EVALUATION OF THE VALIDITY OF THE EUROQOL-5 DIMENSIONS (EQ-5D) AS A MEASURE OF POSTOPERATIVE RECOVERY AFTER PULMONARY RESECTION

Jimmy Bejjani, MD
Department of Surgery
McGill University
Montreal, Quebec, Canada
February 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science (Experimental Surgery).

© Jimmy Bejjani, 2016

ACKNOWLEDGEMENTS

I would like to thank **Dr. Liane S Feldman**, my supervisor, for her guidance during the completion of my research projects, manuscripts and the writing of my thesis. I am also grateful to my colleagues **Dr. Julio F Fiore Jr**, postdoctoral research fellow, and **Ms. Pepa Kaneva**, research group coordinator, for their advices in research methodology and statistical analyses. I also want to convey my appreciation to **Ms. Annie Ncuti** and **Dr. Juan Mata**, former research assistants, for their assistance in patients' recruitment and prospective data collection. As well, I'm thankful to **Dr. Jacques Lapointe**, chair of my research advisory committee, for his academic support and constructive feedback on my thesis content, and **Dr. Kevin Schwarzman** and **Dr. David Bracco**, members of the committee, for their constructive feedback.

Finally, I express my gratitude to **Dr. Jean-Jacques Lebrun**, Associate Dean of Graduate and Postdoctoral Studies at the Faculty of Medicine, and **Dr. Anie Philip**, Program Director of the M.Sc. program in Experimental Surgery, for their administrative support in completing my degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CMLND: Complete ipsilateral mediastinal lymph node dissection
COPD: Chronic obstructive pulmonary disease
CUA: Cost-utility analysis
DALYs: Disability-adjusted life-years
DAT: Diet as tolerated
EORTC-QOL-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire
EQ-5D: EuroQol - 5 dimensions
ERP: Enhanced-recovery pathway
HALYs: Health adjusted life-years
HRQoL: Health-related quality of life
HUI: Health Utilities Index
ICF: International Classification of Functioning, Disability and Health
IQR: Interquartile range
LOS: Length of stay
MCID: Minimal clinically important difference
NICE: UK National Institute of Health and Clinical Excellence
PACU: Post-anesthesia care unit
POD: Postoperative day
QALY: Quality-adjusted life-years
QLI: Spitzer Quality of Life Index
QoL: Quality of life
QoR-40: Quality of recovery-40 questionnaire
QWB: Quality of Well-Being Scale
PRO: Patient-reported outcomes
PSR scale: Postdischarge surgical recovery scale
RCT: Randomized controlled trial
SCLC: Small cell lung cancer
SD: Standard deviation
SF-6D: Short Form 6D
SF-36: Medical Outcome Survey 36-Item Short Form Health Questionnaire
SG: Standard gamble method
SIP: Sickness Impact Profile
TNM: Tumor, Node, Metastasis
TTO: Time Trade-Off method
VAS: Visual analog scale
VATS: Video-assisted thoracoscopic surgery
WHO: World Health Organization
WTD: Worse than death

ABSTRACT

Background: Innovations in thoracic surgery, such as video-assisted surgery (VATS) and enhanced recovery pathways (ERPs) are intended to improve outcomes after pulmonary resection, including patient recovery. However, these innovations have implementation costs. Full economic appraisal requires preference-based measures that reflect the construct of "postoperative recovery". Currently, there is no validated measure of postoperative recovery after pulmonary resection. The EuroQol-5 dimensions (EQ-5D) is a generic standardized measure of health-related quality of life used in a wide variety of conditions for clinical and economic evaluation. Hence, we investigated the responsiveness and construct validity of the EQ-5D as a measure of postoperative recovery after planned pulmonary resection for suspected malignant tumors.

Methods: Patients undergoing pulmonary resection completed the EQ-5D questionnaire and visual analog scales (VAS) for pain and fatigue at baseline (preoperatively) and at one and three months postoperatively. The EQ-5D includes a descriptive health profile comprising 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) weighted from valuation sets based on general populations to provide a summary index score (range: 1 to -0.109). Longitudinal validity (responsiveness) was assessed by testing the *a priori* hypothesis that EQ-5D scores would follow the trajectory of postoperative recovery i.e. decrease at one month after surgery in comparison to baseline, improve from one to three months, and return to baseline levels at three months. Discriminant construct validity was assessed by testing the *a priori* hypotheses that EQ-5D scores would be lower after surgery for patients above 70 years old in comparison to younger patients, for patients undergoing an open surgical approach (thoracotomy) in comparison to those undergoing VATS, and for patients who developed complications in comparison to those who did not. Construct convergent validity was assessed by testing the *a priori* hypothesis that EQ-5D scores would inversely correlate with pain and fatigue levels (Spearman's correlation; *r*). Missing values (5%) were handled with multiple imputations.

Results: Fifty-five patients were analyzed (45% male, 62±12 years, 29% video-assisted). There was no significant difference between median EQ-5D scores obtained at baseline (0.83 [IQR 0.80-1]) compared to one month (0.83 [0.80-1], p=0.86) and three months after surgery (1 [0.83-1]; p=0.09). At one month after surgery, EQ-5D scores were significantly lower in patients undergoing thoracotomy vs. video-assisted surgery (0.82 [IQR 0.77-0.89] vs. 1 [0.83-1], p=0.003), but there were no significant differences between patients ≥70 years old vs. younger (0.95 [IQR 0.82-1] vs. 0.83 [0.77-1], p=0.09) or

between patients with vs. without complications (0.82 [IQR 0.79-0.95] vs. 0.83 [0.80-1], p=0.10). There was a low but significant correlation between EQ-5D and VAS scores of pain and fatigue (Rho -0.30 to -0.47, $p \leq 0.01$).

Conclusion: Despite some evidence of convergent validity, the EQ-5D was not sensitive to the hypothesized trajectory of postoperative recovery and showed limited discriminant validity. Therefore, the EQ-5D may not be a valid measure of postoperative recovery one month after lung resection and should be used with caution for economic evaluations of surgical technologies aimed at improving post-hospital recovery.

RÉSUMÉ

Contexte: Les innovations en chirurgie thoracique, telles que la thoracoscopie assistée par vidéo (VATS) et les plans de rétablissement renforcé, visent à améliorer les résultats après une résection pulmonaire. Cependant, l'implémentation de ces innovations nécessite une augmentation des coûts. Leur évaluation complète d'un point de vue économique requiert des mesures basées sur les préférences et qui reflètent le concept de rétablissement postopératoire, et nécessite la quantification du rétablissement postopératoire. Actuellement, il n'existe pas une mesure validée pour le rétablissement postopératoire après une résection pulmonaire. L'EuroQol-5 dimensions (EQ-5D) est une mesure générique et standardisée de la qualité de vie liée à la santé, utilisée à des fins d'évaluation clinique et économique dans une grande variété de conditions médicales. Par conséquent, nous étudions la validité longitudinale et la validité conceptuelle, ou de construit, du EQ-5D comme mesure de rétablissement postopératoire après une résection pulmonaire élective pour des tumeurs potentiellement malignes.

Méthodologie: Les patients cédulés pour une résection pulmonaire ont complété le questionnaire du EQ-5D et des échelles visuelles analogiques (EVA) pour la douleur et la fatigue à l'état de base (préopératoire) et à un et trois mois après la chirurgie. Le EQ-5D inclut un profil de santé descriptif comprenant 5 dimensions (mobilité, soins personnels, activités usuelles, douleur/inconfort, anxiété/dépression) pondérées à partir d'ensembles de valorisation basés sur des populations générales pour fournir un index sommaire (l'étendu étant de 1 à -0.109). La validité longitudinale a été évaluée en testant l'hypothèse *a priori* que l'index de l'EQ-5D suit la trajectoire du rétablissement postopératoire e.g. diminue à un mois après la chirurgie par rapport à l'état de base, s'améliore d'un à trois mois et retourne aux niveaux de l'état de base à trois mois. La validité conceptuelle discriminatoire a été évaluée en testant les hypothèses *a priori* que les index du EQ-5D seraient plus bas après la chirurgie pour les patients âgés de plus de 70 ans par rapport aux patients plus jeunes, pour les patients opérés par une approche chirurgicale ouverte (thoracotomie) par rapport aux patients opérés par VATS, et pour les patients qui ont développé des complications par rapport à ceux qui n'ont pas eu de complications. La validité conceptuelle de convergence a été évaluée en testant l'hypothèse *a priori* que les index du EQ-5D auraient une corrélation inversement proportionnelle avec les niveaux de douleur et de fatigue (corrélation de Spearman; *r*). Les valeurs manquantes (5%) ont été substituées avec des imputations multiples.

Résultats: Les résultats de cinquante-cinq patients ont été analysés (45% hommes, 62±12 ans, 29% assistés par vidéo). Il n'y avait pas de différence significative entre les médianes des index du EQ-5D à

l'état de base (0.83 [IQR 0.80-1]) par rapport à un mois (0.83 [0.80-1], p=0.86) et trois mois après la chirurgie (1 [0.83-1]; p=0.09). À un mois après la chirurgie, les index du EQ-5D étaient significativement plus bas pour les patients opérés par thoracotomie par rapport aux patients opérés par approche minimalement invasive assistée par vidéo (0.82 [IQR 0.77-0.89] vs. 1 [0.83-1], p=0.003), mais il n'y avait pas de différence significative entre les patients âgés de plus de 70 ans par rapport aux plus jeunes (0.95 [IQR 0.82-1] vs. 0.83 [0.77-1], p=0.09) ou entre les patients avec complications par rapport aux patients sans complication (0.82 [IQR 0.79-0.95] vs. 0.83 [0.80-1], p=0.10). Une corrélation faible mais significative a été observée entre les index du EQ-5D et les scores EVA de douleur et de fatigue (Rho -0.30 to -0.47, $p \leq 0.01$).

Conclusion: Malgré certaines preuves de validité de convergence, le EQ-5D n'était pas sensitif à la trajectoire de rétablissement postopératoire et a montré une validité discriminante limite. Ainsi, le EQ-5D ne serait pas une mesure valide pour le rétablissement postopératoire un mois après résection pulmonaire et devrait être utilisé avec prudence dans l'évaluation économique de technologies chirurgicales visant à améliorer le rétablissement après le congé d'hôpital.

PREFACE AND CONTRIBUTION OF AUTHORS

Overview of research work leading to my thesis topic:

As a Master's student with the Surgical Recovery (SURE) Team at the McGill University Health Centre, my research projects were focused on postoperative recovery. First, I examined the impact of preoperative patient education as a component of an enhanced recovery pathway (ERP) after laparoscopic cholecystectomy. Patients reported high levels of satisfaction with a new educational booklet but it did not reduce patient anxiety compared to patients cared for prior to development of the booklet. This work was presented at the 2013 Annual Meeting of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) in Baltimore. The poster is referenced as: *P168 - **Bejjani J**, Watson D, Capretti G, Kaneva P, Fried GM, Vassiliou MC, Carli F, Feldman LS. Development of a patient educational resource for laparoscopic cholecystectomy: An important component of an Enhanced Recovery After Surgery Pathway* (URL: http://bit.ly/erpsages).

We conducted a systematic review to summarize the evidence regarding the impact of ERPs on clinical and patient-reported outcomes in elective pulmonary resection. One randomized controlled trial (RCT) and five observational studies fulfilled our selection criteria. All the observational studies reported shorter LOS with the use of ERPs, but the RCT reported no difference. However, patients treated with ERPs have had lower rates of pulmonary complications in the RCT and reduced hospitalization costs in two observational studies. These results should be interpreted with caution due to high risk of bias and the limited number of studies. This review is mentioned in section 2.2.2 on recovery after lung resection, and is published in the Journal of Thoracic and Cardiovascular Surgery. *Fiore JF, **Bejjani J**, Conrad K, Niculseanu P, Landry T, Lee L, Mulder DS, Ferri LE, Feldman LS. Systematic review of the influence of enhanced recovery pathways in elective lung resection. J Thorac Cardiovasc Surg. 2015 Oct 3. pii: S0022-5223(15)01821-8.*

Meanwhile, the implementation of an ERP for elective pulmonary resection was ongoing at our hospital. A prospective study to examine its impact on postoperative outcomes was performed. Our study concluded that the pathway was associated with reductions in hospital length of stay (LOS), chest tube duration and overall postoperative complication rate after pulmonary resection. This was published in Surgery: *Madani A, Fiore JF Jr, Wang Y,* ***Bejjani J,*** *Sivakumaran L, Mata J, Watson D, Carli F, Mulder DS, Sirois C, Ferri LE, Feldman LS* *An enhanced recovery pathway reduces duration of stay and complications after open pulmonary lobectomy. Surgery. 2015 Oct;158(4):899-910* (URL: http://bit.ly/erplung).

In order to perform a comprehensive cost-effectiveness analysis of the pulmonary resection ERP from the societal perspective, we chose the widely-used EuroQol - 5 dimensions (EQ-5D) instrument to assess patient-reported quality of life after surgery, and for its potential to generate quality-adjusted life years (QALYs) for the cost-effectiveness analysis of the ERP. At one month after surgery, we noticed that the proportion of patients reporting no problem with anxiety/depression has decreased, compared to the increased proportion of problems reported with the other dimensions. Given these opposite trends, we examined whether the EQ-5D is a valid postoperative recovery measure after elective pulmonary resection. The study was published in Surgical Research: ***Bejjani J****, Fiore FJ, Lee L, Kaneva P, Mata J, Ncuti A, Sirois C, Mulder DS, Ferri LE, Feldman LS. Validity of the EuroQol-5 dimensions as a measure of recovery after pulmonary resection. J Surg Res, 194 (2015), 281-88* (URL: http://bit.ly/EQ5Dlung).

Upon the recommendation of my research supervisor, I have chosen to prepare my thesis on the EQ-5D study as I was the first author of the manuscript. This thesis has therefore been prepared in a manuscript-based format, as an alternative to the traditional thesis format, approved by McGill University as stated in the guidelines on the preparation of a thesis listed below.

"As an alternative to the traditional thesis format, the thesis research may be presented as a collection of scholarly papers of which the student is the ***author or co-author****; that is,* ***it can include the text of one*** *or more manuscripts, submitted or to be submitted for publication, and/or published articles reformatted according to thesis requirements as described below. Manuscripts for publication are frequently very concise documents.* ***The thesis is expected to be a more detailed, scholarly work*** *than manuscripts for publication in journals, and must conform to general thesis requirements. Note: These papers cannot alone constitute the thesis;*

The thesis must contain additional text that will connect them, producing a cohesive, unitary focus, and documenting a single program of research. A Manuscript- (or Article-) based thesis will be judged by the examiners as a unified, logically-coherent document in the same way a traditional thesis is judged.

The structure for the manuscript-based thesis must conform to the following:

- *Just as in the traditional format, the thesis must be presented as a unified whole with respect to font size, line spacing and margin sizes (see thesis format).*

- *The thesis must conform to all other requirements listed under thesis components above.*

- *The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next, providing a cohesive, unitary focus, documenting a single program of research. Connecting text must be provided so that the completed thesis functions as an integrated whole.*

There is no specified number of manuscripts or articles required for a Master's or a Doctoral thesis, nor is prior publication or acceptance for publication of the manuscripts a requirement. Publication or acceptance for publication of research results before presentation of the thesis in no way supersedes the University's evaluation and judgment of the work during the thesis examination process (i.e., it does not guarantee that the thesis will be found acceptable for the degree)."

Given the nature of a thesis based on a single manuscript, the requirements to include a thesis introduction, objectives/hypotheses and conclusion, separately from these same sections within the manuscript, inevitably lead to some degree of repetition. An attempt is however made to minimize such.

Contributions of authors:

Jimmy Bejjani (JB) wrote this thesis and was the principal author of the manuscript included in this thesis. JB was directly involved in the conception and design of the study, acquisition, analysis and interpretation of data, and drafted the manuscript.

Liane S. Feldman (LSF) was the thesis supervisor and primary investigator of the research program. LSF is co-author on the manuscript and contributed significantly in the research process. LSF was involved in the conception and design of the study, interpretation of data, and revised the manuscript.

Julio F Fiore contributed to the conception and design of the study, analysis and interpretation of data, and co-drafted the manuscript;

Pepa Kaneva, **Juan Mata** and **Annie Ncuti** contributed to the conception and design of the study, interpretation of data and revised the manuscript;

Lawrence Lee contributed to the conception and design of the study, interpretation of data and revised the manuscript;

Christian Sirois, **David S. Mulder** and **Lorenzo E. Ferri** performed the lung resections, and contributed to the conception and design of the study, interpretation of data and revised the manuscript.

STATEMENT OF SUPPORT

The salary support for Jimmy Bejjani was provided by the McGill Surgeon Scientist Program. The Steinberg-Bernstein Centre for Minimally-Invasive Surgery and Innovation is supported by an unrestricted education grant from Covidien Canada.

1 - INTRODUCTION

Lung cancer is among the three most common cancers in males and females, and the leading cause of cancer death[1]. An elective resection may be performed in patients who are fit for surgery and whose tumors are diagnosed at an early stage. Beside common postoperative complications, recovery from lung resection may be particularly complicated by air leaks and pain control[2]. Fortunately, advances in anesthesia and surgical care have improved outcomes after lung resection. The development of anesthetics with fast pharmacokinetic properties and minimal side effects have decreased the incidence of postoperative nausea and vomiting. Enhanced analgesic regimens, and better knowledge of pain physiology and management have improved pain control. The advent of minimally-invasive procedures (ie. video-assisted thoracic surgery) and enhanced-recovery pathways (ERPs), also known as fast-track surgery, further reduce the surgical stress response and hasten recovery. ERPs are standardized perioperative care pathways combining evidence-based interventions that reduce surgical stress and support more rapid return to normal functioning[3]. Such interventions include reduced preoperative fasting, multimodal analgesia including intraoperative thoracic epidural, and earlier postoperative feeding, ambulation, and catheter removal. Initially developed for colorectal surgery, ERPs have shown benefits in reducing the overall rate of complications and hospital length of stay (LOS) in a Cochrane systematic review[4]. They have been subsequently implemented across surgical specialties, and are associated with decreased postoperative morbidity and shorter LOS without increasing the risk of major complications or readmissions[5].

While these traditional outcomes measures are important in evaluating postoperative recovery, patients also expect to recover their health and quality of life (QoL) and not just survive the operation[6]. Recovery is a complex construct, encompassing multiple dimensions including symptoms, functional status and QoL, with no standard definition or measurement tool. Patients equate recovery to the absence of symptoms and the return of their ability to perform activities as they could before their

operation[7]. The process of recovery can be modeled as a rapid, transient decline in health status immediately after surgery, followed by a more gradual return to baseline or population norms[8]. The role of patient-reported outcomes (PRO), reports coming directly from patients on their health status, to measure of post-operative recovery is particularly important when evaluating the overall cost-effectiveness of care pathways, as complete recovery in these domains takes months, and is certainly not complete upon hospital discharge. Evaluation of cost-effectiveness for new interventions such as ERPs is important as they are associated with implementation and maintenance costs, but may provide societal benefits in the longer-term.

The Washington Panel on Cost-Effectiveness in Health and Medicine and the UK National Institute of Health and Clinical Excellence (NICE) recommend that economical evaluations of health interventions be expressed in terms of quality adjusted life years (QALYs)[9,10]. A QALY is the arithmetic product of the life expectancy after an intervention and the QoL during these remaining life years. It is a summary measure used as a common 'currency' to compare the health outcomes resulting from competing interventions, while considering both the quantity (years) and quality of life. Should an intervention results in a better QoL than another, patients are likely to resume daily life activities and contribute to society earlier. In being a qualitative representation of the life years gained, compared to life expectancy alone, QALYs are used in cost-effectiveness analyses to guide resource-allocation decisions. Valuing technologies aimed to enhance postoperative recovery in terms of QALY requires preference-based measures that reflect the construct of postoperative recovery.

The EQ-5D is a generic preference-based measure used to generate QALYs[11]. It is the preferred instrument of NICE for estimating the impact of different technologies on QALY[10]. EQ-5Dis a short and cognitively simple questionnaire that makes it attractive for use in research. However, a potential disadvantage of EQ-5D is the lack of sensitivity to detect disease-specific effects of interventions[12]. Research also suggests that it has a ceiling effect limiting its ability to measure health status in

populations with low levels of disability[13,14]. Other instruments used to measure HRQoL, such as the EORTC-QOL-C30, may be more sensitive and specific for assessment of recovery in lung cancer populations[15]. However, they do not generate QALYs for cost-effectiveness analyses. Considering the widespread use of EQ-5D in economic appraisals, and the need for a QALY measure to evaluate ERPs, we evaluated its validity as a measure of recovery after elective lung resection.

2 – LITERATURE REVIEW

2.1 – Lung Cancer and Surgical Treatment

Lung cancer is a malignant growth of abnormal tissue in the respiratory tract. In Canada, it is the leading cause of cancer death[1]. Its broad range of clinical manifestations may negatively affect patients' quality of life (QoL) and limit their physical, mental and functional capacity[16,17].

Factors that determine the treatment of a tumor include its staging, histology and genetics, as well as patient comorbidities and adequacy of lung reserve. Staging based on the current 7th edition of the lung cancer TNM system[18] provides a common language to describe the magnitude of a tumor, and its potential spread beyond the original site, in order to determine the therapy and prognosis. Most common therapeutic options for lung cancer include surgery, radiation and chemotherapy, or combinations of these modalities, to obtain a microscopically margin-negative resection[19,20]. Patients with a high risk of malignancy based on previous history of smoking or cancer, older age, upper lobe location, and nodule larger than 30 mm or with abnormal edge characteristics should undergo a definitive surgical biopsy or resection.

