, and current incontinence status (represented by the UCA) on the original GUTSS. Of interest is that none of the demographic predictors were significant on any of the four patient satisfaction measures. In summary the data suggested that patient satisfaction is a function of information given, perceived treatment effect, expectations and current incontinence status. These predictors were used as criteria for validation tests of patient satisfaction measures.

Table 6  Predicting patient satisfaction

<table>
<thead>
<tr>
<th>Factor (a)</th>
<th>Representative predictors</th>
<th>Consult</th>
<th>CSQ-18</th>
<th>GUTSS (original)</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>β (b)</td>
<td>p</td>
<td>β (b)</td>
<td>p</td>
</tr>
<tr>
<td>F1</td>
<td>UCA – now</td>
<td>-0.03</td>
<td>0.76</td>
<td>-0.09</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>How successful was treatment</td>
<td>-0.26</td>
<td>0.02</td>
<td>-0.42</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>F2</td>
<td>UCA – pre-treatment</td>
<td>0.01</td>
<td>0.96</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Treatment received (c)</td>
<td>-0.13</td>
<td>0.09</td>
<td>-0.05</td>
<td>0.50</td>
</tr>
<tr>
<td>F3</td>
<td>Age group</td>
<td>0.02</td>
<td>0.81</td>
<td>-0.09</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Relationship status</td>
<td>-0.05</td>
<td>0.50</td>
<td>0.01</td>
<td>0.86</td>
</tr>
<tr>
<td>F4</td>
<td>Expected outcomes</td>
<td>0.13</td>
<td>0.09</td>
<td>0.26</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Information given</td>
<td>-0.45</td>
<td>&lt;0.01</td>
<td>-0.47</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Explanatory power (R²ADJ)</td>
<td>0.26</td>
<td>0.46</td>
<td>0.67</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>8,149</td>
<td>8,148</td>
<td>8,150</td>
<td>8,148</td>
</tr>
<tr>
<td></td>
<td>F-value (omnibus test)</td>
<td>7.95</td>
<td>16.97</td>
<td>41.93</td>
<td>9.02</td>
</tr>
<tr>
<td></td>
<td>p-value (omnibus test)</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Notes:

a = From Table 5
b = Standardized coefficient
b = Excludes “Other” (3 cases)
6 Psychometric Evaluation of the Patient Satisfaction Measures

Although Section 5 reports both the levels and the predictors of patient satisfaction for the four study measures, it is apparent from Tables 4 and 6 that they have very different properties. This section examines the internal psychometric properties of each of the scales.

Hawthorne (2006c), which reviewed patient satisfaction scales, recommended that the original GUTSS be revised to remove the two filter questions. This is described in Section 6.1 and the revised GUTSS is used throughout the rest of this section.

6.1 Revision of the original GUTSS

Hawthorne (2006c) recommended that the original GUTSS be revised to remove the two filter questions. To achieve this, two different versions of the GUTSS questions involving the filter items were included in the survey: the original GUTSS items and modified items without the two filter questions. The details are given in Table 7.

Table 7 Modifications to the GUTSS items to remove the filter questions

<table>
<thead>
<tr>
<th>Modifications to the GUTSS items to remove the filter questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 7</strong> Modifications to the GUTSS items to remove the filter questions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>#2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>#3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>#5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>#6</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Importantly, the GUTSS was placed at the beginning of the patient satisfaction questionnaire, whereas the two modified non-filtered items were placed towards the end. This may have affected the results. To investigate this, after recoding, the relationship between the two sets of items was examined. The kappa for item #3 was 0.54 and it was 0.51 for #6, indicating perfect agreement for 85% of cases between the two item #3s and 74% for the item #6s. These results would suggest that comparison of different versions of GUTSS is subject to potential bias in approximately 20% of cases.

In the original GUTSS, for items #3 and #5 where a person did not endorse the previous filter question (thus skipping these two questions), the assigned values were ‘6’ on the assumption that the respondent experienced no incontinence whatsoever. With the revised GUTSS items, the best possible response level was a ‘5’ for these two items. This change, then, materially affected the overall GUTSS score by reducing the maximum possible score from 34 to 32. To ensure equivalence, for all tests between the different versions of the GUTSS, scores were converted to T-scores.

### Table 8  Mokken analysis of the GUTSS and revised GUTSS

<table>
<thead>
<tr>
<th>Item stem (from modified GUTSS)</th>
<th>Original GUTSS</th>
<th>Revised GUTSS</th>
<th>Revised GUTSS if #6 removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_i$ (a)</td>
<td>$H_i$</td>
<td>$H_i$</td>
</tr>
<tr>
<td></td>
<td>Crit-value</td>
<td>Crit-value</td>
<td>Crit-value</td>
</tr>
<tr>
<td>1 How happy are you with effect of treatment?</td>
<td>0.65</td>
<td>0.65</td>
<td>0.68</td>
</tr>
<tr>
<td>2 Over the past 4 weeks, do you still have problems with incontinence?</td>
<td>0.62</td>
<td>0.60</td>
<td>0.64</td>
</tr>
<tr>
<td>3 How satisfied are you with the outcomes from your treatment?</td>
<td>0.66</td>
<td>0.65</td>
<td>0.69</td>
</tr>
<tr>
<td>4 During the past 4 weeks have you been disappointed with the outcome of your treatment?</td>
<td>0.49</td>
<td>0.60</td>
<td>0.64</td>
</tr>
<tr>
<td>5 Before you had the treatment, was the information from your clinician.....</td>
<td>0.44</td>
<td>0.46</td>
<td>0.44</td>
</tr>
<tr>
<td>6 Was the attitude/behaviour of the clinicians.....</td>
<td>0.48</td>
<td>0.50</td>
<td>0.56</td>
</tr>
<tr>
<td>7 How satisfied are you with the explanations your clinician has given you about the results of your treatment?</td>
<td>0.54</td>
<td>0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>8 Are you happy with the care you received in the hospital or clinic?</td>
<td>0.50</td>
<td>0.53</td>
<td>0.49</td>
</tr>
<tr>
<td>Loevinger’s coefficient of homogeneity (H)</td>
<td>0.55</td>
<td>0.58</td>
<td>0.60</td>
</tr>
<tr>
<td>Scale reliability (Mokken $\rho$)</td>
<td>0.89</td>
<td>0.90</td>
<td>0.90</td>
</tr>
</tbody>
</table>

**Notes:**

- $a = H_i$ Item coefficient of scalability
- $b = $Crit-value under P-matrix analysis
- $* =$ worst performing item, which exceeds Crit-value for item inclusion
A Mokken scale analysis was undertaken comparing the original and revised versions of the GUTSS, as shown in Table 8. Both versions reported Loevinger’s H >0.50 (indicating homogeneity of the scale) and excellent reliability. Although all items H-values were above the conventional cutpoint for item scalability (i.e. the cutpoint at which an item would be deemed not part of the scale, <0.30) and both scales were above Mokken’s suggested Loevinger’s coefficient of homogeneity H cutpoint (>0.50) for a strong unidimensional scale (Mokken, 1982), there were a number of violations of the conventional value for inclusion of an item under the assumption of double-monotonicity (the convention is that Crit-value > 80 suggest an item should be removed) (Molenaar et al., 2000). These violations centred around #6 (Attitude/behaviours of clinicians) on both versions of the GUTSS, suggesting possible difficulties with this item. As shown in Table 8 if this item were removed, the properties of the GUTSS would improve marginally. The analysis suggests, effectively, that there was little difference between the original and revised versions.

Figure 2 shows the distribution of T-scores for the original and revised versions of the GUTSS by 0.5 standard deviations from the mean of 50 (i.e. at 5-point intervals). The largest discrepancies were in the 50-54 point range where there were 10 cases more on the revised GUTSS when compared with the original GUTSS. Given that the GUTSS descriptive system confounds current incontinence status with satisfaction, this discrepancy is not surprising. Figure 3 is a scatterplot of the original and revised GUTSS T-scores. The Spearman correlation was 0.97 (p<0.01, N = 177). Importantly, 87% of cases obtained scores on the revised GUTSS which were within 2-points of their score on the original GUTSS.

**Figure 2  Distribution of the original and revised GUTSS T-scores by 0.5 standard deviations**
When the reliability of the two versions of the GUTSS was examined, it was Cronbach $\alpha = 0.87$ for the original GUTSS and 0.89 for the revised GUTSS. As shown in Table 8 the reliability estimated by the Mokken $p$ was slightly higher for both GUTSS versions. The proportion of explained variance was 74% for the original GUTSS and 76% for the revised GUTSS.

Table 9 shows four tests of sensitivity using selected predictors from Table 6: on two of the predictors (treatment effect and information given) the two versions of the GUTSS were expected to be sensitive, and on the other two predictors (pre-treatment status and age group) the two GUTSS versions were expected to be insensitive. It was expected there would be no meaningful difference in T-scores between the original and revised versions of the GUTSS. The findings were consistent with these expectations.
Table 9  Tests of convergent and divergent validity for the original and revised GUTSS

Tests of convergent and divergent validity for the original and revised GUTSS

<table>
<thead>
<tr>
<th>Tests of convergent and divergent validity for the original and revised GUTSS</th>
<th>N</th>
<th>Original GUTSS</th>
<th>Revised GUTSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (a)</td>
<td>SD</td>
</tr>
<tr>
<td>UCA pretreatment symptoms</td>
<td>None/Slight/Mild</td>
<td>22</td>
<td>51.25</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>54</td>
<td>49.94</td>
</tr>
<tr>
<td></td>
<td>Severe</td>
<td>93</td>
<td>49.94</td>
</tr>
<tr>
<td>F-value (b)</td>
<td></td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>How successful was treatment</td>
<td>Cured</td>
<td>62</td>
<td>58.53</td>
</tr>
<tr>
<td></td>
<td>Partly cured</td>
<td>57</td>
<td>48.99</td>
</tr>
<tr>
<td></td>
<td>Improved</td>
<td>42</td>
<td>44.83</td>
</tr>
<tr>
<td></td>
<td>Not helped</td>
<td>13</td>
<td>31.66</td>
</tr>
<tr>
<td>F-value</td>
<td></td>
<td>95.84</td>
<td>82.11</td>
</tr>
<tr>
<td>Age group</td>
<td>31-45 years</td>
<td>31</td>
<td>48.55</td>
</tr>
<tr>
<td></td>
<td>46-60 years</td>
<td>73</td>
<td>50.27</td>
</tr>
<tr>
<td></td>
<td>61-75 years</td>
<td>57</td>
<td>50.36</td>
</tr>
<tr>
<td></td>
<td>75+ years</td>
<td>16</td>
<td>51.00</td>
</tr>
<tr>
<td>F-value</td>
<td></td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>Information given</td>
<td>Excellent</td>
<td>57</td>
<td>55.60</td>
</tr>
<tr>
<td></td>
<td>Very good</td>
<td>62</td>
<td>50.72</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>36</td>
<td>45.08</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>19</td>
<td>41.02</td>
</tr>
<tr>
<td>F-value</td>
<td></td>
<td>23.47</td>
<td>28.31</td>
</tr>
</tbody>
</table>

Notes:

a = non-transformed T-scores
b = ANOVA, all ANOVAs carried out on transformed data

Particularly important were the findings that (a) for all levels of all predictors the mean GUTSS scores did not vary between the two measures by more than 1 T-score point, and (b) that the F-values were very similar (in one test, treatment effect, favouring the original GUTSS and in the other test, information given, favouring the revised GUTSS).

In conclusion, the removal of the two filter items from the original GUTSS shortened the GUTSS without making any appreciable difference to its scores or responsiveness: the two versions are equivalent scales. The revised GUTSS is therefore used throughout the rest of this report.

6.2 Descriptive systems and item sensitivities

The descriptive systems of the four measures were mapped by item content, as shown in Table 10. This revealed that there were substantial differences between the measures. The GUTSS primarily measures treatment outcomes, the Consult SQ the technical skill and relationship with
the treating clinician, the CSQ-18 whether the treatment received was appropriate (general satisfaction) and the relationship with the treating clinician, and the PSI the involvement of the patient and his/her family in the decision-making process. It is likely these differences may partly explain the differences shown in Tables 4 and 6.

Spearman correlations between the four instruments were computed and are shown in Table 10. To test whether all four instruments do measure the same construct, all four were entered into a principal components analysis; the results showed a single factor explaining 77% of the variance (eigenvalue 3.07). The loadings on the factor were 0.93 (CSQ=18), 0.89 (PSI), 0.85 (GUTSS) and 0.83 (Consult SQ). These results suggest that the four measures assess the same construct, albeit very differently due to the different descriptive systems and response scales.

In addition to differences in the descriptive systems shown in Table 10, another potential source of difference is in the item response scales. The GUTSS uses a mixture of 5-point Likert-type response scales (e.g. Very satisfied/Satisfied/Neither/Dissatisfied/Very dissatisfied) and Guttman responses (e.g. Extremely problematic/Very problematic/Problematic/Slightly problematic/Not problematic). The Consult SQ uses standard 5-point Likert scales (e.g. Strongly agree/Agree/Not sure/Disagree/Strongly disagree). The CSQ-18 has modified 4-point Guttman-type response scales (e.g. No, definitely not/No, not really/Yes, generally/Yes, definitely) and the PSI utilises a variety of response scales. Some items on the PSI have modified 7-point Likert-type response scales (e.g. Strongly agree/Mildly agree/Not sure/Very mildly disagree/Mildly disagree/Moderately disagree/Strongly disagree) whereas other items have Guttman-type scales (e.g. All of the time/A good bit of the time/Some of the time/A little of the time/Hardly any of the time/Almost none of the time/None of the time). Almost certainly these differences partly explain the lower scores on the Consult SQ, since end aversion is a well-known phenomenon with Likert scales, but not with Guttman scales⁴.

