Development of a novel measure of surgical recovery using Rasch analysis

Tung Tran

Degree of Master of Science
Department of Epidemiology
Department of Surgery

McGill University
Montreal, Quebec, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master's of Science

August 31, 2010

©Copyright 2010 All rights reserved.

ACKNOWLEDGMENTS

This thesis would not have been possible without the guidance and support of Dr. Nancy Mayo. I would like to thank her for her great supervision, advice and inspiration. Her truly scientific rigor and passion has stimulated my growth as a student and has inspired me to become the great researcher and scientist that she is. Her vision has pushed me to think beyond my limits.

I owe my deepest gratitude to Dr. Liane Feldman, my co-supervisor, for her constant support. She was always accessible and willing to help. Her enthusiasm has motivated me throughout this process. She has been a mentor and a role model for me in her ability to integrate a meaningful research career and a successful clinical practice.

I am grateful to Dr. Gerald Fried who has been a continuous source of inspiration and motivation for our research group. His experience and contacts have allowed me to meet and interact with many of the great leaders in minimally invasive surgery.

It is a pleasure to thank Ms. Pepa Kaneva who has been an invaluable resource throughout this entire journey. From the beginning, she has been there to provide advice and ensure high quality maintenance of the databases and research protocols. She always kept me on track.

I would also like to show my gratitude to Dr Lois Finch who has helped me learn the Rasch model. She is truly an expert and I am fortunate to have had the chance to work with such a knowledgeable person.

I would also like to thank Ms. Ina Vanderspuy who was integral in introducing me to the Rasch model and setting up my study.

I would like to thank the Surgeon Scientist program for providing funding and giving me the opportunity to pursue this research.

The generous support of the unrestricted Covidien Educational Research grant is greatly appreciated and has allowed me to attend and present at many conferences which have been some of the most memorable experiences of these past 2 years. Finally, I would like to thank my amazing friends and family for all of their love and support.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABSTRACT

Development of a novel measure of surgical recovery using Rasch analysis

OBJECTIVE: Innovations in surgery are advocated on the basis of "enhanced recovery", currently measured using a mix of narrowly focused administrative indicators or patient-reported multiple-item questionnaires. Questionnaire length and the difficulty of integrating these different viewpoints limit accurate measurement. The objective of our study was to describe a method to develop a harmonized single linear measure that would be sensitive to expected differences in surgical recovery.

METHODS: The measure was developed and validated using data from 50 patients undergoing laparoscopic cholecystectomy. Patients were assessed preoperatively, 1week and 1month postoperatively by the following measures:health-related quality-of-life[generic(SF-36) and disease-specific(GIQLI)], symptoms (visual analog scales for pain and fatigue), physical activity (questionnaire) and function(6-minute-walk-test). Rasch analysis combined items across domains of the various instruments to develop the new measure on a logit scale which was transformed to a score from 0-100(SD). Reliability and validity were assessed. *P<0.05

RESULTS: A 34-item measure met all model requirements and included items from all domains. Reliability was excellent (0.96). Mean score decreased from 60(18) preoperatively to 55(15) at 1week* and increased above baseline to 68(18) at 1month*. Construct validity was assessed by comparing patients with or without complications. Groups were similar preoperatively (54vs62), but patients with complications had lower scores at 1 week (48vs58*) and 1month (57vs73*). The measure had moderate correlations with all instruments at all time points (r=0.33-0.87).

CONCLUSIONS: A novel measure of surgical recovery was developed encompassing a broad range of domains. Results are compatible with clinically observed postoperative recovery trajectories. This method illustrates that recovery can be quantified with mathematical units.

RÉSUMÉ

Développement d'une nouvelle mesure de rétablissement postopératoire

OBJECTIF: Innovations en chirurgie sont préconisées sur la base de «convalescence amélioré», actuellement mesuré en utilisant un mélange d'indicateurs administratifs étroitement ciblés ou des questionnaires à choix multiples rapportés par les patients. La longueur du questionnaire et la difficulté à intégrer ces différents points de vue limitent la précision de ces mesures. L'objectif de notre étude était de décrire une méthode pour développer une mesure harmonisée et quantitative qui serait sensible aux différences attendues en matière de rétablissement chirurgicale.

MÉTHODES: La mesure a été développé et validé en utilisant des données de 50 patients devant subir une cholécystectomie laparoscopique. Les patients ont été évalués en préopératoire, 1 semaine et 1 mois après l'opération par les mesures suivantes: la santé liée à la qualité de vie [générique(SF-36) ou spécifique à la maladie (GIQLI)], les symptômes (échelles visuelle analogique de la douleur et la fatigue), l'activité physique (CHAMPS questionnaire) et la fonction (6MWT). L'analyse selon le modèle Rasch a combiné à travers les domaines des différents instruments pour développer la nouvelle mesure sur une échelle logit qui a été transformée à un score de 0 à 100 (SD). La fiabilité et la validité ont été évaluées. * P = 0,05

RÉSULTATS: Une mesure de 34 items a satisfait toutes les exigences du modèle et a inclus des items de tous les domaines. La fiabilité a été excellente (0,96). Score moyen a diminué de 60 (18) en préopératoire à 55 (15) * à 1 semaine postop et a augmenté à 68 (18) à un mois *. La validité de construct a été évaluée en comparant des patients avec ou sans complications. Les groupes étaient similaires en préopératoire (54vs62), mais les patients présentant des complications avaient des scores inférieurs à 1 semaine (48vs58 *) et à un mois (57vs73 *). La mesure avait une corrélation modérée avec tous les instruments à tous les points dans le temps (r = de 0,33 à 0,87).

CONCLUSIONS: Une nouvelle mesure de rétablissement chirurgicale a été élaboré englobant un large éventail de domaines. Les résultats sont compatibles avec les trajectoires de récupération postopératoire observées cliniquement. Cette méthode démontre que la récupération peut être quantifiée avec des unités mathématiques.

INTRODUCTION
THE PROBLEM

One of the most significant changes in surgical care over the past decade has been the shift from inpatient to ambulatory surgery. In the past, patients undergoing surgical procedures remained in hospital for prolonged periods of time after surgery. An ambulatory procedure is defined as a nonemergency procedure in which all components of the process including admission, operation and discharge are accomplished on the same day. (Mcgrath 2003) In the United States, ambulatory surgery has been increasing since the 1980s and now account for 70% of all elective surgical procedures while England has seen an increase from 34% in 1990 to 49% in 2001. It is widely believed that the driving force behind this trend is mainly economical, as costs for hospital admission are rising. The shift toward ambulatory surgery has been enabled by recent medical advances in anesthesia and surgery. Improvements in anesthesia allow patients to gain consciousness more quickly after general anesthesia with fewer side effects and better pain control from enhanced analgesic regimens.(Cullen 2009) These advances are due to the development of safer anesthetic agents, improved knowledge of pain physiology and pain management, and incorporation of a better understanding of pathophysiology into perioperative care. Innovations in surgical techniques include the advent of minimally invasive procedures and fast track surgery that have further reduced the surgical stress response and postoperative pain and have opened the door for potentially enhanced recovery. Today, more complex procedures are being performed on sicker patients. (White 1994) The advances in anesthesia and surgical care have allowed an increase in ambulatory surgery and decreased the incidence of adverse events. This has shifted the postoperative recovery process to a home environment.
Complicating further the concept of recovery is that some surgery is performed to improve function and return to a pre-surgical level would be considered sum-

optimal. In general surgery, particularly abdominal surgery, surgery is performed to alter the disease course and patients may be functioning normally up until the time of surgery. Major abdominal surgery is usually followed by a period of disability, traditionally called "convalescence". During this time, patients report symptoms of pain, fatigue, and a reduced participation in physical activity. However, surgical recovery remains a nebulous concept predominantly because there is no accepted definition. Thus, there is a lack of specific validated tools to measure this construct.

Recovery is a complex process during which patients regain their preoperative functions and activities. However, for doctors, nurses, administrators and patients, the term "recovery" may be perceived differently. For example, for the anesthetist, recovery may represent the time when the patient has regained consciousness and stabilized physiological variables. Surgeons may claim recovery to have occurred once the incision is healed and the dressings are removed, while for hospital administrators, length of stay and discharge from the hospital may be important outcomes. From a patient's perspective, however, recovery only occurs once they are able to return to work or their regular daily activities. (Chumbley 1997) In fact, patients do not define recovery as being healed physically; they instead equate recovery with absence of symptoms and return of their ability to perform activities as they could prior to surgery. (S.V.M. Kleinbeck and N. Hoffart 1994)

Definition of Recovery

There is no consistent definition of recovery in the literature. Recovery is in fact a complex and continual process; moving through the moment the surgery has ended to the time the patient regains consciousness from anesthesia, to patient discharge to the floor, to patient discharge from the hospital to patient return to normal activities. The early stages of recovery tend to overlap with intraoperative care during which time the patient is regaining consciousness and recovering from anesthesia. However, recovery is not complete until the patient has returned to a

preoperative physiological or health state. (Stewart 1978) This entire process can last several days or weeks depending on patient factors and surgical factors. Recovery has traditionally been divided in 3 distinct but overlapping phases: early recovery, intermediate recovery and late recovery.

Phase 1 or early recovery occurs as the patient emerges from general anesthesia and begins with the discontinuation of the anesthetic agents, the patient's recovery of protective reflexes and motor function. Because patients require monitoring at this stage of recovery, this process usually occurs in the operating room or the post-anesthesia care unit (PACU) in the presence of trained personnel. There are scoring systems devised to assess the suitability for discharge from the PACU. The Aldrete score is based on physiological endpoints such as motor activity, color, responsiveness, respiration and circulation. This scoring system allows for the evaluation of a patient's readiness for discharge from the PACU. Once a threshold score is achieved, the patient is deemed fit for discharge to a step-down unit or floor. From there, intermediate recovery or phase 2 takes place until the patient is ready for discharge home. Late recovery occurs after discharge until the patient resumes normal everyday activities.(Marshall 1999) A significant proportion of anesthesiology research has focused on the pre-discharge and immediate postoperative period of recovery and has concentrated on minimizing postoperative symptoms to ensure early patient discharge from hospital. (Figueredo ED 1998) There are relatively few studies and, thus, few tools that have been used to evaluate late recovery, the period in which full functional and psychological recovery takes place. During this time, the patient is often outside the hospital and there, it is difficult to measure the impact of surgery on patients` functional status after discharge.

The Merriam Webster medical dictionary defines recovery as: "the act of regaining or returning toward a normal or healthy state". The Saunders Comprehensive Veterinary Dictionary defines surgical recovery as: "the process of healing of a surgical wound and restoration to normal of body functions and

systems, e.g. fluid and acid-base balance that have been disturbed by the original disease or by the surgical procedure."

The implication is that there is a disturbance in health due to a disease process or a surgical procedure and the gradual return to or above one's normal health is called recovery.

Tools Available to Measure Outcomes

Surgical research has traditionally assessed postoperative recovery using mortality, major complications length of hospital stay, length of time until return to work or daily activities and patient satisfaction. These outcomes reflect the "five-D's" originally proposed by Kerr White in the 1960's: death, disease, discomfort, disability, and dissatisfaction(White 1969). More modernly the outcomes of mortality, morbidity, disability (encompassing discomfort and emotional and psychological distress), dissatisfaction, and cost are recognized. (Mayo 2009)

Also, there has been a renewed interest in health care outcomes in recent years and in particular in the importance of incorporating the patient's perspective on outcome through the use of patient reported outcomes or PROs. (Acquadro C 2003)

Mortality and morbidity

Surgical outcome studies have traditionally focused on procedure-based outcomes and were often collected retrospectively from the medical record. The most common outcomes were mortality and morbidity in the form of major complications. The main advantages of these measures is that, for surgeons, they have face validity in that the patient did not have these prior to scheduled surgery and hence they should not have them after "recovery". There outcomes are also readily available. There are however, several limitations. For most surgical procedures, happily there are very few major adverse events but this poses

problems for statistical power if these outcomes are used to test new therapies or for routine monitoring of outcomes. For example, in order to have the 80% power to detect a statistically significant doubling of a baseline mortality rate of 5%, would require a sample size of 474 persons in each of the groups to be contrasted (Dallal 1986). There are very few procedures with a baseline mortality rate of 5% that are performed at such high frequencies and hence an increase of this magnitude might be dismissed as statistical variation rather than a need for action. In addition, the most common procedures are associated with a very low baseline risks which compound the problem of statistical power even further. Thus while the outcomes of morbidity and mortality may be practical for cardiovascular and cancer procedures, they are considerably less useful to assess the outcomes of low-risk operations, or those higher risk procedures that are done to improve quality of life. (Birkmeyer 2004)

While there is a single definition of mortality, it is not so straightforward to measure major complications. While there is no standard for measuring major complications, a common method is to count complications but classifying complications is challenging as there are few grading systems for complications. The most commonly used grading system was developed by Clavien and colleagues in 1992 and revised in 2004. Clavien defined complications as "any deviation from the normal postoperative course". Complications are graded I to V, according to the level of intervention required for treatment. (Dindo 2004) However, the term "complication" encompasses a variety of clinical states. Furthermore, the relative rarity of adverse events as a result of improved care means the majority of patients will have none contributing to a strong floor effect. (Bergman, Feldman et al.) In fact a major study in 38,958 patients undergoing ambulatory surgery found the risk of mortality in the 30 days following surgery to be 1:11,273. (Warner MA 1993) Other major prospective studies involving 13,433 and 6,914 patients, respectively, showed no perioperative deaths within one week of surgery. (Natof 1980; P.G. Duncan 1992) The risk of major

morbidity such as myocardial infarction, stroke and pulmonary emboli was also extremely low.

Disability including discomfort (symptoms), distress and limitations in activity

Because death and major complications are now rare outcomes due in part to improvements in surgical techniques and anesthesia, researchers have started to focus on less severe but more frequent clinical outcomes such as minor complications and sequelae such as postoperative pain, nausea, and vomiting. While these outcomes may not be as serious, they may offer insight into the efficacy of treatment. Minor complications and adverse symptoms or side effects such as postoperative pain or nausea and vomiting can prolong length of stay in PACU and in the hospital. Postoperative nausea and vomiting (PONV) is a common complication after anesthesia and for this reason, it has been frequently used as an outcome measure in ambulatory surgery. Many investigators have studied various regimens of anesthetic and antiemetic agents and their optimal timing to minimize PONV. However, the use of PONV as a surrogate outcome measure is limited because they were not consistent with non-surrogate endpoints or other outcomes such as length of stay in the recovery room, incidence of unplanned hospital admission or patient satisfaction. (Fisher 1999) Furthermore, PONV is a temporary postoperative symptom and often resolves soon after surgery.

Other endpoints used to measure surgical recovery include time-to-awakening and motor coordination after anesthesia. The Aldrete score was one of the first to describe patients' early recovery from anesthesia.(Aldrete 1970) Other studies have focused on the presence or extent of symptoms such as postoperative pain and PONV (Myles 1999)PONV after discharge home which occurs in 35% of patients is also a major obstacle for patients' return to normal daily activities. (Carroll NV 1995)

Postoperative pain may have broad implications on many aspects of recovery. Inadequate postoperative analgesia has been associated with decreased mental and psychological function as well as postoperative delirium (Lynch 1998). It has also been associated with fatigue which may negatively affect recovery, undermining patients' ability to resume daily activities. (DeCherney AH 2002) Furthermore, pain may also contribute to sleep disturbances and limit mobilization and delay return to usual activities. Finally, poorly controlled postoperative pain has also been correlated with poor quality of life and patient dissatisfaction. (Dawson L 1999; Myles PS 2000; Wu CL 2003)

The other common outcome assessed to reflect surgical recovery is return to work or normal activities. This milestone was often used to indicate end of the surgical recovery period and the time frame needed to be defined for purposes of remuneration during sick leave and for accessing disability benefits. However, recommendations for time off work given to patients by surgeons for most procedures are based mostly on tradition rather than strong evidence. While the immediate impact of surgery for the patient includes acute pain and other acute symptoms that gradually resolve over time, the patient can be left with a period of disability or fatigue and this may extend the time required for "convalescence" after surgery. Patients' return to work is often influenced by the manner in which work is remunerated. Patients who are self-employed tend to return to work earlier than patients with salaried jobs. The nature of patients' work also plays a role in their return. Jobs that are more physically demanding will require a longer period of convalescence in order that the person is recovered sufficiently to return to demanding work. Patients may return to work at a reduced level of activity or responsibility or delay their return for other reasons. Therefore, return to work has not been shown to correlate with validated measures of physical function or health status. On other hand, return to normal daily activities is also difficult to measure accurately as this may mean different things to different people and hence it is important to have an accurate portrait of what usual activities are. .

Finally, patients' expectations of how long it will take them to recover after surgery informed by their surgeon, may also play an important role in the time taken to resume full activities.(Bergman, Feldman et al. ; AW Majeed 1995; Bergman, Feldman et al. 2005)

Cost of care: length of stay

One of the most commonly reported measures of recovery in the literature is length of hospital stay after surgery. This outcome is easily collected retrospectively and is very relevant to payers as patients with prolonged length of stay (LOS) have substantially increased resource consumption.(Kalish RL 1995) However, LOS can often be influenced by many health-system factors and non-clinical factors. Health-system based factors such physician or hospital culture and institutional practice patterns can cause a discrepancy in LOS for the same procedure. Patient factors such as patient expectations, insurance status or discharge destination can also be a major player.

The conventional outcomes described above are fraught with numerous limitations in their ability to measure recovery after surgery. Mortality and major morbidity, while very clinically important events, are also very rare outcomes especially among those undergoing low risk procedures. Physiologic endpoints resolve shortly after surgery and thus their relationship with more complex processes such as physical and psychological functioning is not always evident. Finally, return to work or normal activities can often be misclassified and are largely dependent on patient expectations, doctors` recommendations, type of work, and social factors. Length of stay, while being an important economic outcome, can be influenced by many health-system and non-clinical factors.(Bergman, Feldman et al.)

Patient centered outcomes

Because of these limitations, there has been a steady increase in the emphasis of measurement of patient-centered outcomes in the literature. Many medical

interventions are designed to improve a patient's "quality of life rather than extend their duration of life". Quality of life issues have become increasingly important in health care evaluation and in research on evaluation of interventions and technologies. Assessing quality of life as an outcome of health intervention was relatively unknown in the 1970s but since then, there has been an exponential growth the use of quality of life as a an outcome in clinical research. These assessments have played an increasing role in evaluating the best treatment options and influencing clinical practice decisions. In the health field, the term quality of life has been narrowed to the concept of health-related quality of life which can be generic to assess the construct in a general population or disease-specific capturing those aspects of health interference specific to one or more related pathological processes. As this construct can only be captured by asking the patient directly, a number of indices with good psychometric properties have evolved and are in usage clinically.

Measuring Patient Centered Outcomes: Symptoms, Function, Health Perception and Quality of Life (QOL)

Two conceptual frameworks are relevant to classify patient-centered outcomes of health interventions: the Wilson-Cleary Model and the World Health Organization's International Classification of Function, Disability and Health.

Wilson Cleary model

The WHO defines quality of life as "*individuals' perception of their position in life in the context of the culture in which they live and in relation to their goals, expectations, standards and concerns*"(1995). From the perspective of the health care system, many of the components of quality of life, such as jobs, housing, schooling and neighborhood are not attributes of health, and are outside the purview of the health care system. (Ware 1987) As a result, the distinct concept of

health-related quality of life (HRQL) has emerged. Patrick and Erickson's (Patrick 1993), definition of HRQL is particularly important from a health services perspective; ".. *a measure of the value assigned to duration of life as modified by impairments, functional states, perceptions and opportunities, as influenced by disease, injury, treatment and policy*". While the definitions differ, there is an emerging consensus that generic HRQL takes into account levels of physical, mental, social and role functioning, and includes abilities, relationships, perceptions, life satisfaction and well-being (Bowling 1985; Berzon 1993). What distinguishes HRQL measures from others that capture some of these constructs is that HRQL must appraise at least the three core domains: physical, psychological and social.(Berzon 1993)

HRQL typically captures aspects of health ranging from the negative such as anxiety, depression, pain, fatigue, limited mobility, and restricted roles, to the positive including energy and happiness.

Wilson and Cleary described a comprehensive conceptual model for HRQL.(Wilson IB 1995) The model integrates five levels namely physiologic variables, symptom status, functional status, general health perceptions and overall quality of life. The authors suggested that this conceptual model of HRQL could unite biomedical and social science paradigms. The biomedical paradigm focuses on etiologic and pathological processes and its goal is to define and understand causal relationships. The biomedical model leaves no room for contribution of social, psychological or behavioral dimensions to the disease process. On the other hand, the biopsychosocial paradigm focuses on functional status and well-being of the patient. It evaluates all the factors including the social context and the environment that contribute to one's illness.(Engel 1977) The Wilson-Cleary model allows the integration of both these paradigms into a single conceptual framework that is potentially useful for health care providers. The model also incorporates individual and environmental factors. This HRQL conceptual model has been widely utilized in different populations such as patients with cancer, HIV/AIDS, Parkinson's disease and heart disease. In this

model, it is theorized that physiological variables influence symptom status which influences functional health. Functional health then influences general health perceptions that influence overall quality of life. Evaluation of physiological variables is done at the cellular level and centers on organs and organ systems. Symptoms status is assessed through the organism as a whole. Functional health takes into account one's environment and one's ability to adapt within that environment. General health perceptions include the previous concepts as well as mental health. The Wilson Cleary conceptual model of HRQL illustrates a theoretical approach of HRQL as a multidimensional construct. Figure 1 illustrates the Wilson Cleary model for HRQL. (Sousa 2006)

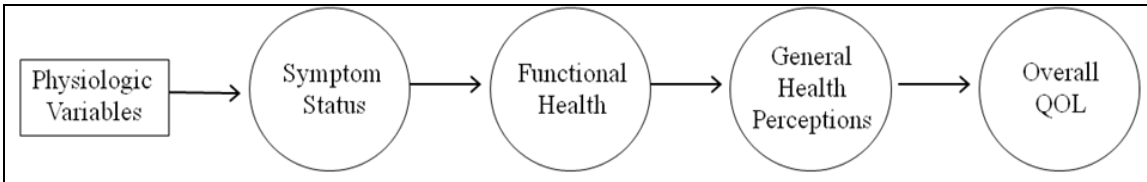


Figure 1. The main pathway in the Wilson Cleary (1995) HRQL conceptual model.

World Health Organization's International Classification of Function, Disability and Health (ICF)

The International Classification of Function, Disability and Health is a classification system of health states proposed by the WHO based on the sociological perspective of health that considers disability along the whole continuum of function. The ICF brings a new perspective to the notion of 'health' and 'disability'. The ICF shifts the focus from cause to impact and thus places all health conditions on equal grounds allowing them to be compared. Moreover the ICF takes into account the social and environmental factors of disability and does not consider disability as purely 'medical' or 'biological'. Thus, with the ICF, one can record the impact of a health condition and the environment on a person's overall functioning.

The overlap and compatibility of these two models has been demonstrated by Valderas and Alonso: biological and physiological impacts and symptoms in the

Wilson-Cleary model are impairments in the ICF and functions in the Wilson-Cleary model are activities and participation in life's roles under the ICF. Both models recognize the importance of personal factors and the environment in modifying these inter-relationships. The authors developed a classification system for PRO based on a valid integrated conceptual model. There were standard definitions for terms used in the classification. Symptom status is defined as a patient`s perception of an abnormal physical, emotional or cognitive state. Functional status is defined as the ability of an individual to perform tasks. Health perceptions is the subjective integration of all information related to symptom status and functional status and finally health related quality of life are the aspects of quality of life that relate specifically to a person`s health. The classification system by Valeras and Alsonso provide the framework for the classification of the most commonly used instruments and may allow for more adequate selection and applications of these instruments. The integrated conceptual model allows for a better method of classifying outcomes of health interventions. (Valderas 2008)

Why measure aspects of quality of life

Assessment of health-related quality of life can be useful in many ways. It allows for the assessment of the impact of chronic disease burden and the assessment of efficacy and effectiveness of therapeutic interventions. While physiological measures may be objective and important for clinicians, patients are not familiar or interested in these endpoints. Furthermore, they correlate poorly with patients' functional ability or well-being. (Guyatt 1993) For example, for patients with chronic lung or heart disease, exercise capacity in the laboratory does not translate to exercise capacity in everyday life.(Guyatt GH 1985) Also, two patients with objectively the same clinical criteria may have very different responses to treatment due to one's experiences, expectations or ability to adapt. For these reasons, patients, clinicians and administrators are becoming more interested in measuring HRQL as an outcome to evaluate novel interventions.(Wennberg 1990)

Generic instruments of HRQL

Measurement of HRQL can be approached in two ways, through the use of generic indices which apply to population at large or disease-specific indices which address health effects typical of specific health conditions.

Generic instruments attempt to measure all important dimensions of HRQL. For example, the Sickness Impact Profile includes a physical and a psychological domain as well as other categories for eating, work and recreation. The main advantage of generic instruments lies in their ability to apply to any population across types and severity of disease and across different medical treatments or health interventions. They allow broad comparisons across diverse groups. Furthermore, there are often scores available for a normative population permitting comparisons to be made to people without the specific health condition. Nonetheless, aspects particular to a specific disease may be missed and generic instruments may not be sensitive enough to detect small but important changes in health of specific diseases.(Patrick DL 1989)

Specific instruments of HRQL

Specific instruments focus on aspects of quality of life specific to a particular disease state, patient population, function ability or problem. These instruments tend to be more intuitive and relevant to clinicians and patients. They allow for potentially increased responsiveness by including aspects of HRQL that are important for the study population. They are developed specifically to capture health states and change in these that are considered clinically relevant for patients or clinicians as they are changes that come about through intervention or changes highly associated with established physiologic measures. Nevertheless, the scope of the instrument may be too narrow to measure unintended changes in the study population. They may also have a narrow range of applicability. For example, if one develops a specific instrument to measure quality of life in patients undergoing chemotherapy, this instrument is only applicable to

chemotherapeutic regimens with a similar side effects profile. Also, different treatments may improve different aspects of HRQL.(Guyatt GH 1986)

Criteria for the evaluation of QOL instruments

Psychometric evaluation of new HRQL scales begins with assessment of validity, reliability and responsiveness or sensitivity to change.

Validity

Criterion validity involves the evaluation of a new instrument against an accepted absolute "gold standard" in the measurement for that construct. In the case of QOL, no such standard exists as responses are experiential and personal. In the assessment of face validity or content validity of an instrument, one reviews whether the items clearly and unequivocally cover the intended spectrum of topics. Face validity can be maximized if there are people from a wide range of backgrounds (physicians, patients, administrators) involved in the assessment. Construct validity is a more formal approach which involves examining the relationship between the quality of life instrument with other established measures. For example, the extent of agreement between HRQL scores and laboratory or clinical measures of disease severity could be examined. Also, one could look at an instrument's ability to distinguish between patient groups with known differences in health status. (Fitzpatrick R 1992)

Reliability

A measure is reliable if it yields the same results in a stable patient. It is best assessed by repeated measuring of patients in whom the clinical status has not changed.

