Health Status Measures for Predicting Postoperative Length of Stay out of Hospital Following Lung Resection Via Thoracotomy

by

Janet Ann Parsons

A thesis submitted in conformity with the requirements for the degree of Master of Science (M.Sc.). Institute of Medical Science. University of Toronto

© Copyright by Janet Ann Parsons. 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
ABSTRACT

Title: Health Status Measures for Predicting Postoperative Length of Stay Out of Hospital Following Lung Resection via Thoracotomy

Student: Janet Ann Parsons
Degree and Year: Master of Science, 1997
Department: Institute of Medical Science
University: University of Toronto

We investigated the hypothesis that preoperative assessment which included a multidimensional health status measure (HSM), when added to FEV₁, DLco and 6-minute walk test, would enhance the ability to predict postoperative outcome as assessed by length of stay out of hospital (LOSOH) in lung resection patients. HSM was evaluated by the EORTC QLQ-C30 questionnaire (global quality of life (QL), social function (SF) and emotional function (EF) scales). For the entire sample (n=70), none of the 6 independent variables evaluated was predictive of LOSOH, while age was a significant predictor. 4 subjects suffered extreme outcomes (LOSOH = 0-5 days), with age the sole factor differentiating these subjects from those with less extreme outcomes (LOSOH = 14-26 days, n=66). In these 66 subjects, QL score and 6-minute walk distance were the strongest predictors of LOSOH. HSM shows a limited ability to predict LOSOH in patients undergoing lung resection.
ACKNOWLEDGMENTS

First, I would like to thank my supervisor, Dr. Arthur Slutsky, for his enthusiasm, encouragement, and sound advice. He is generous in sharing his research expertise and fosters an excellent learning environment for his graduate students. Although he works at a whirlwind pace himself, he never seems too busy to offer a student assistance. I consider myself fortunate indeed to have received such excellent supervision.

Likewise, my programme committee (Dr. Roger Goldstein, Dr. Michael Johnston, Prof. Molly Verrier, and Dr. Charles Chan) were all invaluable to the completion of this research project. They too were always available for consultation when most needed. I would like to express my gratitude to the Thoracic Surgery Division of Mount Sinai Hospital for their continued support of this study (in particular, Dr. Johnston, Dr. Alan Casson, and Dr. Gail Darling, as well as the other staff in the Thoracic Division Office). The assistance offered by the hospital Physiotherapy Department was also outstanding. Many thanks to Ms. Joanne Oakes for her invaluable help with the data collection. Thanks are also due to Dr. Shelly Bull, who provided excellent advice regarding the statistical analysis.

The European Organization for the Research and Treatment of Cancer (EORTC, based in Brussels, Belgium) kindly assisted in this research project, granting me permission to employ their health status instrument, providing scoring manuals and other resources, and allowing their copyrighted instrument to be reproduced in Appendix A of the thesis.

Financial support for this study was generously provided by the Ontario Lung Association through an Ontario Respiratory Care Society Fellowship Award. Additional funding was received from the Mount Sinai Hospital Rose Tomo Bursary programme.

And finally, I wish to thank my husband, Andrew, for his love and support throughout my studies.
# Table of Contents

Abstract  
Acknowledgements  
Table of Contents  
List of Figures  
List of Tables  
List of Abbreviations  
List of Definitions

Chapter 1  
Introduction  
Background  
Lung Resection: a High Risk Surgical Population  
Measurement and Definition of Postoperative Complications Following Lung Resection  
Review of Physiologic Measures Traditionally Employed in the Preoperative Assessment of Lung Resection Candidates

Chapter 2  
Health Status Measurement: Background  
Historical Context of Health Status Research  
Problems of Definition  
Choosing an Appropriate HRQL Measure for an Oncology Population  
Summary of Literature Review  
Research Hypothesis  
Specific Aims of the Study

Chapter 3  
Methods  
Study Design  
Subjects  
Procedure  
Outcome Measures
# Data Analysis

## Chapter 4

### Results

- Primary Analysis: 67
- Examination of Variables: 73
- *Post hoc* Tests: 89
- Summary: 122

## Chapter 5

### Discussion

- Primary Analysis: 128
- Secondary Analyses: 135
- Functional Exercise Testing: 140
- Limitations of the Study: 144
- Implications and Future Directions: 147
- Conclusions: 150

### Appendices

- **Appendix A**
  - EORTC Health Status Instruments: 153
- **Appendix B**
  - Consent Form: 157
- **Appendix C**
  - Modified Borg Rating of Perceived Exertion: 160
- **Appendix D**
  - Diagnostics of Multiple Regression Residuals: 162
- **Appendix E**
  - Instructions to Patients for Performance of the 6-minute walk test: 167

### References

- 169
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 3.1</td>
<td>Timeline for Subject Testing</td>
<td>50</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Frequency Distributions of LOSOH for Entire Sample and Group A</td>
<td>68</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Scatterplots of LOSOH vs. Age</td>
<td>72</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>LOSOH vs. Surgeon</td>
<td>75</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Graphs of Preoperative EF Scores</td>
<td>77</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Graphs of Preoperative SF Scores</td>
<td>78</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Graphs of Preoperative Global QL Scores</td>
<td>79</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Graphs of Preoperative FEV&lt;sub&gt;1&lt;/sub&gt;</td>
<td>80</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Graphs of Preoperative DLco</td>
<td>81</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Graphs of Preoperative 6-minute Walk Distance</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Test-retest Reliability of 6-minute Walk Test (2 Trials)</td>
<td>92</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Intra-rater Reliability of 6-minute Walk (Sample)</td>
<td>93</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>Inter-rater Reliability of 6-minute Walk (Sample)</td>
<td>93</td>
</tr>
<tr>
<td>Figure 4.13</td>
<td>Test-retest Reliability of 6-minute Walk in Normals</td>
<td>96</td>
</tr>
<tr>
<td>Figure 4.14</td>
<td>Scatterplots of LOSOH vs. BMI</td>
<td>104</td>
</tr>
<tr>
<td>Figure 4.15</td>
<td>Scatterplot of Walk Distance vs. PF Score</td>
<td>106</td>
</tr>
<tr>
<td>Figure 4.16</td>
<td>LOSOH vs. Lung Cancer Stage</td>
<td>109</td>
</tr>
<tr>
<td>Figure 4.17</td>
<td>LOSOH based on Administration of Neoadjuvant Therapy</td>
<td>110</td>
</tr>
<tr>
<td>Figure 4.18</td>
<td>Global QL Score vs. Administration of Neoadjuvant Therapy</td>
<td>110</td>
</tr>
<tr>
<td>Figure 4.19</td>
<td>Scatterplots of LOSOH vs. Anaesthesia Times</td>
<td>112</td>
</tr>
<tr>
<td>Figure 4.20</td>
<td>LOSOH vs. Amount of Lung Tissue Resected</td>
<td>113</td>
</tr>
<tr>
<td>Figure 4.21</td>
<td>LOSOH vs. Extent of Resection (Simple or Extended)</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4.22</td>
<td>Recovery Profile of EF Scores over 3 Months Postoperatively</td>
<td>118</td>
</tr>
<tr>
<td>Figure 4.23</td>
<td>Recovery Profile of SF Scores over 3 Months</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Figure 4.24</td>
<td>Recovery Profile of Global QL Scores over 3 Months</td>
<td></td>
</tr>
<tr>
<td>Figure 4.25</td>
<td>Recovery Profile of 6-minute Walk Distance over 3 Months</td>
<td></td>
</tr>
<tr>
<td>Figure 4.26</td>
<td>Recovery Profile of PF Score over 3 Months</td>
<td></td>
</tr>
<tr>
<td>Figure 4.27</td>
<td>Recovery Profile of RF Scores over 3 Months</td>
<td></td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Discharge Date by Day-of-week</td>
<td></td>
</tr>
</tbody>
</table>

118 119 120 121 121 132
List of Tables

1.1 Recent Selected References Evaluating the Predictive Ability of FEV$_1$ and DLco 12

4.1 Reasons for Removal from Study 65

4.2 Characteristics of Sample 66

4.3 Study Sample: Summary Statistics of Primary Outcome Variables 67

4.4 Complications Suffered 67

4.5 Group B: Preoperative Data 71

4.6 Univariate Regression of Independent Variables with LOSOH 84

4.7 Multiple Linear Regression Table with Age as Predictor (Entire Sample) 87

4.8 Multiple Linear Regression Table of Group A (a) and Correlation Matrix of Predictor Variables 88

4.9 One Way Repeated Measures ANOVA on Walk Tests 94

4.10 Reliability Coefficients of the 6-minute Walk Test 94

4.11 Data of Comparator Subjects with COPD 98

4.12 Walk Test Performance Compared with Lung Resection Candidates and COPD Patients (Stratified by FEV$_1$) 99

4.13 Pearson Correlations of Preoperative Stair Climbing Performance with Operative Outcome (LOSOH) and Other Preoperative Measures 101

4.14 Multiple Regression Using Stair Climbing Workload Achieved as a Predictor Variable (Entire Sample and Group A) 102

4.15 Correlations of Additional Preoperative Functional Scale Measurements taken with the QLQ-C30 and the QLQ-LC13 Modular Supplement 107

