

GUTSS
The Genito-Urinary Treatment Satisfaction
Scale Study

Dr Graeme Hawthorne

Deputy Director, Program Evaluation Unit, Centre for Health Program Evaluation,
The University of Melbourne

Ms Cath Harmer

Centre for Health Program Evaluation
The University of Melbourne

December, 2000

ISSN 1325 0663

ISBN 1 876662 33 6

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434 **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au
Web Address <http://ariel.unimelb.edu.au/chpe/>

ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

AUTHOR ACKNOWLEDGMENTS

This report is based on a research project submitted by Ms Cath Harmer in partial fulfilment for the degree of Master of Public Health at The University of Melbourne. Although the original work was undertaken by Ms Cath Harmer, this paper is very different to her original report: it incorporates extensive revisions, omits a lot of material from the original report and includes much new material, particularly regarding the data analyses.

We would like to acknowledge the support of the *Colposuspension, Laparoscopic versus Abdominal Incision Multicentre* (CLAIM) trial team, particularly Dr Marcus Carey and Ms Anne Cornish, and also to the women who participated in the research. We would like to thank Sandy and Alexandra, who made a generous gift of the many hours required to conduct the research and report on the findings.

Ethics approval was obtained from The University of Melbourne's Health Sciences Human Ethics Subcommittee (HREC No. 980114), and from the Royal Women's Hospital's Research and Ethics Committee as an amendment to the original ethics approval granted to the CLAIM trial (Project number 96/48). Approval from other participating centres in the CLAIM trial was not required upon approach to the respective ethics committees, as it was believed that the patient satisfaction research was covered under the original issuing of ethics approval for the clinical trial.

Dr Hawthorne's position at The University of Melbourne is funded by the Victorian Consortium for Public Health.

The work undertaken in the GUTSS study and presented here is the sole responsibility of the authors. It does not represent the views of The University of Melbourne, the Centre for Health Program Evaluation, those of the CLAIM trial members or of the medical centres participating in the trial.

Table of Contents

| | |
|--|-----------|
| Executive summary | i |
| 1 Introduction..... | 1 |
| 2 Background and rationale..... | 2 |
| 2.1 Patient satisfaction | 3 |
| 2.1.1 Socio-political context and use of patient satisfaction measures | 3 |
| 2.1.2 The construct of patient satisfaction | 4 |
| 2.1.3 Validity and reliability in measuring patient satisfaction..... | 5 |
| 2.2 Incontinence and patient satisfaction..... | 8 |
| 2.2.1 Defining incontinence..... | 8 |
| 2.2.2 Prevalence and impact of incontinence | 9 |
| 2.2.3 Surgery and outcomes | 10 |
| 2.2.4 Laparoscopic studies | 13 |
| 2.3 The CLAIM trial and measuring patient satisfaction..... | 14 |
| 2.3.1 Some issues for measuring patient satisfaction | 15 |
| 3 Study method | 16 |
| 3.1 Research methods..... | 16 |
| 3.1.1 Recruitment | 17 |
| 3.1.2 Questionnaire development | 17 |
| 3.1.3 Construction and validation..... | 18 |
| 3.2 Criteria instruments | 19 |
| 4 Construction of the GUTSS..... | 20 |
| 4.1 Focus group findings..... | 20 |
| 4.2 Questionnaire and item description | 22 |
| 4.3 Participant details..... | 23 |
| 4.4 Scale development: reliability and content validity | 27 |
| 4.4.1 The theoretical models..... | 27 |
| 4.4.2 Item properties | 28 |
| 4.4.3 Data reduction and weighting | 29 |
| 4.4.4 Model A: A unidimensional scale | 30 |
| 4.4.5 Model B: The multidimensional scale | 31 |
| 4.5 Scoring the GUTSS..... | 33 |
| 4.5.1 Model A | 33 |
| 4.5.2 Model B | 34 |
| 4.6 Validation of the GUTSS scales | 37 |
| 4.6.1 Construct validation: convergent-discriminant validity..... | 37 |
| 4.6.2 Criterion validity..... | 39 |
| 4.6.3 Open or laparoscopic technique | 40 |
| 5 Discussion | 41 |
| 5.1 Choice of scale..... | 41 |
| 5.2 Limitations | 42 |
| 5.3 Towards a theory of patient satisfaction | 46 |

| | |
|---------------------------|-----------|
| 6 Conclusion | 47 |
| References | 47 |

List of Tables

| | | |
|----------|--|----|
| Table 1 | Details of the women interviewed..... | 24 |
| Table 2 | Health status of participants..... | 25 |
| Table 3 | Post surgical clinical measures..... | 27 |
| Table 4 | Analysis of all items..... | 29 |
| Table 5 | Structure of the unidimensional satisfaction scale, Model A..... | 31 |
| Table 6 | Structure of the multi-dimensional satisfaction scale, Model B..... | 32 |
| Table 7 | Internal structure of the Model B scale..... | 33 |
| Table 8 | Pearson correlations between Model A and Model B and the UDI, IIQ, SF-36, AQL & HPS..... | 37 |
| Table 9 | Criteria evidence for the GUTSS Model A..... | 38 |
| Table 10 | Criteria evidence for the GUTSS Model B..... | 39 |
| Table 11 | Clinical criteria evidence for the GUTSS models..... | 40 |
| Table 12 | Validation of the GUTSS internal structure..... | 44 |
| Table 13 | Validation of the GUTSS at 6 months following surgery..... | 45 |

List of Figures

| | | |
|----------|---|----|
| Figure 1 | Model A histogram..... | 34 |
| Figure 2 | Model B histogram..... | 35 |
| Figure 3 | Model B Outcome satisfaction histogram..... | 36 |
| Figure 4 | Model B Care satisfaction histogram..... | 36 |

Executive summary

Urinary incontinence in adults affects the quality of a person's life and is common within our community, with at least double the number of women enduring the condition at any age when compared with men (Thom 1998). The effectiveness of treatment for urinary stress incontinence as judged by the patient is thus an important measurement of its outcome. However, there exists no known in-depth qualitative research or psychometrically determined measurement of patient satisfaction with outcome of incontinence treatment. The purpose of the *Genito-Urinary Treatment Satisfaction Scale* (GUTSS) study was to develop a reliable and valid measure of women's satisfaction with the outcome of treatment for genuine urinary stress incontinence and other related disorders. This was achieved with the assistance of the research team and women participating in the Colposuspension, Laparoscopic versus Abdominal Incision Multicentre (CLAIM) trial at the Royal Womens' and Mercy Hospitals in Melbourne, Australia. The methods used in the construction of the GUTSS were:

1. A literature review identifying the context and use of patient satisfaction measures; issues in the construction and measurement of patient satisfaction; prior reporting of patient satisfaction in the incontinence field; and the impact and prevalence of urinary incontinence in the population.
2. Two focus groups were held. The first was with CLAIM trial team members and the second with 10 participants enrolled in the CLAIM trial. The purpose was to identify the component areas believed to be important in the measurement of patient satisfaction with the outcome of treatment for urinary incontinence.
3. Development and piloting a 22-item questionnaire covering the following components of patient satisfaction, as synthesised from the literature review and focus groups: adverse effects and pre-existing conditions, care received, pain and discomfort, satisfaction with outcome, current health status and living, and expectations.
4. Conducting telephone interviews using the developed questionnaire with 45 women from the CLAIM trial and who were 6 months or more post-treatment. The responses to the items were assessed through factor analysis, reliability measurement (Cronbach α) and regression coefficients.

Two models were hypothesised from the literature review and the qualitative research. These represented a single dimensional model (Model A) and a multi-dimensional model (Model B). The two scales corresponding to the models performed similarly with respect to the analyses undertaken. However, it was judged that the component areas in the multi-dimensional Model B would better meet the needs of clinicians and researchers. In addition, Model B accorded with the findings of the literature review to a greater extent than the unidimensional Model A. Given these findings, Model B was identified as the *GUTSS* instrument.

The GUTSS instrument (Cronbach $\alpha = 0.83$) comprises two internally reliable and independent sub-scales: *Outcome satisfaction* (Cronbach $\alpha = 0.93$) and *Care satisfaction* (Cronbach $\alpha = 0.76$). Taken together, these two scales incorporate questionnaire items from three of the six component areas identified through the literature review and qualitative research. Both sub-scales were judged to have high ecological validity as a result of the development of the items

with trial participants. Construct validity evidence was determined through correlation with the *Incontinence Impact Questionnaire* (IIQ) and with the *Urogenital Distress Inventory* (UDI); both specific urinary incontinence health-related quality of life (HRQoL) measures. The GUTSS was highly correlated with both these instruments, $r = -0.67$. These findings were supported by moderate correlation with the SF-36, the *Assessment of Quality of Life* (AQoL) instrument (a utility health-related quality of life instrument) and the *Patient Satisfaction Survey* (PSS) measure (a generic patient satisfaction scale). The findings accord with past research that subjective and objective measures of treatment outcome do not necessarily measure the same construct (Hawthorne and Batterham 1996). In addition, six months after operation the GUTSS instrument did not identify a significant difference in patient satisfaction between open incision and laparoscopic surgery used to provide the treatment of Burch colposuspension. This null result was predicted, as the 6-month period between undergoing the procedure and measuring patient satisfaction, set for data collection by the CLAIM trial, was believed to diminish the more immediate outcome benefits of having the laparoscopic technique and to contribute to a greater overall level of satisfaction with treatment. Significant differences were, however, observed between those who reported continuing incontinence and those reporting they were continent, suggesting that satisfaction is a function of absolute health status.

These findings from the construction sample were replicated in an analysis of the GUTSS using data from a second study involving similar surgical procedures. The factorial structure, internal consistency and sensitivity as measured by post-operative absolute health status were all confirmed in this second analysis, suggesting the GUTSS possesses, *prima facie*, sufficient validity, reliability and sensitivity to used with confidence.

The study limitations primarily relate to the small number of participants ($n=45$) in the development of the GUTSS with respect to the psychometric modelling.

From the findings, a *dual role transgression* theory of patient satisfaction is proposed. That is, the level of satisfaction is hypothesised to depend upon the patient's post-operative absolute health status and also her understanding of the health worker's role *vis-a-vis* and her own role as a patient-consumer.

In conclusion, it is hoped that the GUTSS instrument will be a useful clinical and research measure of incontinence treatment outcomes, particularly where new technology and techniques are being introduced. The aim of the GUTSS is to assist with improving the HRQoL of women who have undergone treatment for urinary incontinence through contributing to improving the measurement of their satisfaction with surgical interventions.

1 Introduction

The current study was undertaken in response to a request from the Colposuspension, Laparoscopic versus Abdominal Incision Multicentre (CLAIM) Trial conducted in Melbourne, Australia, to develop a patient satisfaction questionnaire for use by participants. Its primary purpose was to measure the level of treatment outcome expressed by women undergoing the Burch colposuspension procedure for stress urinary incontinence (Hilton and Stanton 1983; Cundiff, Harris et al. 1997). This procedure involves restoration of the neck of the bladder through suturing to Cooper's ligament, and is described as urethrovaginal fixation (Burch 1968). The procedure was either performed laparoscopically, also known as a 'keyhole procedure', or through an open abdominal incision. The comparison of the two different techniques was the aim of the CLAIM trial.

A second purpose was for the patient satisfaction instrument to be of use with other trials in the urogynaecological field and in on-going public and private practice. Thirdly, it was indicated that this subjective measure of treatment outcome would be considered as an alternative to the invasive urodynamics assessment of treatment outcome commonly used in clinical practice.

Supporting the measurement of patient satisfaction in the CLAIM trial is the increasingly broad use of patient satisfaction measures within the health care industry. Measurement of patient satisfaction has become an integral component of the assessment of the overall quality of a service, an individual episode of care, and as an expected outcome of care (Donabedian 1980; Donabedian 1988; Bowling 1991; Draper 1997). It is this construct — the measurement of what are thought to be the dimensions of "patient satisfaction" — which has the potential to determine whether a person will attend for treatment, take treatment, and identify her/himself as recovered (Larsen and Rootman 1976; Bowling 1991). As such, patient satisfaction indicators are used within health care program planning and evaluation (Dull, Lansky et al. 1994; Draper 1997). Another reason for the use of a subjective measure within the CLAIM trial was that subjective measures of treatment do not necessarily equate to objective measures (Ware, Snyder et al. 1983): Hawthorne & Batterham estimated the correlation is approximately 0.35 (Hawthorne and Batterham 1996). That is, patient satisfaction measures may provide different information when compared with objective ratings of health care treatment outcomes. It is important to note, however, that although this finding supports the measurement of patient satisfaction adding to the clinical information regarding the outcome of interventions, it suggests that the third objective proposed by the CLAIM trial for measuring patient satisfaction — an alternative to objective clinical measures — may not be valid. The subjective measure of patient satisfaction cannot replace objective measures, but may usefully complement them.

In assessments of treatment outcome for incontinence, objective measurements are provided with detail based on urodynamic tests. This may be compared with the measurement of patient satisfaction, which has not been conducted using psychometrically validated and reliable instruments (Streiner and Norman 1989; Bowling 1991; Wright and Feinstein 1992; McDowell and Newell 1996; Gormley 1997). Nor is there any evidence of the use of standard qualitative interview techniques and analysis methods being used either. For example, in a retrospective study by Benderev on endoscopic bladder neck suspension modifications, patient satisfaction was reported as being measured via a scale ranging from +3 (extremely satisfied) to -3 (extremely dissatisfied). No further details were provided (Benderev 1992). Similarly, in a

prospective study on orthoptic diversion in women using the Kock Ileal neobladder procedure “patient satisfaction” was reported (Taylor, Beart et al. 1983). However, the questions asked were not reported, and it is unclear if the questions were collected from self-administered questionnaires and/or from personal interviews with health workers involved in the procedure at the hospital. The authors reported that the sample of women *overall* were satisfied. What *overall* relates to was not indicated.

With the advent of laparoscopic technology in the urogynaecological field, treatment outcome is of increased clinical and managerial importance, but in the one randomised control trial reported at the time of writing patient satisfaction was not measured and the “success rate” of the laparoscopic procedure was reported to be lower than for open surgery (Su, Wang et al. 1997). In a non-randomised prospective study where patient satisfaction was measured, the question(s) were not provided, but it was indicated that satisfaction was based on meeting expectations and *quality of life* indicators (Liu 1993). This basis of measurement should not be taken as a valid measure of patient satisfaction, as quality of life scales do not measure patient satisfaction, although they may be confounded by patient expectation. The research, though, reports that expectations account for just 8% of the variance within patient satisfaction scores (Linder-Pelz 1982; Bowling 1997). The Liu study reported a high success rate for the laparoscopic procedure, and substantiated the finding by listing the procedural and quality of life advantages when compared to the open technique; but it is unknown whether the advantages of the laparoscopic technique would be similarly assessed by women who had undergone the open technique.

In contrast to these incontinence studies, in an investigation of cholecystectomy surgery outcome where laparoscopic technology was compared to open incisions it was shown that patients with open cholecystectomies were significantly more satisfied with all dimensions of their care (Kane, Maciejewski et al. 1997). This finding, and the conflicting results to date in the incontinence studies, suggest that the enthusiasm with which laparoscopic technology has been embraced within the urogynaecological field needs to be more vigorously assessed from the patient’s perspective, not just the clinicians’.

As no psychometrically valid and reliable patient satisfaction instrument has been used in the incontinence field, the importance of using such an instrument to measure patient satisfaction is evident for two reasons. The first is to address the significant ecological validity gap in the research due to the lack of appropriate techniques in developing items and questionnaires to ensure they reflect the concerns of patients. The second is that the inconsistent findings suggest that the patient perspective may add additional information to clinical outcomes. If the clinician is satisfied with the objective outcome of treatment but the woman is dissatisfied with the treatment procedure and/or outcome, these divergent perspectives should be reflected in evaluations of new procedures. Where the perspective of the patient is not included, the implications may be that although the clinician may wish to use the technology the woman may not keep her appointment, may reject the use of the technology, and/or may not attend for follow-up care. All of which have possible negative implications for the health of the woman living with incontinence.

2 Background and rationale

In order to measure patient satisfaction with treatment for urinary stress incontinence this study examined the following:

-
- the purpose and use to which patient satisfaction measurements are put;
 - clarification of the construct that is actually being measured as patient satisfaction;
 - identification of how patient satisfaction has been measured in the past, particularly with respect to the validity and reliability of the measurements in relation to incontinence treatment;
 - the benefits to the individual and society in measuring patient satisfaction with incontinence treatment;
 - the limitations and opportunities that the implementation of the CLAIM trial itself presented for the measurement of patient satisfaction.

2.1 Patient satisfaction

2.1.1 Socio-political context and use of patient satisfaction measures

Patient satisfaction is increasingly cited as a key factor in assessing the quality of care within hospitals in Australia (Draper and Hill 1996; Draper 1997), following on from the extensive use of patient satisfaction measurement in hospital services of the United Kingdom and the United States (Cleary, Edgeman-Levitan et al. 1991; McIver 1992). This trend towards including the consumer perspective can be traced to two distinct movements. The first, originating in the 1950s, was to improve clinical outcomes through evaluation; the second, from the 1960s and 1970s, was part of the individual rights movement (Williams 1997). Together these reflected increased concern with consumers' evaluations of interventions. This approach is a central concern of the 'new managerialism' now dominant in the public sector.