Responsiveness

Responsiveness is the ability to detect clinically important changes in quality of life, even if they are small. Instruments need to distinguish between patients at a

point in time and need to be sensitive to changes in patients that occur over time. This information is crucial for clinicians and patients. Responsiveness can be quantified using indicators of effect size. There are several reasons why an instrument may be insensitive to change. A generic instrument may not include several items that are relevant to the specific disease and thus fail to measure a change in condition when one occurs. Also, items included in the instrument may be static; these items may not change or be the target of intervention. Finally, the scoring of a quality of life instrument may be subject to floor or ceiling effects. For example, patients with very high quality of life scores may fail to register an improvement from treatment because of the limitation of the scale. There is unfortunately no standard to which to compare new instruments. (Fitzpatrick R 1992)

Model for Surgical Recovery

Similar to HRQL, there have been models postulated to depict the likely causal pathways between the different dimensions involved in the process of recovery. Carli et al. proposed a simple model in which surgery is a well-known stressor that produces immediate biological changes. (Figure 2)

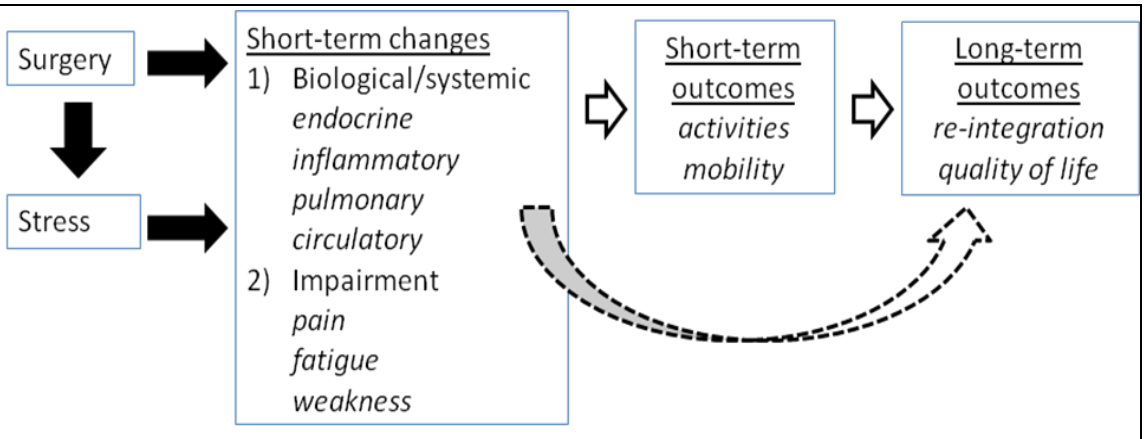


Figure 2. Biological model of Surgical Recovery by Carli et al. 2001

These changes are physiologic and systematic and can be measured in the perioperative period. These biological and physiological changes in turn may impact on short-term outcomes such as mobility or participation in activities.

Finally, long term outcomes such as functional status, quality of life may be affected. While evidence is lacking for some of these associations, it is helpful to consider how the various parameters could be studied. (Carli 2001)

Another model by Carli et al. has been proposed to describe the trajectory of functional ability throughout the surgical process. (Figure 3)

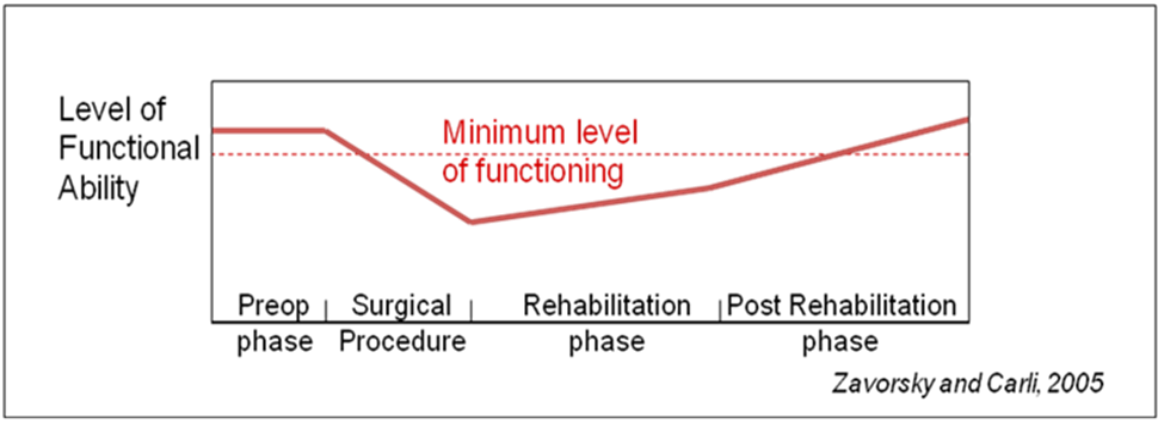


Figure 3. Trajectory of functional ability throughout surgical process

Patients start at a baseline functional status preoperatively. This level is determined by patient factors such as age, comorbidities and disease burden. Patients then undergo a decrease in functional status shortly after surgery before entering the recovery phase. The extent of the decrease and the slope of the recovery phase may be influenced by surgical factors such as invasiveness of the procedure and presence of intraoperative or postoperative complications. (Carli 2005) This model allows us to conceptualize the recovery process as well as identify factors that may modify this process.

Modern Instruments Used to Measure Recovery

With these concepts, many patient reported outcomes have focused on the assessment of HRQL related to surgical procedures. These instruments can be generic or disease specific. Some instruments attempt to measure all important dimensions of QOL while others focus on a particular aspect of QOL such as symptom status or functional status.

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

The short form health survey of the Medical Outcomes Study questionnaire contains 36 items and was designed as a generic index of health status in the general population. It was developed from the Rand Corporation's health insurance experiment in the United States of America. The original general health survey was lengthy with 108 items. In order to develop an instrument that was comprehensive and psychometrically rigorous yet practical, the authors tested various shorter forms. The short form 20 was developed and then the short form 36 was constructed to answer to criticism of its predecessor. It attempts to capture aspects of health important to all patients.(Ware JE Jr 1992; McHomey CA 1993) The tool is useful to compare patients with the same condition or among patients with different conditions. As it was developed to measure health status in the general population, it provides an excellent way comparing a patient population to a "normal" or "healthy" group. (Stewart AL 1989) The SF-36 measures 8 health concepts: 1)physical functioning, 2)role limitations attributed to physical problems, 3)bodily pain, 4)social functioning, 5)mental health, 6)role limitations attributed to emotional problems, 7)vitality and 8)general health perception. Questions are scored on a Likert scale and the 8 subscales are summed to provide 8 scores ranging from 0 to 100. There are also 2 summary scores that aggregate the scales: physical component summary score and mental component score which are standardized to a mean of 50 and a standard deviation of 10. (Ware JE 1994) The SF-36 can be self-administered or used by interview in person or telephone. Questions generally take 5-10 minutes to complete. (Weinberger M 1991) The reliability of the eight scales and the two summary scores have been evaluated using internal consistency and test-retest methods. Most published studies showed a reliability estimate of greater than 0.8 (0.7 is recommended for group comparisons) for the scales and it exceed 0.9 for the physical and mental summary scores. (Ware JE 1993; Ware JE 1994) These results have been replicated across a variety of different patient groups with different diagnoses and

socio-demographic characteristics. Validity has been provided for patient populations in the United States. The SF-36 has been able to classify patients according to the disease severity and detect differences on all 8 subscales. Furthermore, it could distinguish patients with medical problems from patients with psychiatric conditions. (McHomey CA 1992; McHomey CA 1993) Responsiveness to clinical change has also been demonstrated. (Katz JN 1992) The SF-36 has been useful in comparing general and specific populations, estimating the burden of different diseases and evaluating the health benefits of different interventions. (Manocchia M 1998) These findings and those of other studies have provided evidence for the strong psychometric properties of the SF-36. This scale has been used in more than 200 studies in North America.
The SF-36 has also proven valuable in surgical patients. Some studies have used the SF-36 to assess the success of novel surgical interventions. For example, Trus et al. showed an improvement of all eight SF-36 health categories after laparoscopic antireflux surgery. (Trus TL 1999) Others have studied the impact of surgical outcomes on patient quality of life. Camilleri-Brennan et al showed a significantly lower HRQL in patients with recurrent rectal cancer. (Camilleri-Brennan J 2001) Therefore, the SF-36 has become a well-validated, reliable and sensitive generic HRQL questionnaire applicable to the surgical population.

Gastrointestinal Quality of Life Index (GIQLI)

The GIQLI developed by Eypasch is a self-reported system-specific, validated instrument of QOL designed to assess patients with gastrointestinal disorders. It contains 36 multidimensional items that address symptoms, physical, emotional and social dysfunction related to gastrointestinal conditions. Each item is scored on a Likert scale from 0 to 4 with a total score ranging from 0 to 144 with higher scores representing better HRQL. (Eypasch E 1995) It can also be divided into 5 different subscores: 1)gastrointestinal symptoms (0-76 points), 2)emotional status (0-20 points), 3)physical function (0-28 points), 4)social function (0-16 points) and 5)stress due to treatment (0-4 points). For reliability, internal consistency

coefficients were high. Test-retest reliability was assessed in clinically stable patients (ICC= 0.92). The questionnaire was developed by an international board of experts; this comprehensive process ensured content validity. Construct validity was supported with correlations between GIQLI and other appropriate measures. Scores on the GIQLI were able to differentiate between groups of patients with varying levels of function as well as between patients with gastrointestinal disease and those without. Responsiveness was highest in patients with gastroesophageal disorders but there was variable responsiveness for other abdominal procedures.(Bremers AJ 2000; Freys SM 2001; Chen L 2002; Korenkov M 2002; Ludwig K 2002) Thus, the GIQLI questionnaire has been shown to be a potentially useful instrument to measure quality of life in patients with gastrointestinal disorders.

Quality of Recovery Score (QoR and QoR-40)

Myles et al. sought to develop a patient-rated valid, reliable and responsive instrument of quality of recovery after anesthesia and surgery. A 61-item questionnaire was constructed that asked patients, relatives, medical and nursing staff to rate various items describing patient experiences postoperatively. The most highly ranked items were summarized in a nine-point index score called the "QoR score". This new scoring system was then prospectively evaluated. Evidence for good convergent validity and discriminant construct validity was provided from 2 cohorts of surgical patients (n=449). Good interrater agreement ($\rho = 0.55$), test-retest reliability ($\rho = 0.61$) and internal consistency ($\alpha = 0.51$) were demonstrated. This suggested that this instrument was acceptable for group measurements.(Myles 1999) In a follow up study, Myles et al. developed a 40-item questionnaire (QoR-40) with a maximum score of 200. They found good convergent validity between this instrument and the VAS. Evidence construct validity was provided by negative correlation with duration of hospital stay and a lower QoR-40 score in women compared to men. There was good test-retest reliability ($\rho = 0.92$), internal consistency ($\alpha = 0.93$) and split half coefficient ($\alpha =$

0.83).(Myles 2000) The QoR-40 was also studied to measure the association between quality of recovery 3 days after surgery and QoL 3 months postoperatively in patients undergoing cardiac surgery. It was found to be a valid measure of recovery and superior to the SF-36 in measuring early postoperative recovery. It was also found to be predictive of QoL 3 months after surgery. (Myles 2001) Thus the QoR and the QoR-40 are valid and reliable instruments measuring early postoperative recovery and may be useful endpoints in perioperative clinical studies.

Activities Assessment Scale (AAS)

Other questionnaires have addressed a patient's day-to-day functional status and quality of life after surgery. The AAS was designed for a randomized control trial comparing laparoscopic versus open inguinal hernia procedures. A pool of condition-specific and procedure specific items was developed through focus groups with surgeons and with patients who had undergone either laparoscopic or open inguinal hernia repair to develop The final version of the AAS comprises 11 items covering a wide number of sedentary (lying), movement-related (walking), and graded-intensity physical activities (housekeeping to exercise). The scores range from 0 to 100 with higher scores indicating a higher level of activity. It requires 3 to 5 minutes to complete. This scale was then studied in a sample of 2164 patients at baseline and 1562 patients at 3 month follow up. The AAS was found to show internal consistency (α=0.85) and construct validity and known-groups validity were demonstrated through correlation with the physical function subscale of SF-36 and comparisons between clinical subgroups respectively. Thus, the AAS is another instrument measuring a wide spectrum of physical activity, is easily administered and may prove useful to clinicians assessing patient outcomes after surgery. (McCarthy Jr 2005)

Community Health Activities Model Program for Seniors (CHAMPS) questionnaire

The CHAMPS questionnaire is a patient self-report which was designed to estimate physical activity in elderly people. It was developed to provide an outcome of physical activity for the Community Health Activities Model Program for Seniors intervention study.(Stewart 1997) It includes 41 items that estimate the length of time spent on various physical activities in a given week. CHAMPS includes activities of varying intensities ranging from light to vigorous. With the type and the duration of physical activities, one can calculate an estimate of the total caloric expenditure per kilogram per week. The questionnaire can be administered in 10 to 15 minutes by questionnaire or interview. In terms of reliability, the reported intraclass correlation coefficients were reported in the order of 0.58-0.67 over a 6-month interval and good test-retest reliability was found over a 2-week period (ICC=0.62-0.76). Preliminary evidence for construct validity was provided by Harada et al. and Stewart et al. who found correlations 0.22-0.54 between caloric expenditure and four functioning measures (self-reported physical functioning, energy and fatigue, 6-minute walk and lower body functioning tests).(HARADA 1997) CHAMPS data also showed moderate correlations with data from other physical activity questionnaires, the Physical Activity Scale for the Elderly (PASE) and the Yale Physical Activity Survey (YPAS). It was also able to discriminate between less active and more active elderly adults. The instrument was sensitive to expected change with small to moderate effect sizes (0.38-0.64) as it detected changes in patients who underwent interventions to increase physical activity. (Stewart 2001) CHAMPS has also been validated in other populations including elderly Australians and African-Americans. (Resnicow 2003; Cyarto 2006) Therefore, CHAMPS has good psychometric properties and yields estimates of energy expenditure that are sensitive to community-based interventions. In the context of surgical patients, Feldman et al. provided evidence for construct validity for CHAMPS as a measure of surgical

recovery in a population of patients undergoing laparoscopic cholecystectomy. (Feldman, Kaneva et al. 2009)

An additional benefit of this questionnaire is that the activities included are also those relating to socialization and quiet recreation hence capturing return to usual activities, a construct integral to participation.

Six-Minute Walk Test (6MWT)

A common feature of activities involved in functional status one`s capacity for walking. Thus a functional test such as the six-minute walk test (6MWT) may be a useful instrument. This test was initially developed in patients with chronic obstructive pulmonary disease. (Guyatt GH 1985) It assesses a person's ability to maintain a moderate walking speed for a specific period of time. It is a submaximal exercise test that can be performed in patients who cannot tolerate maximal exercise. It is very simple and does not require expensive equipment or specialized training. It has been applied widely and is considered safe in patient populations with walking limitations. In terms of reliability, test-retest coefficients have been reported in the range of 0.73 to 0.99. These results included many different populations such as the elderly, people with chronic cardiac or pulmonary disease, heart failure, chronic obstructive pulmonary disease, fibromyalgia, peripheral arterial disease, and end-stage lung disease. (Enright PL 1998; Gibbons WJ 2001; Finch E 2002; Steffen TM 2002; Kervio G 2003) The measurement error in community-dwelling elderly is estimated to be 20m. (Troosters T 1999) In terms of validity, there is strong correlation (0.64-0.90) between the 6MWT and VO_2max, the most accepted test of exercise tolerance, in patients without health conditions associated with musculoskeletal impairments. (Cahalin L 1995; Montgomery PS 1998; Roul G 1998; Miyamoto S 2000; McElduff P 2002) The 6MWT has also shown good predictive validity. It has been found to be a good predictor of peak oxygen consumption as well as survival in patients with advanced heart failure and lung disease. (Cahalin LP 1996; Kadikar A 1997) Several studies have found the 6MWT to be responsive to

change after exercise intervention, however, there is little data in the clinical setting.(Fitts 1995; Gunnarsson 1997; Hogue CC 1997) In patients with stroke, the value is close to 60m. (Miller PA 2002) In the context of surgical patients, the 6MWT has been used as a marker to guide the assessment of patients for lung transplantation. It has also been used to compare differences in outcomes of patients undergoing lobectomy by thoracoscopic surgery, anterior limited thoracotomy anteroaxillary thoracotomy or posterolateral thoracotomy. (Nomori, Ohtsuka et al. 2003) Finally, Moriello et al. used the 6MWT as a measure to assess recovery after elective colon resection surgery. (Moriello, Mayo et al. 2008) Therefore, the 6MWT may satisfy some of the requirements for an indicator of surgical recovery.

Visual Analog Scale (VAS)

Other instruments such as the visual analog scale are often used to assess the early period of recovery. These focus on symptom status such as anxiety, pain, fatigue and nausea. VAS for these symptoms may contribute to our understanding of the recovery process. Visual analog scales have been used to measure a variety of subjective responses in psychology, education and health research. (Freyd 1923; Aitken 1969) The VAS is a simple tool widely used to assess a subjective "feeling". A patient is asked to indicate their perceived level of symptom along a 10cm horizontal line ranging from 0 to 10. 0 is defined no symptom at all, 10 is defined as unbearable symptom. The score is then recorded. The VAS for pain has been extensively studied. The VAS pain score has been found to correlate well with acute and chronic pain levels by several investigators. The reported measurement error is 20mm. (Scott J 1976; Gaston-Johansson F 1990; Jensen Ml' 1993) Moreover, many studies have found the VAS for anxiety to correlate well with State Anxiety of the Spielbereger State-Trait Anxiety Inventory (STAI), a validated test quantifying state anxiety. (Vogelsang 1988; Arellano R 1989; Millar K 1995) Furthermore, the VAS method has proven to be useful in the assessment of postoperative nausea intensity and for testing the efficacy of medication.

(Boogaerts 2001) Finally, the VAS for fatigue has also been extensively studied in various populations such patients with multiple sclerosis, cancer or undergoing colorectal surgery. (Krupp 1989; Schwenk 1998; Gaston-Johansson F 1999) All these symptoms may play a role in the surgical recovery process and the VAS is one method in which to assess them.

Limitations of Current Instruments

Despite a wide availability of instruments used to measure surgical recovery, they all have several common limitations.

Measurement Theory

Classical test theory (CTT) has dominated the field of measurement and while it has a number of strengths, its limitations have been identified with the application of Item Response Theory to the health field.

In classical test theory, researchers administer a questionnaire or survey containing a number of items intended to assess the same construct. The responses of each item are then aggregated and each weighted equally to form a total score for the scale. The total score is then often treated as if it were situated on an interval scale. However, the classical test theory relies on several assumptions. Firstly, all items are given an equal weighting and thus assumed to have an identical importance to the construct. Classical test theory assumes that each person has a true score for that instrument. Because the instruments are imperfect, the observed score is different from the true score as a result of measurement error. This can be expressed as:

$$X = T + E$$

X: observed score T: true score E: error

The assumption from this theory is that the individuals taking the test are a "representative" sample of a larger and less well-defined population

Another major limitation of CTT is the inability to separately estimate a person's underlying "ability" on the construct and a questionnaire item's "difficulty" along that same construct. That is, a person's score is dependent on the items administered and the items' mean difficulty depends on the sample of people taking the questionnaire. Therefore, the estimates of person ability or item difficulty are dependent on the sample and cannot be compared across groups along the distribution of the underlying attribute. A person ability and item difficulty cannot be estimated separately. Furthermore, CTT only yields one estimate of reliability and standard error of measurement while we know that the precision of an instrument changes along the spectrum of the underlying construct.(DeVellis 2006) Because CTT is based on a test score, the administration of entire instrument is required in order to obtain a score. Given the length of many HRQL questionnaires, the extent to which these can be applied in clinical practice can be limited. (Prieto 2003) Finally, while clearly the intention is that a measure should be independent of specifics surrounding the administration of a particular test, ordinal test scores (obtained by adding the number of items answered correctly) are test-bound. Ordinal test scores are completely dependent on the specifics of a given test. For example, a score of 25 on a particular test has no intrinsic meaning. Even if we have the total number of test items (e.g. 25 out of 30), it is not very helpful because there is a difference between answering 25 easy items correctly and answering 25 more difficult items correctly. Therefore, to have a real understanding of a person's ability on a given test, one would have to individually evaluate each item attempted. Unfortunately, the classical test theory does not provide the adequate tools to address these issues. (Masters 2005)

Most HRQL instruments are comprised of a collection of items which are scored on an ordinal rating scale. A total score is commonly derived by summing the ordinal responses; this process assumes that each item contributes equally to the underlying construct and that the ordinal response options are mathematical values permitting addition.

The latter assumption often does not hold. Ordinal scores are usually assigned to response categories starting with the lowest category and increasing incrementally until the highest category. There is often no rationale for choosing one incremental progression instead of another and in fact, any number progression is satisfactory as long as the numbers get larger. Therefore, these numbers only serve to indicate the order in which the categories are scored and do not represent true measures. Because there is no relationship made between the distances separating the response categories and the underlying construct being measured, one cannot consider these scores to be true measures. (C. Merbitz 1989; Linacre 1989) For example, when considering a 5-point scale for pain: none, mild, moderate, severe, very severe; each with a respective score 0,1,2,3,4, one can only conclude that higher numbers represent "more" of the underlying construct without the ability to quantify it. For example, a reduction of 2 points on this scale, say from the category very severe to moderate, represents a greater reduction in pain than a decrease from the category very severe to severe, however we cannot conclude that it corresponds to twice the effect. Arithmetic operations and parametrical statistics are often performed on the scores obtained from these ordinal scales. Unfortunately, this is not appropriate as these types of manipulation are only valid for interval measures. That is, the scores allocated to each response category of ordinal scales is somewhat arbitrary and do not represent true numbers. Furthermore, most instruments are summed scores which usually include a zero score. This zero score as well as the gradations are arbitrary and instrument-dependent and have no meaningful interpretation. Therefore, misusing these numbers can lead to false conclusions. (Ashby 1984) Without an interval scale, one can only ascertain whether change has occurred however the magnitude of that change is difficult to estimate. The extent of change is an important component when evaluating the effectiveness of alternative interventions. Conversely, interval scales are linear, quantitative and have known and equal intervals. The difference between two units is known and constant along the entire scale of interest. Therefore, arithmetical and statistical operations

provide mathematical units with meaningful interpretations and allow for quantitative comparisons among response scores.

Assessment of change

Because most HRQL instruments lack interval properties, it can be difficult to interpret changes in scores from these indices. One study by Stucki et al. showed evidence of these difficulties. When items of a health status index are concentrated in one part of the continuum, changes may be inflated.(Norquist 2004)

Therefore we feel that surgical recovery is a complex construct that encompasses multiple dimensions including physiological function, symptoms, physical, emotional and social function, health perception and quality of life. Given the limitations of traditional outcomes and current measurement tools to assess quality of life, there is a need to develop a harmonized single linear method of measuring surgical recovery.

THE SOLUTION

To achieve this, a method alternative to the classical test theory needs to be used to allow for the development of an interval-like metric for measuring the underlying construct. We must first define the concept of objective measurement before we can develop a satisfactory measure.

Definition of Objective Measurement

As written by the Program Committee of the Institute for Objective Measurement, December 2000, objective measurement can be defined as “the repetition of a unit amount that maintains its size, within an allowable error, no matter which

instrument, intended to measure the variable of interest, is used and no matter who or what relevant person or thing is measured." An estimate from an objective measure stays constant with time. It is independent of the people being measured, the measuring instrument and the people using the instrument. This is a fundamental requirement for useful measurement. Objective measurement operates under the traditions of fundamental measurement theory, item response theory, and latent trait theory.

Item-response theory (IRT)

In many measurement situations, there is an underlying variable of interest. The development of every measuring instrument begins with the concept of an underlying variable. Contrary to height or weight, such constructs are unobservable and cannot be measured directly. These are what psychometricians call unobservable or latent traits. These latent traits are concepts rather than physical dimensions. The concept of surgical recovery fits this description. In order to measure these underlying traits, one needs a scale of measurement, a ruler with a given metric. The intention of measures for underlying variables is that they should have the same meaning regardless of the instrument used to obtain the results. This is particularly important when comparing results of different tests. Item response theory (IRT) is composed of a set of generalized linear models with their corresponding statistical procedures that describe the relationship between a person's observed responses to a query (item) and their location on an underlying or latent trait. IRT has several advantages over CTT. IRT yields person and item estimates that are independent with respect to the characteristics of the population. Also, unlike CTT, the standard errors are conditional on trait level. (Hays RD 2000) IRT models are mathematical equations that describe the relation between a person's underlying level of latent trait (person ability) and their probability of particular response on a given item of the questionnaire. (Reise SP 1993) The association between the predicted responses to an item and the latent trait is displayed by the item-characteristic curve (ICC). Most models of IRT assume

unidimensionality and local independence. Unidimensionality indicates that all items in a scale are measuring a single construct. For local independence, all items must be uncorrelated from each other once the latent trait has been controlled for. (McDonald 1981) Unidimensionality is an essence of measurement. However, in practice, unidimensionality is more of a qualitative rather than quantitative concept. No actual test can be perfectly unidimensional. However, it is an ideal to which measures must be approximated in order to obtain generalizable results. For example, a test consisting of English and math items with the purpose of making a pass/fail decision is unidimensional inasmuch as this test is used for that purpose. The decision to create such a test does not mean English and math are identical or exchangeable outside of their ability to make that pass/fail decision. Therefore, unless each test item is treated alone as a test, there is always a compromise between the ideal of unidimensionality and the requirements of practice.(Linacre 1989)

When there are dichotomous items, there tends to be an S-shape curve to describe the relation between increasing person ability and increasing probability of endorsing that item. The ICC shows a non-linear regression of the probability of a specific response on the y-axis and a person's ability or trait level on the x-axis. (Figure 4)

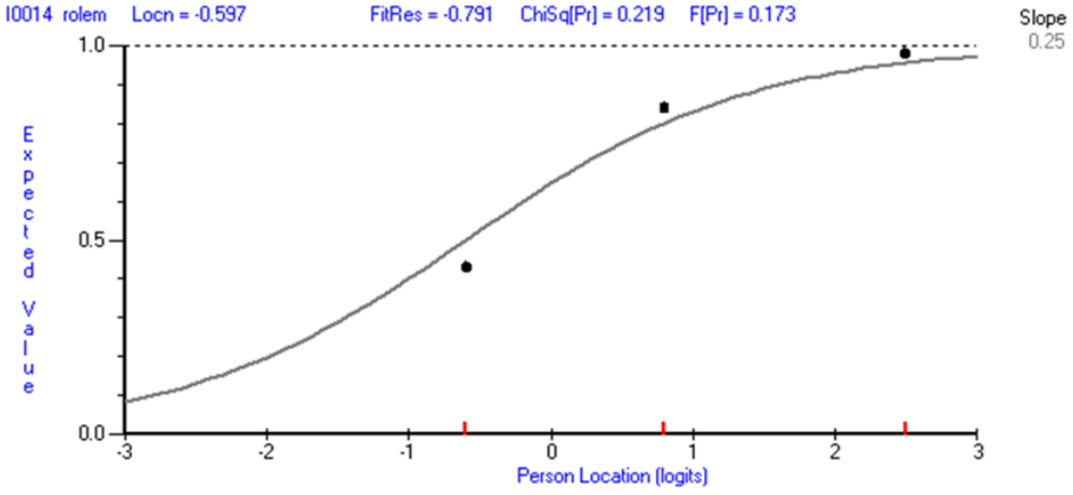


Figure 4. Item Characteristic Curve for a Dichotomous Item (I0014).

The ICC for “easier” items, items with a higher probability of endorsement, will be shifted more to the left of the scale while the ICC for “more difficult” items, items with a lower probability of endorsement are shifted to the right of the scale.