4.16 Profile of Subjects Requiring ICU Admission 116

4.17 Reasons for Readmission 117
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABG</td>
<td>Arterial blood gas</td>
</tr>
<tr>
<td>ADL</td>
<td>Activities of daily living</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATS</td>
<td>American Thoracic Society</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>CHF</td>
<td>Congestive heart failure</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>CRQ</td>
<td>Chronic Respiratory Disease Questionnaire</td>
</tr>
<tr>
<td>DLco</td>
<td>Single-breath diffusing capacity of carbon monoxide</td>
</tr>
<tr>
<td>EBL</td>
<td>Estimated blood loss</td>
</tr>
<tr>
<td>ECOG</td>
<td>Eastern Cooperative Oncology Group</td>
</tr>
<tr>
<td>EF</td>
<td>Emotional Function scale of the EORTC quality of life instrument</td>
</tr>
<tr>
<td>EORTC</td>
<td>European Organization for the Research and Treatment of Cancer</td>
</tr>
<tr>
<td>FACT-L</td>
<td>Functional Assessment of Cancer Therapy (Lung Cancer subscale)</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Forced expiratory volume in 1 second</td>
</tr>
<tr>
<td>FIM</td>
<td>Functional Independence Measure</td>
</tr>
<tr>
<td>FLIC</td>
<td>Functional Living Index Cancer</td>
</tr>
<tr>
<td>FVC</td>
<td>Forced vital capacity</td>
</tr>
<tr>
<td>global QL</td>
<td>Global Quality of Life scale of the EORTC instrument</td>
</tr>
<tr>
<td>HR</td>
<td>Heart rate</td>
</tr>
<tr>
<td>HRQL</td>
<td>Health-related quality of life</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>ICIDH</td>
<td>International Classification of Impairments, Disabilities, and Handicaps</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive Care Unit</td>
</tr>
<tr>
<td>LOS</td>
<td>Length of stay (in hospital, postoperatively)</td>
</tr>
<tr>
<td><strong>LOSOH</strong></td>
<td>Length of stay out of hospital within the first 30 days postoperatively</td>
</tr>
<tr>
<td><strong>MVV</strong></td>
<td>Maximum voluntary ventilation</td>
</tr>
<tr>
<td><strong>OCI</strong></td>
<td>Ontario Cancer Institute</td>
</tr>
<tr>
<td><strong>PaCO₂</strong></td>
<td>Partial pressure of carbon dioxide (measured with ABG)</td>
</tr>
<tr>
<td><strong>PaO₂</strong></td>
<td>Partial pressure of oxygen (as above)</td>
</tr>
<tr>
<td><strong>PFT</strong></td>
<td>Pulmonary function test(ing)</td>
</tr>
<tr>
<td><strong>PF</strong></td>
<td>Physical Function scale of the EORTC instrument</td>
</tr>
<tr>
<td><strong>PPO-</strong></td>
<td>Predicted postoperative-</td>
</tr>
<tr>
<td><strong>QL</strong></td>
<td>Quality of life</td>
</tr>
<tr>
<td><strong>QLQ-C30</strong></td>
<td>EORTC Quality of Life Core Questionnaire (30 items, Version 2)</td>
</tr>
<tr>
<td><strong>QLQ-LC13</strong></td>
<td>EORTC Quality of Life Questionnaire -- Lung Cancer Supplement (13 items)</td>
</tr>
<tr>
<td><strong>RF</strong></td>
<td>Role Function scale of the EORTC instrument</td>
</tr>
<tr>
<td><strong>SF</strong></td>
<td>Social Function scale of the EORTC instrument</td>
</tr>
<tr>
<td><strong>SF-36</strong></td>
<td>Medical Outcomes Study 36-Item Short Form Health Survey</td>
</tr>
<tr>
<td><strong>SIP</strong></td>
<td>Sickness Impact Profile</td>
</tr>
<tr>
<td><strong>SVT</strong></td>
<td>Supraventricular tachycardia</td>
</tr>
<tr>
<td><strong>VATS</strong></td>
<td>Video-assisted thoracoscopic surgery</td>
</tr>
<tr>
<td><strong>VO₂(max)</strong></td>
<td>(Maximum) oxygen uptake</td>
</tr>
<tr>
<td><strong>WHO</strong></td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Definitions:

In the interest of clarity, I have elected to include definitions of some terms commonly used in this thesis. Inexact terminology in the field of health status has led to considerable confusion in the literature. I will try to distinguish between certain terms used in a general sense, and those with very specific meanings. The definitions provided come from a variety of sources (references are cited below), but the following references were particularly helpful and succinct: Patrick DL, Erickson P. Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation. New York: Oxford University Press, 1993 (21); Streiner DL, Norman GR. Health Measurement Scales: a Practical Guide to their Development and Use. Oxford: Oxford University Press, 1994 (68); Streiner DL et al. PDQ Epidemiology. Toronto: BC Decker Inc., 1989 (77); McSweeny AJ, Creer TL. Health-related quality of life assessment in medical care. Disease-a-Month 1995: 41: 6-71 (120); World Health Organization. International Classification of Impairments, Disabilities, and Handicaps. Geneva: World Health Organization, 1980 (140); Ware JE. Conceptualizing disease impact and treatment outcomes. Cancer 1984: Suppl: 2316-2326 (52). Many works by Gordon Guyatt and colleagues also proved helpful (37-38,41,42,46-47) as did various publications of the EORTC (23-24,123,149).

The terms health status and health-related quality of life can be considered synonymous in this study. Both terms encompass a multidimensional conception of health and disease: health status measures evaluate three broad functional domains: physical, psychological and social function. In this thesis, they are defined as subjective measures (patient-evaluated, patient-perceived). These measures should not be confused with another type of evaluation employed in this study, namely objective (researcher or clinician-rated) measures of functional exercise ability (for instance, the 6-minute walk test). Health status measures evaluate both disability and handicap, while functional exercise tests evaluate disability (using ICDH terms).

In this thesis, length of stay out of hospital within the first 30 postoperative days (LOSOH) is a surrogate measure of operative complications (both morbidity and mortality). This measure is taken to represent all possible complications experienced by lung resection patients, along a continuum from mild to severe to fatal. Operative outcome is a general term employed to denote the
postoperative course experienced by a patient. It does not describe a specific outcome measure in this study.

Construct validity involves the degree to which a theoretical construct describes the attribute of interest (77), and the relative strength of the relationship between the new measure and another (established) measure of the 'same' attribute (68,77). Concurrent validity refers to the extent of agreement between the measure of interest and a 'gold standard' measured at the same time (77). Predictive validity also entails comparing two different measures, but the measure being validated is administered at Time 1 and a measure of outcome/status is administered at Time 2; the extent to which these measures are correlated/associated indicates the degree of validity of the measure given at Time 1(68).

The terms predictive utility and predictive ability are used frequently throughout this thesis. These are general terms that indicate the extent to which a particular measure (i.e. an independent variable) predicts the outcome of interest (in our study, LOSOH). These terms should not be confused with the usage of utility as employed by various investigators of the utilitarian school of health status (e.g. measuring standard gamble, health preferences, timed trade-offs) (21,143); nor should these terms be confused with the specific statistical terms of positive and negative predictive values (77). Predictor variable refers to an independent variable contributing significantly to the variance of a dependent variable within a multiple linear regression analysis (160).

The term reliability (test-retest, intra-rater, and inter-rater) refers to the reproducibility of test scores (77). Reliability coefficients employed in this thesis include Pearson's product moment correlation and intraclass correlation.
CHAPTER 1

INTRODUCTION

Lung cancer is the leading cause of cancer death in North America, with an estimated 168,000 new cases occurring in 1994 in the United States (81). Surgical resection provides the best chance for cure for patients suffering from non-small cell forms of the disease. Whereas the vast majority of pulmonary resections are performed for primary lung cancer, an increasing number of patients undergo resection of pulmonary metastases secondary to renal cell cancer, osteosarcoma, soft tissue sarcoma, colon cancer, and various other neoplasms (80). Given the increasing incidence of lung cancer in both industrialized and developing nations, single lung resection via thoracotomy is likely to remain prevalent for the foreseeable future (81). The Centers for Disease Control estimate that approximately 30,000 lung resections are performed annually in the United States (13).

Although the morbidity rates attending this procedure have improved considerably in recent decades, it remains a high-risk intervention (82). Therefore clinicians and researchers remain interested in identifying those candidates at high risk for developing complications. Many of these postoperative complications are associated with smoking-related comorbidities, such as chronic obstructive pulmonary disease (COPD), coronary artery disease, and peripheral vascular disease (81-82). Early identification of high-risk individuals would allow clinicians to offer them a variety of prophylactic perioperative interventions, including preoperative smoking cessation programs, respiratory rehabilitation, bronchodilator therapy, and more intensive postoperative monitoring and therapeutic interventions.

Despite considerable research directed at evaluating predictive measures of surgical fitness, it remains difficult to determine who will and who will not tolerate pulmonary
To date, most studies have focused on physiological measures of respiratory and cardiac impairment (1.5-18.81-82), such as pulmonary function tests (PFTs), arterial blood gas analysis (ABGs), exercise capacity, and ventilation-perfusion scans (7.12-14.17-18.81-82). Unfortunately, some patients appear to be excellent surgical candidates based on the above traditional physiologic parameters, yet suffer major postoperative complications. Other patients -- considered poor candidates based on the same measures -- tolerate pulmonary resection remarkably well.

Clinicians working with thoracic surgery patients have remarked on the discrepancy between their 'clinical impression' of a patient and their ranking based on physiological measures (67.83). Frequently, such 'impressions' arise from patients' subjective reports regarding their past and present states of health, their social support networks, and their general motivation to have the surgery. Such preoperative 'clinical impressions' remain anecdotal, with no prospective studies to support such claims.

Our study evaluated the relative predictive ability of a multidimensional health status instrument, when compared with conventional measures of physiologic function (PFTs), and a measure of functional exercise ability (the 6-minute walk test). We hypothesized that a subjective health status instrument would significantly enhance the preoperative assessment of lung resection patients, and would identify additional, psychosocial factors predictive of postoperative recovery. A secondary purpose of the study was to evaluate the impact of pulmonary resection (via thoracotomy) on postoperative health status and exercise ability. This chapter and the next outline the background to the topic and provide the rationale for our study design.
BACKGROUND

Lung Resection: a High Risk Surgical Population:

Surgical excision of pulmonary malignancies was first attempted during the 19th century, but it was not until 1930 that long-term survival was achieved (84). Many of these early patients succumbed to postoperative complications. As recently as 30 years ago, the perioperative morbidity for pneumonectomy was as high as 20 per cent, with mortality rates of 5 to 10 per cent (84). Fortunately, significant improvements have occurred in the past two decades, both in terms of long-term survival, and in terms of operative mortality and morbidity (84-85). The current 5-year survival rates for pulmonary resection for NSCLCa are as high as 40 per cent overall, almost double the rates quoted in 1960 (85). Advances in staging and preoperative selection are largely credited for such improvements (84-85). Long-term survival rates are highest for Stage 1 and 2 disease (with rates above 50 per cent) (85). Patients with N2 disease experience lower 5-year survival rates, ranging from 9 to 29 per cent, depending on the authors consulted (84-85). Occasionally patients undergo resection for N3 disease or in the presence of isolated distant metastases, but these represent very select groups, with few survivors at 5 years (84-85).

Resection of pulmonary metastases was first attempted in 1927, but it has only become a widely accepted practice in the past 25 years (80). 5-year survival rates vary depending on tumour type: results of resections for metastases from osteosarcoma and soft tissue sarcoma can range from 20 to 40 per cent, while metastasectomy for malignant melanoma is less promising, with rates of 4 to 20 per cent (80). Other important factors to long-term survival from metastatic disease include disease-free interval, the administration of adjuvant chemotherapy, and the total number of pulmonary nodules (depending on the type of tumour encountered) (80).
Modern operative mortality figures for pulmonary resection range from 2 to 12 per cent, while morbidity can range from as low as 7 per cent to as high as 50 per cent, depending on the particular definitions used and the studies consulted (1-6.81-82.86). By contrast, coronary artery bypass grafting enjoys a mortality rate of less than 1 per cent. In one of the few prospective analyses in recent years, Deslauriers and colleagues (1994) documented a 30-day mortality rate of 3.8 per cent for patients undergoing resection for lung cancer, with non-fatal complications occurring at a rate of 27 per cent (86). In the same year, Kearney et al. (1994) documented an overall complication rate of 17 per cent (81). It is evident that, while surgical resection does improve survival and offers the best chance for cure, it is still a procedure characterized by significant risks.