Non-small cell lung cancer (NSCLC) is the only lung cancer subtype amenable to surgical resection. Surgery is the treatment of choice for stage I to IIIA NSCLC tumors, with systematic sampling of mediastinal lymph node to accurately stage the surgically resected lung cancer, or with complete ipsilateral mediastinal lymph node dissection (CMLND) when possible[21]. The resection is performed with a lobectomy, which is the surgical resection of one of the five anatomical components (lobes) of the lung. A wedge resection (segmentectomy) is an alternative to a lobectomy and involves the removal of a wedge-shaped piece of lung that contains the tumor surrounded by a margin of healthy lung parenchyma. Lobectomy by video-assisted thoracoscopic surgery (VATS) is a minimally invasive approach that is oncologically equivalent to open lobectomy (thoracotomy). Either approach allows the

removal of the tumor-containing lobe and the mediastinal lymph node dissection for early-stage NSCLC, with equal safety, local oncological control and survival[22]. The open approach is more suited to large (>3 cm) tumors or tumors located close to the major blood vessels or airways. The use of VATS is also limited in patients with neoadjuvant radiation therapy, parietal pleural adhesions (in stage II) or prior chest surgery.

2.2 – Surgical Recovery

2.2.1 – Definition of Surgical Recovery

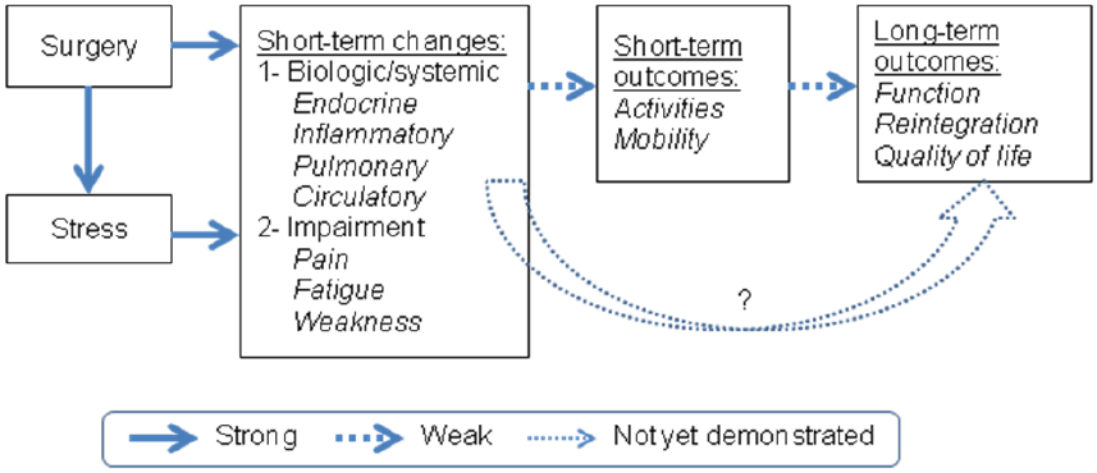
The concept of surgical recovery or postoperative "convalescence" has been used as a postoperative outcome since the 1980s without standard definition. When limited to measurement of clinical outcomes, complete assessment of recovery is incomplete. While we reported shorter LOS and reduced complication rate after elective lung surgery with an ERP[23], these outcomes do not reflect the impact of the ERP from the patients' perspective.

In a literature review of postoperative recovery, Allvin et al. describe surgical recovery as a complex energy-requiring process to return to the preoperative level of activities of daily living, and optimum level of psychological well-being, achieved by regaining control over four functions: physiological, psychological, social and habitual[24]. Hence, comprehensive assessment requires assessment of the multiple dimensions of recovery.

Carli and Mayo proposed a biological model to describe the hierarchical pathway relating these dimensions in the process of recovery[25] (Figure 1). It is suggested that the immediate short-term changes stemming from the surgical stress response result in biologic effects (hormonal and inflammatory changes, and organ dysfunction) that are manifested as impairments (pain, fatigue, lack of energy) that then impact on short-term outcomes (such as activities of daily living or walking) and

then longer-term outcomes (functional status, return to baseline leisure or economic activities, QoL). These outcomes can each be measured to obtain a fuller assessment of the (negative) impact of the perioperative period on patients and compare outcomes between new techniques purported to improve patient recovery.

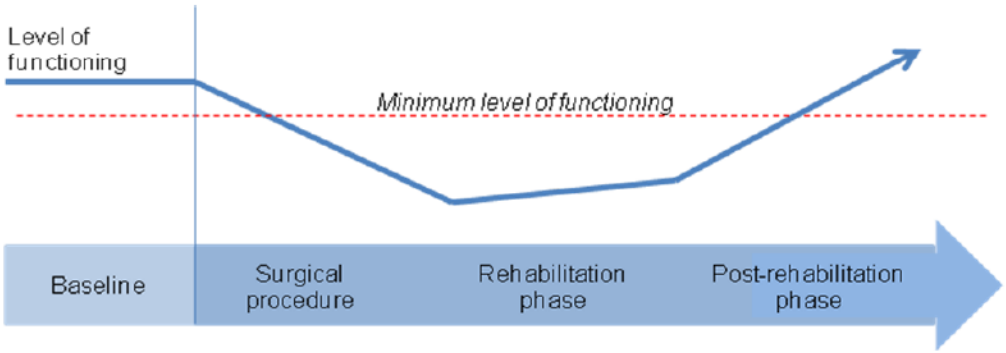
Figure 1 - Carli and Mayo's model for measuring the outcomes of surgical procedures



Source: Carli F, Mayo N. Measuring the outcome of surgical procedures: what are the challenges? British journal of anaesthesia 2001;87:531-3.

To quantify recovery (implying a "return" to baseline), it is necessary to define patient status prior to surgery. A theoretical model of the trajectory of functional ability throughout the surgical process was suggested by Carli and Zavorsky[26] (Figure 2). Before surgery, patients start at a baseline functional status determined by factors such as fitness level, comorbidities and disease burden. A rapid, transient decline in health status occurs with the surgery, followed by a gradual rehabilitation to (or above) baseline levels of health. The extent of the decline and the slope of the recovery phase may be influenced by the nature of the surgery, surgical stress response, patient comorbidities, postoperative complications and the perioperative care, suggesting several potential strategies to improve recovery.

Figure 2 - Carli and Zavorsky's model of trajectory of recovery



Source: Carli F, Zavorsky GS. Optimizing functional exercise capacity in the elderly surgical population. Current opinion in clinical nutrition and metabolic care 2005;8:23-32.

Lee et al. further divide recovery into three distinct phases, early, intermediate and late[27], and recommend outcomes that may be measured in each.. The early phase, from the operating room to discharge from the post-anesthesia care unit (PACU), can be assessed using physiologic and biologic outcomes. The intermediate phase, from PACU to discharge from hospital, can be evaluated using symptoms scores and impairments in activities of daily living. Finally, the late phase of recovery, from hospital discharge to return to usual function and activities, is best evaluated using measures of function status and HRQoL.

2.2.2 – Recovery after Lung Resection

Innovations in thoracic surgery and anesthesia have improved recovery after lung resection. The use of VATS has reduced soft tissue trauma and pain, leading to quicker rehabilitation[28]. Extubation in the OR has facilitated patients' discharge from the PACU to their rooms, avoiding intensive-care stays. However, postoperative morbidity is still common after lung resection and is reported to occur at a rate up to 40%[29]. Complications impair recovery, increase LOS[30], and delay return to regular activities and to the expected postoperative quality of life[31].

Prolonged postoperative recovery imposes a significant economic burden in an era of budget

constraints[32]. Improving perioperative care is therefore not only desirable for patients, but also for the efficient utilization of healthcare resources. There has been increased interest in the use of multimodal care plans to hasten recovery, reduce morbidity and facilitate early hospital discharge.[33] This concept of care is often referred to as "fast-track", Enhanced Recovery After Surgery (ERAS) or "enhanced recovery program" (ERP). ERPs combine multiple evidence-based elements encompassing all phases of perioperative care[34]. These care elements may have modest benefits when used alone, but are believed to have a synergistic effect to attenuate surgical stress and postoperative organ dysfunction, thereby reducing complications and facilitating postoperative recovery.[33]

Improving recovery begins preoperatively with patient education to increase engagement with accelerated recovery, decreased preoperative fasting and use of prophylactic antibiotics. Intra-operatively, use of an epidural and multimodal analgesia, the use of VATS or a muscle-sparing surgery and the prevention of hypothermia are beneficial. After surgery, early patient mobilization and removal of tubes, catheters, IVs and oxygen support, multimodal analgesia allowing earlier removal of the epidural, and early resumption of oral intake are emphasized[35]. Some authors also advocate for outpatient treatment of air leaks with the use of a Heimlich valve[36,37]. Daily reinforcement with patients and their families of the planned events for each day contribute to better prepare them for early discharge[38].

The scientific rationale for those elements is that a quicker recovery of muscle strength, pulmonary function and tissue oxygenation, and more adequate nutrition result in decreased general morbidity. As well, the pulmonary condition of patients asked to ambulate the same day after surgery tend to be better than for those who start walking the next day[39]. In addition, physiotherapy after lung resection reduces the incidence of postoperative complications and improves lung function[40]. Earlier dietary intake after surgery improves the recovery of gastrointestinal function and can be expected to help recover immunity against bacterial infection after surgery[39].

A systematic review by our research group identified only six studies (one RCT and 5 observational) investigating the impact of an ERP for pulmonary resection [35]. In comparison to traditional care, lower rates of pulmonary complications and reduced hospitalization costs were shown in one RCT[41] and two observational studies[42,43], respectively. Additionally, no study reported any difference in readmission, mortality or overall complication rates. Nevertheless, there is an overall high risk of bias and a limited number of studies comparing traditional care to ERPs with lung resection. In addition, there was significant heterogeneity in the number of ERP elements used by each study.

Prior to implementation of our 4-day multidisciplinary ERP for lung wedge resection and lobectomy, median hospital stay after major lung resection was 7 days. Table 6 outlines the key interventions and milestones of our ERP. During the preoperative testing day, a detailed booklet about lung surgery and the daily ERP milestones is given to patients as part of the teaching and counseling (link: http://bit.ly/GuideLungSurg). It is designed to be an important reference to perioperative care with text at at an appropriate health literacy level, the use of pictures and alignment with the rest of the pathway. Posters are also displayed on the wards to remind patients about the importance of breathing exercises and of daily milestones for ambulation, pain control, nutrition, and removal of tubes and intravenous lines. The pathway was the standard of care, with standard orders initiating each postoperative intervention unless changed by the physician. The pathway specifies urinary catheter removal on POD 1, which is prior to removing the thoracic epidural, as a previous trial demonstrated that this can be accomplished safely when with use of a bladder-scan based urinary retention protocol is also implemented [44]. After a literature review, the ERP steering committee came to consensus on a threshold of 300 cc per 24h for chest tube removal, since we identified little data supporting one threshold over another, and readmission because of recurrent effusions using thresholds as high as 500 cc per 24h are uncommon[45,46]. The target discharge day was postoperative day (POD) 3 for patients with 1 chest tube and POD 4 for patients with 2 chest tubes, provided they met predefined discharge

criteria: pain ≤ 4/10, ambulating well, voiding adequately, diet well tolerated, normal vital signs and wound healing well. Removal of the urinary catheter on POD 1 and removal of the last chest tube on or before POD 3 were independent predictors of decreased LOS[23].

Table 6 - Key interventions & milestones - MUHC ERP for lung wedge resection and lobectomy

Preoperative phase		
- Standardized teaching and counseling to patient/family Review/discuss: prescriptions, inspirometer use, clinical care pathway, discharge plans, expected hospital length of stay; - Educational booklet with daily goals; - Shorter preoperative fasting Clear Fluids until 2h before admission - Protocolized prophylactic antibiotics; - Antiembolic stockings on call to operating room.		
Intraoperative phase		
- Thoracic epidural analgesia; - Prevention of hypothermia using active warming; - Extubation in the operating room or in the post-anesthesia care unit.		
Postoperative phase		
- Spirometry 10 times per hour while awake and chest physiotherapy every 4 hours.		
POD 0	Chest tube	Suction at -20 cmH_2O.
	Diet	Clear Fluids
	Mobilization	Up in chair with assistance as tolerated.
POD 1	Chest tube	Remove suction, follow by Chest X-ray.
	Diet	Diet as tolerated - N/S lock IV if tolerating diet.
	Mobilization	Up in chair 3 times per day for all meals + 30-60 minutes each time, ambulate in hallway 2 times per day with assistance. Remove antiembolic stockings when fully ambulating.
	O_2 support	Remove if SpO_2 on room air is > 92% or at preoperative baseline
	Urinary catheter	Remove if urine output is adequate (≥200cc/8h) - Bladder scan and urinary retention protocol if no urine output 8 hours after removal
POD 2	Chest tube	Remove #1 if <300cc/24hrs, non-chylous and no air leak. Follow by Chest X-ray.
	Mobilization	Out of bed for all meals and at least 8 hours during the day, walking

		in hallway 17.5-35 meters 3 times per day with assistance.
POD 3	Chest tube	Remove #2 if <300cc/24hrs, non-chylous and no air leak. Follow by Chest X-ray.
	Mobilization	Increase ambulation to 75 meters 3-5 times per day.
	Pain control	Epidural stop test on the day the last chest tube is removed.
	Discharge home	If patient had 1 chest tube and meets discharge criteria.
POD 4	Discharge home	If patient had 2 chest tubes and meets discharge criteria.

POD: postoperative day; N/S: Normal saline; SpO_2: peripheral capillary oxygen saturation

Patient-reported outcomes (PRO) add valuable information to clinical indicators of surgical recovery and allow us to calculate QALYs and the cost-utility of ERPs. PRO are obtained from HRQoL measures and assess recovery in greater depth as they relate to symptoms (e.g. pain, fatigue, nausea/vomiting, anxiety/depression), functional status (e.g. mobilization, return to work, cognitive function), health perceptions and quality of life.

2.3 – Measures of Health-Related Quality of Life (HRQoL)

2.3.1 – Definition of HRQoL

The World Health Organization (WHO) defines health as a *state of complete physical, mental and social well-being and not only the absence of disease and infirmity*[47]. Quality of life (QoL) is defined as a multidimensional dynamic construct which is influenced by variables such as stress, depression, cognitive appraisal, and coping[48]. The WHO adds that QoL is influenced by the cultural context in which individuals live and in relation to their goals, expectations, standards and concerns[49].

The concept of HRQoL describes the subjective perception of an individual regarding his physical, mental and social functions, and the impact of disease and treatment on his ability to live a fulfilling life[50]. The physical, psychological and social core domains are what distinguish HRQoL measures from other measures capturing some of these constructs[51]. Along with somatic sensations, these factors are

the most contributive to HRQoL and are impacted on by cancer and surgery. Patrick and Erickson's definition of HRQoL 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*"[52].

Physicians have traditionally relied on clinical assessments, laboratory values and clinical outcomes to determine patients' recovery status after surgery. Yet clinical outcomes do not capture daily functioning problems and the resulting burden of care on patients and their care-givers, which are captured by PROs. HRQoL in patients after lung resection is influenced by higher symptom burden, limitations in physical functioning and clinical depression, anxiety and stress[53,54]. The International Classification of Functioning, Disability and Health (ICF) describes thoroughly the impact of cancer on physical function and subsequently HRQoL[55]. Cancer patients are also frequently more concerned about QoL and disability than about longevity[56].

Therefore, when assessing recovery, HRQoL endpoints should be documented to supplement traditional clinical outcomes. HRQoL can be assessed with several generic and disease-specific instruments. Both types have different advantages and disadvantages, and there is no clear recommendation on which PRO is most appropriate for recovery.

2.3.2 – Generic Measures of HRQoL

Generic measures are used to compare HRQoL between different population groups. Most of them rely on self-rated assessments based on the patient's perception of his/her capabilities. There is evidence for validatity, reliability and applicability in many different diseases, treatments or interventions, and populations[57]. They are straightforward to administer and are available in many languages. They can be used clinically and in population health and health services research.

Generic measures are generally composed of a descriptive profile and of a summary index. The first reports the abilities and limitations in several health domains, and the second reports the overall health status as one summary index[58]. While profiles capture more extensive descriptive information, they do not balance gains and losses between dimensions on which they are based. This limitation can be an important constraint if change in health status is the desired objective. However, gains and losses between domains can be balanced with the summary index which reflect health with the concept of health utility or preference as described in 2.3.5. A utility represents the quality of the remaining life-years. It is the strength of preference a respondent has for a given health state, with a single cardinal value between 0 (death) and 1 (full health)[59]. Utilities can in turn be used for the calculation of QALYs, a key measure used in cost effectiveness analyses. QALYs are calculated by multiplying the duration of time spent in a health state (quantity) by the utility score associated with that health state (quality).

The two extreme response options in a domain of a given generic measure (e.g. no problems vs extreme problems in the EQ-5D) discriminate appropriately between health states. However, patients may not be able to discriminate between the middle options and may choose a response arbitrarily[60]. Furthermore, although most PROs have proven to be valid and reliable indices, patients may have different perceptions of the questions being asked. The evaluation of one's own health and QoL is actually related to one's "goals, expectations, standards, concerns and symptoms" at that particular moment, all of which may also change over the course of time[49]. The resultant variability in an individual's self-evaluation can result from a concept known as 'response shift' phenomenon.[61]. This concept is related to three factors: (1) scale recalibration, i.e. change in respondent's internal standards, (2) change in respondent's values or priorities, i.e. the importance of domains constituting the construct of recovery, or (3) reconceptualization, i.e. redefinition of the construct of recovery.

For example, (1) a lung cancer patient with symptoms of anxiety before his lung resection may have evaluated his HRQoL as good. After uncomplicated surgery, his anxiety may be better controlled and,

looking back, he may now evaluate his initial HRQoL as poor in comparison, and still rate his present health as good. (2) Should an elderly patient's physical function not improve after surgery, he may reconsider goals that were important and reachable, such as emotional and spiritual experiences, which results in a change in his values. He may now evaluate his health state as better, based on emotional and not physical function. (3) Finally, a patient's overall definition of recovery may change as he learns that it is possible to have a reasonable QoL with a worse condition. This variability in response is normal during an illness. Since the resultant change in health may no longer be related to the effects of the treatment only, it could become difficult to evaluate the reason behind the change in patients' perceived health.

While generic HRQoL allows for comparisons between diseases, a limitation is that they may not address particular aspects relevant to HRQoL in specific conditions such as lung cancer, and may consequently provide an incomplete picture on how a condition affects a patient's QoL. This has lead to the development of disease-specific instruments.

2.3.3 – Disease-Specific Measures of HRQoL

In comparison to generic measures, disease-specific measures do not compare HRQoL across different patient populations. Their primary focus is QoL aspects considered to be most relevant to a particular group of patients with a specific condition, such as lung cancer.

They provide a more complete picture of the impact of the condition on QoL. Consequently, disease-specific measures usually have a better sensitivity to clinically important changes in specific conditions[57,62]. Studies have found that these measures also tend to be more responsive compared to generic measures[63,64]. This can be explained by the greater depth of the description of domains in disease-specific measures for a particular group of patients. They may relate more closely with traditional clinical measures. As well, disease-specific instruments tend to be more intuitive and

relevant to clinicians and patients. Nonetheless, their scope may be too narrow to measure unintended changes and their range of applicability may be limited.

2.3.4 – Patient-Reported Outcomes (PRO) Used in Surgical Recovery

There is no single PRO instrument designed and validated using modern psychometric techniques to measure the multidimensional construct of surgical recovery after hospital discharge. However, several generic PRO measures have been used to measure surgical recovery. Based on a broad review of the literature, table 7 summarizes the dimensions and psychometric properties of common measures that have been used in this setting. In a systematic review of outcomes used to evaluate recovery in ERPs after abdominal surgery, the EQ-5D and SF-36 (through the SF-6D utility index) were the only generic measures that can be used to compute QALYs[65]. Besides, among generic PRO measures, the Health Utilities Index (HUI) and the Quality of Well Being scale (QWB) are found to be used in surgical recovery in only one study for each[66,67].

However, generic HRQoL instruments may both omit important concepts for surgical recovery and cover concepts that are not relevant to this construct[68] and there is a need for specific measures. Systematic reviews on instruments specifically designed to measure surgical recovery identified a total of 16 PRO measures[69,70]. None of them was fully validated but two instruments demonstrate superior psychometric properties and were recommended for future studies - i.e. the Postdischarge surgical recovery (PSR) scale and 40-item Quality of recovery (QoR-40) - but these cannot be used to calculate QALYs.

Systematic reviews have also looked at PRO measures used to evaluate lung cancer therapies, but have not been validated in the context of surgical recovery[71,72]. Nonetheless, EQ-5D and SF-6D remain the recommended measures in studies aiming for cost-effectiveness assessments.

EQ-5D has been widely used with over 120 different medical conditions in clinical trials, observational studies and other health surveys[73,74]. Indeed, it can be used directly for cost-utility analyses, while it is necessary to first derive SF-36 health states to SF-6D in order to generate a preference-based index. As well, EQ-5D has been increasingly used to evaluate postoperative recovery. For instance, the search platform of the EuroQol group lists over 60 manuscripts reporting EQ-5D results in this specific context since 2010 (www.euroqol.org - tab: "EQ-5D References" - search keywords: "postoperative recovery").