Rasch Model

Overview

The Rasch model is a 1-parameter logistic function model. It “provides practitioners with a basis for : establishing the extent to which a set of test items work together to provide measures of just a single variable, defining a unit of measurement for the construction of interval level measures and constructing numerical measures that have meaning independent of a particular set of items used.”(Masters 2005) They were originally designed and used for educational assessment but in recent years, their use has increased in health research. Rasch models are probabilistic measurement models which can be applied in various disciplines including psychology, marketing, economics, social sciences and are being increasingly used in areas such as health care. They provide the basis for the measurement of quantitative attributes or traits on a continuous scale based on ordinal or categorical data obtained from responses of persons to items. In the Rasch model, the probability of a specific response to an item (e.g. right/wrong answer) is modeled under the relationship between person and item parameters. More specifically, the probability is modeled under a logistic function of the difference between a person’s ability and an item’s difficulty. (Equation 1) (Andrich 1988)

Let $X_{ni} = x \in \{0, 1\}$ be a dichotomous random variable where, for example, $x = 1$ denotes a correct response and $x = 0$ an incorrect response to a given assessment item. In the Rasch model for dichotomous data, the probability of the outcome $X_{ni} = 1$ is given by:

$$\Pr\{X_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}},$$

where β_n is the ability of person *n* and δ_i is the difficulty of item *i*.

Equation 1. Rasch model for dichotomous data.

These parameters are estimates of the quantitative level of trait of a person or item. For example, in educational tests, item parameters represent the "difficulty" of an item; that is, the higher the difficulty, the lower the likelihood a given person will obtain a correct answer. Person parameters represent the ability of a person; the higher the ability, the higher the likelihood that person will obtain a correct answer on a given item. When a person's ability is equal to an item's difficulty, there is by definition a 50% probability of a correct answer according to the Rasch model. (Bond 2001)This model allows for the possibility of obtaining measurements from categorical or ordinal response data.

There is an important distinction between a Rasch model and the perspective of a statistical model. In statistical models, the purpose of the model is to describe the data. Parameters of a statistical model are accepted or rejected on the basis of how well they fit the data. A Rasch model on the other hand, is a model in the sense of an ideal or standard and is distinctly different. In a Rasch model, the purpose is to find data that fit the model. The rationale for this is based on the requirements of objective measurement as it is understood in physical sciences.

A useful analogy to understand this rationale is to consider 2 objects on a weighting scale. If object A was measured as having "more weight" than object B, but immediately after, object B was measured as having "more weight" than object A, we would consider that there was an error with the scale. A requirement for measurement is that comparison between objects should be the same and independent of other factors such as time. This requirement is found within the

Rasch structure and therefore the Rasch model does not change to suit the data. Therefore, an important advantage of Rasch models is that measurement can be considered independent of the sample. Thus if the data fits the Rasch model, item parameters are not dependent on the persons being measured.(Andrich 1988) The Rasch model also allows one to assess the extent to which the set of items in the measure are sample-independent.(Ellis 1992) These item parameters should remain relatively constant in the context of an invariant construct and thus allow comparisons of person abilities to be made across different samples.

Scaling

In CTT, when all the test-takers attempt all the items, a test score is attributed to an estimated of person ability. A greater total test score represents a person with greater ability. However, as mentioned earlier, because the items are on ordinal scales, the relationship between test score and person ability is not linear. Also, the relationship between test score and person ability is dependent on the distribution of the items of that particular test.

In contrast, the Rasch model defines the scale for the item locations first. For example, in educational test, an item with few correct responses is considered a more "difficult" item and thus has a higher item location. This process is called item calibration. Once all the items are scaled, the person abilities are located on the same scale. The Rasch model allows for the measurement of both item difficulty and person ability along a common measurement continuum.

As mentioned earlier, for dichotomous data, the difficulty of item is by definition the location on the scale at which a person the same corresponding ability has a 50% probability of obtaining a correct response. The probability of a correct response is lower than 50% when item difficulty exceeds person ability and is higher than 50% when person ability is greater than item difficulty. Therefore, this results in a linear interval scale that allows one to determine which set of items corresponds to a person's ability for the underlying construct. And given the instrument-independent characteristics of a Rasch model, a person's ability can be

estimated independently of the particular set of items administered. It is therefore possible for two persons to have similar scores or ability location after having responded to a completely different set of items.(Wright 1979; Wright BD 1982) Furthermore, each item or person parameter has a standard error of measurement associated with it as an expression of the precision of that estimate. Measurement error is not uniform throughout the range of test scores and tends to be higher at both extremes of the scale.

Statistical and graphical tests are performed to evaluate the fit of the data to the model. The response patterns obtained from the sample are tested against what is expected, a probabilistic form of the Guttman scale, and with other fit statistics. Some tests are global while some assess certain groups of people or items. These statistics determine fit to the model by testing how far the observed responses are from the responses expected from the model. Because the Rasch model defines measurement requirements, the data is tested to see if they fit the model`s expectations. This is different from statistical modeling where a model is developed to best represent the observed data. (Andersen 1977)

Polytomous Rasch model

The model discussed above was applicable for dichotomous items. There is also a Rasch model for polytomous items in which the objective is to measure a latent trait in through responses to items that are scored with successive integers on an ordinal scale. This model becomes applicable to commonly used Likert scales. The polytomous Rasch model was derived by Andrich in 1978. It is also sometimes referred to as the Rating Scale Model or the Partial Credit Model. (Andrich 2005) It is still a general probabilistic measurement model similar to the Rasch model for dichotomous data preserving its distinctive properties at the same time allowing for the use of sequential integers. (Equation 2) (Andrich 1978)

Firstly, let

$$X_{ni} = x \in \{0, 1, \ldots, m_i\}$$

be an integer random variable where m_i is the maximum score for item i. That is, the variable X_{ni} is a random variable that can take on integer values between 0 and a maximum of m_i.

In the polytomous Rasch model, the probability of the outcome $X_{ni} = x$ is

$$\Pr\{X_{ni} = x, x > 0\} = \frac{\exp\sum_{k=1}^{x}(\beta_n - \tau_{ki})}{1 + \sum_{x=1}^{m}\exp\sum_{k=1}^{x}(\beta_n - \tau_{ki})}; \Pr\{X_{ni} = 0\} = \frac{1}{1 + \sum_{x=1}^{m}\exp\sum_{k=1}^{x}(\beta_n -$$

where τ_{ki} is the kth threshold location of item i on a latent continuum, β_n is the location of person n on the same continuum, and m is the maximum score for the item. Applied in a given empirical context, the model can be considered a mathematical hypothesis that the probability of a given outcome is a probabilistic function of these person and item parameters. The graph showing the relation between the probability of a given category as a function of person location is referred to as a *Category Probability Curve* (CPC). An example of a CPC for an item with 3 categories, scored from 0 to 2, is shown in Figure 12.

Equation 2. Rasch model for polytomous data.

Differential item functioning (DIF)

Within the context of Rasch measurement, the model should work the same way regardless of the group being assessed. People from different groups (eg. gender) with the same latent trait (ability) should have the same probability of providing a certain response. DIF occurs when people from different groups with the same underlying ability have different probabilities of giving a certain response. For example, if one gender did not display the same probability of positively responding to a given item, this item would be considered to have DIF and this would violate the unidimensionality requirement. (Fischer 1995)

Advantages of Rasch

Rasch models have several advantages. As previously mentioned, ordinal or cardinal data can be linearly transformed and organized on an interval scale which then allows for parametric manipulations. Testing for unidimensionality as well as other measurement issues such as category ordering (do the categories of an item work the way they are intended to?) or differential item functioning can all be

assessed within the framework of a Rasch model. Unlike traditional psychometric methods, the items fitting the Rasch model do not all have to be administered as they are ordered by difficulty. This feature would reduce the burden on the respondent while retaining the strong psychometric properties. Therefore, fitting a Rasch model is a method to address many methodological measurement aspects of scale development and validation and providing a linear transformation for ordinal data. (Pallant 2007)

Examples of instruments developed using Rasch analysis

For these reasons, recent studies in health research have begun to investigate the use of Rasch models to develop new instruments (Bode RK 2003), to modify existing questionnaires (Pallant JF 2006), and to compare results from different instruments and different languages. (Holzner B 2006) These studies have started to show the advantages of Rasch models compared to traditional psychometric instruments. For example, Norquist et al. used a Rasch model to assess change over time of an outcome measure used in total hip replacement surgery. They found some gains in responsiveness of outcome measures from Rasch-based approaches compared to standard Likert scoring. (Norquist 2004) Thus the authors were able to improve their ability to assess change. As another example, Raczek et al. used Rasch measurement models to evaluate the comparability of forms of SF-36 physical functioning across countries. They showed that Rasch models could be used to assess the equivalence of item calibrations in different contexts such as different languages and to create a common measure of physical functioning that incorporates items from different instruments.(Raczek 1998) Finally, Pallant et al. used Rasch analysis to evaluate the use of the Hospital Anxiety and Depression Scale (HADS) as a global measure of psychological distress. Although HADS contained two aspects, anxiety and depression, their results helped conceptualize the construct of psychological distress and provided evidence to support the use of HADS as a measure for this unidimensional construct.(Pallant 2007) These are some examples of the recent applications of the

Rasch models that have helped redefine the way measurement is undertaken in health research.

PURPOSE OF THE STUDY

Rationale and Objectives

There is a gap in understanding the construct of surgical recovery. It is a complex construct with multiple dimensions, many involving domains with HRQL. Because surgical recovery is such a nebulous construct, there is a lack of comprehensive, validated, hierarchical instruments to measure it. Conventional outcomes are fraught with limitations. Patient-reported outcomes may be more appropriate but most of these instruments have been developed under classical test theory. Unfortunately, these outcomes have limitations intrinsic to the classical test theory related to ordinal scales, inability to adequately assess change and instrument- and person-dependency of the score.

Working within the concept of objective measurement as understood in physical sciences, we can develop methods for measuring complex constructs under the Rasch model. Rasch models allow harmonization across different measures by hierarchically calibrating the different items to a linear scale; the resultant method now fits the requirements necessary for quantitative meaningful measurement. Within the Rasch model, one can assess for unidimensionality, differential item functioning and category ordering while working on an interval scale. Furthermore, there have been many recent studies exploring this method to develop, modify or validate new tools in health research.

Therefore, the aim of our study is develop a harmonized single linear method of measuring surgical recovery that would be sensitive to expected changes arising from surgery.

THE NEW METHOD OF MEASURING METHODS

Subjects

A longitudinal study was carried out of patients greater than 18 years of age scheduled for ambulatory or short-stay (overnight) laparoscopic cholecystectomy. The study was approved by the MUHC-Montreal General Hospital Research Ethics Committee (GE#04-012) and informed consent was obtained prior to entry into the study. Patients were recruited in the preoperative clinic between September 2005 and August 2006. Exclusion criteria included: poor understanding of English or French, inability to walk, emergency surgery, a psychiatric condition precluding participation in the study, a procedure converted to open and the presence of an immediate complication.

Data Collection

Patients were evaluated preoperatively and at 1 week and 4 weeks postoperatively. Patients were assessed with the CHAMPS physical activity questionnaire, a generic HRQL questionnaire (SF-36), a system-specific HRQL questionnaire (GIQLI), a functional walking capacity test (6MWT) and a visual analog scale (VAS) for pain, anxiety, nausea and fatigue. Data on intraoperative and postoperative adverse events were also collected from the medical records. Patients were instructed to return to their usual daily and physical activities whenever they felt comfortable. Patients who required insurance forms to be completed were given 1 week off from work after surgery. Patients who failed to present for scheduled follow-up visits were contacted by telephone in an attempt to complete the questionnaires (n = 12 at 1 week and n = 13 at 1 month).

Indices

The method of measuring surgical recovery used items from valid and reliable indices and tests that are commonly used to evaluate patients after surgery. Physical activity was evaluated through the CHAMPS questionnaire. CHAMPS was first developed to assess the efficacy of interventions to increase physical activity level in the elderly. It is a 41-item questionnaire in which patients report the frequency and duration of a wide variety of physical activities. These activities range in intensity from low (walking leisurely, light housework) to vigorous (running, heavy lifting). Patient may also add other activities they have performed that are not listed in the questionnaire. Each activity is assigned a metabolic equivalent (MET) or kcal/kg/hr. For example, walking leisurely for 1 hour is equivalent to 2.5 kcal/kg while a more vigorous activity such as jogging is associated with 7 kcal/kg. With the type and duration of activities, one can calculate a patient's total energy expenditure for a typical week. (Stewart 2001) However, several modifications were made to the original CHAMPS questionnaire for the purposes of our study. Firstly, the original CHAMPS questionnaire asks subjects to report their physical activity in "a typical week in the last 4 weeks". Patients in our study were asked to report their physical activity in the preceding week. Secondly, for entry into the Rasch model, patients` responses were dichotomized according to their participation in each item during the preceding week. Thirdly, the International Classification of Functioning, Health and Disease (ICF) was used to recode the items from CHAMPS. This was necessary as many items refer to quite similar activities. A total of 18 ICF coded activities were identified. The intensity of activity was quantified depending on the physical intensity of the item using metabolic equivalents (MET) values based on values reported by Ainsworth et al. (Ainsworth 1993). For example, items involving sporting activities were recoded into the item d9201 according to the ICF classification. In this recoded item, the categories: bowling, playing golf while riding a cart, playing golf while walking, doubles tennis, downhill skiing,

skating, sports (basketball, soccer, racquetball), singles tennis and ice hockey were given a ranked intensity from 1 to 9 respectively on an ordinal scale. The intensity level of 0 was given to a subject who participated in none of these activities. Thus, one item, engagement in sports, was created and this was the item used in the Rasch model.

Functional walking capacity was assessed with the performance of the six-minute-walk test. (6MWT). According to guidelines for the test, patients were instructed to walk back and forth in a 15-m stretch of hallway for 6 minutes, at a speed that would make them tired at the end of the walk. It is a submaximal exercise test originally developed for patients with chronic obstructive pulmonary disease. (Guyatt GH 1985) It has since been validated in a variety of patient population including surgical patients undergoing colorectal surgery. (Moriello, Mayo et al. 2008) Continuous data from the 6MWT was categorized into 9 categories (0-200m, 201-300m, 301-350m, 351-400m, 401-450m, 451-500m, 501-600m, 601-800m, >800m) corresponding to scores from 0 to 8 respectively on an ordinal scale.

The Medical Outcomes Study 36-item Short-form health survey (SF-36) was used as a generic tool to assess HRQL. It has been extensively studied and has been found to be reliable and valid and many populations. It measures 8 health concepts, each scored on scale ranging from 0 to 100. Higher scores indicate higher quality of life. These health concepts include 1)physical functioning, 2)role limitations attributed to physical problems, 3)bodily pain, 4)social functioning, 5)mental health, 6)role limitations attributed to emotional problems, 7)vitality and 8)general health perception. There are also 2 summary scores that aggregate the 8 subscales: physical component summary score (PCS) and mental component score (MCS) which are standardized to a mean of 50 and a standard deviation of 10. (Ware JE 1994)

The Gastrointestinal Quality of Life Index (GIQLI) is a specific tool designed to assess HRQL in patients with gastrointestinal disorders. It comprises 36 items; each is scored on a Likert scale from 0 to 4 with a total score ranging from 0 to

144 with higher scores representing better HRQL. It can also be divided into 5 different subscores: 1) gastrointestinal symptoms (0-76 points), 2) emotional status (0-20 points), 3) physical function (0-28 points), 4) social function (0-16 points) and 5) stress due to treatment (0-4 points). (Eypasch E 1995)
Visual analog scale items (0-10) were used to assess pain at rest, with cough or with movement as well as anxiety, nausea and fatigue. These items were scored ordinally from 0 to 10. The individual items of the SF-36 and the GIQLI were tested for fit to the Rasch model and sub-scales were used to provide evidence for the validity of the new measurement method.

Data Analysis

Descriptive statistics were used to characterize our study population. Baseline data and data over time were presented as median and interquartile range for values of the indices used to measure recovery after surgery. Parametric and non parametric tests were performed in the statistical analysis software R version 2.8.0.
A Rasch analysis was conducted to construct a measure from items from the indices described above. The Rasch method was used to further confirm item hierarchy and dimensionality. The Rasch model specifies the relationship between the probability of a person`s response to a particular item and the interaction between a person`s level of functional ability and the level of functioning that item represents. (Wright 1979) Each person's ability and each item's level of difficulty are estimated in the Rasch model with a standard error (SE) and are organized hierarchically on the same scale which is based on the natural log of odds ratio or logits. The Rasch model transforms the response of each item onto an interval scale using a logit transformation with the average item difficulty set at "zero" on the logit scale. When a person's ability is equal to an item's difficulty, the probability of success on that item is 50%. The Rasch allows an easy interpretation of the estimates of the measure. One can estimate a person's ability

given their score by calculating the antilog of the logit score and computing the probability of success on that item.

The extended logistic Rasch model for ordered response categories was computed with Rasch Unidimensional Measurement Model programme (RUMM 2020). (Andrich D 2005) Criteria for fit included: standardized fit residuals greater than 2.0 or less than -2.0, a significant χ^2 or *F*-statistic. (Andrich 1988) To ensure proper structure, the difficulty levels of each item's response options must be ordered. Disordered responses can mean that patients were not able to distinguish the response categories as finely as the response options allow. (Linacre 2002; Andrich 2005) For example, a person with less ability may be found to have a higher probability of success on a more difficult item that a person with higher ability. Thus disordered response categories were inspected and rescored in a manner to optimize the rating scale category effectiveness. (Linacre 2002; Andrich 2005) The items were co-calibrated concurrently using the partial credit model, a method that estimates the difficulty level of items from different indices measuring the same construct, despite differences in number or types of response categories.(McHorney 2000) Items were deleted iteratively based on fit to the model, association with the construct and precision. Rasch analysis provides a standard error for each item and person as a measure of precision as well as overall reliability and separation indices. The closer the reliability index is to 1.0 (range 0.0-1.0), the better. (Smith 2001)

Differential item functioning

Once the data fit the model, a two-way analysis of variance of the person-item residuals was performed to determine whether each item's location was stable across different influencing factors (gender, time). (Andrich 2004) To quantify a construct like surgical recovery, a measure must be invariant; that is, for persons with the same level of ability, the probability of successful responses must be unaffected by characteristics such as gender or time of assessment. For DIF analysis, the patients were divided into three groups of equal ability and then by

gender or time within that group. The difference in the level of difficulty for each item was assessed. The significance level ($p<0.05$) was adjusted for multiple comparison with Bonferroni correction. (Streiner 1995)

There were a total of 150 assessments (50 patients at 3 time points each) available for analysis. All assessments were analyzed together. We hypothesize that the items will fit the model and have similar location no matter what time (measurement invariance across time) the assessments were done. We then tested for DIF across time to ensure stability of the data.

Validity

There is no "gold standard" test or index for surgical recovery. Therefore, content validity and three aspects of construct validity, (discriminant validity, longitudinal validity and known groups validity) were assessed. The validity of the scale is recognized if the items fit the Rasch model, if they are found to form a single unidimensional linear construct fulfilling the requirement that persons at any ability level are more likely to score higher on easy items than on more difficult items. The extent to which the surgical recovery measure evaluates what it is intended to measure was examined by discriminant validity (convergent and divergent). For convergent validity the surgical recovery measure should correlate highly (>0.7) with indices closely related to recovery such as symptom status and functional status (SF-36 Pain, physical function, role physical, vitality subscales, GIQLI symptoms, physical function subscales) at each point in time. Longitudinal validity (the ability of the measure to detect change in the perioperative period) was assessed by comparing patients' status on the measure preoperatively, 1 week and 1 month postoperatively. We hypothesized that there would be a decline in patients' status at 1 week followed by an improvement during the recovery period at 1 month consistent with the proposed surgical recovery model. For known groups validity, patients with complications were expected to have lower scores on the measure compared to patients free of complications.

To evaluate the responsiveness of the surgical recovery measure, effect size was calculated using Cohen's D. Effect size provides a standardized value for determining the size of change. This was compared to other measures such as the SF-36 physical function.

When the data are found to fit the Rasch model with acceptable fit statistics and absence of DIF, unidimensionality of the construct 'surgical recovery' is confirmed by examining the distribution of the standardized residuals via a principal component analysis (PCA). This identifies any subsets of items that may load together and therefore represent a different construct. Independent t-tests were used to ascertain whether person locations were significantly different if subsets identified on PCA were treated as separate scales. (Smith 2002)

RESULTS

Descriptive Statistics

Sixty patients with symptomatic cholelithiasis entered into the study. Seven patients were excluded: two patients had operations at other hospitals, four had conversion to open cholecystectomy and one had an early complication (abdominal hematoma requiring admission). Three patients refused follow-up visits. Thus, 50 patients remained for analysis. Ambulatory surgery was scheduled for 46 patients and 43 were discharged the day of their operation. 3 patients were admitted: one patient was admitted for pain control (1 day), another to undergo endoscopic retrograde cholangiopancreatogrophy for treatment of a non obstructive stone in the common bile duct found on intraoperative cholangiogram (2 days) and a third for observation after a difficult procedure (4 days). The first postoperative evaluation occurred at a median of 8 [8;12] days and the second at 29.5 [22;34] days after surgery. Patient and operative characteristics are presented in Table 1.

Table 1. Characteristics of Study Population

Characteristics	Mean (SD) (n = 50)
Age, years	51 (16.8)
Male	28%
Working at time of surgery	57%
Body mass index, kg/m^2	26.9 (6.1)
Obese (BMI ≥30)	38%
Indication - Biliary Colic	80%
ASA I/II/III	24%/62%/8%
Duration of surgery, min	66 (19)
Pts with ambulatory surgery, %	86%
Complications, %	30%
Wound infection	8%
Worsening of other health condition	8%
Accidents	4%
Urinary tract infection/retention	4%
Cough	2%
Partial small bowel obstruction	2%
Fever	2%

ASA, American Society of Anaesthesia Score

The mean age of the cohort was 51 (SD15) years; 28% were men, 38% were obese, the majority (86%) were American Society of Anesthesia (ASA) score 1 or 2 and 15 patients reported complications or adverse events during the recovery period. Of the patients reporting complications, 9 were also documented in the medical record: four patients had umbilical wound infections: 2 patients received antibiotics from their family physician or surgeon, 1 patient treated herself with antibiotics for “inflammation” without examination from a physician and 1 patient had drainage from the wound which was observed by the surgeon without additional intervention; two patients had urinary complications (urinary retention and pyelonephritis); one patient was treated with antibiotics for fever for which no cause was identified; one patient was readmitted for partial small bowel obstruction and one patient received antibiotics from a family physician for a productive cough and difficulty breathing. Six patients reported problems during

the recovery period not documented in the medical record. Two patients had lower extremity injuries, three patients had worsening of chronic pain conditions and 1 patient reported increased fatigue and anxiety from caring for her spouse's medical condition.

Table 2 demonstrates the range of values for the indices used to estimate surgical recovery.

Table 2. Median and Interquartile range of values on the indices used to estimate recovery after surgery

Measures	Preop	1 week	1 mo
Generic HRQL (SF-36)			
Physical function	82 [50-99]	60 [45-85]	90 [65-100]
Role physical	87.5 [25-100]	0 [0-25]	75 [0-100]
Bodily pain	58 [45-90]	45 [32-68]	80 [68-100]
General health perceptions	70 [55-80]	70 [60-80]	70 [60-85]
Vitality	52 [28-80]	45 [21-65]	55 [36-80]
Social function	88 [62-100]	75 [50-100]	94 [66-100]
Role emotional	100 [67-100]	100 [8-100]	100 [100-100]
Mental health	78 [64-88]	84[72-92]	84 [76-92]
Physical health (PCS)	47 [36-54]	35 [28-41]	50 [38-55]
Mental health (MCS)	54 [43-57]	53 [43-60]	55 [50-60]
GI Specific HRQL (GIQLI)			
Symptoms (0 to 76)	55 [45-68]	56 [48-67]	66 [58-70]
Emotions (0 to 20)	15 [11-17]	16 [12-19]	17 [14-19]
Physical function (0 to 28)	20 [12-23]	16 [10-20]	20 [14-24]
Social function (0 to 16)	14 [11-16]	10 [9-12]	13 [11-16]
Impairment due to medical tx (0 to 4)	4 [3-4]	3 [2-4]	4 [3-4]
Total (0 to 144)	109 [90-127]	99 [85-121]	119 [103-131]
Physical Activity (CHAMPS)			
Caloric expenditure (kcal/kg/wk)			
All activities	42 [25-61]	18 [9-30]	30 [16-48]
Moderate or greater intensity	18 [0-37]	0 [0-8]	2 [0-28]
Frequency per week			
All activities	14 [7-19]	10 [7-14]	14 [8-18]
Moderate or greater intensity	2 [0-7]	0 [0-2]	0 [0-6]
Hours per week			
All activities	13 [7-18]	6 [4-10]	10 [6-17]
Moderate or greater intensity	4 [0-6]	0 [0-2]	0 [0-6]
Walking capacity (6MWT)	494 [416-540]	410 [339-475]	509 [435-555]
Symptoms (0 to 10 VAS)			
Anxiety	1 [0-3]	0 [0-0]	0 [0-1]
Fatigue	3 [1-5]	2 [0-5]	1 [0-5]
Nausea	0 [0-2]	0 [0-1]	0 [0-0]
Pain at rest	0 [0-3]	1 [0-2]	0 [0-0]
Pain with coughing	0 [0-1]	1 [0-4]	0 [0-0]
Pain with movement	0 [0-3]	1 [0-3]	0 [0-1]

SF-36: subscales range from 0 to 100; PCS and MCS are standardized to have a mean of 50 and a SD of 10.

Over the 150 assessments (50 patients over 3 time points), there was missing data for 35. 2 patients did not complete the GIQLI questionnaire preoperatively and data for 6MWT was missing for 33 evaluations. Compared to baseline assessments, at 1 week postoperatively, patients had significantly decreased scores in the physical function, role physical, vitality, bodily pain subscales and physical component scores of the SF-36, the 6MWT, and the physical function and social function subscores of the GIQLI ($p<.05$). Patients also had less physical activity as measured by total caloric expenditure by CHAMPS 1 week postoperatively ($p<.05$). One month after surgery, while patients still reported lower levels of physical activity compared with baseline, scores for most indices were either at baseline or had increased above their preoperative scores. Figure 5 shows our study population at all time points relative to Canadian norms for SF-36 scores. (Hopman and Group 2000)

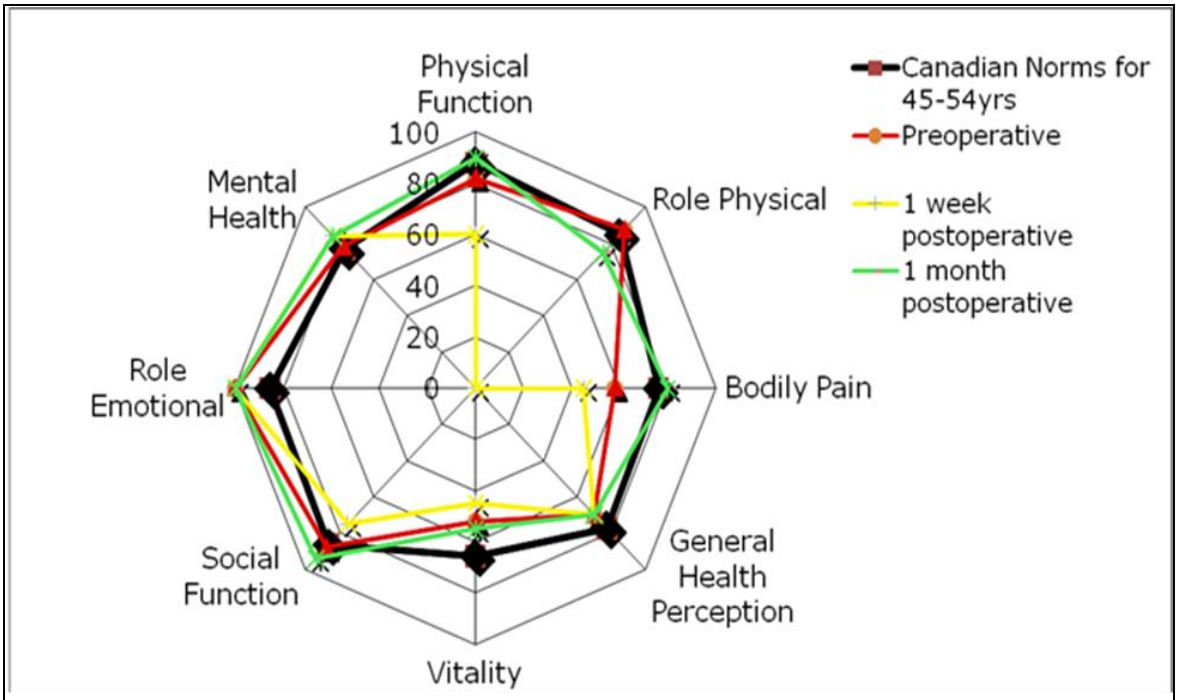


Figure 5. Radar Plot for Scores of SF-36 Subscales in Study Population and Canadian Norms.