Its use in Australia can be traced to national evaluation policy: since 1984 the objectives of Australian federal evaluation policy have emphasised efficiency, effectiveness and accountability, with patient satisfaction being an indicator of accountability (Department of Finance and Australian Public Service Board 1987). Importantly, the development and ongoing monitoring of the evaluation policy was located within the Department of Finance. This strengthened the emphasis on the objective of economic efficiency, including the need to justify and validate initiatives undertaken in the health care system by using patient satisfaction measures. The context in which patient satisfaction is being undertaken in the public health sector in Australia, and particularly in Victoria, is further enhanced by an understanding of the central tenet of 'new managerialism'. Originating in private sector practices, its central tenet is that of continuous reform to ensure responsiveness to the customers' needs (Shafritz and Russell 1997; Corbett 1998).

Hence the need to assess the patient's (consumer's) satisfaction with health care as a means of legitimising health policy and improving those areas where consumer dissatisfaction is expressed (although the extent to which the consumer perspective is acted upon has been contested (Shafritz and Russell 1997)). In order to ensure that the interests of the patient are being represented and that evaluations are not just satisfying the needs of vested interests some researchers have suggested that the ways patient satisfaction is constructed, measured, and reported needs to occur as independently as possible from the socio-political context outlined above (McTaggart and Blackmore 1990; House 1993; Scriven 1993).

*

In the current context, the importance of patient satisfaction is its contribution to evaluation; although once completed, evaluations directly provide evidence for the socio-political context outlined above. The aim of the CLAIM trial was to determine differences between the two techniques previously described, particularly to provide information on the benefits and cost-effectiveness of the laparoscopic procedure. Assuming there were no clinical outcome differences between the two interventions, that there were cost-benefits associated with the laparoscopic procedure, and that patient satisfaction remained the same, then the preferred procedure would be laparoscopy. If there was significantly greater dissatisfaction with one of the procedures, then the situation would need to be reassessed.

2.1.2 The construct of patient satisfaction

Carr-Hill stated the obvious when he asserted that consumer satisfaction was an outcome of the health care system (Carr-Hill 1992); Donabedian, with a focus on quality assessment, saw patient satisfaction arising from the medical infrastructure, the treatment process and its outcomes (Donabedian 1980; Donabedian 1988). But these generic statements are not very informative and need to be further examined. Locker & Dunt, in their seminal review twenty years ago which analysed British work in measuring patient satisfaction, reported that patient satisfaction was a product of the amount of information received (Locker and Dunt 1978), while Pascoe defined patient satisfaction as a person's "reactions to salient aspects of the contexts, process and results of their experience" (Pascoe 1983, p 189).

Regarding the constituent components, Ware *et al* argued these comprised access to care, financial resources and their distribution, continuity of care, interpersonal manner, quality of care and treatment, and the conduct of doctors (Ware, Snyder et al. 1983). This classification was used by Loeken *et al* in their patient satisfaction instrument for women undergoing mammography; they defined the components as accessibility, interpersonal and technical aspects of care and patient education/information (Loeken, Steome et al. 1997). Hardy *et al* defined three areas of satisfaction related to hospital care as being the process of care itself (components), the patient's improvement (determinants), understanding of their health (components and determinants), and their psychological wellbeing (determinants) (Hardy, West et al. 1996). Similarly, Sitzia & Wood made the distinction between satisfaction *determinants* (the patient variables) and *components* (care variables). They grouped determinants into expectations, patient characteristics and psychosocial determinants (Sitzia and Wood 1997). In short, patient satisfaction is thought to be an artefact of the patient's perceptions of the different dimensions relevant to it (Linder-Pelz 1982).

The models of satisfaction above have been questioned by Williams who reviewed of some of the earlier work which advanced psychological theories of satisfaction, including Lawler's *discrepancy theory*, *fulfilment theory* and *equity theory* (Williams 1997). Linder-Pelz, drawing on the work of Fishbein & Ajzen (Fishbein and Ajzen 1975) proposed a *value-expectancy* model, which when tested failed to support the Fishbein & Azjen model; satisfaction was unrelated to fulfilment but inversely correlated with discrepancy. Linder-Pelz concluded that prior expectations were the main determinants of subsequent expressed satisfaction, irrespective of whether the expectation was fulfilled or not (Linder-Pelz 1982). Unfortunately, the independent effects of expectations accounted for only 8% of the variance in satisfaction, and values and perceived occurrences only 10% (Linder-Pelz 1982). In research examining the relationship between health outcomes and

satisfaction, Kane, Maciejewski & Finch found that outcomes explained less than 8% of the variance in patient satisfaction. They concluded that the time elapsed since treatment diminished the patient's ability to distinguish between the various components of satisfaction (Kane, Maciejewski et al. 1997). This implies that patient satisfaction should be measured at multiple points in time. Such findings have led to others hypothesising that patient satisfaction is affected by variables outside of health care service control. Just how these variables interact with satisfaction remains unknown: in Thompson's study of these variables, it was shown they explained 20% of the variance (Williams 1997).

Another consideration is the expectation that the person receiving care has about the role performance of the caregiver, including the communication of information (Carr-Hill 1992). Larsen & Rootman found a close correlation with the physician's role performance and patient satisfaction (Larsen and Rootman 1976). Twenty years later Owens & Batchelor reported similar findings (Owens and Batchelor 1996).

The issues described above relating to defining the universe of patient satisfaction are paralleled by a lack of agreement on what ought to be measured. As Williams argued, researchers need to know what people mean when they state they are satisfied (Williams 1997). To interpret and use patient satisfaction information and remain true to the respondent's intent the researcher must have an understanding of how the patient evaluated the treatment, including any sources of dissatisfaction (Carr-Hill 1992). Williams argued that dissatisfaction is not expressed unless negative expectations have been exceeded. It follows that where there is no room for the expression of dissatisfaction, measures of satisfaction may not be true evaluations (Williams 1997). Thus qualitative studies and open-ended questions have elicited more negative experiences than surveys using closed dichotomous questions (such as those answered with a tick or "yes" or "no") (Locker and Dunt 1978; Carr-Hill 1992; Williams 1997; McIver and Meredith 1998). One of the difficulties is that only extreme behaviour results in expressions of dissatisfaction, thus explaining why most people express satisfaction when responding to patient satisfaction questionnaires. As Bowling noted, a failure to detect dissatisfaction has foiled the measurement of patient satisfaction (Bowling 1997).

Part of the reason for this is that patients build up relationships with their doctors; ie. dependency may invoke a reluctance to criticise, especially where the relationship is continued through ongoing treatment. Often the patient cannot easily request services elsewhere and may be highly dependent on health care workers (Hardy, West et al. 1996). This dependency suggests that patients are not market consumers in the usual sense of the term, and that they do not make critical evaluations (Drummond, O'Brien et al. 1997; Williams 1997), especially where they are older and have chronic health care needs (Owens and Batchelor 1996). In addition, dependency may be stronger where the patient feels he/she has imperfect information about his/her health state and health needs (Drummond, O'Brien et al. 1997). This suggests that the information provided to the patient may play a critical role in the formation of satisfaction evaluation.

2.1.3 Validity and reliability in measuring patient satisfaction

The use of patient satisfaction measures assumes that a patient's satisfaction is measurable. Furthermore, the act of measurement itself confers upon the patient's satisfaction a score that is often assumed to be a reliable and valid measure simply because it was "measured". These assumptions are based on the following premises: (a) what is measured accurately represents

the patient's beliefs; (b) that the measurement processes do not impact on these beliefs; and (c) that the measurement of satisfaction is an accurate evaluation of the care received and the subsequent health outcome. Where these premises can be shown to be true, we may claim validity evidence for the instrument used.

In general terms validity can be defined as ensuring that an instrument measures what it purports to measure. Regarding patient satisfaction specifically, as the accuracy of the numerical values obtained cannot be established by direct observation or by comparison with some 'gold standard' of satisfaction, instrument reliability and validity should be determined by establishing a nomological net of evidence (Cronbach and Meehl 1955). This implies that validation of a satisfaction instrument is an ongoing process and an instrument may prove to be valid in one context but not in another. The nomological net of validity evidence includes instrument content (does it adequately reflect the theoretical universe upon which it is premised?), criteria (how well does the measurement match values obtained from other instruments purporting to measure the same thing?) and construct validity (does the instrument provide scores from which inferences about the defined universe can be made?). In addition, instruments need to be reliable (is the measurement stable?). With respect to the construction of a patient satisfaction instrument, each of these is briefly discussed below.

The following discussion is based on this explication of the premises underlying the assumed isomorphic relationship between the concept of patient satisfaction, the measurement instrument used and the observed (constructed) model of patient satisfaction deriving from instrument scores. A critical issue in ensuring validity is the incorporation into this model of patient perspectives, particularly where the form of collecting information about patient satisfaction is through a formal questionnaire. This is the preferred approach, in part due to its low cost, time efficiency and standardisation. Criticisms of this method include that questionnaires may be designed to reflect particular aspects of care (Williams 1997) and that limited response sets may not cover the universe of experiences.

Content validity

Given patients are the experts regarding their satisfaction with services provided and treatment given, it would seem that an ecological approach to content validation is necessary.

A starting point in ensuring ecological validity is to enlist the active participation of the population to be surveyed in the construction of the questionnaire; thus patients-as-consumers are co-producers who have their perspective incorporated into what is evaluated (Palmer, Donabedian et al. 1991; Wadsworth 1991; Alford 1997; Yeatman 1997). Where this occurs it is argued that patients are empowered (their expertise is recognised and acknowledged) (Fetterman, Kaflarian et al. 1996; Smith 1998), essentially increasing the 'face' validity of the questionnaire items. This is not to argue that the inclusion of patients' perspectives should be at the expense of other stakeholders' (eg. surgeons, nursing staff, and administration personnel; hospital managers, funding bodies and future users of the patient satisfaction instrument), but rather that the process should be inclusive as outlined in 'stakeholder theory' (Guba and Lincoln 1990). Such an inclusive process ought to improve instrument content validity, as the resulting items should cover all the components important to the interested parties (Bowling 1991). There is an argument that, within this inclusive process, the number of items in each dimension to be measured should

increase with the value assigned to the dimension by the interested parties (Streiner and Norman 1989).

Construct validity

Questionnaires must also have 'construct validity'; which is the extent to which the instrument accurately measures the construct of patient satisfaction. Construct validity is comprised of *discriminant* and *convergent* validity. To assess construct validity the scale(s) within an instrument should be tested against scales that measure or predict similar constructs (convergent validity) and scales/factors where the correlation is expected to be divergent (discriminant validity).

Regarding convergent validity, within the CLAIM trial measures of HRQoL were undertaken through use of the *Incontinence Impact Questionnaire* (IIQ) and the *Urogenital Distress Inventory* (UDI). We would expect the correlations between these instruments and patient satisfaction to be moderate, based on the argument that the construct of HRQoL, although is not the same thing, is related to patient satisfaction (Bowling 1991).

Discriminant validity, on the other hand, might be demonstrated where patient satisfaction instrument scores are independent of socio-demographic factors such as income, age, and education level. This argument is predicated on the assumption that socio-demographic variables do not independently affect satisfaction with a service or procedure. If this assumption is false, patient satisfaction scores will vary by these predictors. In the life satisfaction literature there is some evidence that these factors do affect estimates of life satisfaction (Cummins 1995; Cummins 1996).

Criterion validity

The final requirement for reliable instruments is 'criterion validity'; that is, the extent to which the measure correlates with a 'gold standard' (Bowling 1997). It consists of two types: *concurrent* and *predictive* validity.

Concurrent validity is where the new instrument being developed can be substituted for a past measure. Reasons for wishing to substitute one measure for another include reducing the invasiveness of the measurement process; improving the ease of administration; and cost minimisation. As measures of patient satisfaction in the incontinence field have not been previously developed through psychometric methods, the establishment of criterion validity with an instrument that measures patient satisfaction in this field is not possible. However, comparison of patient satisfaction scores with a general patient satisfaction instrument can be made, as can comparisons with clinical incontinence measures. The correlation coefficients would be expected to be moderate, 0.30–0.60, as the former would not have the specificity of a condition specific instrument and the latter would be measuring a different construct of outcome (Hawthorne and Batterham 1996).

Predictive validity is the capacity of new measurement to predict future differences. In this study this was not investigated.

Reliability

A questionnaire must have, in addition to the qualities described above, reliability. The reliability of a measure lies in its capacity to consistently produce the same results from the same respondents under the same conditions (Ware, Snyder et al. 1983; Streiner and Norman 1989; Bowling 1997). This should occur when the instrument is given to the same people at different points in time where there has been no change in circumstances that would predict a change in score (Bowling 1991). The various forms of reliability include internal consistency, multiple form reliability, test-retest reliability, intra-rater and inter-rater agreements and sensitivity to change (Bowling 1991). Most usually, however, in cross-sectional studies, reliability is assessed via internal consistency (Cronbach α).

*

If the measurement of patient satisfaction with incontinence treatment can achieve reliability, content, construct and criterion validity as described above, then a valid and reliable instrument would have been constructed.¹

2.2 Incontinence and patient satisfaction

In order to measure patient satisfaction with incontinence treatment processes, a clear understanding of stress incontinence, the impact of incontinence on women's lives, its prevalence in the community, and the outcomes expected from surgery in relation to patient satisfaction is required. The following sections establish the need for measuring patient satisfaction in relation to incontinence treatment processes, and the possible component areas and determinants to be used in these measurements.

2.2.1 Defining incontinence

The International Incontinence Society defines urinary incontinence as being *a condition in which involuntary loss of urine is a social or hygienic problem and is objectively demonstrable* (Bates, Bradley et al. 1979; Walter 1983). Stress incontinence occurs where there is a disproportion between the pressure in the bladder and resistance in the urethra, causing the intravesical pressure to exceed the urethral closure pressure. This is distinguished from urge incontinence, where involuntary loss of urine is associated with a strong desire to void and uninhibited detrusor contractions (*motor urge incontinence*), which may or may not be associated with uninhibited detrusor contractions at cystometry (*sensory urge incontinence*) (Walter 1983).

A further definition is often assumed in investigations into the effectiveness of surgical and non-surgical procedures for incontinence, that is of *genuine urinary stress incontinence*. This is where the involuntary loss of urine occurs, but in the absence of detrusor activity. Genuine urinary stress incontinence is identified through the use of urodynamic assessment procedures, principally urethral pressure measurement. This involves an invasive procedure whereby a silicon-rubber coated dual-sensor microtransducer catheter is inserted into the woman's bladder

¹

This does not imply that validity or reliability will necessarily have been established. Validity and reliability are not fixed properties, but are evolving and may be demonstrated under particular circumstances or with particular populations. When used with a different population the validity and reliability of an instrument needs to be re-investigated; it cannot be taken as a given.

through the urethra to measure *resting* urethral closure pressure, a *stress* profile is then taken when the woman is asked to cough, and mid-urethral pressure responses measured (Hilton and Stanton 1983). This technique is also used to determine the outcome — the objective success or otherwise — of an operation for incontinence.

The causes of incontinence include: (a) increasing age associated with falling intra-urethral pressure; (b) hormonal changes during the menses; (c) the effect of genuine estrogens on the bladder-urethra function in postmenopausal women; (d) previous pelvic surgery where unsuccessful past surgery may be associated with low urethral pressure and short urethral length; and (e) parity (Hilton and Stanton 1983; Walter 1983). Regarding parity, it should be noted that Hilton identified no significant differences in incontinent women with no birthing history and increasing number of live births (parity) (Hilton and Stanton 1983). Cundiff *et al*, in their examination of clinical predictors of urinary incontinence in women, found that reliance on symptoms as compared to urodynamic testing for stress incontinence resulted in a misdiagnosis of 13% of women with stress incontinence without urge incontinence, and of 59% of women with urge incontinence without stress incontinence. They concluded that urodynamic testing seemed to be warranted for all patients (Cundiff, Harris *et al*. 1997). This highlights three issues: (1) the need to specify diagnostic techniques for stress incontinence, (2) the impact that relying on particular techniques may have on measured outcomes, and (3) the need to compare pre- and post-surgical measures of incontinence. In relation to measuring patient satisfaction with outcome of treatment for incontinence it suggests that a higher correlation between the urodynamics results as opposed to other measured outcomes should be identified. This is particularly true if women have been indicated for surgery on the basis of a prior urodynamics assessment where genuine urinary stress incontinence was diagnosed.

2.2.2 Prevalence and impact of incontinence

The prevalence of incontinence in the population varies markedly between studies, with estimates ranging from 2–57% (Ashworth and Hagan 1993; Brocklehurst 1993; Thom 1998). In a recent review of population-based surveys (from Scandinavian countries, the UK, US, New Zealand, Spain and Japan) Thom found that the reported prevalence varied with respect to method, although not significantly so. Interviews conducted in person elicited higher levels of reported incontinence than mailed questionnaires. Other factors such as location, study year, and response rate had minimal impact on the reported prevalence. Thom noted that other researchers found that asking probe questions following a negative response resulted in an additional 10% of respondents reporting incontinence. His review reported that urinary incontinence in younger women is 5–6 times higher than in younger men, and that this reduced to approximately double the rate in older women when compared to older men (Thom 1998).