The complications experienced by patients undergoing lung resection range along a continuum from relatively benign (e.g. prolonged air leak, tachydysrhythmias) to the severe (e.g. lobar atelectasis, pneumonia, bronchopleural fistula) to potentially fatal (respiratory or cardiac failure, myocardial infarction, pulmonary embolism) (82.84.87). Atelectasis and pneumonia account for over one third of all complications in this population (1). The majority of major complications documented in Deslauriers' series were respiratory in nature (e.g. bronchopleural fistula, respiratory failure, pneumonia, pulmonary embolus, and atelectasis) (86).

The mechanism of pulmonary complications associated with lung resection has been well-described in the literature (1-3.57.84.88). One of the most significant factors is the surgical approach. Posterolateral thoracotomy is the preferred incision, as it permits the greatest exposure of the involved hemithorax (84). Thoracotomy alters the normal mechanics of the chest wall, impairing the patient's ability to cough, deep breathe and mobilize postoperatively (3.14.88-90). Impaired ventilation and neuroreflexive inhibition of the diaphragm result in postoperative pulmonary dysfunction (67). The cascade of ventilation-perfusion mismatching, impaired gas exchange, and altered pulmonary defense mechanisms predispose the patient to postoperative atelectasis and pneumonia (67.89-91).
Both surgical exposure of the lungs as well as impaired mobility postoperatively increase the risk of pulmonary embolus (84,87). In a retrospective study by Ziomek et al. (1992), pulmonary embolus accounted for 52 per cent of operative mortality (87). Postresectional pulmonary edema (particularly in a setting of pneumonectomy) is yet another potential complication (92).

Besides pulmonary complications, patients may experience a variety of major and minor cardiac dysrhythmias (atrial fibrillation and supraventricular tachycardia (SVT) being the most common forms) (84,93). The etiology of supraventricular tachydyssrhythmias remains unclear (93). A recent prospective study by Amar et al. (1995) reported an incidence of SVT of 18 per cent following lung resection (93). The authors suggest that increased right heart pressure may contribute to the development of SVT (93). They also found that the patients who developed postoperative arrhythmias experienced greater intraoperative blood losses and received larger amounts of fluid intraoperatively (93).

Some authors have noted that postoperative complication rates were related to the extent of lung tissue resected (3,67,81,84,93-94), while others have found no such relationship (86,95). The majority of authors indicated that pneumonectomy was characterized by a higher complication rate than subtotal resections (3,81-82,84,93): however. Deslauriers’ recent prospective analysis of 783 lung resection patients revealed no significant difference in either morbidity or mortality for those undergoing lobectomy versus pneumonectomy (86). There has been disagreement as to whether wedge resection confers any advantage over lobectomy (85-86). Deslauriers and colleagues (1994) found that extended procedures (e.g. intrapericardial dissection, chest wall resection) posed no greater risk than simple resection (86). Other authors have found that extended resections were associated with higher incidences of postoperative complications (82,93). For instance, Busch and colleagues (1994) documented an 82 per cent rate of complications in those undergoing extended procedures, versus an overall complication rate of 39 per cent for their entire sample \((n = 104)\) (82).
Additional risk factors investigated include smoking history, age, sex, and administration of neoadjuvant chemotherapy. Amar et al. (1995) found that preoperative chemotherapy in general (and mitomycin, in particular) was associated with a higher 30-day operative mortality rate (93). Other authors have reported similar results (82, 86, 93). Smoking history has been evaluated by several investigators (81-82, 132-133). There is disagreement concerning whether cessation date significantly affects complication rates, but overall pack-year history does appear to influence operative outcome (81-82). Unfortunately, most studies looking specifically at smoking history have been performed in patients other than those undergoing pulmonary resection (e.g. general and cardiac surgery patients) (82, 96, 132). Some authors argue that age greater than 70 years is a significant risk factor for developing postresectional complications (81, 86), while others refute this claim (82, 93). There are studies indicating that male patients are at greater risk than females (81, 86), while others fail to confirm this hypothesis (82, 93).

Given the strong association between smoking history and lung cancer, it is hardly surprising that lung cancer patients frequently suffer from concomitant chronic obstructive pulmonary disease (COPD). It is estimated that 90 per cent of lung resection candidates present with some signs and symptoms of COPD (13, 81-82). Many of the traditional preoperative tests try to identify those patients suffering from severe degrees of airflow limitation. It is of little use to offer a patient a potentially curative resection only to have him suffer an untimely death (or severely curtailed function) secondary to surgical complications and poor presurgical pulmonary reserve (13). Hence the emphasis in previous studies on the predictive abilities of pulmonary function tests (PFTs) and aerobic capacity measures, which are evaluations commonly used in both the COPD and lung resection populations.

Prior to a consideration of currently employed predictive measures, it is essential to review an issue central to surgical research, namely the measurement (and definition) of postoperative complications. This is the focus of the next section.
Measurement and Definition of Postoperative Complications Following Lung Resection:

Even a cursory review of the surgical literature demonstrates that (with the exception of death), there is little agreement as to the exact definition of postoperative complications (91). As outlined above, non-fatal complications following surgical lung resection may be related to pulmonary or cardiovascular compromise, may be related to intraoperative mishaps, may be secondary to various comorbidities (COPD and coronary artery disease), and so forth (81-84, 86, 91-93, 128, 170). Even the criteria for determining whether a patient has developed a postoperative pneumonia is a topic of considerable controversy (61). The complications experienced by lung resection recipients range along a continuum from the relatively benign (prolonged air leak or atrial fibrillation) to the severe (pneumonia or respiratory failure) to the fatal. Given the wide array of complications suffered and their highly variable degrees of severity, investigators lack a single, easily definable outcome measure. The obvious exception is death, but this is a relatively rare complication, and the sample sizes necessitated by mortality studies are unwieldy. What is required is a measure that will encompass both morbidity and mortality within a single measure, and that will reflect the relative severity of complications experienced.

Numerous authors have employed the surrogate measure in-hospital length of stay (LOS) to measure operative outcome in a variety of surgical populations. A MEDLINE search of 602 articles (1992-1996) with LOS as a primary measure of outcome (varied surgical and critical care populations) indicates that many authors have used it as a surrogate measure reflecting complication rates, with varying degrees of success. However this measure is notoriously poor at accounting for early operative mortality. A patient who dies on postoperative day 3 (POD#3) would receive a LOS of 3 which would appear a remarkably successful outcome. It appears that LOS is a suitable measure of morbidity, but not mortality. It is common practice in surgical studies to consider the
morbidity and mortality rates within the first 30 postoperative days (3.6-7.9, 82.84.86, 93.95, 128.170).

In our study, it was important to employ a single measure of operative outcome -- one which would reflect complication rates (inclusive of morbidity and mortality). Our intent to use multiple linear regression analysis of the data made it desirable to employ a single dependent variable. We elected to evaluate a new surrogate measure of complications, length of stay out of hospital within the first 30 days postoperatively (LOSOH). In some respects, it is the reciprocal of the concept LOS. If a patient is discharged home from hospital on POD#5, he would receive a LOSOH score of 25 days. However, if he were to be readmitted on POD#10 and stayed in hospital a further 10 days, his LOSOH would drop to 15. If a patient were to die on POD#2, he would receive a LOSOH of 0 (the worst possible outcome). This demonstrates the advantages of LOSOH: it incorporates morbidity and mortality within a single measure, and it accounts for hospital readmissions. Higher scores represent better operative outcome (fewer days of hospitalization).

This approach is different from the way LOS has traditionally been applied and interpreted in other surgical studies. While 30-day operative morbidity and mortality figures are commonly employed, investigators have noted that early postoperative mortality is not accounted for when LOS is used as a surrogate measure of outcome. Our literature review failed to identify any study which applied a maximum LOS of 30 days as the maximum possible score and therefore, representative of the worst operative outcome (it might be suggested that a death on POD#2 should receive a score of 30 automatically: none of the studies consulted employed this approach). Moreover, readmission patterns were rarely addressed in the LOS studies consulted. Because LOS has never been employed in this fashion, we chose to adopt LOSOH as our surrogate measure of outcome to distinguish it from the traditional models of LOS usage.
Obviously, several criticisms can be levelled at both LOSOH and LOS as surrogate measures of postoperative complications. First, other factors other than complications could be responsible for the number of days during which patients are hospitalized. The availability of community resources (e.g., home care services), family supports/living arrangements, socioeconomic status, administrative policies, surgeon preferences, and the like could all impact hospital discharge and readmission patterns. Second, it does not address the issue of quality of time spent out of hospital (e.g., days of normal activity, functional losses). Despite these shortcomings, LOSOH was adopted as a surrogate measure of operative outcome for our study, with appropriate attention paid to factors (other than complications) which might potentially affect length of hospitalization. Patterns of discharge based on administrative policies, attending surgeon, discharge by day-of-week (among other factors) were tracked in our study. Because of its relative advantages (as outlined above), LOSOH was chosen to represent (albeit in a simplistic fashion) postoperative complications. It was necessary to investigate the construct validity of LOSOH in our study. Further consideration of the strengths and weaknesses of this measure appears in the Methods, Results, and Discussion chapters of this thesis.

Despite the significant attention devoted to predictive measures in the surgical literature, there remains considerable debate as to which preoperative physiologic tests are best for identifying high-risk candidates (13.82-83). Moreover, there may be some complications that are virtually impossible to predict based on preoperative patient assessment: for instance, unexpected massive intraoperative hemorrhage, sudden onset of life-threatening arrhythmias intraoperatively, or surgical mishaps (67.170). For obvious reasons, most investigators have examined the ability of various preoperative measures to predict postoperative complications (usually cardiopulmonary in nature). As the following literature review will demonstrate, no gold standard for preoperative testing has yet been identified (13.82-83.170).
Review of Physiologic Measures Traditionally Employed in the Preoperative Assessment of Lung Resection Candidates:

As outlined above, patients undergoing single lung resection represent a particularly high-risk surgical population. This procedure entails removing lung tissue which may (or may not) participate significantly in overall cardiopulmonary function – depending on the degree of parenchymal destruction by tumour, the presence of shunt, emphysematous changes, and so forth. Such factors differentiate this surgical population from those receiving, for instance, upper abdominal surgeries. This is one of the primary reasons that the lung resection candidate is at significant risk for developing pulmonary complications. In a setting of cancer, surgeons are reluctant to deny patients a potential cure based on unreliable preoperative outcome measures. A review of the recent literature regarding physiologic measures used in the preoperative assessment of lung resection candidates appears below.

Pulmonary Function Testing:

Pulmonary function tests (PFTs) have the longest history and are the most commonly cited preoperative measures in the lung resection literature. They also enjoy a long history of predictive testing. The greatest advantages of PFTs are that they are readily available to most thoracic surgeons, they are objective and reproducible measures, they have undergone a rigorous standardization process, and there are normative values available (97). The American Thoracic Society (ATS), the European Community for Coal and Steel (ECCS), and the European Society for Clinical Respiratory Physiology have all published guidelines regarding testing standards (97-100). A Canadian study of inter- and intra-laboratory variability of PFT measures by Kangalee and Abboud (1992) revealed that the forced expiratory volume in 1 second (FEV₁) and forced vital capacity (FVC) were the most reproducible measures, while diffusing capacity for carbon monoxide (DLCO) was the
least reproducible (101). The authors noted that there was greater variability between laboratories than within laboratories (101).