Patient scores were close to or slightly below population norms preoperatively, however, by 1 week, they had decreased below these levels. They returned to

baseline or improved at 1 month. The results were similar for GIQLI in which patients were below populations norms (total score 109 vs. 125) preoperatively. (Eypasch et al. 1995) Patients again decreased at 1 week (99) and improved by 1 month (119). There is an equation using gender and age to compute expected 6MWT distance in a healthy population. (Equation 3) (Enright PL 1998)

Equation 3. Reference Equation for the 6-Min Walk Distance in Healthy Adults.

Men:

$$6MWD = (7.57 \times height_{cm}) - (5.02 \times age) - (1.76 \times weight_{kg}) - 309 \text{ m}.$$

Alternate equation using BMI[*]:

$$6MWD = 1{,}140 \text{ m} - (5.61 \times BMI) - (6.94 \times age)$$

When using either equation, subtract 153 m for the LLN

Women:

$$6MWD = (2.11 \times height_{cm}) - (2.29 \times weight_{kg}) - (5.78 \times age) + 667 \text{ m}.$$

Alternate equation using BMI:

$$6MWD = 1{,}017_{m} - (6.24 \times BMI) - (5.83 \times age)$$

When using either equation, subtract 139 m for the LLN

Definition of abreviations: BMI = body mass index; 6MWD = 6-min walk distance; LLN = lower limit of normal.

[*] BMI in kg/m^2.

Using this equation, study patients were within the lower limits of expected 6MWT distance preoperatively for their age and gender. Their performance declined below the lower limits at 1 week postop before returning to baseline at 1 month. While there are no population norms for the CHAMPS questionnaire, preoperatively, the patients in our study group were within the range of values reported in the literature for this index. (Stewart 1997; Stewart 2001; Stewart 2001; Lusardi 2003; Resnicow 2003)

Measure Development

Recoding

Disordered thresholds were observed for 54 of 76 polytomous items. The categories were not ordered as expected from low to high. Several of the items had less than 10 responses in the lower categories and these categories were recoded. Furthermore, few patients responded to the middle categories of the items reflecting trouble differentiating between the response options of these items. Therefore, 54 items were rescored by collapsing the middle categories. (Appendix)

Item reduction

After rescoring 71% (54 items) of the polytomous items with disordered response thresholds, the fit statistics were evaluated and the poorest fitting items were removed iteratively. 45 items were removed (the list of deleted items and the reasons for deletion are given in the appendix). After each deletion, all fit statistics and response options were re-examined. The standardized residual fit statistics of the deleted items ranged from (6.24 to -3.58). They included items related to the extent to which patients restricted the kinds of food they eat or their frequency of awakening in the night. Items irrelevant to the construct were also deleted. Items with inter-item residual correlations >0.5 were also deleted. After deletions, 34 items remained. Together, these items formed a pool of calibrated items with good fit to the model representing the construct of surgical recovery (mean item-standardized residuals = -0.015 (0.811), mean person residuals = -0.181 (0.99)). There was good match between person ability and item difficulty (mean person ability was 0.877, SD=1.44). The final surgical recovery measure consisting of 34 items was chosen based on relevance to the underlying construct, lack of DIF and precision.

Properties and structure of surgical recovery measure

After deletions, thirty-four items fit the model with a global fit statistics of an item-trait interaction χ^2 of 48.1 (df:68 , p>0.97), person and item fit residuals of -0.18 (0.99) and -0.02 (0.81) respectively. All item and person fit statistics met the requirements of the Rasch model. The item fit statistics met the criterion value of +/-2.0 (range -1.64 to 1.50).

Item precision varied from 0.09 logits SE to 0.28 logit SE. The 34 items listed in Table 3 operate together and define the spectrum of surgical recovery without floor or ceiling effects.

Table 3. Items from SF-36, GIQLI, VAS, CHAMPS and 6MWT retained in the Rasch measure of Surgical Recovery

Item	Source	Item #
1. How often during the past week, have you been troubled by **uncontrolled stool?**	GIQLI Q36	I0056
2. How often during the past week, have you been **sad about being ill?**	GIQLI Q11	I0033
3. **VAS Pain with coughing**	VAS	I0063
4. Does your health now limit: **Bathing and Dressing yourself**	SF36 Q12	I0010
5. Since becoming ill, have you been troubled by **changes in your appearance?**	GIQLI Q18	I0040
6. **VAS Pain with movement**	VAS	I0052
7. **VAS Pain at rest**	VAS	I0062
8. During the past week, how well have you been able to **cope with everyday stress?**	GIQLI Q10	I0032
9. How often during the past week, have you been **happy with life in general?**	GIQLI Q13	I0035
10. How often during the past week, have you been **frustrated about your illness?**	GIQLI Q14	I0036
11. During the past week, have you had to **cut down the amount of time spent on work or activities because of emotional problems?**	SF36 Q14	I0014
12. During the past week, how much have you been troubled by **medical treatment of your illness?**	GIQLI Q24	I0046
13. How often during the past week, have you been troubled by **nausea?**	GIQLI Q33	I0055
14. To what extent has your **sexual life been impaired (harmed)** because of your illness?	GIQLI Q26	I0048
15. How often during the past week, have you had a feeling of **fullness in the upper abdomen?**	GIQLI Q2	I0024
16. During the past week, to what extent has your physical health or emotional problems **interfered with your normal social activities with family, friends**, neighbors, or groups?	SF36 Q15	I0015
17. Does your health now limit: **Bending, kneeling, or stooping**	SF36 Q8	I0006
18. How often during the past week, have you had **bloating?**	GIQLI Q3	I0025
19. How much **bodily pain** have you had during the past week?	SF36 Q16	I0016
20. Does your health now limit: **Climbing several flights of stairs**	SF36 Q6	I0004
21. During the past week, how often have you been able to **take part in your usual patterns of leisure or recreational activities?**	GIQLI 23	I0045
22. Because of your illness, to what extent do you **feel unfit?**	GIQLI 21	I0043
23. How often during the past week, have you had **pain in the abdomen?**	GIQLI 1	I0023
24. How often during the past week, have you had found **eating to be a pleasure?**	GIQLI 8	I0030
25. During the past week, how much of the time did you **feel worn out?**	SF36 Q29	I0021
26. During the past week, how much did the **pain interfere with your normal work or activities?**	SF36 Q22	I0017
27. During the last week, how much of the time did you **feel tired?**	SF36 Q31	I0022
28. **In general, would you say your health is:**	SF36 Q1	I0011
29. During the past week, how much of the time did you **have a lot of energy?**	SF36 Q27	I0020
30. Does your health now limit: **Vigorous activities (running, lifting heaving objects, participating in strenous sports)**	SF36 Q3	I0001
31. During the past week, how much of the time did you **feel full of pep?**	SF36 Q23	I0018
32. 6MWT	6MWT	I0083
33. Over the last week, did you **do any volunteer work?**	CHAMPS D855	I0066
34. Over the last week, did you **play any sports? (bowling, golf, ski, skate, basketball, soccer, racquetball, tennis, hockey)**	CHAMPS D9201	I0072

Source: Original or recoded item from source questionnaire

Each item had n thresholds where n is defined as number of response categories minus 1. The items retained for the surgical recovery measurement method are depicted in Figure 6 with the person distribution on top and the item distribution on the bottom and conceptualizes and quantifies recovery after surgery. The distribution of the 50 patients throughout 3 time points across the measure of surgical recovery is shown. The horizontal axis is scaled in logits and represents recovery with the least level of functioning at the left to the highest level of function at the right. The vertical axis represents a frequency distribution of patients and item thresholds at each location. The precision of the measurement of person ability by the items is demonstrated by the Test Information Function line in top part of Figure 6. (Penfield 2005)

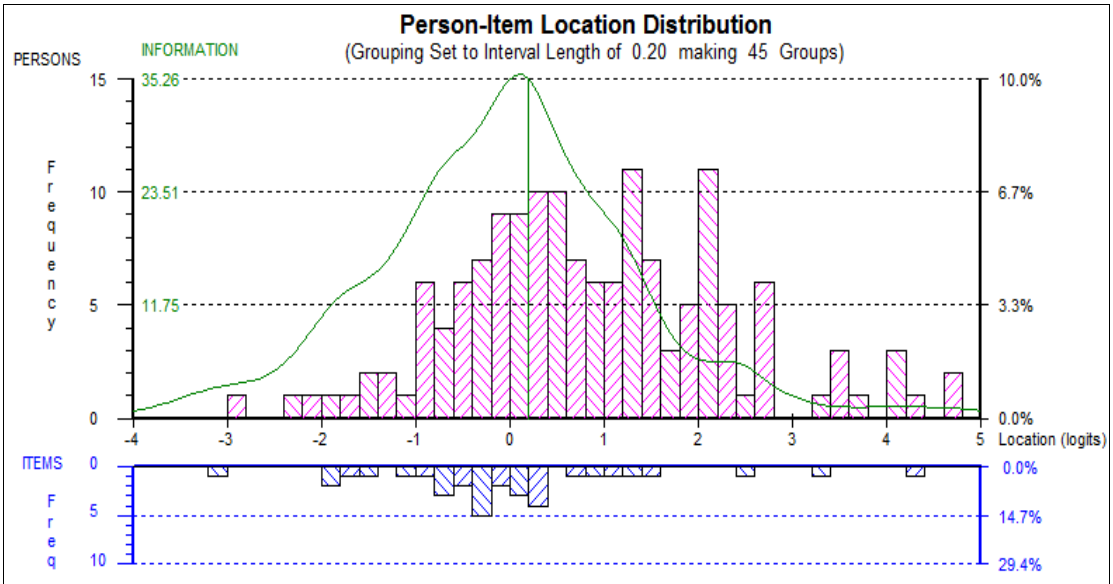


Figure 6. Person-Item Location Distribution of Rasch Measure of Surgical Recovery.

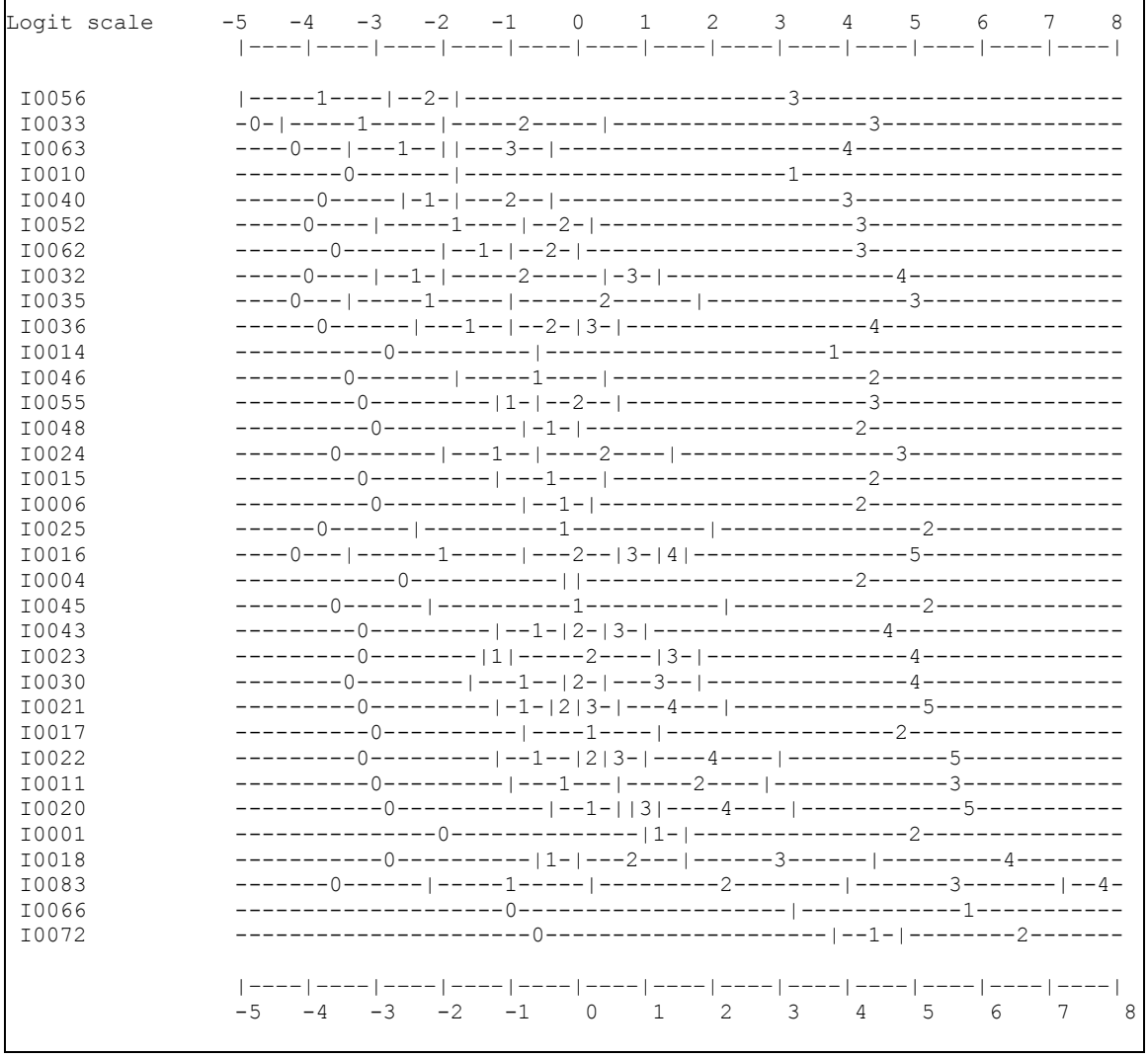
Person ability is well estimated between -2 logits and +2 logits, however the precision decreases at +2 logits. The internal consistency and Cronbach's alpha could not be calculated due to the presence of missing data. Reliability of the person hierarchy was measured with a Person Reliability Index of 0.96.

The item threshold difficulty ranges from -3.1 logits for the item "uncontrolled stool" to 7.4 logits for the item "6MWT > 800m". The range of person abilities

ranged from across 7.4 logits from -2.8 logits of ability to +4.6 logits. The majority of patients were situated between -1 and +2 logits.

Figure 7 shows the item map with the location of each item threshold based on and the rescored response options.

Figure 7. Item Threshold Map of Rasch Measure of Surgical Recovery.



The distance between the response options represents the range of difficulty between the two. Difficulty increases from left to right. The small vertical line is a threshold point, the location at which a person with that level of ability (on the

logit scale) has an expected 50% probability of responding to one category or the other.

No patients were at the 'ceiling' or the floor of the measure as the items spanned a greater range of difficulty than the patients.

The unidimensionality was demonstrated by overall model fit and by a principal component analysis of the Rasch person-item residuals. The first component explained 10.1 % of the remaining variance. (Smith 2002) There were two secondary patterns found within the residuals of 8 items. When person estimates of these items were compared to the total item set, 89% and 88% of the t tests were insignificant respectively. (Tennant 2006) The distribution of residual variance was random in nature, though not as low as would be expected if the measure was perfectly unidimensional.

The measure was examined for DIF. The difficulty level of the items did not fluctuate with gender. Three items had evidence of DIF with time of assessment. The items "how often have you been nervous or anxious about you illness?" from the GIQLI, "does your health now limit you in moderate activities?" from the SF-36 and fatigue from the VAS all had evidence of DIF with time. These items did not yield the same item response function; that is, patients with the same ability who would be expected to answer these items in the same way did not. Therefore these items did not work the same way for patients at the preoperative, 1 week or 1 month postoperative time points. At equivalent levels of ability, patients at 1 week postop were more likely to have less anxiety or fatigue as compared to patients preoperatively or 1 month postop. Also, patients preoperatively were less likely to be limited by moderate activities as compared to patients postoperatively. It can be hypothesized that patients may have answered these items differently depending on postoperative expectations. For this reason, the items were excluded from the scale. Both items and patients appear along the same scale, the Rasch logit scale. We then performed a linear transformation of the logit scale to fit a calibrated scale from 0-100. Higher values on the 0 to 100 scale represent more difficult items and higher functional ability. The scale was calibrated from 0 to

100 to reflect the range of -4 or +4 logits. There were some items and some patients outside of this range thus with scores greater than 100 or lesser than 0 however, they represented less than 4% of the sample. (Figure 8)

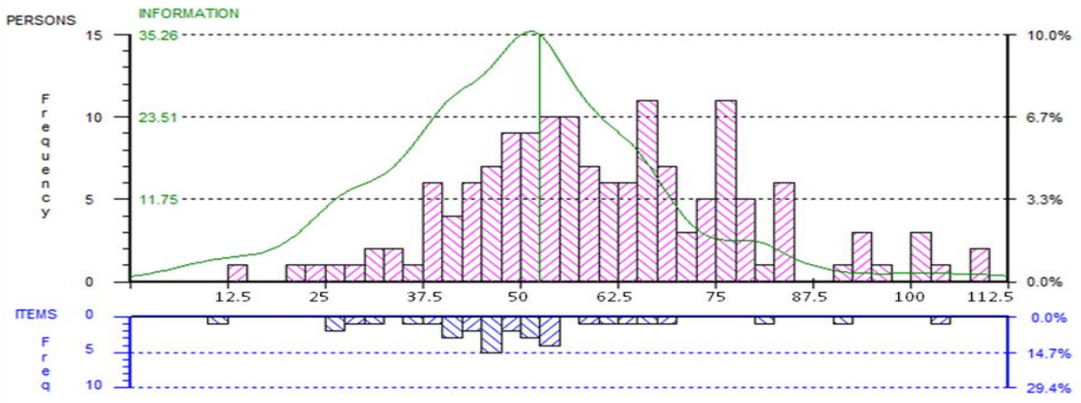


Figure 8. Person-Item Location Distribution of Rasch Measure of Surgical Recovery with Rescaled Score from 0-100.

The average score on the transformed surgical recovery scale was 61.0 (18.0) or 0.877 (1.44) logits. Therefore the average patient likely had positive responses to the first 30 items and negative responses to the last 8.

Psychometric Properties of New Measure

Content Validity

This measurement method for surgical recovery covers a broad spectrum of difficulty across 10 logits with items from all indices relating to: symptoms (9); physical function (10); psychological function (12) and social function (2). Symptoms included in the measure were generic like pain and nausea and gastrointestinal-specific such as uncontrolled stool and bloating. The items are organized as theorized from low level symptom items (VAS pain at rest: 38.6) to more difficult activity items (Vigorous activity SF-36: 66.9) Items were arranged from lower psychological states (frustrated: 42.4) to higher states (full of pep: 68). The hierarchy of activity items in the Rasch analysis supports the content validity

of this surgical recovery measure method. Basic activities such as ‘bathing’ or ‘dressing’ (score 28) were easier than more vigorous activity items such as ‘climbing several flights of stairs’ (50) which were easier than more intense activities such as ‘sports’ (104.4).

Convergent construct validity

Table 4 presents Pearson correlations between the new method to measure surgical recovery and the other indices.

Table 4. Spearman rank correlation coefficients between the Surgical Recovery Measure and other indices at all time points

Index	Surgical recovery Measure	Preop	1 week	1month
Generic HRQL (SF-36)				
Physical function	0.68	0.58	0.72	0.72
Role physical	0.71	0.81	0.57	0.74
Bodily pain	0.78	0.85	0.81	0.63
General health perceptions	0.61	0.63	0.64	0.63
Vitality	0.88	0.91	0.83	0.88
Social function	0.72	0.77	0.70	0.70
Role emotional	0.54	0.62	0.51	0.45
Mental health	0.68	0.74	0.68	0.68
Physical health (PCS)	0.77	0.77	0.79	0.74
Mental health (MCS)	0.67	0.71	0.70	0.64
GI Specific HRQL (GIQLI)				
Symptoms (0 to 76)	0.83	0.88	0.85	0.78
Emotions (0 to 20)	0.78	0.82	0.77	0.81
Physical function (0 to 28)	0.85	0.80	0.83	0.83
Social function (0 to 16)	0.75	0.82	0.71	0.74
Impairment due to medical tx (0 to 4)	0.62	0.62	0.55	0.62
Total (0 to 144)	0.92	0.94	0.93	0.89
Physical Activity (CHAMPS)				
Caloric expenditure (kcal/kg/wk)				
All activities	0.32	0.40	0.26	0.41
Moderate or greater intensity	0.32	0.47	0.25	0.38
Frequency per week				
All activities	0.27	0.25	0.39	0.38
Moderate or greater intensity	0.30	0.39	0.15	0.41
Hours per week				
All activities	0.31	0.38	0.28	0.35
Moderate or greater intensity	0.36	0.51	0.34	0.38
Walking capacity (6MWT)	0.57	0.55	0.49	0.56
Symptoms (0 to 10 VAS)				
Anxiety	-0.51	-0.55	-0.53	-0.52
Fatigue	-0.69	-0.76	-0.66	-0.67
Nausea	-0.55	-0.66	-0.53	-0.46
Pain at rest	-0.62	-0.67	-0.64	-0.54
Pain with coughing	-0.59	-0.62	-0.64	-0.45
Pain with movement	-0.65	-0.67	-0.67	-0.59

The total scores from this method correlated well with all indices at all time points and this demonstrating convergent construct validity. The Rasch measure had low

to high correlations for most indices ranging from 0.25 to 0.94. All correlations were statistically significant ($p<0.05$). The measure also had low correlations with gender (0.24-0.38).

Longitudinal validity

The distribution of the surgical recovery measure at all time points is depicted in Figure 8. Compared to baseline score of 60 (18), patients decreased to a score of 55 (15) on the scale at the 1-week postoperative evaluation ($P<0.02$). This increased during the recovery period to 68 ($P<0.01$) and this score was significantly above the baseline level ($p<0.001$). (Figure 9)

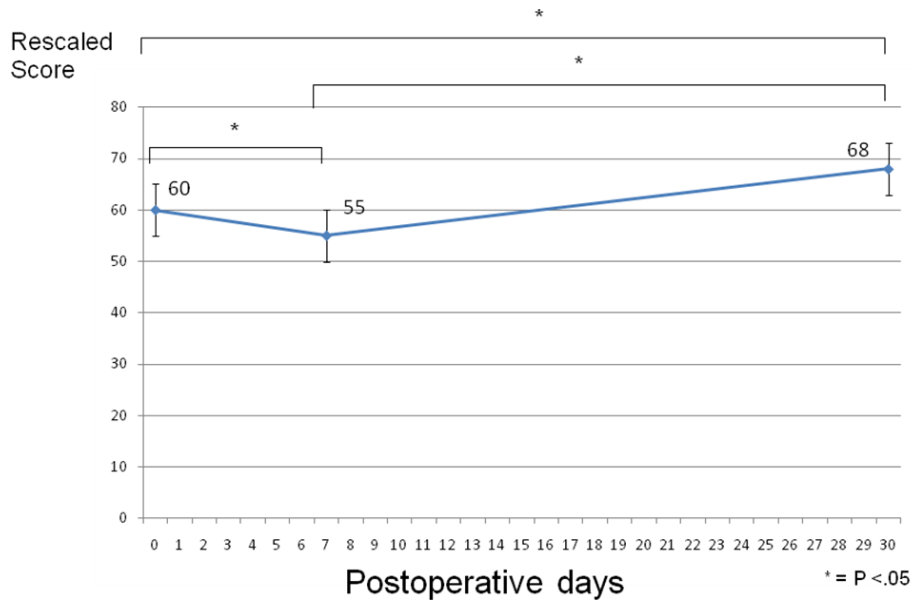


Figure 9. Changes in scores of Surgical Recovery Measure during recovery from operation.

Patients in the study sample were distributed over a large range of scores, from a minimum of 15 to a maximum of 100. Postoperative recovery can be estimated as the percentage of patients returning to baseline scores. Return to baseline was defined as +/- 10% of baseline scores. At 1 week postop, 48% of patients were below their baseline, 30% were at baseline and 22% were above. At 1 month, 14% were still below baseline levels, 26% were at baseline and 60% were above.

Known-groups construct validity

Fifteen patients reported complications during the follow up period. Baseline scores of patients with or without complications were not statistically different (54 (22) vs. 62 (16) respectively, P=0.19). One week postoperatively, patients with complications had a lower score as compared to patients who were free of complications (48 (14) vs. 59 (14), P<0.02). At one month postoperatively, the difference between the groups persisted (57 (19) vs. 73(46), P<0.01). Though patients with complications returned to their baseline scores (P=0.66), patients without complications surpassed their baseline (P<0.01). (Figure 10)

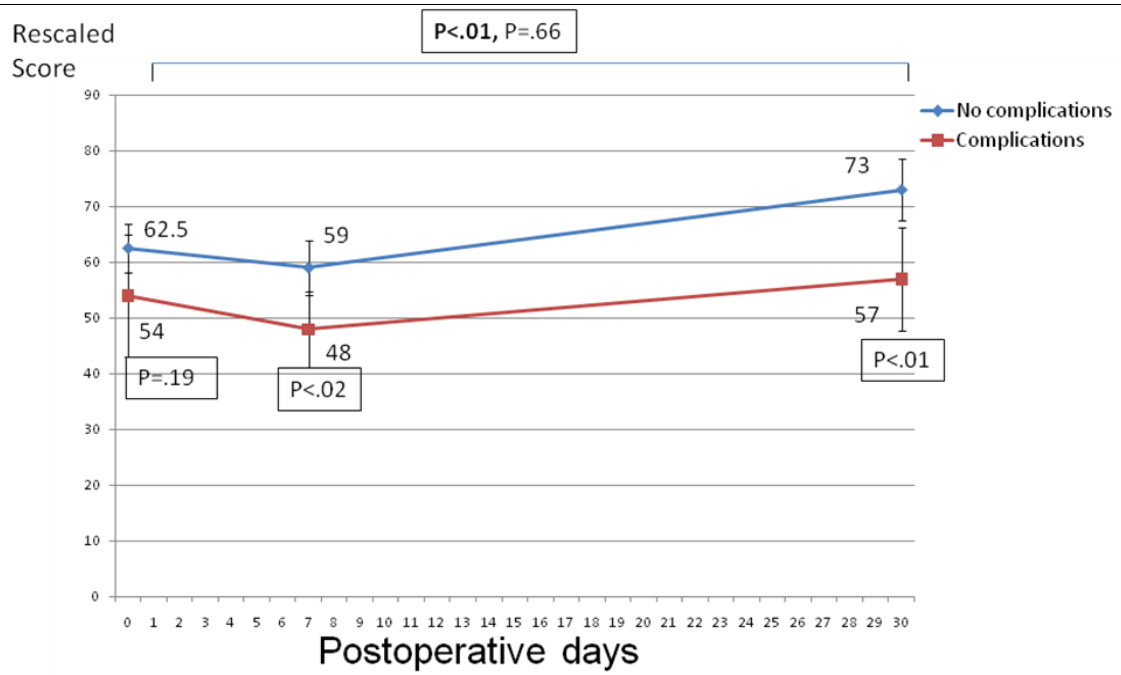


Figure 10. Score of Surgical Recovery Measure in Patients with or without Complications.