---

⁴ Additionally, the instrument designers set different time periods for respondents to reflect upon. The GUTSS refers to the previous 4-weeks, the Consult SQ refers to the very last visit to the clinician, no timeframe is set for the CSQ-18, and the PSI asks respondents to reflect on their care over the past 6 to 8 months. In this study we attempted to minimise these differences through removing all references to timeframes, other than when they were embedded within item stems. It is unlikely that the instrument developers instructions regarding timeframes played any part in the results.
Table 10  Patient satisfaction measures descriptive systems

<table>
<thead>
<tr>
<th>Dimension (a)</th>
<th>Consult SQ</th>
<th>CSQ-18</th>
<th>GUTSS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access &amp; facilities</td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Information</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>****</td>
</tr>
<tr>
<td>Relationship</td>
<td>****</td>
<td>**</td>
<td>**</td>
<td>*****</td>
</tr>
<tr>
<td>Participation</td>
<td></td>
<td>**</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Technical skill</td>
<td>****</td>
<td>**</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Effectiveness</td>
<td></td>
<td>**</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Satisfaction general</td>
<td>***</td>
<td>****</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>***</td>
<td></td>
<td>**</td>
<td></td>
</tr>
</tbody>
</table>

Spearman correlations between scales (b)

<table>
<thead>
<tr>
<th></th>
<th>Consult SQ</th>
<th>CSQ-18</th>
<th>GUTSS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSQ-18</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUTSS</td>
<td>0.48</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSI</td>
<td>0.64</td>
<td>0.74</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

a = Adapted from Hawthorne 2006. Based on examination of instrument items.

b = All significant, p < 0.01.

Understandably differences in item response scales can affect how respondents react to them. Table 11 shows the mean item response by item type across the four instruments. The means, standard deviations and the proportion of cases endorsing ceiling scores are consistent with the literature which suggests that over 80% of patients are satisfied with their health care in the short term (Cartwright, 1967; Garcia-Aguilar et al., 1996; Gamagami et al., 1999; McConnell and Khubchandani, 1983; Marcello et al., 1993).

The average IRTCs (average rest of test correlations) were highest for the PSI (0.76), then the GUTSS (0.67), the Consult SQ (0.64) and the CSQ-18 (0.58). These results reflect the average interitem correlations, which were 0.59 (SD = 0.12) for the PSI, 0.55 (SD = 0.14) for the GUTSS, 0.45 (SD = 0.11) for the Consult SQ and 0.41 (SD = 0.14) for the CSQ-18. To examine the sensitivity of the items, the stronger representative predictor from factor 1, the factor which explained the highest proportion of variance in the factor analysis (Tables 5 and 6) at 31%, was used as the criterion against which items were tested. This was self-reported treatment effectiveness. The results (Table 11) showed that at the p<0.001 level all items from the GUTSS were definitely sensitive, as were 9/18 for the CSQ-18, 7/23 for the PSI and 2/18 items from the Consult SQ. The content of these items was (by instrument):

- Consult SQ: Satisfaction with this visit and will follow the clinician’s advice;
- CSQ-18: Satisfied with amount of help received, did service help you deal with health problem, did you get the service you needed, are there services you needed and didn’t get, rate the quality of the service, generally satisfied with received services, recommend service to friend, did the treatment meet your needs, would you go back to the same clinic;

---

5 The number of tests carried out on the items was 65, and if α = 0.05 then 5% of all significant results in Table 8 could be Type I errors. Based on a Bonferroni adjustment where \( \alpha = \frac{\alpha}{C} \), where \( C \) = the number of comparisons, to avoid the probability of any Type I errors, \( \alpha < 0.001 \).
- GUTSS: Happy with treatment effect, still having health problems, satisfaction with treatment outcome, disappointment, pre-treatment information, attitude/behaviour of health professionals, explanations of treatment results and satisfaction with clinic care;
- PSI: Explanations were not understood, feeling helpless, feeling out of control of the situation, satisfied with cooperation from health care providers, satisfied with how decisions were made, satisfied with level of care received and feel respected by health professionals.

Examination of the items which were definitely non-sensitive (i.e. all those where \( p > 0.05 \)) showed that these were:
- Consult SQ: Telling the clinician personal things, that the time with the clinician was too short, that there were some things the clinician did not know about the patient, that the patient was examined very thoroughly, that the clinician noticed the patient as a person, that the clinician viewed the patient as a person, that the clinician knew all about the patient, and that the patient found it difficult to tell the clinician private things;
- CSQ-18: Being seen promptly, attractiveness of the clinic, characteristics of the clinic building, and individual rights being respected;
- PSI: That the patient went through more than they expected, and satisfaction with the involvement of the patient's family.
<table>
<thead>
<tr>
<th>Item</th>
<th>Consult SQ</th>
<th>CSQ-18</th>
<th>GUTSS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Mean SD %</td>
<td>Mean SD %</td>
<td>Mean SD %</td>
<td>Mean SD %</td>
</tr>
<tr>
<td>1</td>
<td>0.84 0.22 56%</td>
<td>0.77 0.26 48%</td>
<td>0.78 0.26 46%</td>
<td>0.54 0.37 27%</td>
</tr>
<tr>
<td>2</td>
<td>0.85 0.18 52%</td>
<td>0.65 0.36 37%</td>
<td>0.73 0.27 35%</td>
<td>0.56 0.39 30%</td>
</tr>
<tr>
<td>3</td>
<td>0.83 0.19 46%</td>
<td>0.88 0.23 74%</td>
<td>0.76 0.26 43%</td>
<td>0.77 0.28 46%</td>
</tr>
<tr>
<td>4</td>
<td>0.81 0.22 44%</td>
<td>0.71 0.35 47%</td>
<td>0.84 0.25 61%</td>
<td>0.81 0.26 56%</td>
</tr>
<tr>
<td>5</td>
<td>0.58 0.31 15%</td>
<td>0.81 0.27 57%</td>
<td>0.82 0.22 51%</td>
<td>0.76 0.29 46%</td>
</tr>
<tr>
<td>6</td>
<td>0.83 0.18 44%</td>
<td>0.85 0.23 63%</td>
<td>0.90 0.18 69%</td>
<td>0.82 0.27 59%</td>
</tr>
<tr>
<td>7</td>
<td>0.61 0.31 21%</td>
<td>0.85 0.23 63%</td>
<td>0.82 0.21 50%</td>
<td>0.83 0.26 61%</td>
</tr>
<tr>
<td>8</td>
<td>0.57 0.28 14%</td>
<td>0.82 0.23 53%</td>
<td>0.68 0.19 63%</td>
<td>0.81 0.27 55%</td>
</tr>
<tr>
<td>9</td>
<td>0.83 0.17 40%</td>
<td>0.71 0.25 31%</td>
<td>0.57 0.25 31%</td>
<td>0.86 0.24 61%</td>
</tr>
<tr>
<td>10</td>
<td>0.77 0.23 37%</td>
<td>0.80 0.26 54%</td>
<td>0.56 0.26 54%</td>
<td>0.89 0.23 73%</td>
</tr>
<tr>
<td>11</td>
<td>0.66 0.27 18%</td>
<td>0.87 0.21 65%</td>
<td>0.60 0.21 65%</td>
<td>0.76 0.28 44%</td>
</tr>
<tr>
<td>12</td>
<td>0.79 0.21 36%</td>
<td>0.85 0.22 61%</td>
<td>0.82 0.22 61%</td>
<td>0.78 0.26 41%</td>
</tr>
<tr>
<td>13</td>
<td>0.67 0.26 21%</td>
<td>0.85 0.23 64%</td>
<td>0.80 0.23 64%</td>
<td>0.77 0.26 39%</td>
</tr>
<tr>
<td>14</td>
<td>0.51 0.28 9%</td>
<td>0.87 0.22 69%</td>
<td>0.74 0.22 69%</td>
<td>0.77 0.25 43%</td>
</tr>
<tr>
<td>15</td>
<td>0.55 0.24 7%</td>
<td>0.86 0.18 60%</td>
<td>0.59 0.18 60%</td>
<td>0.78 0.25 37%</td>
</tr>
<tr>
<td>16</td>
<td>0.57 0.29 10%</td>
<td>0.81 0.24 56%</td>
<td>0.66 0.24 56%</td>
<td>0.78 0.24 38%</td>
</tr>
<tr>
<td>17</td>
<td>0.69 0.28 25%</td>
<td>0.87 0.19 65%</td>
<td>0.49 0.19 65%</td>
<td>0.83 0.21 44%</td>
</tr>
<tr>
<td>18</td>
<td>0.68 0.27 24%</td>
<td>0.88 0.20 72%</td>
<td>0.72 0.20 72%</td>
<td>0.85 0.21 54%</td>
</tr>
</tbody>
</table>

Notes:
- a = Mean score as proportion of available scale range.
- b = Standard deviation.
- c = Item rest of test correlation.
- d = Kruskall-Wallis $\chi^2$ test by self-reported perceived treatment success (Table 5). * = p < 0.05, ** = p < 0.01, *** = p < 0.001.
A difficulty is that on items with similar content for the Consult SQ there were some that were sensitive and others that were not. Consider the two items “The time I was able to spend with the clinician (doctor, physiotherapist or nurse) was a bit too short” (Consult SQ item 5 in Table 8) with “The time I was allowed to spend with the clinician (doctor, physiotherapist or nurse) was not long enough to deal with everything I wanted” (Consult SQ item 11 in Table 9). Precisely why one of these two items should have proved sensitive to treatment effect and the other not sensitive is unknown. It may reflect that respondents are more likely to agree with a negative item than reject a positive one (Reiser et al., 1986). The agreement between the two items was \( \kappa = 0.42 \), suggesting a fair/moderate level of agreement (Landis and Koch, 1977). There was agreement by just 104/177 cases.

Subject to this caveat, in general it could be expected that non-sensitive items cover areas of health care and patient satisfaction that are not especially pertinent to the treatment of incontinence in Australia, at least in this study sample.

### 6.3 Scale properties

Table 12 presents summary data on the psychometric properties of each of the patient satisfaction measures, including percentage mean scores and standard deviations, internal consistency (the Cronbach alpha), the percentage of explained variance and the Loevinger H test for unidimensionality. The scale means were similar across three of the instruments, but were lower for the Consult SQ. The standard deviations were similar across all four measures. The extent of skew within the measures can be gauged by the proportion of cases obtaining ceiling scores; for three of the measures between 21% to 28% of scores were within the top 5% of the scale ranges. For the Consult SQ the proportion was 3%. The Cronbach alphas were all above 0.90 for these instruments. Nunnally & Bernstein (1994) argue that for group measurement an \( \alpha \) of 0.80 is adequate, but that for individual assessments an \( \alpha \) of >0.90 should be desired. The Cronbach alphas, therefore, suggest that each of the measures would be potentially suitable for individual use. The percentage of variance explained ranged from 59% for the CSQ-18 through to 77% for the GUTSS – which may suggest that the GUTSS is measuring a narrower construct than the other instruments. All four instruments were unidimensional as assessed by the Loevinger H statistic, although only the GUTSS and PSI would be regarded as very strong scales on this criterion.
### Table 12 Scale properties

<table>
<thead>
<tr>
<th>Scale properties</th>
<th>Consult SQ</th>
<th>CSQ-18</th>
<th>GUTSS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (percentage)</td>
<td>70.43</td>
<td>81.79</td>
<td>81.50</td>
<td>79.06</td>
</tr>
<tr>
<td>SD</td>
<td>16.69</td>
<td>15.16</td>
<td>17.54</td>
<td>19.87</td>
</tr>
<tr>
<td>% Ceiling (a)</td>
<td>3%</td>
<td>21%</td>
<td>28%</td>
<td>25%</td>
</tr>
<tr>
<td>Average IRTC (b)</td>
<td>0.45</td>
<td>0.41</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>Cronbach α</td>
<td>0.93</td>
<td>0.90</td>
<td>0.90</td>
<td>0.97</td>
</tr>
<tr>
<td>% variance</td>
<td>65%</td>
<td>59%</td>
<td>77%</td>
<td>70%</td>
</tr>
<tr>
<td>Loevinger H</td>
<td>0.51</td>
<td>0.42</td>
<td>0.58</td>
<td>0.63</td>
</tr>
</tbody>
</table>

**Notes:**

a = Top 5% of scale range.

b = item rest of test correlation

These conclusions, however, must be reviewed in light of instrument length and acquiescent response set bias (ARSB). The data in Table 11, the inter-item correlations and Cronbach alphas reported in Table 12 are notable for showing that; generally, instrument length is associated with higher alphas and smaller IRTC standard deviations. The interpretation would be that these are above those recommended in the literature for homogenous scales; certainly as shown in section 5 this was the case for the GUTSS. Accepted values for scale homogeneity are in the 0.2 to 0.4 inter-item correlation ranges. Where mean inter-item correlations are higher than 0.5 there is likely to be redundancy (double-counting) (Briggs and Cheek, 1986). This suggests redundancy or double-counting within each of the four measures, particularly the PSI. It may be that the very high alpha- and H-values for this measure reflect similarity among respondents’ selections. An undue number of similar selections across consecutive items is indicative of ARSB. The presence of ARSB using two methods was accordingly calculated for each of the instruments.

The first criterion was that any 6 consecutive items answered identically, regardless of the content or direction of the item (positive or reversed), would be indicative of ARSB, excluding those cases who answered all items at the ceiling level (i.e. excluding those who indicated extreme satisfaction). The results showed that suspected ARSB would be present in 10% of GUTSS respondents, 11% for the Consult SQ, 25% for the CSQ-18 and in 29% of cases for the PSI.
A second test of ARSB was to examine changes in standard deviations to items, by the item administration order, regardless of item content or direction. A decline in standard deviations across an instrument is suggestive of ARSB because as respondents work through an instrument they may pay less attention to individual items and tend to suffice leading to artificial similarity in their responses. The results, shown in Figure 2, show significant declines in standard deviations for the CSQ-18 (first 9 items compared with second 9 items, Welch’s approximate t-test, $t = 2.68$, $p = 0.02$, df = 11) and the PSI (first 11 items compared with second 12 items, Welch’s approximate t-test, $t = 3.12$, $p < 0.01$, df = 15). The was no significant difference for the Consult SQ (first 9 items compared with second 9 items, Welch’s approximate t-test, $t = 1.11$, $p = 0.18$, df = 11) or GUTSS (first 4 items compared with second 4 items, Welch’s approximate t-test, $t = 1.10$, $p = 0.93$, df = 5).