Among women, stress incontinence is reported more frequently in the younger population, whereas stress and urge incontinence predominate in older women (Thom 1998). Overall, the range of prevalence of ‘ever experienced urinary incontinence’ reported in the various studies was 17–55% for women aged 65 years or more and 12–42% for women aged 15–64 (Ashworth and Hagan 1993; Brocklehurst 1993; Thom 1998). This level of prevalence in the community, combined with the social and emotional effects of incontinence, clearly identifies urinary incontinence as having a significant health impact on the quality of women’s lives.

However, it must be noted that definitions of stress incontinence in population surveys do not necessarily equate with the diagnosis of genuine urinary stress incontinence used for purposes of surgical intervention, nor is the prevalence known with respect to the definition of incontinence devised by the International Incontinence Society.

Regardless of how it is diagnosed or classified, the symptoms of urinary incontinence have been characterised as a 'hidden handicap' (Berglund, Eisemann et al. 1996). One study of women aged 25–55 years who predominantly had not sought help from a doctor or other health care worker, identified that the topic of incontinence was taboo. Not only was it a socially unacceptable conversation topic, but the women themselves found the topic difficult. These women saw incontinence as a problem related to personal control, often viewed as possibly being their own fault. Low self-image and restricted daily activities were common and for many women considerable time and planning were needed to try to maintain a normal lifestyle (Ashworth and Hagan 1993). Similarly, a UK survey found that embarrassment was common, with serious impact on sufferers' social lives. Incontinence was reported as leading to restrictions on activities, including the amount of liquid drunk, the need to know the locations of public toilets, going out less often, and restrictions on lifting (Brocklehurst 1993). In a New Zealand survey, only a third of women with regular incontinence had sought medical help. Many lacked awareness of management techniques and available surgical interventions (Holst and D 1988).

To the researchers' knowledge, no similar surveys have been conducted in Australia.

2.2.3 Surgery and outcomes

Treatments for women with stress urinary incontinence include a number of surgical techniques categorised under the following seven types of intervention:

- bladder buttress operations associated with Kelly & Pacey;
- the Marshall-Marchetti-Krantz procedure;
- colposuspension as described by Burch;
- long-needle suspensions with endoscopic control, associated with Pereyra;
- long-needle suspensions under endoscopic control, associated with Stamey;
- sling procedures;
- periurethral injections (Jarvis 1994).

In 1994 Jarvis reviewed these techniques to determine which was preferred. He noted the poor comparability of the studies undertaken to that time, the lack of a minimum follow-up period, a lack of standardisation of inclusion and exclusion criteria, and poor notification of small technique variations (Jarvis 1994). He concluded that there was no single genuine stress incontinence operation which should be offered as a first choice. He did report, however, that the Burch colposuspension, the Marshall-Marchetti-Krantz procedure, endoscopic bladder neck suspension and bladder sling operations produced mean continence rates in excess of 85% on clinical measures. In reviewing subjective patient reports of outcomes, he observed that these results gave a greater cure rate for incontinence than clinical measures, and suggested this was the key reason subjective measures were unsatisfactory in scientific assessment. Yet others have relied upon subjective assessments. For example, in a population-based survey ($n=1,148$) that examined long-term satisfaction with the outcome of surgical treatment, Dionko *et al* asked for

details of how much and how frequently urine was being lost. They reported that of the 46 women who were 2 years or more post-surgery, only 39% reported that they were absolutely continent (Diokno, Brown et al. 1989).

In a more recent review by the American Urological Association similar difficulties were also identified, although it was reported that, based on long-term cure/dry rates of at least 48 months, retropubic suspensions were the most efficacious procedures (these include Burch colposuspension, the Marshall-Marchetti-Krantz procedure (84% cure rate respectively), and slings (83% cure rate) (Leach, Dmochowski et al. 1997). No mention was made of whether the techniques within the studies reviewed were performed with open-type surgery or using laparoscopic techniques. The Association did note that the recommended procedures were associated with slightly higher complication rates, including post-operative voiding dysfunction, and longer convalescence. The definition of 'cure' was not total clinical continence, as in Jarvis' review, but both objective and subjective cure definitions were accepted. The review recognised that 'cure' does not always equate to 'dry', and included a cure/dry/improved category so that the women's' judgments of successful operations could be recorded.

In a 10–20 year follow-up study by Alcalay *et al* of women who, between 1974 and 1983 underwent a modified Burch colposuspension, the 'objective cure' of incontinence (as defined by inability to demonstrate stress incontinence during clinical examination and provocative urodynamics) was estimated at 69% 10–12 years after surgery, whereupon a plateau was reached (Alcalay, Monga et al. 1995). Two features of this study are pertinent to the measurement of patient satisfaction. The study tracked women longer than the 2-year minimum period for long-term follow-up set by the American Urological Association, thus demonstrating the changing nature of incontinence post-surgery over an extended period of time (Cundiff, Harris et al. 1997; Leach, Dmochowski et al. 1997). On analysis of subjective and objective cure rates no difference was found, which was inconsistent with the findings of Jarvis reported above. In a trial of the Contelle device (which provides intravaginal electronic stimulation for female stress and urge incontinence) it was reported that there were some discrepancies between objective urodynamic tests and the patients' ratings of treatment efficacy (Fall, Ahlstrom et al. 1986). Nowhere in the paper was it reported how women were asked about the efficacy of the treatment, or what the discrepancies were indicating in relation to the efficacy of the treatment. This is of particular concern as the trial was for new technology: out of the original 51 patients offered long-term intravaginal electronic stimulation four withdrew as the device was uncomfortable, two withdrew because of slight vaginal bleeding and five withdrew because it was an unaesthetic mode of treatment. This represented 22% of the original 51 patients asked to participate. Although two studies have reported on the social impact of surgery for stress incontinence, neither investigated patient satisfaction (Berglund, Eisemann et al. 1996; Black, Bowling et al. 1998). One study used a subjective cure rate measure, but its determination was not described and the data were collected within the hospital environment using non-validated scales (Berglund, Eisemann et al. 1996). The other study concluded, based on the stability of the outcomes, that assessment is only required once and it may be at any time from 3–12 months after the operation; the extent of improvement depended on pre-operative severity (Black, Bowling et al. 1998). Whether pre-operative severity relates to a woman's satisfaction with treatment outcome is unknown, as is the stability of patient satisfaction during the 3–12 month period.

Regarding the relationship between clinical cure and satisfaction the literature poses more questions that it answers due to the failure to adequately report how satisfaction is defined or measured. Litwiller *et al* , when assessing the vaginal wall sling operation, asked women for their current satisfaction with urinary status, their willingness to undergo the operation again and willingness to recommend the operation to someone with similar problems. Although no validation evidence for these questions was presented, it was reported that 62% (26) of the women were satisfied with their urinary status after 31 months post-operation. Post-operative urge incontinence was the single most important factor affecting patient satisfaction; it was a slightly better predictor than resolution of genuine stress urinary incontinence (Litwiller, Nelson et al. 1997). A similar line of questioning by Walker & Texter when evaluating the Stamey procedure found that 65% would undergo the procedure again if necessary. They reported that incontinence was highly correlated with satisfaction estimates (Walker and Texter 1992). Again, however, no psychometric data on validity or reliability were presented. Zorzos & Paterson in their evaluation of the Marshall-Marchetti-Krantz procedure 2–13 years post-surgery asked two dichotomous questions probing satisfaction. They reported that 58% of women were satisfied; again no psychometric evidence on the questions were presented (Zorzos and Paterson 1996). In a study of endoscopic bladder neck suspension modifications, patient satisfaction was reported using a scale ranging from +3 (extremely satisfied) to -3 (extremely dissatisfied), but no further information was provided (Benderev 1992).

What is evident from these studies is that there are two critical issues which need to be addressed. One is in relation to the definition of satisfaction (ie. what is being measured?) and the other is in relation to the validity of the measurement (ie. can we trust the measurement?). These fundamental questions can be illustrated by reviewing the Jarvis study described above (Jarvis 1994). Although Jarvis reported on subjective measures, no definitions were reviewed or provided and it is unclear whether like measures were being compared, whether psychometrically developed instruments or qualitative techniques were being used. The limitations of this position at a theoretical level would include questions around whether the measures were evaluating different aspects of outcome and whether they were valid and reliable indicators. The position taken in Jarvis' study leads to a minimalist construction of cure-outcome in relation to surgical treatment of incontinence. In addition, none of the review studies sighted by the researchers indicated whether subjective measures were taken prior to or after the objective measures; the order of measurement may have influenced the subjective estimates. Although Alcalay *et al* stated that subjective measures were taken at the hospital, it is not clear whether the person who undertook the outcome measurements was a person who operated on any of the women (Alcalay, Monga et al. 1995); the same issue of dependent relationships is also evident in the Litwiller *et al* study (Litwiller, Nelson et al. 1997). Although this analysis has been primarily made with respect to Jarvis' study, it could equally apply to the other studies reviewed above: as reported by Gormley no psychometrically determined instrument or in-depth qualitative studies have been used to measure patient satisfaction with outcome (Gormley 1997).

The difficulty this poses can be illustrated by reviewing the findings reported by Zorzos & Paterson. They found that 42% of women with an unsuccessful self-rated outcome gave a negative satisfaction rating (Zorzos and Paterson 1996). This indicates that even when the operation was considered unsuccessful by a woman — as judged by quality of life questions and whether she was still *troubled* by her bladder — she will, in many cases, rate the outcome as satisfactory on a dichotomous scale. This may mean: (a) that women regard an improvement

(but not a cure) as a successful outcome (which is consistent with the American Urological Association's position outlined above); (b) that satisfaction may reside with having had all that can be done (irrespective of the outcome); (c) that the dichotomous scale commonly used for patient satisfaction evaluation unduly restricts the responses that can be given; or (d) that the validity and reliability of a dichotomous measure is limited (the accepted practice in scale construction is to have at least three items). It may also be the case that there is a need to measure dissatisfaction as well as satisfaction.

Recent work by Williams supports the notion that negative concerns are not always reflected in patient satisfaction measures. His work highlighted the need for monitoring and responding to negative concerns raised by patients independently of their rated satisfaction with outcomes (Williams, Coyle et al. 1998). He suggested that *duty* and *culpability* act as filters of experiences in the individual's critical evaluation process. Consistent with this, another recent finding on dissatisfaction suggested that dissatisfaction is a separate and discrete construct from satisfaction; Mulcahy & Tritter stated that less of one does not necessarily mean more of the other; and reported that both may exist simultaneously as well as being discrete constructs (Mulcahy and Tritter 1998). They concluded that more research should investigate how people cope with dissatisfaction, thus implying support for Williams' suggestion that cultural filters may be operating (Williams, Coyle et al. 1998).

2.2.4 Laparoscopic studies

As the laparoscopic procedure is relatively new technology and the CLAIM trial was established to undertake a comparison between open and laparoscopic techniques, a separate investigation into patient satisfaction and laparoscopic surgery for stress urinary incontinence was undertaken. In three different retrospective review articles of laparoscopic procedures, patient satisfaction with outcome was rated. The rates of satisfaction were given as 87% (98 women) after an average of 8.4 (1-28) months follow-up (Cooper, Carlo et al. 1996), 94.83% (55 women) over a 6–22 month period (Liu 1993), and the third study indicated that 7 out of 8 women were satisfied at 3 months post-operation (Langebrenke, Dahlstrom et al. 1995). Little or no indication is provided in any of the studies as to how the women's satisfaction was measured. Of equal importance is that minor modifications in surgery were not consistently reported across studies, which may mean that comparison of different studies is limited. Specifically, one editorial comment indicated that when undertaking the laparoscopic Burch colposuspension procedure it is necessary to insert a standard two sutures as with the open procedure for comparison between these different techniques; it also stated that many surgeons only apply one in the laparoscopic technique (Dwyer 1996).

In a retrospective investigation by Lam *et al* it was reported that for all 15 patients treated, including two with pre-operative detrusor instability, all were continent, were satisfied with the operation, and there were no voiding difficulties (Lam, Jenkins et al. 1995). How satisfaction was measured is not indicated nor was it indicated how continence was determined. The follow-up period ranged from between six weeks to nine months which indicates that all women were not followed across the 2-year period commonly accepted as being needed for measuring final surgical treatment outcomes (Lam, Jenkins et al. 1995; Black, Griffiths et al. 1996; Leach, Dmochowski et al. 1997).

Another recent study of genuine stress urinary incontinence compared the outcomes from laparoscopic Burch colposuspension, abdominal colposuspension, and vaginal needle suspension. The women were asked if they would have the same operation again and if they were satisfied (Das 1998). At 36-months post-operation 90%, 60% and 60% were satisfied with the surgery, and 90%, 70% and 80% would opt for the same surgery again, respectively. As the participant numbers were small (10 in each group), and assignment to surgery was not random or blinded these results should be interpreted cautiously. The patient satisfaction questions asked were similar to those used in past urogynaecological studies and highlighted the importance researchers (and presumably clinicians) placed on an overall rating of satisfaction with treatment outcome.

In contrast to these studies, one study comparing laparoscopic with open surgery examined three dimensions of patient satisfaction (quality of care, hospital care, and physician time), with two ways of looking at outcomes: absolute (status at 6-months after surgery) and relative (difference between baseline and follow-up status) (Kane, Maciejewski et al. 1997). Each of the outcomes was related significantly to the satisfaction scales, which were psychometrically developed. Satisfaction was shown to be more closely associated with absolute outcomes, although none of the variables within the satisfaction scores explained more than 8% of the variance. This was consistent with previous findings in the patient satisfaction literature; that the construction of patient satisfaction measures remains elusive and may be relatively independent of clinical changes. This study also identified that patients preferred the open rather than the laparoscopic technique

*

In summary, the review of incontinence surgical treatment outcomes and the measurements of patient satisfaction indicated both the need to define incontinence by urodynamics testing, and how little understood are the determinants of patient satisfaction. Overall, this review of urinary incontinence patient satisfaction highlights the limited conceptualisation of what comprises a subjective 'satisfaction' outcome from a surgical procedure, as well as the lack of rigour in its measurement. In the studies reviewed, patient satisfaction was generally not investigated with regard to either validity or reliability. In similar reviews undertaken for surgical procedures for other urological conditions, no qualitative investigations of patient satisfaction or psychometrically determined measurements were found (Taylor, Beart et al. 1983; Marcello, Roberts et al. 1993; Garcia-Aguilar, Belmonte et al. 1996). Although the research field has placed importance on measuring satisfaction, whether this is congruent with how women perceive satisfaction (or dissatisfaction) with incontinence treatment is unknown and no rigorous or substantiated psychometric and qualitative investigations into patient satisfaction appears to have been conducted.

Thus there is a lack of reliable and valid measures which evaluate patient satisfaction with surgery that treats a condition which greatly impacts on the quality of a person's life. This implied the need for an instrument that could compare patients' satisfaction with the intervention and the outcome from a range of treatments for incontinence; an instrument which could be used in clinical practice and within clinical trials testing new procedures and technology.

2.3 The CLAIM trial and measuring patient satisfaction

Issues identified from the review of the incontinence literature need to be related to the CLAIM trial, as the trial procedures materially affected the investigation into patient satisfaction.

The CLAIM trial compared open and laparoscopic Burch colposuspension surgeries used to treat women with urodynamically demonstrated genuine urinary stress incontinence. The protocol for the trial indicated that the Burch colposuspension procedure was to be performed in a similar manner regardless of which technique was used. Two to three sutures, of specified type, were used and the position of suspension, the iliopectineal ligament, was identified. The suspensions were done without tension, and the third suture was placed only if a high cystocele was present. A catheter was inserted at the completion of the operation.

These details regarding the operative technique ensured that a consistent procedure was undertaken regardless of the technique used. The involvement of urodynamically identified genuine urinary stress incontinence patients ensured that women with a similar condition were being treated. All women in the trial were required to undergo physiotherapy and to have experienced no improvement in their incontinence problems to become eligible for surgery.

One of the CLAIM objectives was to assess the long-term subjective and objective success of the open and laparoscopic procedures.

2.3.1 Some issues for measuring patient satisfaction

The participant selection process for the CLAIM trial incorporated voluntary agreement. Voluntarism in clinical trials is known to be associated with lower morbidity and mortality rates when compared with those who do not volunteer, regardless of the treatment provided. It is also often associated with participant attributes such as age, sex, socioeconomic status and education (Aday 1996). Voluntarism may affect the measurement of patient satisfaction, making generalisations less valid with regard to the greater population of women with urinary stress incontinence.

The CLAIM trial was controlled and partially blinded, in that the patients and ward staff did not know which operation had been conducted until the dressings were removed. But there was no way of maintaining this blinding: the patients knew which treatment they had received at six months, when subjective outcome data were collected. To ensure objective data on patient satisfaction rating with knowledge of the procedure and/or with respect to time elapsed, measurements should be collected throughout the post-operative period of follow-up. This was not possible, as 25% of the women were 6 months or more post-operation at the commencement of negotiations to develop the patient satisfaction measure.

The collection of HRQoL information prior to surgery and at 6 months follow-up may have reduced the impact of memory inaccuracy and loss related to the retrospective collection of information. It would have therefore increased introspection about the value of the treatment. It is likely this increased validation in the development of the patient satisfaction instrument. Unfortunately, pre-surgery patient satisfaction information was not collected in the CLAIM trial. Hence, memory may have influenced the data on women's expectations of treatment outcome.