FEV₁ is the most frequently employed PFT measure for predicting operative outcome following pulmonary resection, either alone, or as part of the FEV₁/FVC ratio (1.5-6, 11-13, 15, 67, 97). This test assesses airway and bellows function (13). It is common for definitive cut-points to be quoted in the literature, beyond which patients are thought unlikely to tolerate lobectomy or pneumonectomy (13, 16, 64, 67). This is sometimes expressed as an actual value in litres (L) (1.7 to 2.0 L prior to surgery or 0.8 to 1.0 L following surgery are commonly quoted estimates) (12, 15-16, 67, 82), based on the argument that subjects with absolute values lower than these will be left with unacceptable levels of respiratory impairment and disability. Besides being expressed as an absolute value, FEV₁ can also be expressed as a percent of predicted normal values (% predicted) -- based on the subject's age, sex, height and weight (16, 18, 60, 64, 67, 18; 97-98). This approach permits more realistic intersubjective comparisons: a diminutive elderly female subject may experience less impairment with an FEV₁ of 1.3 L than does a younger, large-framed male subject with the exact same value (67). However, based on normative tables for relative size and age of the patient, we understand that an FEV₁ of 50 per cent of predicted means much the same to both these subjects despite their discrepant statures (1.64, 67, 81). Despite the advantages of using the FEV₁, its use as a predictor is limited (6, 13, 16, 82). Several reports indicate that protracted postoperative dyspnea is neither related to preoperative PFT scores, nor to the amount of parenchyma resected (12, 70, 82).

Gilbreth and Weisman (1994) argue that preoperative PFT values should not be used to exclude patients from lung resection (given that this usually represents the only hope of cure): rather they should indicate a need for further in-depth preoperative investigations (16, 67).

Recent work by Kearney et al. (1994) employed a different transformation of PFT data (81). They determined that the best predictor of postoperative complications was a
measure of predicted postoperative FEV₁ (PPO-FEV₁), based on the preoperative measure and the number of bronchopulmonary segments to be resected (81). They found that a preoperative FEV₁ of less than 1 L was not predictive of postoperative complications, although PPO-FEV₁ was predictive. Unfortunately, the clinical utility of this form of measure is questionable, given that it is difficult to determine the exact extent of resection required for cure based on preoperative radiographic techniques. Direct visualization is

Table 1.1: Recent Selected References Evaluating the Predictive Ability of FEV₁ and DLco

<table>
<thead>
<tr>
<th>Study</th>
<th>PFT Measure Evaluated</th>
<th>Predictive Utility (+ or -)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nagasaki et al. (1982)</td>
<td>FEV₁ (% predicted)</td>
<td>+</td>
</tr>
<tr>
<td>Colman et al. (1982)</td>
<td>FEV₁, FVC (% pred.)</td>
<td>+</td>
</tr>
<tr>
<td>Smith et al. (1984)</td>
<td>FEV₁-PPO</td>
<td>-</td>
</tr>
<tr>
<td>Ferguson et al. (1988)</td>
<td>DLco (% predicted)</td>
<td>+</td>
</tr>
<tr>
<td>Markos et al. (1989)</td>
<td>FEV₁-PPO</td>
<td>+</td>
</tr>
<tr>
<td>Vance et al. (1991)</td>
<td>DLco and DLco (% pred.)</td>
<td>+</td>
</tr>
<tr>
<td>Schramel et al. (1992)</td>
<td>FEV₁-PPO</td>
<td>+</td>
</tr>
<tr>
<td>Holden et al. (1992)</td>
<td>FEV₁-PPO</td>
<td>-</td>
</tr>
<tr>
<td>DeFilippi &amp; Ferguson (1992)</td>
<td>FEV₁-PPO</td>
<td>-</td>
</tr>
<tr>
<td>Kearney et al. (1994)</td>
<td>FEV₁-PPO</td>
<td>+</td>
</tr>
<tr>
<td>Deslauriers et al. (1994)</td>
<td>FEV₁ (absolute)</td>
<td>-</td>
</tr>
<tr>
<td>Busch et al. (1994)</td>
<td>various spirometric measures, incl. FEV₁-PPO</td>
<td>-</td>
</tr>
<tr>
<td>Bolliger et al. (1995)</td>
<td>FEV₁ (absolute, L)</td>
<td>-</td>
</tr>
<tr>
<td>Amar et al. (1995)</td>
<td>FEV₁ (% predicted)</td>
<td>-</td>
</tr>
<tr>
<td>Epstein et al. (1993)</td>
<td>FEV₁-FVC. ratio (absolute)</td>
<td>-</td>
</tr>
<tr>
<td>Reilly (1995) (review article)</td>
<td>DLco (% predicted)</td>
<td>-</td>
</tr>
<tr>
<td>Gass &amp; Olsen (1986)</td>
<td>DLco (% predicted)</td>
<td>+</td>
</tr>
</tbody>
</table>

**This table is not intended as a meta-analysis, but is rather a summary of studies consulted. The +/- symbols denote the studies' conclusions.**
usually required. Moreover, these authors recognized that the low number of patients in their study with an FEV$_1$ of less than 1 L may represent a selection bias. The results of these preoperative assessments may have been further biased by the lack of blinding of the surgeons (81).

The results of studies examining the relative predictive ability of PFTs remain inconclusive. As shown in Table 1.1, for every study supporting their utility as a predictive measure, there is a study refuting this claim.

Another measure of interest to investigators is DLco, which measures pulmonary gas exchange (reflecting the alveolar surface area available for gas exchange, the integrity of the alveolar membrane, as well as capillary blood volume) (13). DLco is a measure commonly included with laboratory PFTs. The DLco test does not measure diffusion directly but rather measures the rate at which inhaled carbon monoxide gas is cleared from an inhaled gas mixture of known proportions (12). Carbon monoxide absorption is dependent on haemoglobin content (results of this test are usually corrected based on the patient's own haemoglobin level) and alveolar-capillary membrane diffusion (12). A number of investigators have found preoperative measures of DLco to be predictive of postoperative complications (9, 12-13, 16, 39, 60, 64, 95, 104, 131). In 1988, Ferguson et al. noted that a DLco of less than 60 per cent predicted correlated with increased rates of postoperative morbidity and mortality among lung resection recipients (104). A 1989 study by Markos et al. determined that a PPO-DLco of less than 40 per cent was associated with increased mortality rates (reported mortality rate of 5.7 per cent) (9). Bolliger's results (1995) suggest that DLco is predictive of mortality but not morbidity (95). Conversely, studies by Vance et al. (1991) and Epstein et al. (1993) found that DLco was not predictive of operative outcome (6, 126).

A number of other components of laboratory PFTs have also been considered in the literature, but are not a major focus of this thesis. However, it is fitting to discuss them briefly and provide a rationale for focusing on FEV$_1$ and DLco. One of the measures
mentioned (and evaluated in the literature) is maximum voluntary ventilation (MVV). This test not only evaluates parenchymal function, but also reflects overall strength and endurance of the chest wall musculature (12). Because it requires a maximal effort from the patient, it is also thought to reflect patient motivation (12). Unfortunately this test lacks reproducibility (only one trial is possible during a single testing session) as well as specificity (it does not identify suboptimal patient effort) (12). Because of the inherent difficulty of the test, many laboratories do not routinely include MVV as part of the standard PFT work-up. Nevertheless, some authors do provide information regarding its predictive ability for lung resection. Several authors consider an MVV of less than 50 per cent predicted to be associated with increased risk for postoperative complications (12,81,105). Gaensler's seminal work (1955) on predictive measures for pulmonary resection in a setting of tuberculosis set the stage for many of the oft-quoted figures employed today for acceptable surgical risk; he included findings on MVV, noting a 40 per cent mortality rate in those with preoperative MVV scores of less than 50 per cent predicted (81,106). Later authors cited similar results (81,107). However, a more recent study by Miller and Hatcher (1987) found that patients could tolerate thoracotomy well with MVV scores below 50 per cent predicted (81, 108).

Although several physiological measures of lung volume, flow rates, and diffusing capacity are of interest, it would appear that FEV₁ and DLco are the 2 measures most likely to be of predictive value in our study. These 2 measures each represent distinct aspects of respiration, namely bellows and gas exchange functions. Given the constraints of sample size and the statistical analysis employed (multiple linear regression), it was not possible to study all PFT parameters as part of the primary analysis in the current investigation.

**Arterial Blood Gas Analysis:**

Of specific interest to most authors is arterial blood gas analysis (ABGs) -- in particular, an elevated partial pressure of carbon dioxide (PaCO₂). The range of PaCO₂ is
quite narrow, and is thought to be a reliable indicator of severe pulmonary disease (including the control of breathing). A preoperative PaCO$_2$ greater than or equal to 45 mmHg is commonly assumed to pose an increased risk for lung resection (12.81). This test is routinely ordered in most facilities prior to lung resection, yet there is inconclusive evidence that it is highly predictive of postoperative complications (83). Kearney et al. (1994) evaluated elevated PaCO$_2$ in their predictive study, and found no difference in complication rates between patients with elevated PaCO$_2$ versus those with normal levels. Morice et al. (1992) found similar results (72,103). Busch et al. (1994) likewise documented no association between postoperative complication rates and elevated PaCO$_2$ (82).

In the present study, we documented preoperative ABG results (where available). For the same reason that we chose to limit the PFT parameters used in the primary statistical analysis (namely a restricted number of independent variables based on sample size available), we chose to exclude this measure from the primary statistical analysis. However, ABGs are considered in the secondary descriptive analysis of the results.

Exercise Testing:

For many years, clinicians have extolled the virtues of exercise testing for predicting postoperative complications in lung resection candidates. Even prior to the availability of modern PFTs, surgeons recognized that the ability to climb stairs appeared to predict operability (12). In recent years, numerous studies have been undertaken to investigate the predictive abilities of various forms of exercise testing in thoracotomy/lung resection candidates. The complexity and sophistication of these tests varies widely, from relatively simple stair-climbing and timed walking tests to laboratory measures of maximum oxygen uptake (VO$_2$ max). A brief literature review of both forms of testing is warranted given that simple functional exercise tests were employed in the present study (6-minute walk test and stair-climbing test).
Comprehensive cardiopulmonary exercise testing in a laboratory setting has been advocated by numerous researchers (6-7,13,16,64,67,109,134). The rationale for this type of testing is that it stresses the entire cardiopulmonary/oxygen transport system, yet is refined enough to permit the identification of specific areas of dysfunction (109). One of the first studies to evaluate the predictive ability of a 6-stage treadmill exercise test in a setting of pneumonectomy was that of Reichel in 1972 (13,134). In his retrospective analysis, he found that patients who were unable to complete the entire test suffered a morbidity rate of 57 per cent, while all patients able to complete the exercise test tolerated pneumonectomy without complications (13,134). More recently, Epstein and colleagues determined that simply the inability to exercise (for a variety of reasons: musculoskeletal, neurologic, vascular, and behavioural), served as a risk factor for developing postoperative complications following pulmonary resection (110). Numerous authors have examined the merits of exercise testing for assessing operative risk (6-10,14,16,64,72,74,95,110-111,115-117). Some researchers have employed treadmill testing, while others have used cycle ergometry. Cycle ergometry is the most popular of the two. Variables commonly assessed include oxygen uptake (VO₂), carbon dioxide output, expired air flow, blood pressure, electrocardiography (ECG), and ABGs (109).