Responsiveness

To assess responsiveness to change over time, Table 5 shows a comparison of the effect size for the surgical recovery measure and other indices. The Rasch measure has a smaller effect size to assess changes from baseline to 1 week

postop compared to the SF-36 physical function subscale and the CHAMPS caloric expenditure. It is slightly more responsive during the recovery period from 1 week to 1 month after surgery.

Table 5. Effect size of Surgical recovery measure and other indices

	RASCH measure	**SF36 PF**	**CHAMPS**
Preop to 1 wk	-0.256	-0.3395	-0.91
1 wk to 1mo	0.74	0.65	0.64
Preop to 1mo	0.44	0.26	-0.29

DISCUSSION

THE ROADMAP AHEAD

There is a need for a relevant, comprehensive, easy to administer, valid method of measuring recovery after surgery in order to better prepare and educate patients and evaluate surgical innovations.

A method comprising 34-items drawn from existing indices and tests calibrated reliably to a linear scale was devised for patients undergoing laparoscopic cholecystectomy using Rasch methodology. The new method demonstrated no ceiling or floor effects and had initial evidence for content, cross-sectional, longitudinal and known-groups construct validity. The method was clinically responsive and could discriminate between patients with complications from those without. It had moderate correlations with indices used in its development. The method expands on the range of assessment tools currently available for surgical recovery by encompassing multiple dimensions including symptoms, physical, psychological and social functions. The level of function of any individual can be measured and thus their recovery trajectory can be described. The Rasch logit

scale developed has been converted to a scale ranging from 0 to 100. A score of 50 represents a functional ability of 0.0 logits. This person is likely respond favorably to the items below this level but would probably not respond favorably to items above this level.

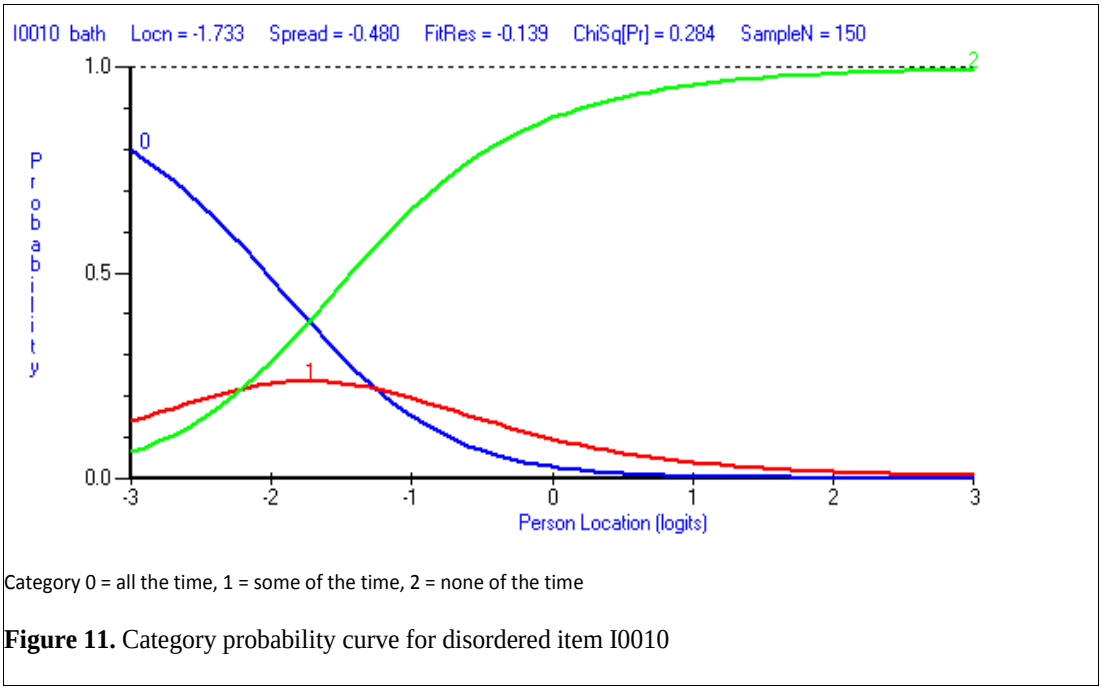
Patients' scores on the indices commonly used to estimate recovery after surgery are shown in Table 2. They demonstrate the difficulty in describing the recovery process. Most scores show similar patterns of trajectory, they decrease at 1 week postoperatively before returning to baseline or above at 1 month. However, they do not all show this trajectory and the amount of change is variable between scores. Furthermore, because the scores are largely based on ordinal categories, it is difficult to attribute meaningful interpretations to these changes. For example, there are many ways to obtain a score of 82 on the physical function subscale of the SF-36. This subscale is a mean summary score of 10 items each scoring 0, 50 or 100. Therefore an average total score of 82 on this subscale is indicative of limitations some of the time to some activities and this can be obtained from various combinations of responses. Without analyzing the items individually, it is challenging to determine which items are more problematic. Based on the total score, we are unable to determine the level of recovery for any single individual. These limitations apply to most scores that are based on ordinal categories which make up a large part of the instruments used today. In contrast, this method of measuring surgical recovery provides a transparent single linear measure with item hierarchy that allows the required information to interpret a person`s ability based on their score.

The difficulty levels of the 34 items cover a wide range of symptoms, levels of physical, psychological and social functions, Although some patients had maximum or minimum scores on some of the indices, this was not the case with this measurement method. The items were able to cover the broad range of difficulty with no important gaps. There were a few items in the higher end of the scale (>4 logits or score >100) however, the end of the spectrum represented very high functioning activities and did not apply to the majority of the patients in the

sample. Thus with the majority of patients situated in the middle range of the scale (-2 to +2 logits or score of 25 to 75), the measure cannot be said to be too easy or too difficult.

The choice of instruments used for the development of the surgical recovery measure was based on previous studies. A recent review by Korolija et al. suggested the use of the SF-36 (generic instrument) in conjunction with the GIQLI (disease-specific instrument) questionnaires for assessment of quality of life for patients undergoing laparoscopic cholecystectomy based on several trials. The postoperative time points suggested for QOL assessments after this type of surgery were 1 month and 6 months. (Korolija 2004) We therefore included the items from these indices in the development of the method.

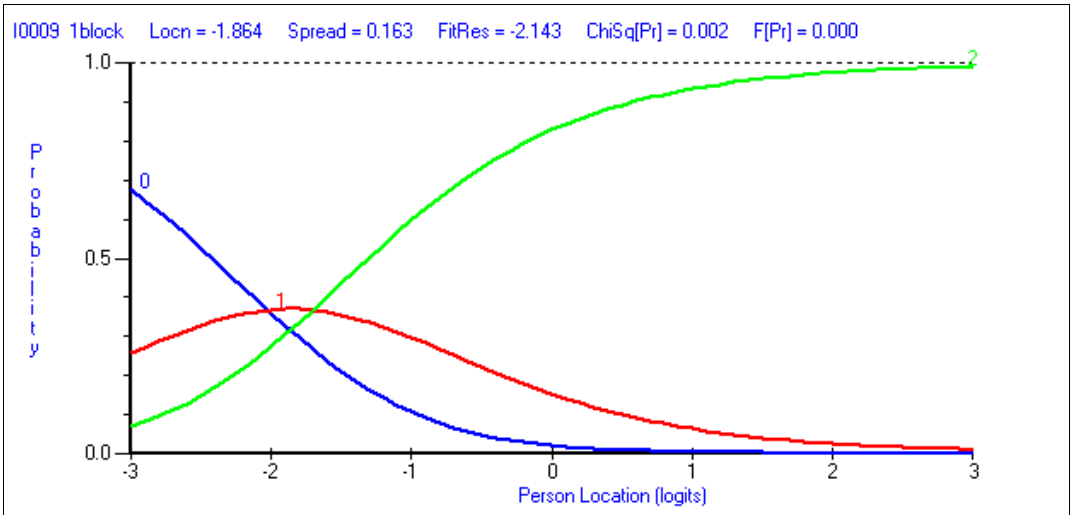
The Rasch analysis revealed items with disordered response options and these were rescored. Disordered categories indicate that response choices may be poorly worded and confusing for patients or items are multidimensional or irrelevant. (Wright BD 1982) Collapsing of response categories especially middle categories was necessary as patients were not able to discern between these response options. For example, the physical function item from the SF-36 “Does your health currently limit you in these activities: bathing or dressing yourself?” was initially scored 0 for the response “all the time”, 1 for “some of the time” and 2 for “none of the time”. There were six responses for the category “all of the time”, 12 for “some of the time” and the other 132 responses were “none of the time. Through the Rasch analysis, it was found that this item was disorded. Figure 11 shows the category probability curve for this item. It is evident that there is disordered threshold for item.



Category 0 = all the time, 1 = some of the time, 2 = none of the time

Figure 11. Category probability curve for disordered item I0010

A person with low function is likely to receive a score of 0. As a person's functional status increases along the x-axis, they are more and more likely to score 1 or 2. At the approximate location of -1.7, there is an equal probability of scoring 0 or 2. This is the threshold between scores 0 and 2. Patients with higher ability are then more likely to score 2. There is however, no threshold for the category 1. Thus this item is disordered and patients are not able to discern between the category "some of the time" and the other categories. Thus this item needs to be rescored and it was done by collapsing the categories "some of the time" and "all of time". Conversely, Figure 12 shows the category probability curve for the item:" "Does your health currently limit you in these activities: waking 1 block?" which is not disordered. Decreasing the number of response options did not compromise the ability of the recovery measure to assess postoperative functioning.



Category 0 = all the time, 1 = some of the time, 2 = none of the time

Figure 12. Category probability curve for nondisordered item I0009.

Evidence for validity was presented by the fit of the data, the item-person-hierarchy, high correlations between the measure and other indices, and variation longitudinally, cross-sectionally and by contrasting sub-groups known to differ (known-groups validity). The ordering of the items within a dimension was as expected. For example, there have been studies in which the Rasch analysis was used to evaluate the hierarchy of the SF-36 physical function subscale. (Jenkinson 2001) The ordering of the items from this subscale in the Rasch measure was consistent with previous studies. Across dimensions, it was more difficult to predict as little is known how these dimensions interact during the process of recovery. However, the final ordering of items showed initial content validity. The final method included 13 items from the SF-36, 15 from the GIQLI, the 6MWT, 3 VAS items and 10 items from CHAMPS which were recoded as 2 items for the Rasch analysis

In qualitative research studies, patients define recovery after surgery as absence of symptoms and return to “normal activities”. Likewise, we believe the process of recovery involves multiple dimensions for which there are no validated instruments. HRQL instruments capture many important aspects of this process

but do not integrate all the relevant aspects on an interval scale. The surgical recovery measurement method enables quantification of both these concepts on a single linear scale. The scores on this scale are meaningful and directly attributable to patients` symptoms status and functional status. We chose a study population consisting of patients undergoing laparoscopic cholecystectomy because it is a well-defined procedure with a somewhat homogenous population which would allow us to focus on the properties of the method. As a group, patient scores on the surgical recovery measure decreased one week after surgery before improving above baseline at 1 month. While recovery is a dynamic process, for practical purposes, we were only able to assess two time points. We are therefore only obtaining a snapshot of patients' along their recovery trajectory. The results from these time points suggest that much of the decline in patient function has occurred during the 1st week postop. By 1 month, patients are above their baseline function. This also suggests that biliary colic which accounted for 80% of the indications for surgery is a chronic disease process. Patients' disease burden is demonstrated through lower scores at baseline as compared to 1 month postoperatively. Thus, the disease process prior to surgery clearly has an impact on patients' functional status suggesting that preoperatively patients are below their "true" baseline.

Limitations

The surgical recovery method was developed and validated in a single sample of 50 patients undergoing a specific procedure. Central to the development of a measure is the concept of unidimensionality. Surgical recovery is a complex process composed of several dimensions with items from physical, psychological and social functioning. Our measure was tested for evaluated for unidimensionality by a principal component analysis of the Rasch person-item residuals. There were two secondary patterns found within the residuals of 8 items suggesting these items correlate highly to each other potentially forming a second construct. Thus, surgical recovery as a concept defined by our measure is not perfectly

unidimensional. Unidimensionality however, is an ideal and similar constructs such as health-related quality of life cannot be perfectly unidimensional. Nevertheless, for practical purposes, development of measures can still prove useful.

Information from repeated measurements from 3 different time points was treated as independent observations. A multi level item response theory analysis might have been an appropriate strategy in this case however, given the complexity of this type of model and the limitations of our sample size, this type of analysis was not considered.

The same data set was used to validate the new measure. Ideally, the validation study for the new measure should be conducted with a different data set and this may be the focus of further studies.

Although our patients have demographic characteristics typical of those undergoing laparoscopic cholecystectomy, it is difficult to generalize beyond our sample since few studies have incorporated all these dimensions together. However, previous studies have shown similar trajectories after surgery. Similar to our results, a randomized trial by Barkun et al. showed significant improvement in QOL as measured by the GIQLI questionnaire at 1 month. (Barkun JS 1992) Burney et al. also showed similar SF-36 scores preoperatively and 2 months postoperatively after laparoscopic cholecystectomy. (Burney 2002) Thus, while there have been studies investigating QOL after laparoscopic cholecystectomy using these instruments, there have been no previous study to incorporate multiple dimensions along a single unidimensional continuum. We feel our sample adequately represents the patient population undergoing laparoscopic cholecystectomy and the Rasch measure reflects expected changes after surgery. While the sample seems representative of the larger patient population undergoing laparoscopic cholecystectomy, it only consisted of 50 patients with some missing data. Further testing and validation with other surgical patients with gastrointestinal conditions is needed. Many items needed rescoring including the items from the CHAMPS questionnaire and the 6MWT. A limitation of the RUMM program is its inability to incorporate continuous data. Therefore, in the

case of the 6MWT, clinically meaningful thresholds were created and the item was rescored with 8 response options. Similarly, information regarding energy expenditure from the CHAMPS questionnaire was not included in the analysis. Instead, items were recoded using ICF classification. This led to 13 recoded items. Thus only information pertaining to patients' participation to the items was included. This was the first time that the items from the CHAMPS have been coded to the ICF and it was done by only two persons and hence there may be some unreliability in the assignment of codes.

The method incorporated mostly of self-report items. There was only one capacity item, the 6MWT. The capacity item has the disadvantage of requiring assessment of patients by trained personnel in the hospital. This accounted for much of the missing data, as the self-report indices could be completed by telephone.

Finally, the change in functional status as seen with the surgical recovery measure was smaller than hypothesized. This method detected only a 10% decrease in score from a mean of 60 preoperatively to a mean of 55 at 1 week postoperatively. There are several possible explanations for this observation. Firstly, patients consider recovery as an absence of symptoms and a return to normal activities. A sizeable proportion of the items related to symptoms relevant to the postoperative process. The change seen at 1 week postop may reflect the relatively quick resolution of symptoms after this type of surgery. Furthermore, the 1 week time point may be too late. The steepest decline in functional status for this type of surgery may in fact take place within the first few hours or days after surgery. Thus, at 1 week, patients may be at on the flatter part of the recovery trajectory. To better elucidate this trajectory, patients need to be evaluated at earlier time points in future studies.

Future Directions

Knowledge translation

We have outlined a process to develop a method for measuring a complex construct like surgical recovery. With further validation of this method, it could be used to educate patients as to expected recovery post-surgery. Because the items are organized on a linear scale, we can better define patient functional status at different points during the recovery phase. Given a patient's baseline functional status, postoperatively, we can inform them of expected symptoms and of activities they should be able to perform. In this way, patients and their physicians can quantitatively track their progress through their recovery from surgery. Moreover, with further studies, we can obtain population based expectations at different time points. If a patient does not meet these expectations (i.e. functional status is not improving as anticipated or even regresses), it may be a signal for patients to seek medical attention. In our study, patients with complications were at significantly lower levels as compared to patients who were free of complications. Thus this method could ultimately assist patients and physicians in the prompt diagnosis of postoperative complications or other factors that may affect recovery. We feel this method allows patients to be proactive about their postoperative course.

Computer adaptive testing (CAT)

The Rasch model is the one-parameter IRT model and is used to develop measures. By using all the parameters of the IRT model it is possible to identify a more generalized method to measure people with a wide range of abilities. This process of measuring people is facilitated by computer technology, computer adaptive testing (CAT). CAT has changed the process of estimating latent traits or abilities that cannot be directly observed. CAT was originally developed in the setting of educational testing (Lord 1968) but has since emerged in medical and rehabilitation studies. (Ware 2000; Dijkers 2003) It has several advantages over

traditional pencil and paper instruments or computer administered tests. It allows one to administer items that are matched to the person's level of ability thereby avoiding unnecessary items that may be uninformative and extend unnecessarily to test taking time. In contrast to a test that requires a person to respond to all the items, CAT selects only items that are appropriate to the person's ability forgoing items that are either too easy or too difficult. (Revicki 1997) This allows one to administer fewer items thus reducing the burden for the person while maintaining precision of the estimates. It also allows the easy introduction and testing of new items into a measure. (Sands 1997) At this time, it is unlikely that busy surgical clinics will move to CAT measures but it is possible even at this stage with only the one-parameter model to present the items in such a way that patients and surgeons can quickly situate the patient on the linear hierarchy. All of the items would be presented and expected progression can be mapped on paper

Validation in other populations

The Rasch models may be useful to develop methods for measuring surgical recovery in other populations. The next step is to extend this method to different populations (eg. elderly, pediatric, patients with specific conditions) undergoing different types of procedures. There will likely be common items among these measures; however we may find different items from different indices or tests to be relevant. As the concept of recovery has not yet been consistently defined, Rasch models can help us gain an understanding of this complex process. Thus we have described a method to develop a quantitative and harmonized measure of surgical recovery in a specific population. Further studies with larger sample sizes are needed to further validate this method. To validate this method for other populations (e.g. patients undergoing inguinal hernia repair), two different approaches could be applied. The existing method could be administered to a sample of people undergoing a different surgical procedure and the method evaluated for reliability, validity and responsiveness. This would provide evidence to support generalizability of the measure in this sample of patients.

Alternatively, all the indices and tests used in the development of the original Rasch measure could be administered to this new sample of patients. Using the Rasch model, one can then create a measure of recovery from the data of this new sample of patients. The differences in items and item locations can be qualitatively and quantitatively compared. As most of the indices and tests used in this method were generic or gastrointestinal-specific instruments, many of the items could be applicable to quantifying the recovery from other procedures in general surgery. Because the concept of recovery is still obscure, determining which indices or tests to use in the development of a Rasch measure is uncertain. In order to delineate aspects of the recovery process that may be important to patients, qualitative studies may prove useful. In previous studies, patients defined recovery from surgery as absence of symptoms and return to normal activities. For this reason, we felt it was important to include these aspects into this method. While this measure is the most comprehensive to date, there may be aspects of recovery not well represented by the instruments used in the analysis. Social and economic aspects may play an important role in recovery and warrant further study in future measures. Other instruments might be considered in the development of future measure to cover these domains. Thus, further qualitative studies may build on our understanding.

Defining trajectories after surgery

Not all patients experience the recovery period in the same way. Firstly, individually, patients start at different baseline functional levels. As they progress through the recovery process after surgery, they follow different trajectories. In our study, we have outlined the trajectory for patients as a group. However, there are patients that are at various levels of function along the scale. While most patients improved above their baseline levels at 2 months, there remained a number of patients who were well below these levels. In this study, 3 time points were covered: preoperatively, 1 week and 1 month postoperatively. We are able to observe patients' progression as a group through the recovery period at these

times. However, I was unable to model the extent to which individuals vary in their trajectory and why they might vary. With more longitudinal data and a greater sample size, we can define the different trajectories that patients experience after surgery using multilevel regression models. These models have the ability to investigate differences between individual trajectories and factors that might influence them. (Pastor 2003) For example, researchers could explore whether patients with high baseline functional status recover at a faster rate than patients with low initial status. Thus with Rasch measures like the one developed, complex multilevel models can be used to study patient postoperative trajectories over time and their predictors.

Evaluating outcomes

Finally, the most direct use of the surgical recovery measure is to quantitatively compare outcome of surgical innovations. As our understanding in anesthesia and surgery increases, improvements in surgical care are becoming more refined. Traditional outcomes and instruments do not have the sensitivity to identify the improvements from these innovations. Objective evaluation has been hampered by lack of a suitable outcome of recovery. Thus when evaluating surgical innovations such as minimally invasive surgery and fast track surgery which are designed to “shorten recovery”, this measure or other measures developed using the Rasch methods can provide an objective outcome of surgical recovery. Recovery after surgery is an outcome that is important to patients and the surgical recovery measure captures all relevant dimensions involved in this process on a hierarchical scale.

CONCLUSION

In conclusion, the Rasch analysis has identified 34 items for a measure of recovery after surgery. The measure demonstrates good initial psychometric properties without floor or ceiling effects and encompasses a large range of dimensions by including many HRQL instruments. It covers a broad range of clinical conditions during the surgical recovery process. It has expanded our understanding of recovery after surgery. Rasch models illustrate that recovery can be quantified in mathematical units and show promise in the development of future measures.

GLOSSARY

The terminology used in the field of measurement is often unfamiliar to clinicians and hence terms that have specific scientific meaning are often misused or used inconsistently. In this thesis, we will use the following definitions.

- **Constructs/domain** An *intangible, theoretical entity that is operationalized into one or more items* ***(Sloan JA 2002)***
- **Item** – *A single question which can be used as a stand alone question, as part of a series of loosely affiliated questions, or as part of a psychometrically sound measurement index* ***(Sloan JA 2002)***
- **Index** – *psychometrically sound collection of items with an underlying theoretical framework that distinguishes between inter-related constructs relevant to a given health conditions (Sloan JA 2002)*
- **Attributes** - *tangible entity standard definition (height, weight, age, a health condition*) (Mayo 2004-2009)
- **Tests** - *direct indicator of the attribute (ie. math test, blood test, dexterity test)(Mayo 2004-2009)*
- Scale - the response options or units for an item, an index or a test (Sloan JA 2002)
- Measure - is a verb, an action (it is not an entity unless its has demonstrated unidimensionality and measurement invariance) (Mayo 2004-2009)
- Instrument – is a device or a questionnaire that is used to measure a construct or an attribute; when the items of a questionnaire (Mayo 2004-2009)
- Outcome is an aspect of an individual's physical, emotional, mental or social health that is expected to change or to vary owing to a deliberate intervention or to the presence of another personal, health or environmental factor (Mayo 2004-2009)

- Patient reported outcome is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. (FDA Guidance for Industry)

APPENDIX

Appendix 1
Short-Form-36 Health Survey (SF-36)

SF-36 HEALTH SURVEY

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

(circle one)

Excellent 1

Very good 2

Good 3

Fair 4

Poor 5

2. Compared to one week ago, how would you rate your health in general now?

(circle one)

Much better now than one week ago 1

Somewhat better now than one week ago 2

About the same as one week ago 3

Somewhat worse now than one week ago 4

Much worse now than one week ago 5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

(circle one number on each line)

ACTIVITIES	**Yes, Limited A Lot**	**Yes, Limited A Little**	**No, Not Limited At All**
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
b. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing **several** flights of stairs	1	2	3
e. Climbing **one** flight of stairs	1	2	3
f. Bending, kneeling, or stooping	1	2	3
g. Walking **more than a mile**	1	2	3
h. Walking **several blocks**	1	2	3
i. Walking **one block**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

	YES	**NO**
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Were limited in the **kind** of work or other activities	1	2
d. Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

5. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

	YES	NO
a. Cut down the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Didn't do work or other activities as **carefully** as usual	1	2

6. During the past week, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

(circle one)

Not at all 1

Slightly 2

Moderately 3

Quite a bit 4

Extremely 5

7. How much bodily pain have you had during the past weeks?

(circle one)

None 1

Very mild 2

Mild 3

Moderate 4

Severe 5

Very severe 6

8. During the past week, how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all 1

A little bit 2

Moderately 3

Quite a bit 4

Extremely 5

9. These questions are about how you feel and how things have been with you during the past week. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past week -

(circle one number on each line)

		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
a.	Did you feel full of pep?	1	2	3	4	5	6
b.	Have you been a very nervous person?	1	2	3	4	5	6
c.	Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d.	Have you felt calm and peaceful?	1	2	3	4	5	6
e.	Did you have a lot of energy?	1	2	3	4	5	6
f.	Have you felt downhearted and blue?	1	2	3	4	5	6
g.	Did you feel worn out?	1	2	3	4	5	6
h.	Have you been a happy person?	1	2	3	4	5	6
i.	Did you feel tired?	1	2	3	4	5	6

10. During the past week, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

(circle one)

All of the time 1

Most of the time 2

Some of the time.....3

A little of the time 4

None of the time.....5

11. How TRUE or FALSE is each of the following statements for you?

(circle one number on each line)

		Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a.	I seem to get sick a little easier than other people	1	2	3	4	5
b.	I am as healthy as anybody I know	1	2	3	4	5
c.	I expect my health to get worse	1	2	3	4	5
d.	My health is excellent	1	2	3	4	5

Appendix 2
The Gastrointestinal Quality of Life Index (GIQLI)

The Gastrointestinal Quality of Life Index (GIQLI)

1. How often during the past 2 weeks have you had pain in the abdomen?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

2. How often during the past 2 weeks have you had a feeling of fullness in the upper abdomen?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

3. How often during the past 2 weeks have you had bloating (sensation of too much gas in the abdomen)?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

4. How often during the past 2 weeks have you been troubled by the excessive passage of gas through the anus?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

5. How often during the past 2 weeks have you been troubled by strong burping or belching?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

6. How often during the past 2 weeks have you been troubled by gurgling noises from the abdomen?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

7. How often during the past 2 weeks have you been troubled by frequent bowel movements?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

8. How often during the past 2 weeks have you found eating to be a pleasure?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

9. Because of your illness, to what extent have you restricted the kinds of food you eat?

( )	( )	( )	( )	( )
very much	much	somewhat	a little	not at all

10. During the past 2 weeks, how well have you been able to cope with everyday stresses?

( )	( )	( )	( )	( )
extremely poorly	poorly	moderately	well	extremely well

11. How often during the past 2 weeks have you been sad about being ill?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

12. How often during the past 2 weeks have you been nervous or anxious about your illness?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

13. How often during the past 2 weeks have you been happy with life in general?

( )	( )	( )	( )	( )
never	a little of the time	some of the time	most of the time	all of the time

14. How often during the past 2 weeks have you been frustrated about your illness?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

15. How often during the past 2 weeks have you been tired or fatigued?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

16. How often during the past 2 weeks have you felt unwell?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

17. Over the past week, have you woken up in the night?

( )	( )	( )	( )	( )
every night	5-6 nights	3-4 nights	1-2 nights	never

18. Since becoming ill, have you been troubled by changes in your appearance?

( )	( )	( )	( )	( )
a great deal	a moderate amount	somewhat	a little bit	not at all

19. Because of your illness, how much physical strength have you lost?

( )	( )	( )	( )	( )
a great deal	a moderate amount	somewhat	a little bit	not at all

20. Because of your illness, to what extent have you lost your endurance?

( )	( )	( )	( )	( )
a great deal	a moderate amount	somewhat	a little bit	not at all

21. Because of your illness, to what extent do you feel unfit?

( )	( )	( )	( )	( )
a great deal	a moderate amount	somewhat	a little bit	not at all

22. During the past 2 weeks, how often have you been able to complete your normal daily activities (school, work, household)?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

23. During the past 2 weeks, how often have you been able to take part in your usual patterns of leisure or recreational activities?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

24. During the past 2 weeks, how much have you been troubled by the medical treatment of your illness?

( )	( )	( )	( )	( )
very much	much	somewhat	a little	not at all

25. To what extent have your personal relations with people close to you (family or friends) worsened because of your iliness?