Another area of potential difficulty was dependency. The HRQoL data were collected by both practice nurses and clinicians. It was possible that information collected post-examination may have been influenced by comments made by the urologist, thereby confounding the subjective evaluations of the women with clinical outcomes, or through the women basing their responses

on beliefs induced by the clinician, rather than reflecting their own beliefs. The CLAIM trial requested that the patient satisfaction data be collected when a woman was 6 months post surgery; a timeframe which was incongruent with the literature regarding the timeframe necessary for long-term changes to have occurred. This indicates that the patient satisfaction scores reported in the current study represented intermediate rather than long-term evaluations.

Finally, one of the objectives of the CLAIM trial was to eliminate the need to conduct the urodynamics assessment post-surgery. As suggested in the literature review, this was not possible due to the different constructs being measured. Although clinical measures of outcome can be compared to patient satisfaction, only moderate correlations should be expected. In addition, the categorisation used to assess urodynamic results formed no logical scale, so comparison was meaningless. For example, '0 = no significant abnormality', '4 = genuine stress incontinence' yet '7 = voiding difficulty'. The other clinical outcome scales of stress incontinence, urinary incontinence, and self-described urgency after surgery do form logical scales, and can be used to ascertain criterion validity. For example, urinary incontinence was recorded as '0 = none', '1 = occasional' and '2 = frequent'.

*

In conclusion, a number of factors were linked to the design and implementation of the CLAIM trial that affected the interpretation of the psychometric analyses conducted on the patient satisfaction data. These were, in the main, outside the control of the researchers and their effect upon the findings is uncertain.

3 Study method

The purpose of this study was to develop a patient satisfaction instrument for use with women who have had surgery for incontinence; specifically an instrument that met with the standard psychometric criteria for latent variable measurement. The study objectives were:

- 1 To achieve ecological validity for all items in the instrument through development involving the women participants, the CLAIM trial members as experts in the incontinence treatment field, and through a critical review of the literature.
- 2 To achieve construct validity as identified through convergent-discriminant validity using standard psychometric procedures, and through comparison with other associated scales and variables.
- 3 To determine the internal consistency of the items used in the scale measuring patient satisfaction.
- 4 To determine the criterion validity of the resultant instrument against post surgical 6-month clinical outcome measures.

3.1 Research methods

The research methods are described under the headings recruitment, questionnaire development, construction and validation.

3.1.1 Recruitment

It was assumed that the CLAIM trial sample size ($n=200$) would be available for the construction study. However, only 75 women had reached 6 months post-operation when data were collected. The majority of these women were from the Royal Women's Hospital (48% (36)) and the Mercy Hospital (44% (33)). Seven participants resided outside the Melbourne metropolitan area, with 1 being from a state other than Victoria. As access to interpreters had been agreed by the CLAIM trial members² all women who were 6 months or more post-surgery (regardless of language spoken) were invited to participate. All 75 eligible women were posted a plain language statement outlining the research and inviting participation. A consent form was included; where a woman agreed to participate she was asked to sign the consent form and return it to the research team in a pre-stamped and addressed envelope. In addition, a randomly chosen sub-sample were asked to participate in a focus group assisting in the initial item development process. If no response was received by a set date, the woman was contacted by telephone to ascertain the reason, and where appropriate, the woman was encouraged to return the consent form.

3.1.2 Questionnaire development

Literature search and review

A literature search on Medline, Psychlit and The University of Melbourne's library catalogue was undertaken using the key words and combinations of "patient satisfaction, Burch colposuspension, and urinary stress incontinence". The references from the materials resulting from this search were inspected for other sources not identified in the electronic search. Following review of abstracts, all articles which seemed relevant were extracted. The resulting literature review is presented above in *Section 2 Background and Rationale*. The following five key component areas relating to the construct of patient satisfaction were identified:

- Structure: this includes physical surroundings and access issues (Donabedian 1980; Ware, Snyder et al. 1983; Carr-Hill 1992; Loeken, Steome et al. 1997; Sitzia and Wood 1997).
- Process: interpersonal aspects of care, perceived technical care, and information transfer (Donabedian 1980; Ware, Snyder et al. 1983; Carr-Hill 1992; Loeken, Steome et al. 1997; Sitzia and Wood 1997).
- Pain and discomfort: this included both physical and psychological aspects (Donabedian 1980; Ware, Snyder et al. 1983; Loeken, Steome et al. 1997).
- Satisfaction with outcome: incorporated here are expected future behaviours as well as current satisfaction and health status (Donabedian 1980; Ware, Snyder et al. 1983; Walker and Texter 1992; Hardy, West et al. 1996; Zorzos and Paterson 1996; Leach, Dmochowski et al. 1997; Litwiller, Nelson et al. 1997; Loeken, Steome et al. 1997).
- Expectations: this referred to expectations held by a patient in relation to the outcome of the procedure. It was included even though past research has shown that little variance within the construct of patient satisfaction is accounted for by this dimension. This was

² In the event these interpreters were generally unavailable at the time of data collection.

included to confirm the findings and to determine if the resulting instrument should be weighted (Donabedian 1980; Linder-Pelz 1982; Sitzia and Wood 1997; Williams 1997).

Focus groups

Following the literature review, focus group discussions were held with CLAIM trial members followed by the focus group with the women. This order provided participants with the opportunity to confirm or reject the component areas proposed from both the literature and the CLAIM trial members. Where concepts considered important to patient satisfaction were raised by the CLAIM trial members but not by the patients, these were explicitly raised and probed by the researchers. This order also permitted the testing of a first draft of items and discussion on response formats with the women during follow-up telephone calls. Concepts and logistic considerations in conducting the focus group discussions were based on theoretical and tested practices in the field (Morgan 1993). Discussion continued until saturation was reached, and all the component areas identified from the literature review and CLAIM trial members' suggestions had been covered.

Nine of the 11 members of the CLAIM trial team, surgeons and nurses, participated in the clinician focus group. The purpose was to ascertain the clinical and health care factors that the trial team believed would impact on the outcome of treatment and the women's satisfaction with this outcome. Participants were specifically asked to identify factors relating to the different surgical techniques that might impact on a woman's satisfaction with treatment outcome. A series of questions based on the possible component areas thought to be associated with patient satisfaction, identified in the literature review, were used as prompts to facilitate discussion. The focus group was taped and transcribed.

Ten women, at 6 months post-surgery, were randomly chosen within hospital strata. Six agreed to participate, but only five attended. In addition, two other women were interviewed in their homes due to transport difficulties, and one woman was interviewed through an interpreter. All eight participants gave permission to tape the discussion. The discussions were to determine the areas that were important to the women in evaluating their satisfaction with treatment outcomes. At the beginning of the interview the women were encouraged to relate their experiences of the entire process of deciding to have the surgery and participating in the trial. This was done in order to identify areas that might impact on their evaluation of the treatment outcome, but which were not incorporated into the prepared questionnaire. To test draft items, telephone follow-up interviews were conducted with all participants, except for the woman who spoke a language other than English; she did not feel that a telephone interview would be adequate and because the hospital interpreter service was not available she was unable to participate.

3.1.3 Construction and validation

Once refined the item bank was administered to CLAIM trial participants, along with other questions probing aspects of their treatment, their demographic details and health status.

The data were entered into EpiInfo (Dean, Dean et al. 1994), checked and verified prior to data analysis, which was undertaken with EpiInfo (Dean, Dean et al. 1994) and SPSS (SPSS 1998). The analyses were the conventional procedures used in instrument construction: item

examination, and scale construction using exploratory factor, internal consistency and logical analyses.

Once constructed, preliminary scale validation was carried out through comparison of the results with the other measures available in the study: incontinence measures (UDI, IIQ), a general health measure (SF-36), a HRQoL measure (AQoL) and a generic patient satisfaction questionnaire (PSS). In addition, GUTSS scores were also examined by changes in the level of incontinence ('relative incontinence') and the level of incontinence at 6 months post-operation ('absolute incontinence').

Accepted test size

Of major concern was the small numbers of participants in the study and the number of analyses undertaken on the dataset. Under these conditions where the conventional test size is accepted for significance testing, Type I errors may occur; *viz.*, where $\alpha = 0.05$, this assumes that 1/20 tests could be 'significant' by chance. In order to reduce the likelihood of Type I errors researchers may either adjust all analyses by the number of tests (using a standard adjustment procedure that estimates the appropriate test size), or they may make an *a priori* decision about the accepted test size. We adopted the latter course and set the test size at $\alpha = 0.01$ in order for significant associations to be accepted; *viz.*, under this test size 1/100 tests could be expected to be a Type I error due to chance. Where p-values between 0.01–0.05 were obtained these are reported as being 'suggestive' of a relationship.

The exception to this are the analyses reported in Table 13. The data in this table were from a different dataset to the data reported in all other analyses. Given that only four tests were conducted on this dataset the conventional test size ($\alpha = 0.05$) was accepted.

3.2 Criteria instruments

The instruments used to collect HRQoL information were the *Incontinence Impact Questionnaire* (IIQ) and the *Urogenital Distress Inventory* (UDI) (Uebersax, Wyman et al. 1995). The IIQ and UDI validity and internal-consistency reliability were tested by Shumaker *et al* and were found to be psychometrically strong with the exception of the stress subscale in the UDI which had a reliability of Cronbach $\alpha = 0.48$ (Shumaker, Wyman et al. 1994). The study did not, however, investigate test-retest reliability, although similar long-form scales tested much earlier by Wyman *et al* found values ranging from $\alpha = 0.52$ – 0.70 (Wyman, Harkins et al. 1987). In the latter study, the research population was white, well-educated, North American women thus making generalisations less valid for an Australian population with varying educational, cultural and ethnic backgrounds. Short forms of these two instruments were then developed by Uebersax et al, and significant correlations with the long-form subscales were found (Uebersax, Wyman et al. 1995). When reading the findings regarding the correlation between patients satisfaction and the UDI these caveats should be borne in mind.

In addition, a measure of general health status, the SF-36 was also used (Ware, Snow et al. 1993). The SF-36 is a generic multi-dimensional health status measure that comprises eight scales measuring Physical Function, Role Physical, Bodily Pain, General Health, Vitality, Social Function, Role Emotional, and Mental Health. These scales are collapsed into two summary scales, the Physical Component Scale (PCS) and the Mental Health Component Scale (MCS).

The scores on the eight scales are transformed into 0–100 point scales, where 100 represents the best health state. The PCS and MCS are presented as T-scores, with a mean of 50 and a standard deviation of ± 10 (Ware, Kosinski et al. 1994). This report only presents results at the PCS and MCS levels due to sampling limitations.

The *Assessment of Quality of Life (AQoL)* was also administered to participants. This is a newly developed HRQoL utility instrument, which was designed to overcome some of the limitations of earlier instruments (Hawthorne and Richardson 1995). The AQoL descriptive system comprises 15 items in 5 dimensions. Item responses are all ordinal scales with four levels per item. The dimensions are Illness, Independent Living, Social Relationships, Physical Senses and Psychological Wellbeing (Hawthorne, Richardson et al. 1999). The utility weights were derived from an Australian population sample using time-trade off (TTO). During the calculation of the utility index, the Illness dimension score is not used. A multiplicative function is used to combine the remaining four dimensions into the utility index. The range of scores is between -0.04 (health states worse than death) to 0.00 (death) to 1.00 (full health) (Hawthorne, Richardson et al. 2000). The higher the score the better the HRQoL.

Finally, the *Patient Satisfaction Survey (PSS)* scale was used as a generic patient satisfaction instrument (Hawthorne 1993). This comprises five items probing a person's concerns following treatment with their physical health, ability to perform their everyday tasks, their emotional health, personal relationships and other general health issues. Although this instrument has not been extensively validated, a pilot study of 25 depressed women participating in a psycho-social support group provided a good internal consistency estimate ($\alpha = 0.74$), while when the PSS was administered to 21 mothers experiencing difficulties after childbirth, the results suggested the PSS had good test-retest characteristics (Spearman-Brown $r = 0.75$) (Rowe, Temple et al. 1996).

4 Construction of the GUTSS

4.1 Focus group findings

CLAIM trial members identified the following areas that they believed affected satisfaction with the outcome of the treatment:

- Care within the hospital and from clinicians might impact on satisfaction with treatment outcome. This included that patients were in receipt of a lot of attention, even though the non-private patients within the trial did not have continuity of clinician.
- Related urogynaecological conditions might affect the women's satisfaction, regardless of the clinical outcome on urinary stress incontinence. That is, prolapse or urge incontinence would not be addressed by the operation, but the operation might still be perceived as unsuccessful if no positive alteration to these conditions occurred.
- Prior expectations were believed to be a large factor contributing to the patients' satisfaction with treatment outcome.

-
- The number of items on the questionnaire needed to be minimal to ensure participation . Thirty items was thought to be too many. The questionnaire should be suitable for use as a self-administered questionnaire or by an interviewer over the telephone.
 - An instrument that could be used across the field of urogynaecological procedures was requested, as it would be used in other trials.

An initial draft of items was generated using the component areas that CLAIM trial members thought contributed to women’s satisfaction with the outcome of the procedure, and based on the five possible component areas identified from the literature review. Each member of the team was asked to provide comments on wording, location and content of the questions. Adjusted items were then trialed with the subsample of women who formed the participant focus group. The findings from the focus group of women and the interviews revealed their concerns were similar to those previously reported in population surveys and in-depth interviews with women who had not sought treatment (Ashworth and Hagan 1993). For example,

“It was a constant worry (I) would not leave the house without a pad on, worried about not making it to a toilet or no toilet being there... ”

Women also commented on how difficult it was to get information, and one woman stated:

“... it’s criminal that those TV ads for pads don’t have a message saying see your doctor for an operation as an alternative.”

When asked to consider what determined their satisfaction with the outcome of the surgery the women identified the following areas of concern:

- Adverse effects: these related not only to the surgical procedure, which were consistent with past findings of complications resulting from the Burch colposuspension technique, but extended to infections resulting from the urodynamics assessment:
“I had to have that test before and then three times after the operation and had to keep going back because of infections”
- Confidence in the long-term outcome: some women commented that because they had only recently had their 6-month follow-up consultation and testing they felt more confident about the longevity of the outcome. The comments from women also confirmed past reports that women were satisfied with the outcome of the surgical intervention even if improvement was only partial (Leach, Dmochowski et al. 1997).
- Care received: the women were quick to distinguish between hotel services and care/treatment. Hotel services — the facilities and food for example — were consistently described as not significant in evaluating satisfaction with the outcome of treatment, whereas information provision, follow-up services and discharge-related issues were considered important. Regarding the latter some women believed that being discharged with a catheter in place may have reduced the benefit of the operation.

“I don’t think they should send you home with a catheter. It was too hard; I had problems after...”

Some women noted that the level of care in the hospital would have influenced their satisfaction with the treatment outcome more if discussion had taken place within the month after the operation, particularly during the first week after the operation.

- Expectations: all women confirmed that they would not have had the operation unless they had felt moderately *hopeful* that the outcome would be successful.
- Health-related quality of life: most women related stories about improved HRQoL since the operation, the most frequent being not having to wear pads or not to use as many. The issue of sexual activity before and after the operation was raised by the participants in the focus group discussion. It was reported to be an important consideration in deciding their level of satisfaction, but cautioned against asking all women how the operation had impacted on their sex life.
- Pain and discomfort: interestingly, all women regardless of operation technique (open or laparoscopic) mentioned that early in their recovery they had felt pain and/or discomfort associated with the scar(s) from the surgery. One woman expressed gratitude about the minimal scarring from the laparoscopic procedure. One woman, who still had an infection, identified strongly with judging satisfaction in relation to pain and discomfort. Other women expressed the belief that if you have an operation you have to expect a certain amount of pain and discomfort.

In response to questioning in relation to participation in the trial, response scales to be used and draft item testing, the following summary of responses was obtained.

- Participation in the trial: some women were very happy to participate and saw their participation as helping other women. Others were not, as they had not been informed prior to the operation that they were involved in a trial. Some felt that they would have had a better outcome with an alternative procedure. It is important to note here that all participants were sent a brief written explanation about the CLAIM trial, which formed part of the consent form that they were required to sign to enrol in the trial.

With respect to the testing of draft items for the questionnaire, the women rejected two items as being repetitive. Some commented that they felt there should be room for comments at the end. They preferred discussing the areas of concern and satisfaction with the researcher rather than answering a questionnaire. Minor word changes were made to some items to correlate with words more commonly understood by the patients.

4.2 Questionnaire and item description

From the literature review and focus group discussion process a total of 22 items were written *de novo* for trialing in the final telephone questionnaire, 7 items had a filter question with a ‘Yes/No’ option preceding the item of interest. The items covered the component areas of *adverse effects and pre-existing conditions* (ie. conditions other than genuine urinary stress incontinence); *care*

received which incorporated the constructs of structure and process from the literature review; *pain and discomfort, satisfaction with outcome; and expectations.*

To minimise response set bias (Aday 1996), seven items were written so they needed to be reversed prior to scoring. Although a variety of response sets were used, all items used a 5-point response scale. A further 14 questions were asked to identify determinants that might contribute to patient satisfaction and to the effect of the use of the questionnaire and participation in the trial. For the purposes of validation, two other instruments were included: (a) the *Patient Satisfaction Survey* (PSS) scale (Hawthorne 1993) and (b) the *Assessment of Quality of Life* (AQoL) questionnaire respectively (Hawthorne, Richardson et al. 1997).