Table 7 - Common generic PROs used to report surgical recovery

Instrument	Domains (# items)	Psychometric properties	Examples in surgical recovery
EuroQol 5-dimensions (EQ-5D)	- Mobility (3) - Self-Care (3) - Usual Activity (3) - Pain/Discomfort (3) - Anxiety/Depression (3) - VAS Global health status	Reliability: test-retest[75], internal consistency[76] Validity[75]: construct, (convergent, discriminant), concurrent Responsiveness[75]	RCT on ERP vs. standard care following open liver resection[77]
Medical Outcome Survey 36-Item Short Form Health Questionnaire (SF-36)	Physical Health Subscale: - Physical Functioning (10) - Role-Physical (4) - Bodily Pain (2) - General Health (5) Mental Health Subscale: - Vitality (4) - Social Functioning (2) - Role-Emotional (3) - Mental Health (5)	Reliability[75]: internal consistency, test- retest, inter-rater[78] Validity[75]: content[79], criterion, construct (convergent, discriminant), predictive Responsiveness[75]	RCT of controlled rehabilitation with early ambulation and diet vs. traditional care following radical total gastrectomy[80]
Short Form SF-12 Health Survey	Physical Health Subscale: - Physical Functioning (2) - Role-Physical (2) - Bodily Pain (1) - General Health (1) Mental Health Subscale: - Vitality (1) - Social Functioning (1) - Role-Emotional (2) - Mental Health (2)	Reliability[75]: test-retest, internal consistency Validity[75]: construct (convergent, discriminant)	Emotional and physical recovery after coronary bypass surgery[81]

Health Utilities Index (HUI)	HUI Mark 2: - Sensation (4) - Mobility (5) - Emotion (5) - Cognitive (4) - Self-care (4) - Pain (5) - Fertility (3) HUI Mark 3: - Vision (6) - Hearing (6) - Speech (5) - Ambulation (6) - Dexterity (6) - Emotion (5) - Cognition (6) - Pain (5)	Reliability[75]: test-retest, inter-rater Validity[75]: construct (convergent, discriminant), concurrent Responsiveness[75]	Responsiveness of HUI in osteoarthritis patients undergoing total hip arthroplasty[66]
Quality of Well-Being scale (QWB)	- Mobility scale (3) - Physical activity scale (3) - Social activity scale (5) - Symptom and problem complexes (27)	Reliability: test-retest[82], inter-rater[75] Validity[75]: content, construct (convergent)	QWB vs. SF-36 in patients with total hip or knee replacement[67]
Spitzer Quality of Life Index (QLI)	- Health and functioning (13) - Psychological/Spiritual (7) - Social and Economic (8) - Family (5)	Reliability[75]: test-retest, internal consistency, inter-rater Validity[75]: content, construct (convergent, discriminant) Responsiveness[83]	RCT of laparoscopy with ERP for colonic surgery[84]

Nottingham Health Profile (NHP)	- Physical mobility (8) - Pain (8) - Social isolation (5) - Emotional reaction (9) - Energy (3) - Sleep (5)	Reliability[75]: test-retest, internal consistency Validity[75]: construct (convergent, discriminant) content concurrent Responsiveness[75]	Criterion validation of NPH before/after transurethral resection of the prostate[85] Comparison of NHP and SF-36 before/after cardiac surgery[86]
Sickness Impact Profile (SIP)	Physical (45): - Ambulation (12) - Mobility (10) - Body Care and Movement (23) Psychosocial (48) - Social Interaction (20) - Communication (9) - Alertness Behavior (10) - Emotional Behavior (9) Independent categories: - Sleep and Rest (7) - Eating (9) - Home Management (10) - Recreation and Pastimes (8) - Employment (9)	Reliability[75]: internal consistency, test- retest, inter-rater. Validity[75]: construct (convergent, discriminant) Responsiveness[75]	SIP vs. SF-36 in prolonged surgical intensive-care stay[87]

2.3.4.1 – Description of the EQ-5D

EuroQol-5 dimensions is a HRQoL instrument designed for self-completion by respondents (appendix 8.1). It is quick and cognitively simple with instructions included in the questionnaire[88]. It is suitable for use in postal surveys, in clinics and face-to-face interviews, and has been used in the general population and many different illness populations[12]. It has become one of the most widely used generic measures of health in Europe and in health economic evaluation. The EQ-5D is designed to be used as a complement alongside other HRQoL measures, and to facilitate the collection of a common dataset for reference purposes[11,88,89].

The EQ-5D is comprised of two parts: a descriptive health state system and a visual analog scale (VAS). The first part tests the domains of physical, mental and social functioning by emphasizing five dimensions: mobility, self-care, usual activities, pain/discomfort, as well as anxiety /depression. On the EQ-5D-3L, they each have three levels (1: no problem; 2: moderate problems; 3: extreme problems) which defines 243 health states. The second part assesses self-perception of overall health on a vertical 20-cm VAS, labeled from the best imaginable health state (100) to the worst imaginable health state (0). Patients draw a line on the scale ("thermometer") to indicate how good or bad they perceive their health state to be.

The EQ-5D provides utilities as an indirect measurement method. Patients complete the descriptive profile of the EQ-5D questionnaire to reflect their health state. The response to the questionnaire provides a sequence of 5 digits corresponding to levels (1, 2, 3) for each dimension in this specific order: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. This sequence of digits is applied in a valuation algorithm that calculates a utility (EQ-5D index). Finally, this utility is multiplied by the number of life years to obtain a QALY value.

Development and valuation of the EQ-5D:

The EQ-5D was developed by the EuroQol Group through a review of six existing HRQoL instruments[11].: the Quality of Well-Being Scale, the Rosser Index, the 15D, the Sickness Impact Profile, the Nottingham Health Profile and the Health Measurement Questionnaire. The EuroQol Group included dimensions that are wide in content, suitable for different health states and usable by the general population. It was based on a compromise between a comprehensive instrument with as many dimensions as other instruments, and a simple instrument to generate a small number of different health states.

In 1988, the first version consists of six dimensions (6D) with two to three levels for each dimension, generating 216 health states. These dimensions were *mobility*, *self-care*, *main activity*, *social relationships*, *pain* and *mood*. Levels were on an ordinal scale except for *self-care*, which was on a nominal scale (*no problems with self-care, unable to dress self*, and *unable to feed self*). A new version was ratified in 1990, including five dimensions: *mobility*, *self-care*, *usual activities*, *pain/discomfort* and *anxiety/depression*. All levels have been changed to an ordinal scale: *no problems*, *some* or *moderate problems*, and *extreme problems*. This latest version generates 243 health states, a limited number to enable the use of a comprehensive valuation method. A health state results from combining one level to each dimension.

In its development, direct methods were used to value health states defined by its descriptive system and generate index scores. Forty-five of the 243 possible health states were valued by respondents in the UK with the TTO (time trade-off) method, using 15 different health states (13 EuroQol states plus immediate death and unconscious). These 45 health states were used to create a regression model to estimate an index score for the remainder of the 243 health states. The algorithm provides 84/243 negative utilities corresponding to health states worse than death (WTD) and no utility between 0.883

and 1.0. The UK tariff (so-called A1 tariff) yields a range from -0.594 to 1[9]. Algorithms with different weights are available for different countries considering the variability in preference-based weights. As the complete Canadian valuation of EQ-5D health states is not yet available[90], our study used the US valuation algorithm, the so-called US D1 tariff[91]. The lowest utility score in the D1 tariff is -0.102[92].

2.3.4.2 – Description of Visual Analog Scales

A visual analog scale (VAS) (appendix 8.2 and 8.3) is a simple tool used to assess a subjective perception or feeling. Patients indicate their perceived degree of a symptom along a vertical or horizontal line, ranging from 0 to 10. The endpoint 0 defines the absence of the symptom, and 10 defines the symptom as unbearable. They are often used to assess the early period of recovery since they focus on symptom status, such as for anxiety, pain, fatigue and nausea. VAS for these symptoms may contribute to our understanding of postoperative recovery, as the degree of these symptoms influences this construct. VAS have been shown to be validate and reliable for surgical patients, and responsive to change with recovery[93].

They are also useful for frequent measurements in postoperative patients. VAS for pain has been extensively reported and is found to correlate with acute and chronic pain levels by several studies[94-96]. Several studies have found that VAS for anxiety correlates with the State-Trait Anxiety Inventory (STAI), a validated test quantifying state anxiety[97,98]. As well, the VAS method has proven to be useful in the assessment of postoperative nausea intensity and for testing the efficacy of medication[99]. In addition, VAS for fatigue has also been studied in various populations such patients with cancer or undergoing surgery[100,101].

2.3.5 – Health Utilities, Health Preferences and Quality-Adjusted Survival

In health economics, health utilities and health preferences are cardinal values that summarize

information from several health domains or attributes. In the strict sense, health utilities are elicited in conditions of uncertainty, while health preferences are elicited in conditions of certainty. These values are derived by valuation techniques such as Time Trade-Off (TTO), the Visual Analog Scale (VAS) or the Standard Gamble (SG). Utility/preference weights lie along an interval scale (0-1) with the same magnitude of change valued equally across the scale. Therefore, low values mean greater losses for younger individuals than for older individuals.

Utilities and preferences valuation methods:

Three main methods are used to value utilities and preferences, each having drawbacks and complexities. In general, health economists support the use of choice-based methods: SG or TTO. A review of utilities by Marimoto and Fukuis has found a strong tendency for VAS to yield the lowest, TTO the middle and SG the highest utility values for the same health states[102].

The SG is the classic method of measuring utilities (in the strict sense) in conditions of uncertainty[103]. Individuals face a choice between a certain health state and a gamble. SG tries to measure utilities by considering the risk behavior of individuals involved. It is consistent with the expected utility theory which proposes, among three assumptions, that the individual will decide rationally to achieve the maximum benefit (higher utilities for the health state than the risk-seeking behavior). The SF-6D, a derivate from the SF-36 used to measure recovery after colorectal surgery, has been valued with the SG method[59].

The TTO is a more easily understood alternative to the SG. Respondents are asked to choose between two certain outcomes[103]. Respondents are asked how many years in a healthy state would be equivalent to x years in a poorer state of health. In other words, how many years would they be willing to sacrifice to avoid a certain poorer health state. In this case, years are the unit of comparison and the value for each outcome can then be calculated. As a result, utility weights indicate the trade-off between the

quality and the length of life. The EQ-5D is valued with TTO, method which is preferred for QALYs calculation and cost-analyses[9]. Compared to VAS, TTO produces higher utility in mild and moderate health states, and lower utility in severe health states[59].

Quality-Adjusted Survival:

Health utility and preference values lead to Health adjusted life-years (HALYs), which include Quality adjusted life-years (QALYs) and Disability-adjusted life-years (DALYs). QALYs and DALYs are population health measures combining mortality and morbidity outcomes in a single indicator, while using utilities and preferences highlighting populations' choices. They are used to estimate the burden of disease and its impact on communities, and in economic analyses.

Different patient populations and health-care interventions can be compared with QALYs. The foundation of QALYs lies in the social theory of utilitarianism, in which policies to improve social welfare should maximise benefits by doing the greatest good for the greatest number of people[104]. Given healthcare budget constraints, QALYs are maximized by improving patients' QoL, hence their utility weights at different periods. As well, interventions that are relatively inexpensive (low cost per QALY) are generally prioritized over those that are relatively expensive (high cost per QALY).

The incremental cost-effectiveness ratio (ICER) is used to summarize the cost-effectiveness of a health care intervention. It is defined as the ratio of the difference in cost between two different interventions to the difference in their effect, such as:

$$ICER_{ERP} = (C_{ERP} - C_{TC}) / (E_{ERP} - E_{TC})$$

In our context, C_{ERP} and C_{TC} represent the cost of the ERP and traditional care, and E_{ERP} and E_{TC} represent the effectiveness of the ERP and traditional care on an outcome, respectively.

When measured in terms of QALY, ICER is also referred to as cost-utility analysis (CUA):

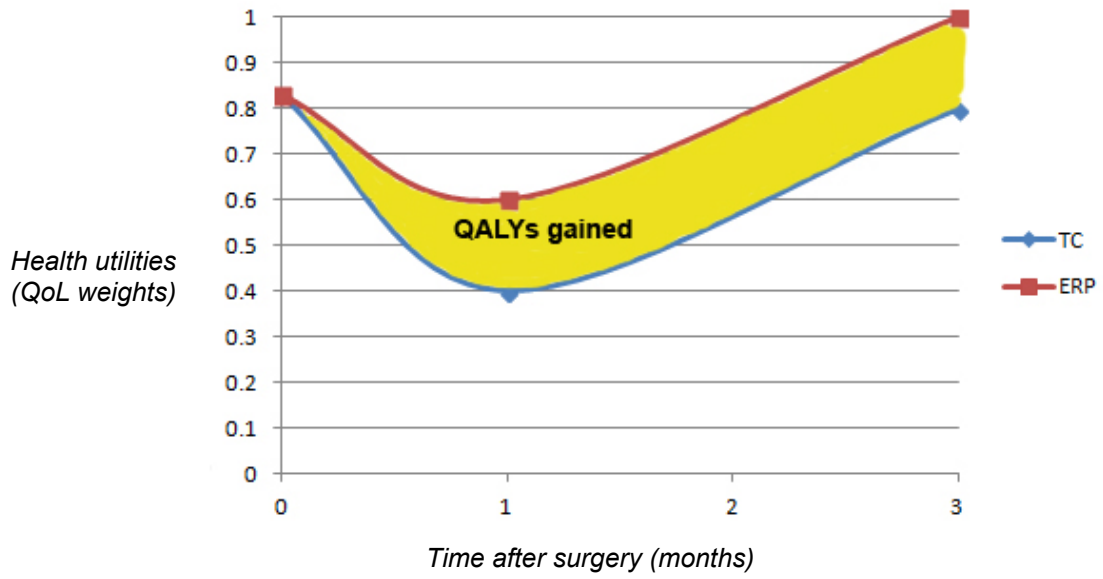
$CUA_{ERP} = (C_{ERP} - C_{TC}) / (QALYs\ produced\ by\ ERP - QALYs\ produced\ by\ TC)$

CUA are suitable when QoL is an important outcome and when health economists aim to compare interventions with a common unit of measurement. QALYs are calculated as the arithmetic product of the number of years of life gained and the utility (or preference) weight of the QoL in each of those years. The ERP would be considered *dominant* if less costly and more effective, and *cost-effective* if more costly and more effective.

For example, if intervention A provides a gain of 5 years at a steady health state of 0.50 until sudden death, it produces 2.5 QALYs. If intervention B provides a gain of 4 years at a steady health state 0.75 until sudden death, it produces 3 QALYs. Intervention B has therefore led to an overall gain of 0.5 QALY compared to intervention A.

Figure 3 shows a hypothetical example of QALYs gained between a group of patients treated with an ERP and another group treated with traditional care with utilities at baseline, 1 month and 3 months after surgery. In this situation, ERP has consistently a greater area under the QALY-time curve than traditional care. QALYs gained can be visualized between both curves (yellow area).

Figure 3 - QALYs gained between traditional care and ERP (hypothetical example)



We aimed to measure QALYs in the context of enhanced recovery to perform a cost-effectiveness analysis of the ERP. Evidence of the economic benefits of an ERP is necessary for several reasons. While ERPs are often associated with shorter LOS[4], this may not have an important impact, as the costs attributable to the last day of a long admission are economically insignificant in the total costs[105]. Compared to LOS, postoperative complications are more significant cost-drivers, our systematic review did not show reveal advantages for ERPs in overall complication rates[106]. Finally, ERPs require resources to develop, implement, audit and maintain[107].

In view of this objective, we designed a cost-effectiveness study to compare ERP to traditional care for lung resection, in which the main exposure variable would be the use of ERP for perioperative management and the main outcome variable would be the ICER from the institutional, healthcare system and societal perspectives using QALYs as the measure of effectiveness. We chose the EQ-5D as the QoL measure since it is very commonly used and recommended for its ability to generate QALY. Nevertheless, we needed first to assess the validity of the EQ-5D as a measure of postoperative recovery after lung resection.

2.4 – Validity Assessment

The validity of EQ-5D refers to whether the instrument adequately measures what it purports to measure (ie, postoperative recovery). The validation process is based on hypotheses testing, aiming to prove (or refute) that EQ-5D measures recovery after lung resection. Criterion validity involves the evaluation of the instrument against an accepted absolute "gold standard" in the measurement for that construct. However, there is no such standard in measuring recovery, hence longitudinal and construct validity are used instead.

2.4.1 - Longitudinal Validity: Responsiveness and minimal clinically important differences (MCID)

An important psychometric property is longitudinal validity, defined as responsiveness over time. Responsiveness is the ability of an instrument to measure a meaningful change by detecting minimal clinically important differences (MCID), even if changes are small[108-112]. In the context of recovery, the score should follow the model of surgical recovery, namely a rapid decline from baseline immediately after surgery, followed by slow rehabilitation over weeks to months.

"Responsiveness" should be distinguished from "sensitivity", often used interchangeably. Sensitivity actually refers to the usefulness of the instrument to measure any change over time, regardless of whether the change is clinically significant[113]. An instrument may be insensitive to change if it misses several relevant items for the specific disease being studied (ie. generic instruments), if the items included in the instrument are static (not the target of the intervention), or if the scoring is subject to floor or ceiling effects.

Responsiveness can be assessed with different methods: either by demonstrating changes in instrument scores before and after an intervention of known efficacy[113], by correlating score changes between functional scales and physiologic measures[112], or by calculating the sensitivity and specificity of scales to develop Receiver Operating Characteristic (ROC) curves[110].

The MCID is defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient' s management"[108]. It is specific to each domain of a patient-centered outcome profile and depends upon the disease process and its severity, and demographic characteristics of the study population. The MCID is computed with anchor-based methods (patient-centered, self-perspective of the health improvement attributable to an intervention)

or distribution-based methods (derived from the standard deviation, standard error of the mean and effect size). There is no consensus (Delphi method) for the EQ-5D regarding the optimal technique, nor value estimate by the EuroQol Group. Coretti et al. highlights the heterogeneity of methods in a recent critical review of the MCID for EQ-5D index[114].

In a literature review, the MCID derived from both the anchor- and the distribution-based methods in different population groups was summarized. Depending on patients' groups, the anchor-based MCID varies from -0.011 (COPD group) to 0.139 (leg ulcer group) and the distribution-based MCID ranges from 0.11 (leg ulcer group I) to 0.14 (leg ulcer group II). For the entire patient population, the mean and median MCID were 0.074 (range -0.011 - 0.140) and 0.081, respectively, when estimated using the anchor-based method[115].

In addition, in a retrospective cohort study by Pickard et al.[116], with 534 patients having advanced cancer - from any primary source - who had undergone at least two cycles of chemotherapy, the MCID was estimated at 0.07 to 0.08 for lung cancer using anchor-based methods, and 0.10 to 0.12 for all subgroups of cancer using distribution-based methods.

2.4.2 – Construct Validity

Construct validity is the degree to which an instrument measures what it claims to be measuring[117]. Constructs are abstractions made to conceptualize variables that cannot be observed or measured directly. This approach involves examining the correlations of the measure with other established measures known to be related to the construct.

The multitrait-multimethod matrix or MTMM, developed by Campbell and Fiske, stresses the importance to demonstrate both convergence and discrimination validity, the two subtypes forming construct validity, in order to establish construct validity[48,118,119]

2.4.2.1 – Discriminant Validity

Discriminative validity, also called known-group validity, assesses whether concepts that theoretically should not be related are, in fact, unrelated. It is the ability of an instrument to distinguish between populations of individuals with known differences in health status[119-122].

2.4.2.2 – Convergent Validity

Convergent validity is concerned with the extent to which the instrument scores are related to other measures, intended to measure the same construct, to which they should theoretically be related. If there is some level of correlation, it should result in the expected direction. That correlation, measured with Pearson or Spearman's rank test, may have positive or negative coefficient[118].

3 – OBJECTIVE AND HYPOTHESIS

The aim of our study is to examine the validity of a questionnaire of health status (EQ-5D) as a measure of postoperative recovery after an elective pulmonary resection. Since there is no gold standard measure of "postoperative recovery", the goal is to evaluate validity by testing *a priori* hypotheses based on the assumption that the EQ-5D index scores represent this construct, hence follow the expected trajectory of recovery after lung resection, discriminate groups known to have different recovery patterns and correlate with other measures of postoperative recovery.

The longitudinal validity (responsiveness) is assessed by testing the hypothesis that the EQ-5D index score at baseline will be higher than at 30 days post-operatively, and the score at 30 days will be lower than at 90 days post-operatively.

The construct convergent validity is assessed by testing the hypothesis that higher EQ-5D index scores will correlate with lower pain levels (VAS) and lower fatigue levels (VAS) at baseline, 30 days and 90 days after surgery.

Finally, the construct discriminant validity is assessed by testing three hypotheses supported by previous literature on lung resection[54,123,124]:

1) Patients with postoperative complications will have a lower EQ-5D index score at 30 and 90 days after surgery than patients without postoperative complications;
2) Patients who underwent a thoracotomy will have a lower EQ-5D index score at 30 and 90 days after surgery than patients who underwent a VATS;
3) Patients above 70 years old will have a lower EQ-5D index score at 30 and 90 days after surgery than younger patients.

4 – MANUSCRIPT

Validity of the EuroQol-5 dimensions (EQ-5D) as measure of recovery after pulmonary resection

In: J Surg Res, 194 (2015), 281-88.

[1]Jimmy Bejjani MD, [1]Julio F Fiore Jr PhD, [1]Lawrence Lee MD MSc, [1]Pepa Kaneva MSc, [1]Juan Mata MD, [1]Annie Ncuti BSc, [2]Lorenzo E Ferri MD PhD, [1]Liane S Feldman MD

[1]*Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, McGill University Health Centre, Montreal, Qc*

[2]*Division of Thoracic Surgery, McGill University Health Centre, Montreal, Qc*

Corresponding author:
Liane S. Feldman, MD FRCSC FACS
Montreal General Hospital 1650 Cedar Ave, L9-309
Montreal, QC H3G 1A4
Tel: (514) 934-1934 ext. 44004
Fax: (514) 934-8210
Email: liane.feldman@mcgill.ca

Validity; EQ-5D; Recovery; Utility; Thoracic Surgery; Lung Cancer; Video-Assisted Thoracic Surgery; Thoracotomy;

Authors contribution: JB contributed to the conception and design of the study, acquisition of data, analysis and interpretation of data, and drafted the article; JFF contributed to the conception and design of the study, analysis and interpretation of data, and drafted the article; LL contributed to the conception and design of the study, interpretation of data and revised the article; PK contributed to the conception and design of the study, acquisition of data and revised the article; JM contributed to the conception and design of the study, acquisition of data and revised the article; AN contributed to the conception and design of the study, acquisition of data and revised the article; CS contributed to the conception and design of the study, interpretation of data and revised the article; DSM contributed to the conception and design of the study, interpretation of data and revised the article; LEF contributed to the conception and design of the study, interpretation of data and revised the article; LSF contributed to the conception and design of the study, interpretation of data, and revised the article.