In a study by Smith et al. (1984), patients who had achieved a VO₂max of 20 mL/kg/min had a low risk of postoperative morbidity, while those whose VO₂max was less than 15 mL/kg/min were at significantly higher risk (7). Olsen and colleagues (1989) found similar results with submaximal values of VO₂ (14). A 1990 study by Gerson et al. concluded that bicycle ergometry was predictive of pulmonary and cardiac complications in elderly subjects undergoing elective upper abdominal and noncardiac thoracic surgery (10). Unfortunately, they did not identify which form of surgery held the highest morbidity and mortality rates. Bechard and Wetstein (1987), as well as Bolliger (1995), set the threshold of prediction at 10 mL/kg/min (95,103,109,135). Other authors have not found VO₂max to be prognostic of postoperative complications (6,64,116). Miyoshi et al. (1987) found
that VO₂ was predictive of mortality but not morbidity (136). As with other physiologic measures employed with lung resection candidates, a definitive statement regarding the value of preoperative laboratory cardiopulmonary exercise testing remains elusive.

Simplified forms of exercise testing are popular among clinicians evaluating lung resection candidates. These tests have the benefits of ease of administration and low financial burden (when compared with laboratory exercise testing). Tests of stair-climbing and level walking distance (with or without pulse oximetry measures) are the most commonly employed tests in the daily clinical practice of thoracic surgery teams. Because these measures are functional in nature (reflecting activities commonly performed by the majority of patients) they have meaning for operative candidates (75-76, 112).

In recent years, comparative studies of laboratory-type tests and walk tests have revealed reasonable correlations between the two. Unfortunately, most validation studies of level-walking tests and VO₂max have been performed in patients suffering from COPD and chronic heart failure, and not in a lung resection population. In a small study of the 12-minute walk test (in patients with COPD), Bernstein et al. (1994) found that the distances walked during consecutive intervals of the 12-minute walk were moderately correlated (r=0.72) with VO₂ achieved using cycle ergometry (113). McGavin et al. (1976) also found a correlation between 12-minute walk distance and VO₂max achieved by cycle ergometry (114). Interestingly, they found poor correlations between FEV₁ and distance ambulated (114). Correlations between the 6-minute walk and VO₂ were found to be lower, but significant (r=0.64) (113). Guyatt et al. (1985) correlated the 6-minute walk with cycle ergometry results in studies of patients with chronic heart disease and those with COPD (38,42). In a study of patients with chronic heart failure, Lipkin et al. (1986) found a curvilinear relationship between VO₂max and 6-minute walk distance (112). They found less variability in distance walked in those with a VO₂max greater than 20 mL/kg/min (112). In this study, patients expressed a preference for the walk test over treadmill testing because the former related well to their normal activities of daily living (112).
Holden et al. (1992) compared three different forms of exercise testing in a lung resection population. The authors found that 6-minute walking distance and stair climbing results were correlated with VO₂max by cycle ergometry (115). Of interest was the finding that 6-minute walk distance and number of stairs climbed were both predictive of operative outcome, while VO₂max was not (115). One reason for this may have been the reluctance of some patients to perform the maximal cycle ergometry test (115).

As mentioned previously, surgeons have noted for years that the ability to climb stairs appeared to be positively correlated with operative outcome (12). Four studies have examined the relative predictive ability of stair climbing as a preoperative measure (14,57,115,117). Bolton and colleagues (1987) found that male patients who were able to climb 127 stairs or more were at lower risk for developing complications (57,137). They determined that the number of steps climbed correlated well with FEV₁ and FVC spirometric measures (16,57,137). Unfortunately, not all patients in their study were operative candidates, so they were unable to draw firm conclusions as to the predictive ability of this test (16,57,137). Holden et al. (1992) (in the study mentioned previously) asked patients to perform a timed, symptom-limited, self-paced stair climb (115). Subjects were monitored using pulse oximetry. Knowing the dimensions of the steps, the number of steps ascended, the time taken to climb the steps, and the subject's weight, they employed the following formulae to calculate workload and VO₂ achieved:

\[
\text{Work} = \text{step height} \times \text{no. steps per minute} \times \text{weight (kg)} \times 0.1635 \\
\text{VO}_2 = (5.8 \times \text{subject weight}) + 151 + 10.1 \times \text{Work} \quad (115)
\]

The same formulae are employed in our study. Besides VO₂ and workload measurements, the number of stairs climbed proved to be predictive of mortality within the first 90 days in Holden's study (115). They determined a threshold value of 44 steps for identifying high-risk surgical candidates (115). Both stair climbing and 6-minute walk distance were found...
to be highly predictive measures in a lung resection population (115). As other investigators working with COPD patients have found, the VO2max calculated from stair climbing was higher than that achieved using a conventional cycle ergometry test (79, 138).

Holden employed the same formulae and testing procedure as Olsen and colleagues had in 1991 (14, 115). Olsen’s retrospective analysis of 54 male subjects revealed a predictive threshold of 75 steps for identifying those patients at significantly higher risk of major postoperative morbidity (14).

In a recent Canadian study, Rao et al. (1995) performed a retrospective analysis of 299 patients who had been evaluated preoperatively using pulse oximetry and stair climbing, and compared this with conventional spirometric measures (117). Specifically, they were interested in comparing FEV1 and oxygen desaturation during stair climbing as relative predictive measures (117). Their results indicated that pulse oximetry during stair climbing (2 flights) had greater predictive ability than FEV1. Desaturation (below 90%) during stair climbing identified 50 per cent of perioperative deaths, while all of the patients who died had an FEV1 of greater than 1.5 L (their prespecified cutoff for poor spirometry) (117). Moreover, oximetry was significantly more predictive of home oxygen requirements, prolonged hospital length of stay (LOS), development of respiratory failure, and intensive care unit (ICU) admission. The authors identified the need for a prospective trial (117).

Functional exercise tests have the advantage of being inexpensive, readily available, and simple to perform. These measures are presently undergoing more rigorous standardization. They are well-tolerated by patients, are familiar and meaningful to the majority of lung resection candidates, and include elements of the postoperative rehabilitation programs that most patients receive following thoracotomy (early ambulation). Since these simple tests have been found to correlate reasonably well with more complex forms of exercise testing, they appear to be acceptable screening tools for identifying patients at high risk of postoperative complications. The 6-minute walk test
was selected as the primary measure of functional exercise capacity in this study. A detailed consideration of the 6-minute walk test is the focus of the next section.

6-Minute Walk Test:

In 1976, McGavin et al. published an article advocating the adoption of a new method of clinical exercise testing (114). They developed a 12-minute walk test for estimating exercise tolerance in patients with chronic bronchitis (113-114). The idea was adapted from Cooper's work concerning a 12-minute running test for healthy normals (it correlated with VO2max obtained via more sophisticated testing techniques) (113, 151). McGavin et al. found that the distance walked in 12 minutes correlated poorly with FEV1, yet correlated significantly (if only moderately) with VO2max achieved with cycle ergometry (r = 0.52) (114). In a later study of 62 patients with either COPD or pulmonary infiltration, McGavin and colleagues (1978) compared the 12-minute walk distance with subjective ratings of breathlessness using an oxygen-cost diagram and with Borg's rating of perceived exertion (RPE), and with measures of pulmonary function (namely FEV1, FVC and DLco) (129). They found the highest correlations between the distance walked in 12-minutes and the subjective scores on the oxygen-cost diagram and the Borg RPE, as well as the DLco (129). While they found statistically significant correlations with FEV1 and FVC and 12-minute distance, the correlations were not as strong as with the former measures and did not appear to have clinical significance (129). In both studies, the test demonstrated high test-retest reliability (r = 0.97), but there was significant improvement on the second attempt (129). For this reason, they advocated two trials to establish a baseline and eliminate the practice effect (114, 129).

Since these initial reports, there have been numerous studies employing the 12-minute walk test, mostly in the COPD patients. A Dutch study by Dekhuyzen et al. (1986) demonstrated that the 12-minute walking distance was well correlated with FEV1, PaO2, and other measures of pulmonary function (152). ZuWallack and colleagues (1991)
investigated the relative improvement in 12-minute walking distance and numerous
predictors of this improvement, following an outpatient pulmonary rehabilitation program
(153). They found that patients with a greater ventilatory reserve and a larger FEV$_1$
showed greater improvements in the distance ambulated (153).

A few investigators have chosen to evaluate the 12-minute walk in patients
undergoing lung resection. Bagg (1984) did not find the test helpful in identifying the
patients who died postoperatively (74). Markos et al. (1989) employed the 12-minute walk
test in their study of preoperative predictors in lung resection candidates (9). They also
subjected their patients to cycle ergometry. They do not provide any information regarding
correlation between cycle ergometry results and the walk test; however, a perusal of their
data table indicates that there was little relationship between the two measures (9). They do
state that neither workload achieved during cycle ergometry, nor distance ambulated were
predictors of postoperative complications (9).

Most studies have employed level walking in a quiet corridor, but some have used
the 12-minute walk on a self-paced treadmill. A 1990 study by Swerts and colleagues
revealed that walking speeds for the corridor tests were significantly higher than for the
treadmill tests (40). In 1985, researchers compared the 12-minute walk, a fixed pace step
test, and duration of cycle ergometry in patients with COPD (79). The investigators found
that improvements in performance were seen after four trials, regardless of the type of test
administered. These authors recommended three practice attempts at any exercise test.
Interestingly, they also found weak correlations between exercise performance and FEV$_1$
(79).

Because the 12-minute test was thought to be too time-consuming (especially given
the need for repeated measures), investigators became interested in shorter versions of the
test. In 1982, Butland and colleagues compared 2- and 6-minute versions with the 12-
minute walk in COPD patients (39). They found them all to be highly correlated with one
another: 6-minute vs. 12-minute ($r = 0.955$), 2-minute vs. 12-minute ($r = 0.864$), and 2-
minute vs. 6-minute ($r = 0.892$) (39). The authors recommend the 6-minute test as a sensible compromise between the longest and shortest versions (39). In a recent reanalysis of the 12-minute walk, Bernstein et al. (1994) found that changes in VO$_2$ were most closely correlated with changes seen during the 12-minute version, than with those seen in the 2-, 4-, and 6-minute versions (113). They advocated the 12-minute test as the most reliable indicator of exercise ability (113).