( )	( )	( )	( )	( )
very much	much	somewhat	a little	not at all

26. To what extent has your sexual life been impaired (harmed) because of your illness?

( )	( )	( )	( )	( )
very much	much	somewhat	a little	not at all

27. How often during the past 2 weeks, have you been troubled by fluid or food coming up into your mouth (regurgitation)?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

28. How often during the past 2 weeks, have you felt uncomfortable because of your slow speed of eating?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

29. How often during the past 2 weeks, have you had trouble swallowing your food?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

30. How often during the past 2 weeks, have you been troubled by urgent bowel movements?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

31. How often during the past 2 weeks, have you been troubled by diarrhea?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

32. How often during the past 2 weeks, have you been troubled by constipation?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

33. How often during the past 2 weeks, have you been troubled by nausea?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

34. How often during the past 2 weeks, have you been troubled by blood in the stool?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

35. How often during the past 2 weeks, have you been troubled by heartburn?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

36. How often during the past 2 weeks, have you been troubled by uncontrolled stools?

( )	( )	( )	( )	( )
All of the time	most of the time	some of the time	a little of the time	never

Appendix 3
Communities Health Activities Model Program for Seniors (CHAMPS) questionnaire

Appendix. *Continued.*

This questionnaire is about activities that you may have done in the past 4 weeks. The questions on the following pages are similar to the example shown below.

INSTRUCTIONS

If you DID the activity in the past 4 weeks:

Step #1 **Check the YES box.**

Step #2 **Think about how many TIMES a week you usually did it, and write your response in the space provided.**

Step #3 **Circle how many TOTAL HOURS in a typical week you did the activity.**

Here is an example of how Mrs. Jones would answer question #1: Mrs. Jones usually visits her friends Maria and Olga twice a week. She usually spends one hour on Monday with Maria and two hours on Wednesday with Olga. Therefore, the total hours a week that she visits with friends is 3 hours a week.

In a typical week during the past 4 weeks, did you…							
1. Visit with friends or family (other than those you live with)? ☒ **YES** How many TIMES a week? 2 ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	(3-4½ hours)	5-6½ hours	7-8½ hours	9 or more hours

If you DID NOT do the activity:

- **Check the NO box and move to the next question**

In a typical week during the past 4 weeks, did you …							
1. Visit with friends or family (other than those you live with)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
2. Go to the senior center? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
3. Do volunteer work? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
4. Attend church or take part in church activities? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
5. Attend other club or group meetings? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
6. Use a computer? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
7. Dance (such as square, folk, line, ballroom) (do not count aerobic dance here)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
8. Do woodworking, needlework, drawing, or other arts or crafts? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
9. Play golf, carrying or pulling your equipment (count walking time only)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
10. Play golf, riding a cart (count walking time only)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
11. Attend a concert, movie, lecture, or sport event? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
12. Play cards, bingo, or board games with other people? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
13. Shoot pool or billiards? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
14. Play singles tennis (do not count doubles)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
15. Play doubles tennis (do not count singles)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
16. Skate (ice, roller, in-line)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
17. Play a musical instrument? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
18. Read? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
19. Do heavy work around the house (such as washing windows, cleaning gutters)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
20. Do light work around the house (such as sweeping or vacuuming)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
21. Do heavy gardening (such as spading, raking)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
22. Do light gardening (such as watering plants)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
23. Work on your car, truck, lawn mower, or other machinery? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
****Please note: For the following questions about running and walking, include use of a treadmill.**							
24. Jog or run? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
25. Walk uphill or hike uphill (count only uphill part)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
26. Walk fast or briskly for exercise (do not count walking leisurely or uphill)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
In a typical week during the past 4 weeks, did you …							
27. Walk to do errands (such as to/from a store or to take children to school (count walk time only)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
28. Walk leisurely for exercise or pleasure? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
29. Ride a bicycle or stationary cycle? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
30. Do other aerobic machines such as rowing, or step machines (do not count treadmill or stationary cycle)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
31. Do water exercises (do not count other swimming)? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
32. Swim moderately or fast? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
33. Swim gently? ☐ **YES** How many TIMES a week?____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
34. Do stretching or flexibility exercises (do not count yoga or Tai-chi)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
35. Do yoga or Tai-chi? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
36. Do aerobics or aerobic dancing? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
37. Do moderate to heavy strength training (such as hand-held weights of more than 5 lbs., weight machines, or push-ups)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
38. Do light strength training (such as hand-held weights of 5 lbs. or less or elastic bands)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
39. Do general conditioning exercises, such as light calisthenics or chair exercises (do not count strength training)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
40. Play basketball, soccer, or racquetball (do not count time on sidelines)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
41. Do other types of physical activity not previously mentioned (please specify)? _____ ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

Thank You

Appendix 4
Recoding CHAMPS

The items from the CHAMPS questionnaire were recoded according to the ICF. Because there exist no standard ICF codes for CHAMPS, two independent researchers identified the most appropriate ICF category corresponding to the given CHAMPS item.

During the past week, did you...			
	CHAMPS item	ICF recoding	ICF definition
Q1	Visit with friends or family (other than those you live with)?	d9205 Socializing	Engaging in informal or casual gatherings with others, such as visiting friends or relatives or meeting informally in public places.
Q2	Go to a senior or social centre?	d9205 Socializing	Engaging in informal or casual gatherings with others, such as visiting friends or relatives or meeting informally in public places.
Q3	Do volunteer work?	d855 Non-remunerative employment	Engaging in all aspects of work in which pay is not provided, full-time or part-time, including organized work activities, doing the required tasks of the job, attending work on time as required, supervising other workers or being supervised, and performing required tasks alone or in groups, such as volunteer work, charity work, working for a community or religious group without remuneration, working around the home without remuneration.
Q4	Attend church or take part in church activities?	d9300 Organized religion	Engaging in organized religious ceremonies, activities and events.
Q5	Attend other club or group meetings?	d9205 Socializing	Engaging in informal or casual gatherings with others, such as visiting friends or relatives or meeting informally in public places.
Q6	Use a computer?	d3601 Using writing machines	Using machines for writing, such as typewriters, computers and Braille writers, as a means of communication.
Q7	Dance such as square, folk, line, ballroom (do not count aerobic dance here)?	d9208 Recreation and leisure, other specified	Engaging in any form of play, recreational or leisure activity, such as informal or organized play and sports, programmes of physical fitness, relaxation, amusement or diversion, going to art galleries, museums, cinemas or theatres; engaging in crafts or hobbies, reading for enjoyment, playing musical instruments; sightseeing, tourism and travelling for pleasure.
Q8	Do woodworking, needlework, drawing or other arts and crafts?	d9203 Crafts	Engaging in handicrafts, such as pottery or knitting.
Q9	Play golf, carrying and pulling your equipment?	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.

Q10	Play golf, riding a cart?	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.
Q11	Attend a concert, movie, lecture or sport event?	d9202 Arts and culture	Engaging in, or appreciating, fine arts or cultural events, such as going to the theatre, cinema, museum or art gallery, or acting in a play, reading for enjoyment or playing a musical instrument.
Q12	Play cards, bingo or board games with other people?	d9200 Play	Engaging in games with rules or unstructured or unorganized games and spontaneous recreation, such as playing chess or cards or children's play.
Q13	Shoot pool or billiards?	d9202 Arts and culture	Engaging in, or appreciating, fine arts or cultural events, such as going to the theatre, cinema, museum or art gallery, or acting in a play, reading for enjoyment or playing a musical instrument.
Q14	Play single tennis?	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.
Q15	Play double tennis?	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.
Q16	Skate (ice, roller, in-line)?	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.
Q17	Play a musical instrument?	d9202 Arts and culture	Engaging in, or appreciating, fine arts or cultural events, such as going to the theatre, cinema, museum or art gallery, or acting in a play, reading for enjoyment or playing a musical instrument.
Q18	Read?	d9202 Arts and culture	Engaging in, or appreciating, fine arts or cultural events, such as going to the theatre, cinema, museum or art gallery, or acting in a play, reading for enjoyment or playing a musical instrument.
Q19	Do heavy work around the house (such as washing windows, cleaning gutters)?	d640 Doing housework	Managing a household by cleaning the house, washing clothes, using household appliances, storing food and disposing of garbage, such as by sweeping, mopping, washing counters, walls and other surfaces; collecting and disposing of household garbage; tidying rooms, closets and drawers; collecting, washing, drying, folding and ironing clothes; cleaning footwear; using brooms, brushes and vacuum cleaners; using washing machines, driers and irons. Inclusions: washing and drying clothes and garments; cleaning cooking area and utensils; cleaning living area; using household appliances, storing daily necessities and disposing of garbage
Q20	Do light work around the house (such as sweeping or vacuuming)?	d640 Doing housework	Managing a household by cleaning the house, washing clothes, using household appliances, storing food and disposing of garbage, such as by sweeping, mopping, washing counters, walls and other surfaces; collecting and disposing of household garbage; tidying rooms,

			closets and drawers; collecting, washing, drying, folding and ironing clothes; cleaning footwear; using brooms, brushes and vacuum cleaners; using washing machines, driers and irons. Inclusions: washing and drying clothes and garments; cleaning cooking area and utensils; cleaning living area; using household appliances, storing daily necessities and disposing of garbage
Q21	Do heavy gardening (such as spading, raking)?	d650 Caring for household objects	Maintaining and repairing household and other personal objects, including house and contents, clothes, vehicles and assistive devices, and caring for plants and animals, such as painting or wallpapering rooms, fixing furniture, repairing plumbing, ensuring the proper working order of vehicles, watering plants, grooming and feeding pets and domestic animals. Inclusions: making and repairing clothes; maintaining dwelling, furnishings and domestic appliances; maintaining vehicles; maintaining assistive devices; taking care of plants (indoor and outdoor) and animals
Q22	Do light gardening (such as watering plants)?	d650 Caring for household objects	Maintaining and repairing household and other personal objects, including house and contents, clothes, vehicles and assistive devices, and caring for plants and animals, such as painting or wallpapering rooms, fixing furniture, repairing plumbing, ensuring the proper working order of vehicles, watering plants, grooming and feeding pets and domestic animals. Inclusions: making and repairing clothes; maintaining dwelling, furnishings and domestic appliances; maintaining vehicles; maintaining assistive devices; taking care of plants (indoor and outdoor) and animals
Q23	Work on your car, truck; lawn mower, or other machinery?	d650 Caring for household objects	Maintaining and repairing household and other personal objects, including house and contents, clothes, vehicles and assistive devices, and caring for plants and animals, such as painting or wallpapering rooms, fixing furniture, repairing plumbing, ensuring the proper working order of vehicles, watering plants, grooming and feeding pets and domestic animals. Inclusions: making and repairing clothes; maintaining dwelling, furnishings and domestic appliances; maintaining vehicles; maintaining assistive devices; taking care of plants (indoor and outdoor) and animals
Q24	Jog or run?	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming
Q25	Walk uphill or hike uphill?	d450 Walking	Moving along a surface on foot, step by step, so that one foot is always on the ground, such as when strolling, sauntering, walking forwards, backwards, or sideways. Inclusions: walking short or long distances; walking on different surfaces; walking around obstacles
Q26	Walk fast or briskly for exercise?	d450 Walking	Moving along a surface on foot, step by step, so that one foot is always on the ground, such

			as when strolling, sauntering, walking forwards, backwards, or sideways. Inclusions: walking short or long distances; walking on different surfaces; walking around obstacles
Q27	Walk to do errands (such as to/from store or take children to school)?	d450 Walking	Moving along a surface on foot, step by step, so that one foot is always on the ground, such as when strolling, sauntering, walking forwards, backwards, or sideways. Inclusions: walking short or long distances; walking on different surfaces; walking around obstacles
Q28	Walk leisurely for exercise or pleasure?	d450 Walking	Moving along a surface on foot, step by step, so that one foot is always on the ground, such as when strolling, sauntering, walking forwards, backwards, or sideways. Inclusions: walking short or long distances; walking on different surfaces; walking around obstacles
Q29	Ride a bicycle or stationary cycle?	d4750 Driving human-powered transportation	Driving a human-powered vehicle, such as a bicycle, tricycle, or rowboat.
Q30	Do other aerobic machine such as rowing or stepping machines?	d4750 Driving human-powered transportation	Driving a human-powered vehicle, such as a bicycle, tricycle, or rowboat.
Q31	Do water exercises (do not count other swimming)?	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming
Q32	Swim moderately or fast?	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming
Q33	Swim gently?	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming
Q34	Do stretching or flexibility exercises?	d410 Changing basic body position	Getting into and out of a body position and moving from one location to another, such as getting up out of a chair to lie down on a bed, and getting into and out of positions of kneeling or squatting. Inclusion: changing body position from lying down, from squatting or kneeling, from sitting or standing, bending and shifting the body's centre of gravity
Q35	Do yoga or tai-chi?	d410 Changing basic body position	Getting into and out of a body position and moving from one location to another, such as getting up out of a chair to lie down on a bed, and getting into and out of positions of

			kneeling or squatting. Inclusion: changing body position from lying down, from squatting or kneeling, from sitting or standing, bending and shifting the body's centre of gravity
Q36	Do aerobics or aerobic dancing?	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming
Q37	Do moderate to heavy strength training (such as hand-held weights of more than 5lbs., weight machines, or push-ups)?	d430 Lifting and carrying objects	Raising up an object or taking something from one place to another, such as when lifting a cup or carrying a child from one room to another. Inclusions: lifting, carrying in the hands or arms, or on shoulders, hip, back or head; putting down
Q38	Do light strength training (such as hand-held weights of 5lbs or less or elastic bands)?	d430 Lifting and carrying objects	Raising up an object or taking something from one place to another, such as when lifting a cup or carrying a child from one room to another. Inclusions: lifting, carrying in the hands or arms, or on shoulders, hip, back or head; putting down
Q39	Do general conditioning exercises, such as light calisthetics or chain exercises (do not count strength training)?	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming
Q40	Play basketball, soccer or racquetball ?	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.
*Q41	Do other types of physical activity not previously mentioned (please specify)? Bowling Ski Play hockey	d9201 Sports	Engaging in competitive and informal or formally organized games or athletic events, performed alone or in a group, such as bowling, gymnastics or soccer.
*Q41	Do other types of physical activity not previously mentioned (please specify)? Play with kids Play darts	d9200 Play	Engaging in games with rules or unstructured or unorganized games and spontaneous recreation, such as playing chess or cards or children's play.
*Q41	Do other types of physical activity not previously mentioned (please specify)? Use a treadmill	d455 Moving around	Moving the whole body from one place to another by means other than walking, such as climbing over a rock or running down a street, skipping, scampering, jumping, somersaulting or running around obstacles. Inclusions: crawling, climbing, running, jogging, jumping, and swimming

*Q41	Do other types of physical activity not previously mentioned (please specify)? Sexual activity	d7702 Sexual relationships	Creating and maintaining a relationship of a sexual nature, with a spouse or other partner.
*Q41	Do other types of physical activity not previously mentioned (please specify)? Go to work	d850 Remunerative employment	Engaging in all aspects of work, as an occupation, trade, profession or other form of employment, for payment, as an employee, full or part time, or self-employed, such as seeking employment and getting a job, doing the required tasks of the job, attending work on time as required, supervising other workers or being supervised, and performing required tasks alone or in groups.

*responses for item 41 of CHAMPS were included and coded to the most appropriate ICF category

18 new ICF coded activities were identified and were recoded as items for entry into the RUMM2020 program to perform the Rasch analysis. Similar activities were grouped and recoded into 1 ICF item. Responses were categorized on the ordinal scale quantified by the physical intensity of the item using metabolic equivalents (MET) values based on values reported by Ainsworth et al. (Ainsworth 1993). An activity with a higher MET value was given a higher response score compared an activity with a lower MET value within the same ICF item. Each activity's MET value is shown in ( ).

During the past week, did you do any of these activities...										
New ICF recoded item	Scoring									
	0	1	2	3	4	5	6	7	8	9
d9205 Socializing	none	Visit with friends or family (other than those you live with)? (1.5)	Go to a senior or social centre? (2.0)	Attend other club or group meetings? (2.0)						
d855 Non-remunerative employment	none	Do volunteer work?								
d9300 Organized	none	Attend church or take part in church								

religion		activities? (1.0)								
d3601 Using writing machines	none	Use a computer?								
d9208 Recreation and leisure, other specified	none	Dance such as square, folk, line, ballroom (do not count aerobic dance here)? (4.5)								
d9203 Crafts	none	Do woodworking, needlework, drawing or other arts and crafts? (2.0)								
d9201 Sports	none	Do other types of physical activity not previously mentioned (please specify)? -Bowling (3.0)	Play golf, riding a cart? (3.5)	Play golf, carrying and pulling your equipment? (4.5)	Play double tennis? (5.0)	Do other types of physical activity not previously mentioned (please specify)? -Ski (7.0)	Skate (ice, roller, in-line)? (7.0)	Play basketball, soccer or racquetball? (7.0)	Play single tennis? (8.0)	Do other types of physical activity not previously mentioned (please specify)? -Hockey (8.0)
d9202 Arts and culture	none	Read? (1.3)	Attend a concert, movie, lecture or sport event? (1.5)	Shoot pool or billiards? (2.5)	Play a musical instrument? (3.0)					
d9200 Play	none	Play cards, bingo or board games with other people? (1.5)	Do other types of physical activity not previously mentioned (please specify)? -Play with kids (2.5)	Do other types of physical activity not previously mentioned (please specify)? -Play darts (2.5)						
d640 Doing	none	Do light work	Do heavy work							

housework		around the house (such as sweeping or vacuuming)? (2.5)	around the house (such as washing windows, cleaning gutters)? (4.0)							
d650 Caring for household objects	none	Do light gardening (such as watering plants)? (2.5)	Work on your car, truck; lawn mower, or other machinery? (3.0)	Do heavy gardening (such as spading, raking)? (4.5)						
d455 Moving around	none	Do water exercises (do not count other swimming)? (4.0)	Do general conditioning exercises, such as light calisthetics or chain exercises (do not count strength training)? (3.5)	Do aerobics or aerobic dancing? (6.5)	Swim gently? (6.0)	Do other types of physical activity not previously mentioned (please specify)? -Use a treadmill (9.0)	Jog or run? (9.0)	Swim moderately or fast? (10.0)		
d450 Walking	none	Walk to do errands (such as to/from store or take children to school)? (3.5)	Walk leisurely for exercise or pleasure? (3.5)	Walk fast or briskly for exercise? (3.8)	Walk uphill or hike uphill? (6.0)					
d4750 Driving human-powered transportation	none	Ride a bicycle or stationary cycle? (8.0)	Do other aerobic machine such as rowing or stepping machines? (8.0)							
d410 Changing basic body position	none	Do stretching or flexibility exercises? (2.5)	Do yoga or tai-chi? (2.5)							
d430 Lifting and carrying objects	none	Do light strength training (such as hand-held weights of 5lbs or less or elastic bands)? (3.5)	Do moderate to heavy strength training (such as hand-held weights of more than 5lbs., weight machines, or push-ups)? (8.0)							

d7702 Sexual relationships	none	Do other types of physical activity not previously mentioned (please specify)? -Sexual activity (1.3)								
d850 Remunerative employment	none	Do other types of physical activity not previously mentioned (please specify)? -Go to work								

Appendix 5
Rescoring of Items

List of all included items with **original** scoring:

Item #	Source	Item	Scoring 0	1	2	3	4	5	6	7	8	9	10
I0001	SF36 (physical function subscale)	Does **your health now limit you** in these activities? If so, how much? – Q3: **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0002	SF36 (physical function subscale)	Q4: **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0003	SF36 (physical function subscale)	Q5: Lifting or carrying groceries		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0004	SF36 (physical function subscale)	Q6: Climbing **several** flights of stairs		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0005	SF36 (physical function subscale)	Q7: Climbing **one** flight of stairs		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0006	SF36	Q8: Bending,		Yes,	Yes,	No, Not							

	(physical function subscale)	kneeling, or stooping		Limited a Lot	Limited a Little	limited at All							
I0007	SF36 (physical function subscale)	Q9: Walking **more than a mile**		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0008	SF36 (physical function subscale)	Q10: Walking **several blocks**		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0009	SF36 (physical function subscale)	Q11: Walking **one block**		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0010	SF36 (physical function subscale)	Q12: Bathing or dressing yourself		Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All							
I0011	SF36 (general health perception)	**Q1:** In general, would you say your health is:		Excellent	Very good	Good	Fair	Poor					
I0012	SF36 (health change)	**Q2: Compared to one week ago**, how would your rate your health in general **now**?		Much better now than one week ago	Somewhat better now than one week ago	About the same	Somewhat worse now than one week ago	Much worse now than one week ago					
I0013	SF36 (Role physical)	During the **past week**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**?		Yes	No								

		- Q13: Cut down the amount of time you spent on work or other activities											
I0014	SF36 (Role emotional)	During the **past week,** have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious) - Q14: Cut down the amount of time you spent on work or other activities		Yes	No								
I0015	SF36 (Social function)	During the **past week,** to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?		Not at all	Slightly	Moderately	Quite a bit	Extremely					
I0016	SF36 (Bodily pain)	How much **bodily** pain have you had during the **past week**?		None	Very mild	Mild	Moderate	Severe	Very severe				
I0017	SF36 (Bodily pain)	Q22: During the **past week,** how much did **pain** interfere with your normal work		Not at all	Slightly	Moderately	Quite a bit	Extremely					

		(including both work outside the home and housework)?											
I0018	SF36 (Vitality)	How much of the time during the **past week** -Q23: Did you feel full of pep?		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time				
I0019	SF36 (Mental Health)	-Q24: Have you been a very nervous person?		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time				
I0020	SF36 (Vitality)	-Q27: Did you have a lot of energy?		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time				
I0021	SF36 (Vitality)	-Q29: Did you feel worn out?		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time				
I0022	SF36 (Vitality)	-Q31: Did you feel tired?		All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time				
I0023	GIQLI	Q1: How often during the past week have you had pain in the abdomen?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0024	GIQLI	Q2: How often during the past week have you had a feeling of fullness in the upper abdomen?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0025	GIQLI	Q3: How often during the past week have you had bloating (sensation of too much gas in the abdomen)?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0026	GIQLI	Q4: How often	All of the	Most of the	Some of the	A little of	never						

		during the past week have you been troubled by the excessive passage of gas through the anus?	time	time	time	the time							
I0027	GIQLI	Q5: How often during the past week have you been troubled by strong burping or belching?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0028	GIQLI	Q6: How often during the past week have you been troubled by gurgling noises from the abdomen?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0029	GIQLI	Q7: How often during the past week have you been troubled by frequent bowel movements?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0030	GIQLI	Q8: How often during the past week have you found eating to be a pleasure?	Never	A little of the time	Some of the time	Most of the time	All of the time						
I0031	GIQLI	Q9: Because of your illness, to what extent have you restricted the kinds of food you eat?	Very much	Much	Somewhat	A little	Not at all						
I0032	GIQLI	Q10: During the past week, how well have you been able to cope with everyday stresses?	Extremely poor	Poorly	Moderately	Well	Extremely well						

I0033	GIQLI	Q11: How often during the past week have you been sad about being ill?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0034	GIQLI	Q12: How often during the past week have you been nervous or anxious about your illness?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0035	GIQLI	Q13: How often during the past 2 weeks have you been happy with life in general?	Never	A little of the time	Some of the time	Most of the time	All of the time						
I0036	GIQLI	Q14: How often during the past week have you been frustrated about your illness?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0037	GIQLI	Q15: How often during the past week have you been tired or fatigued?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0038	GIQLI	Q16: How often during the past week have you felt unwell?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0039	GIQLI	Q17: Over the past week, have you woken up in the night	Every night	5-6 nights	3-4 nights	1-2 nights	never						
I0040	GIQLI	Q18: Since becoming ill, have you been troubled by changes in your appearance?	A great deal	A moderate amount	Somewhat	A little bit	Not at all						
I0041	GIQLI	Q19: Because of	A great	A moderate	Somewhat	A little bit	Not at all						

		your illness, how much physical strength have you lost?	deal	amount									
I0042	GIQLI	Q20: To what extent have you lost your endurance?	A great deal	A moderate amount	Somewhat	A little bit	Not at all						
I0043	GIQLI	Q21: Because of your illness, to what extent do you feel unfit?	A great deal	A moderate amount	Somewhat	A little bit	Not at all						
I0044	GIQLI	Q22: During the past week how often have you been able to complete your normal daily activities (school, work, household)?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0045	GIQLI	Q23: During the past week how often have you been able to take part in your usual patterns of leisure or recreational activities?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0046	GIQLI	Q24: During the past week how much have you been troubled by the medical treatment of your illness?	Very much	Much	Somewhat	A little	Not at all						
I0047	GIQLI	Q25: To what extent have your personal relations with people close to you (family or friends)	Very much	Much	Somewhat	A little	Not at all						

		worsened because of your illness ?											
I0048	GIQLI	Q26: To what extent has your sexual life been impaired (harmed) because of your illness?	Very much	Much	Somewhat	A little	Not at all						
I0049	GIQLI	Q27: How often during the past week, have you been troubled by fluid or food coming up into your mouth (regurgitation)?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0050	GIQLI	Q28: How often during the past week, have you felt uncomfortable because of your slow speed of eating?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0051	GIQLI	Q29: How often during the past week, have you had trouble swallowing your food?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0052	GIQLI	Q30: How often during the past week, have you been troubled by urgent bowel movements?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0053	GIQLI	Q31: How often during the past week, have you been troubled by diarrhea?	All of the time	Most of the time	Some of the time	A little of the time	never						