Word count: 3306

ABSTRACT

Background: Surgical innovations advocated to improve patient recovery are often costly. Economic evaluation requires preference-based measures that reflect the construct of patient recovery. We investigated the responsiveness and construct validity of the EuroQol-5 dimensions (EQ-5D) as a measure of postoperative recovery after planned pulmonary resection for suspected malignant tumors.

Methods: Patients undergoing pulmonary resection completed the EQ-5D questionnaire and visual analog scales (VAS) for pain and fatigue at baseline (preoperatively) and at one and three months postoperatively. Responsiveness and construct validity (discriminant and convergent) were investigated by testing *a priori* hypotheses.

Results: Fifty-five patients were analyzed (45% male, 62±12 years, 29% video-assisted). There was no significant difference between median EQ-5D scores obtained at baseline (0.83 [IQR 0.80-1]) compared to one month (0.83 [0.80-1], p=0.86) and three months after surgery (1 [0.83-1]; p=0.09). At one month after surgery, EQ-5D scores were significantly lower in patients undergoing thoracotomy vs. video-assisted surgery (0.82 [IQR 0.77-0.89] vs. 1 [0.83-1], p=0.003), but there were no significant differences between patients ≥70 years old vs. younger (0.95 [IQR 0.82-1] vs. 0.83 [0.77-1], p=0.09) or between patients with vs. without complications (0.82 [IQR 0.79-0.95] vs. 0.83 [0.80-1], p=0.10). There was a low but significant correlation between EQ-5D and VAS scores of pain and fatigue (Rho -0.30 to -0.47, p≤0.01).

Conclusions: Despite evidence of convergent validity, the EQ-5D was not sensitive to the hypothesized trajectory of postoperative recovery and showed limited discriminant validity. This study suggests that the EQ-5D may not be appropriate to value recovery after lung resection.

INTRODUCTION

Evidence suggests that innovations in thoracic surgery, such as the use of video-assisted thoracic surgery (VATS), robotics and enhanced-recovery pathways (ERP), improve postoperative recovery,[125,126] however, they can be costly to implement and to maintain.[127,128] Given constraints in health-care budgets, economic appraisal of these strategies is increasingly important.

The Panel on Cost-Effectiveness in Health and Medicine recommends that economical evaluations of health interventions be expressed in terms of quality-adjusted life years (QALY).[129] This standardized methodological approach is also endorsed by the UK National Institute of Health and Clinical Excellence (NICE).[130] QALY is a summary measure used in cost-effectiveness analyses (termed "cost-utility analyses" when QALYs are included) to inform resource-allocation decisions. It is calculated by multiplying the time spent in a health state by the quality-of-life weight of that health state measured in terms of "utilities". A utility represents the preference of individuals for being in a particular health state and is measured on an interval scale from zero to one (zero reflects states of health equivalent to death and one reflects full health).[131]

The EuroQol-5 dimensions (EQ-5D) is arguably the utility measure most commonly used to generate QALYs.[132,133]. It comprises a short and cognitively simple questionnaire containing 5 items.[134] There are translations available in 169 languages and country-specific preference weights have been established in 18 countries.[133] The ease with which the EQ-5D is administered, scored, and interpreted possibly accounts for its popularity over other utility measures (e.g. Health Utilities Index Mark 2 and Mark 3 (HUI2 and HUI3), Short Form 6D (SF-6D)). A potential disadvantage of the EQ-5D, however, is that its profile classification may be too crude to capture the health state of populations with low levels of disability.[13,14]. Also, as with any generic measure of health status, the EQ-5D may lack sensitivity to detect condition-specific effects of interventions.[135]

Valuing strategies aimed to enhance postoperative recovery in terms of QALY requires utility measures that reflect postoperative recovery. Recovery after surgery is a complex construct that is poorly defined and measured. The trajectory of recovery is characterized by rapid, transient decline in health status, followed by a gradual rehabilitation towards or beyond baseline levels of health.[27] There is limited research evaluating the measurement properties of utility measures within this specific context.

The objective of this study was to contribute evidence for the responsiveness (i.e. the ability to detect change over time in the construct of interest), discriminant validity (i.e. ability to discriminate between groups hypothesized to differ in the construct of interest) and convergent validity (i.e. overlap between measures that presumably represent the same construct) of the EQ-5D as a measure of postoperative recovery after elective lung resection for suspected malignancy. To address this aim, we tested specific hypothesis based on the assumption that EQ-5D scores follow the expected trajectory of recovery after lung resection, discriminate groups hypothesized to have different recovery patterns and correlate with other measures of postoperative recovery.

METHODS

Patients

Adult patients (18 years of age or older) undergoing elective lobectomy or wedge resection for suspected malignancy at a single university-affiliated institution between September 2011 and August 2012 were considered for inclusion. These patients comprised the control arm of a historical-controlled study assessing the outcomes of the implementation of an Enhanced Recovery Pathway at our thoracic surgery unit. Patients unable to understand English or French and with neuropsychiatric conditions precluding cooperation with the study were excluded. Surgeries were performed using posterolateral (non-muscle sparing) thoracotomy or VATS according to the surgeon's discretion. Demographic data

(age, sex, pulmonary function tests, ASA score and comorbidities) and surgical outcome data (type of surgery, surgical access and complications within 30 days after surgery) were prospectively recorded. Intraoperative and postoperative complications were classified using the Thoracic Surgery Morbidity & Mortality Assessment Tool[136]: Grade I, are complications that do not require pharmacologic treatment or other intervention; Grade II, are complications that require pharmacologic treatment or minor intervention; Grade III, are complications that require surgical, radiologic, endoscopic intervention, or multi-therapy and; Grade IV, are complications that require intensive care treatment and life support. The McGill University Health Centre Research Ethics Board approved the study (ref. 11-006-SDR) and written informed consent was obtained from each participant.

Measures

All participants completed the EQ-5D questionnaire at baseline (preoperatively), and at one- and three-months after surgery. The EQ-5D comprises five descriptive items and a 20-cm visual analog scale to rate overall health (EQ-VAS). The five descriptive items (mobility; self-care; usual activities; pain/discomfort and anxiety/depression) are graded at three levels (no problems, some problems, major problems), classifying responders into one of 243 possible health states. A scoring algorithm based on valuations of general population samples is used to transform these health states into a single utility value (EQ-5D index). As the study on the Canadian valuation of EQ-5D heath states is at preliminary stage,[90] we used the algorithm defined for the US population. Using this algorithm, scores range from -0.109 (extreme problems in all five dimensions) to 1 (full health) (i.e. higher scores indicate better health). As only the descriptive items of the EQ-5D were used for utility evaluation, EQ-VAS scores were not considered in our analysis.

Pain and fatigue were evaluated at the same time points (baseline, one and three months) using visual analog scales (VAS) graded from zero (no pain or fatigue) to 10 (extreme pain or fatigue).

Responsiveness analysis

To test responsiveness, we evaluated whether the EQ-5D index followed the expected trajectory of recovery after a surgical intervention, i.e. a period of immediate deterioration of health status followed by a progressive return to baseline.[25] We tested the *a priori* hypothesis that EQ-5D index scores would be significantly lower at one month after surgery in comparison to baseline, improve from one to three months and return to baseline levels at three months. This trajectory of recovery has been demonstrated in previous studies evaluating postoperative quality of life in patients undergoing lung resection.[137-140]

Validity analysis

As there is no gold standard measure of 'postoperative recovery', validity was evaluated by testing *a priori* hypotheses assuming that the EQ-5D index represents this construct (i.e. construct validity). Discriminant construct validity was assessed by testing the hypotheses that EQ-5D index scores obtained at one month after surgery would be lower in: (1) patients >70 years old in comparison to younger patients, (2) in patients undergoing thoracotomy in comparison to VATS and (3) in patients who developed postoperative complications in comparison to those who did not. These hypotheses have been supported by previous literature.[54,123,124] Convergent construct validity was investigated by testing the hypothesis that EQ-5D scores at one and three months after surgery would inversely correlate with VAS scores of pain and fatigue.

Statistical analysis

We estimated that at least 51 patients would be required to detect a minimal important difference of 0.06 (SD 0.15)[116] in EQ-5D index scores between the assessments at baseline and at 1 month after surgery (responsiveness analysis). This sample size was calculated to yield 80% power with level of significance of 0.05.

Data were expressed as proportion (n), mean (SD), or median [IQR], as appropriate. Histograms of EQ-5D index scores showed that ceiling effect was common at all time points making the data positively skewed. The frequency of missing data was 5% (one patients missed the follow up at one month and two patients missed the follow up at three months). We conducted a sensitivity analysis using chained equations to impute these missing data.[141] As the study results would not be influenced by the use of multiple imputations (possibly due to the low percentage of missing data[142]), we opted to perform a complete case analysis as it facilitates interpretation and allowed the use of non-parametric tests to account for the abnormal distribution of EQ-5D index scores. Wilcoxon's signed-rank test was used to determine whether EQ-5D index scores differed over the course of postoperative recovery (responsiveness analysis). We also performed an exploratory analysis of responsiveness stratifying patients by operative approach and extent of resection. Mann–Whitney U test was used to compare mean EQ-5D index scores between patients >70 years old vs. younger patients, undergoing thoracotomy vs. VATS and with vs. without complications (discriminant validity). The relationship between EQ-5D index scores and VAS scores of pain and fatigue (convergent validity) was assessed using Spearman's correlation (Rho). Correlation coefficients were interpreted as very high (1.00-0.90), high (0.89-0.70), moderate (0.69-0.50), low (0.49-0.26) and little if any (0.25-0.00).[143] Data analysis was performed using Stata 12 (StataCorp, College Station, TX).

RESULTS

Sixty-one patients were considered for inclusion in the study; six were withdrawn due to unplanned pneumonectomy and bilobectomy. Fifty-five patients completed the baseline evaluation and at least one follow up assessment. Demographic and operative characteristics of the patients included in the analysis are reported in Table 1. Mean age was 62±12 years old and 25 patients (45%) were male. Sixteen patients (29%) underwent VATS. Specific procedures included lobectomies (n=31, 56%) and wedge resections (n=24, 44%). Lobectomies were performed via thoracotomy in 27 patients (87%) and

via VATS in 4 patients (13%). Wedge resections were performed via thoracotomy in 12 patients (50%) and via VATS in 12 patients (50%). Neoadjuvant and adjuvant therapy were given to 2 (4%) and 4 (8%) patients, respectively. Complications occurred in 25 patients (45%). These complications were classified as grade I in 10 patients (40%), grade II in 8 patients (32%), grade III in 2 patients (8%), grade IV in 5 patients (20%) (Table 2). There were no postoperative deaths during the study period.

We observed considerable ceiling effects of EQ-5D index scores at all time points. Prior to surgery, 25 patients (45%) had scores at the upper limit of the scale (1 = full health). The frequency of ceiling effect was slightly lower at one month after surgery (n=20, 37%); at three months, 57% of patients (n=30) had EQ-5D index scores clustered at the upper limit. Floor effects were not observed at any time point.

Responsiveness: The hypothesis that the EQ-5D index is responsive to the expected trajectory of recovery after lung resection was not supported by the data (Table 3), as there was no significant decline in median EQ-5D index scores from baseline to one month after surgery. However, scores obtained at three months were significantly higher compared to one month. Similar trends were observed when patients were stratified by operative approach and extent of resection.

Discriminant validity: Baseline EQ-5D index scores were not significantly different between the patient groups tested for discriminant validity (Table 4). Scores at one month after surgery were significantly lower in patients undergoing thoracotomy vs. VATS, but there was no significant difference in scores obtained from patients ≥70 years old vs. younger or patients with vs. without complications. Therefore, our results supported only one of the three hypotheses tested for discriminant validity. At 3 months after surgery, EQ-5D scores were significantly higher in older patients compared to younger patients.

Convergent validity: Results of the convergent validity analysis are presented in Table 5. The EQ-5D index had a low but significant correlation with VAS scores of pain at one and three months after

surgery. The correlation between EQ-5D and VAS scores of fatigue was also low and significant at all time points.

Table 1 - Patient demographics

	All cohort (n=55)	Operative approach		Extent of resection	
		Thoracotomy (n=39)	VATS (n=16)	Lobectomy (n=31)	Wedge resection (n=24)
Age, years	**62±12**	60±12	65±11	65±11	58±12
Sex M:F	**25:30**	18:21	7:9	16:15	9:15
BMI, kg/m^2 *	**26 [22-29]**	26 [22-29]	27 [23-30]	26 [22-39]	27 [22-30]
FEV1, % predicted **	**89±27**	89±27	91±27	89±28	90±26
DLCO, % predicted **	**77±19**	78±20	79±15	77±20	78±17
ASA status					
I	**2 (4)**	2 (5)	0	2 (6)	0
II	**34 (62)**	25 (64)	9 (56)	20 (64)	14 (58)
III	**19 (34)**	12 (31)	7 (44)	9 (29)	10 (42)
Smoking status					
Never smoker	**16 (29)**	9 (23)	7 (44)	10 (32)	6 (25)
Ex-smoker	**29 (53)**	22 (56)	7 (44)	15 (48)	14 (58)
Current smoker	**10 (18)**	8 (20)	2 (12)	6 (19)	4 (17)
Comorbidities					
COPD	**13 (24)**	10 (26)	3 (19)	7 (37)	6 (25)
Diabetes	**10 (47)**	9 (23)	1 (6)	8 (26)	2 (8)
Hypertension	**26 (47)**	16 (41)	10(62)	15 (48)	11 (46)
CAD	**7 (13)**	4 (10)	3 (19)	2 (6)	5 (21)
DLPD	**20 (39)**	13 (33)	7 (44)	10 (32)	10 (42)
Pathologic result					
No malignancy	**8 (14)**	4 (10)	4 (25)	2 (6)	6 (25)
Primary tumor only	**35 (64)**	28 (72)	7 (44)	22 (71)	13 (54)
Primary tumor and nodal metastasis	**7 (13)**	6 (15)	1 (6)	6 (19)	1 (4)
Distal metastasis/secondary tumor	**5 (9)**	1 (2)	4 (25)	1 (3)	4 (16)

Data expressed as mean±SD, median [IQR] or n (%). BMI= body mass index, FEV1 = forced expiratory volume in the first second, DLCO = Carbon Monoxide Diffusing Capacity, ASA = American Society of Anesthesiologists; VATS = video-assisted thoracic surgery; COPD = Chronic obstructive pulmonary disease; CAD = Coronary artery disease; DLPD = Dyslipidemia

* Excluding 2 patients with missing BMI

** Excluding 14 patients with missing FEV1

*** Excluding 15 patients with missing DLCO

Table 2. Description and classification of 30-day postoperative complications

	All cohort (n=55)	Operative approach		Extent of resection	
		Thoracotomy (n=39)	VATS (n=16)	Lobectomy (n=31)	Wedge resection (n=24)
Postoperative complications	**25 (45)**	20 (51)	5 (31)	15 (48)	10 (42)
Prolonged air leak	**14 (25)**	7 (18)	0	5 (16)	2 (8)
Atelectasis	**4 (7)**	1 (2)	1 (6)	1 (3)	1 (4)
Pneumonia	**4 (7)**	1 (2)	1 (6)	0	2 (8)
COPD exacerbation	**4 (7)**	1 (2)	1 (6)	1 (3)	1 (4)
Pleural effusion	**2 (4)**	1 (2)	0	0	1 (4)
Respiratory failure	**6 (11)**	1 (2)	2 (12)	1 (3)	2 (8)
Pulmonary edema	**4 (7)**	2 (5)	0	2 (6)	0
Pulmonary embolism	**3 (5)**	1 (2)	0	1 (3)	1 (4)
Empyema and lobe necrosis	**2 (4)**	1 (2)	0	0	1 (4)
Chyle leak	**2 (4)**	1 (2)	0	0	1 (4)
Hemoptysis	**2 (4)**	1 (2)	0	1 (3)	0
Heart failure	**2 (4)**	1 (2)	0	1 (3)	0
Acute MI	**2 (4)**	1 (2)	0	1 (3)	0
Arrhythmia	**2 (4)**	1 (2)	1 (6)	1 (3)	0
SIADH	**2 (4)**	1 (2)	0	1 (3)	0
Postoperative hemorrhage	**4 (7)**	2 (5)	0	2 (6)	0
Urinary tract infection	**4 (7)**	2 (5)	0	2 (6)	0
Urinary retention	**2 (4)**	0	1 (6)	0	1 (4)
Lumbar radiculopathy	**2 (4)**	1 (2)	0	1 (3)	0
Classification (most severe complication)					
Grade I	**10 (40)**	9 (23)	1 (6)	7 (22)	3 (12)
Grade II	**8 (32)**	6 (15)	2 (12)	5 (16)	3 (12)
Grade III	**2 (8)**	2 (5)	0	0	2 (8)
Grade IV	**5 (20)**	3 (8)	2 (12)	3 (10)	2 (8)

Table 3- Hypothesis testing for responsiveness: Change of EQ-5D scores at 1 and 3 months after surgery in comparison to baseline

	n	Median [IQR]	p-value*
All cohort			
Baseline vs. 1 month after surgery	54	0.83 [0.80-1] vs. 0.83 [0.80-1]	0.86
1 month vs. 3 months after surgery	51	0.83 [0.80-1] vs. 1 [0.83-1]	0.001
Baseline vs. 3 months after surgery	52	0.83 [0.80-1] vs. 1 [0.83-1]	0.09
Lobectomy through thoracotomy			
Baseline vs. 1 month after surgery	26	0.83 [0.81-1] vs. 0.83 [0.80-0.89]	0.49
1 month vs. 3 months after surgery	25	0.83 [0.80-0.89] vs. 1 [0.83-1]	0.009
Baseline vs. 3 months after surgery	26	0.83 [0.81-1] vs. 1 [0.83-1]	0.25
Lobectomy through VATS			
Baseline vs. 1 month after surgery	4	0.91 [0.69-1] vs. 0.91 [0.83-1]	0.70
1 month vs. 3 months after surgery	4	0.91 [0.83-1] vs. 0.93 [0.85-1]	0.46
Baseline vs. 3 months after surgery	4	0.91 [0.69-1] vs. 0.93 [0.85-1]	0.16
Wedge resection through thoracotomy			
Baseline vs. 1 month after surgery	12	0.91 [0.55-1] vs. 0.80 [0.69-0.93]	0.81
1 month vs. 3 months after surgery	12	0.80 [0.69-0.93] vs. 0.85 [0.80-1]	0.009
Baseline vs. 3 months after surgery	12	0.91 [0.55-1] vs. 0.85 [0.80-1]	0.62
Wedge resection through VATS			
Baseline vs. 1 month after surgery	12	0.83 [0.81-1] vs. 1 [0.91-1]	0.09
1 month vs. 3 months after surgery	10	0.1 [0.91-1] vs. 1 [1-1]	0.65
Baseline vs. 3 months after surgery	10	0.83 [0.81-1] vs. 1 [1-1]	0.36

EQ-5D = EuroQol-5D; IQR = interquartile range; VATS = video-assisted thoracic surgery
* Wilcoxon's signed-rank test

Table 4 – Hypotheses testing for discriminant validity: EQ-5D scores at baseline, 1 month and 3 months after surgery in patients hypothesized to have slower postoperative recovery

Comparison	Baseline			1 month			3 months		
	n	**Median [IQR]**	**p-value***	**n**	**Median [IQR]**	**p-value***	**n**	**Median [IQR]**	**p-value***
Older (≥ 70 years old) vs. younger (< 70 years old)	16 vs. 39	0.91 [0.80-1] vs. 0.83 [0.80-1]	0.73	16 vs. 38	0.95 [0.82-1] vs. 0.83 [0.77-1]	0.09	14 vs. 38	1 [0.81-1] vs. 0.93 [0.81-1]	0.02
Thoracotomy vs. VATS	39 vs. 16	0.83 [0.81-1] vs. 0.83 [0.80-1]	0.98	38 vs. 16	0.82 [0.77-0.89] vs. 1 [0.83-1]	0.003	38 vs. 14	1 [0.85-1] vs. 1 [0.83-1]	0.19
Complications vs. no complications	25 vs. 30	0.83 [0.83-1] vs. 0.83 [0.80-1]	0.62	24 vs. 30	0.82 [0.74-0.95] vs. 0.83 [0.82-1]	0.10	24 vs. 28	1 [0.82-1] vs. 1 [0.83-1]	0.66

EQ-5D = EuroQol-5D; VATS = video-assisted thoracic surgery; IQR = interquartile range
*Mann–Whitney U test

Table 5 – Hypothesis testing for convergent validity: Relationship between the EQ-5D scores and VAS scores of pain and fatigue

	1 month after surgery (n=54)		**3 months after surgery (n=52)**	
	Rho	**p-value***	**Rho**	**p-value***
VAS Pain	- 0.45	<0.001	-0.39	0.003
VAS Fatigue	-0.30	0.02	-0.37	0.006

EQ-5D = EuroQol-5D; VAS = visual analog scale; Rho = spearman's correlation coefficient
* Spearman's test

DISCUSSION

The results from the present study do not support the validity and responsiveness of the EQ-5D index as a measure to value recovery after lung resection. Despite demonstrating convergent validity, the EQ-5D was not sensitive to the expected trajectory of postoperative recovery and showed limited discriminant validity. Significant ceiling effect was observed for EQ-5D index scores at all assessment points. This may in part be responsible for the instrument's limited ability to detect differences in health status within and between patients recovering from surgery.

The ceiling effect of the EQ-5D has been recognized by several studies in the past. This issue is often attributed to the fact that the EQ-5D has only three response categories for each of the five dimensions, which limits its ability to discriminate degrees of health status.[13,144] In a population based study from the UK (n=1980), 54% of people who responded the EQ-5D reported full health in spite of having an underlining medical condition.[13] A similar result were observed in a population survey from the US (n=11248), where the rate of ceiling effect was 47%.[144] Having a large proportion of responses at the upper limit of the scale makes the EQ-5D less responsive to changes in health status in conditions associated with low morbidity. Therefore, this instrument may be more suitable for populations with major disability, where the distribution of EQ-5D scores is less skewed.[13,14] The substantial ceiling effect observed in our study suggests that level of disability experienced by patients undergoing lung resection is not high enough to be adequately measured by the EQ-5D.