Despite such findings, the walk test that currently enjoys the greatest popularity is the 6-minute version. As with the 12-minute test, the majority of studies of the 6-minute walk have been undertaken with COPD patients, with a subsection of investigations relating to the chronic heart failure (CHF) population (38,42). Studies by Mungall et al. (1979) and by Guyatt et al. (1985) concluded that the 6-minute walk test possessed comparable (and acceptable) properties of reliability, when compared with the 12-minute version (164-165). A 1986 study by Lipkin et al. (1986) found that the relationship between distance ambulated in 6 minutes and VO$_2$ max was not directly linear but rather traced an exponential curve, whereby little variation was seen in distance walked in normal subjects: yet those with significant (New York Heart Association (NYHA) Class II and III CHF) heart failure showed the greatest variability between the two types of tests (112). They argue that maximal as opposed to submaximal testing may be warranted in patients with mild heart failure, since maximal tests are more likely to elicit symptoms in these subjects (112). Of interest, Lipkin and colleagues are the only investigators to have examined the upper limits of the 6-minute walk test (in terms of distance ambulated). They only tested 10 normals, and compared their walk distances with those of patients suffering varying degrees of heart failure (112). They found normals to be highly consistent in their distances ambulated (clustering between 650 and 750 m) (112).

Much of the validation and standardization of 6-minute walk testing comes from the work of Guyatt and colleagues over the past decade or more (38,41-42,61,66). In 1984, Guyatt et al. studied patients suffering from either COPD or CHF, and found that
encouragement was a significant factor in walking test performance, and could be responsible for gains of as much as 30.5 m (66). The authors therefore cautioned researchers intending to employ this outcome measure to standardize their approach -- either giving encouragement in a standard fashion, or giving no encouragement whatsoever (66). Consistency is the key above all. In a 1985 study of patients with CHF, Guyatt and colleagues found that the 6-minute walk correlated well, not only with maximal exercise tests (cycle ergometry), but also with the NYHA functional classification (38). In the same study they recommended 2 practice walks, prior to taking a baseline measure (38). In another report (42), the same authors noted high correlations between cycle ergometry and walk distance, and similar correlations between the walk distance and scores on four functional status questionnaires (Oxygen-Cost Diagram, the Rand Instrument, the Baseline Dyspnea Index, and the Specific Activity Scale) (42). Of particular interest was the finding that the cycle ergometry measures and results from the questionnaires correlated (42). In addition, all four questionnaires were highly correlated with one another (42). Because of the highly reproducible results, these authors have employed the 6-minute walk as an outcome measure in a number of clinical trials (29,35,41).

6-minute walk testing is commonly used as an outcome measure in studies evaluating new therapies for COPD. These articles are of interest to authors employing these measures in their own studies, as they documented various correlations between the walk test measurements and subjective measures. Unfortunately, the 6-minute walk has been used infrequently in trials with the lung resection population, which seems odd, given the significant interest exercise testing has elicited over the years. The study by Holden et al. (1992) indicated that the 6-minute walk distance can be predictive in high-risk subsets of lung resection candidates (115). The ability to walk further than 1000 ft. (304.80 m) in 6 minutes was predictive of successful operative outcome (115). As a predictor of 90-day survival, the walk test displayed a sensitivity of 100 per cent, a positive predictive value of 85 per cent, and a negative predictive value of 100 per cent (115). The
authors agreed that previous investigations of the 12-minute test's predictive powers were disappointing (9.74), but argued persuasively for further investigations of the 6-minute test (115).

There is increasing interest is validating the 6-minute walk test as an outcome measure in disparate patient populations, not just in those suffering from chronic ventilatory and cardiac insufficiency. The test has been used as an outcome measure for assessing rehabilitation of the orthopaedic surgical population, amputees, and the frail elderly (164). Other types of walk tests are also being investigated; for example, the self-paced walk test, the 2-minute walk test, the 100 m walk, and various tests on self-paced treadmills have been used to evaluate disability and function in a wide array of patients (154-156).

Given the extensive reliability studies already undertaken with the 6-minute walk, and the interest in exercise testing in general in the lung resection population, we chose this test as an outcome measure for the current study. We felt that since we were investigating measures of preoperative function and health-related quality of life, it would be valuable to have a measure that assessed exercise performance by a method that had meaning to the subjects. Since it was impossible for us to perform 3 preoperative trials of the test, we were concerned with the relative reproducibility of the test in this particular patient population. We decided to investigate the test-retest reliability, as well as intra- and inter-rater reliability of the 6-minute walk test during the course of this investigation. Because so many subjects exhibited highly consistent results with repeated measures (more so than the results reported in the literature on COPD patients), we felt this issue worth pursuing further. We decided to test a small sample of healthy normals by way of comparison with the study sample. We also chose to compare the results seen in our sample with 80 patients who had documented COPD, to see if there were any differences in their walk profiles.
Summary of Currently Available Preoperative Measures:

PFTs are the current gold standard for preoperative assessment. While FEV₁ has been the most extensively researched measure, DLco has also undergone intensive scrutiny. Each of these tests measures different aspects of pulmonary function, namely bellows and gas exchange functions. Another preoperative measure that has enjoyed considerable attention is exercise testing -- both sophisticated laboratory evaluations and simple tests of functional ability. The 6-minute walk and stair climbing tests are two examples of commonly employed functional exercise tests. Such “low-tech” exercise testing is appealing because of its low cost, ease of administration, and the fact that the tasks performed are meaningful to most patients (16).

The Necessity of New Preoperative Measures:

Conventional physiologic parameters used to evaluate lung resection candidates have not proven effective for identifying patients at increased risk of postoperative complications. Functional exercise testing holds promise, but requires further validation. Recently, there have been suggestions that health status (health-related quality of life) measures should be investigated in patients undergoing lung resection. In a recent editorial comment on the articles by Kearney et al. and Busch et al. in the March 1994 issue of Chest, Cykert suggests that a major outcome which should be evaluated in the lung resection population is health-related quality of life (83). In their response to this editorial, Busch and colleagues agree that this factor should be given more importance in surgical decision-making.

The goal of our study is to assess the relative predictive ability of subjective measures of health status when compared with PFTs (the current gold standard) and functional exercise testing. Before delineating the specific aims of the study and the
research hypothesis. It is important to review the literature on health status measures. Chapter 2 reviews the current literature on health status assessment, and will consider the historical context of its development.
CHAPTER 2

HEALTH STATUS MEASUREMENT: BACKGROUND

The past decade has seen an explosion in the area of health status measurement. While traditional measures of health and disease have focused on certain physiologic functions (as delineated above) and on measures such as survival (or quantity of life) and disease-free interval, there has been new interest in measuring quality of life and the relative impact of various medical and surgical interventions on same (68). Increasing acceptance of subjective measures of health status means that clinicians are attempting “to measure what was previously thought to be unmeasurable” (68).

What follows is an overview of the health status literature, followed by a consideration of the health status instrument selected for this study. But first, it is necessary to provide the historical context of the health status field and to define some of the terms that will be employed throughout this thesis.

Historical Context of Health Status Research:

In recent decades, there has been increasing recognition that the determinants of health and disease are not solely physiological, but include social and psychological factors as well. There is a shift from biomedical models to more holistic models of medicine. Many date this shift in attitudes toward health to the World Health Organization’s (WHO) 1947 definition of health as a “state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity” (21, 139). In 1980, the WHO published its International Classification of Impairments, Disabilities, and Handicaps (ICIDH), which recognized that disease processes occur not only within an individual, but within the context of his/her society and cultural environment (118, 140). In 1977, quality of life was added as a keyword to the MEDLINE computerized search system (119). The
last few years have seen a dramatic increase in publications and research related to health status evaluation, with new journals devoted to the topic (Quality of Life Research), and the foundation of the International Society for Health-Related Quality of Life in 1994 (120).

Although the interest in health-related quality of life measurement has been a relatively recent trend in medicine, general interest in health and disease in its broader social context is longstanding. The theoretical foundations of the health status field are quite old. Interest in relating personality traits and emotional states to physical health dates back at least as far as the ancient Greeks and Romans and their preoccupation with the four humors (21). Many anthropological, sociological, and psychological theorists have been interested in the impact of disease on social and role function, subjective well-being, and health perception (21). One could argue that modern clinicians have been remarkably slow to study health and disease in their broader social and cultural contexts, being caught up in the modern wave of empiricism.

Theoretical contributions to the field of health-related quality of life can be traced to numerous sources. Jeremy Bentham's utilitarian philosophy (advocated in the 18th century) held that the central motivating force to all human action is the quest for pleasure and the avoidance of pain (21.141). Moreover, the utilitarians argued that preferences could be quantified by interval or ratio scales (known as Bentham's hedonic calculus) (21.141). In this century, modern utilitarians have focused on the assignment of values to particular or probable outcomes, giving rise to the field of timed-tradeoff and standard-gamble models of health preferences (21.142). Torrance, Lane, Llewellyn-Thomas and others have employed utilitarian principles to their research regarding medical decision-making (142-144).

The work of Talcott Parsons and the functionalist school of sociology emphasized the interrelatedness between human biological/physiological function, and social and
psychological function (21.145). Parsons defined illness in the following way:

"...a state of disturbance in the normal functioning of the total human individual including both the state of the organism as a biological system, and of his personal and social adjustments." (The Social System. 1951, p.431) (145)

These sociological theorists also addressed the difference between capacity and ability to fulfill one's 'normal' or expected roles.

As mentioned at the outset of this section, the advent of the WHO's ICIDH classification system in 1980 was a tremendous boon to the field of multidimensional health status measurement. With such internationally sanctioned support, multinational organizations have developed for evaluating various aspects of health status among normals and among populations suffering from diverse pathologies. The health status instrument used in the present study came from just such an organization, namely the European Organization for the Research and Treatment of Cancer (EORTC). Recent publications from the 1995 conference on the ICIDH emphasize that the relationship between the concepts of disease, impairment, disability, and handicap should not be viewed as causal or linear (118.121). Indeed, the very wording of the original WHO document of the ICIDH contained the subheading “the consequences of disease”, which has fostered this (mistaken) linear reasoning -- from etiology to pathology to manifestations (121.146). As Halbertsma points out. clinicians and researchers should consider these concepts to be interdependent, characterized by fluid relationships (121). Hence the rationale for the present study: the psychosocial factors evaluated are assumed to stem neither directly from the lung pathology. nor from the prospect of the surgical resection itself. They are evaluated preoperatively, based on the assumption that psychosocial factors themselves may influence outcome (besides simply being influenced by the surgical intervention). The latter is the more common assumption in traditional medical models. Our rationale is more in keeping with the spirit of the ICIDH.

The various outcome measures evaluated in this study range across the spectrum outlined by the WHO and the ICIDH (121). The functional exercise tests employed (e.g.
6-minute walk and stair-climbing tests) are measures of disability, while the questionnaire’s social functioning scale is a measure of handicap. Pulmonary function tests are measures of impairment.