I0054	GIQLI	Q32: How often during the past week, have you been troubled by constipation?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0055	GIQLI	Q33: How often during the past week, have you been troubled by nausea?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0056	GIQLI	Q34: How often during the past week, have you been troubled by blood in the stool?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0057	GIQLI	Q35: How often during the past week, have you been troubled by heartburn?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0058	GIQLI	Q36: How often during the past week, have you been troubled by uncontrolled stool?	All of the time	Most of the time	Some of the time	A little of the time	never						
I0059	VAS	This is a scale to help express the way you feel. For each one please indicate, using the scale from 0-10, how you feel? 6. Anxiety	0 (None)	1	2	3	4	5	6	7	8	9	10 (Worst possible)
I0060	VAS	4. Fatigue	0 (None)	1	2	3	4	5	6	7	8	9	10 (Worst possible)
I0061	VAS	5.Nausea	0 (None)	1	2	3	4	5	6	7	8	9	10 (Worst

													possible)
I0062	VAS	1. Pain at rest	0 (None)	1	2	3	4	5	6	7	8	9	10 (Worst possible)
I0063	VAS	2. Pain with coughing	0 (None)	1	2	3	4	5	6	7	8	9	10 (Worst possible)
I0064	VAS	3. Pain with Movement	0 (None)	1	2	3	4	5	6	7	8	9	10 (Worst possible)
I0065	CHAMPS	Q (D9205) During the past week, did you do any of these activities?	No	Visit with family and friends (other than those you live) (yes)	Go to a social/senior center (yes)	Attend other club or group meetings (yes)							
I0066	CHAMPS	Q (D855) During the past week, did you do any of these activities?	No	Do volunteer work (yes)									
I0067	CHAMPS	Q (D850) During the past week, did you do any of these activities?	No	Go to work (yes)									
I0068	CHAMPS	Q (D9300) During the past week, did you do any of these activities?	No	Attend church or take part in church activities (yes)									
I0069	CHAMPS	Q (D3601) During the past week, did you do any of these activities?	No	Use a computer									
I0070	CHAMPS	Q (D9208) During	No	Dance such as									

		the past week, did you do any of these activities?		square, folk, line, ballroom, do not count aerobic dance here) (yes)									
I0071	CHAMPS	Q (D9203) During the past week, did you do any of these activities?	No	Do woodworking, needlework, drawing or other arts and crafts (yes)									
I0072	CHAMPS	Q (D9201) During the past week, did you do any of these activities?	No	Bowling (yes)	Play golf, riding a cart (yes)	Play golf, carrying and pulling your equipment (yes)	Play double tennis (yes)	Ski (yes)	Skate (ice, roller, in-line) (yes)	Play basketball, soccer or racquetball (yes)	Play single tennis (yes)	Play hockey (yes)	
I0073	CHAMPS	Q (D9202) During the past week, did you do any of these activities?	No	Read (yes)	Attend a concert, movie, lecture or sport event (yes)	Shoot pool or billiards (yes)	Play a musical instrument (yes)						
I0074	CHAMPS	Q (D9200) During the past week, did you do any of these activities?	No	Play cards, bingo or board games with other people (yes)	Play with kids (yes)	Play darts (yes)							
I0075	CHAMPS	Q (D640) During the past week, did you do any of these activities?	No	Do light work around the house (such as sweeping or vacuuming) (yes)	Do heavy work around the house (such as washing windows, cleaning gutters)								

					(yes)								
I0076	CHAMPS	Q (D650) During the past week, did you do any of these activities?	No	Do light gardening (such as watering plants) (yes)	Work on your car, truck; lawn mower, or other machinery (yes)	Do heavy gardening (such as spading, raking)							
I0077	CHAMPS	Q (b4551/d455) During the past week, did you do any of these activities?	No	Do water exercises (do not count other swimming) (yes)	Do general conditioning exercises, such as light calisthetics or chain exercises (do not count strength training) (yes)	Do aerobics or aerobic dancing (yes)	Swim gently (yes)	Use a treadmill (yes)	Jog or run (yes)	Swim moderately or fast (yes)			
I0078	CHAMPS	Q (D450) During the past week, did you do any of these activities?	No	Walk to do errands (such as to/from store or take children to school) (yes)	Walk leisurely for exercise or pleasure (yes)	Walk fast or biskly for exercise (yes)	Walk uphill or hike uphill (yes)						
I0079	CHAMPS	Q (D4750) During the past week, did you do any of these activities?	No	Ride a bicycle or stationary cycle (yes)	Do other aerobic machine such as rowing or stepping machines (yes)								
I0080	CHAMPS	Q (D41) During the past week, did you do any of these	No	Do stretching or flexibility exercises	Do yoga or tai-chi (yes)								

		activities?		(yes)									
I0081	CHAMPS	Q (D0) During the past week, did you do any of these activities?	No	Do light strength training (such as hand-held weights of 5lbs or less or elastic bands)	Do moderate to heavy strength training (such as hand-held weights of more than 5lbs., weight machines, or push-ups (yes)								
I0082	CHAMPS	Q (D7702) During the past week, did you do any of these activities?	No	Sexual activity (yes)									
I0083	6MWT	Distance walked in 6 minutes	0-200m	201-300m	301-350m	351-400m	401-450m	451-500m	501-600m	601-800m	>801m		

List of all included items with ***rescoring***:

Item #	Source	Item	Scoring 0	1	2	3	4	5
I0001	SF36 (physical function	Does **your health now limit you** in these activities? If so, how much? –	Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All			

	subscale)	Q3: **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports						
I0002	SF36 (physical function subscale)	Q4: **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All			
I0003	SF36 (physical function subscale)	Q5: Lifting or carrying groceries	Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All			
I0004	SF36 (physical function subscale)	Q6: Climbing **several** flights of stairs	Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All			
I0005	SF36 (physical function subscale)	Q7: Climbing **one** flight of stairs	Yes, Limited a Lot; Yes, Limited a Little	No, Not limited at All				
I0006	SF36 (physical function subscale)	Q8: Bending, kneeling, or stooping	Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All			
I0007	SF36 (physical function subscale)	Q9: Walking **more than a mile**	Yes, Limited a Lot; Yes, Limited a Little	No, Not limited at All				
I0008	SF36 (physical function subscale)	Q10: Walking **several blocks**	Yes, Limited a Lot; Yes, Limited a Little	No, Not limited at All				
I0009	SF36 (physical function subscale)	Q11: Walking **one block**	Yes, Limited a Lot	Yes, Limited a Little	No, Not limited at All			
I0010	SF36 (physical	Q12: Bathing or dressing yourself	Yes, Limited a Lot; Yes,	No, Not limited at All				

	function subscale)		Limited a Little					
I0011	SF36 (general health perception)	**Q1:** In general, would you say your health is:	Poor; Fair	Good	Very good	Excellent		
I0012	SF36 (health change)	**Q2: Compared to one week ago**, how would your rate your health in general **now**?	Much worse now than one week ago	Somewhat worse now than one week ago	About the same	Somewhat better now than one week ago	Much better now than one week ago	
I0013	SF36 (Role physical)	During the **past week**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**? - Q13: Cut down the amount of time you spent on work or other activities	Yes	No				
I0014	SF36 (Role emotional)	During the **past week,** have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious) - Q14: Cut down the amount of time you spent on work or other activities	Yes	No				
I0015	SF36 (Social function)	During the **past week,** to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?	Extremely	Quite a bit; Moderately	Slightly; Not at all			
I0016	SF36 (Bodily pain)	How much **bodily** pain have you had during the **past week**?	Very severe	Severe	Moderate	Mild	Very mild	None
I0017	SF36	Q22: During the **past week,** how	Extremely;	Moderately; Slightly	Not at all			

	(Bodily pain)	much did **pain** interfere with your normal work (including both work outside the home and housework)?	Quite a bit					
I0018	SF36 (Vitality)	How much of the time during the **past week** . . . -Q23: Did you feel full of pep?	None of the Time	A Little of the Time	Some of the Time; A Good Bit of the Time	Most of the Time	All of the Time	
I0019	SF36 (Mental Health)	How much of the time during the **past week** . . . -Q24: Have you been a very nervous person?	All of the Time	Most of the Time	A Good Bit of the Time; Some of the Time	A Little of the Time	None of the Time	
I0020	SF36 (Vitality)	How much of the time during the **past week** . . . -Q27: Did you have a lot of energy?	None of the Time	A Little of the Time	Some of the Time	A Good Bit of the Time	Most of the Time	All of the Time
I0021	SF36 (Vitality)	How much of the time during the **past week** . . . -Q29: Did you feel worn out?	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
I0022	SF36 (Vitality)	How much of the time during the **past week** . . . -Q31: Did you feel tired?	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
I0023	GIQLI	Q1: How often during the past week have you had pain in the abdomen?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0024	GIQLI	Q2: How often during the past week have you had a feeling of fullness in the upper abdomen?	All of the time	Most of the time	Some of the time; A little of the time	never		
I0025	GIQLI	Q3: How often during the past week have you had bloating (sensation of too much gas in the abdomen)?	All of the time	Most of the time; Some of the time; A little of the time	never			
I0026	GIQLI	Q4: How often during the past week have you been troubled by the excessive passage of gas through the anus?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0027	GIQLI	Q5: How often during the past week have you been troubled by strong burping or belching?	All of the time	Most of the time	Some of the time	A little of the time	never	

I0028	GIQLI	Q6: How often during the past week have you been troubled by gurgling noises from the abdomen?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0029	GIQLI	Q7: How often during the past week have you been troubled by frequent bowel movements?	All of the time	Most of the time	Some of the time; A little of the time	Never		
I0030	GIQLI	Q8: How often during the past week have you found eating to be a pleasure?	Never	A little of the time	Some of the time	Most of the time	All of the time	
I0031	GIQLI	Q9: Because of your illness, to what extent have you restricted the kinds of food you eat?	Very much	Much	Somewhat; A little	Not at all		
I0032	GIQLI	Q10: During the past week, how well have you been able to cope with everyday stresses?	Extremely poor	Poorly	Moderately	Well	Extremely well	
I0033	GIQLI	Q11: How often during the past week have you been sad about being ill?	All of the time	Most of the time	Some of the time; A little of the time	Never		
I0034	GIQLI	Q12: How often during the past week have you been nervous or anxious about your illness?	All of the time	Most of the time; Some of the time	A little of the time	never		
I0035	GIQLI	Q13: How often during the past 2 weeks have you been happy with life in general?	Never	A little of the time; Some of the time	Most of the time	All of the time		
I0036	GIQLI	Q14: How often during the past week have you been frustrated about your illness?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0037	GIQLI	Q15: How often during the past week have you been tired or fatigued?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0038	GIQLI	Q16: How often during the past week have you felt unwell?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0039	GIQLI	Q17: Over the past week, have you woken up in the night	Every night	5-6 nights; 3-4 nights; 1-2 nights	never			
I0040	GIQLI	Q18: Since becoming ill, have you been troubled by changes in your	A great deal	A moderate amount	Somewhat; A little bit	Not at all		

		appearance?						
I0041	GIQLI	Q19: Because of your illness, how much physical strength have you lost?	A great deal	A moderate amount; Somewhat	A little bit	Not at all		
I0042	GIQLI	Q20: To what extent have you lost your endurance?	A great deal	A moderate amount; Somewhat; A little bit	Not at all			
I0043	GIQLI	Q21: Because of your illness, to what extent do you feel unfit?	A great deal	A moderate amount	Somewhat	A little bit	Not at all	
I0044	GIQLI	Q22: During the past week how often have you been able to complete your normal daily activities (school, work, household)?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0045	GIQLI	Q23: During the past week how often have you been able to take part in your usual patterns of leisure or recreational activities?	All of the time	Most of the time; Some of the time; A little of the time	Never			
I0046	GIQLI	Q24: During the past week how much have you been troubled by the medical treatment of your illness?	Very much; Much	Somewhat; A little	Not at all			
I0047	GIQLI	Q25: To what extent have your personal relations with people close to you (family or friends) worsened because of your illness ?	Very much; Much	Somewhat; A little	Not at all			
I0048	GIQLI	Q26: To what extent has your sexual life been impaired (harmed) because of your illness?	Very much; Much	Somewhat; A little	Not at all			
I0049	GIQLI	Q27: How often during the past week, have you been troubled by fluid or food coming up into your mouth (regurgitation)?	All of the time	Most of the time; Some of the time; A little of the time	Never			
I0050	GIQLI	Q28: How often during the past week, have you felt uncomfortable because of your slow speed of eating?	All of the time	Most of the time	Some of the time; A little of the time	never		
I0051	GIQLI	Q29: How often during the past	All of the	Some of the time; A little of the	Never			

		week, have you had trouble swallowing your food?	time; Most of the time	time				
I0052	GIQLI	Q30: How often during the past week, have you been troubled by urgent bowel movements?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0053	GIQLI	Q31: How often during the past week, have you been troubled by diarrhea?	All of the time	Most of the time	Some of the time	A little of the time	never	
I0054	GIQLI	Q32: How often during the past week, have you been troubled by constipation?	All of the time	Most of the time	Some of the time; A little of the time	Never		
I0055	GIQLI	Q33: How often during the past week, have you been troubled by nausea?	All of the time; Most of the time	Some of the time	A little of the time	never		
I0056	GIQLI	Q34: How often during the past week, have you been troubled by blood in the stool?	NA	NA	NA	NA	NA	
I0057	GIQLI	Q35: How often during the past week, have you been troubled by heartburn?	All of the time; Most of the time	Some of the time; A little of the time	never			
I0058	GIQLI	Q36: How often during the past week, have you been troubled by uncontrolled stool?	All of the time	Most of the time	Some of the time; A little of the time	never		
I0059	VAS	This is a scale to help express the way you feel. For each one please indicate, using the scale from 0-10, how you feel? 6. Anxiety	9; 10 (Worst possible)	4; 5; 6; 7; 8	1; 2; 3	0 (None)		
I0060	VAS	4. Fatigue	9; 10 (Worst possible)	6; 7; 8	3; 4; 5	1; 2	0 (None)	
I0061	VAS	5.Nausea	8; 9; 10 (Worst possible)	5; 6; 7	1; 2; 3; 4	0 (None)		
I0062	VAS	1. Pain at rest	8; 9; 10 (Worst possible)	4; 5; 6; 7	1; 2; 3	0 (None)		
I0063	VAS	2. Pain with coughing	9; 10 (Worst	8	5; 6; 7	1; 2; 3; 4	0 (None)	

			possible)					
I0064	VAS	3. Pain with Movement	9; 10 (Worst possible)	4; 5; 6; 7; 8	1; 2; 3	0 (None)		
I0065	CHAMPS (	Q (D9205) During the past week, did you do any of these activities?	No	Visit with family and friends (other than those you live) (yes)	Go to a social/senior center (yes); Attend other club or group meetings (yes)			
I0066	CHAMPS	Q (D855) During the past week, did you do any of these activities?	No	Do volunteer work (yes)				
I0067	CHAMPS	Q (D850) During the past week, did you do any of these activities?	No	Go to work (yes)				
I0068	CHAMPS	Q (D9300) During the past week, did you do any of these activities?	No	Attend church or take part in church activities (yes)				
I0069	CHAMPS	Q (D3601) During the past week, did you do any of these activities?	No	Use a computer				
I0070	CHAMPS	Q (D9208) During the past week, did you do any of these activities?	No	Dance such as square, folk, line, ballroom, do not count aerobic dance here) (yes)				
I0071	CHAMPS	Q (D9203) During the past week, did you do any of these activities?	No	Do woodworking, needlework, drawing or other arts and crafts (yes)				
I0072	CHAMPS	Q (D9201) During the past week, did you do any of these activities?	No	Bowling (yes); Play golf, riding a cart (yes); Play golf, carrying and pulling your equipment (yes); Play double tennis (yes); Ski (yes); Skate (ice, roller, in-line) (yes); Play basketball, soccer or racquetball (yes); Play single tennis (yes)	Play hockey (yes)			
I0073	CHAMPS	Q (D9202) During the past week, did you do any of these activities?	No	Read (yes)	Attend a concert, movie, lecture or sport event (yes)	Shoot pool or billiards (yes); Play a musical		

						instrument (yes)		
I0074	CHAMPS	Q (D9200) During the past week, did you do any of these activities?	No	Play cards, bingo or board games with other people (yes); Play with kids (yes)	Play darts (yes)			
I0075	CHAMPS	Q (D640) During the past week, did you do any of these activities?	No	Do light work around the house (such as sweeping or vacuuming) (yes)	Do heavy work around the house (such as washing windows, cleaning gutters) (yes)			
I0076	CHAMPS	Q (D650) During the past week, did you do any of these activities?	No	Do light gardening (such as watering plants) (yes); Work on your car, truck; lawn mower, or other machinery (yes)	Do heavy gardening (such as spading, raking)			
I0077	CHAMPS	Q (d455) During the past week, did you do any of these activities?	No	Do water exercises (do not count other swimming) (yes); Do general conditioning exercises, such as light calisthetics or chain exercises (do not count strength training) (yes); Do aerobics or aerobic dancing (yes); Swim gently (yes); Use a treadmill (yes); Jog or run (yes)	Swim moderately or fast (yes)			
I0078	CHAMPS	Q (D450) During the past week, did you do any of these activities?	No	Walk to do errands (such as to/from store or take children to school) (yes); Walk leisurely for exercise or pleasure (yes)	Walk fast or biskly for exercise (yes); Walk uphill or hike uphill (yes)			
I0079	CHAMPS	Q (D4750) During the past week, did you do any of these activities?	No	Ride a bicycle or stationary cycle (yes); Do other aerobic machine such as rowing or stepping machines (yes)				
I0080	CHAMPS	Q (D41) During the past week, did you do any of these activities?	No	Do stretching or flexibility exercises (yes)	Do yoga or tai-chi (yes)			
I0081	CHAMPS	Q (D430) During the past week, did you do any of these activities?	No	Do light strength training (such as hand-held weights of 5lbs or less or				

				elastic bands); Do moderate to heavy strength training (such as hand-held weights of more than 5lbs., weight machines, or push-ups (yes)				
I0082	CHAMPS	Q (D7702) During the past week, did you do any of these activities?	No	Sexual activity (yes)				
I0083	6MWT	Distance walked in 6 minutes	0-200m	201-300m; 301-350m; 351-400m; 401-450m	451-500m; 501-600m	601-800m	>801m	

Appendix 6
Item Deletion

After items were reordered, potential items for deletion were identified using both statistical and conceptual approaches. We relied on an iterative use of Rasch analysis to identify items that could be eliminated without loss of precision and reliability. Potential items were deleted one at a time. With each item deletion, the item scale locations were recalibrated and the person reliability was reevaluated. Item fit values between -2.0 and +2.0 were considered adequate.

The deleted items in the order in which they were deleted and their fit characteristics are shown here:

Items deleted							
Order of Deletion	Item #	Source	Item	Fit Residuals	X^2**	F-statistic	Comment/Rationale
1	I0082	CHAMPS	During the past week, did you do any of these activities? Q (D7702) Sexual activity				This item was deleted for conceptual reasons. Only 1 person provided a response for this item. Not included in the original questionnaire and conceptually different from other activities.
2	I0067	CHAMPS	During the past week, did you do any of these activities? Q (D850) Go to work				This item was deleted for conceptual reasons. Only 1 person provided a response for this item. Not included in the original questionnaire and conceptually different from other activities.
3	I0012	SF36	**Q2: Compared to one week ago**, how would your rate your health in general **now**?	**+6.24**	**56.37**	**20.32**	This item is conceptually different from other items. It is the only one that evaluates change in status
4	I0044	GIQLI	Q22: During the past week how often have you been able to complete your normal daily activities (school, work, household)?	**+5.21**	**16.71**	**5.06**	Item does not fit the data. Patients may have had a different understanding of "normal" daily activities.
5	I0031	GIQLI	Q9: Because of your illness, to what	**+4.08**	**16.98**	**6.26**	Will likely be due to other symptoms such as bloating

			extent have you restricted the kinds of food you eat?				(I0025) and correlates with other items like finding eating to be a pleasure (I0030)
6	I0080	CHAMPS	During the past week, did you do any of these activities? Q (D410) Changing basic body position	**+3.83**	**45.85**	**12.08**	Likely correlates with other low intensity activities like climbing stairs (I0004)
7	I0039	GIQLI	Q17: Over the past week, have you woken up in the night	**+3.96**	**22.44**	**8.68**	Likely caused by physical symptoms like nausea or mental symptoms like frustration
8	I0053	GIQLI	Q31: How often during the past week, have you been troubled by diarrhea?	**+3.80**	**28.98**	**8.74**	High correlation with symptom of uncontrolled stool (I0001) and may not be reflective of symptoms after laparoscopic cholecystectomy
9	I0026	GIQLI	Q4: How often during the past week have you been troubled by the excessive passage of gas through the anus?	**+3.42**	4.64	2.5	May not be reflective of symptoms after laparoscopic cholecystectomy
10	I0052	GIQLI	Q30: How often during the past week, have you been troubled by urgent bowel movements?	**+3.20**	**14.57**	**5.78**	May not be reflective of symptoms after laparoscopic cholecystectomy
11	I0029	GIQLI	Q7: How often during the past week have you been troubled by frequent bowel movements?	**+3.14**	**13.98**	**5.25**	High correlation with symptom of uncontrolled stool (I0001) and may not be reflective of symptoms after laparoscopic cholecystectomy
12	I0078	CHAMPS	During the past week, did you do any of these activities? Q (D450) Walking	**+3.04**	**21.70**	**9.10**	May be reflected in other low intensity activities like climbing stairs (I0004)
13	I0073	CHAMPS	During the past week, did you do any of these activities? Q (D9202) Arts and culture	**+2.65**	**29.37**	**12.86**	Likely not related or affected during recovery phase after surgery
14	I0019	SF36	How much of the time during the **past 4 weeks** . . . -Q24: Have you been a very nervous person?	**+2.67**	**7.09**	2.59	May not well reflect the process of recovery. Patients may be nervous for different reasons pre and post surgery
15	I0027	GIQLI	Q5: How often during the past week	**+2.45**	3.4	1.51	May not be reflective of symptoms after laparoscopic

			have you been troubled by strong burping or belching?				cholecystectomy
16	I0069	CHAMPS	During the past week, did you do any of these activities? Q (D3601) Using writing machines	**+2.36**	5.83	2.41	Not likely affected by recovery process after surgery
17	I0028	GIQLI	Q6: How often during the past week have you been troubled by gurgling noises from the abdomen?	**+2.32**	**15.53**	**5.85**	May not be reflective of symptoms after laparoscopic cholecystectomy
18	I0037	GIQLI	Q15: How often during the past week have you been tired or fatigued?	**-3.27**	**6.42**	**5.95**	Correlates with other items on fatigue and energy (I0022)
19	I0008	SF36	Q10: Walking **several blocks**	**-2.67**	**11.95**	**10.60**	May be reflected in other low intensity activities like climbing stairs (I0004)
20	I0041	GIQLI	Q19: Because of your illness, how much physical strength have you lost?	**-2.43**	3.06	2.51	Item is conceptually different from others as it assesses change from previous state.
21	I0013	SF36	Q13: Cut down the amount of time you spent on work or other activities	**-2.33**	4.94	**3.88**	Likely to be caused by symptoms and correlate with other activities
22	I0042	GIQLI	Q20: To what extent have you lost your endurance?	**-2.23**	**11.61**	**8.64**	Item is conceptually different from others as it assesses change from previous state.
23	I0071	CHAMPS	During the past week, did you do any of these activities? Q (D9203) Crafts	+1.82	**24.94**	**8.14**	Not likely affected by recovery process after surgery
24	I0076	CHAMPS	During the past week, did you do any of these activities? Q (D650) Caring for household objects	+1.73	**23.71**	**8.08**	Correlates highly with other tasks of daily life and leisure (I0045)
25	I0077	CHAMPS	During the past week, did you do any of these activities? Q (D455) Moving around	+1.63	**18.54**	**6.26**	Correlates highly with other vigorous activities (I0001) or sports (I0072)
26	I0070	CHAMPS	During the past week, did you do any of these activities? Q (D9208) Recreation and leisure, other specified	+0.66	**18.39**	**7.30**	Correlates highly with other vigorous activities (I0001) or sports (I0072)
27	I0057	GIQLI	Q35: How often during the past week,	+1.23	**11.18**	**4.50**	Misfit to the model, symptom not likely related to

			have you been troubled by heartburn?				recovery process
28	I0054	GIQLI	Q32: How often during the past week, have you been troubled by constipation?	+1.97	**11.43**	**4.22**	Misfit to the model, symptom not likely related to recovery process
29	I0065	CHAMPS	During the past week, did you do any of these activities? Q (D9205) Socializing	+1.93	**10.21**	**4.23**	Misfit to the model, very low intensity activity not likely affected by recovery process
30	I0075	CHAMPS	During the past week, did you do any of these activities? Q (D640) Doing housework	**+2.11**	**9.70**	**4.09**	Misfit to the model, activity correlates with other items of daily life and leisure (I0045)
31	I0050	GIQLI	Q28: How often during the past week, have you felt uncomfortable because of your slow speed of eating??	-1.76	5.78	**4.87**	Misfit to the model, activity not likely related to recovery process
32	I0081	CHAMPS	During the past week, did you do any of these activities? Q (D430) Lifting and carrying objects	+1.54	**11.20**	**3.90**	Misfit to the model, activity correlates with other high intensity activity (I0072)
33	I0074	CHAMPS	During the past week, did you do any of these activities? Q (D9200) Play	+1.10	**12.31**	**3.42**	Misfit to the model, activity correlates with other items of daily life and leisure (I0045)
34	I0079	CHAMPS	During the past week, did you do any of these activities? Q (D4750) Driving human-powered transportation	+1.39	**11.85**	**3.71**	Misfit to the model, activity not likely related to recovery process
35	I0049	GIQLI	Q27: How often during the past week, have you been troubled by fluid or food coming up into your mouth (regurgitation)?	**+2.04**	1.44	0.47	Misfit to the model, correlates with other symptoms (I0045)
36	I0009	SF36	Q11: Walking **one block**	-1.43	**6.26**	**6.47**	Misfit to the model, activity correlates with other items of low and moderate intensity (I0004 & I0006)
37	I0047	GIQLI	Q25: To what extent have your personal relations with people close to you (family or friends) worsened	-1.24	**5.55**	**4.56**	Misfit to the model, item correlates with other items assessing social function (I0015)

			because of your illness ?				
38	I0051	GIQLI	Q29: How often during the past week, have you had trouble swallowing your food?	-0.98	4.57	**3.75**	Misfit to the model, correlates with other symptoms (I0045)
39	I0003	SF36	Q5: Lifting or carrying groceries	-1.42	**6.10**	**4.45**	Misfit to the model, activity correlates with other items of moderate intensity (I0004)
40	I0068	CHAMPS	During the past week, did you do any of these activities? Q (D9300) Organized religion	+1.07	**7.83**	2.46	Misfit to the model, low intensity activity, likely not related to construct
41	I0007	SF36	Q9: Walking **more than a mile**	-1.47	**6.27**	**3.88**	Misfit to the model, activity correlates with other items of moderate intensity (I0004)
42	I0005	SF36	Q7: Climbing **one** flight of stairs				Redundant item with I0004 with correlation 0.54 but higher SE 0.22 vs 0.13
43	I0055	GIQLI	Q33: How often during the past week, have you been troubled by nausea?				Redundant item with I0061 with correlation 0.50 but higher SE 0.15 vs 0.11
44	I0059	VAS	This is a scale to help express the way you feel. For each one please indicate, using the scale from 0-10, how you feel? 6. Anxiety	+1.18	**6.71**	2.82	May not well reflect the process of recovery. Patients may have anxiety for different reasons pre and post surgery
45	I0034	GIQLI	Q12: How often during the past week have you been nervous or anxious about your illness?				Differential item functioning across time of evaluation(with Bonferroni adjustment)
46	I0060	VAS	Fatigue				Differential item functioning across time of evaluation(with Bonferroni adjustment)
47	I0038	GIQLI	Q16: How often during the past week have you felt unwell?	-1.68	**6.32**	**4.28**	Misfit to model, likely affected by many other items
48	I0002	SF36	Q4: **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf				Differential item functioning across time of evaluation(with Bonferroni adjustment)

LIST OF REFERENCES

Acquadro C, B. R., Dubois D et al. (2003). "Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. ." Value Health **6**(5): 522-531.

Ainsworth, B., Haskel,l WL, Leon, AS, Jacobs, DR, Montoye, HJ, Sallis, JF, Paffenbarger, RS. (1993). "Compendium of physical activities: classification of energy costs of human physical activities." Med. Sci. Sports Exerc. **25**: 71-80.

Aitken, R. (1969). "Measurement of feelings using visual analogue scales." J Roy Soc Med **62**: 989-993.

Aldrete, J., Kroulik, D. (1970). "A postanesthetic recovery score." Anesth Analg **49**: 924-933.