Patients undergoing surgical interventions experience a period of deterioration followed by a gradual return to baseline levels of health[27]. After lung resection, studies using instruments such as the Short Form-36 Health Survey (SF-36)[137,140] and the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire[138,139] have demonstrated that the health status of patients is still impaired at one month after surgery with return to baseline levels at three months. In the

present study, the EQ-5D was not responsive to this trajectory as the scores obtained at one month after surgery were not significantly different from baseline scores. We hypothesize that the greater descriptive ability the SF-36 and the EORTC Quality of Life Questionnaire make these instruments more likely to detect postoperative changes in health status and, therefore, to represent a complex construct such as 'postoperative recovery'. Although these instruments can be used to extract utility scores[27,145], this use has not been validated in the context of recovery from lung resection. In our study, EQ-5D scores obtained at three months after surgery were significantly higher compared to one month. Scores also tended to be higher at 3 months in comparison to baseline, although this difference was not statistically significant. This result possibly reflects the higher rate of ceiling effect at 3 months, which may be related to improvements in cancer-related symptoms and/or reduced psychological stress (anxiety/depression) after surgery.

Our results showed that the EQ-5D has limited ability to discriminate between groups of patients previously shown in the literature to have slower postoperative recovery. Only one of the three *a priory* hypothesis tested for discriminant validity was confirmed by the data. Patients undergoing VATS had significantly higher EQ-5D index scores at one month after surgery in comparison to those undergoing thoracotomy. This corroborates with available evidence that, in comparison to open surgery, VATS is associated with better health status for up to six months after surgery.[123] Although previous research have suggested that patients ≥70 years old[124] and patients who develop complications[54] have slower postoperative recovery, the EQ-5D could not discriminate these patient groups. Surprisingly, at one month after surgery, patients ≥70 years old tended to have higher EQ-5D scores in comparison to younger patients. At 3 months, scores in older patients were significantly higher. These results were possibly confounded by the fact that older patients tended to have higher scores at baseline. This also could be a result of response shift (i.e. older patients may conceptualize their health and quality of life differently after surgery) or selection bias (i.e. certain comorbidities may contraindicate surgery for

older but not for younger patients). The relationship between development of complications and EQ-5D index scores was in the expected direction (i.e. patients with complications tended to have lower EQ-5D scores) but there was no statistical significance.

We identified a low but significant inverse correlation between EQ-5D and VAS scores of pain and fatigue at all assessment times. This confirms the hypotheses tested for convergent validity. The correlations coefficients were all in the expected direction with higher levels of pain and fatigue indicating lower EQ-5D scores. The low levels of correlation identified were not unexpected, as pain and fatigue do not necessarily affect all the health dimensions covered by the EQ-5D.

In order to properly assess the cost-utility of interventions hypothesized to improve postoperative recovery, it is important that preference-based measures adequate capture this construct. The simplicity and condensed format of the EQ-5D undoubtedly contributed to its global dissemination, however, our study confirms the findings from previous research showing its lack of sensitivity in patients with low levels of disability.[13,14] There is currently an array of generic preference measures described in the literature[146] and future research should focus on assessing if any of these measures have superior psychometric properties when evaluating postoperative recovery. In contrast to the EQ-5D, other utility measures such as the Health Utilities Index Mark 2 and Mark 3 (HUI2 and HUI3) and the Short Form 6D (SF-6D) have greater descriptive ability and lower rates of ceiling effect[144,146-148], likely due to the larger number of health states defined by these instruments. A new version of the EQ-5D using five levels of responses for each dimension (resulting in a total of 683 possible health states) has been proposed by the EuroQol Group[146,149], but it is not yet widely used. A recent study by our research group provided evidence for the responsiveness, and discriminant and convergent validity of the SF-6D as measure of recovery major elective colorectal surgery.[150] We believe that the performance of this measure may be similar in pulmonary resection, but this hypothesis should be tested in future research.

The present study has potential limitations. We investigated a sample of patients undergoing elective lobectomy or wedge resections for suspected malignancy; therefore, our results cannot be extrapolated to emergency surgeries, larger pulmonary resections (i.e. bilobectomies, pneumonectomy) or surgeries for other pulmonary diseases. As we used secondary data from a historical controlled study, our assessment of convergent validity was limited to the measures of recovery available from the primary study (VAS scores of pain and fatigue). Assessing the relationship between EQ-5D index scores with other patient-reported (e.g. SF-36) and performance-based measures (e.g. six-minute walk test[151]) that have been validated as measures of postoperative recovery would have provided valuable information in regards to convergent validity. In addition, our analysis was limited to measures obtained at baseline, 1 month and 3 months after surgery. Although previous studies have shown that the trajectory of recovery can be detected by evaluating patients at these three time points, we cannot exclude that the EQ-5D is more responsive to change if measured within the first days/weeks after surgery[77] (when patients are more likely to have issues with the domains evaluated by the EQ-5D). Another important limitation that needs to be noted is that this study was powered to detect within-group differences in EQ-5D scores (responsiveness analysis); therefore, our sample may not have been large enough to detect differences between groups (discriminant validity). Our stratified analysis of responsiveness is also potentially underpowered. Although the discriminatory ability of the EQ-5D could have increased if we analyzed a larger sample, this likely would not have changed our conclusions in regards to the usefulness of this instrument as a measure of recovery. The poor responsiveness and high rate of ceiling effect provides evidence that this instrument may not be psychometrically appropriate to represent this construct.

Our study provides avenues for future research. Establishing the psychometric properties of an outcome measure is a continuous process of investigation, therefore, we encourage that our findings be confirmed in studies involving larger samples and with EQ-5D score obtained at shorter time points.

Other hypotheses regarding the discriminant validity of the EQ-5D should be tested (e.g. using pulmonary function data and comorbidity indexes as predictors). Also, future research should focus on assessing whether other preference-based measures have superior psychometric properties when measuring recovery after lung resection.

In conclusion, the EQ-5D was not sensitive to the expected trajectory of postoperative recovery and showed limited discriminant validity in patients undergoing elective pulmonary resection. These results suggest that the EQ-5D may not be an adequate instrument to generate QALYs in studies assessing the cost-utility of interventions intended to enhance recovery.

5- FURTHER DISCUSSION

Our study is the first to assess the validity of a HRQoL instrument as measure of recovery in a population undergoing elective lung resection. Prior to evaluating the cost-effectiveness of "Enhanced Recovery Programs", it is essential that the chosen measure reflects the outcome being examined, that is recovery. Furthermore, we have chosen to define recovery as a multidimensional construct, as patients do, emphasizing return to baseline functioning and activities[7]. Although EQ-5D is a widely used generic preference-based measure recommended for the estimation of QALY[10], our results cannot support its use as a measure of recovery and caution its use for cost-effectiveness analysis of ERPs in this surgical population. Several limiting factors may have influenced our results, for which alternatives are presented for future studies.

5.1 – Patients' selection

Between September 2011 and August 2012, a total of 242 patients underwent elective lung resection at our institution. For the present study, patients were recruited at the pre-operative clinic where study eligibility was completed for 119 patients (49%), of which 78 consented for the study. The other patients treated during this time were not seen in our clinic and instead were prepared for surgery in referring hospitals. These patients therefore were not included in the study. While the only factor precluding patients from inclusion was geographic, whether they differ in other characteristics that could affect recovery was not evaluated. Among consenting patients, 23 were later excluded from statistical analysis because their surgery was canceled (2), a greater extent of resection was performed (6), or their follow-up was incomplete (15 unreachable or withdrew from the study). There was no difference between patients included and those excluded from the analysis (table 8).

This study was performed in a relatively homogenous population of patients, excluding patients operated in an emergent setting. Patients undergoing a pneumonectomy were also excluded as their

recovery differs, since they do not require chest tubes, resulting in better pain control and a shorter LOS. However, their respiratory functioning after discharge may be significantly different than patients having more limited resection. While limiting the generalizability of our results, this approach allows for a more focused assessment of the EQ-5D in the context of more limited lung resection.

5.2 – Impact of the anxiety/depression dimension

The contribution of the anxiety/depression dimension in the summary index is one of the main reasons for which the overall EQ-5D score did not follow the trajectory of recovery. Our results have shown that a higher proportion of patients report problems with pain/discomfort, usual activities, mobility and self-care at 30 days after surgery compared to baseline. However, fewer patients report problems with anxiety/depression at 30 days ($p=0.03$, Chi-square test). Figure 4 describes the anxiety/depression curve compared to the other dimensions. Also, anxiety/depression is among three dimensions with a significant change at 30 days from baseline, along with usual activities and pain/discomfort (Table 9). Moreover, the weights associated with anxiety/depression (-0.156 for level 2 and -0.450 for level 3) are stronger than those associated with usual activities (-0.140 for level 2 and -0.374 for level 3). The synergy of these three reasons may contribute to the ability of the anxiety/depression dimension to counter-balance the impact of other dimensions in the summary index.

Table 8 - Comparison between patients included vs. excluded from the analysis

	Patients consented (n=78)		
Demographics	Included (n=55)	Excluded (n=23)	p-value
Age	62 ± 12	66 ± 11	0.18
Sex M:F	25:30	11:12	0.85
BMI, kg/m^2 *	26 (22-29)	27 (24-29)	0.77
FEV1, % predicted **	89 ± 27	93 ± 19	0.77
DLCO, % predicted ***	77 ± 19	73 ± 15	0.34
ASA status			
I	2 (4)	1 (4)	0.94
II	34 (62)	15 (65)	
III	19 (34)	7 (30)	
Smoking status			
Never smoker	16 (29)	9 (39)	0.39
Ex-smoker	29 (53)	8 (35)	0.15
Current smoker	10 (18)	6 (26)	0.48
Comorbidities			
COPD	13 (24)	2 (9)	0.13
Diabetes	10 (18)	4 (17)	0.93
Hypertension	26 (47)	13 (57)	0.46
CAD	7 (13)	4 (17)	0.59
DLPD	20 (39)	7 (30)	0.52
Pathologic result			
No malignancy	8 (14)	2 (9)	0.48
Primary tumor only	35 (64)	17 (74)	0.38
Primary tumor and nodal metastasis	7 (13)	2 (9)	0.61
Distal metastasis/secondary tumor	5 (9)	2 (9)	0.96

Data expressed as mean ± SD, median [IQR] or n (%) and statistical significance with Student t-test, Mann-Whitney U test and Chi-square test, respectively. BMI= body mass index, FEV1 = forced expiratory volume in the first second, DLCO = Carbon Monoxide Diffusing Capacity, ASA = American Society of Anesthesiologists; COPD = Chronic obstructive pulmonary disease; CAD = Coronary artery disease; DLPD = Dyslipidemia

* Excluding 2 patients from the study cohort with missing BMI.

** Excluding 14 patients from the study cohort and 7 among excluded patients with missing FEV1.

*** Excluding 15 patients from the study cohort and 8 among excluded patients with missing DLCO.

Figure 4 - Proportions (%) of patients reporting no problem, per dimension and time point

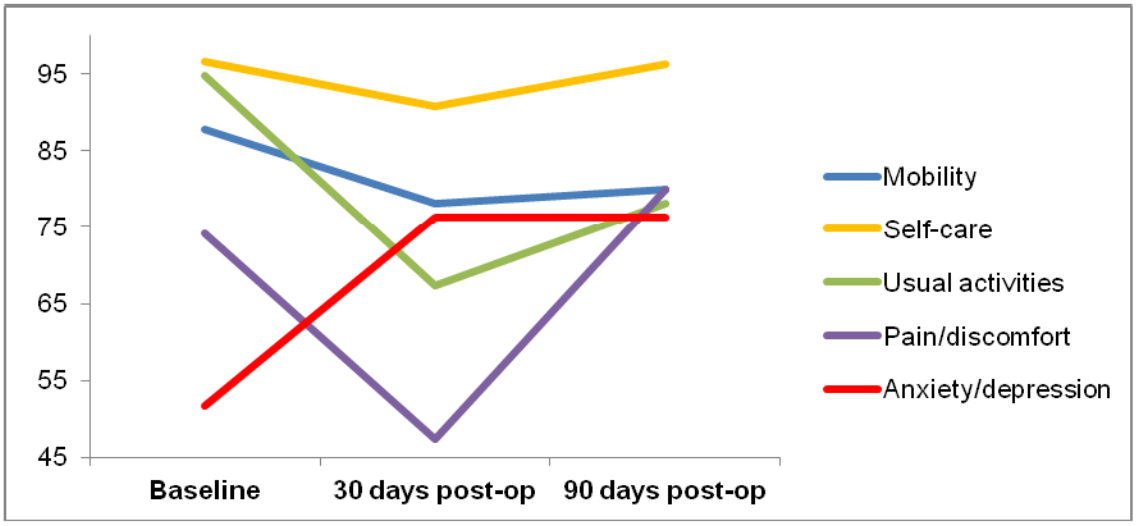


Table 9 - Percentage of change between baseline and 30 days after surgery, per dimension and level

EQ-5D DIMENSIONS		% at baseline (n=55)	% at 30d post-op (n=55)	p-value*
MO	No problem	88.9	77.4	0.11
	Some problems	11.1	22.6	
	Extreme problems	0	0	
SC	No problem	96.4	90.6	0.22
	Some problems	3.6	9.4	
	Extreme problems	0	0	
UA	No problem	94.5	96.8	**< 0.01**
	Some problems	5.5	28.3	
	Extreme problems	0	1.9	
PD	No problem	72.7	49.1	**0.02**
	Some problems	20	45.3	
	Extreme problems	7.3	5.7	
AD	No problem	53.7	77.4	**0.03**
	Some problems	33.3	18.9	
	Extreme problems	13	3.8	

*Chi-square test

MO: mobility; SC: self-care; UA: usual activities; PD: pain/discomfort; AD: anxiety/depression

5.3 – Inconsistency between EQ-5D index utility and EQ-VAS

Patients rate psychological issues after surgery as less relevant to their postoperative recovery process compared to functional status and energy level [68]. In contrast to the EQ-5D index, the EQ-VAS allows patients to define 'health state' in their own terms which may therefore emphasize physical status over mental health during recovery. In further statistical analysis, the EQ-VAS better conformed to the recovery trajectory, showing a statistically significant decline from baseline to 30 days after surgery (table 10), in contrast to the EQ-5D index. Both the EQ-VAS and the EQ-5D index detected improvements from 30 to 90 days. Changes in the EQ-5D index and in the EQ-VAS are not necessarily consistent.

Table 10 - Change of EQ-VAS at 1 and 3 months after surgery in comparison to baseline

	n	Median [IQR]	p-value*
All cohort			
Baseline vs. 1 mo after surgery	54	80 [70; 90] vs. 75 [60; 89]	**0.03**
1 mo vs. 3 mo after surgery	51	75 [60; 89] vs. 80 [74; 90]	**< 0.01**
Baseline vs. 3 mo after surgery	52	80 [70; 90] vs. 80 [74; 90]	0.76

* Wilcoxon Signed Rank test

Additionally, the EQ-VAS scores discriminated between patients with and without complications at 30 days (table 11), which was not seen with the EQ-5D index. Again, the possible omission of anxiety/depression in patients' self-rating of their health status could underlie this finding, in contrast to the EQ-5D index. Interestingly, the EQ-VAS was not been found to be responsive to the trajectory of recovery in colorectal surgery[152]. This underlines the need to evaluate the responsiveness of measures in specific contexts.

Table 11 - Difference in EQ-VAS between patients with and without complications

	n	Median [IQR]	p-value*
All cohort			
Baseline	25 vs. 30	80 [73; 93] vs. 80 [70; 92]	0.65
1 month	24 vs. 30	70 [50; 85] vs. 80 [70; 90]	**0.03**
3 months	24 vs. 28	80 [70; 90] vs. 80 [80; 95]	0.23

* Mann-Whitney test

5.4 – Considerations relating to the ceiling effect

We observed considerable ceiling effects of EQ-5D index scores at all time points. Prior to surgery, 25 patients (45%) had scores at the upper limit of the scale (1 = full health). The frequency of ceiling effect was slightly lower at one month after surgery (n=20, 37%); at three months, 57% of patients (n=30) had EQ-5D index scores clustered at the upper limit. This ceiling effect suggests the possibility that the morbidity after lung resection is not sufficiently high to be measured by the EQ-5D. However, morbid populations such as Hodgkin's disease and advanced HIV disease have also shown evidence of ceiling effect[153,154]. Therefore, the non-parametric distribution of EQ-5D index should be reported since the mean value may not provide an adequate picture of patients' HRQoL. The distribution is usually such (e.g., discontinuous, with peaks or with a high ceiling effect) that conventional statistical methods are not suitable for data analysis. Such distribution is influenced by the inclusion of only 3 levels per dimension in the standard format of the EQ-5D. This characteristic has lead to the development of a recent US median model (MM-OC model), instead of the D1 model based on mean TTO valuations, with scores ranging from -0.81 to 1[155]. The EuroQol Group has also recently introduced a 5-level EQ-5D, given the ceiling effect of the standard format (3-level) and its restricted ability to discriminate small to moderate changes in health status[156]. The five levels include no problems, slight problems, moderate problems, severe problems and extreme problems. Preliminary studies have shown a reduced ceiling effect, an increased reliability, an improved ability to discriminate between health statuses, and established convergent and discriminant validity[149,157]

5.5 - Assessment time points

As well, the appropriate timing in the assessment of QoL is essential with the dynamic construct that is postoperative recovery. In order to use the EQ-5D for cost-utility analysis, and considering the particularity of the anxiety/depression dimension, the choice of postoperative time points should be altered in future studies using the EQ-5D to assess postoperative recovery. Our time points have contributed to the ceiling effect and difficulties in identifying the change among and within patients. The selection of one and three months after surgery was based on the literature of lung cancer surgery with the EORTC QoL Questionnaire-C30[138,139] and the SF-36 Health Survey[137,140].

Considering the short timeframe of an ERP, evaluating the validity of EQ-5D for use in the economic evaluation of an ERP could have included assessments before the discharge day. However, we have planned the first postoperative assessment at one month after surgery considering the positive validation study of SF-6D after elective colorectal resection; postoperative assessments were at 1 and 2 months after surgery, while the median hospital LOS was 5 days[150].

Nonetheless, for a single dimension among five (i.e. anxiety/depression), the sensitivity of the EQ-5D index is much greater than the sensitivity of indexes from measures with thirty items and over. Using measurements at considerably shorter time points (2, 4, 6, 8 days after surgery), one study reported EQ-5D scores after open liver resection that follow the trajectory of recovery[77]. The assessment of recovery from a moderate COPD exacerbation has also shown the greatest improvement to occur in the acute phase, within 14 days[158].

5.6 - Impact of the EQ-5D content density

The EQ-5D presents a high content density, that is the ratio of the number of meaningful concepts to the number of items[68]. This may cause problems in items interpretation between patients or time points, and may consequently provide an imperfect representation of recovery. For example, item 3: *Usual*

Activities (e.g. work, study, housework, family or leisure activities) contains 5 different meaningful concepts. For this item, a patient may report no problem at baseline but some problems at 30 days after surgery should he/she encounter limitations in performing any of these 5 meaningful concepts. As well, the spectrum of physical ability on the EQ-5D is much narrower than activities reported by patients as affected after major interventions[159]. For instance, *mobility* should include more activities than walking, since impairments in mobility could result in problems maintaining an upright posture or transfer activities. This may contribute to the observation seen on figure 4 and table 9, that the *usual activities* dimension does significantly change between baseline and 1 month after surgery, but the *mobility* dimension does not.

In addition, regarding its content, the EQ-5D does not assess patients' cognitive status (concentration, memory, language or executive function) and may not capture all important psychological changes given the lack of specific questions on psychological changes outside those on anxiety and depression. However, no recovery measure can be both comprehensive and simple to complete[11].

5.7 – Sample size, discriminatory power and responsiveness

Our sample size of 51 patients was estimated using a within-group design, with a MCID of 0.06 and a standard deviation (SD) of 0.15, to achieve a power of 80% and α of 0.05. A sample size calculation with a between-group design would have been more appropriate to assess discriminant validity, which would result in a sample size of 144 patients. Regardless of discriminant validity, in light of the impact of the anxiety/depression dimension, our time points selection and the ceiling effect, a larger sample would not have improved responsiveness to the trajectory of recovery. In addition, the MCID was expressed in terms of means and SD, the sample size was calculated with these statistical indices despite the non-parametric distribution of the data with the EQ-5D.

Compared to other measures of recovery, the lower discriminatory power and responsiveness of the EQ-5D is also described the literature. The better psychometric properties of other generic measures are due to richer health state descriptive systems. Patients reporting a health problem have significantly higher mean scores on the EQ-5D compared to scores on SF-12 and SF-36 for all dimensions[13,14]. The restricted ability to discriminate between mild to moderate changes in health status is mentioned with several conditions, and partly explained by the pronounced ceiling effect[12]. When a clinically important change occurs, its consequent change in the index is large. This effect results from the inclusion of only 3 levels per dimension in the descriptive system, and from variables in valuation algorithms changing the index widely should a change takes place from or to level 3 on any dimension (D1, I3 and $I3^2$ in the US algorithm; N3 in the UK algorithm).

In the evaluation of convergent validity, we have noted a statistically significant but clinically weak result. The more simple and focused fatigue VAS and pain VAS scores are not measuring the psychological distress from anxiety that is an important component of EQ-5D. Therefore, the lack of more significant correlation may be due to the limitations of the VAS physical scores themselves while the EQ-5D could potentially be the more appropriate score in this situation. It is a possible limitation when choosing a comparator that is not validated as a gold standard.