The interest in quality-of-life/health status evaluation originated with investigations of certain chronic disease processes for which therapeutic success was questionable (in terms of symptomatic relief and survival: diseases such as chronic obstructive pulmonary disease (COPD), cancer, and arthritis) (21,120). For instance, a 1990 study by Kazis et al (1990) reported that a major predictor of 5-year survival among rheumatoid arthritis patients was patient-reported health perception (119,166). Only recently has the surgical literature begun to reflect the multifactorial nature of disease and recovery following surgical interventions (44,51). Studies are now being published which indicate that subjective ratings of health status, past experiences, and expectations can be associated with both objective and subjective responses to treatment. A 1991 study by Haut et al. documented that patients’ past experiences of motion sickness, as well as subjective ratings of their expectations to treatment are correlated not only with subjective experience of nausea and vomiting following chemotherapy, but objective measures of these sequelae as well (125). If pre-treatment measures of health status (or expectations of same) can predict physical effects of treatment, it may be that preoperative measures of health status can predict the postoperative recovery of patients undergoing lung resection.

Following lung resection, patients must undergo intensive rehabilitation. If patients rate their health status poorly preoperatively, perhaps they will be less motivated to engage in programs of early ambulation, programs of deep breathing and coughing exercises, and so forth. A failure to mobilize in the early postoperative period increases the patient’s risk for deep vein thrombosis and pulmonary embolus. Reluctance to comply with the regimen of cardiorespiratory physical therapy could result in unresolved atelectasis and the potential for pneumonia secondary to secretion retention. Such reluctance could conceivably delay
recovery. In this way, we believe that preoperative health status could influence rates of medical complications post-lung resection.

It makes sense that early surgical interest in health status evaluation occurred amongst clinicians operating to relieve pain and suffering: orthopaedic surgeons performing arthroplasties, as well as general surgeons performing resections for ulcerative colitis and Crohn's disease are but two populations that have been evaluated in terms of health status and operative outcome (122). Surgeries that are aimed at improving survival are now beginning to be investigated in this regard, of which lung resection for primary lung cancer is a prime example. Because such procedures represent the best hope for cure, it is only recently that quality-of-life issues have been addressed; previously, quality of survival seemed of secondary importance to survival itself. But given the high perioperative morbidity and mortality associated with pulmonary resection, it is vital to ensure that this potentially curative procedure does not result in unacceptable levels of disability and handicap for the patient. Given the limitations of traditional physiological measures (as outlined above), it is likely that hitherto unidentified health status factors may help in identifying those patients at risk for postoperative complications. Moreover, the use of a subjective measure of health status is warranted, as a patient’s perception of his/her health status may have a major impact on postoperative recovery.

The traditional preference for 'objective' physiological measures stems from a value judgment by the medical and scientific community: that the clinician's/researcher's impartial perspective must be the most accurate depiction of a patient's health. McCullough (1984) argues that this value judgment in favour of objective measures arises from a paternalistic or "beneficence model" of health care (see the wide array of literature in biomedical ethics) (120,167). The researcher/physician is assumed to 'know best'. In the fields of health-care ethics and health status evaluation there is a move toward a patient-centred model of health-care, and an "autonomy model" of research, whereby patients' subjective reports are given increasing importance in terms of building knowledge about health and disease.
We are becoming aware that the patient's perspective may be equally valid when compared with that of the 'impartial observer' (120). This shift in perspective is evident when comparing early HRQL research (when clinicians were asked to rate a patient's quality of life) with more recent research which stresses the patient's perspective. The question is "who should evaluate patients' quality of life?" While historically it was clinicians who evaluated patients' health statuses (using performance status ratings such as the Karnofsky Index, ECOG scales, etc.), several studies have shown little correlation between patients' subjective ratings and clinicians' objective ratings of health status (120,123). The current trend is to employ patient self-reports of health status, and the present study reflects this trend.

Problems of Definition:

The terms health status and quality of life frequently are employed interchangeably (20-21). Considerable confusion has ensued owing to the breadth of the terms employed and the rapidity with which this term has been adopted (20-21, 54). Quality of life can be taken to mean many things, so it is essential to restrict its definition, if it is to be a useful concept in clinical measurement. Quality of life can refer to diverse environmental factors (such as air pollution, urban planning, or adequate nutrition), poverty levels, as well as ill-defined notions of happiness and satisfaction (20-21, 26, 52, 68, 120). Indeed some of the literature equates quality of life with well-being, an equally nebulous term. Fortunately, a number of sources have provided meaningful and workable definitions of the concept (20-21, 120, 123). Patrick and Erickson (1993), in their book Health Status and Health Policy define health-related quality of life as

the value assigned to duration of life as modified by the impairments, functional states, perceptions, and social opportunities that are influenced by disease, injury, treatment, or policy.

(21: p. 22)
In this definition, the authors summarize the WHO's definition of quality of life, and of health (21, 121). It reflects the multidimensional nature of health and disease, of function and dysfunction, and the principles expressed in the ICIDH (121). Other authors have incorporated the notion of *perceived* health or health status as the most salient feature of the definition (21, 52, 120). Some authors employ the term *health status*, which may be considered as interchangeable with health-related quality of life (HRQL) in this thesis. This is in keeping with the majority of health status/HRQL researchers today. Unlike Gill and Feinstein (1994), I do not think that the term *health status* ignores the personal or subjective nature of the measure, nor does it have to connote a value imposed from outside the subject (124). I do agree with their criticism that many studies fail to distinguish between overall quality of life and *health-related* quality of life (124). The type of measure that we would define as one of health status or HRQL employs a subjective rating and is multidimensional in nature. Health status is not purely a rating of physical performance (20-21, 120, 124).

By definition, HRQL measures should evaluate three broad functional domains: physical, psychological, and social (20-21, 52, 76, 120, 124). While some authors separate this out further into role function, pain, cognitive status, and emotional function, these are rather subdivisions within the three domains, whereby positive symptomatology may be thought of as measuring physical distress, with cognitive and emotional function comprising part of overall psychological function. Essentially, health status/HRQL instruments should evaluate these three domains. In this thesis, they are taken to be subjective measures. (This goes back to the philosophical debate concerning perspective and the beneficence versus autonomy debate over health status evaluation: the central question being *health for whom* (120)). Moreover, it is important to distinguish subjective health status measures from measures of physical performance status, which are frequently measured by health-care providers—so-called 'objective' measures of performance (48). Examples of such measures of physical capacity are the Karnofsky Index, the Functional Independence Measure (FIM), the Eastern Cooperative Oncology Group (ECOG) Scale.
and the Katz Activities of Daily Living Index (48). We chose to employ a subjective health status measure, because subjects' self-reports of what they actually do is an important measure of performance, more so than a simple measure of capacity (76). Patients may be capable of climbing 8 flights of stairs, but if they are sedentary and never do this outside of the research/clinical setting, the task will mean little to them. Their usual preoperative level of daily activity is more likely to affect their postoperative recovery than is an arbitrary test in a laboratory. Studies in gerontology indicate that the context and meaningfulness of an activity/task are of considerable importance to patient performance status (75-76). In this study, we asked patients to perform physical tasks (6-minute walk and stair climbing tests), but we also asked them to report on their own level of function.

**Choosing an Appropriate HRQL Measure for an Oncology Population:**

The HRQL literature is replete with cautions concerning the difficulties inherent in choosing an appropriate measurement tool (20-21.23-24.37.46.68.76.120). Until recently, many investigators seemed comfortable inventing new instruments for each study, some paying little attention to the chosen measures' properties of validity, reliability, and responsiveness (20.120). As research has progressed in the field of health status measurement, multicentre and multinational cooperative study groups have arisen to promote a more standardized approach (149). It is now expected that new instruments will meet minimum standards of validity, reliability and responsiveness (20-21.68.120). Test-retest reliability, inter-rater reliability, internal consistency, and alternative-form reliability should all be addressed in evaluative studies of a new instrument so that the reader may judge the amount of measurement error inherent in the instrument (68.120). While reliability is a necessary condition for establishing an instrument as valid, it is not a sufficient condition (120). We want to know if a potentially useful measure actually evaluates the properties it purports to measure (i.e. we must establish its validity). Face validity, content validity, construct and criterion validity must be addressed in studies
evaluating new measurement tools (20.68.120). Finally, a health status instrument should be able to detect meaningful change in the disease process, whether in describing the natural history of a disorder, or measuring the effects of a new treatment regimen (20.68.120).

**Generic versus Disease-Specific Measures:**

Researchers face a formidable task when choosing among the many measures currently available. Having outlined the ideal qualities of an instrument in the previous paragraph, it becomes apparent that no one available measure will be perfect. The researcher is faced with a series of compromises regarding the measurement properties of the existing instruments. Unless the investigator wishes to invent his/her own measure, he/she will have to tolerate the shortcomings of the best available tool. A number of choices are crucial. First, one must decide whether a single *indicator* or *index* is more suitable to the research question, whether a *profile* of scores is preferable, or whether a *battery* of different health status measures is the most acceptable approach (21). In addition, investigators must choose between generic and disease-specific measures, and between discriminative versus evaluative instruments (20-21.46-47.68.120). The primary benefit of generic measures is that they permit comparison between different types of patients (21.120). A number of these measures exist and have been validated to varying degrees. Examples of this form of health status measure are the Sickness Impact Profile (SIP), the McMaster Health Index Questionnaire, the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36), the Nottingham Health Profile, and the Quality of Well-Being Scale (QWB) (21.120).

The advantage of disease-specific measures is that they are designed to evaluate a particular patient population and pose specific questions regarding the disorder (21.120). Disease-specific measures typically possess high levels of content validity, and are more responsive to changes in the subject’s disease status over time (120). However, the
increased responsiveness may necessitate a trade-off in terms of reduced test-retest reliability, according to Patrick and Deyo (1989) (147.120). For the present study, we chose a disease-specific measure because the subjects present with a specific diagnosis -- cancer -- and most often with a *recent* diagnosis. Such a diagnosis is likely to impact significantly on all dimensions of HRQL -- physical, psychological, and social (45). Patients’ perceptions of a cancer diagnosis will be different from those of osteoarthritis, for example (52). Because of the unique aspects of cancer diagnosis, the choice of a disease-specific instrument, validated in a cancer population, seemed appropriate (20-21.52-53.120).

Cancer (along with other chronic diseases such as chronic obstructive pulmonary disease and rheumatoid arthritis) was one of the conditions first targeted by quality-of-life researchers (148). Yet, it differs from these more benign diseases in several important ways. Perceptions of life expectancy, associated pain, and the relative discomfort accompanying common treatments, coupled with the fear associated with a diagnosis of cancer are some of the features that set it apart from other diagnoses. There is a sense of immediacy (and urgency) that differentiates it from many of the other chronic conditions traditionally focused on by health status investigators. These perceptions regarding cancer are held not only in Western industrialized societies but in many non-Western cultures as well (63). A diagnosis of cancer (for instance lung cancer) conjures up very different images for patients than does a diagnosis of emphysema or arthritis. It makes intuitive sense that health status instruments devised for cancer populations will include items reflecting the unique nature of the diagnosis (and patients’ perceptions thereof) as well as the treatment regimens employed (surgery, chemotherapy, and radiation therapy). In few other patient populations is the assumption that “we have to make you feel a lot worse before we can make you feel better” so strongly held. While inroads have been made in terms of survival and treatment effectiveness with certain forms of the disease (for
instance, breast cancer), treatment in only marginally successful for other malignancies (e.g. lung cancer) (28).