4.3 Participant details

Of the 75 women contacted 58 (77%) consented to be interviewed. Of the 17 not consenting, 2 had moved and were not contactable; 1 was unavailable during the study timeframe; 12 did not return a consent form after follow-up telephoning and 2 women returned their form indicating that they did not wish to participate. Of the women who returned a consent form, 45 (78%) were interviewed; this was primarily due to the difficulty in obtaining interpreters for 7 women (1 Spanish, 4 Italian, 1 Greek, and 1 Turkish). Three telephone numbers had been disconnected, one woman was overseas, one did not provide a telephone number, and another was on holidays. This small number meant the cell sizes within the factor analyses during construction of the patient satisfaction measure were marginal to allow valid statistical conclusions to be drawn (Guadagnoli and Velicer 1988).

The basic socio-demographic details of participants are provided in Table 1. Generally, the women were middle-aged, born in Australia, had high school education and came from households with lower incomes.

Table 1: Details of the women interviewed

| | | <i>N. (a)</i> | % |
|------------------|-------------------|---------------------|-----|
| Age (mean, sd) | | 54 years (11 years) | |
| Country of birth | Australia | 30 | 67% |
| | Overseas | 15 | 33% |
| Education level | Primary | 9 | 20% |
| | High | 21 | 47% |
| | Technical/Further | 3 | 7% |
| | University (b) | 12 | 27% |
| Income (c) | \$15,000 or less | 17 | 38% |
| | \$16,000–\$25,000 | 8 | 18% |
| | \$26,000–\$35,000 | 5 | 11% |
| | \$36,000–\$45,000 | 6 | 13% |
| | \$46,000+ | 6 | 13% |

Notes:

a = Except where otherwise stated

b = Including nursing qualifications

c = Data missing from 3 cases

Parity, on average, was 2.8 (range: 1–6 children), most were pre-menopausal and on hormone replacement therapy. Regarding their incontinence, most were suffering stress incontinence and urinary incontinence. Urodynamic diagnosis indicated that most were suffering genuine stress incontinence. Regarding the surgical procedure undergone, the women were evenly split between open surgery and laparoscopic surgery. Finally, most of the women were public patients.

Table 2: Health status of participants

| | | <i>N. (a)</i> | % |
|--------------------------|-----------------------------|---------------|-----|
| Parity | 1 | 5 | 13% |
| | 2 | 11 | 28% |
| | 3 | 16 | 40% |
| | 4+ | 8 | 20% |
| Menstrual status | Pre-menopause | 20 | 51% |
| | Menopausal | 17 | 44% |
| | Post-menopause (b) | 2 | 5% |
| HRT (c) | Yes | 36 | 84% |
| | No | 7 | 16% |
| Stress incontinence | None | 1 | 3% |
| | Occasional | 5 | 14% |
| | Frequent | 30 | 83% |
| Urinary incontinence | None | 7 | 22% |
| | Occasional | 9 | 28% |
| | Frequent | 16 | 50% |
| Urodynamic diagnosis (d) | No significant abnormality | 1 | 2% |
| | Hypersensitive bladder | 2 | 4% |
| | Genuine stress incontinence | 42 | 84% |
| | Detrusor instability | 5 | 10% |
| Procedure | Open procedure | 21 | 49% |
| | Laparoscopic procedure | 22 | 51% |
| Type of patient | Public | 40 | 89% |
| | Private | 5 | 11% |

Notes:

a = Except where otherwise stated

b = Or hysterectomy

c = Hormone replacement therapy

d = Women could be diagnosed with up to three conditions. In the sample, seven women had multiple conditions.

Table 3 shows the 6-month follow-up from surgery. This reveals that approximately one-quarter of the women were still diagnosed with genuine urinary stress incontinence by the urodynamics assessment, and that about one-third reported urinary and urge incontinence.

Table 3: Post surgical clinical measures

| | | <i>N.</i> | % |
|--------------------------|-----------------------------|-----------|-----|
| Stress incontinence | None | 33 | 87% |
| | Occasional | 5 | 13% |
| Urinary incontinence | None | 26 | 68% |
| | Occasional | 8 | 21% |
| | Frequent | 4 | 11% |
| Urge incontinence | None | 23 | 61% |
| | Occasional | 13 | 34% |
| | Frequent | 2 | 5% |
| Urodynamic diagnosis (a) | No significant abnormality | 24 | 50% |
| | Hypersensitive bladder | 2 | 4% |
| | Genuine stress incontinence | 12 | 25% |
| | Detrusor instability | 4 | 8% |
| | Voiding difficulty | 6 | 13% |

Note:

a = Women could be diagnosed with up to three conditions. Six women were diagnosed post-operation with multiple conditions

4.4 Scale development: reliability and content validity

4.4.1 The theoretical models

Since there was uncertainty arising from differences between theoretical models of patient satisfaction and the findings of the focus group interviews, two different models of satisfaction were constructed during analysis. It was hypothesised that for each of the component areas (adverse effects and pre-existing conditions, care received, pain and discomfort, satisfaction with outcome, and expectations) there would be at least one item in the final instrument.

Model A. A single-dimension model of satisfaction incorporating items from all six component areas. The assumption here was that the six areas would be related. Areas which were not related would be excluded from the scale in order to minimise measurement error.

Model B. A multi-dimensional model, in which any of the six areas could be subsumed within a single unidimensional scale representing that area, or *n...* number of unidimensional scales where the items covered several of the areas (this model assumed that some areas would be related to others; eg. expectations with outcome satisfaction etc.).

4.4.2 Item properties

Responses to the 22 items were examined, and are presented in Table 5. This shows the content area each item represented, item content, means and standard deviations. Negative items were reversed prior to data analysis and where filter questions (“Yes”/“No”) were used those responding “No” were assigned the best health state described in the linked 5-point scale. This was done following investigation of the alternative scoring method, where filter “No” responses were assigned a value better than the best health state described in the 5-point scale. Factor analysis of the two scoring methods revealed no difference in the proportion of variance explained by the two methods.

For the purposes of instrument construction, it is generally accepted that items should discriminate between individuals through adequately covering the potential range. In this case all items were deemed adequate with the exception of Item 17 (where the actual range was under 50% of the potential range).

Turning to item means we adopted a cut-off rule where the means were proportional to the range and should not be located within 0.4 standard deviations of a scale endpoint. Where this phenomenon occurs it indicates that respondents are selecting an endpoint value and the item is displaying a severe ceiling or floor effect. The poor items in this case were Q17 and Q18.

Finally, the standard deviations were inspected. Narrow standard deviations suggest that items fail to discriminate between individuals (ie. everyone tends to select the same value). On a 5-point scale a commonly used standard is that the standard deviations should exceed 0.5. Question 17 was the only item failing this test, although problems may have been apparent with Q12, Q18 and Q36.

Based on item inspection, Q17 and Q18 were the potential items to be discarded. In terms of item content, this would have left no items probing scarring. Since it seemed to be an important issue, as per the models outlined above, these items were retained.

Table 4: Analysis of all items

| <i>Items (a)</i> | <i>Content</i> | <i>Content area (b)</i> | <i>Negative/ Positive (N/P) Item</i> | <i>Potential Range</i> | <i>Actual Range</i> | <i>Mean</i> | <i>Sd</i> |
|------------------|--|-------------------------|--------------------------------------|------------------------|---------------------|-------------|-----------|
| Q9 | General health rating | CH | P | 1—5 | 100% | 2.02 | 0.92 |
| Q10 | Prior expectations of operation | EX | P | 1—5 | 60% | 1.71 | 0.66 |
| Q11 | Prior information about the operation | CR | P | 1—5 | 100% | 2.07 | 1.03 |
| Q12 | Prior expectations of outcomes | EX | P | 1—5 | 60% | 1.82 | 0.54 |
| Q13 | Satisfaction with operation | SO | P | 1—5 | 100% | 1.91 | 1.18 |
| Q14 | Attitudes/Behaviours of doctors & nurses | CR | P | 1—5 | 80% | 1.49 | 0.84 |
| Q16* | Effect of operation on intimate relationships | CH | P | 1—5 | 100% | 1.69 | 1.15 |
| Q17 | Scar from the operation being painful | PD | N | 1—5 | 40% | 1.20 | 0.41 |
| Q18 | Scar bothers you | PD | N | 1—5 | 80% | 1.13 | 0.51 |
| Q19 | Length of stay in hospital | AE | P | 1—5 | 80% | 3.64 | 0.77 |
| Q21* | After operation, problems with water works/prolapse | AE | N | 1—5 | 83% | 1.87 | 1.10 |
| Q22 | Recommend operation to friends | SO | P | 1—5 | 100% | 1.89 | 1.40 |
| Q24* | Disappointed with operation outcome | SO | N | 1—5 | 100% | 1.69 | 1.22 |
| Q26* | Do you have other problems (water works/prolapse) the operation was not meant to fix | AE | N | 1—5 | 100% | 1.33 | 0.91 |
| Q28* | Do you have different/new problems from the operation | AE | N | 1—5 | 100% | 1.71 | 1.10 |
| Q30* | Has discomfort limited your life enjoyment | PD | N | 1—5 | 80% | 1.38 | 0.75 |
| Q31 | Satisfaction with doctor's explanations about results of operation | CR | P | 1—5 | 100% | 2.29 | 1.33 |
| Q32 | Rest of life with problem as is now | CH | P | 1—5 | 100% | 2.22 | 1.36 |
| Q33 | Satisfaction with outcome of operation on personal relationships | CH | P | 1—5 | 100% | 1.98 | 1.03 |
| Q34 | How happy with the care received in hospital | CR | P | 1—5 | 100% | 1.71 | 1.12 |
| Q35* | Pain from the operation | PD | N | 1—5 | 60% | 1.24 | 0.53 |
| Q37 | Happy with the effect of the operation | SO | P | 1—5 | 100% | 1.93 | 1.18 |

Notes:

a = Question number on the questionnaire.

b = CH = Current health status and living, EX = Expectations, SO = Satisfaction with outcomes, PD = Pain and discomfort, CR = Care received, AE = Adverse effects and pre-existing conditions.

* = Asterisks show items with a dichotomous filter.

4.4.3 Data reduction and weighting

Conventional procedures for data reduction were employed using exploratory factor analysis (both principal components and varimax rotation), reliability analyses (using item-rest-of-test and

Cronbach α), and logical item content analysis to ensure that the component areas comprising patient satisfaction were adequately represented. An iterative procedure involving all three methods was used until the most parsimonious solution was derived consistent with the measurement principles outlined by Thurstone, Rummel and Pedhazur and Schmelkin when developing scales using factor analytic techniques (Thurstone 1947; Rummel 1970; Pedhazur and Schmelkin 1991). This resulted in two solutions, matching the two hypothesised models outlined above. These are described as Model A and Model B respectively.

To control for expectancy bias, two of the items (Q10 and Q12) probed respondents' expectations: one covered expectations of the intervention itself and the other expectations of the outcome. Responses to these items were combined to form an expectancy score on a scale of 0.00–1.00, where 1.00 represented a highly favourable expectation.

This expectancy score was then used to weight responses to other items. When comparisons of derived scales, as described below, were made using weighted and unweighted items very few differences occurred: the factorial structures were the same; and there was very little difference in reliability or explained variance estimates. For example, for Model A (Table 5) the effect of weighting was to alter the eigenvalue from 4.51 to 4.53, so that the percentage of variance explained rose from 75.1% to 75.6% and the Cronbach α from 0.93 to 0.94. Weighting was therefore discarded, and the two items pooled with the database.

4.4.4 Model A: A unidimensional scale

Table 5 shows the derived unidimensional scale. This comprised six items, covering the component areas of satisfaction with the outcomes (Q37, Q22 and Q24), the effect on current health status and living (Q33 and Q32) and the presence of adverse effects and pre-existing conditions (Q21). The psychometric data in Table 5 suggests these comprised a unidimensional scale explaining 75% of the variance.

The very high Cronbach α suggests there may be some redundancy in the scale, although the alpha-if-item-removed statistic (not shown in the table) suggested each item was making a unique contribution. If any item were to be considered for removal it would perhaps be Q22, on content grounds, thereby retaining the breadth of content.

Table 5: Structure of the unidimensional satisfaction scale, Model A

| <i>Item</i> | <i>Factor analysis</i> | | | <i>IRT (a)</i> | <i>Cronbach a</i> |
|---|------------------------|--------------------|-----------------------------|----------------|-------------------|
| | <i>PC (b)</i> | <i>Eigen-value</i> | <i>% variance explained</i> | | |
| Happy with the effect of the operation | 0.92 | | | 0.88 | |
| Satisfaction with outcomes on personal relationships | 0.87 | | | 0.80 | |
| Living the rest of life with the condition as it is now | 0.87 | | | 0.80 | |
| Post-operation, continuing problems with incontinence | 0.86 | | | 0.80 | |
| Prepared to recommend the operation to friends | 0.84 | | | 0.76 | |
| Disappointed with the operation outcome | 0.83 | 4.51 | 75.1% | 0.77 | 0.93 |

Notes:

a = Item-rest-of-test correlation

b = Principal components analysis

4.4.5 Model B: The multidimensional scale

Table 6 presents a more complicated instrument comprising two factors. In Table 6 the properties of the whole instrument are presented, assuming the computation of a single score (justified by the principal components loadings: all items load greater than 0.30). This assessment was further augmented by the satisfactory IRT loadings and the Cronbach α , which was excellent for this type of instrument (suggesting breadth of measurement without redundancy). As shown, the instrument model explained 72% of the variance.

The factor analysis varimax rotation showed the presence of two factors within the instrument. That there were no cross-loadings suggested that the measurement was not confounded and that each factor possessed statistical independence, allowing them to be scored separately.

Table 6: Structure of the multi-dimensional satisfaction scale, Model B

| <i>Items</i> | <i>Factor analysis</i> | | | | <i>IRT (c)</i> | <i>Cronbach a</i> | |
|--|------------------------|--------------------|-----------|---------------------|----------------|-------------------|-----------------------------|
| | <i>PC (a)</i> | <i>Varimax (b)</i> | | <i>Eigen-values</i> | | | <i>% Variance explained</i> |
| | | <i>F1</i> | <i>F2</i> | | | | |
| Happy with the effect of the operation | 0.91 | 0.90 | | | 0.79 | | |
| Satisfaction with operation | 0.87 | 0.92 | | | 0.70 | | |
| After operation, problems with water works/ prolapse | 0.81 | 0.90 | | | 0.63 | | |
| Disappointed with operation outcome | 0.81 | 0.89 | | 3.79 | 47.4% | 0.62 | |
| Satisfaction with doctor's explanations about results of operation | 0.58 | | 0.71 | | | 0.52 | |
| How happy with the care received in hospital | 0.44 | | 0.85 | | | 0.40 | |
| Attitudes/Behaviours of doctors & nurses | 0.51 | | 0.79 | | | 0.47 | |
| Prior information about the operation | 0.34 | | 0.69 | 1.98 | 24.7% | 0.32 | |
| Total (whole scale) | | | | | 72.1% | 0.83 | |

Notes:

a = Principal components analysis

b = Loadings <0.30 not shown

c = Item-rest-of-test correlation

The properties of each of the sub-scales are presented in Table 7. For *Factor 1: Outcome satisfaction*, the item content measures satisfaction with the outcome from the intervention (Q13, Q37 and Q24), and the presence of adverse effects and pre-existing conditions (Q21). The principal components analysis showed these items comprised a single factor, explaining 86% of the variance. The high Cronbach α (0.93) suggests the factor may possess some redundancy.

Regarding *Factor 2: Care satisfaction*, the content is concerned with issues around the care received (Q11, Q14, Q31 and Q34). Again, these comprised a single factor, explaining 59% of the variance. The Cronbach α (0.76) was, perhaps, a little low for this type of factor, but was within conventional limits.

Table 7: Internal structure of the Model B scale

| <i>Items</i> | <i>Factor analysis</i> | | | <i>IRT (a)</i> | <i>Cronbach a</i> |
|--|------------------------|--------------------|-----------------------------|----------------|-------------------|
| | <i>PC (b)</i> | <i>Eigen-value</i> | <i>% variance explained</i> | | |
| <i>Factor 1: Outcome satisfaction</i> | | | | | |
| Satisfaction with operation | 0.93 | | | 0.88 | |
| Happy with the effect of the operation | 0.93 | | | 0.87 | |
| After operation, problems with water works/prolapse | 0.90 | | | 0.82 | |
| Disappointed with operation outcome | 0.89 | 3.34 | 83.5% | 0.81 | 0.93 |
| <i>Factor 2: Satisfaction with care received</i> | | | | | |
| How happy with the care received in hospital | 0.85 | | | 0.62 | |
| Attitudes/Behaviours of doctors & nurses | 0.80 | | | 0.59 | |
| Satisfaction with doctor's explanations about results of operation | 0.75 | | | 0.55 | |
| Prior information about the operation | 0.68 | 2.37 | 59.3% | 0.51 | 0.76 |

Notes:

a = Item-rest-of-test correlation

b = Principal components analysis

4.5 Scoring the GUTSS

Prior to scoring the GUTSS, two items needed to have values imputed and the scores reversed.

- Items Q21 and Q24 (see Table 4) needed to have values imputed for women who did not have any symptoms of incontinence (Q21) or who were not at all disappointed with the outcomes of surgery (Q24). In each case the imputed values were set at the highest score; ie. a raw value of '5' was assigned.
- Items Q21 and Q24 also needed the scoring reversed.

Once these were completed, the procedures below were used to compute GUTSS scores for both models.