In the evaluation of discriminant validity, the difference in EQ-5D index score between patients who underwent a thoracotomy and patients who underwent a VATS might have been greater if measures were done earlier. Such difference might also have even more convincing should EQ-5D be compared to other scores that did vary at these time points. As a word of caution, no methodology has been used to control for tumor size and other confounding factors; it is not useful for the validity assessment of EQ-5D but remains important should results of this study be used to show actual differences between thoracotomy and VATS. Besides, the expected difference in outcomes between younger and older patients is our weakest *a priori* hypothesis. It is supported by only one trial that shows a difference at 6

months[124]; two studies show no difference[54,140]. An instrument, even if it is valid, cannot be expected to identify a difference in QoL between these two groups if such difference is not present.

In addition, we noted that older patients reported higher EQ-5D index scores at baseline. Salati et al. have observed a similar effect in evaluating recovery in the elderly after major lung resection, using the SF-36[140]. At baseline, the elderly group had higher scores on the mental health (MH) item and the mental component summary (MCS), and it has been hypothesized that older patients are more prepared to face the challenge of cancer and its consequent surgery. This possibility is an additional factor to the response shift phenomenon and the selection bias related to the appropriate selection of patients fit for surgery. The selection of surgical candidates is also an underlying bias for the absence of negative health states (WTD) in our cohort, even with the elderly. With the US valuation set, a health status WTD would have occurred if a patient reports extreme problems on 4 dimensions out of five, including either mobility or pain/discomfort. Both of these dimensions are necessary given their strong weights at level 3 (-0.537 and -0.558, respectively).

5.8 - Mapping PRO measures to preference-based measures

Other PRO measures have been studied in the context of recovery but are not commonly reported to evaluate ERP after surgery[65]. Since recovery is a multidimensional construct, the assessment of ERP should not be limited to clinical effectiveness and should involve PRO measures to reflect outcomes of greatest importance to patients. Currently, these measures - whether generic or recovery-specific QoL instruments - have limitations preventing their use in our study population. They have either not been validated as measure of recovery after lung resection or cannot be directly applied for cost-effectiveness analysis and require "mapping" models. Mapping enables the use of health status data from non-preference based QoL measures to be transformed into utility data for the calculation of QALY. Using empirical data, regression models or algorithms are developed to estimate the

relationship between target preference-based measures (e.g. EQ-5D or SF-6D) and other indicators or measures of health (e.g. QLQ-C30)[160].

Among PRO measures used to compare ERPs to traditional care, the European Organization for Research and Treatment of Cancer Quality-of-Life Questionnaire-C30 (QLQ-C30) and the SF-36 are the most common[65,161]. Both are reported with patients undergoing lung resection[137-140]. The QLQ-C30 and SF-36 scoring systems are not preference-based and have no valuation; they are then unsuitable to calculate QALYs and be used in CUA directly. However, they can be mapped to a preference-based generic instrument, such as the Short Form 6D (SF-6D), and used indirectly in CUA provided satisfactory validity of the "target" measure (e.g. SF-6D) as measure of recovery after lung resection.

5.9 – The EORTC QLQ-C30, SF-36 and SF-6D

The EORTC QLQ-C30 is a QoL instrument that incorporates 30 items in five functional scales, three symptom scales, a global health and QoL scale, and six single-item symptom measures[15]. QLQ-C30 includes several aspects of recovery important to patients [68]. Although it is uncommonly used to evaluate recovery after surgery, it is the predominant instrument to measure HRQoL in lung cancer studies[71]. Given its strong psychometric properties with lung cancer patients, QLQ-C30 was identified as the best developed of 50 instruments measuring dimensions of QoL in this setting[162]. The QLQ-C30 was not correlated with EQ-5D since it was not planned, and we could not obtain scores retrospectively. The SF-36 is a widely applied generic measure of QoL[163]. Its physical and mental health subscales are described in Table 6. It has evidence of validity and is responsive to the expected trajectory of surgical recovery after lung resection[164] and it may be more sensitive than EQ-5D[13].

The SF-6D is a generic preference-based health measure derived from 11 items of the SF-36. It covers six domains: physical functioning, role limitations, social functioning, pain, mental health and vitality; with four to six severity levels, and presents 18,000 unique health states. Hence, SF-6D is more

sensitive in detecting smaller changes than EQ-5D. As drawback, its description of health states is limited at the lower end of the scale, and it is valued by the SG technique which tend to produce higher utilities than the TTO technique. Consequently, its most severe state generates a utility score of 0.291 and it may underestimate the magnitude of changes in QoL for patients starting at the lower end[165].

SF-6D provides a means for using the SF-36 and QLQ-C30 data to generate utilities and calculate QALYs for cost-utility analyses[166,167]. The evidence to support mapping models from QLQ-C30 to SF-6D with different cancer patients is recent[168]. While the most appropriate method for mapping is uncertain, the latest study including lung cancer patients has shown that four regression models performed similarly within each cancer type[167].

5.10 – Discriminatory power of SF-6D between ERP and traditional care groups

Recently, the SF-6D has been validated as measure of postoperative recovery after elective colorectal resection by Lee et al.[150] and shown to be superior to EQ-5D[152]. Specifically, the SF-6D follows the expected trajectory of postoperative recovery after colorectal surgery and a ceiling effect was observed in less than 1% of values. In contrast, the EQ-5D was not responsive to this trajectory and showed a significant ceiling effect. Therefore, Lee et al. selected the SF-6D to examine the cost-effectiveness of ERP in colorectal surgery[107]. Nonetheless, while the colorectal ERP was been associated with shorter LOS, lower care-giver burden, less time off work and subsequently lower overall societal cost, there were no differences in QoL as measured by SF-6D Therefore, while the SF-6D may be a valid and responsive measure of recovery, it did not discriminate between the ERP and traditional care groups.

The psychometric properties after colorectal surgery cannot be generalized to other populations. For example, the SF-6D was found to be less responsive to recovery than the EQ-5D after liver transplantation and arthroscopic partial meniscectomy surgery[165,169]. Assessing its validity as measure

of recovery after lung resection and its discriminatory power between ERP and traditional care groups should be considered for future research.

5.11 – Envisioning other preference-based measures: HUI and QWB Scale

The Health Utilities Index (HUI) and the Quality of Well-Being Scale (QWB) are two other preference-based measures which psychometric properties that could also be considered to assess recovery after lung resection, and have the benefit of direct use in cost-effectiveness analyses.

HUI is a generic multi-attribute classification system consisting of 2 systems, Mark 2 (HUI2) and Mark 3 (HUI3)[170]. Together HUI2 and HUI3 describe almost 1 million unique health states and their scoring systems provide utility scores for calculating QALYs. Along with EQ-5D, HUI is considered to be the best health status measure and recommended in studies intended to be used in economic evaluation[59]. A study from Statistics Canada has also shown HUI as superior to EQ-5D in distinguishing between minor levels of impairment[171]. Although used in a wide variety of health problems, only one study is found to report its use to assess surgical recovery[66]. Its validity may be studied in recovery as it incorporates relevant attributes for recovery such as mobility, emotion, cognition, self-care and pain.

The QWB is the first generic instrument specifically designed to measure HRQoL for the estimation of QALYs[172]. The QWB questionnaire includes 71 items divided into three scales (mobility, physical activity and social activity) with 3 to 5 levels of function; and a symptom problem complex score. The inclusion of the effect of symptoms on patients' life explains the lack of ceiling effect[173]. However, its use in surgical populations has been infrequent. This is likely because of its lengthy completion and the relative complexity of interviewer training required until the self-administered version was developed.

5.12 – Recovery-specific HRQoL instruments: PSR scale and QoR-40

Systematic reviews of recovery specific QoL instruments have recommended the use of the Postdischarge surgical recovery (PSR) scale and the Quality of recovery-40 (QoR-40) instrument in

studies measuring recovery[69,70]. Their psychometric properties were reportedly better than other recovery-specific instruments, but further assessment is required to examine their responsiveness and MCID. Like the QLQ-C30 and SF-36, their use for CUA requires mapping to a preference-based measure valid as measure of recovery after lung resection.

There are limitations with both instruments. The PSR scale has only been used with short-stay surgical patients; it incorporates 15 items and is mainly focused on daily life activities[174,175]. After lung resection, it may be appropriate in the weeks after surgery. However, it does not consider mental health.

The QoR-40 includes 40 items in five dimensions: emotional state, physical comfort, psychological support, physical independence, pain[176,177]. A meta-analysis reports its validity in measuring postoperative recovery in several types of surgeries[178]. However, it was designed to measure early recovery and normalizes within days to weeks. No study with lung resection was available before publication, but the QoR-40 has been recently used with thoracoscopic (VATS) lung resection until 48h after surgery[179,180].

6- FINAL CONCLUSION

The study has fostered an understanding of the advantages and drawbacks of EQ-5D in patients undergoing pulmonary resection. The EQ-5D is widely used and recommended in economic appraisal of healthcare technologies, and provides valuable insight into patient perceptions of health and quality of life. However, with our current limitations, it does not reflect appropriately the construct of recovery in this population; its lack of sensitivity and its limited discriminant validity remain important barriers. As part of the construct validity assessment, if the *a priori* discriminant validity hypothesis regarding the elderly vs. young patients was not an appropriate presumption, if VAS scores including anxiety were used - reinforcing the convergent validity to a more convincing correlation - and if the measured time points have been, indeed, chosen too late, the EQ-5D could potentially be an appropriate tool. Nonetheless, our current results caution the use of EQ-5D in cost-utility analyses of technologies aimed at enhancing recovery after lung resection.

Among future considerations, in addition to considering larger samples and shorter time points, the new EQ-5D version with 5 levels, the EQ-5D-5L, should be assessed to examine whether the reduced ceiling effect and the improved discrimination ability reported with other diseases are also reflected in this population. As well, the future Canadian valuation set for EQ-5D and the US median model can be considered to observe their effect on the statistical distribution of utilities and the trajectory of recovery. Other preference-based instruments, such as the SF-6D, HUI and QWB, may be assessed further to examine whether they better reflect the construct of postoperative recovery in this context, while remaining careful to differences in utilities resulting from their respective valuation techniques. Notwithstanding the benefit of any of generic instruments in health economics and resource-allocation, complementing their use with disease-specific instruments would better describe patients' recovery. Mapping non-preference based instruments that are recommended to measure recovery, such as the SF-36, QLQ-C30, PSR scale and QoR-40, can also be envisioned in future studies. As well, EQ-5D may be correlated to these recovery tools at earlier time points in future trials. Finally, developing and

validating a specific preference-based health index would provide an innovative approach to measure recovery with greater accuracy.

7- REFERENCES

1. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Canadian Cancer Statistics 2014. Toronto, ON: Canadian Cancer Society; 2014. May 2014. ISSN 0835-2976.
2. Campos JH. Fast track in thoracic anesthesia and surgery. Current opinion in anaesthesiology 2009;22:1-3.
3. White PF, Kehlet H, Neal JM, et al. The role of the anesthesiologist in fast-track surgery: from multimodal analgesia to perioperative medical care. Anesthesia and analgesia 2007;104:1380-96.
4. Spanjersberg WR, Reurings J, Keus F, van Laarhoven CJ. Fast track surgery versus conventional recovery strategies for colorectal surgery. The Cochrane database of systematic reviews 2011:CD007635.
5. Nicholson A, Lowe MC, Parker J, Lewis SR, Alderson P, Smith AF. Systematic review and meta-analysis of enhanced recovery programmes in surgical patients. The British journal of surgery 2014;101:172-88.
6. Cheema FN, Abraham NS, Berger DH, Albo D, Taffet GE, Naik AD. Novel approaches to perioperative assessment and intervention may improve long-term outcomes after colorectal cancer resection in older adults. Ann Surg 2011;253:867-74.
7. Kleinbeck SV, Hoffart N. Outpatient recovery after laparoscopic cholecystectomy. AORN journal 1994;60:394, 7-8, 401-2.
8. Lee L, Tran T, Mayo NE, Carli F, Feldman LS. What does it really mean to "recover" from an operation? Surgery 2013.
9. Brazier JE, Rowen, D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for generating health state utility values. Available from http://wwwnicedsuorguk 2011.
10. NICE. Guide to the Methods of Technology Appraisal. National Institute for Clinical Exellence. 2013.
11. Paul Kind RB, Rosalind Rabin. EQ-5D concepts and methods: A developmental history. 2005.
12. Saarni SI, Harkanen T, Sintonen H, et al. The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2006;15:1403-14.
13. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 1998;7:155-66.
14. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 1993;2:169-80.
15. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute 1993;85:365-76.
16. Smith EL, Hann DM, Ahles TA, et al. Dyspnea, Anxiety, Body Consciousness, and Quality of Life in Patients with Lung Cancer. Journal of Pain and Symptom Management Journal of Pain and Symptom Management 2001;21:323-9.
17. Tanoue L. Chapter 66 - Lung Cancer: Clinical Evaluation and Staging. In: Agustí SGSAS, ed. Clinical Respiratory Medicine (Fourth Edition). Philadelphia: W.B. Saunders; 2012:788-800.
18. Detterbeck FC, Postmus PE, Tanoue LT. The stage classification of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e191S-210S.

19. Burdett SS, Stewart LA, Rydzewska L. Chemotherapy and surgery versus surgery alone in non-small cell lung cancer. The Cochrane database of systematic reviews 2007:CD006157.
20. Douillard J-Y, Rosell R, De Lena M, et al. Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB–IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. The Lancet Oncology 2006;7:719-27.
21. Manser R, Wright G, Hart D, Byrnes G, Campbell DA. Surgery for early stage non-small cell lung cancer. The Cochrane database of systematic reviews 2005:CD004699.
22. Yan TD, Black D, Bannon PG, McCaughan BC. Systematic review and meta-analysis of randomized and nonrandomized trials on safety and efficacy of video-assisted thoracic surgery lobectomy for early-stage non-small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2009;27:2553-62.
23. Madani A, Fiore JF, Jr., Wang Y, et al. An enhanced recovery pathway reduces duration of stay and complications after open pulmonary lobectomy. Surgery 2015;158:899-910.
24. Allvin R, Berg K, Idvall E, Nilsson U. Postoperative recovery: a concept analysis. Journal of Advanced Nursing 2007;57:552-8.
25. Carli F, Mayo N. Measuring the outcome of surgical procedures: what are the challenges? British journal of anaesthesia 2001;87:531-3.
26. Carli F, Zavorsky GS. Optimizing functional exercise capacity in the elderly surgical population. Current opinion in clinical nutrition and metabolic care 2005;8:23-32.
27. Lee L, Tran T, Mayo NE, Carli F, Feldman LS. What does it really mean to "recover" from an operation? Surgery 2014;155:211-6.
28. Cerfolio RJ, Bryant AS. Does minimally invasive thoracic surgery warrant fast tracking of thoracic surgical patients? Thoracic surgery clinics 2008;18:301-4.
29. Allen MS, Darling GE, Pechet TT, et al. Morbidity and mortality of major pulmonary resections in patients with early-stage lung cancer: initial results of the randomized, prospective ACOSOG Z0030 trial. The Annals of thoracic surgery 2006;81:1013-9; discussion 9-20.
30. Wright CD, Gaissert HA, Grab JD, O'Brien SM, Peterson ED, Allen MS. Predictors of Prolonged Length of Stay after Lobectomy for Lung Cancer: A Society of Thoracic Surgeons General Thoracic Surgery Database Risk-Adjustment Model. The Annals of Thoracic Surgery;85:1857-65.
31. Handy JR, Jr., Asaph JW, Skokan L, et al. What happens to patients undergoing lung cancer surgery? Outcomes and quality of life before and after surgery. Chest 2002;122:21-30.
32. Lee L, Mata J, Ghitulescu GA, et al. Cost-effectiveness of Enhanced Recovery Versus Conventional Perioperative Management for Colorectal Surgery. Ann Surg 2014.
33. Kehlet H, Wilmore DW. Evidence-based surgical care and the evolution of fast-track surgery. Annals of surgery 2008;248:189-98.
34. Adamina M, Kehlet H, Tomlinson GA, Senagore AJ, Delaney CP. Enhanced recovery pathways optimize health outcomes and resource utilization: A meta-analysis of randomized controlled trials in colorectal surgery. Surgery 2010;149:830-40.
35. Fiore JF, Jr., Bejjani J, Conrad K, et al. Systematic review of the impact of enhanced recovery pathways in elective lung resection. The Journal of thoracic and cardiovascular surgery Epub ahead of print.
36. McKenna RJ, Jr., Fischel RJ, Brenner M, Gelb AF. Use of the Heimlich valve to shorten hospital stay after lung reduction surgery for emphysema. The Annals of thoracic surgery 1996;61:1115-7.
37. Ponn RB, Silverman HJ, Federico JA. Outpatient chest tube management. The Annals of thoracic surgery 1997;64:1437-40.

38. Cerfolio RJ, Pickens A, Bass C, Katholi C. Fast-tracking pulmonary resections. The Journal of thoracic and cardiovascular surgery 2001;122:318-24.
39. Kaneda H, Saito Y, Okamoto M, Maniwa T, Minami K, Imamura H. Early postoperative mobilization with walking at 4 hours after lobectomy in lung cancer patients. General thoracic and cardiovascular surgery 2007;55:493-8.
40. Agostini P, Calvert R, Subramanian H, Naidu B. Is incentive spirometry effective following thoracic surgery? Interactive cardiovascular and thoracic surgery 2008;7:297-300.
41. Muehling BM, Halter GL, Schelzig H, et al. Reduction of postoperative pulmonary complications after lung surgery using a fast track clinical pathway. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2008;34:174-80.
42. Maruyama R, Miyake T, Kojo M, et al. Establishment of a clinical pathway as an effective tool to reduce hospitalization and charges after video-assisted thoracoscopic pulmonary resection. The Japanese journal of thoracic and cardiovascular surgery : official publication of the Japanese Association for Thoracic Surgery = Nihon Kyobu Geka Gakkai zasshi 2006;54:387-90.
43. Zehr KJ, Dawson PB, Yang SC, Heitmiller RF. Standardized clinical care pathways for major thoracic cases reduce hospital costs. The Annals of thoracic surgery 1998;66:914-9.
44. Carli F, Charlebois P, Baldini G, Cachero O, Stein B. An integrated multidisciplinary approach to implementation of a fast-track program for laparoscopic colorectal surgery. Canadian journal of anaesthesia = Journal canadien d'anesthesie 2009;56:837-42.
45. Cerfolio RJ, Bryant AS. Results of a prospective algorithm to remove chest tubes after pulmonary resection with high output. The Journal of thoracic and cardiovascular surgery 2008;135:269-73.
46. Zhang Y, Li H, Hu B, et al. A prospective randomized single-blind control study of volume threshold for chest tube removal following lobectomy. World journal of surgery 2014;38:60-7.
47. WHO. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948 http://www.who.int/about/definition/en/print.html.
48. Allison PJ, Locker D, Feine JS. Quality of life: A dynamic construct. Social Science & Medicine 1997;45:221-30.
49. Group TW. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Social science & medicine (1982) 1995;41:1403-9.
50. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 1993;2:451-9.
51. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 1993;2:367-8.
52. Patrick DL EP. Health Status and Health Policy: Health status and health policy- quality of life in health care evaluation and resource allocation. New York, New York: Oxford University Press 1993.
53. Schulte T, Schniewind B, Walter J, Dohrmann P, Kuchler T, Kurdow R. Age-related impairment of quality of life after lung resection for non-small cell lung cancer. Lung cancer (Amsterdam, Netherlands) 2010;68:115-20.
54. Ferguson MK, Parma CM, Celauro AD, Vigneswaran WT. Quality of life and mood in older patients after major lung resection. The Annals of thoracic surgery 2009;87:1007-12; discussion 12-3.
55. Bornbaum CC, Doyle PC, Skarakis-Doyle E, Theurer JA. A critical exploration of the International Classification of Functioning, Disability, and Health (ICF) framework from the

perspective of oncology: recommendations for revision. Journal of multidisciplinary healthcare 2013;6:75-86.
56. McNeil BJ, Weichselbaum R, Pauker SG. Speech and survival: tradeoffs between quality and quantity of life in laryngeal cancer. The New England journal of medicine 1981;305:982-7.
57. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Medical care 1989;27:S217-32.
58. Mayo N GM, Kind P. Performance of the EuroQOL EQ-5D in a Canadian Population. EuroQOL Plenary Meeting (Rotterdam) 1997.
59. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. Journal of health services research & policy 1999;4:174-84.
60. Meyer T, Deck R, Raspe H. Problems completing questionnaires on health status in medical rehabilitation patients. Journal of rehabilitation medicine 2007;39:633-9.
61. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Social science & medicine (1982) 1999;48:1531-48.
62. Guyatt GH, Feeny DH, Patrick DL. Measuring Health--related Quality of Life. Annals of Internal Medicine 1993;118:622.
63. Wright JG, Young NL. A comparison of different indices of responsiveness. Journal of clinical epidemiology 1997;50:239-46.
64. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. Journal of clinical epidemiology 2003;56:52-60.
65. Neville A, Lee L, Antonescu I, et al. Systematic review of outcomes used to evaluate enhanced recovery after surgery. The British journal of surgery 2014;101:159-70.
66. Blanchard C, Feeny D, Mahon JL, et al. Is the Health Utilities Index responsive in total hip arthroplasty patients? Journal of clinical epidemiology 2003;56:1046-54.
67. Shields RK, Enloe LJ, Leo KC. Health related quality of life in patients with total hip or knee replacement. Archives of physical medicine and rehabilitation 1999;80:572-9.
68. Lee L, Dumitra T, Fiore JF, Jr., Mayo NE, Feldman LS. How well are we measuring postoperative "recovery" after abdominal surgery? Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2015.
69. Kluivers KB, Riphagen I, Vierhout ME, Brölmann HAM, de Vet HCW. Systematic review on recovery specific quality-of-life instruments. Surgery 2008;143:206-15.
70. Herrera FJ, Wong J, Chung F. A systematic review of postoperative recovery outcomes measurements after ambulatory surgery. Anesth Analg 2007;105:63-9.
71. Damm K, Roeske N, Jacob C. Health-related quality of life questionnaires in lung cancer trials: a systematic literature review. Health economics review 2013;3:15.
72. Carolina Casañas i Comabella EG, Ray Fitzpatrick. A structured review of patient-reported outcome measures for patients with lung cancer. Patient-reported Outcome Measurement Group, Department of Public Health, University of Oxford 2010.
73. Mandy van Reenen MO. EQ-5D-3L User Guide - Basic information on how to use the EQ-5D-3L instrument2015.
74. Who is using EQ-5D (Examples of clinical areas where EQ-5D is used). 2016. 2016, at http://www.euroqol.org/about-eq-5d/how-to-use-eq-5d/who-is-using-eq-5d.html.)
75. McDowell I. Measuring Health: A Guide to Rating Scales and Questionnaires, Third Edition. Oxford University Press 2006.
76. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Annals of medicine 2001;33:337-43.
77. Jones C, Kelliher L, Dickinson M, et al. Randomized clinical trial on enhanced recovery versus standard care following open liver resection. The British journal of surgery 2013;100:1015-24.