The only instrument to exhibit ARSB on both these tests was the PSI. Since the instruments were administered in the order GUTSS, CSQ-18, Consult SQ, PSI it is likely that the presence of ARSB was a function of the measure rather than its place in the questionnaire. (There was evidence for ARSB on the 2nd and 4th administered instruments. If it was solely a function of instrument order, then it could be expected that it would be present on the 3rd instrument rather than the 2nd). Where there is evidence of ARSB, reliability and explained variance estimates will be inflated especially in long instruments – and the instrument with the highest proportion of cases exhibiting some ARSB, the PSI, was also the longest of the instruments.

Additionally, it may be that the long and convoluted item stems and response options contributed to the ARSB in the PSI. For example, item #16 in the PSI probes: ‘During the treatment you received (for your incontinence) how satisfied were you with the amount of choice you had in decisions affecting your health care?’ with the response options of ‘Not really satisfied/Generally satisfied but significant areas of dissatisfaction/Generally satisfied but minor areas of dissatisfaction/Quite satisfied/Mostly satisfied/Very satisfied/Completely satisfied’. The presence of ARSB in the PSI may partly account for the very high Cronbach alpha and Loevinger H obtained by the PSI (Table 9). It may also have played a part in the CSQ-18.
The four measures were subject to several tests of discrimination by the predictors of patient satisfaction shown in Tables 6 and 9, by treatment type (physiotherapy, physiotherapy and surgery and surgery) and by the UCA then-test treatment outcomes (Table 2). As expected, all four instruments discriminated between groups by effect of treatment. For the self-report effectiveness item the scores on the Consult SQ followed a different pattern than the other instruments; whereas the other instruments all gave monotonic scores, on the Consult SQ those who reported some improvement obtained reported higher satisfaction than did those who reported they were partly cured of their incontinence.

All four instruments also discriminated by pre-treatment information given to the patients, and none of the instruments discriminated by age group. For treatment type, the only instrument that significantly discriminated was the GUTSS, and for the UCA then-test the instruments that discriminated were the CSQ-18 and the GUTSS. It is worthwhile noting that on the UCA then-test for all instruments the group reporting the highest level of dissatisfaction were those participants who were classified as having no changes in incontinence status from their health care. Although this may be due to the small sample size, it may be that those who had worse outcomes were grateful for their health care whereas those for whom there was no change may have been disappointed and felt their treatment was a waste of time and effort. The details are given in Table 13.
### Table 13  Tests of convergent and divergent validity for the four incontinence measures, T-scores

<table>
<thead>
<tr>
<th>Tests of convergent and divergent validity for the four incontinence measures, T-scores</th>
<th>N</th>
<th>Consult SQ</th>
<th>CSQ-18</th>
<th>GUTSS</th>
<th>PSI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiotherapy</td>
<td>45</td>
<td>50.49</td>
<td>8.90</td>
<td>48.10</td>
<td>9.09</td>
</tr>
<tr>
<td>Physiotherapy &amp; surgery</td>
<td>55</td>
<td>48.57</td>
<td>10.98</td>
<td>49.12</td>
<td>11.03</td>
</tr>
<tr>
<td>Surgery</td>
<td>69</td>
<td>50.00</td>
<td>10.01</td>
<td>51.79</td>
<td>9.60</td>
</tr>
<tr>
<td>F-value (b)</td>
<td>0.35</td>
<td>2.57</td>
<td>4.90**</td>
<td>1.44</td>
<td></td>
</tr>
<tr>
<td>UCA then treatment test (c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worse outcomes</td>
<td>10</td>
<td>48.55</td>
<td>11.31</td>
<td>45.45</td>
<td>9.68</td>
</tr>
<tr>
<td>No change</td>
<td>20</td>
<td>48.51</td>
<td>13.19</td>
<td>44.04</td>
<td>13.97</td>
</tr>
<tr>
<td>Improved</td>
<td>135</td>
<td>49.99</td>
<td>9.40</td>
<td>51.09</td>
<td>8.86</td>
</tr>
<tr>
<td>F-value</td>
<td>0.10</td>
<td>4.47*</td>
<td>12.40***</td>
<td>2.42</td>
<td></td>
</tr>
<tr>
<td>How successful was treatment (self-report)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cured</td>
<td>62</td>
<td>53.39</td>
<td>7.38</td>
<td>54.84</td>
<td>6.52</td>
</tr>
<tr>
<td>Partly cured</td>
<td>55</td>
<td>48.28</td>
<td>9.81</td>
<td>49.94</td>
<td>9.05</td>
</tr>
<tr>
<td>Improved</td>
<td>41</td>
<td>49.48</td>
<td>8.82</td>
<td>46.68</td>
<td>8.63</td>
</tr>
<tr>
<td>F-value</td>
<td>5.93***</td>
<td>13.72***</td>
<td>79.16***</td>
<td>11.13***</td>
<td></td>
</tr>
<tr>
<td>Information given</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>57</td>
<td>55.16</td>
<td>8.04</td>
<td>55.43</td>
<td>7.67</td>
</tr>
<tr>
<td>Very good</td>
<td>61</td>
<td>49.35</td>
<td>9.38</td>
<td>50.43</td>
<td>8.65</td>
</tr>
<tr>
<td>Good</td>
<td>36</td>
<td>47.63</td>
<td>7.98</td>
<td>45.87</td>
<td>7.37</td>
</tr>
<tr>
<td>Fair</td>
<td>17</td>
<td>38.17</td>
<td>11.15</td>
<td>38.44</td>
<td>13.03</td>
</tr>
<tr>
<td>F-value</td>
<td>16.57***</td>
<td>20.38***</td>
<td>27.45***</td>
<td>17.40***</td>
<td></td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-45 years</td>
<td>31</td>
<td>49.03</td>
<td>11.03</td>
<td>48.65</td>
<td>11.45</td>
</tr>
<tr>
<td>46-60 years</td>
<td>71</td>
<td>49.01</td>
<td>10.75</td>
<td>50.26</td>
<td>10.95</td>
</tr>
<tr>
<td>61-75 years</td>
<td>56</td>
<td>50.86</td>
<td>8.81</td>
<td>50.29</td>
<td>8.53</td>
</tr>
<tr>
<td>75+ years</td>
<td>16</td>
<td>52.06</td>
<td>8.71</td>
<td>50.82</td>
<td>7.30</td>
</tr>
<tr>
<td>F-value</td>
<td>0.41</td>
<td>0.17</td>
<td>0.63</td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

- a = non-transformed T-scores
- b = ANOVA, all ANOVAs carried out on transformed data. * = p<0.05, ** = p<0.01, *** = p<0.001.
- c = Present – Retrospective scores. See Table 2.
7 Deriving a New Short Patient Satisfaction Scale

Although Section 6 showed that each of the four instruments examined was reliable and sensitive, the results also suggested that there were issues with coverage, with the items and item response scales and data distribution, and that each included redundant items. These issues suggest improvements could be made. Additionally, there is need for a short patient satisfaction instrument to be used in the Australian research context and for a single item for use in everyday primary care consultations.

7.1 Some issues arising from the analysis of instruments in Section 6

7.1.1 A theoretical model of patient satisfaction and its measurement: issues in coverage

Representational theory postulates that an instrument (the manifest model) should adequately represent the construct universe – which is the underlying latent concept that is being measured (in this case, a theoretical model of patient satisfaction). Validity is where there is an isomorphic match between the manifest and latent models (Pedhazur and Schmelkin, 1991; Michell, 1986). Hawthorne (2006c) discussed a model of patient satisfaction based on Donabedian’s notion that:

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. (Donabedian, 1988, p. 1746)

The implication is that the construct of patient satisfaction covers all aspects of care quality, particularly the interpersonal processes. Hawthorne’s review identified the dimensions of this care quality as being those presented in Table 14 (Hawthorne, 2006c). Dissatisfaction will occur where there is a cluster of small transgressions or a major failure in service provision. This model of patient satisfaction postulates that in a comprehensive assessment of patient satisfaction all seven dimensions will contribute and should be measured on the a priori axiom of equal weighting. If this axiom is not supported, then a smaller number of dimensions can be measured without compromising the validity of the measurement model.

<table>
<thead>
<tr>
<th>Table 14 Patient satisfaction measures descriptive systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient satisfaction measures descriptive systems</strong></td>
</tr>
<tr>
<td><strong>Dimension (a)</strong></td>
</tr>
<tr>
<td>Access &amp; facilities</td>
</tr>
<tr>
<td>Information</td>
</tr>
<tr>
<td>Relationship</td>
</tr>
<tr>
<td>Participation</td>
</tr>
<tr>
<td>Technical skill</td>
</tr>
<tr>
<td>Effectiveness</td>
</tr>
<tr>
<td>Satisfaction general</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

Notes:

a = Adapted from Hawthorne 2006. Based on examination of instrument items.
For example, of the six items measuring access and facilities (Table 14), those covering the attractiveness and characteristics of the clinic (CSQ-18, items #2 and #3 in Table 11) were non-responsive which suggests that these issues were not important in the discrimination of patient satisfaction in the study population. In contrast, 3 of the 4 other items measuring access and facilities discriminated; these covered being seen promptly (CSQ-18, item #1 in Table 11), the consultation being not long enough and needing a little longer with the clinician (Consult SQ #11 and #16 in Table 11). That the 6th item, that the consultation was too short (Consult SQ #5 in Table 11), did not discriminate may reflect that it is a reversed item. This finding is consistent with Reiser et al.'s (1986) observation that respondents are more likely to agree with a negative item than reject a positive one. In summary, the important aspect of access and facilities to capture is time with the clinician rather than an assessment of the physical environment.

As well as ensuring adequate coverage for validity purposes, scales should be parsimonious: the longer each instrument in a questionnaire, the longer the questionnaire. Long questionnaires may increase both participation and response rates (Dillman, 1978; Dillman et al., 1996; Day et al., 1995; Pedhazur and Schmelkin, 1991; WHOQoL Group, 1998; Ware et al., 1996; Berwick et al., 1991; Yammarino et al., 1991). Indeed, the data from this study would suggest that ARSB was present in the longest instrument, the PSI (see Figure 2). Parsimony is also required to meet classic test theory axioms which state that instrument scores should be isomorphic with the underlying construct. This implies that scales should be unidimensional to minimise error (Anastasi, 1976; Pedhazur and Schmelkin, 1991; Streiner and Norman, 1995).

These axioms suggest that any new measure of patient satisfaction should be as short as possible, with probably just 1 item from each of the dimensions reported in Table 14 that are deemed important.

### 7.1.2 Item response scales, data distribution and item stems

**Item response scales and data distributions**

Item response scales should contain sufficient categories so that all respondents have the option of a category that fits their situation. Subject to this requirement, response categories should be limited to those that report meaningful differences or which do not contribute to increased error (typically response bias, as shown above for the PSI items). Generally, 5 response categories have been the convention for self-report scales, at least since Likert (1932).

The argument for increasing this number relies upon achieving greater sensitivity through allowing a more refined match between respondent condition and item response, thus reducing measurement error and increasing sensitivity. However, this argument is subject to channel capacity limitation, where channel capacity is the ability to meaningfully discriminate between different choices (Hawthorne et al., 2006b). Where channel capacity is exceeded, measurement error increases (Miller, 1956). Most people can reliably discriminate between 7 levels (plus or minus 2) (Miller, 1956; Molenaar, 1982). Although there is little research confirming this in the health sciences, it has been used as the standard for identifying minimum important differences (Norman et al., 2003). The implication is that the number of response categories should not exceed channel capacity.

In the four patient satisfaction instruments assessed, the number of response levels was 5 for the GUTSS, 4 for the CSQ-18, 5 for the Consult SQ and 7 for the PSI. When the data in Table 11 was used to compute the means across each instrument (after transformation) there were no significant differences (ANOVA, $F = 0.84$, $df = 3.63$, $p = 0.48$). Additionally, an item analysis revealed that while an average of 16% of cases selected the bottom two categories for the Consult SQ, the comparable figures for the CSQ-18, GUTSS and PSI were 9%, 6% and 10% respectively (Table 15). These findings suggest that there was no clear-cut pattern of responsiveness associated with increased number of response categories. This conclusion, however, is subject to the finding that there was evidence of ARSB in the PSI, suggesting perhaps that the combination...
of instrument length and the number of response categories being near the limit of channel capacity may induce sufficing and bias. The implication is that any future patient satisfaction instrument should use <7 levels on response scales.