4.5.1 Model A

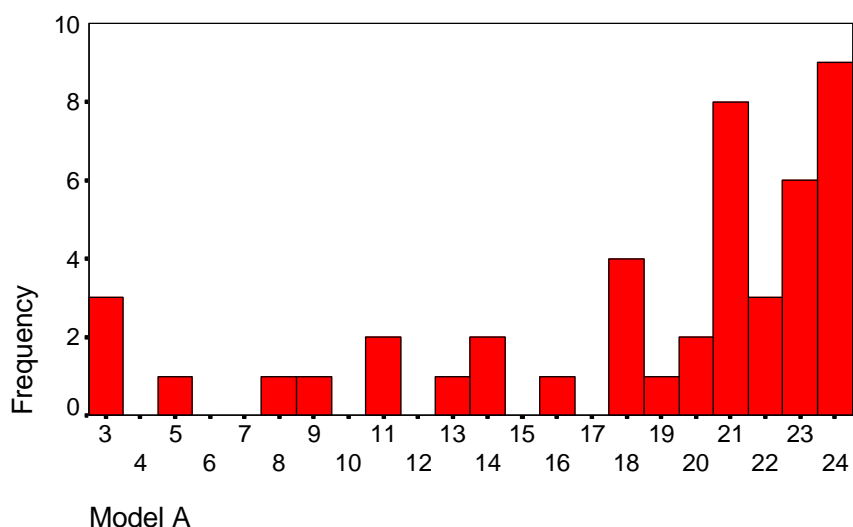
Scores on Model A were computed by simply adding the item values. The raw score was placed on a possible scale from 0–24 by subtracting 6, the lowest possible score, and then reversing the score so that a high score on the scale corresponded to a high level of satisfaction. The formula, using the item numbers from Table 4:

$$Model\ A = 24 - [\sum (Q21, Q22, Q24, Q32, Q33, Q37) - 6]$$

Equation 1

For example, if a woman scored '9' as a raw score her score would be: $24 - (9-6) = 21$. The results are presented in Figure 1. The mean score was 18.4 (± 6.4), the median was 21, and the range was from 3–24. This indicates a skewed distribution towards women being highly satisfied.

Figure 1 Model A histogram



4.5.2 Model B

Again simple summation was used for each of the sub-scales. These were then added together to form the GUTSS score. For each of the sub-scales, 4 was subtracted and the scores reversed, thus providing scores within the range 0–16. For Model B the range therefore was 0–32. The formulae were:

$$OS = 16 - [\sum(Q13, Q21, Q24, Q37) - 4] \quad \text{Equation 2}$$

$$CS = 16 - [\sum(Q11, Q14, Q31, Q34) - 4] \quad \text{Equation 3}$$

$$\text{Model B} = OS + CS \quad \text{Equation 4}$$

Figure 2 shows the distribution of scores. The mean score was 25 (± 6.1), the median 27, and the range was from 5 to 32. The figure also suggests that the distribution of scores was not as skewed as Model A (see Figure 1). Figures 3 and 4 show the distribution of scores for each of the two sub-scales, satisfaction with outcomes and satisfaction with care respectively.

Figure 2 **Model B histogram**

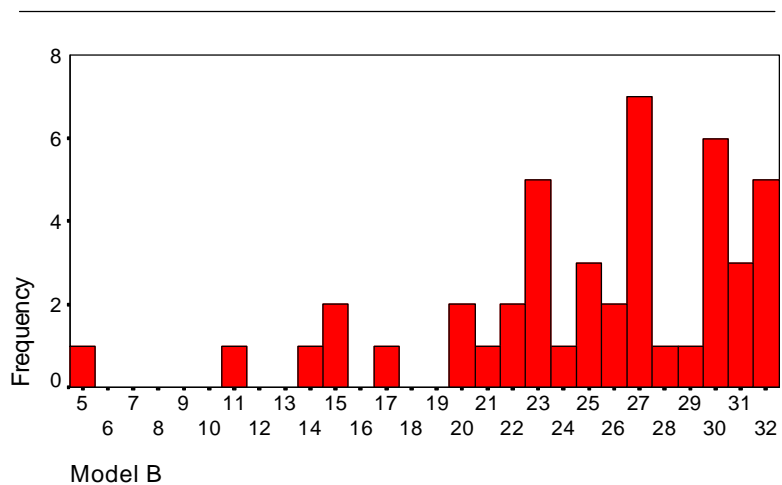


Figure 3 Model B Outcome satisfaction histogram

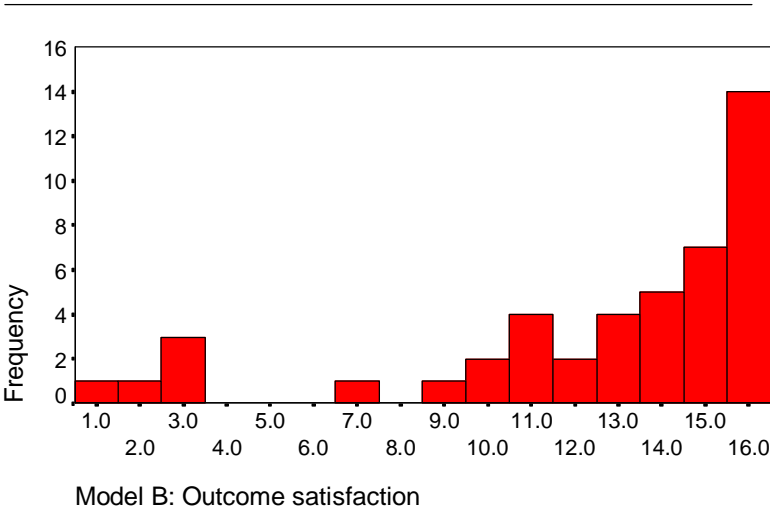
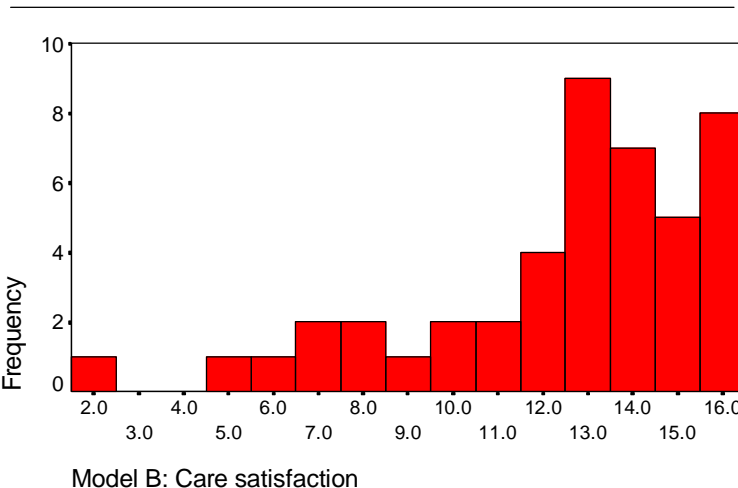


Figure 4 Model B Care satisfaction histogram



4.6 Validation of the GUTSS scales

4.6.1 Construct validation: convergent-discriminant validity

The GUTSS was assessed against the other instruments used in the CLAIM trial. For a description of these, see Section 3.2. It was hypothesised that the GUTSS Model A, Model B and Model B *Factor 1 Outcome satisfaction* would correlate highly with the UDI and the IIQ, but that Model B, *Factor 2 Care satisfaction* would not since it was measuring a different construct. With regard to the SF-36 it was hypothesised there would be moderate correlations with the two summary scales. A similar hypothesis was made for comparison with the AQoL utility scores. The GUTSS was expected to correlate more highly with the PSS, because this was a generic patient satisfaction scale.

Table 8: Pearson correlations between Model A and Model B and the UDI, IIQ, SF-36, AQoL & HPS

| | <i>Model A</i> | | <i>Model B</i> | | | | | |
|----------------|----------------|----------|--------------------------|----------|--------------------|----------|-----------------|----------|
| | | | <i>Model B (F1 + F2)</i> | | <i>F1: Outcome</i> | | <i>F2: Care</i> | |
| | <i>r</i> | <i>p</i> | <i>r</i> | <i>p</i> | <i>r</i> | <i>p</i> | <i>r</i> | <i>p</i> |
| UDI | -0.70 | <0.01* | -0.67 | <0.01* | -0.66 | <0.01* | 0.17 | 0.01* |
| IIQ | -0.72 | <0.01* | -0.67 | <0.01* | -0.68 | <0.01* | 0.15 | 0.02 |
| SF-36 Physical | 0.33 | 0.04 | 0.32 | 0.06 | 0.24 | 0.16 | 0.09 | 0.07 |
| SF-36 Mental | 0.59 | <0.01* | 0.45 | <0.01* | 0.55 | <0.01* | 0.01 | 0.43 |
| AQoL utility | 0.32 | <0.01* | 0.28 | <0.01* | 0.23 | <0.01* | 0.04 | <0.01* |
| HPS | 0.65 | <0.01* | 0.51 | <0.01* | 0.59 | <0.01* | 0.03 | 0.25 |

Notes:

* = significant difference, $p < 0.05$

Table 8 shows the results. The strongest correlations were with the UDI and IIQ for each of Models A and B. Moderate correlations were also found with the two summary scales of the SF-36 and with the AQoL. The correlations with the PSS were moderate to high.

Turning to the two sub-scales for Model B, for *Factor 1 Outcome satisfaction* the correlations were as mainly as expected: there were strong correlations with the UDI, IIQ and HPS, moderate correlations with the SF-36 Mental Health summary score and lower correlations with the SF-36 Physical Health summary score. The lower correlations with the AQoL were unexpected. Model B, *Factor 2 Care satisfaction* was moderately correlated with the UDI, IIQ and SF-36 Physical Health summary scale, and weakly correlated with the AQoL, HPS and SF-36 Mental Health summary scale.

Table 9 shows the relationship between the GUTSS Model A and the demographic characteristics of participants. It was expected that there would be no significant association between these variables and the GUTSS on the argument that satisfaction with treatment

outcomes or care should be independent of these factors. As shown in the table, generally, this was the case. The only variable on which GUTSS Model A scores significantly varied was education: women who were more highly educated obtained scores suggesting they were significantly more satisfied. Table 10 shows the same analysis, but for GUTSS Model B. Unlike Model A, there was no significant difference by education level. These findings suggest that background factors played little part in determining the level of satisfaction expressed by the women participating in the CLAIM trial.

Table 9: Criteria evidence for the GUTSS Model A

| | | <i>Median</i> | <i>IQR (a)</i> | <i>Statistics (b)</i> |
|------------------------|---------------------------|---------------|----------------|-----------------------------|
| Age group | <=50 years | 21.0 | 10.0 | $\chi^2 = 0.41, p = 0.52$ |
| | 51+ years | 21.0 | 7.5 | |
| Education level | Primary/High | 18.5 | 12.0 | $\chi^2 = 7.40, p < 0.01^*$ |
| | Technical/University | 22.0 | 3.0 | |
| Income | \$35,000 | 20.0 | 10.5 | $\chi^2 = 2.17, p = 0.14$ |
| | \$36,000 | 22.0 | 2.8 | |
| Parity | 1-2 births | 20.5 | 13.8 | $\chi^2 = 1.71, p = 0.19$ |
| | 3+ births | 21.0 | 5.8 | |
| Menstrual status | Pre-menopause | 21.0 | 6.5 | $\chi^2 < 0.01, p = 0.93$ |
| | Menopausal/Post-menopause | 21.0 | 10.0 | |
| HRT | Yes | 21.0 | 9.8 | $\chi^2 = 0.09, p = 0.77$ |
| | No | 21.0 | 5.0 | |
| Incontinence level (c) | None/Occasionally | 21.5 | 9.3 | $\chi^2 = 0.20, p = 0.66$ |
| | Frequently | 21.0 | 6.0 | |
| Procedure | Open procedure | 21.0 | 10.5 | $\chi^2 = 0.09, p = 0.77$ |
| | Laparoscopic procedure | 21.0 | 5.3 | |

Notes:

a = Inter-quartile range

b = Kruskal-Wallis 1-way ANOVA

c = Based on combined diagnoses of 'None', 'Occasional', 'Frequent' for stress incontinence and urinary incontinence

* = Significant difference, $p < 0.05$

Table 10: Criteria evidence for the GUTSS Model B

| | | <i>Median</i> | <i>IQR (a)</i> | <i>Statistics (b)</i> |
|------------------------|---------------------------|---------------|----------------|---------------------------|
| Age group | <=50 years | 26.0 | 10.0 | $\chi^2 = 0.37, p = 0.55$ |
| | 51+ years | 27.0 | 7.0 | |
| Education level | Primary/High | 25.5 | 10.0 | $\chi^2 = 0.68, p = 0.41$ |
| | Technical/University | 27.0 | 7.0 | |
| Income | \$35,000 | 26.5 | 8.3 | $\chi^2 < 0.01, p = 0.98$ |
| | \$36,000 | 26.5 | 6.5 | |
| Parity | 1-2 births | 24.0 | 12.8 | $\chi^2 = 1.70, p = 0.19$ |
| | 3+ births | 27.0 | 7.0 | |
| Menstrual status | Pre-menopause | 27.0 | 7.8 | $\chi^2 = 0.26, p = 0.61$ |
| | Menopausal/Post-menopause | 27.0 | 8.0 | |
| HRT | Yes | 26.0 | 8.0 | $\chi^2 = 0.89, p = 0.35$ |
| | No | 27.0 | 8.0 | |
| Incontinence level (c) | None/Occasionally | 27.0 | 9.8 | $\chi^2 = 0.11, p = 0.74$ |
| | Frequently | 27.0 | 7.0 | |
| Procedure | Open procedure | 25.0 | 8.5 | $\chi^2 = 0.12, p = 0.73$ |
| | Laparoscopic procedure | 27.0 | 7.3 | |

Notes:

a = Inter-quartile range

b = Kruskal-Wallis 1-way ANOVA

c = Based on combined diagnoses of 'None', 'Occasional', 'Frequent' for stress incontinence and urinary incontinence

4.6.2 Criterion validity

Although the literature review indicated that there is no 'gold standard' for measuring patient satisfaction with the outcomes from surgery for incontinence, it was expected that satisfaction would be strongly related to either the degree of improvement in incontinence experienced or the woman's post-operative incontinence level. The gold standard for measuring urinary incontinence in women is the urodynamics tests described in the literature review (see also Tables 2 and 3 for the results of pre- and post-urodynamic testing). However, these clinical findings do not form a logical scale, and so cannot be used as a criteria for providing validity evidence for the GUTSS.

A measure, however, could be constructed by combining the different levels of stress and urinary incontinence (both change scores and post-operatively), and then using these as the criteria. The results are presented in Table 11 for each of the two GUTSS models. This revealed that on both GUTSS models there was no significant difference by the improvement in relative incontinence, but that there was a significant effect by post-operative absolute incontinence level. Women who, post-operation, were continent reported greater levels of satisfaction when compared with women who were still incontinent.

Table 11 Clinical criteria evidence for the GUTSS models

| | | <i>Median</i> | <i>IQR (a)</i> | <i>Statistics (b)</i> |
|---------------------------|-------|---------------|----------------|------------------------------|
| Model A | | | | |
| Relative incontinence (c) | Small | 19.0 | 15.0 | |
| | Large | 21.0 | 5.0 | $\chi^2 = 1.60, p = 0.21$ |
| Absolute incontinence (d) | No | 22.0 | 3.5 | |
| | Yes | 15.0 | 15.0 | $\chi^2 = 11.78, p < 0.01^*$ |
| Model B | | | | |
| Relative incontinence (c) | Small | 25.0 | 14.5 | |
| | Large | 27.0 | 7.0 | $\chi^2 = 1.20, p = 0.27$ |
| Absolute incontinence (d) | No | 27.0 | 7.0 | |
| | Yes | 22.5 | 12.0 | $\chi^2 = 7.97, p < 0.01^*$ |

Notes:

a = Inter-quartile range

b = Kruskal-Wallis 1-way ANOVA

c = Pre-post incontinence change. Small: where there was some evidence of improvement (1,2); Large: where there was evidence of improvement (3,4). Computed from pre-operation and post-operation incontinence severity for stress incontinence and urinary incontinence (See Tables 2 & 3). The derived values were: 1 = slight improvement, 2 = some improvement, 3 = improvement, 4 = large improvement. One case was discarded where her incontinence became worse (-1). Due to the small numbers of those with post-operative incontinence, where data were missing these were modelling using regression and the critical demographic & health status variables presented in Tables 1 & 2.

d = Post-operation incontinence. Dichotomised, based on combined diagnoses of 'None', 'Occasional', 'Frequent' for post-operation stress incontinence and urinary incontinence (See Tables 2 & 3). Due to the small numbers of those with post-operative incontinence, where data were missing these were modelling using regression and the critical demographic & health status variables presented in Tables 1 & 2.

* = Significant difference, $p < 0.05$

4.6.3 Open or laparoscopic technique

Our hypothesis, informed by the literature and focus group interviews, was that there would be no significant difference between satisfaction levels based on the type of surgical technique. This

was based on the assumption that the benefits associated with the laparoscopic technique of decreased hospitalisation and minimal scarring would be considered of minimal importance at 6 months post-surgery. Further, if both techniques were technically well performed they would result in the same procedure being performed (the Burch colposuspension) and the outcomes should therefore be similar for all patients. As shown in Table 10 this proved to be the case.