78. Cabral DL, Laurentino GE, Damascena CG, Faria CD, Melo PG, Teixeira-Salmela LF. Comparisons of the Nottingham Health Profile and the SF-36 health survey for the assessment of quality of life in individuals with chronic stroke. Revista brasileira de fisioterapia (Sao Carlos (Sao Paulo, Brazil)) 2012;16:301-8.
79. Eshaghi SR, Ramezani MA, Shahsanaee A, Pooya A. Validity and Reliability of the Short Form-36 Items Questionnaire as a Measure of Quality of Life in Elderly Iranian Population. AMERICAN JOURNAL OF APPLIED SCIENCES 2006;3:1763-6.
80. Delaney CP, Zutshi M, Senagore AJ, Remzi FH, Hammel J, Fazio VW. Prospective, randomized, controlled trial between a pathway of controlled rehabilitation with early ambulation and diet and traditional postoperative care after laparotomy and intestinal resection. Diseases of the colon and rectum 2003;46:851-9.
81. Doering LV, Moser DK, Lemankiewicz W, Luper C, Khan S. Depression, healing, and recovery from coronary artery bypass surgery. American journal of critical care : an official publication, American Association of Critical-Care Nurses 2005;14:316-24.
82. Anderson JP, Kaplan RM, Berry CC, Bush JW, Rumbaut RG. Interday reliability of function assessment for a health status measure. The Quality of Well-Being scale. Medical care 1989;27:1076-83.
83. McEntee DJ, Badenhop DT. Quality of life comparisons: gender and population differences in cardiopulmonary rehabilitation. Heart & lung : the journal of critical care 2000;29:340-7.
84. Vlug MS, Wind J, Hollmann MW, et al. Laparoscopy in combination with fast track multimodal management is the best perioperative strategy in patients undergoing colonic surgery: a randomized clinical trial (LAFA-study). Ann Surg 2011;254:868-75.
85. Doll HA, Black NA, Flood AB, McPherson K. Criterion validation of the Nottingham Health Profile: patient views of surgery for benign prostatic hypertrophy. Social science & medicine (1982) 1993;37:115-22.
86. Falcoz PE, Chocron S, Mercier M, Puyraveau M, Etievent JP. Comparison of the Nottingham Health Profile and the 36-item health survey questionnaires in cardiac surgery. The Annals of thoracic surgery 2002;73:1222-8.
87. Lipsett PA, Swoboda SM, Campbell KA, Cornwell E, 3rd, Dorman T, Pronovost PJ. Sickness Impact Profile Score versus a Modified Short-Form survey for functional outcome assessment: acceptability, reliability, and validity in critically ill patients with prolonged intensive care unit stays. The Journal of trauma 2000;49:737-43.
88. Group E. EuroQol--a new facility for the measurement of health-related quality of life. Health policy (Amsterdam, Netherlands) 1990;16:199-208.
89. Brooks R. EuroQol: the current state of play. Health policy (Amsterdam, Netherlands) 1996;37:53-72.
90. Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. PloS one 2012;7:e31115.
91. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Medical care 2005;43:203-20.
92. Heijink R, van Baal P, Oppe M, Koolman X, Westert G. Decomposing cross-country differences in quality adjusted life expectancy: the impact of value sets. Population health metrics 2011;9:17.
93. McCarthy M, Jr., Chang CH, Pickard AS, et al. Visual analog scales for assessing surgical pain. Journal of the American College of Surgeons 2005;201:245-52.
94. Scott J, Huskisson EC. Graphic representation of pain. Pain 1976;2:175-84.
95. Gaston-Johansson F, Gustafsson M. Rheumatoid arthritis: determination of pain characteristics and comparison of RAI and VAS in its measurement. Pain 1990;41:35-40.
96. Jensen MP, McFarland CA. Increasing the reliability and validity of pain intensity measurement in chronic pain patients. Pain 1993;55:195-203.

97. Vogelsang J. The Visual Analog Scale: an accurate and sensitive method for self-reporting preoperative anxiety. Journal of post anesthesia nursing 1988;3:235-9.
98. Millar K, Jelicic M, Bonke B, Asbury AJ. Assessment of preoperative anxiety: comparison of measures in patients awaiting surgery for breast cancer. British journal of anaesthesia 1995;74:180-3.
99. Boogaerts JG, Vanacker E, Seidel L, Albert A, Bardiau FM. Assessment of postoperative nausea using a visual analogue scale. Acta Anaesthesiologica Scandinavica 2000;44:470-4.
100. Schwenk W, Bohm B, Muller JM. Postoperative pain and fatigue after laparoscopic or conventional colorectal resections. A prospective randomized trial. Surgical endoscopy 1998;12:1131-6.
101. Gaston-Johansson F, Fall-Dickson JM, Bakos AB, Kennedy MJ. Fatigue, pain, and depression in pre-autotransplant breast cancer patients. Cancer practice 1999;7:240-7.
102. Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. Journal of epidemiology / Japan Epidemiological Association 2002;12:160-78.
103. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res 1972;7:118-33.
104. Dolan P. Utilitarianism and the measurement and aggregation of quality--adjusted life years. Health care analysis : HCA : journal of health philosophy and policy 2001;9:65-76.
105. Taheri PA, Butz DA, Greenfield LJ. Length of stay has minimal impact on the cost of hospital admission. Journal of the American College of Surgeons 2000;191:123-30.
106. Fiore JF BJ, Conrad K, Niculiseanu P, Landry T, Lee L, Mulder DS, Ferri LE. Feldman LS. Systematic review of the impact of enhanced recovery pathways in elective lung resection. Submitted to JTCVS.
107. Lee L, Mata J, Ghitulescu GA, et al. Cost-effectiveness of Enhanced Recovery Versus Conventional Perioperative Management for Colorectal Surgery. Ann Surg 2015.
108. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials 1989;10:407-15.
109. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. The Journal of rheumatology 1993;20:429-36.
110. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. Journal of Chronic Diseases 1986;39:897-906.
111. Deyo RA, Inui TS. Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. Health Services Research 1984;19:275-89.
112. Meenan RF, Anderson JJ, Kazis LE, et al. Outcome assessment in clinical trials evidence for the sensitivity of a health status measure. Arthritis & Rheumatism 1984;27:1344-52.
113. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Medical care 2000;38:II84-90.
114. Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. Expert Rev Pharmacoecon Outcomes Res 2014;14:221-33.
115. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2005;14:1523-32.
116. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and quality of life outcomes 2007;5:70.
117. Westen D, Rosenthal R. Improving construct validity: Cronbach, Meehl, and Neurath's ship. Psychological assessment 2005;17:409-12.

118. David L.Streiner GRN. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford Medical Publications 1995.
119. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 1959;56:81-105.
120. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I: Applications and issues in assessment. BMJ (Clinical research ed) 1992;305:1074-7.
121. American Educational Research Association APA, & National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association 2014.
122. Goodwin LD. Changing conceptions of measurement validity: an update on the new standards. Journal of Nursing Education 2002;41:100-6.
123. Handy JR, Jr., Asaph JW, Douville EC, Ott GY, Grunkemeier GL, Wu Y. Does video-assisted thoracoscopic lobectomy for lung cancer provide improved functional outcomes compared with open lobectomy? European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2010;37:451-5.
124. Leo F, Scanagatta P, Vannucci F, Brambilla D, Radice D, Spaggiari L. Impaired quality of life after pneumonectomy: who is at risk? The Journal of thoracic and cardiovascular surgery 2010;139:49-52.
125. Jones NL, Edmonds L, Ghosh S, Klein AA. A review of enhanced recovery for thoracic anaesthesia and surgery. Anaesthesia 2013;68:179-89.
126. Whitson BA, Groth SS, Duval SJ, Swanson SJ, Maddaus MA. Surgery for early-stage non-small cell lung cancer: a systematic review of the video-assisted thoracoscopic surgery versus thoracotomy approaches to lobectomy. The Annals of thoracic surgery 2008;86:2008-16; discussion 16-8.
127. Van Schil P. Cost analysis of video-assisted thoracic surgery versus thoracotomy: critical review. The European respiratory journal 2003;22:735-8.
128. Varadhan KK, Lobo DN, Ljungqvist O. Enhanced recovery after surgery: the future of improving surgical care. Critical care clinics 2010;26:527-47, x.
129. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA : the journal of the American Medical Association 1996;276:1172-7.
130. Sorenson C, Drummond M, Chalkidou K. Comparative effectiveness research: the experience of the National Institute for Health and Clinical Excellence. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2012;30:4267-74.
131. Robinson R. Cost-utility analysis. BMJ (Clinical research ed) 1993;307:859-62.
132. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. Medical care 2002;40:442-6.
133. Devlin NJ, Krabbe PF. The development of new research methods for the valuation of EQ-5D-5L. The European journal of health economics : HEPAC : health economics in prevention and care 2013;14 Suppl 1:S1-3.
134. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. Journal of clinical epidemiology 2003;56:317-25.
135. Dyer MT, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. Health and quality of life outcomes 2010;8:13.
136. Seely AJ, Ivanovic J, Threader J, et al. Systematic classification of morbidity and mortality after thoracic surgery. The Annals of thoracic surgery 2010;90:936-42; discussion 42.

137. Brunelli A, Socci L, Refai M, Salati M, Xiume F, Sabbatini A. Quality of life before and after major lung resection for lung cancer: a prospective follow-up analysis. The Annals of thoracic surgery 2007;84:410-6.
138. Balduyck B, Hendriks J, Lauwers P, Van Schil P. Quality of life evolution after lung cancer surgery: a prospective study in 100 patients. Lung cancer (Amsterdam, Netherlands) 2007;56:423-31.
139. Balduyck B, Hendriks J, Lauwers P, Sardari Nia P, Van Schil P. Quality of life evolution after lung cancer surgery in septuagenarians: a prospective study. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2009;35:1070-5; discussion 5.
140. Salati M, Brunelli A, Xiume F, Refai M, Sabbatini A. Quality of life in the elderly after major lung resection for lung cancer. Interactive cardiovascular and thoracic surgery 2009;8:79-83.
141. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in medicine 2011;30:377-99.
142. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. American journal of epidemiology 2004;160:34-45.
143. Munro BH. Statistical Methods for Health Care Research, 3rd ed. Philadelphia: Lippincott.
144. Bharmal M, Thomas J, 3rd. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2006;9:262-71.
145. Kim SH, Jo MW, Kim HJ, Ahn JH. Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. Health and quality of life outcomes 2012;10:151.
146. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2008;17:485-95.
147. Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DP. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. The Journal of rheumatology 2008;35:592-602.
148. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health economics 2004;13:873-84.
149. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2008;11:275-84.
150. Lee L, Elfassy N, Li C, et al. Valuing postoperative recovery: validation of the SF-6D health-state utility. The Journal of surgical research 2013;184:108-14.
151. Moriello C, Mayo NE, Feldman L, Carli F. Validating the six-minute walk test as a measure of recovery after elective colon resection surgery. Archives of physical medicine and rehabilitation 2008;89:1083-9.
152. Lee L, Mata J, Augustin BR, et al. A comparison of the validity of two indirect utility instruments as measures of postoperative recovery. The Journal of surgical research 2014;190:79-86.
153. Wu AW, Jacobson KL, Frick KD, et al. Validity and responsiveness of the euroqol as a measure of health-related quality of life in people enrolled in an AIDS clinical trial. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2002;11:273-82.
154. Norum J, Angelsen V, Wist E, Olsen JA. Treatment costs in Hodgkin's disease: a cost-utility analysis. European journal of cancer (Oxford, England : 1990) 1996;32A:1510-7.

155. Shaw JW, Pickard AS, Yu S, et al. A median model for predicting United States population-based EQ-5D health state preferences. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2010;13:278-88.
156. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2011;20:1727-36.
157. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2013;22:1717-27.
158. Goossens LM, Nivens MC, Sachs P, Monz BU, Rutten-van Molken MP. Is the EQ-5D responsive to recovery from a moderate COPD exacerbation? Respiratory medicine 2011;105:1195-202.
159. Chin Lim W. Understanding the concept of health related quality of life in adult, general critical care survivors [Doctoral]. http://researchonline.lshtm.ac.uk/682439/: London School of Hygiene & Tropical Medicine; 2011.
160. Longworth L, Rowen, D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. Available from http://wwwnicedsuorguk 2011.
161. Bartels SA, Vlug MS, Ubbink DT, Bemelman WA. Quality of life after laparoscopic and open colorectal surgery: a systematic review. World journal of gastroenterology 2010;16:5035-41.
162. Montazeri A, Gillis CR, McEwen J. Quality of life in patients with lung cancer: a review of literature from 1970 to 1995. Chest 1998;113:467-81.
163. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. Journal of clinical epidemiology 1998;51:903-12.
164. Mangione CM, Goldman L, Orav EJ, et al. Health-related quality of life after elective surgery: measurement of longitudinal changes. Journal of general internal medicine 1997;12:686-97.
165. Longworth L, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health economics 2003;12:1061-7.
166. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of health economics 2002;21:271-92.
167. Kontodimopoulos N. The potential for a generally applicable mapping model between QLQ-C30 and SF-6D in patients with different cancers: a comparison of regression-based methods. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2015;24:1535-44.
168. Kontodimopoulos N, Aletras VH, Paliouras D, Niakas D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2009;12:1151-7.
169. Goodwin PC, Ratcliffe J, Morris J, Morrissey MC. Using the knee-specific Hughston Clinic Questionnaire, EQ-5D and SF-6D following arthroscopic partial meniscectomy surgery: a comparison of psychometric properties. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2011;20:1437-46.
170. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI(®)): concepts, measurement properties and applications. Health and quality of life outcomes 2003;1:54-.
171. Bélanger A BJ-M, Guimond E, et al. A head-to-head comparison of two generic health status measures in the household population: McMaster Health Utilities Index (Mark3) and the EQ-5D. Statistics Canada Report Ottawa, Ontario 2002;1-62.

172. Sieber WJ GE, David KM, Ganiats TG, Kaplan RM. Quality of Well Being Self-Administered (QWB-SA) Scale - User's Manual. Health Services Research Center, University of California, San Diego 2008.
173. Kaplan RM, Ganiats TG, Sieber WJ, Anderson JP. The Quality of Well-Being Scale: critical similarities and differences with SF-36. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua 1998;10:509-20.
174. Berg K, Idvall E, Nilsson U, Årestedt KF, Unosson M. Psychometric evaluation of the post-discharge surgical recovery scale. Journal of Evaluation in Clinical Practice 2010;16:794-801.
175. Kleinbeck SV. Self-reported at-home postoperative recovery. Research in nursing & health 2000;23:461-72.
176. Myles PS, Hunt JO, Nightingale CE, et al. Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. Anesth Analg 1999;88:83-90.
177. Myles PS, Weitkamp B, Jones K, Melick J, Hensen S. Validity and reliability of a postoperative quality of recovery score: the QoR-40. British journal of anaesthesia 2000;84:11-5.
178. Gornall BF, Myles PS, Smith CL, et al. Measurement of quality of recovery using the QoR-40: a quantitative systematic review. British journal of anaesthesia 2013;111:161-9.
179. Ambrogi V, Mineo TC. VATS biopsy for undetermined interstitial lung disease under nongeneral anesthesia: comparison between uniportal approach under intercostal block vs . three-ports in epidural anesthesia. Journal of Thoracic Disease 2014;6:888-95.
180. Guilherme Dal Agnol JB, Yves Lacasse, Nathalie Gagné, Paula A Ugalde. Randomized Study of Preoperative Dexamethasone for Quality of Recovery in VATS Lung Resection Patients (APV). ClinicalTrialsgov Identifier: NCT02275702 (ongoing recruitment).

EQ5D

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

MOBILITY
I have no problems in walking about ❑

I have some problems in walking about ❑

I am confined to bed ❑

SELF-CARE
I have no problems with self-care ❑

I have some problems washing or dressing myself ❑

I am unable to wash or dress myself ❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*
I have no problems doing my usual activities ❑

I have some problems with performing my usual activities ❑

I am unable to perform my usual activities ❑

PAIN / DISCOMFORT
I have no pain or discomfort ❑

I have moderate pain or discomfort ❑

I have extreme pain or discomfort ❑

ANXIETY / DEPRESSION
I am not anxious or depressed ❑

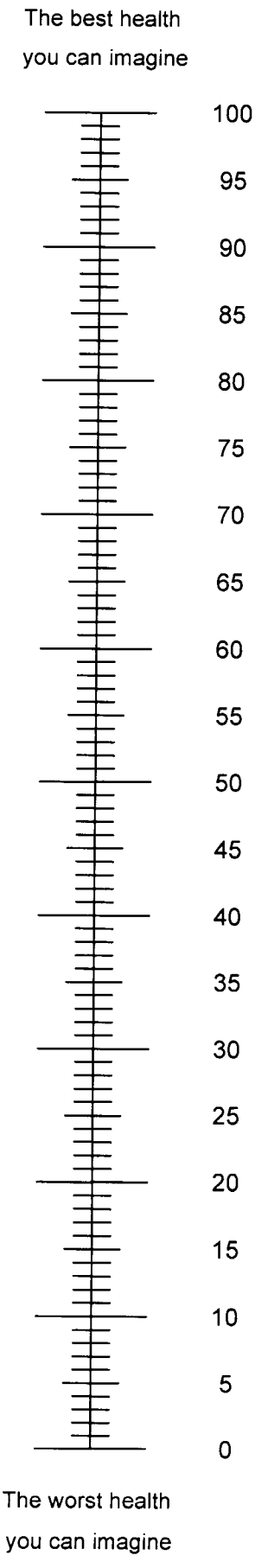
I am moderately anxious or depressed ❑

I am extremely anxious or depressed ❑

© 1990 EuroQol Group EQ-5D™ is a trade mark of the EuroQol Group

- To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.
- We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today

YOUR HEALTH TODAY = ☐



© 1990 EuroQol Group EQ-5D™ is a trade mark of the EuroQol Group

EQ5D

Pour chaque catégorie, cochez l'énoncé qui décrit le mieux votre état de santé actuel.

MOBILITÉ

Je n'ai aucune difficulté à marcher ❑

J'ai de la difficulté à marcher ❑

Je suis obligé(e) de rester au lit ❑

SOINS AUTONOMES

Je n'ai pas de difficulté à prendre soin de moi-même ❑

J'ai de la difficulté à me laver ou à m'habiller seul(e) ❑

Je suis incapable de me laver ou de m'habiller seul(e) ❑

ACTIVITÉS HABITUELLES *(ex. travail, études, tâches ménagères, activités familiales ou loisirs)*

Je n'ai aucune difficulté à faire mes activités habituelles ❑

J'ai de la difficulté à faire mes activités habituelles ❑

Je suis incapable de faire mes activités habituelles ❑

DOULEUR/MALAISES

Je ne ressens pas de douleur ou de malaise ❑

Je ressens des douleurs ou des malaises légers ❑

Je ressens des douleurs ou des malaises intenses ❑

ANXIÉTÉ / DÉPRESSION

Je suis ni anxieux(se), ni déprimé(e) ❑

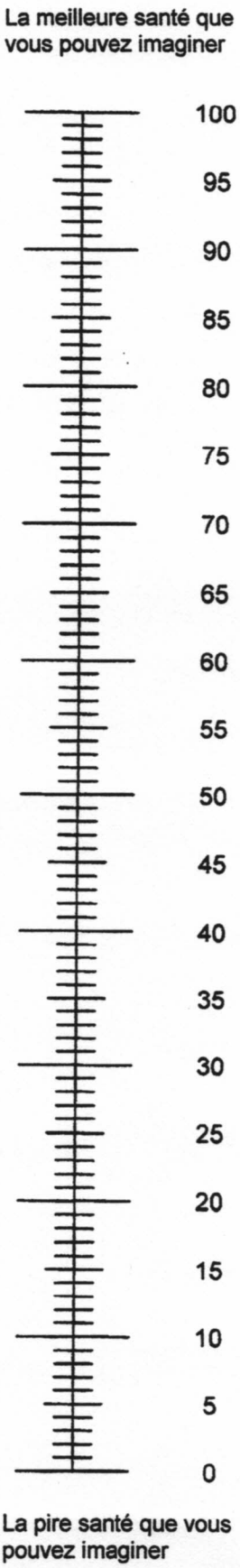
Je suis légèrement anxieux(se) ou déprimé(e) ❑

Je suis très anxieux(se) ou déprimé(e) ❑

© 1990 EuroOol Group EQ-SDTM is a trade mark of the EuroQol Group

- Pour vous aider à exprimer votre état de santé, nous vous présentons une échelle (qui ressemble à un thermomètre) sur laquelle le meilleur état de santé imaginable correspond à 100 tandis que le pire état de santé imaginable correspond à 0.
- Nous voudrions que vous indiquiez sur cette échelle, selon votre opinion, comment bonne ou mauvaise est votre santé aujourd'hui. Faire cela en traçant une ligne de la boîte ci-dessous jusqu'à n'importe quel point de l'échelle qui indique le niveau de votre santé aujourd'hui.