**Table 15 Item response distributions (not recoded)**

<table>
<thead>
<tr>
<th>Item response distributions (not recoded)</th>
<th>Response level (percentages)</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
<th>Level 6</th>
<th>Level 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consult SQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>56%</td>
<td>33%</td>
<td>6%</td>
<td>4%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>52%</td>
<td>39%</td>
<td>7%</td>
<td>2%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>46%</td>
<td>43%</td>
<td>8%</td>
<td>2%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>44%</td>
<td>41%</td>
<td>11%</td>
<td>3%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15%</td>
<td>42%</td>
<td>12%</td>
<td>20%</td>
<td>11%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>44%</td>
<td>44%</td>
<td>10%</td>
<td>2%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>21%</td>
<td>37%</td>
<td>14%</td>
<td>21%</td>
<td>7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>14%</td>
<td>38%</td>
<td>15%</td>
<td>31%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>40%</td>
<td>53%</td>
<td>5%</td>
<td>2%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>37%</td>
<td>44%</td>
<td>12%</td>
<td>5%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>18%</td>
<td>53%</td>
<td>12%</td>
<td>12%</td>
<td>6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>36%</td>
<td>50%</td>
<td>8%</td>
<td>5%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>21%</td>
<td>43%</td>
<td>21%</td>
<td>12%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>10%</td>
<td>30%</td>
<td>22%</td>
<td>33%</td>
<td>5%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>7%</td>
<td>33%</td>
<td>36%</td>
<td>21%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>10%</td>
<td>47%</td>
<td>12%</td>
<td>24%</td>
<td>7%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>25%</td>
<td>48%</td>
<td>10%</td>
<td>11%</td>
<td>6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>24%</td>
<td>46%</td>
<td>11%</td>
<td>17%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSQ-18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48%</td>
<td>38%</td>
<td>12%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>37%</td>
<td>39%</td>
<td>7%</td>
<td>18%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>74%</td>
<td>20%</td>
<td>3%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>47%</td>
<td>34%</td>
<td>5%</td>
<td>15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>57%</td>
<td>35%</td>
<td>2%</td>
<td>6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>63%</td>
<td>32%</td>
<td>2%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>63%</td>
<td>32%</td>
<td>2%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>53%</td>
<td>41%</td>
<td>3%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>31%</td>
<td>53%</td>
<td>11%</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>54%</td>
<td>33%</td>
<td>11%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>65%</td>
<td>30%</td>
<td>3%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>61%</td>
<td>34%</td>
<td>2%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>64%</td>
<td>30%</td>
<td>4%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>69%</td>
<td>26%</td>
<td>2%</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>60%</td>
<td>39%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>56%</td>
<td>34%</td>
<td>9%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>65%</td>
<td>34%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>72%</td>
<td>22%</td>
<td>6%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Table continued next page.)
<table>
<thead>
<tr>
<th>Item stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>An axiom of measurement is that item stems should be short and free of obvious defects (Foddy, 1993). Unfortunately, a few items failed this basic criterion. The key problems were in relation to unnecessarily long complex item stems involving double-barrelled statements, as these examples show:</td>
</tr>
<tr>
<td>▪ How clearly did the clinician (doctor, physiotherapist or nurse) with whom you worked most closely understand your problem and how you felt about it? (CSQ-18, #10 in Table 11);</td>
</tr>
<tr>
<td>▪ The clinician (doctor, physiotherapist or nurse) was interested in me as a person, and not just my illness (Consult SQ, #13 in Table 11);</td>
</tr>
<tr>
<td>▪ People sometimes feel that they have gone through a lot when receiving health care. Do you agree that you have gone through a lot in relation to the health care you have received for your incontinence? (PSI, #1 in Table 11);</td>
</tr>
</tbody>
</table>

When all such items were tabulated, the results suggested that there were no item stems with particular difficulty in the revised GUTSS, at least 1 for the CSQ-18 (CSQ-18, #10 in Table 11), 2 for the Consult SQ (#3 and #13 in Table 11), and 3 for the PSI (#1, #13, and #15 in Table 11).
7.2 Constructing a short assessment of patient satisfaction scale

Given the very small numbers of cases endorsing lowest order levels (Levels 4 through 7 depending upon the instrument), in the interests of a stable model, these levels were collapsed prior to further analysis (for example, the Consult SQ levels 4 & 5 were collapsed). Similarly, where there were inconsistent responses as identified through partial credit item response theory (IRT) analysis, these were also combined. Non-sensitive items were removed; these were the items that failed to discriminate between known groups (see Table 11). Additionally, items with wording difficulty were also removed. The remaining items (N = 49) were pooled with the intent of creating a short assessment of patient satisfaction scale.

7.2.1 Data analysis procedure

Using the full pooled dataset of items (N = 49), iterative Mokken and partial credit item response theory (IRT) analyses were conducted with the removal of an item at each iteration until the most parsimonious solution was obtained consistent with measuring the seven dimensions outlined in Table 14. Although Mokken analysis is appropriate for examining the properties of a scale (and in particular the relationship between an item and a scale) it does not enable examination of individual item thresholds (Molenaat et al., 2000), for which purpose Rasch IRT analysis was used.

The criteria for item removal following Mokken analysis were (in hierarchical order of importance):

- Statistical comparison of item fit, assessed by Crit-values (where two items were compared, the item with the higher Crit-value was considered for removal) under both P-matrix and rest-of-test score models (the sample was randomly split into three groups for this latter test);
- Comparison of Loevinger H-values, where of items being compared the item with the lowest H-value was considered for removal;
- Examination of item data distributions (see Table 15), where the item with the worst distribution was considered for removal
- Examination of the item stem and response categories where those items with long stems or with meanings judged to have non-Australian emic effects were considered for removal.

Where two items performed similarly on these criteria, each was removed in turn and the model tested where the criteria were the scale Loevinger H-value and the Mokken $\rho$-reliability estimate. The model with the higher properties was accepted, subject to the findings from a partial credit IRT analysis of the same items.

The items in the item pool after each iterative Mokken analysis were examined using a partial credit IRT model (Andrich et al., 2004; Bond and Fox, 2001) for point biserial correlations, disordered probability thresholds, differential item analysis, and item fit residuals. This information was used in conjunction with that obtained from the Mokken analysis to inform decisions regarding item removal or retention.

The Mokken and IRT iterations were performed such that the number of items measuring each dimension on Table 15 was maintained ±1 item. This ensured equal and consistent representation of the dimensions within the ever diminishing item pool. This procedure was an attempt to achieve a balance between fidelity to the model of patient satisfaction shown in Table 14, the statistical requirements for a strong homogenous scale and the measurement properties of individual items.
Table 16  **Final model of a unidimensional Short Assessment of Patient Satisfaction scale (SAPS)**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>N</th>
<th>Item stem (abbreviated)</th>
<th>Item source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness</td>
<td>1</td>
<td>Happy with the effect of your treatment</td>
<td>GUTSS</td>
</tr>
<tr>
<td>Information</td>
<td>2</td>
<td>Satisfaction with explanations of treatment results</td>
<td>GUTSS</td>
</tr>
<tr>
<td>Technical skill</td>
<td>3</td>
<td>The clinician was careful to check everything</td>
<td>Consult SQ</td>
</tr>
<tr>
<td>Participation</td>
<td>4</td>
<td>Satisfaction with health care choices</td>
<td>PSI</td>
</tr>
<tr>
<td>Relationship</td>
<td>5</td>
<td>How much were you respected</td>
<td>PSI</td>
</tr>
<tr>
<td>Access &amp; facilities</td>
<td>6</td>
<td>The time with the clinician was not long enough</td>
<td>Consult SQ</td>
</tr>
<tr>
<td>Satisfaction general/Other</td>
<td>7</td>
<td>Happy with the care received</td>
<td>GUTSS</td>
</tr>
</tbody>
</table>

7.2.2 Findings

The final double-monotone model is shown in Tables 16 and 17. Table 16 shows the final 7 items, one for each of the dimensions of patient satisfaction, an abbreviated item stem to convey the content of the item and the source instrument. Table 17 shows the psychometric properties of the items, and of the whole scale.

The Loevinger H, fit of each item exceeded the conventional cutpoint for inclusion in a homogenous scale (0.40). As shown there were almost no violations of Guttman monotonicity. The Crit-values implied that these violations may be due to sampling errors. The scale Loevinger H exceeded the conventional requirement (0.50) for a evidence of a strong unidimensional scale (Mokken, 1982). The partial credit IRT analysis results show a very consistent relationship between the items (the point biserial correlations), a good spread of logit locations suggesting that a wide range of respondents will find the items acceptable – although it should be noted that the logits here suggest that most respondents were well satisfied with their health care. The easiest item for respondents to endorse was clearly #6 (The time with the clinician was not long enough); perhaps this should not be unexpected! The fit residuals, $\chi^2$'s and p-values indicate that no item was misfitted to the model. For convenience, the scale is referred to as the SAPS (Short Assessment of Patient Satisfaction).
Table 17  Psychometric properties of SAPS items and the SAPS scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Mokken analysis</th>
<th>Partial credit IRT analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hi (a)</td>
<td>Crit-value (b)</td>
</tr>
<tr>
<td>1</td>
<td>0.52</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.55</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.56</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.55</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0.58</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.51</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>0.56</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes:

- Scale statistics: Loevinger H: 0.55, ρ = 0.86, Cronbach α = 0.86
- a = Hi = Item coefficient of scalability
- b = Crit-value under P-matrix analysis
- c = point biserial correlation
- d = In logits
- e = standard error

7.2.3 Scoring the SAPS

As indicated above, many items in the dataset (and in the SAPS) suffer sparse data and were therefore recoded. Additionally, as shown in Table 15 there are differing response scales used in the four patient satisfaction instruments. These two issues pose particular problems for scoring of composite instruments.

The assumption behind summated scales is that each item is weighted equally, i.e. it contributes as much to the scale score as any other item. This assumes an equal number of response categories. Where there are discrepancies between response categories, this effectively weights the responses by the category number. For example, consider a patient satisfaction instrument with two items; one is dichotomous (0, 1) and the other has a 5-point response scale (0, 1, 2, 3, 4). Under summation, the maximum scale score is 5, but most of this score is due to the second item. Effectively, this scale would be weighted 20% and 80% for the 1st and 2nd items respectively; with the 2nd item exerting undue leverage on scale scores.

In the case of the items contributing to the SAPS, the original instrument number of response categories were 5 (for the GUTSS and Consult SQ items, #1, #2, #3, #6 and #7 in Tables 16 and 17) and 7 (for the PSI items, #4 and #5 in Tables 16 and 17). Following recoding for the psychometric analyses to deal with the sparse data these were 3 points for item #6, 4 points for items #1, #2, #3 and #7, and 5 points for items #4 and #5. Although the seemingly obvious solution to the leverage issue posed by these differences is to weight the item responses, the literature suggests that weighting achieves little; the general advice is not to weight (Nunnally and Bernstein, 1994; Trauer and MacKinnon, 2001). In this study there is an even more compelling reason not to weight the items: most respondents were well satisfied, thus any recoding or weighting of the original items may render the response scales inapplicable in situations where participants have reason to be dissatisfied with their health care.
Consequent upon these considerations, the scale score for the SAPS was simply the summed value of responses. Following summation the potential score range was from 0 through to 39, where a high score represents a high level of satisfaction. Figure 5 shows the distribution of SAPS scores. This suggests that the SAPS spread cases out reasonably evenly across the scale range. Most cases were generally satisfied (those with high scores), there were a smaller number who were somewhat dissatisfied (with modest scores) and some who were dissatisfied (with low scores).

**Figure 5  Score distribution for the Short Assessment of Patient Satisfaction scale**

The basic properties of the SAPS are given in Table 18, along with those of the other patient satisfaction instruments. As shown, the mean SAPS score was similar to the other measures, except for the Consult SQ, and the standard deviation was marginally smaller. The percentage within the top 5% of the scale range was slightly smaller than for any of the other measures, again except for the Consult SQ. SAPS scores correlated with the Consult SQ scores $r_s = 0.73$, the CSQ-18 $r_s = 0.78$, the GUTSS $r_s = 0.83$ and the PSI $r_s = 0.83$ ($p < 0.01$ for all). These Spearman correlations are suggestive that the SAPS measures the same construct as the other measures.

**Table 18  Patient satisfaction scores, percentage scores**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>% at ceiling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consult SQ</td>
<td>177</td>
<td>70.28</td>
<td>16.72</td>
<td>3%</td>
</tr>
<tr>
<td>CSQ-18</td>
<td>176</td>
<td>81.63</td>
<td>15.16</td>
<td>21%</td>
</tr>
<tr>
<td>GUTTS (modified version)</td>
<td>177</td>
<td>81.58</td>
<td>17.58</td>
<td>31%</td>
</tr>
<tr>
<td>Patient Satisfaction Index (PSI)</td>
<td>176</td>
<td>78.89</td>
<td>19.92</td>
<td>24%</td>
</tr>
<tr>
<td>SAPS</td>
<td>176</td>
<td>82.52</td>
<td>15.12</td>
<td>20%</td>
</tr>
</tbody>
</table>

Notes

$a =$ Ceiling defined as top 5% of scale range
When SAPS scores were examined by the predictors of patient satisfaction from Table 13, the data showed that it was sensitive to treatment type, the UCA then-test treatment effect, the self-report treatment effectiveness and to information given. It was not sensitive to age group. On the UCA then-test the same pattern as reported on the other instruments (Table 13) regarding those who had not changed in their incontinence status was also reported on the SAPS: these cases obtained the lowest satisfaction scores.
### Table 19  Tests of convergent and divergent validity for the SAPS, T-scores

<table>
<thead>
<tr>
<th>Tests of convergent and divergent validity for the SAPS, T-scores</th>
<th>N</th>
<th>SAPS Mean</th>
<th>SD (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Physiotherapy</td>
<td>45</td>
<td>47.54</td>
</tr>
<tr>
<td></td>
<td>Physiotherapy &amp; surgery</td>
<td>55</td>
<td>49.23</td>
</tr>
<tr>
<td></td>
<td>Surgery</td>
<td>69</td>
<td>52.25</td>
</tr>
<tr>
<td>F-value (b)</td>
<td></td>
<td></td>
<td>3.12*</td>
</tr>
<tr>
<td>UCA then test (c)</td>
<td>Worse outcomes</td>
<td>10</td>
<td>46.82</td>
</tr>
<tr>
<td></td>
<td>No change</td>
<td>20</td>
<td>41.72</td>
</tr>
<tr>
<td></td>
<td>Improved</td>
<td>135</td>
<td>51.39</td>
</tr>
<tr>
<td></td>
<td>F-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How successful was treatment (self-report)</td>
<td>Cured</td>
<td>62</td>
<td>55.83</td>
</tr>
<tr>
<td></td>
<td>Partly cured</td>
<td>55</td>
<td>49.30</td>
</tr>
<tr>
<td></td>
<td>Improved</td>
<td>41</td>
<td>46.95</td>
</tr>
<tr>
<td></td>
<td>Not helped</td>
<td>12</td>
<td>35.11</td>
</tr>
<tr>
<td></td>
<td>F-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information given</td>
<td>Excellent</td>
<td>57</td>
<td>55.83</td>
</tr>
<tr>
<td></td>
<td>Very good</td>
<td>61</td>
<td>50.04</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>36</td>
<td>45.82</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>17</td>
<td>39.11</td>
</tr>
<tr>
<td></td>
<td>F-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age group</td>
<td>31-45 years</td>
<td>31</td>
<td>48.58</td>
</tr>
<tr>
<td></td>
<td>46-60 years</td>
<td>71</td>
<td>50.39</td>
</tr>
<tr>
<td></td>
<td>61-75 years</td>
<td>56</td>
<td>50.33</td>
</tr>
<tr>
<td></td>
<td>75+ years</td>
<td>16</td>
<td>50.34</td>
</tr>
<tr>
<td></td>
<td>F-value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

- a = non-transformed T-scores
- b = ANOVA, all ANOVAs carried out on transformed data. * = p<0.05, ** = p<0.01, *** = p<0.001.
- c = Present – Retrospective scores. See Table 2.
Table 20  Test of responsiveness to patient satisfaction, five measures, T-scores

<table>
<thead>
<tr>
<th>Pooled percentage quartiles</th>
<th>N</th>
<th>Consult</th>
<th>CSQ-18</th>
<th>GUTSS</th>
<th>PSI</th>
<th>SAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>SD</td>
<td>SD</td>
<td>SD</td>
<td>SD</td>
</tr>
<tr>
<td>1 (least satisfied)</td>
<td>42</td>
<td>38.63</td>
<td>8.83</td>
<td>37.62</td>
<td>9.23</td>
<td>38.36</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>47.92</td>
<td>7.10</td>
<td>48.58</td>
<td>5.45</td>
<td>47.68</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>52.78</td>
<td>5.08</td>
<td>53.67</td>
<td>5.18</td>
<td>55.32</td>
</tr>
<tr>
<td>4 (most satisfied)</td>
<td>44</td>
<td>59.13</td>
<td>5.22</td>
<td>59.55</td>
<td>2.19</td>
<td>59.22</td>
</tr>
<tr>
<td>F-values (b)</td>
<td></td>
<td>70.18***</td>
<td>117.65***</td>
<td>112.96***</td>
<td>80.89***</td>
<td>152.88***</td>
</tr>
</tbody>
</table>

Notes:

a = non-transformed T-scores
b = ANOVA, all ANOVAs carried out on transformed data. * = p<0.05, ** = p<0.01, *** = p<0.001.