5 Discussion

The literature review indicated that the measurement of patient satisfaction following surgical intervention for incontinence was generally lacking in rigour and, where dichotomous questions are used, may lack sensitivity. From this literature and our focus groups with both clinicians and women who had undergone incontinence surgery, we identified six key areas of satisfaction which could be measured: current health status and living capacity, expectations and their fulfilment, satisfaction with the outcomes from the intervention, the level of pain and discomfort suffered, the care received, and the effect of any adverse outcomes or pre-existing conditions. We developed 22 items measuring these constructs, and administered them to 45 women at 6 months post-surgery. Following item examination, we constructed two models of patient satisfaction using an interactive process of exploratory factor, reliability and logical analyses. We then assessed our two models against very basic demographic characteristics, measures of health status and HRQoL, self-completed tests of incontinence, change scores from pre- to post-operation incontinence levels (ie. relative incontinence) and against 6-month follow-up incontinence status (ie. absolute incontinence).

In general, the findings suggested that both models of satisfaction possessed appropriate psychometric properties and can be used with confidence.

5.1 Choice of scale

The GUTSS Model A scale contained six items covering satisfaction with current health (2 items), outcomes (3 items) and adverse events (2 items); Model B contained eight items covering satisfaction with outcomes (3 items), adverse events (1 item) and care received (4 items).

The results of the analyses suggested that the Model A and Model B scales measured satisfaction with two slightly different constructs. The Pearson correlation between Model A and Model B was 0.84; but between Model A and *Factor 1: Outcome satisfaction* in Model B it was 0.93 compared with 0.31 between Model A and the Model B *Factor 2: Care satisfaction*. These correlations suggest that Model A and Model B *Factor 1: Outcome satisfaction* were virtually identical (as expected since they shared three common items). Model A did not measure satisfaction with care at all; as such Model B *Factor 2: Care satisfaction* represents an additional independent dimension that was being considered by women in their satisfaction with the entirety of the intervention process.

In assessing the relationship between the IIQ, UDI and the GUTSS Models A and B, it should be borne in mind that the IIQ and UDI measure issues concerned with HRQoL. It is possible this explains the slightly higher correlations between the IIQ, UDI and Model A ($r = -0.70$ & 0.72) when compared with Model B ($r = -0.67$ for both). It should be noted that Model A includes two items from the *current health status and living* dimension; the GUTSS Model B scale does not include any items from this dimension. However, the good correlations obtained for both GUTSS

models would also suggest that the HRQoL measures, the UDI and IIQ, measure factors congruent with the construct of patient satisfaction (as measured by the GUTSS) rather than those issues predominantly relating to HRQoL. This explanation may have some validity, as a review of these instruments by Renck-Hooper *et al* concluded that few of the items in the UDI and IIQ were relevant to HRQoL (Renck-Hooper, McKenna et al. 1997)

Turning to Model B, *Factor 2: Care satisfaction*, it is interesting that it did not include the only item (Q19) addressing the physical and structural issues of care. This confirms the comments made by the women in the focus group that these issues are not of primary importance in determining satisfaction with outcomes of an intervention. As this question related to length of stay in hospital this issue may be of greater importance where the decision to go home is not at the discretion of the woman. That this item was not identified as important during construction of a patient satisfaction scale raises questions about the involvement of patient satisfaction measures within hospital quality improvement and monitoring processes; questions which could be pursued by other researchers.

In determining whether to recommend the GUTSS Model A or Model B it is important to note that in previous investigations into hospital patient satisfaction — and in other specific disease/intervention patient satisfaction instruments — the dimension of *care received* has been strongly identified (Ware, Snyder et al. 1983; McCarthy, Shroyer et al. 1995; Kane, Maciejewski et al. 1997). It would seem important to include this in a measure of women's satisfaction with incontinence treatment. Furthermore, the durability of the *care received* dimension over a 6-month period implies that information pre- and post-surgery, the attitude/behaviour of doctors and nursing staff and care within the hospital have considerable impact on satisfaction with surgical care. In particular, consistent with the literature (Larsen and Rootman 1976), the content of these items indicates that the role performance of health care workers is an important dimension in patient satisfaction. In addition, it is possible that the *care received* dimension may be even more important where patients assess their satisfaction with incontinence interventions closer to the time of the treatment. This is when the immediate effect of *care received* on patient satisfaction would be more strongly felt.

For these reasons the GUTSS Model B scale is the recommended scale for measuring genito-urinary treatment satisfaction and is referred to as the *GUTSS* instrument. Each of the two factors are described as sub-scales measuring *Outcome satisfaction* (F1) and *Care satisfaction* (F2).

5.2 Limitations

In recommending the use of the GUTSS instrument it is important to note the limitations of its construction. The limitations of the study include a response rate of 60%, which although low is acceptable for questionnaire development. Of greater concern is the relatively small number of women (45) who participated in the GUTSS development. This number is well below that recommended for scale development, particularly where factor analytic techniques are used (Guadagnoli and Velicer 1988). This limits the strength of the conclusions drawn regarding the identification of the scales and the reliability of the instrument.

That the development sample also included the women who participated in the focus group is a matter of concern. It is possible this may have led to artificially constrained variance thereby causing over-estimation of the GUTSS' internal factorial structure and reliability (see Tables 5, 6 and 7). This was a practical necessity brought about by the few women available for participation in the study and the tight study deadlines for instrument development.

That it was not possible to identify why particular women chose not to participate limited our understanding of the population profile, the distribution of scores, and subsequent analyses of the data. Also we were unable to interview the non-English speaking women. These restrictions may have skewed the results in some way, since it may be that other dimensions of satisfaction would have been of greater importance to these women. The overall low income status of the women interviewed may also limit the wider application of the findings, although we found no evidence income *per se* led to differences in obtained GUTSS scores (see Table 9). We did find that women with higher educational qualifications reported higher levels of satisfaction; whether this was due to the GUTSS behaving differently by educational status or if this was a reflection of some other factor is unknown.

A remaining limitation is that the collection of data at a single point in time did not allow test-retest reliability to be assessed.

Despite these limitations, there is some external evidence regarding the validity of the GUTSS. In addition to the CLAIM trial construction sample reported in this paper, the GUTSS was also used in a further trial of laparoscopic versus open surgery treatment for incontinence ($n = 152$ cases available for analysis). Analysis of these data are presented in Tables 12 and 13. Table 12 shows the internal structure of the GUTSS with this new sample. In the comparable analysis, the four items in the *Outcomes satisfaction* sub-scale explained 77% of the variance compared with 47% for the construction sample, and the *Care satisfaction* sub-scale 63% compared with 25% for the construction sample (see Table 6). Examination of item loadings confirmed the original factor analysis: all items primarily loaded on their designated factor, although two items cross-loaded, whereas they hadn't with the construction sample. These statistics provide evidence confirming the internal structure of the GUTSS. The internal consistency estimate of $\alpha = 0.84$ was virtually identical to the original estimate ($\alpha = 0.83$).

Table 12 Validation of the GUTSS internal structure

| <i>Items</i> | <i>Factor analysis</i> | | | | | <i>IRT (c)</i> | <i>Cronbach a</i> |
|--|------------------------|--------------------|-----------|---------------------|-----------------------------|----------------|-------------------|
| | <i>PC (a)</i> | <i>Varimax (b)</i> | | <i>Eigen-values</i> | <i>% Variance explained</i> | | |
| | | <i>F1</i> | <i>F2</i> | | | | |
| Happy with the effect of the operation | 0.81 | 0.89 | | | | 0.80 | |
| Satisfaction with operation | 0.86 | 0.88 | | | | 0.79 | |
| After operation, problems with water works/ prolapse | 0.67 | 0.81 | | | | 0.83 | |
| Disappointed with operation outcome | 0.75 | 0.86 | | 3.07 | 77% | 0.81 | |
| Satisfaction with doctor's explanations about results of operation | 0.57 | (0.49) | 0.58 | | | 0.82 | |
| How happy with the care received in hospital | 0.44 | | 0.87 | | | 0.85 | |
| Attitudes/Behaviours of doctors & nurses | 0.47 | | 0.88 | | | 0.84 | |
| Prior information about the operation | 0.68 | (0.35) | 0.72 | 2.15 | 63% | 0.82 | |
| Total (whole scale) | | | | | 72% | | 0.84 |

Notes:

a = Principal components analysis

b = Loadings <0.30 not shown

c = Item-rest-of-test correlation

Table 13 shows GUTSS scores broken down by IIQ and UDI scores for women for whom these measures were available at 6-month follow-up (IIQ, $n = 47$; UDI, $n = 82$). Two sets of scores were calculated: (a) the absolute 6-month follow-up post-surgery IIQ and UDI scores and (b) the relative change IIQ and UDI scores from baseline. For the IIQ, due to the small number of cases, scores were dichotomised at the 50th percentile; for the UDI scores were recoded into quartiles. The hypothesis was that there would be a monotonic relationship between the recoded IIQ and UDI scores and median GUTSS scores. As shown in Table 13 this was the case for the UDI and IIQ absolute scores, and the differences were statistically significant, although in the case of the IIQ only just so ($p = 0.05$). Turning to the relative scores, the data showed that there was a monotonic relationship with the IIQ and they were suggestive for the UDI (although neither of these attained statistical significance). These findings are highly suggestive that the GUTSS is sensitive to post-operation incontinence health status scores.

Table 13 Validation of the GUTSS at 6 months following surgery

| | | <i>N.</i> | <i>Median</i> | <i>IQR (a)</i> | <i>Statistics (b)</i> |
|---------------------------------|-----------------|-----------|---------------|----------------|------------------------------|
| UDI – Absolute incontinence (c) | 1 st | 22 | 29.5 | 6.0 | $\chi^2 = 18.21, p < 0.01^*$ |
| | 2 nd | 14 | 29.5 | 6.0 | |
| | 3 rd | 23 | 26.0 | 6.0 | |
| | 4 th | 23 | 24.0 | 14.0 | |
| UDI – Relative incontinence (d) | 1 st | 17 | 24.0 | 13.5 | $\chi^2 = 2.84, p = 0.42$ |
| | 2 nd | 16 | 26.0 | 5.8 | |
| | 3 rd | 28 | 28.0 | 5.5 | |
| | 4 th | 19 | 27.0 | 10.0 | |
| IIQ – Absolute incontinence (e) | Lo | 25 | 27.0 | 7.5 | $\chi^2 = 3.70, p = 0.05^*$ |
| | High | 22 | 24.5 | 9.0 | |
| IIQ – Relative incontinence (f) | No/Lo | 18 | 24.0 | 10.3 | $\chi^2 = 1.24, p = 0.27$ |
| | High | 27 | 26.0 | 7.0 | |

Notes:

a = GUTSS inter-quartile range

b = Kruskal-Wallis 1-way ANOVA

c = Based on UDI quartiles, where the 1st quartile is those reporting little or no incontinence.

d = Quartile change scores computed from pre- minus post-surgery scores on the UDI respectively, where 1st quartile is those reporting no or very small changes, and the 4th quartile those reporting large positive changes in distress due to incontinence.

e = Dichotomised due to small numbers at the 50th percentile score. Lo = low level of stress, Hi = high level of stress due to incontinence impact.

f = Dichotomised due to small numbers at the 50th percentile score. No/Lo = low level in the change scores, Hi = high level of change scores indicating large improvement in incontinence impact status.

* = Significant difference, $p < 0.05$

The findings presented in Tables 12 and 13 suggest that the limitations of the study described above did not play an important role in the development of the GUTSS nor do they seem to have undermined the generalisability of the GUTSS to other samples; indeed, they provide *prima facie* evidence for the sensitivity, validity and reliability of the GUTSS. The findings do suggest, however, that, consistent with the literature, satisfaction may change where the GUTSS instrument is used at different follow-up time points; eg. say, at 3 months post-surgery and then again at 6 months (Black, Bowling et al. 1998). This hypothesis needs to be tested, as does an assessment for long term follow-up after at least a 2 year interval from date of surgery.

5.3 Towards a theory of patient satisfaction

The findings of this study, in relation to the components of patient satisfaction, suggest that *care received* and *satisfaction with outcomes* are the two primary components contributing to the construct of patient satisfaction. This is consistent with previous findings in relation to absolute outcomes and patient satisfaction (Locker and Dunt 1978; Linder-Pelz 1982; Hardy, West et al. 1996; Kane, Maciejewski et al. 1997), particularly with Sitzia & Wood's classification into the determinants and components of satisfaction (Sitzia and Wood 1997); the *Outcomes* sub-scale would fall within their 'determinants' rubric and the *Care* sub-scale 'components'.

That expectations did not impact greatly on a woman's satisfaction confirms previous investigations reporting that although important, they explain little of the satisfaction variance (Linder-Pelz 1982; Kane, Maciejewski et al. 1997). That GUTSS satisfaction scores were not sensitive to health improvement (ie. relative scores pre- and post-treatment) is also consistent with the literature (Kane, Maciejewski et al. 1997). Also, despite the literature (Donabedian 1980; Ware, Snyder et al. 1983; Loeken, Steome et al. 1997), pain and discomfort failed to explain a significant amount of the variance in either model. It may be that at 6 months post-operation any pain had been forgotten or was discounted due to its temporary nature; or, as expressed in the focus group, a certain amount of pain and discomfort was expected by those undergoing surgery. However, it should be cautioned that the data were collected 6 months after the intervention, which was the clinical period considered appropriate for judging the CLAIM trial outcome; expectations and pain may play a greater role in judgements made closer to surgery.

The finding that women's absolute health status at 6-month follow-up was important in predicting satisfaction levels (see Tables 11 and 13) was consistent with previous work suggesting that urge incontinence post-intervention was a strong indicator of patient satisfaction (Litwiller, Nelson et al. 1997). This finding suggests that the women were not judging the treatment as consumers,³ but rather as patients, particularly as some were still diagnosed with genuine stress urinary incontinence after surgery. If they were acting as consumers, it might be expected that these participants would have indicated greater dissatisfaction. The general lack of scores in the lowest range quartile of the GUTSS (see Figure 2) may be a reflection that women have very limited options as to where and from whom they can receive incontinence treatment in the public health system. Similarly it may be related to their satisfaction in receiving any relief at all, having been unsuccessfully through physiotherapy in order to become eligible for the Burch colposuspension procedure.

That satisfaction with care was identified as an important dimension of satisfaction, suggests, as per the work of Carr-Hill (Carr-Hill 1992), that the principles upheld by the consumer advocacy movement may be important to treatment outcomes. Perhaps women expect that information provision is a part of the role of the health worker, and that, regardless of their market powerlessness, they believe they are entitled to receive it.

The hypothesised theory is then one of *dual role transgression*. Patient satisfaction may operate as a function of the absolute health status of the patient following an intervention, and also as a

³ If they were judging it as consumers, it could be expected that their satisfaction scores would systematically vary by relative changes in their incontinence status between pre- and post-surgery measures.

function of the health care worker in providing treatment and in relation to the patient as a consumer of this medical care. Together, these two factors — post-treatment health status and role perception — may influence how a person evaluates their level of satisfaction. Further to this theory, given the very high levels of satisfaction expressed it may be that in order for dissatisfaction to be indicated on an instrument such as the GUTSS, transgression of both of these factors must take place.

6 Conclusion

The estimated prevalence of incontinence in the community and the socio-psychological impact it has on women's lives highlights the negative effects on HRQoL that incontinence can cause (Ashworth and Hagan 1993; Brocklehurst 1993; Berglund, Eisemann et al. 1996). Although there is a range of surgical and non-surgical treatments for stress incontinence, there has been minimal evaluation and comparison of these techniques within randomised clinical trials. Where outcomes from surgical incontinence treatment have been studied, definitions of follow-up periods and measures used to assess satisfaction with treatment have not been conducted in a standardised manner or with psychometrically developed measures. Similarly, no detailed studies have been conducted to identify the components and determinants of satisfaction with these interventions. The process undertaken in this study to develop the GUTSS instrument is the first step in addressing this ecological gap through providing a providing a detailed measure of satisfaction with outcomes, the GUTSS instrument.

The dimensions of *outcome* and *care satisfaction* which comprise the GUTSS, accord with past findings regarding factors statistically associated with patient satisfaction. That the women indicated limited dissatisfaction may represent the quality of the care and/or the limitations inherent in the GUTSS instrument. It is therefore recommended that future work pursue identification of the role of health status in determining treatment satisfaction, and the role of the health worker *vis-a-vis* that of the patient.

Although there is sufficient nomological evidence to suggest that the GUTSS is a sensitive, valid and reliable patient satisfaction measure, given the limitations of this study the GUTSS instrument must be further evaluated to determine its validity and test-retest reliability with other populations.

It is hoped that use of the GUTSS instrument will facilitate the measurement of satisfaction with incontinence treatment outcomes; the comparison of different treatments, particularly where new techniques and technology are being introduced; and improve the delivery of care in the urogynaecological field. This in turn should help improve the HRQoL for those receiving surgery for urinary incontinence.

References

- Aday L 1996, *Designing and Conducting Health Surveys: A Comprehensive Guide*, San Francisco, Jossey Bass.
- Alcalay M, Monga A & Stanton S 1995 'Burch colposuspension: A 10-20 year follow up', *British Journal of Obstetric Gynaecology*, vol 102, pp 740-745.