Andersen, E. B. (1977). "Sufficient statistics and latent trait models." Psychometrika **42**: 69-81.

Andrich, D. (1978). "A rating formulation for ordered response categories." Psychometrika **43**: 357-374.

Andrich, D., Sivakumar Alagumalai, David D Durtis, and Njora Hungi (2005). "The Rasch model explained." Applied Rasch Measurement: A book of exemplars. **Springer-Kluwer. Chapter 3**: 308-328.

Andrich, N. (1988). "Rasch Models for Measurement " Sage Publications, Newbury Park, CA

Andrich, D. (2005). "Rasch models for ordered response categories." Encyclopedia Stat BehavSci **4**: 1698-1707.

Andrich, D., Hagquist, C. (2004). "Detection of differential item functioning using analysis of variance." Personality and Individual Difference **36**(955-968).

Andrich D, S. B., editors (2005). "Interpreting RUMM2020." Perth: RUMM Laboratory Pty Ltd.

Arellano R, C. C., Chung F (1989). "Timing of the anesthetist's preoperative outpatient interview." Anesth Analg **68**: 645-8.

Ashby, J. T. T. a. F. G. (1984). "Measurement scales and statistics: the misconception misconceived." Psychol Bull **96**: 394-401.

AW Majeed, S. B., N Williams, DR Hannay and AG Johnson (1995). "Variations in medical attitudes to postoperative recovery period." BMJ 311 (1995), p. 296.

Barkun JS, B. A., Sampalis JS, Fried G, Taylor B, Wexler MJ, Goresky CA, Meakins JL (1992). "Randomised controlled trial of laparoscopic versus mini cholecystectomy. The McGill Gallstone Treatment Group." Lancet **340**: 1116-1119.

Bergman, S., L. S. Feldman, et al. "Evaluating surgical outcomes." Surgical Clinics of North America **86**(1): 129-49.

Bergman, S., L. S. Feldman, et al. (2005). "Measuring surgical recovery: the study of laparoscopic live donor nephrectomy." American Journal of Transplantation **5**(10): 2489-95.

Bergner M, B. R., Carter WB, Gilson BS. (1981). "The Sickness Impact Profile: Development and final revision of a health status measure. ." Med Care **19**: 787-805.

Berzon, R., Hays, R.D. and S. A. Shumaker (1993). "International use, application and performance of health-related quality of life instruments." Quality of Life Research **2**(6): 367-368.

Birkmeyer, J., Justin B. Dimick MD*, † and Nancy J. O. Birkmeyer PhD† (2004). "Measuring the quality of surgical care: structure, process, or outcomes?" Journal of the American College of Surgeons, Volume 198, Issue 4, April 2004, Pages 626-632.

Bode RK, C. D., Lai JS, Heinemann AW (2003). "Developing an initial physical function item bank from existing sources." Journal of Applied Measurement **4**: 124-136.

Bond, T. G., Fox, C.M. (2001). "Applying the Rasch Model: Fundamental measurement in the human sciences." Lawrence Erlbaum.

Boogaerts, J., Vanacker, E, Seidel, L, Albert, A and Bardiau, FM (2001). "Assessment of postoperative nausea using a visual analogue scale." Acta Anaesthesiologica Scandinavica **44**(4): 470-474.

Bowling, A. (1985). "MEASURING DISEASE: A REVIEW OF DISEASE-SPECIFIC QUALITY OF LIFE MEASUREMENT SCALES." Open University Press, Buckingham · Philadelphia.

Bremers AJ, R. J., Vijn A, Janss RA, Bemelman WA (2000). "Laparoscopic adhesiolysis for chronic abdominal pain: an objective assessment." J Laparoendosc Adv Surg Tech A **10**: 199-202.

Burney, R., Jones, KR (2002). "Ambulatory and admitted laparoscopic cholecystectomy patients have comparable outcomes but different functional health status." Surgical Endoscopy **16**: 921-926.

C. Merbitz, J. M. a. J. C. G. (1989). "Ordinal scales and foundations of misinference." Arch Phys Med Rehabil **70**: 308-312.

Cahalin L, P. P., Prevost S, Wain J, Ginns L. (1995). "The relationship of the 6-min walk test to maximal oxygen consumption in transplant candidates with end-stage lung disease." Chest **108**: 452-9.

Cahalin LP, M. M., Semigran MJ, et al. (1996). "The six-minute walk test predicts peak oxygen uptake and survival in patients with advanced heart failure. ." Chest **110**: 325-332.

Camilleri-Brennan J, a. S. R. (2001). "The impact of recurrent rectal cancer on quality of life." EJSO **27**: 349-353.

Carli, F. a. M., N (2001). "Measuring the outcome of surgical procedures: what are the challenges?" British Journal of Anaesthesia **87**(4): 531-533.

Carli, F. Z., Gerald S (2005). "Optimizing functional exercise capacity in the elderly surgical population." Current Opinion in Clinical Nutrition and Metabolic Care **8**(1): 23-32.

Carroll NV, M. P., Cox FM, et al (1995). "Postoperative nausea and vomiting after discharge from outpatient surgery centers." Anesth Analg 1995;80:903–9.

Chen L, D. N., Shi X, Tao S, Zhang W (2002). "Life quality of patients after cholecystectomy." Zhonghua Wai Ke Za Zhi **40**: 762-765.

Chumbley, G., Hall, GM (1997). "Recovery after major surgery: does the anaesthetic make any difference?" British Journal of Anaesthesia, 1997 **78**(4): 347-349.

Chung, F. a. M., G (1999). "Factors contributing to a prolonged stay after ambulatory surgery." Anesth Analg 89 6 (1999), pp. 1352–1359.

Clifton JC, F. R. (2001). "Quality-of-Life Measurement in Surgical Randomized Controlled Trials." Journal of Investigative Surgery **14**(5): 253-258.

Cullen, K. (2009). "Ambulatory Surgery in the United States, 2006." National Health Statistics Reports **11**: 1-28.

Cyarto, E., Marshalla, AL, Dickinsona, RK and Brown, WJ (2006). "Measurement properties of the CHAMPS physical activity questionnaire in a sample of older Australians." Journal of Science and Medicine in Sport **9**(4): 319-326

Dallal, G. (1986). "PC-SIZE: A Program for Sample-Size Determinations." The American Statistician **50**: 52.

Dawson L, B. K., Carr EC, Barrett RF. (1999). "Improving patients' postoperative sleep: a randomized control study comparing subcutaneous with intravenous patient-controlled analgesia." J Adv Nurs 1999; 30:875-881.

DeCherney AH, B. G., Isaacson K, Gall S. (2002). "Postoperative fatigue negatively impacts the daily lives of patients recovering from hysterectomy." Obstet Gynecol 2002; 99:51-57.

DeVellis, R. (2006). "Classical Test Theory." Med Care **44**(11): S50-59.

Dijkers, M. (2003). "A computer adaptive testing simulation applied to the FIM instrument motor component." Arch Phys Med Rehabil **84**: 384-393.

Dindo, D., Demartines, N, Clavien, PA (2004). "Classification of Surgical Complications: A New Proposal With Evaluation in a Cohort of 6336 Patients and Results of a Survey." Ann Surg **240**(2): 205-213.

Donabedian, A. (1985). "Explorations in quality assessment and monitoring." Ann Arbor, MI: Health Administration Press **1**: 256.

Ellis, B., Kimmel, HD (1992). "Identification of unique cultural response patterns by means of Item Response Theory." J of Applied Psychology **77**: 177-184.

Engel, G. (1977). "The need for a new medical model: A challenge for biomedicine." Science **196**: 129-136.

Enright PL, S. D. (1998). "Reference equations for the six-minute walk in healthy adults." Am J Respir Crit Care Med **158**: 1384-7.

Eypasch E, W. J., Wood-Dauphinee S et al. (1995). "Gastrointestinal Quality of Life Index:

Development, validation and application of a new instrument." Br J Surg **82**: 216-22.
Feldman, L. S., P. Kaneva, et al. (2009). "Validation of a physical activity questionnaire (CHAMPS) as an indicator of postoperative recovery after laparoscopic cholecystectomy." Surgery **146**(1): 31-9.
Figueredo ED, C. L. (1998). "Ondansteron in the prophylaxis of postoperative vomiting: A meta-analysis." J Clin Anesth 1998; 10: 211-21
Finch E, B. D., Stratford PW, Mayo NE. (2002). "Physical rehabilitation outcome measures." 2nd ed. Hamilton: BC Decker;.
Fischer, G. H. M., I.W. (1995). "Rasch models: foundations, recent developments and applications. ." New York: Springer-Verlag.
Fisher, D. M. (1999). "Surrogate outcomes: meaningful not!." Anesthesiology 90 2 (1999), pp. 353–359.
Fitts, S., Guthrie, MR. (1995). "Six-minute walk by people with chronic renal failure. Assessment of effort by perceived exertion." Am J Phys Med Rehabil **74**: 54-8.
Fitzpatrick R, F. A., Gore S, Jones D, Spiegelhalter D, Cox D (1992). "Quality of life measures in health care. I: Applications and issues in assessment." BMJ **305**: 1074-1077.
Freyd, M. (1923). "The graphic rating scale." J Educ Psychol **43**: 83-102.
Freys SM, T. H., Heimbucher J, Fuchs KH, Fein M, Thiede A (2001). "Quality of life following laparoscopic gastric banding in patients with morbid obesity." J Gastrointest Surg **5**: 401-407.
Gaston-Johansson F, F.-D. J., Bakos AB, Kennedy MJ. (1999). "Fatigue, pain, and depression in pre-autotransplant breast cancer patients." Cancer Pract **7**(5): 240-7.
Gaston-Johansson F, G. M. (1990). "Rheumatoid arthritis: determination of pain characteristics and comparison of RAI and VAS in its measurement." Pain **41**: 35-40.
Gibbons WJ, F. N., Sloan S, Levy RD. (2001). "Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years." J Cardiopulm Rehabil **21**: 87-93.
Gunnarsson, O., Judge, JO, Earles, DR, Marcella, GR. (1997). "A comparison of walking programs for older adults: effects on six minute walking distance " Gerontologist **37**(126).
Guyatt, G. F., DH; Patrick DL (1993). "Measuring Health-Related Quality of Life." Annals of Internal Medicine **118**(8): 622-629.
Guyatt GH, B. C., Tugwell PX (1986). "Measuring disease-specific quality of life in clinical trials." CMAJ **134**: 889-895.
Guyatt GH, S. M., Thompson PJ, et al. (1985). "The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure." Can Med Assoc J **132**: 919-23.
Guyatt GH, T. P., Berman LB, Sullivan MJ, Townsend M, Jones NL, et al. (1985). "How should we measure function in patients with chronic heart and lung disease?" J Chron Dis. **38**: 517-24.
Harada, N., V. Chiu, A. King, O. Scremin, and A. L. Stewart (1997). "Validity of three self-report physical activity measures in older adults." *Gerontologist* **37**: 331.
Hays RD, M. L., Reise SP (2000). "Item Response Theory and Health Outcomes Measurement in the 21st Century." Med Care **38**(S9): II28-II42.
Hogue CC, B. M., DeVellis BM, Riegger-Krugh C, Rosemond, C. (1997). "Effects of an integrated exercises intervention on sedentary older adults with arthritis " Gerontologist **37**: 126.
Holzner B, B. R., Hahn EA, Cella D, Kopp M, Sperner-Unterweger B, Kemmler G (2006). "Equating EORTC QLQ-C30 and FACT-G scores and its use in oncological research. ." European Journal of Cancer **42**: 3169-3177.
Hopman, W., Towheed,T, Anastassiades,T, Tenenhouse, A, Poliquin,S, Berger,C, Joseph,L, Brown, JP,Murray,TM, Adachi,JD, Hanley,DA, Papadimitropoulos, E and the Canadian and M. O. S. R. Group (2000). "Canadian normative data for the SF-36 health survey " CMAJ **163**(3): 265-271.
Jenkinson, C., Fitzpatrick,F, Garratt,A, Peto,V,Stewart-Brown, S (2001). "Can item response theory reduce patient burden when measuring health status in neurological disorders?
Results from Rasch analysis of the SF-36 physical functioning scale (PF-10)." J Neurol Neurosurg Psychiatry **71**: 220-224.

Jensen Ml', M. C. (1993). "Increasing the reliability and validity of pain intensity measurement in chronic pain patients." Pain **55**: 195-203.
Kadikar A, M. J., Kesten S. (1997). "The six-minute walk test: a guide to assessment for lung transplantation. ." J Heart Lung Transplant. **16**: 313-319.
Kalish RL, D. J., Duncan CC, Davis RB, Coffman GA, Iezzoni LI. (1995). "Costs of potential complications of care for major surgery patients." Am J Med Quality 1995; 10: 48–54.
Katz JN, L. M., Phillips CB, Fossel AH, Liang MH (1992). "Comparative measurement sensitivity of short and longer health status instruments." Med Care **30**: 917-925.
Kervio G, C. F., Ville NS. (2003). "Reliability and intensity of the six-minute walk test in healthy elderly subjects." Med Sci Sports Exerc **35**: 169-174.
Korenkov M, S. S., Arndt M, Bograd L, Neugebauer EAM, Troidl H (2002). "Randomised clinical trial of suture repair,polypropylene mesh or autodermal hernioplasty for incisional hernia." Br J Surg **89**: 50-56.
Korolija, K., Sauerland, S., Wood-Dauphine, S., Abbou, CC, Eypasch,E., et al. (2004). "Evaluation of quality of life after laparoscopic surgery." Surg Endosc **18**: 879-897.
Krupp, L. L., NG; Muir-Nash,J; Steinberg, AD (1989). "The Fatigue Severity Scale: Application to Patients With Multiple Sclerosis and Systemic Lupus Erythematosus " Arch Neurol. **46**(10): 1121-1123.
Linacre, B. D. W. a. M. (1989). "Observations are always ordinal; measurements, however, must be interval." Arch Phys Med Rehabil **70**: 857-860.
Linacre, J. (2002). "Optimizing rating scale category effectiveness." J Appl Meas **3**: 85-106.
Lord, F. a. N., MR (1968). "Statistical theories of mental test scores." Addison-Wesley, Reading, MA.
Lohr, K. (1988). "Outcome measurement: concepts and questions." Inquiry 25 (1988), pp. 37–50.
Ludwig K, P. K., Wilhelm L, Bernhardt J (2002). "Prospektive Analyse zur Outcomebewertung nach laparoskopischer versus konventioneller Cholecystektomie." Zentralbl Chir **127**: 41-47.
Lusardi, M., Pellecchia,GL, Schulman, M (2003). "►Functional performance in community living older adults." Journal of Geriatric Physical Therapy **26**(3): 14-22.
Lynch, E., Lazor, MA, Gellis,JE, Orav,J, Goldman, L and Marcantonio, ER (1998). "The impact of postoperative pain on the development of postoperative delirium." Anesthesia & Analgesia **86**: 781-785.
Manocchia M, B. M., Connor J, Keller SD, Shiely JC, Tasai C, et al. (1998). "SF-36 Health Survey Annotated Bibliography: Second Edition (1988–1996) " The Health Assessment Lab, New England Medical Center, Boston, MA
Marshall, S. (1999). "Discharge Criteria and Complications After Ambulatory Surgery." Anesth Analg 1999;88:508.
Masters, G. (2005). "Objective Measurement." Applied Rasch Measurement: A Book of Exemplars: 15-25.
Mayo, N. (2004-2009). "Choosing a Measurement Plan for Clinical Rehabilitation Research." Classnotes; POTH 610A, McGill University
Mayo, N. (2009). "Randomized Trials and Other Parallel Comparisons of Treatment (Chapter 4). ." In J Bailar and D Hoaglin, Medical Uses of Statistics. **J Wiley and Sons.** .
McCarthy Jr, M., Jonasson, O, Chang, CH, Pickard, AS, Giobbie-Hurder, A, Gibbs, J, Edelman, P, Fitzgibbons, R, Neumayer, L (2005). "Assessment of Patient Functional Status after Surgery." J Am Coll Surg **201**(2): 171-178.
McDonald, R. (1981). "The dimensionality of tests and items." Br J Math Stat Psychol. **34**: 100-117.
McElduff P, A. J., Ewald B, Cockburn J, Heller R (2002). "Estimating the contribution of individual risk factors to disease in a person with more than one risk factor." J Clin Epidemiol **55**: 588-592.
Mcgrath (2003). "postoperative recovery and discharge." anesthesiology Clin N Am **21**: 337-386.
McHomey CA, W. J., Raczek AE. (1993). "The MOS 36-item short-form health survey: II. Psychometric and clinical tests of validity in measuring physical and mental health constructs." Med Care **31**: 247-63.
McHomey CA, W. J., Rogers W, Raczek AE, Lu JFR. (1992). "The validity and relative precision

of MOS short- and long-form health status scales and Dartmouth COOP charts: results from the medical outcomes study." Med Care **30**: MS253-65.

McHorney, C., Cohen, AS (2000). "Equating health status measures with item response theory: illustrations with functional status items." Med Care **38**(Suppl 9): II43-II59.

Millar K, J. M., Bonke B, Asbury AJ. (1995). "Assessment of preoperative anxiety: comparison of measures in patients awaiting surgery for breast cancer." Br J Anaesth **74**: 180-3.

Miller PA, M. J., Stevenson TJ. (2002). "Measurement properties of a standardized version of the two-minute walk test for individuals with neurological dysfunction." Physiother Can **54**: 241-8.

Miyamoto S, N. N., Satoh T, et al. (2000). "Clinical correlates and prognostic significance of six-minute walk test in patients with primary pulmonary hypertension. Comparison with cardiopulmonary exercise testing." Am J Respir Crit Care Med **161**: 487-92.

Montgomery PS, G. A. (1998). "The clinical utility of a six-minute walk test in peripheral arterial occlusive disease patients." J Am Geriatr Soc **46**: 706-11.

Moriello, C., N. E. Mayo, et al. (2008). "Validating the six-minute walk test as a measure of recovery after elective colon resection surgery." Archives of Physical Medicine & Rehabilitation **89**(6): 1083-9.

Myles, P. S. (1999). "Development and Psychometric Testing of a Quality of Recovery Score After General Anesthesia and Surgery in Adults." Anesth Analg 1999;88:83-90.

Myles PS, W. D., Hendrata M, et al. (2000). "Patient satisfaction after anaesthesia and surgery: results of a prospective survey of 10811 patients." Br J Anaesth 2000; 84:6-10.

Myles, P., Hunt,JO, Fletcher,H, Solly,R., Woodward,D., Kelly, S. (2001). "Relation between Quality of Recovery in Hospital and Quality of Life at 3 Months after Cardiac Surgery." Anesthesiology **95**: 862-867.

Myles, P., Weitkamp, B., Jones, K., Melick, J., Hensen, S. (2000). "Validity and reliability of a postoperative quality of recovery score: the QoR-40." Br J Anaesth **84**: 11-15.

Natof, H. E. (1980). "Complications associated with ambulatory surgery." JAMA 244 10 (1980), pp. 1116–1118.

Nomori, H., T. Ohtsuka, et al. (2003). "Difference in the impairment of vital capacity and 6-minute walking after a lobectomy performed by thoracoscopic surgery, an anterior limited thoracotomy, an anteroaxillary thoracotomy, and a posterolateral thoracotomy." Surgery Today **33**(1): 7-12.

Norquist, J., Fitzpatrick, R, Dawson, J, Jenkinson, C (2004). "Comparing Alternative Rasch-Based Methods vs Raw Scores in Measuring Change in Health." Medical Care **42**(1): 1-25.

P.G. Duncan, M. M. C., W.A. Tweed et al. (1992). "The Canadian four-centre study of anaesthetic outcomes: III. Are anaesthetic complications predictable in day surgical practice?." Can J Anaesth 39 5 Pt 1 (1992), pp. 440–448. .

Pallant, J. a. T., A (2007). "An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS)." British Journal of Clinical Psychology **46**: 1-18.

Pallant JF, M. R., Tennant A (2006). "Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis." BMC Psychiatry **6**: 28.

Pastor, D., Beretvas, SN (2003). "Running Head: LONGITUDINAL RASCH MODELING: An Illustration of Longitudinal Rasch Modeling in the Context of Psychotherapy Outcomes Assessment " National Council of Measurement in Education, Chicago, April 21-25, 2003.: 1-30.

Patrick DL, B. M. (1990). "Measurement of health status in the 1990s." Annu Rev Public Health. 1990; 11:165-83.

Patrick DL, D. R. (1989). "Generic and Disease-Specific Measures in Assessing Health Status and Quality of Life." Med Care **27**: S217-S232.

Patrick, D., Erickson, P. (1993). "Health Status and Health Policy. Allocating Resources to Health Care. ." New York: Oxford University Press;

Penfield, R. (2005). "Unique properties of Rasch model item information functions." J Appl Meas **6**: 355-65.

Prieto, L., Alonso, J, Lamarca, R (2003). "Classical test theory versus Rasch analysis for quality of life questionnaire reduction." Health and Quality of Life Outcomes **1**(27).

Raczek, A. e. a. (1998). "Comparison of Rasch and Summated Rating Scales Constructed from SF-36 Physical Functioning Items in Seven Countries: Results from the IQOLA Project." Journal of Clinical Epidemiology **51**(11): 1203-1214.

Reise SP, W. K., Pugh RH (1993). "Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance." Psychol Bull. **114**: 552-566.

Resnicow, K. M., F; Blissett, D; Wang, T; Heitzler, C; Lee, RE (2003). "Validity of a Modified CHAMPS Physical Activity Questionnaire among African-Americans." Medicine & Science in Sports & Exercise **35**(9): 1537-1545.

Revicki, D. a. C., CF (1997). "Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing." Qual Life Res **6**: 595-600.

Roul G, G. P., Bareiss P. (1998). "Does the 6-minute walk test predict the prognosis in patients with NYHA class II or III chronic heart failure?" Am Heart J **136**: 449-57.

S.V.M. Kleinbeck and N. Hoffart (1994). "Outpatient recovery after laparoscopic cholecystectomy." AORN J 60 (1994), pp. 394–402.

Sands, W., Waters, and McBride, JR (1997). "Computerized adaptive testing. From inquiry to operation,." American Psychological Association, Washington, DC.

Schwenk, W., Böhm, B and Müller, JM (1998). " Postoperative pain and fatigue after laparoscopic or conventional colorectal resections: A prospective randomized trial " Surgical Endoscopy **12**: 1131-1136.

Scott J, H. E. (1976). "Graphic representation of pain." Pain **2**: 175-184.

Sloan JA, A. N., Cappelleri JC, Fairclough DL, Varricchio C; (2002). "Assessing the Clinical Significance of Single Items Relative to Summated Scores." Mayo Clinic Proceedings **77**: 479-487.

Smith, E. J. (2001). "Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective." J Appl Mea **2**: 281-311.

Smith, E. V. (2002). "Detecting and evaluating the impact of multidimensionality using item fit statistics and principle components analysis of residuals." Journal of Applied Measurement **3**(2): 205-231.

Sousa, K., Kwok OM (2006). "Putting Wilson and Cleary to the test: analysis of a HRQOL conceptual model using structural equation modeling." Quality of Life Research **15**: 725-737.

Steffen TM, H. T., Mollinger L. (2002). "Age- and gender-related test performance in community-dwelling elderly people: Six- Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds." Phys Ther **82**(128-37).

Stewart AL, G. S., Hays RD, Wells K, Rogers MH, Berry SD, at al (1989). "Functional status and well-being of patients with chronic conditions: results from the medical outcomes study." JAMA **262**: 907-13.

Stewart, A. L., K. M. Mills, P. G. Sepsis, et al. (1997). "Evaluation of CHAMPS, a physical activity promotion program for older adults." Ann. Behav. Med. **19**: 353-361.

Stewart, A. M., KM.; King, AC.; Haskell, WL.; Gillis, D; Ritter, PL. (2001). "CHAMPS Physical Activity Questionnaire for Older Adults: outcomes for interventions." Medicine and Science in Sports and Exercise **33**(7): 1126-1141.

Stewart, D., Volgyesi G (1978). "Stabilometry: a new tool for measuring recovery following general anaesthesia." Can Anesth Soc J 1978;25:4–6.

Stewart, A., Verboncoeur, CJ, McLellan, BY et al. (2001). "Physical Activity Outcomes of CHAMPS II: A Physical Activity Promotion Program for Older Adults." Journals of Gerontology **56A**(8): M465-M470.

Streiner, D., Norman, GR, editors (1995). "Health measurement scales: a practical guide to their development and use." Oxford: Oxford University Press.

Tennant, A., Pallant, JF (2006). " Unidimensionality matters! (A tale of two Smiths?)." Rasch Measure Trans **20**: 1048-51.

Troosters T, G. R., Decramer M. (1999). "Six minute walking distance in healthy elderly subjects." Eur Respir J **14**: 270-4.

Trus TL, L. W., Waring JP, Branum GD, and Hunter JG (1999). "Improvement in quality of life measures after laparoscopic antireflux surgery." Ann Surg **229**(3): 331-336.

Valderas, J., Alonso, J (2008). "Patient reported outcome measures: a model-based classification

system for research and clinical practice." Qual Life Res **17**(1): 1125-1135.
Vogelsang, J. (1988). "The visual analog scale: an accurate and sensitive method for self-reporting preoperative anxiety." J Post Anesth Nursing **3**: 235-9.
Ware JE Jr, S. C. (1992). "The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. ." Med Care **30**: 472-483.
Ware JE, K. M. a. K. S. (1994). "SF-36 Physical and Mental Health Summary Scales: A User's Manual "The Health Institute, Boston, MA.
Ware JE, S. K., Kosinski M and Gandek B (1993). "SF-36 Health Survey Manual and Interpretation Guide " New England Medical Center, The Health Institute, Boston, MA.
Ware, J., Bjorner , JB and Kosinski, M (2000). "Practical implications of Item Response Theory and computerized adaptive testing. A brief summary of ongoing studies of widely used headache impact scales." Med Care **38**: (9, Suppl II) II-73–82.
Ware, J. E. (1987). "Standards for validating health me asure s: Definition and content." *J. Chron.Dis.* **40**: 473-480.
Ware JE, S. K., Kosinski M and Gandek B (1993). "SF-36 Health Survey Manual and Interpretation Guide " New England Medical Center, The Health Institute, Boston, MA.
Warner MA, S. S., Chute CG (1993). "Major morbidity and mortality within 1 month of ambulatory surgery and anaesthesia." JAMA 1993;270:1437–41.
Weinberger M, S. G., Hanlon JT et al. (1991). "An evaluation of a brief health status measure in elderly veterans. ." J Am Geriatr Soc **39**: 691-694.
Wennberg, J. (1990). "Outcomes research, cost containment, and the fear of health care rationing." N Engl J Med. **323**: 1202-4.
White, K. (1969). "Discussion of An Historical View of Teaching of Medical Care Administration." American Journal of Public Health and the Nations Health **59**(1P2S): 66-70.
White, P. (1994). "Anaesthesia for day surgery: past, present, and future." Curr Anaesth Crit Care 1994;5:123–6.
Wilson IB, C. P. (1995). "Linking clinical variables with health-related quality of life." JAMA **273**: 59-65.
Wright, B. a. S., MH (1979). "Best Test Design: Rasch measurement." MESA Press, Chicago.
Wright BD, M. G. (1982). "Rating Scale Analysis." Chicago: MESA Press.
Wu CL, N. M., Rowlingson AJ, et al. (2003). "The effect of postoperative analgesia on quality-of-life measurements. ." Anesth Analg 2003; 97:1078-1085.
"The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. ." Soc Sci Med **41**(10): 1403-1409. 1995