The limitation of tests is that they are tests against patient satisfaction predictors, not against the patient satisfaction construct itself. An estimate of the construct itself was derived through computing the mean percentage score across all the four patient satisfaction measures on the assumption that between the four instruments the construct was adequately represented. As shown in section 6.2 all four instruments loaded onto a single underlying factor in a principal components analysis suggesting this assumption is supported. The mean percentage scores were then quartiled and these used as the criterion for assessing the sensitivity of the instruments. The results showed that (as expected) all instruments were highly sensitive, but that the SAPS was more sensitive than any of the other measures. The results are given in Table 20.

7.3  A single measure for immediate assessment of patient satisfaction

Six generic items measuring global satisfaction were identified from the instruments and were examined using partial credit IRT following recoding to remove sparse data. The items were:

- PS#A  How happy are you with the effect of your treatment (GUTSS, #1)
- PS#B  How satisfied are you with the outcome of your treatment (GUTSS, #4)
- PS#C  Are you happy with the care you received (GUTSS, #10)
- PS#D  How satisfied are you with the amount of help received (CSQ-18, #4)
- PS#E  In an overall sense, how satisfied are you with the service you have received? (CSQ-18, #13)
- PS#F  I am totally satisfied with my visit to this clinician (Consult SQ, #1)

The Consult SQ #1 item (PS#F: I am totally satisfied with my visit to this clinician) obtained disordered thresholds. It was therefore discarded. All the other items had ordered probability thresholds.

The remaining items were then examined for differential item functioning (DIF). DIF describes the extent to which two or more groups of respondents interpret an item differently (i.e. whether the item has significantly different meaning for the different groups). Good items do not exhibit significant DIF. The criteria used here were:

---

6  PS = Patient Satisfaction.
(a) That there would be statistically different scores on the item by treatment group (Physiotherapy/Physiotherapy & surgery/Surgery) and by the then-test changes in UCA scores (worse/no change/better). This is a test of item sensitivity by known groups;

(b) That there would be no significant difference by DIF across these tests. Different groups should interpret the item similarly, if the item is valid; and that

(c) There would be no interaction between treatment group, the then-test UCA scores and DIF. The results are given in Table 21. This shows that items PS#A, PS#B, PS#D and PS#E were sensitive to group membership (Physiotherapy/Physiotherapy & surgery/Surgery) and also to the UCA then-test (Worse/No change/Better). Item PS#C was therefore discarded. Regarding evidence of DIF, this was apparent for PS#A on the UCA then-test; it was therefore discarded. The only item with a significant interaction between group status and DIF was item PS#E for the UCA then-test and DIF; it was therefore discarded leaving PS#B and PS#D as possible items with apparently very similar properties on the recoded variables.

Table 21  Item response theory analysis of selected global items

<table>
<thead>
<tr>
<th>Item response theory analysis of selected global items</th>
<th>Treatment group analysis</th>
<th>UCA then-test analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A How happy are you with the effect of your treatment</td>
<td>3.65*</td>
<td>0.79</td>
</tr>
<tr>
<td>B How satisfied are you with the outcome of your treatment</td>
<td>5.14**</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>C Are you happy with the care you received</td>
<td>0.17</td>
<td>2.26</td>
</tr>
<tr>
<td>D How satisfied are you with the amount of help received</td>
<td>5.89**</td>
<td>1.69</td>
</tr>
<tr>
<td>E In an overall sense, how satisfied are you with the service you have received</td>
<td>3.98*</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Notes:
All values are ANOVA F-values. * = p<0.05, ** = p<0.01, *** = p<0.001.
A = Test of significant differences between groups
B = Test of significant differences by DIF among groups
C = Test of significant interaction between groups and DIF
The original response categories for these two items (i.e. the unrecoded response categories shown in Table 15) were 5 for item PS#B and 4 for PS#D. A second IRT analysis was run using the unrecoded variables for these two items. The results showed disordered thresholds for PS#D. The probability thresholds are shown in Figures 6 and 7; while the probability thresholds are ordered for PS#B they are disordered for PS#D. It would appear that PS#B is the more robust item and therefore could be considered as a single patient satisfaction item.

In summary, the IRT analysis (Figure 6) and the responsiveness tests (Table 22) suggest that the single patient satisfaction item PS#B has excellent measurement properties.
### Table 22  Responsiveness of the single patient satisfaction item, PS#B

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Category</th>
<th>Fisher Exact Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Physiotherapy/Physio &amp; Surgery/ Surgery</td>
<td>21.00</td>
</tr>
<tr>
<td>UCA then test (a)</td>
<td>Worse/No change/Better</td>
<td>29.35</td>
</tr>
<tr>
<td>How successful was treatment (self-report)</td>
<td>Cured/Partly cured/Improved-not helped (b)</td>
<td>97.42</td>
</tr>
<tr>
<td>Information given</td>
<td>Excellent/Very good/Good-fair (b)</td>
<td>44.12</td>
</tr>
<tr>
<td>Age group</td>
<td>31-45/46-60/61+ years (b)</td>
<td>8.35</td>
</tr>
<tr>
<td>Pooled patient satisfaction quartiles</td>
<td>Lo/Hi dichotomized at 50th percentile (b)</td>
<td>65.86</td>
</tr>
</tbody>
</table>

Notes:

- a = Present – Retrospective scores. See Table 2.
- b = Categories combined due to sparse data.
8 Discussion and Recommendations

This study follows on from the Hawthorne (2006c) review of patient satisfaction instruments which reported there was no ‘stand out’ instrument and that none could be considered truly satisfactory. Four instruments, however, were recommended for further investigation. They were the:

- The CSQ-18 (Client Satisfaction Questionnaire) developed by Larsen et al. (1979);
- The Consult SQ (Consultation Satisfaction Questionnaire) which assesses satisfaction with a consultation with a general practitioner (Baker, 1990);
- The GUTSS (Genito-Urinary Treatment Satisfaction Scale) which was designed for use in incontinence trials (Hawthorne and Harmer, 2000); and
- The PSI (Patient Satisfaction Inventory) which was designed to discriminate between patients with a life-threatening illness (Guyatt et al., 1995).

This study has directly compared these four instruments through a direct head-to-head comparison in a sample of female patients between 6 to 12 months post incontinence treatment. It also reports on changes to the GUTSS to improve its scoring, the development of a short assessment of patient satisfaction scale, the SAPS, and on the properties of several single global satisfaction items.

8.1 Revision of the GUTSS

The GUTSS was revised through removal of the two filter questions. Examination of the original and revised GUTSS showed that this change made no practical difference whatsoever – indeed the psychometric analysis suggested it may have slightly improved the GUTSS (Table 8) – whilst having the benefit of shortening the GUTSS to just 8 items. It is possible that the GUTSS could be revised even further through removal of #6 (Was the attitude/behaviour of the clinicians…), for this item had more monotonic violations than the acceptable standard (Table 8).

When examined by the patient satisfaction predictors, the revised GUTSS was as sensitive as the original GUTSS (Table 9), and the revised GUTSS was used in the rest of this report.

8.2 Construction of a short assessment of patient satisfaction measure, the SAPS

Hawthorne recommended the development of a short, valid and reliable generic patient satisfaction instrument for use in Australian studies (Hawthorne, 2006c). The first steps towards this were taken in this study through an analysis of all the pooled items from the four published patient satisfaction instruments.

The model used for this exercise was based on Donabedian’s argument that patient satisfaction was an outcome of care, particularly interpersonal processes (Donabedian, 1988). Seven dimensions contributing to this were identified by Hawthorne in his review of patient satisfaction instruments (Table 14) (Hawthorne, 2006c). During this modelling exercise, care was taken to ensure representation of all 7 dimensions. An iterative procedure was followed using Mokken analysis and partial credit IRT by which the worst-fitting items were identified and removed at each iteration, subject to maintaining fidelity to the patient satisfaction model. The final model consisted of 7 representative items, one for each of the identified dimensions (Table 16). Although the initial psychometric properties were excellent (Table 17), it should be borne in mind that the SAPS represents a compromise between fidelity to the model of patient satisfaction, the need for a homogenous scale and good item measurement properties. It is not a de novo purpose-designed
measure of patient satisfaction, but is rather a composite measure cobbled together from available items’.

Participant scores on the SAPS were over a wide range of the possible scale (Figure 5). The SAPS proved sensitive when tested by treatment, the UCA then-test, self-reported treatment effectiveness, and information given by the clinician. When tested against a pooled patient satisfaction estimate, the SAPS was more sensitive than any of the other instruments (Table 20).

8.3 **Comparison of five patient satisfaction measures**

The three generic patient satisfaction instruments above, the revised GUTSS and the SAPS were directly compared on several criteria.

Regarding the descriptive systems, the Consult SQ seemed to be mainly measuring the technical skill and relationship with the treating clinician, the CSQ-18 whether the treatment received was appropriate to the needs of the participant, the GUTSS satisfaction with treatment outcomes and processes, the PSI the patient as a consumer of health services and his/her involvement in the decision-making process, and the SAPS provides a broad patient satisfaction perspective.

Item examination revealed that the only instruments on which all the items statistically discriminated (by self-reported treatment success) was the GUTSS and the SAPS. For the PSI and CSQ-18 three items were definitely insensitive, as were 8 for the Consult SQ (Table 11). Despite these findings, the average IRTCs suggested redundancy within these instruments (Table 12); it is possible that the high reliability and unidimensional estimates reflect a considerable level of double-counting. This empirical evidence is consistent with Hawthorne’s observation of the descriptive systems in which he identified this, particularly for the Consult SQ (Hawthorne, 2006c). For the PSI, redundancy was compounded by the presence of ARBS, and it may have been present in the CSQ-18 as well.

The difficulty for the CSQ-18 may be that it probes some areas of patient satisfaction which may not be particularly pertinent in Australia given the nature of the Australian health care system, as previously noted by Hawthorne (2006c). These areas include access to health care services and health care facilities. Additionally, many of the items within the CSQ-18 are presented from a health service consumer perspective. They are about the patient getting the health care he/she wants, for example ‘Considering your particular needs, how appropriate were the services you received?’ (CSQ-18 #5).

The PSI contained many long and convoluted item stems and responses – it is likely this contributed to the presence of ARSB in this instrument (Figure 2). It is also possible that the perspective of the PSI contributed to ARSB. The PSI is written from the perspective of the patient as a consumer of health services who has to cope with a life-threatening illness, for example ‘During the treatment you received (for your incontinence), how satisfied were you with the amount of choice you had in decisions affecting your health care?’). The number of items covering participation in decision-making and that the patient is treated with respect and consideration by the clinician is 15/23 items (Hawthorne, 2006c). If the study participants (Australian females who had received treatment for incontinence) did not think of themselves as health service consumers, their responses to many of the items on the CSQ-18 and PSI may be partly determined by having never thought about patient satisfaction from this viewpoint.

---

7 For example, although #6 (The time with the clinician was not long enough) was retained as a representative item for the dimension Access, it may not adequately represent this dimension. The psychometric evidence suggests that it was the worst fitting item in the SAPS (Table 17).

8 For the SAPS this was by design: non-responsive items were excluded from the item pool from which the SAPS was developed. See section 7.2 for details.
In distinct contrast to the CSQ-18 and PSI, the Consult SQ items imply considerable respect for the treating clinician; there are no items suggesting the health care should meet the patient’s expectations or that the patient should be involved in medical decision-making. Sample items are ‘I will follow this clinician’s advice because I think he/she is absolutely right’, or ‘This clinician told me everything about my treatment’. Another important difference between the other instruments and the Consult SQ is that the Consult SQ items are all Likert items which have known end-aversion effects, whereas the other instruments have either Guttman or a mixture of Guttman and Likert response scales. It is likely that this difference is the cause of the lower mean scores on the Consult SQ (Table 12).

The perspective of the GUTSS is different again. Of the eight items, four form a sub-dimension assessing satisfaction with the impact of treatment on incontinence: there is no mention here of either patient participation or respect for clinicians. The closest the GUTSS gets to these concepts is in relation to assessing the attitude of the clinical staff (‘Was the attitude/behaviour of the ...’) and patient satisfaction with information provided (‘How satisfied are you with the explanations your clinician has given you about the results of your treatment?’).

Because the SAPS covers all seven patient satisfaction dimensions and draws items from the Consult SQ, GUTSS and PSI it has a broader perspective than any of the other individual instruments.

The extent to which these various perspectives in the instrument descriptive systems reflect cultural differences is unknown, although the discussion above suggests that these are definitely present. Given that there are known cultural effects on satisfaction measures emic effects cannot be ruled out as a source of difference (Hawthorne, 2006c). Even with these differences, however, the correlations between the instruments and a principal component analysis suggested that the four measures were assessing the same construct, albeit differently. Furthermore, all instruments were unidimensional as assessed by the Loevinger H. A summary of these differences and similarities is given in Tables 10, 12 and 16.