-
- Alford J 1997, 'Towards a new public sector management model: Beyond managerialism and its critics', M. Considine & M. Painter, *Managerialism the Great Debate*, Melbourne, Melbourne University Press, pp 152-172.
- Ashworth P & Hagan T 1993, 'The meaning of incontinence: A qualitative study of non-geriatric urinary incontinence sufferers', *Journal of Advanced Nursing*, vol 18, p 1415.
- Bates P, Bradley W & Glen E 1979, 'Standardization of terminology of lower urinary tract function', *Journal of Urology*, vol 121, pp 551-554.
- Benderev T 1992, 'Anchor fixation and other modifications of endoscopic bladder neck suspension', *Urology*, vol 40, no 5, pp 409-418.
- Berglund A, Eisemann M, Lalos A & Lalos O 1996, 'Social adjustment and spouse relationships among women with stress incontinence before and after surgical treatment', *Social Science and Medicine*, vol 42, 11, pp 1537-1544.
- Black N, Bowling A, Griffiths J, Pope C & Abel P 1998, 'Impact of surgery for stress incontinence on the social lives of women', *British Journal of Obstetrics and Gynaecology*, vol 105, no 6, pp 605-612.
- Black N, Griffiths J & Pope C 1996, 'Development of a symptom severity index and a symptom impact index for stress incontinence in women', *Neurourology and Urodynamics*, vol 15, no 6, pp 630-640.
- Bowling A 1991, *Measuring Health: A Review of Quality of Life Measurement Scales*, Milton Keynes, Open University.
- Bowling A 1997, *Measuring Disease: A Review of Disease-specific Quality of Life Measurement Scales*, Buckingham, Open University Press.
- Brocklehurst J 1993, 'Urinary incontinence in the community-analysis of a MORI poll', *British Medical Journal*, vol 306, pp 832-834.
- Burch J 1968, 'Cooper's ligament urethrovesical suspension for stress incontinence', *American Journal of Obstetric Gynaecology*, vol 100, p 764.
- Carr-Hill R 1992, 'The measurement of patient satisfaction', *Journal of Public Health Medicine*, vol 14, no 3, pp 236-249.
- Cleary P, Edgeman-Levitan D, Roberts M, Maloney T, McMullan W, Walker J & Delbanco T 1991, 'Patients evaluate their hospital care: A national survey', *Health Affairs*, vol 10 no 4, pp 254-267.
- Cooper M, Carlo G, Lam A & Carlton M 1996, 'A review of results in a series of 113 laparoscopic colposuspensions', *Australian & New Zealand Journal of Obstetric Gynaecology*, vol 36, pp 44-48.
- Corbett D 1998, *Australian Public Sector Management*, St Leonards, Allen & Unwin.
- Cronbach J & Meehl P 1955, 'Construct validity in psychological tests', *Psychological Bulletin*, vol 52, no 4, pp 281-302.
- Cummins R 1995, 'On the trail of the gold standard for subjective wellbeing', *Social Indicators Research*, vol 35, pp 179-200.

-
- Cummins R 1996, 'The domains of life satisfaction: An attempt to order chaos', *Social Indicators Research*, pp 303-332.
- Cundiff G, Harris R, Coates K & Bump R 1997, 'Clinical predictors of urinary incontinence in women', *American Journal of Obstetric Gynaecology*, vol 177-2, pp 262-266.
- Das S 1998, 'Comparative outcome analysis of laparoscopic colposuspension, abdominal colposuspension and vaginal needle suspension for femal urinary incontinence', *Journal of Urology*, vol 160, no 2, pp 368-371.
- Dean A, Dean J, Coulombier D, Brendel K, Smith D, Burton A, Dicker R, Sullivan K, Fagan R & Arner T 1994, *Epi Info, Version 6: A Word Processing, Database, and Statistics Program for Epidemiology on Microcomputers*, Atlanta, Centers for Disease Control and Prevention.
- Department of Finance and Australian Public Service Board 1987, *Evaluating Government Programs: Financial Management Improvement Program*, Canberra, Australian Government Printing Service.
- Diokno A, Brown M, Brock B, Herzog A & Normolle D 1989, 'Prevalence and outcome of surgery for female incontinence', *Urology*, vol 33 , no 4, pp 285-290.
- Donabedian A 1980, *Explorations in Quality Assessment and Monitoring: Vol 1 The Definition of Quality and the Approaches to its Assessment*, Ann Arbor, Health Administration Press.
- Donabedian A 1988, 'The quality of care: How can it be assessed?', *Journal of the American Medical Association*, vol 260, no 12, pp 1743-1748.
- Draper M 1997, *Involving Consumers in Improving Hospital Care: Lessons from Australian Hospitals*, Canberra, Australian Government Publishing Service.
- Draper M & Hill S 1996, *The Role of Patient Satisfaction Surveys in a National Approach to Hospital Quality Management*, Canberra, Australian Government Publishing Service.
- Drummond M, O'Brien B, Stoddart G & Torrance G 1997, *Methods for the Economic Evaluation of Health Care Programmes*, New York, Oxford University Press.
- Dull V, Lansky D & Davis N 1994, 'Evaluating a patient satisfaction survey for maximum benefit', *Journal on Quality Improvement*, vol 20, no 8, pp 444-453.
- Dwyer P 1996, Editorial Comment, *Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol 36, no 1, pp 44-45.
- Fall M, Ahlstrom K, Carlsson C, Anders E, Erlandson B, Frankenberg S & Mattiasson A 1986, 'Contelle: Pelvic floor stimulation for female stress-urge incontinence', *Urology*, vol 27, no 3, pp 282-287.
- Fetterman D, Kaflarian S & Wandersman A 1996, *Empowerment Evaluation*, Thousand Oaks, Sage.
- Fishbein M & Ajzen I 1975, *Belief, Attitude, Intention, and Behaviour: An Introduction to Theory and Research*, Reading, Addison-Wesley.
- Garcia-Aguilar J, Belmonte C, Wong W, Lowry A & Madoff R 1996, 'Open vs. closed sphincterotomy for chronic anal fissure: Long term results', *Diseases of the Colon and Rectum*, vol 39, pp 440-443.

-
- Gormley E 1997, Editorial: Assessment of outcome and patient satisfaction in the treatment of stress incontinence-uses and problems, *The Journal of Urology*, vol 157, p 1287.
- Guadagnoli E & Velicer W 1988, 'Relation of sample size to the stability of component patterns', *Psychological Bulletin*, vol 103, no 2, pp 265-275.
- Guba E & Lincoln Y 1990, 'Fourth generation evaluation: An 'interview' with Egon Guba and Yvonna Lincoln', *Evaluation Journal of Australasia*, vol 2, pp 3-14.
- Hardy G, West M & Hill F 1996, 'Components and predictors of patient satisfaction', *British Journal of Health Psychology*, vol 1, pp 65-85.
- Hawthorne G 1993, *Patient Satisfaction Survey*, Melbourne, Centre for Health Program Evaluation.
- Hawthorne G & Batterham R 1996, *Australian Validation of the Quality of Life in Depression Scale*, Melbourne, Centre for Health Program Evaluation.
- Hawthorne G & Richardson J 1995, *An Australian MAU/QALY Instrument: Rationale and Preliminary Results*, Proceedings Health Outcomes and Quality of Life Measurement Conference, Canberra.
- Hawthorne G, Richardson J, Day N, Osborne R & McNeil H. 2000, *Construction and Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument*, Melbourne, Centre for Health Program Evaluation.
- Hawthorne G, Richardson J & Osborne R 1999, 'The Assessment of Quality of Life (AQoL) Instrument: A psychometric measure of health related quality of life', *Quality of Life Research*, vol 8, pp 209-224.
- Hawthorne G, Richardson J, Osborne R & McNeil H 1997, *The Assessment of Quality of Life (AQoL) Instrument: Construction, Initial Validation & Utility Scaling*, Melbourne, Centre for Health Program Evaluation.
- Hilton P & Stanton S 1983, 'Urethral pressure measurement by microtransducer: The results in symptom-free women and in those with genuine stress incontinence', *British Journal of Obstetrics and Gynecology*, vol 90, pp 919-993.
- Holst K & Wilson P 1988, 'The prevalence of female urinary incontinence and reasons for not seeking treatment', *New Zealand Medical Journal*, vol 101, pp 756-758.
- House E 1993, *Professional Evaluation: Social Impact and Political Consequences*, Newbury Park, Sage Publications.
- Jarvis G 1994, 'Surgery for genuine stress incontinence', *British Journal of Obstetric Gynecology*, vol 101, pp 371-374.
- Kane R, Maciejewski M & Finch M. 1997, 'The relationship of patient satisfaction with care and clinical outcomes', *Medical Care*, vol 35, no 7, pp 714-730.
- Lam A, Jenkins G & Hyslop R 1995, 'Laparoscopic Burch colposuspension for stress incontinence: Preliminary results', *Medical Journal of Australia*, vol 162, no 2, pp 18-21.
- Langebrenne A, Dahlstrom B, Eraker R & Urnes A 1995, 'The laparoscopic Burch procedure: A preliminary report' *Acta Obstetrica et Gynecologica Scandinavica*, vol 74, no 2, pp 153-155.

-
- Larsen D & Rootman I 1976, 'Physician role performance and patient satisfaction', *Social Science and Medicine*, vol 10, pp 29-32.
- Leach B, Dmochowski R, Appell R, Blaivas J, Hadley H, Lubner K, Mostwin J, O'Donnell P & Roehrborn C 1997, 'Female stress urinary incontinence clinical guidelines. Panel summary report on surgical management of female stress urinary incontinence', *Journal of Urology*, vol 158, (3 Pt 1), pp 875-880.
- Linder-Pelz S 1982, 'Social psychological determinants of patient satisfaction: A test of five hypotheses', *Social Science and Medicine*, vol 16, pp 583-589.
- Linder-Pelz S 1982, 'Toward a theory of patient satisfaction', *Social Science and Medicine*, vol 16, pp 577-582.
- Litwiller S, Nelson R, Fone P, Kim K & Stone A 1997, 'Vaginal wall sling: Long-term outcome analysis of factors contributing to patients satisfaction and surgical success', *Journal of Urology*, vol 157, no 4, pp 1279-1282.
- Liu C 1993, 'Laparoscopic retropubic colposuspension (Burch procedure). A review of 58 cases', *Journal of Reproductive Medicine*, vol 38, pp 526-530.
- Locker D & Dunt D 1978, 'Theoretical and methodological issues in sociological studies of consumer satisfaction with medical care', *Social Science and Medicine*, vol 12, pp 283-292.
- Loeken K, Steome S, Sandvik L & Laerum E 1997, 'A new instrument to measure patient satisfaction with mammography', *Medical Care*, vol 35, no 7, pp 731-741.
- Marcello P, Roberts P, Schoetz D, Collier J, Murray J & Veidenheimer M 1993, 'Long-term results of the ileoanal pouch procedure', *Archives of Surgery*, vol 128, pp 500-504.
- McCarthy M, Shroyer L, Sethi G, Moritz T, Henderson T, Grover F, London M, Gibbs J, Lansky D, Miller D, Clark J & Hammermeister K 1995, 'Self-report measures for assessing treatment outcomes in cardiac surgery patients', *Medical Care*, vol 33, no 10, 0S76-0S85.
- McDowell I & Newell C 1996, *Measuring Health: A Guide to Rating Scales and Questionnaires*, New York, Oxford University Press.
- McIver S 1992, *Obtaining the Views of Inpatients and Users of Casualty Departments*, London, Kings Fund Centre.
- McIver S & Meredith P 1998, 'There for the asking: Can the government's planned annual surveys really measure patient satisfaction?', *Health Service Journal*, 19 February, pp 26-27.
- McTaggart R & Blackmore J 1990, 'Government control of evaluation research', *Evaluation Journal of Australasia*, vol 2, no 3, pp 41-49.
- Morgan D, Ed. 1993, *Successful Focus Groups: Advancing the State of the Art*, Newbury Park, Sage Publications.
- Mulcahy L & Tritter J 1998, 'Pathways, pyramids and icebergs? Mapping the links between dissatisfaction and complaints', *Sociology of Health and Illness*, vol 20, no 6, pp 825-847.
- Owens D & Batchelor C 1996, 'Patient satisfaction and the elderly', *Social Science and Medicine*, vol 42, no 11, pp 1483-1491.
- Palmer R, Donabedian A & Povar G 1991, *Striving for Quality in Health Care*, Ann Arbor, Health Administration Press.

-
- Pascoe G 1983, 'Patient satisfaction in primary health care: A literature review and analysis', *Evaluation and Program Planning*, vol 6, p 185.
- Pedhazur E. & Schmelkin L 1991, *Measurement, Design and Analysis: An Integrated Approach*, Hillsdale, Lawrence Erlbaum.
- Renck-Hooper U, McKenna S & Whalley D 1997, 'Measuring quality of life in female urinary urge incontinence: Development and psychometric properties of the IQoLI', *Journal of Outcomes Research*, vol 1, pp 1-8.
- Rowe L, Temple S & Hawthorne G 1996, 'Mothers' emotional needs and difficulties after childbirth', *Australian Family Physician*, vol 25, (9 Supplement 2), S53-S58.
- Rummel R 1970, *Applied factor analysis*, Evanston, Northwestern University Press.
- Scriven M 1993, *Hard Won Lessons in Program Evaluation*, San Francisco, Jossey Bass Publishers.
- Shafritz J & Russell E 1997, *Introducing Public Administration*, Reading, Addison-Wesley.
- Shumaker S, Wyman J, Uebersax J, McClish D & Fantl J 1994, 'Health-related quality of life measures for women with urinary incontinence: The Incontinence Impact Questionnaire and the Urogenital Distress Inventory', *Quality of Life Research*, vol 3, pp 291-306.
- Sitzia J & Wood N 1997, 'Patient satisfaction: A review of issues and concepts', *Social Science and Medicine*, vol 45, no 12, pp 1829-1843.
- Smith M 1998, 'Empowerment evaluation: Theoretical and methodological considerations', *Evaluation and Program Planning*, vol 21, pp 255-261.
- SPSS 1998, *SPSS for Windows*, Chicago, SPSS Inc.
- Streiner D & Norman G 1989, *Health Measurement Scales: A Practical Guide to their Development and Use*, New York, Oxford Medical Publications.
- Su T, Wang K, Hsu C & Hong B 1997, 'Prospective comparison of laparoscopic and traditional colposuspension in the treatment of genuine stress incontinence', *Acta Obstetrica et Gynecologica Scandinavica*, vol 76, no 6, pp 576-582.
- Taylor B, Beart R, Dozois R, Kelly K & Phillips S 1983, 'Straight ileoanal anastomosis: Ileal pouch-anal anastomosis after colectomy and mucosal proctectomy', *Archives of Surgery*, vol 118, pp 696-701.
- Thom D 1998, 'Variation in estimates of urinary incontinence prevalence in the community: Effects of differences in definition, population characteristics, and study type', *Journal of the American Geriatrics Society*, vol 46, no 4, pp 473-480.
- Thurstone L 1947, *Multiple Factor Analysis*, Chicago, University of Chicago Press.
- Uebersax J, Wyman J, McClish D, Shumaker S, McKlish J & Fantl J 1995, 'Short forms to assess life quality and symptom distress for urinary incontinence in women: The Incontinence Impact Questionnaire and the Urogenital Distress Inventory', *Neurourology and Urodynamics*, vol 14, pp 131-139.
- Wadsworth Y 1991, *Everyday evaluation on the run*, Melbourne, Action Research Issues Association Incorporated.

-
- Walker G & Texter J 1992, 'Success and patient satisfaction following the Stamey procedure for stress urinary incontinence', *Journal of Urology*, vol 147, pp 1521-1523.
- Walter S 1983, *Urinary Incontinence and Genital Descensus: Clinical and Urodynamic Findings*, Copenhagen, Laeforeningens Forlag.
- Ware J, Snow K, Kosinski M & Gandek B 1993, *SF-36 Health Survey: Manual and Interpretation Guide*, Boston, The Health Institute, New England Medical Centre.
- Ware J, Snyder M, Wright R & Davies A 1983, 'Defining and measuring patient satisfaction with medical care', *Evaluation and Program Planning*, vol 6, pp 247-263.
- Ware JE, Kosinski M & Keller SD 1994, *SF-36 Physical and mental health summary scales: A user's manual*, Boston, The Health Institute, New England Medical Centre.
- Williams B 1997, 'Patient satisfaction: A valid concept? *Social Science and Medicine*, vol 38, no 4, pp 509-516.
- Williams B, Coyle J & Healy D 1998, 'The meaning of patient satisfaction: An explanation of high reported levels', *Social Science and Medicine*, vol 47, no 9, pp 1351-1359.
- Wright J & Feinstein A 1992, 'A comparative contrast of clinometric and psychometric methods for constructing indexes and rating scales', *Journal of Clinical Epidemiology*, vol 45, no 11, pp 1201-1218.
- Wyman J, Harkins S, Choi S, Taylor J & Fantl J 1987, 'Psychosocial impact of urinary incontinence in women', *Obstetric Gynaecology*, vol 70, pp 378-381.
- Yeatman A 1997, 'The Reform of Public Management: An Overview',. M Considine and M Painter, *Managerialism the Great Debate*, Melbourne, Melbourne University Press, pp 173-187.
- Zorzos I & Paterson P 1996, 'Quality of life after a Marshall-Marchetti-Krantz procedure for stress urinary incontinence', *Journal of Urology*, vol 155, no 1, pp 259-262.