VOTRE SANTÉ AUJOURD'HUI = ☐

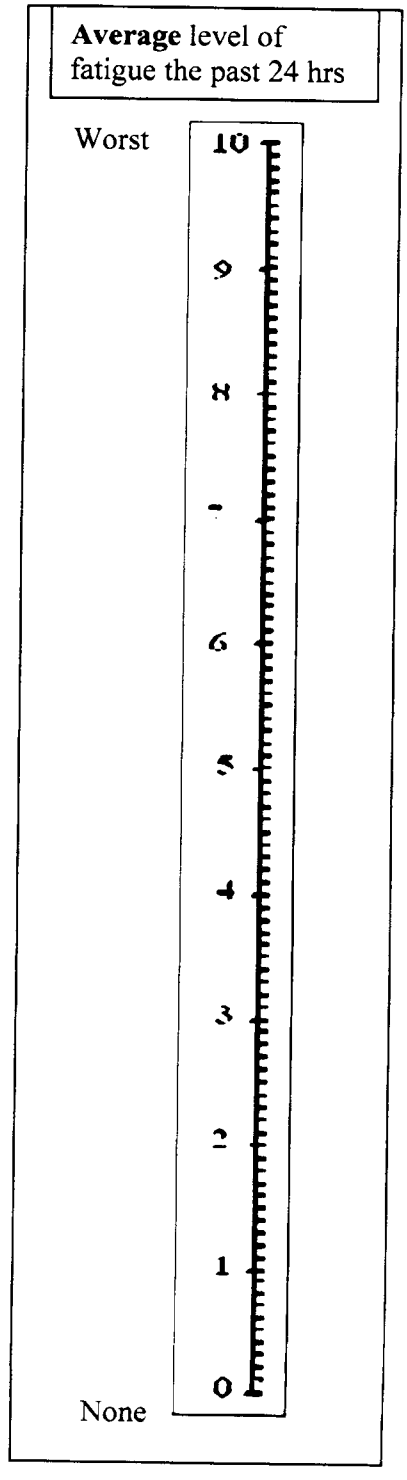


© 1990 EuroQol Group EQ-5D™ is a trade mark of the EuroQol Group

8.2 – Fatigue VAS

Fatigue Scale

We would like to know your level of your fatigue. We ask you to indicate **your average level** of fatigue in the past 24 hours. Please put a line through the scales below to indicate the level of fatigue. The higher numbers indicate more fatigue. **0** indicates that you do not experience fatigue.



8.3 – Pain VAS

Section 11

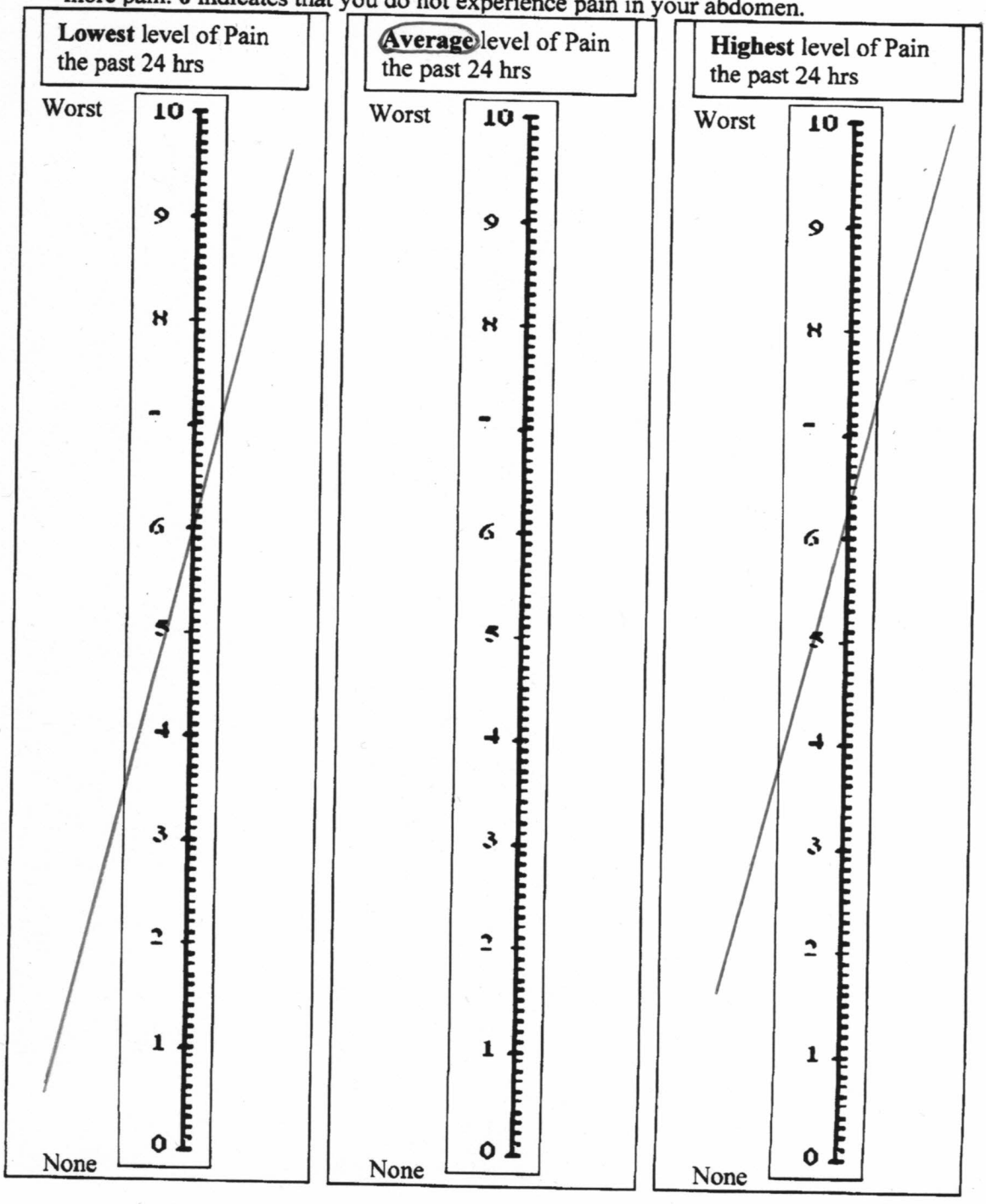
Patient ID:..... Date:.....

Pain Scale

We would like to know your level of Pain ~~arising from your abdomen (belly).~~

We ask you to indicate ~~lowest,~~ **average** ~~and highest~~ levels of pain in the past 24 hours.

Please put a line through the scales below to indicate the level of pain. The higher numbers indicate more pain. **0** indicates that you do not experience pain in your abdomen.



8.3 – Consent Form: English / French

INFORMATION AND CONSENT FORM

STUDY TITLE: Impact of the McGill Multimodal Surgical Recovery (SURE) Pathway on Perioperative Processes and Outcomes

FUNDED BY: The Steinberg-Bernstein Centre for Minimally Invasive Surgery

PRINCIPAL INVESTIGATOR: Dr. Liane S. Feldman

CO-INVESTIGATORS: Dr Franco Carli; Dr Lorenzo Ferri; Dr. David S. Mulder; Dr. Christian Sirois; Pepa Kaneva MSc; Annie Ncuti BSc.

PARTICIPATING INSTITUTIONS: McGill University Health Centre (MUHC).

INTRODUCTION

You are being invited to participate in this study because you are going to undergo an operation.

Before deciding to participate in the study, you should clearly understand its requirements, the risks and benefits. This document provides information about the study, and it may contain words you do not fully understand. Please read it carefully and ask the study staff any questions you may have. They will discuss the study with you in detail. You may take this form with you and discuss the study with anyone else before making your decision. If you decide to participate, you will be asked to sign this form and a copy will be given to you. Participation in this study will not interfere or change the preoperative and postoperative care given to the patient.

PURPOSE OF THE STUDY

The primary goal of this study is to evaluate how health-care programs in Montreal General Hospital affect recovery, level of pain and health-related quality of life after thoracic (surgery in the chest) and abdominal surgery. The secondary goal is to evaluate the health-related cost associated with these operations.

STUDY DESCRIPTION

The technique for your operation has been well established for many years. New recovery programs have been developed aiming to shorten the recovery time by decreasing pain, discomfort, and postoperative nausea and vomiting. These programs include the preoperative (before surgery), intraoperative (during surgery), and postoperative (after surgery) periods. This study will enable the researchers to evaluate how effective these programs are. To do this, we will ask you to fill out questionnaires evaluating your quality of life, physical activity, anxiety, fatigue, nausea and pain levels before and after the operation.

STUDY PROCEDURES

If you agree to participate in this study, you will be asked to do the following evaluations before and after your surgery. The research team will consult your medical file to take note of information relevant data to this research project. The information collected will be gender, age, weight, details about your medical history and the current surgery, and health services used

to estimate the cost of these services. We will also ask you about if someone, for example, a family member or a nurse, help you in the period before and after the operation.

Before surgery

You will be asked to complete several short questionnaires that assess: - how much pain and nausea you have; - how tired you feel; how much help you need before the operation; - what is your overall general health; - what food can you eat.

After Surgery - In Hospital

During your hospital stay you will be asked to fill in a short questionnaire to measure: - how much pain and nausea you have; - how tired you feel; - how much time you spend walking or sleeping; - what is your overall general health; - what food can you eat.

After Surgery - 30 and 90 days after surgery

Over the telephone, you will be asked to complete the questionnaires to measure: - how much pain and nausea you have; - how tired you feel; - what is your overall general health; - what food can you eat; how much help you needed after the hospital discharge and what additional health care services you used.

The preoperative, 30 and 90 days evaluations will take between 15-20 minutes to complete. The in-hospital evaluations will take between 5-10 minutes per day.

POTENTIAL RISKS AND DISCOMFORTS

There are no anticipated risks associated with participating in this study.

POSSIBLE BENEFITS

You are not expected to directly benefit from participating in this study. However, the information gathered from this study may educate the medical field and provide information for future subjects undergoing such procedures.

COMPENSATION

You will not receive any financial compensation to participate in this study.

VOLUNTARY PARTICIPATION

Your participation in this study is entirely voluntary. You may refuse to participate or you may discontinue your participation at any time without explanation, and without penalty or loss of benefits to which you are otherwise entitled. If you decide not to participate, or if you discontinue your participation, you will suffer no prejudice regarding your medical care or your participation in any other research studies.

CONFIDENTIALITY

All information obtained during this study will be kept strictly confidential. Your name will be coded and the code list will be locked in a filing cabinet in the investigator's office with only the investigators having access. The records will be kept for seven years and then they will be destroyed. The results from this study may be published; however, your identity will not be revealed in the combined results. In order to verify the research study data, monitors from the Quality Assurance Officer at the MUHC-Research Ethics Boards may review these records.

SIGNIFICANT FINDINGS

During the course of this study, investigators may generate new research findings from this cohort or new information about an individual participant. The research findings will be shared with you and you are welcome to discuss these findings with your treating physician or the investigators.

QUESTIONS/CONTACT INFORMATION

If you have questions regarding the study, you should contact:
Dr. Liane Feldman, at (514) 934-1934, ext. 44004. If you have any questions about your rights as a study participant, you should contact the hospital Ombudsman at (514) 934-1934, ext 48306.

DECLARATION OF CONSENT

I have read the contents of this consent form, and I agree to participate in this research study. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction. I have been given sufficient time to consider the above information and to seek advice if I choose to do so. I will be given a copy of this signed consent form. By signing this consent form, I have not given up any of my legal rights.

_____

Participant's signature Date (M/D/Y)

_____

Participant's name (please print)

_____

Signature of Person Obtaining Consent Date (M/D/Y)

_____

Name of Person Obtaining Consent

FORMULAIRE DE RENSEIGNEMENTS ET DE CONSENTEMENT ÉCLAIRÉ

TITRE DE L'ÉTUDE : Effet du parcours multimodal de rétablissement chirurgical de McGill sur les processus péri-opératoires et les résultats

FINANCÉE PAR : Le Centre de chirurgie à invasion minimale Steinberg-Bernstein

CHERCHEUSE PRINCIPALE : Dre Liane S. Feldman

CHERCHEURS ASSOCIÉS : Dr Franco Carli; Dr Lorenzo Ferri; Dr. David S. Mulder; Dr. Christian Sirois; Pepa Kaneva MSc; Annie Ncuti BSc.

INSTITUTIONS PARTICIPANTES : Centre universitaire de santé McGill (CUSM)

INTRODUCTION

Vous êtes invité à participer à cette étude parce que vous allez subir une chirurgie. Avant de décider de participer à l'étude, vous devriez clairement comprendre les exigences, les risques et les avantages qui s'y rattachent. Ce document fournit des informations au sujet de l'étude; il est possible qu'il renferme des mots que vous ne comprenez pas clairement. Veuillez le lire attentivement et poser toutes les questions que vous pourriez avoir au personnel de l'étude. Ils discuteront de l'étude en détail avec vous. Vous pouvez apporter une copie de ce document avec vous et en discuter avec la personne de votre choix avant de prendre votre décision. Si vous décidez de participer, nous vous demanderons de signer ce formulaire de consentement et une copie vous sera remise. La participation à cette l'étude n'affectera ni ne changera les soins qui vous seront prodigués avant et après la chirurgie.

BUT DE L'ÉTUDE

L'objectif principal de cette étude est d'évaluer l'effet des programmes de soins de santé sur le rétablissement, le niveau de douleur, ainsi que la qualité de vie des patients après une chirurgie thoracique (chirurgie du thorax) et une chirurgie abdominale. Le second objectif est d'évaluer les coûts en soins de santé liés à de telles opérations.

DESCRIPTION DE L'ÉTUDE

La technique utilisée pour votre chirurgie est employée depuis de nombreuses années. De nouveaux programmes de rétablissement ont été développé afin de raccourcir la période de convalescence en réduisant la douleur, l'inconfort ainsi que les épisodes postopératoires de nausée et vomissement. Ces programmes sont divisés en périodes préopératoire (avant l'opération), intra-opératoire (pendant l'opération) et postopératoire (après l'opération). Cette étude permettra aux chercheurs d'évaluer l'efficacité de ces programmes. Pour ce faire, nous vous demanderons de compléter, avant et après votre opération, des questionnaires évaluant votre qualité de vie, vos activités physiques, votre niveau d'anxiété, de fatigue, de nausée et de douleur.

PROCÉDURES DE L'ÉTUDE

Si vous acceptez de participer à cette étude, nous vous demanderons de remplir des questionnaires avant et après votre chirurgie. L'équipe de recherche consultera votre dossier médical pour recueillir des informations pertinentes pour cette étude, telles que : sexe, âge, poids ainsi que des détails au sujet de vos antécédents médicaux et la chirurgie que vous allez subir. Aussi, l'équipe prendra note de votre usage des services de santé afin d'estimer les coûts qui y sont reliés. Nous demanderons également si quelqu'un, par exemple un membre de la famille ou une infirmière, vous a aidé avant ou après la chirurgie.

Avant la chirurgie

Nous vous demanderons de remplir plusieurs courts questionnaires qui évaluent : votre niveau de douleur et de nausée, votre niveau de fatigue, le type d'assistance dont vous avez eu besoin avant la chirurgie, votre état de santé général et votre régime alimentaire.

Après la chirurgie : à l'hôpital

Durant votre séjour à l'hôpital nous vous demanderons de remplir plusieurs courts questionnaires qui évaluent : votre niveau de douleur et de nausée, votre niveau de fatigue, le temps passé à marcher ou à dormir, votre état de santé général et votre régime alimentaire.

30 et 90 jours après la chirurgie

Nous vous demanderons de remplir, au téléphone, des questionnaires évaluant: votre niveau de douleur et de nausée, votre niveau de fatigue, votre état de santé général, votre régime alimentaire, le type d'assistance dont vous avez eu besoin après votre congé de l'hôpital ainsi que les services de santé que vous avez reçus.

Les évaluations en période préopératoires ainsi que celles à 30 et 90 jours après la chirurgie pourront durer entre 15 à 20 minutes. Quant aux évaluations durant votre hospitalisation, elles prendront entre 5 à 10 minutes par jour.

RISQUES ET INCONFORTS POTENTIELS

Nous ne prévoyons aucun risque associé à votre participation à cette étude.

BÉNÉFICES POTENTIELS

Vous n'obtiendrez aucun bénéfice direct de votre participation à cette étude. Toutefois, les informations recueillies à la suite de cette étude pourraient s'avérer éducatives pour le domaine médical et fournir d'importantes informations pour les patients qui subiront ce type de chirurgie dans le futur.

COMPENSATION

Vous ne recevrez aucune compensation financière pour votre participation à cette étude.

PARTICIPATION VOLONTAIRE

Votre participation à cette étude est entièrement volontaire. Vous pouvez refuser de participer ou encore vous retirer de l'étude à tout moment sans justification. Ceci n'entraînera aucune pénalité ou perte d'avantages auxquels vous avez normalement droit. Si vous décidez de ne pas participer, ou si vous retirez votre participation, vous ne subirez aucun préjudice concernant vos soins médicaux ou votre participation à d'autres études de recherche.

CONFIDENTIALITÉ

Toute information obtenue pendant l'étude demeurera strictement confidentielle. Votre nom sera codé, la liste des codes sera conservée sous clé dans le bureau du chercheur et seuls les chercheurs y auront accès. Les dossiers seront conservés pendant sept ans avant d'être détruits. Les résultats de cette étude pourraient être publiés; toutefois, votre identité ne sera pas révélée dans les résultats combinés. Dans le but de vérifier les données de l'étude, les responsables du Bureau de l'assurance qualité du Comité d'éthique de la recherche du CUSM pourraient consulter ces dossiers.

INFORMATIONS IMPORTANTES

Au cours de l'étude, de nouvelles informations scientifiques portant sur la cohorte entière ou un sur un participant de l'étude pourraient être générées. Ces informations vous seront transmises et vous pourrez en discuter avec votre médecin traitant ou avec les chercheurs chargés de l'étude.

QUESTIONS ET/ OU PERSONNES-RESSOURCES

Si vous avez des questions au sujet de l'étude, vous pouvez communiquer avec :

Dre Liane Feldman au 514-934-1934, poste 44004.

Pour toute question concernant vos droits en tant que participant de l'étude, veuillez contacter l'ombudsman de l'hôpital au 514-934-1934, poste 48306.

DÉCLARATION DE CONSENTEMENT

J'ai lu le contenu du présent formulaire de consentement et j'accepte de participer à cette étude de recherche. J'ai eu l'occasion de poser des questions auxquelles j'ai obtenu des réponses satisfaisantes. On m'a donné suffisamment de temps pour réfléchir aux renseignements ci-dessus et demander conseil à mon entourage si tel était mon choix. En outre, une copie signée du présent formulaire de consentement me sera remise. En signant le présent formulaire de consentement, je ne renonce à aucun de mes droits légaux.

_____

Signature du participant Date (J/M/A)

_____

Nom du participant (en majuscules)

_____

Signature de la personne obtenant le consentement Date (J/M/A)

_____

Nom de la personne obtenant le consentement (en majuscules)

L'Institut de recherche
du Centre universitaire de santé McGill
The Research Institute
of the McGill University Health Centre

8.4 – Ethics Committee Approval

July 11, 2011

Dr. Liane Feldman
MUHC - MGH
Room L9-417

Re: MUHC Authorization to Conduct Human Subjects Research 11-006-SDR

Dear Dr. Feldman:

We are writing to confirm that the study titled "*Impact of the McGill Multimodal Surgical Recovery (SURE) Pathway on Perioperative Processes and Outcome*" was submitted for all institutional reviews required by McGill University Health Centre policy.

The Surgical Techniques/Medical Devices/Reproductive Technologies (SDR) Research Ethics Board (REB) has notified us that ethical approval to conduct your study has been provided.

Please refer to the MUHC Study Code **11-006-SDR** in all future correspondence relating to this study.

Important Note: You are required to advise the MUHC once the study has been initiated. Please complete the Study Status Report through the *eReviews* system to indicate the date the study became active. Instructions for accessing and using the *eReviews* system are available on the RI MUHC website.

On behalf of the MUHC, we wish you every success with the conduct of the research.

Sincerely,

Lamarche for:

Miguel Burnier, MD, PhD
Associate Director for Clinical Research
The Research Institute of the McGill University Health Centre

cc: REB Study File
RI MUHC Study File

Centre universitaire de santé McGill
McGill University Health Centre
Les meilleurs soins pour la vie
The Best Care for Life

Bureau d'éthique de la recherche
Research Ethics Office

May 24, 2012

Dr. Liane Feldman
MUHC – MGH
Room L9.309

Re: "Impact of the McGill Multimodal Surgical Recovery (SURE) Pathway on Perioperative Processes and Outcome"

Dear Dr. Feldman:

We have received an Application for Continuing Review for the research study referenced above. The report was presented for Full Board review at the convened meeting of the SDR Committee on May 23, 2012 and was found to be acceptable for ongoing conduct at the McGill University Health Centre. This was entered accordingly, into the minutes of the meeting. At the MUHC, sponsored research activities that require US federal assurance are conducted under Federal Wide Assurance (FWA) 00000840.

Approval for the study was provided until May 22, 2013. The Committee noted that no revision to the research protocol (4/19/2011) and the informed consent document (6/17/2011) is required at this time.

All research involving human subjects requires review at a recurring interval. It is the responsibility of the investigator to submit an Application for Continuing Review to the REB prior to expiration of approval to comply with the regulation for continuing review of "at least once per year".

However, should the research conclude for any reason prior to the next required review, you are required to submit a Termination Report to the Committee once the data analysis is complete to give an account of the study findings and publication status.

Should any revision to the study, or other unanticipated development occur prior to the next required review, you must advise the REB without delay. Regulation does not permit initiation of a proposed study modification prior to REB approval for the amendment.

We trust this will prove satisfactory to you.

Sincerely,

Sonya Page, PDt, MSc,
Co-Chair,
SDR Committee

Cc: 11-006-SDR