Regarding instrument responsiveness, the data suggested that most participants in this study reported a high level of satisfaction with their incontinence treatment. This is reflected in the mean scores and the proportion of cases obtaining ceiling scores on all measures, except the Consult SQ (Table 12). At the same time, however, there were very clear differences in instrument responsiveness (Tables 13, 18, 19 and 20). To assist with interpretability, the findings from these tables are summarized in Table 23 using the relative efficiency statistic (Fayers and Machin, 2000). The results show that the least sensitive instrument, overall, was the Consult SQ. This was followed by the PSI, the CSQ-18, and the SAPS, and it was found that the GUTSS was the most sensitive instrument. In explaining these findings, it is pertinent to reflect upon the nature of each instrument.

* The Consult SQ probes the patient’s response to the previous visit to a health service. The items are concerned with the personal relationship between the patient and clinician (Baker, 1990). As such it may be that the broader aspects of health care that contribute to patient satisfaction are not adequately represented. Even within the dimensions covered by the Consult SQ there appear to be problems. The data in Table 11 shows that 8/18 Consult SQ items are insensitive while Table 15 shows that the proportion of cases endorsing the highest level ranged from 56% to just 7%. These two findings suggest that the Consult SQ may be internally inconsistent thus explaining its lack of responsiveness when compared with the other instruments.

* The CSQ-18 is primarily concerned with assessing satisfaction with health services (Larsen et al., 1979). As shown in Table 11, three of the items were insensitive and it had the highest proportion of respondents endorsing the highest level (58%); the range of respondents endorsing the highest level was from 72% to 31% (Table 15). As with the Consult SQ, these findings are suggestive that some of the items are insensitive and that the internal

---

9 This helps to explain why the SAPS contains two highly sensitive items from the Consult SQ.
structure of the instrument may be inconsistent – despite the obvious double-counting within
the instrument. The CSQ-18 was far more sensitive than the Consult SQ and the PSI,
particularly on the UCA then-test.

- The reason for the very high sensitivity of the GUTSS is that it comprises two factors, one of
  which (4 items) assesses satisfaction with treatment outcomes (Hawthorne and Harmer,
  2000). As shown in Table 21 the GUTSS performed extremely well on tests involving
  outcomes – when compared with the UCA then-test, for example, it was 124 times more
  sensitive than the Consult SQ – but it performed slightly less well on the test of patient
  satisfaction. The responsiveness of the GUTSS therefore depends on the assumption that
  favourable outcomes are strongly predictive of satisfaction. As shown in Hawthorne’s review
  this assumption should not be uncritically accepted (Hawthorne, 2006c).

- The PSI was designed to assess patient satisfaction with medical care where the patients had
  a life-threatening illness (Guyatt et al., 1995). Although the PSI has more items, more item
  response levels and the highest average item correlation than any of the other instruments
  (Table 12), the evidence suggests that, overall; it was the second least responsive instrument.
  This observation is surprising given that 21/23 items were sensitive to self-reported treatment
  effects (Table 11). There are two possible reasons. Table 15 shows that the proportion of
  cases endorsing the highest level ranged from 73% to 23%, and that the PSI also had the
  highest standard deviation of any instrument (Table 12); these two findings suggest an
  inconsistent internal structure in this study population. The second reason may be to do with
  the presence of ARSB among the items, thus reducing the variance between groups (i.e. by
  making the item responses too similar to each other) and thus causing artificially high scale
  psychometric properties (Table 12) but simultaneously reducing its sensitivity.

Table 23 Comparative sensitivity of patient satisfaction instruments, relative efficiency
estimates

<table>
<thead>
<tr>
<th>Comparative sensitivity of patient satisfaction instruments, relative efficiency estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consult SQ</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>UCA then test</td>
</tr>
<tr>
<td>How successful was treatment (self-report)</td>
</tr>
<tr>
<td>Information given</td>
</tr>
<tr>
<td>Age group</td>
</tr>
<tr>
<td>Pooled patient satisfaction quartiles</td>
</tr>
</tbody>
</table>

Notes:
Source: the relative efficiency estimates are based on the F-values in Tables 13 and 19.

- The SAPS performed extremely well on all tests in Tables 18 and 19 – it discriminated as
  expected and failed to discriminate where appropriate. Overall, it was the second most
  sensitive instrument behind the GUTSS and proved the most sensitive measure on the patient
  satisfaction test (Table 23). Given the conscious effort made during construction to ensure
  fidelity to all dimensions of the model of patient satisfaction, these results were excellent.

8.4 Measuring patient satisfaction with a single item

Five global patient satisfaction items were identified from the various instruments and were
examined using partial credit IRT. Initial statistics suggested two items were to be preferred, and
sub-group examination revealed that one of these suffered DIF (Figures 15 and 16). The
remaining item, PS#B, was subjected to the same responsiveness tests as the patient satisfaction
instruments had been: it was found to be sensitive (Table 22).
In summary, the IRT analysis and responsiveness tests suggested that the single patient satisfaction item PS#B has excellent measurement properties and that it could be used with confidence.

8.5 **Suggested text revisions to the GUTSS, SAPS and the single item PS#B**

As discussed throughout this report there were language and response category issues with many of the items from the different instruments. Although this report has been highly statistical in nature, it is possible that language and response category differences may have played an important part in how the participants responded to the questionnaire. There are clear implications for the recommended instruments.

In order to ensure consistency of language and response categories, the items in the revised GUTSS, the SAPS and the single patient satisfaction item PS#B were revised. These revisions involved:

1. Simplifying the language in the item stems to remove unnecessarily complex descriptions and to simplify key descriptors (e.g. ‘doctor/other health professional’)
2. Making the voice consistent throughout all items, thus using ‘you’ instead of ‘I’ or ‘my’.
3. Using 5-point response scales for all items to avoid exceeding channel capacity.
4. Re-ordering the response scales so that the SAPS has a mixture of positive and negative items.

There are implications under #3 for the scoring of the GUTSS and the SAPS.

- For the GUTSS the removal of the two filter items and their replacement changed the overall GUTSS score by reducing the maximum possible score from 34 to 32. It should be noted that throughout this report the revised GUTSS was used and that the maximum score reported was 32, unless otherwise stated.

- For the SAPS replacing the 7-point response scales for items #4 and #5 with 5-point scales has changed the maximum scoring range from 39 to 35. It is important to note that throughout this report the SAPS scores are on the original scale, i.e. the maximum score for the SAPS in this report was 39.

The instruments with these text revisions can be found in Appendices 1, 2 and 3. The original items can be found in Appendix 4. Table 24 provides the key for matching the revisions with the original items.
Table 24  GUTTS, SAPS and patient satisfaction items and their original source

<table>
<thead>
<tr>
<th>GUTTS, SAPS and patient satisfaction items and their original source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>GUTTS</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>SAPS</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>Single patient satisfaction item PS#B</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

8.6  Study limitations

This study compared different patient satisfaction instruments in a sample of Australian women who had received treatment for their urinary incontinence. The main threat to the study findings is in relation to the study sample.

Because of the study constraints, sampling was limited to the two largest Australian cities; whether patient satisfaction varies by location across Australia is unknown. Sampling was also limited to those who had treatment for urinary incontinence. Whether similar findings would be found among those who had treatment for faecal incontinence is unknown. In the sample itself, there was gender bias because the participating clinics mostly treated females (e.g. in the Melbourne the Royal Women’s Hospital participated). Regarding the gender balance in study participants, the percentage of females was 97% and because of this, males were excluded from the data analysis.

The crude response rate was 44%, which was lower than expected. Because of the methodology the number of patients out of scope is unknown: no questionnaires were returned to the researchers because of unknown addresses. Although lower than expected, this response rate is consistent with the literature for postal surveys in health studies where there are no follow-up
telephone calls eliciting participation (Fowler et al., 2002). Follow-up telephone calls could not be made since the researchers had no knowledge of patients’ names, addresses or other contact details. There were implications for the data analysis arising from this response rate which materially affected the study findings and it may be that with a higher response rate different findings would be apparent (e.g. consider the situation where the respondents were mainly those who were satisfied and that those who were dissatisfied did not participate). Finally, there are potential issues around the demographic profile of those who participated. Given Australia’s population, there was a disproportionate number of Australian-born participants (81%), suggesting that the results may not be generalizable to immigrants.

These limitations suggest that the study findings, which reflect the preferences of Australian-born females who had treatment for urinary incontinence, should be interpreted cautiously.

### 8.7 Recommendations

This study was conducted to inform the NCMS about the possible inclusion of patient satisfaction measures in the proposed national suite of continence measures to be used by clinicians and researchers in Australia (Thomas et al., 2006). It specifically addressed the recommendations of the Hawthorne (2006c) review of patient satisfaction instruments which were to:

- Report on the feasibility of drawing a single item patient satisfaction measure from the continence-specific or generic patient satisfaction instruments;
- Revise the GUTSS continence-specific instrument through removal of the two filter items; and to
- Develop a short, valid and reliable generic patient satisfaction measure.

As reported in the body of this work all three of these objectives were achieved, subject to the study limitations. The findings from this study give rise to the following recommendations.

1. Several single item patient satisfaction measures were examined. These were drawn from the four instruments tested in this study. Of these, 1 item, PS#B, was found to have excellent measurement properties.

   - It is recommended that where a single item of patient satisfaction is required, such as in clinical consultations where immediate feedback is required, the PS#B is used. A copy of this item can be found in Appendix 3.

2. The revision of the GUTSS has simplified its scoring and removed two of the item, thus shortening it. This revision has had no discernible impact on its psychometric properties; if anything it has improved its measurement.

   - It is recommended that where a continence specific patient satisfaction measure is required, such as in clinical research, the revised GUTSS is used, subject to recommendation #4 below. A copy of the GUTSS can be found in Appendix 1.

3. Three generic patient satisfaction instruments were examined in this study. The findings suggest that none are particularly suitable for use in an Australian setting, as determined by detailed inspection of their descriptive systems and assessment of their psychometric properties. Drawing on items from these three instruments and the GUTSS, a short generic patient satisfaction instrument was constructed, the SAPS. Evaluation of the SAPS suggests that it possesses a superior descriptive system than any of the four tested instruments (including the GUTSS) in that it has items covering all dimensions of patient satisfaction reported in the literature. Assessment of its psychometric properties suggests a nomological net of validity evidence supporting it, and that it is reliable.

   - It is recommended that where a short generic patient satisfaction instrument is needed, as in epidemiological or clinical research, the SAPS is used. A copy of the SAPS, with revisions to give the items a common feel, can be found in Appendix 2.
4. There are descriptive system limitations with the GUTSS in that half of the items measure satisfaction with treatment outcomes rather than satisfaction with health care and the interpersonal processes at the heart of care. As such the validity of the GUTSS is open to question, despite its excellent psychometric properties. Given that the correlation between the GUTSS and the SAPS is 0.83 (section 7.2), that the SAPS has greater descriptive system validity, that it is shorter and that it has excellent psychometric properties in its own right, there is a prima facie case that the SAPS is used in preference to the GUTSS. Furthermore, use of the SAPS, the items of which are generic, would enable direct comparison across different disease interventions thus considerably enhancing understanding of patient satisfaction. The only reason for not making an outright recommendation that the SAPS should be the first instrument of choice wherever an assessment of patient satisfaction is required is that the measurement properties of the text revised SAPS (Appendix 2) need to be field tested.

- It is recommended that the text revision SAPS is tested in field settings.
9 References


Donabedian A (1980) *Explorations in Quality Assessment and Monitoring: Vol 1 The Definition of Quality and the Approaches to its Assessment*. Health Administration Press, MI.


GraphPad (2003) GraphPad Software, San Diego.


Guttman L (1944) *A basis for scaling qualitative data.* American Sociological Review. Vol.9, No.2, pp.139-150.


Appendix 1: Copy of the Genito-Urinary Treatment Satisfaction Scale

Instructions: After reading each question, circle the answer that best describes your situation. We know that sometimes answers may not describe you exactly, so please pick the answer that most closely describes you.

When you have finished, please check that you have answered all questions.

Q1. How happy are you with the effect of your treatment?
   Very happy ......................................................... 0
   Happy ................................................................. 1
   Neither happy nor unhappy ............................... 2
   Unhappy ............................................................. 3
   Very unhappy .................................................... 4

Q2. Over the past 4 weeks do you still have problems with incontinence?
   Yes, extreme problems ...................................... 0
   Very much so...................................................... 1
   Some problems .................................................. 2
   Slight problems .................................................. 3
   No problems at all............................................... 4

Q3. How satisfied are you with the outcome of your treatment?
   Very satisfied ..................................................... 0
   Satisfied ............................................................ 1
   Neither satisfied nor dissatisfied ........................... 2
   Dissatisfied ......................................................... 3
   Very dissatisfied ................................................ 4

Q4. During the past 4 weeks have you been disappointed with the outcome of your treatment?
   Extremely disappointed ..................................... 0
   Very disappointed ............................................... 1
   Disappointed ...................................................... 2
   Slightly disappointed ......................................... 3
   Not at all disappointed ........................................ 4

Q5. Before you had the treatment was the information from your doctor or other health professional about the treatment...
   Very good .......................................................... 0
   Good ................................................................. 1
   Fair ................................................................. 2
   Poor ................................................................. 3
   Very poor .......................................................... 4

Q6. Was the attitude/behaviour of the doctor or other health professional...
   Very good .......................................................... 0
   Good ................................................................. 1
   Fair ................................................................. 2
   Poor ................................................................. 3
   Very poor .......................................................... 4

Q7. How satisfied are you with the explanations your doctor or other health professional has given you about the results of your treatment?
   Very satisfied ..................................................... 0
Satisfied .............................................................1
Neither satisfied nor dissatisfied ......................2
Dissatisfied ..........................................................3
Very dissatisfied ..................................................4

Q8. Are you happy with the care you received in the [hospital/clinic]?
   Very unhappy ..................................................0
   Unhappy ..........................................................1
   Neither happy nor unhappy .....................2
   Happy ............................................................3
   Very happy .....................................................4

PLEASE CHECK THAT YOU HAVE ANSWERED ALL THE QUESTIONS.
THANK YOU VERY MUCH FOR YOUR TIME IN COMPLETING THIS QUESTIONNAIRE.

Scoring the GUTSS:
1. Reverse the scores for #1, #3, #5, #6, #7.
2. Sum all scores. The score range is from 0 (extremely dissatisfied) to 32 (extremely satisfied).
Appendix 2: Copy of the SAPS (Revised Items)\textsuperscript{10}

Instructions: After reading each question, circle the answer that best describes your situation. We know that sometimes answers may not describe you exactly, so please pick the answer that most closely describes you.

When you have finished, please check that you have answered all questions.

Q1. How happy are you with the effect of your treatment?
   Very happy ................................................... 0
   Happy .......................................................... 1
   Neither happy nor unhappy .......................... 2
   Unhappy ...................................................... 3
   Very unhappy .............................................. 4

Q2. How satisfied are you with the explanations the {doctor/other health professional} has given you about the results of your treatment?
   Very dissatisfied .......................................... 0
   Dissatisfied ................................................ 1
   Neither satisfied nor dissatisfied .................. 2
   Satisfied .................................................... 3
   Very satisfied ............................................. 4

Q3. The {doctor/other health professional} was very careful to check everything when examining you.
   Strongly agree ........................................... 0
   Agree ........................................................ 1
   Not sure .................................................... 2
   Disagree .................................................... 3
   Strongly disagree ........................................ 4

Q4. How satisfied were you with the choices you had in decisions affecting your health care?
   Very dissatisfied .......................................... 0
   Dissatisfied ................................................ 1
   Neither satisfied nor dissatisfied .................. 2
   Satisfied .................................................... 3
   Very satisfied ............................................. 4

Q5. How much of the time did you feel respected by the {doctor/other health professional}?
   All of the time ............................................. 0
   Most of the time ........................................... 1
   About half the time ..................................... 2
   Some of the time ......................................... 3
   None of the time .......................................... 4

Q6. The time you had with the {doctor/other health professional} was not long enough.
   Strongly agree ........................................... 0
   Agree ........................................................ 1
   Not sure .................................................... 2
   Disagree .................................................... 3

\textsuperscript{10} These items are the suggested items for the SAPS after text revision. The original items can be found in Appendix 3.
Q7. Are you happy with the care you received in the {hospital/clinic}?

- Very happy .................................................. 0
- Happy .......................................................... 1
- Neither happy nor unhappy. ............................. 2
- Unhappy ....................................................... 3
- Very unhappy ............................................... 4

Scoring the SAPS:
1. Reverse the scores for #1, #3, #5, #7.
2. Sum all scores. The score range is from 0 (extremely dissatisfied) to 28 (extremely satisfied).
Appendix 3: Copy of the Recommended Single Patient Satisfaction Item PS#B

PS#B. How satisfied are you with the outcome of your treatment?

- Very dissatisfied: ................................................ 1
- Dissatisfied: ...................................................... 2
- Neither satisfied nor dissatisfied: .......................... 3
- Satisfied: .......................................................... 4
- Very satisfied: .................................................... 5
Appendix 4: Copy of the Study Questionnaire

Key to instruments and items:

<table>
<thead>
<tr>
<th>Item numbers</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.2.1 to A.2.2</td>
<td>ISI retrospective incontinence assessment</td>
</tr>
<tr>
<td>A.3.1 to A.3.6</td>
<td>UDI retrospective incontinence assessment</td>
</tr>
<tr>
<td>A.11.1 to A.11.2</td>
<td>ISI current incontinence assessment</td>
</tr>
<tr>
<td>A.12.1 to A.12.6</td>
<td>UDI current incontinence assessment</td>
</tr>
<tr>
<td>C.1 to C.10</td>
<td>GUTSS original items</td>
</tr>
<tr>
<td>C.11 to C.28</td>
<td>CSQ-18 items</td>
</tr>
<tr>
<td>C.29 to C.46</td>
<td>Consult SQ items</td>
</tr>
<tr>
<td>C.47 to C.48</td>
<td>GUTSS revised items</td>
</tr>
<tr>
<td>C.49 to C.71</td>
<td>PSI items</td>
</tr>
<tr>
<td>D.1 to D.15</td>
<td>AQoL quality of life items</td>
</tr>
<tr>
<td>D.16 to D.21</td>
<td>Friendship Scale social isolation items</td>
</tr>
<tr>
<td>D.22 to D.26</td>
<td>WHO5 wellbeing items</td>
</tr>
</tbody>
</table>
This questionnaire asks about your satisfaction with your recent treatment for incontinence. It is part of a study examining ways in which the treatments for incontinence can be improved. This questionnaire is 22 pages long, and it will take you about 45 minutes to complete.

- Please read each question carefully, and then tick or fill in the circle next to the answer which best describes you e.g. the circle like this ○, becomes this ● or ✓.

- Remember, only select one circle for each question to fill in,

- If you make a mistake, just cross out your answer and select the answer that best describes you. There are no right or wrong answers.

- After you have answered all the questions, please post the questionnaire back with the consent form using the enclosed reply paid envelope.

Thank you for your help.

Assoc.Prof. Graeme Hawthorne
Principal Research Fellow
Department of Psychiatry
University of Melbourne

Ms Laura Hayes
Research Fellow
Department of Psychiatry
University of Melbourne
Part A: Some questions concerning your treatment for incontinence

A.1 What type of incontinence did you seek treatment for?
   - Urinary
   - Faecal
   - Both

This part asks about your incontinence BEFORE YOUR TREATMENT

A.2 Please think back to BEFORE you had your treatment.

A.2.1 How often did you experience urine leakage?
   - Never
   - Less than once a month
   - Several times a month
   - Several times a week
   - Every day or night

A.2.2 How much urine did you lose each time?
   - None, I did not leak urine at all
   - Drops
   - Small splashes
   - More

A.3 Please think back to BEFORE you had your treatment.

Did you experience, and if so, how much were you bothered by:

A.3.1 Frequent urination
   - Not at all
   - Slightly
   - Moderately
   - Greatly

A.3.2 Urine leakage related to the feeling of urgency (sudden desire to urinate)?
   - Not at all
   - Slightly
   - Moderately
   - Greatly

A.3.3 Urine leakage related to physical activity, coughing, or sneezing?
   - Not at all
   - Slightly
   - Moderately
   - Greatly

A.3.4 Small amounts of urine leakage (drops)?
   - Not at all
   - Slightly
   - Moderately
   - Greatly
A.3.5 Difficulty emptying your bladder?
- Not at all
- Slightly
- Moderately
- Greatly

A.3.6 Pain or discomfort in the lower abdominal or genital area?
- Not at all
- Slightly
- Moderately
- Greatly

This part asks ABOUT YOUR TREATMENT

A.4 What month and year did you first seek treatment for your incontinence?

__________ Month ____________ Year

A.5 When you sought treatment, what did you expect to be the outcome?
- The incontinence would be cured, i.e. that there would be no incontinence
- The incontinence would be partly cured, enough so it no longer bothered you in everyday life
- That there would be some improvement in your incontinence
- The incontinence would not be helped at all

A.6 Which clinic did you have your treatment at:
- St George Hospital (Sydney)
- Royal Womens' Hospital (Melbourne)

A.7 Before you had the treatment, was the information from your clinician (doctor, physiotherapist or nurse) about the treatment:
- Excellent
- Very good
- Good
- Fair
- Poor
- Very poor

A.8 What was the main treatment you had for your incontinence?
- Physiotherapy
- Surgery
- Physiotherapy, then surgery
- Surgery, then physiotherapy
- Other, please describe: ___________________________________________

A.9 In what month and year did you finish your treatment? __________ Month ________ Year

A.10 How successful was your treatment?
- The incontinence was cured, i.e. you no longer have any incontinence
- The incontinence was partly cured, enough so it no longer bothers you in everyday life
- There was some improvement in your incontinence
The incontinence was not helped at all; i.e. the treatment failed to help you

This part asks about your incontinence AFTER YOUR TREATMENT

A.11 Please think about your incontinence NOW.

A.11.1 How often do you experience urine leakage?
- Never
- Less than once a month
- Several times a month
- Several times a week
- Every day or night

A.11.2 How much urine do you lose each time?
- None, I do not leak urine at all
- Drops
- Small splashes
- More

A.12 Please think about your incontinence NOW.

Do you experience, and if so, how much are you bothered by:

A.12.1 Frequent urination
- Not at all
- Slightly
- Moderately
- Greatly

A.12.2 Urine leakage related to the feeling of urgency (sudden desire to urinate)?
- Not at all
- Slightly
- Moderately
- Greatly

A.12.3 Urine leakage related to physical activity, coughing, or sneezing?
- Not at all
- Slightly
- Moderately
- Greatly

A.12.4 Small amounts of urine leakage (drops)?
- Not at all
- Slightly
- Moderately
- Greatly

A.12.5 Difficulty emptying your bladder?
- Not at all
- Slightly
- Moderately
- Greatly
A.12.6 Pain or discomfort in the lower abdominal or genital area?

- Not at all
- Slightly
- Moderately
- Greatly
Part B: Some questions about your background

To assist us with the study we need to know a little about your background. This will help us to better understand your answers to the other questions. The information you give is personal, and will remain so. To protect your privacy please do not write your name on this questionnaire.

B.1 You are a:
   ○ Male
   ○ Female

B.2 Your age is: ____________ years.

B.3 In which country were you born? __________________

B.4 Your relationship status is:
   ○ Never married
   ○ Married/de facto
   ○ Divorced
   ○ Separated
   ○ Widowed

B.5 What is your highest completed education level?
   ○ Primary school
   ○ Trade/Apprenticeship Certificate
   ○ High school
   ○ Technical & Further education qualification
   ○ University/College degree

B.6 What is your current working status?
   ○ Working full-time or part-time
   ○ Unemployed, or looking for work
   ○ Homemaker
   ○ Student
   ○ Retired or on sickness benefits
Part C:  Satisfaction with the health care you received for your incontinence

This part asks you a number of questions about how satisfied you feel with the health care you received for your incontinence.

- Please read each question carefully, and then tick the circle which best describes you.
- Remember, only select one circle for each question.
- If you make a mistake, just cross out your answer and select the answer that best describes you. There are no right or wrong answers.

C.1  How happy are you with the effect of your treatment?
- Very happy
- Happy
- Neither happy nor unhappy
- Unhappy
- Very unhappy

C.2  Since your treatment do you still have problems with incontinence?
- Yes, you still have problems (Go to Question C.3)
- No, you have no problems (Go to Question C.4).

C.3  Over the past 4 weeks have these problems been...........
- Extremely problematic
- Very problematic
- Problematic
- Slightly problematic
- Not problematic

C.4  How satisfied are you with the outcome of your treatment?
- Very satisfied
- Satisfied.
- Neither satisfied nor dissatisfied
- Dissatisfied
- Very dissatisfied

C.5  Over the past 4 weeks have you been disappointed with the outcome of the treatment?
- Yes, you have been disappointed (Go to Question C.6)
- No, you have not been disappointed (Go to Question C.7).

C.6  How disappointed are you?
- Extremely disappointed
- Very disappointed
- Disappointed
- Slightly disappointed
- Not disappointed

C.7  Before you had the treatment was the information from your clinician (doctor/physiotherapist or nurse) about the treatment........
- Very good
- Good
- Fair
C.8 Was the attitude/behaviour of the doctors, physiotherapists or nurses at the hospital or clinic........
- Very good
- Good
- Fair
- Poor
- Very poor

C.9 How satisfied are you with the explanations your clinician (doctor or physiotherapist or nurse) has given you about the results of your treatment?
- Very satisfied
- Satisfied
- Neither satisfied nor dissatisfied
- Dissatisfied
- Very dissatisfied

C.10 Are you happy with the care you received in the hospital or clinic?
- Very happy
- Happy
- Neither happy nor unhappy.
- Unhappy
- Very unhappy

C.11 When you first attended hospital for treatment, were you seen as promptly as you felt necessary?
- Yes, very promptly
- Yes, promptly
- No, there was some delay
- No, it seemed to take forever

C.12 In general, how satisfied are you with the comfort and attractiveness of the clinic you attended?
- Quite dissatisfied
- Indifferent, or mildly dissatisfied
- Mostly satisfied
- Very satisfied

C.13 Did the characteristics of the clinic building detract from the services you received?
- Yes, they detracted very much
- Yes, they detracted somewhat
- No, they did not detract much
- No, they did not detract at all

C.14 How satisfied are you with the amount of help you have received?
- Quite dissatisfied
- Indifferent, or mildly satisfied
- Mostly satisfied
C.15 Considering your particular needs, how appropriate were the services you received?
- Very satisfied
- Highly appropriate
- Generally appropriate
- Generally inappropriate
- Highly inappropriate

C.16 Have the services you received helped you to deal more effectively with your problems?
- Very satisfied
- Highly appropriate
- Generally appropriate
- Generally inappropriate
- Highly inappropriate

C.17 When you talked to the clinician (doctor or physiotherapist or nurse) with whom you have worked most closely, how closely did he or she listen to you?
- Not at all closely
- Not too closely
- Fairly closely
- Very closely

C.18 Did you get the kind of service you wanted?
- Yes, definitely not
- Yes, not really
- Yes, generally
- Yes, definitely

C.19 Are there other services you need, but have not received?
- Yes, there definitely were
- Yes, I think there were
- No, I don't think there were
- No, there definitely were not

C.20 How clearly did the clinician (doctor, physiotherapist or nurse) with whom you worked most closely understand your problem and how you felt about it?
- Very clearly
- Clearly
- Somewhat clearly
- Very unclearly

C.21 How competent and knowledgeable was the clinician (doctor, physiotherapist or nurse) with whom you have worked closely?
- Poor abilities at best
- Only of average ability
- Competent and knowledgeable
- Highly competent and knowledgeable

C.22 How would you rate the quality of the service you have received?
- Excellent
- Good
C.23 In an overall general sense, how satisfied are you with the service you have received?
- Fair
- Poor

C.24 If a friend were in need of similar help, would you recommend the clinic you used to him or her?
- No, definitely not
- No, I don't think so
- Yes, I think so
- Yes, definitely

C.25 Did the people at the clinic you used generally understand the kind of help you wanted?
- No, they misunderstood almost completely
- No, they seemed to misunderstand
- Yes, they seemed to generally understand
- Yes, they understood almost perfectly

C.26 To what extent did the clinic's treatment meet your needs?
- Almost all of my needs have been met
- Most of my needs have been met
- Only a few of my needs have been met
- None of my needs have been met

C.27 Have your rights as an individual been respected?
- No, almost never respected
- No, sometimes not respected
- Yes, generally respected
- Yes, almost always respected

C.28 If you were to seek help again, would you go back to the same clinic?
- Definitely not
- No, I don't think so
- Yes, I think so
- Yes, definitely

C.29 I am totally satisfied with my visit to this clinician (doctor, physiotherapist or nurse)
- Strongly agree
- Agree
- Not sure
- Disagree
- Strongly disagree

C.30 This clinician (doctor, physiotherapist or nurse) was very careful to check everything when examining me
C.31 I will follow this clinician’s (doctor, physiotherapist or nurse) advice because I think he/she is absolutely right
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.32 I felt able to tell this clinician (doctor, physiotherapist or nurse) about very personal things
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.33 The time I was able to spend with the clinician (doctor, physiotherapist or nurse) was a bit too short
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.34 This clinician (doctor, physiotherapist or nurse) told me everything about my treatment
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.35 Some things about my consultation with the clinician (doctor, physiotherapist or nurse) could have been better
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.36 There are some things this clinician (doctor, physiotherapist or nurse) does not know about me
   ○ Strongly agree
   ○ Agree
C.37 This clinician (doctor, physiotherapist or nurse) examined me very thoroughly
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.38 I thought this clinician (doctor, physiotherapist or nurse) took notice of me as a person
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.39 The time I was allowed to spend with the clinician (doctor, physiotherapist or nurse) was not long enough to deal with everything I wanted
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.40 I understand my condition much better after seeing this clinician (doctor, physiotherapist or nurse)
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.41 The clinician (doctor, physiotherapist or nurse) was interested in me as a person, and not just my illness
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.42 This clinician (doctor, physiotherapist or nurse) knows all about me
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree
C.43 This clinician (doctor, physiotherapist or nurse) really knew what I was thinking
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.44 I wish it had been possible to spend a little longer with the clinician (doctor, physiotherapist or nurse)
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.45 I am not completely satisfied with my visit to the clinician (doctor, physiotherapist or nurse)
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.46 I would find it difficult to tell this clinician (doctor, physiotherapist or nurse) about some private things
   ○ Strongly agree
   ○ Agree
   ○ Not sure
   ○ Disagree
   ○ Strongly disagree

C.47 Over the past 4 weeks, do you still have problems with incontinence?
   ○ Yes, extreme problems
   ○ Very much so
   ○ Some problems
   ○ Slight problems
   ○ No problems at all

C.48 During the past 4 weeks have you been disappointed with the outcome of your treatment?
   ○ Extremely disappointed
   ○ Very disappointed
   ○ Disappointed
   ○ Slightly disappointed
   ○ Not at all disappointed

C.49 People sometimes feel that they have gone through a lot when receiving health care. Do you agree that you have gone through a lot in relation to the health care you have received for your incontinence?
   ○ Strongly agree that I have been through a lot
   ○ Moderately agree
C.50 In terms of the health care you received for your incontinence, do you agree you went through more than you expected?
- Strongly agree that I went through more than expected
- Mildly agree
- Not sure
- Very mildly disagree
- Mildly disagree
- Moderately disagree
- Strongly disagree

C.51 How often during the treatment you received for your incontinence did you feel that your medical problems were not explained in a way you understood?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.52 How often during the treatment you received for your incontinence, did you feel helpless in regard to decision-making related to your health care?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.53 When you think of the health care you have received for your incontinence, how often did you feel out of control of your situation?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.54 How often during the treatment you received for your incontinence did you feel that decisions were made by the health care providers without involving you as much as you would have liked?
- All of the time
C.55 How often during the treatment you received for your incontinence did you feel that you were not firm enough about what you wanted in relation to your health care?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.56 How often during the treatment you received for your incontinence did you feel overwhelmed with decisions to be made about your health care?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.57 How often during the treatment you received for your incontinence did you feel upset because you did not understand what was going on with your health care?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.58 How often during the treatment you received for your incontinence did you feel that you were involved too late in decisions related to your health care?
- All of the time
- A good bit of the time
- Some of the time
- A little of the time
- Hardly any of the time
- Almost none of the time
- None of the time

C.59 During the treatment you received for your incontinence, how satisfied were you with the way the options regarding your health care were explained to you?
- Not really satisfied
Generally satisfied, but significant areas of dissatisfaction
Generally satisfied, but minor areas of dissatisfaction
Quite satisfied
Mostly satisfied
Very satisfied
Completely satisfied

C.60 During the treatment you received for your incontinence, how satisfied were you with the co-operation you received from the health care providers?
Not really satisfied
Generally satisfied, but significant areas of dissatisfaction
Generally satisfied, but minor areas of dissatisfaction
Quite satisfied
Mostly satisfied
Very satisfied
Completely satisfied

C.61 Some people want decisions related to their health care made by the physician who they think knows what is best for them. Other people want to be involved in the decisions related to their health care. Based on your experiences during the treatment you received for your incontinence, how satisfied were you with the way decisions related to your health care were made?
Not really satisfied
Generally satisfied, but significant areas of dissatisfaction
Generally satisfied, but minor areas of dissatisfaction
Quite satisfied
Mostly satisfied
Very satisfied
Completely satisfied

C.62 During the treatment you received for your incontinence, how satisfied were you with your family’s involvement in your health care?
Not really satisfied
Generally satisfied, but significant areas of dissatisfaction
Generally satisfied, but minor areas of dissatisfaction
Quite satisfied
Mostly satisfied
Very satisfied
Completely satisfied

C.63 Some people want everything possible done for their health problems while others do not want a lot done. How satisfied were you with the level of health care you received during the treatment you received for your incontinence?
Not really satisfied
Generally satisfied, but significant areas of dissatisfaction
Generally satisfied, but minor areas of dissatisfaction
Quite satisfied
Mostly satisfied
Very satisfied
Completely satisfied
C.64 During the treatment you received for your incontinence, how satisfied were you with the amount of choice you had in decisions affecting your health care?
- Not really satisfied
- Generally satisfied, but significant areas of dissatisfaction
- Generally satisfied, but minor areas of dissatisfaction
- Quite satisfied
- Mostly satisfied
- Very satisfied
- Completely satisfied

C.65 How often during the treatment you received for your incontinence, did you feel understood by the health care providers?
- None of the time
- A little of the time
- Some of the time
- A fair bit of the time
- Most of the time
- Almost all of the time
- All of the time

C.66 How often during the treatment you received for your incontinence, did you feel that you and the health care providers shared the same goals in relation to your health care?
- None of the time
- A little of the time
- Some of the time
- A fair bit of the time
- Most of the time
- Almost all of the time
- All of the time

C.67 How often during the treatment you received for your incontinence, did the health care providers clarify your wishes?
- None of the time
- A little of the time
- Some of the time
- A fair bit of the time
- Most of the time
- Almost all of the time
- All of the time

C.68 Based on the health care you received for your incontinence, how much of the time did you feel respected by the health care providers?
- None of the time
- A little of the time
- Some of the time
- A fair bit of the time
- Most of the time
- Almost all of the time
C.69  How often during the treatment you received for your incontinence, did you feel that the health care providers cared about you?
  ○  None of the time
  ○  A little of the time
  ○  Some of the time
  ○  A fair bit of the time
  ○  Most of the time
  ○  Almost all of the time
  ○  All of the time

C.70  How often during the treatment you received for your incontinence, were you able to understand what the health care provider said to you?
  ○  None of the time
  ○  A little of the time
  ○  Some of the time
  ○  A fair bit of the time
  ○  Most of the time
  ○  Almost all of the time
  ○  All of the time

C.71  During the treatment you received for your incontinence, how comfortable did you feel with your involvement in making decisions related to your health care?
  ○  Extremely uncomfortable
  ○  Very uncomfortable
  ○  Quite uncomfortable
  ○  Somewhat uncomfortable
  ○  Quite comfortable
  ○  Very comfortable
  ○  Extremely comfortable
Part D: Some questions about other aspects of your life

The questions in this section ask about your quality of life. Your answers will help us to understand the impact of incontinence and its treatment on peoples’ lives.

For these questions please circle the alternative that best describes you during the last week.

D.1 Concerning my use of prescribed medicines:
- I do not or rarely use any medicines at all.
- I use one or two medicinal drugs regularly.
- I need to use three or four medicinal drugs regularly.
- I use five or more medicinal drugs regularly.

D.2 To what extent do I rely on medicines or a medical aid? (NOT glasses or a hearing aid.) (For example: walking frame, wheelchair, prosthesis etc.)
- I do not use any medicines and/or medical aids.
- I occasionally use medicines and/or medical aids.
- I regularly use medicines and/or medical aids.
- I have to constantly take medicines or use a medical aid.

D.3 Do I need regular medical treatment from a doctor or other health professional?
- I do not need regular medical treatment.
- Although I have some regular medical treatment, I am not dependent on this.
- I am dependent on having regular medical treatment.
- My life is dependent upon regular medical treatment.

D.4 Do I need any help looking after myself?
- I need no help at all.
- Occasionally I need some help with personal care tasks.
- I need help with the more difficult personal care tasks.
- I need daily help with most or all personal care tasks.

D.5 When doing household tasks: (For example, preparing food, gardening, using the video recorder, radio, telephone or washing the car)
- I need no help at all.
- Occasionally I need some help with household tasks.
- I need help with the more difficult household tasks.
- I need daily help with most or all household tasks.

D.6 Thinking about how easily I can get around my home and community:
- I get around my home and community by myself without any difficulty.
- I find it difficult to get around my home and community by myself.
- I cannot get around the community by myself, but I can get around my home with some difficulty.
- I cannot get around either the community or my home by myself.

D.7 Because of my health, my relationships (eg: with my friends, partner or parents) generally:
- Are very close and warm.
- Are sometimes close and warm.
- Are seldom close and warm.
- I have no close and warm relationships.
D.8 Thinking about my relationship with other people:
- I have plenty of friends, and am never lonely.
- Although I have friends, I am occasionally lonely.
- I have some friends, but am often lonely for company.
- I am socially isolated and feel lonely.

D.9 Thinking about my health and my relationship with my family:
- My role in the family is unaffected by my health.
- There are some parts of my family role I cannot carry out.
- There are many parts of my family role I cannot carry out.
- I cannot carry out any part of my family role.

D.10 Thinking about my vision, including when using my glasses or contact lenses if needed:
- I see normally.
- I have some difficulty focusing on things, or I do not see them sharply. For example: small print, a newspaper, or seeing objects in the distance.
- I have a lot of difficulty seeing things. My vision is blurred. For example: I can see just enough to get by with.
- I only see general shapes, or am blind. For example: I need a guide to move around.

D.11 Thinking about my hearing, including using my hearing aid if needed:
- I hear normally.
- I have some difficulty hearing or I do not hear clearly. For example: I ask people to speak up, or turn up the TV or radio volume.
- I have difficulty hearing things clearly. For example: Often I do not understand what is said. I usually do not take part in conversations because I cannot hear what is said.
- I hear very little indeed. For example: I cannot fully understand loud voices speaking directly to me.

D.12 When I communicate with others: (For example: by talking, listening, writing or signing)
- I have no trouble speaking to them or understanding what they are saying.
- I have some difficulty being understood by people who do not know me. I have no trouble understanding what others are saying to me.
- I am only understood by people who know me well. I have great trouble understanding what others are saying to me.
- I cannot adequately communicate with others.

D.13 If I think about how I sleep:
- I am able to sleep without difficulty most of the time.
- My sleep is interrupted some of the time but I am usually able to go back to sleep without difficulty.
- My sleep is interrupted most nights but I am usually able to go back to sleep without difficulty.
- I sleep in short bursts only. I am awake most of the night.

D.14 Thinking about how I generally feel:
- I do not feel anxious, worried or depressed.
- I am slightly anxious, worried or depressed.
- I feel moderately anxious, worried or depressed.
- I am extremely anxious, worried or depressed.
D.15 How much pain or discomfort do I experience?
- None at all.
- I have moderate pain.
- I suffer from severe pain.
- I suffer unbearable pain.

**For these questions please circle the alternative that best describes you during the last four weeks.**

D.16 It has been easy to relate to others
- Almost always
- Most of the time
- About half the time
- Occasionally
- Not at all

D.17 I felt isolated from other people
- Almost always
- Most of the time
- About half the time
- Occasionally
- Not at all

D.18 I had someone to share my feelings with
- Almost always
- Most of the time
- About half the time
- Occasionally
- Not at all

D.19 I found it easy to get in touch with others when I need to
- Almost always
- Most of the time
- About half the time
- Occasionally
- Not at all

D.20 When with other people, I feel separate from them
- Almost always
- Most of the time
- About half the time
- Occasionally
- Not at all

D.21 I felt alone and friendless
- Almost always
- Most of the time
- About half the time
- Occasionally
Not at all

D.22 I have felt cheerful and in good spirits
- All of the time
- Most of the time
- More than half of the time
- Less than half of the time
- Some of the time
- At no time

D.23 I have felt calm and relaxed
- All of the time
- Most of the time
- More than half of the time
- Less than half of the time
- Some of the time
- At no time

D.24 I have felt active and vigorous
- All of the time
- Most of the time
- More than half of the time
- Less than half of the time
- Some of the time
- At no time

D.25 I woke up feeling fresh and rested
- All of the time
- Most of the time
- More than half of the time
- Less than half of the time
- Some of the time
- At no time

D.26 My daily life has been filled with things that interest me
- All of the time
- Most of the time
- More than half of the time
- Less than half of the time
- Some of the time
- At no time

Now that you have finished, please check that you have answered all the questions.

Thank you very much for completing and returning this survey. If you do not have the pre-paid return envelope, please return this form to:

Assoc/Prof Graeme Hawthorne