Traditional models of cancer treatment often assume that the symptoms (and distress) experienced by patients are the direct result of pathology, treatment toxicity (e.g. chemotherapy regimens), or the extent of surgical resection. Clinicians might wonder why we would choose to evaluate quality of life, assuming that all cancer patients would rank their HRQL very low. However, a number of recent oncologic studies reveal that this is an inaccurate inference. Roberts et al. (1992) found that women undergoing radical surgery for gynaecologic cancers reported a reasonably good quality of life, contrary to the assumptions of many medical caregivers and the general public (168). Another recent study evaluated the impact of pretreatment (nonpharmacologic) psychological factors on patients’ experiences of postchemotherapy nausea and vomiting (125). While 53 per cent of subjects developed postchemotherapy nausea, and 36 per cent reported vomiting following chemotherapy, the investigators found that patients’ pretreatment expectations concerning the severity of nausea and vomiting was significantly predictive of the frequency and severity of symptoms reported posttreatment ($R^2 = 0.32, p = 0.001$) (125).

Given such findings, it is possible that the expectations and perceptions of oncology patients undergoing lung resection may play a vital role in operative outcome (specifically the development of postoperative complications). These patients must participate fully in their postoperative rehabilitation program in order to prevent postoperative complications. If the patient anticipates severe pain and discomfort he/she may be less willing to comply with such treatment (early mobilization, deep breathing and coughing exercises, etc.) and may experience a protracted recovery. Such perceptions and expectations may be a product of background culture, personal or family history, media coverage, and other factors.

A number of HRQL measures have been employed in patients with cancer (including lung cancer) in recent years: the Functional Living Index – Cancer (FLIC), the
Nottingham Health Profile, the Rotterdam Symptom Checklist, the Symptom Distress Scale, the Breast Cancer Chemotherapy Questionnaire, the Functional Assessment of Cancer Therapy (FACT) Scales, and the Ontario Cancer Institute (OCI) Quality of Life Scale are just a few of the instruments that have been employed in oncology clinical trials (21,36,48-50, 120,169). The instrument chosen for this study is the European Organization for the Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire (QLQ-C30) and its lung module supplement (QLQ-LC13) (23-24,123,149).

A brief consideration of some of the other measures is in order, prior to a detailed consideration of the measure chosen.

A number of studies have been undertaken with the FLIC, most notably, the studies by Schipper et al. (1984) and Ganz and colleagues (1988) (49-50). Ganz et al. found that the FLIC did not respond sufficiently to changes in patients' physical capacities, and that it did not correlate well with significant declines in recordings of physical status (50). The authors expressed concern over the instrument's responsiveness (50). Maguire and Selby (1989) raised similar concerns (48). The OCI Quality of Life Scale needs to undergo further validation in lung cancer patients (it has not been used widely in this patient population) (48). The Spitzer Quality of Life Index has not demonstrated responsiveness to changes in patients' statuses over time (48), and it shows poor correlation between self-administered versions and interviewer-administered versions of the questionnaire (48). The FACT has a lung cancer subscale (the FACT-L) which looks promising, but it has not undergone the same degree of testing (for validity, reliability and responsiveness) as the EORTC instruments (169).

The Instrument Chosen:

The EORTC is one of the oldest and largest international clinical trials groups devoted to the research and treatment of cancer patients (123). To date, 16 European countries are represented within its Quality of Life Study Group (a subsection of the
EORTC, as well as Australia, Canada, and the United States (123). The organization has undertaken some major multicentre validation trials of their core questionnaire (the QLQ-C30) and its modular supplements (for head and neck, breast, colorectal, esophageal, prostate, and lung cancers) (149). Its lung modular supplement is the only one to date to have undergone widespread cross-cultural validation (149). The other modules are currently under development (149).

In 1982, Aaronson, Bergman, and colleagues (23-24, 123,149) began developing and validating their core questionnaire, specifically in patients with inoperable lung cancer (21, 23-24,45,48,123). They have conducted numerous descriptive and methodological investigations (including randomized controlled trials) of their measure (21. 123,149). The first version of the core instrument (the QLQ-C36) contained 36 items, with 4 functional scales (physical, role, emotional, and social functioning), 2 symptom scales, and a global perceived health status scale (24,123). A rigorous process of forward and backward translation procedures has been performed and has been extensively documented by the Study Group (24,123,149). Version 1 was pretested on 537 patients with lung cancer (123). Further refinements were required, including revision of the emotional functioning scale (123).

The second generation of the questionnaire is the one employed in the present study, namely the QLQ-C30 (refer to Appendix A). It has essentially the same structure as the first version, but includes 5 functional scales (with the addition of a cognitive function scale), 3 symptom scales (with the addition of a pain scale to supplement the original fatigue and nausea scales). The original global health status scale was retained, as were single-item scales assessing numerous symptoms (e.g. sleep disturbance, bowel function), along with the item relating to perceived financial impact (24,123). The primary changes to the instrument entailed a reduction in the number of items in the emotional function scale from 8 items to 4, the addition of an item to the pain scale, and the refinement of the cognitive function scale to a 2 item structure (123,149). The second generation instrument
was tested on 305 patients with unresectable lung cancer (24,123), from 12 different countries, including Canada. It demonstrated acceptable standards of reliability/internal consistency (Cronbach’s alpha > 0.70) on all multi-item scales, with the exception of the role functioning scale (24,68). Cross-cultural and cross-linguistic reliability was tested by comparing scale responses across 3 broad cultural and linguistic subgroups: respondents from English-speaking countries, from Northern Europe, and from Southern Europe. Scale reliability was very similar in all 3 groups, with the exception of the nausea and vomiting scale which was found to have significantly lower scores among Southern European respondents. Aaronson et al. (1994) note that this may well reflect differences in prescription patterns of emesis-inducing treatments in Southern European countries (123).

Our study employs 3 of the functional scales within the core questionnaire as primary outcome measures: the global health status (QL) scale, the Emotional Function (EF) scale and the Social Function (SF) Scale (24,123). The second generation global QL scale demonstrated the highest level of internal consistency (0.89). Of 6 studies undertaken with the instrument (oncology populations only), this scale has consistently demonstrated the highest levels of internal consistency (ranging from 0.85 to 0.94) (123). These high levels of consistency are maintained in both longitudinal and cross-sectional study designs (123). In the same series, the EF scales also exhibited strong internal consistency (0.73 to 0.85). Results for the SF scales are similar, with the 5 most recent studies receiving ratings of 0.72 to 0.83 (earlier versions of this scale had not been as promising). Analysis of trials undertaken with patients receiving treatment for inoperable cancer indicate that the QL and SF scores demonstrate high levels of known-groups validity (123).

Validation studies (7 have been undertaken) of the core questionnaire have included evaluations of clinical validity and known-groups comparisons (24). The grouping variables employed were primary diagnosis, stage of disease, performance status, weight loss, treatment status, treatment toxicity, and prognosis (123). High levels of validity
have been consistently found for the physical, role, fatigue, and overall health status scales (123). Moderate scale validity have been demonstrated for the emotional and social functioning scales as well as for the pain scale (123). Known-groups validity was demonstrated for the emotional and social functioning scales, as well as for the pain scale (123). Concurrent validity was evaluated by comparing the emotional functioning scale with the General Health Questionnaire (an accepted measure of psychological distress) \( r = 0.70 \) (24,123). Aaronson and colleagues say that this represents construct validity, but I interpret this as more in keeping with principles of concurrent validity, as discussed by several authors (68,120). The concurrent validity of the QLQ-C30’s pain scale was demonstrated by comparing scale responses with the pain intensity subscales of the McGill Pain Questionnaire (123).

The QLQ-C30 showed responsiveness to change in disease status over time, and correlated well with separate measures of performance status (the ECOG performance status scales and specific subscales of the WHO’s Acute and Subacute Toxicity Scales) (24). Thus it is suitable for use as an evaluative instrument. Moreover, it is able to distinguish between subjects with different clinical and/or treatment statuses, demonstrating its discriminative ability (24,123). Both discriminative and evaluative properties were important for our study, as we wished to distinguish between patients preoperatively (those at higher risk for complications versus those at lower risk for complications), as well as to track change in health status postoperatively over a period of 3 months.

In previous studies with the QLQ-C30, the instrument was acceptable to patients in that it took a reasonable time interval to complete (11 to 12 minutes on average) (21). The method of administration (self- or interviewer-administered) made no significant difference to the results (24). In their work with asthmatics, Cook et al. (1993) found a high degree of correlation (ICC = 0.84) between self- and interviewer-administered answers, however they caution that self-reports consistently yielded higher levels of dysfunction (46.9 % vs. 35.8 %, \( p<0.0001 \)) (150). This suggests that perhaps subjects are more ‘honest’ when
completing questionnaires without the influence of an interviewer (i.e. there is less social desirability bias) (68).

Unfortunately, there is no data concerning test-retest reliability of the EORTC instruments, and this is a major shortcoming. Data should be available on this form of reliability in the near future, as studies of test-retest reliability are currently underway in Canada, Norway, and the Netherlands. The EORTC recognizes this limitation of the instrument (149). Guyatt (1986), and Streiner and Norman (1994) rightly point out that measures of internal consistency by themselves are not sufficient for demonstrating an instrument's reliability (47,68). Results from three studies of the QLQ-C30 employing longitudinal designs indicate that reliability of the scales improves with repeated measurement (123). The EORTC Study Group points out that this may reflect changes influenced by treatment, rather than any property inherent in the questionnaire. A copy of the core questionnaire and the lung modular supplement appears in Appendix A.

The supplementary lung module (QLQ-LC13) was validated during the initial field tests of the core instrument (both versions), meaning it was tested with over 700 subjects (23-24). This module is composed of 13 items evaluating lung cancer symptomatology (cough, hemoptysis, dyspnea, pain) as well as sequelae of chemotherapy and radiation therapy (mouth soreness, alopecia, peripheral neuropathy, dysphagia) (23,149).

The lung module demonstrated the ability to discriminate between patients with different functional statuses (ECOG and WHO toxicity scales) (23,123). It also showed responsiveness to clinically significant change over time (123). As with the core questionnaire, the QLQ-LC13 displayed acceptable reliability via internal consistency measures (Cronbach’s alpha > 0.70) (23, 123). Taken alone, the Cronbach’s alpha of the dyspnea scale was 0.83, and when combined with the core questionnaire reached 0.86 (23). The module’s pain scale showed relatively poor internal consistency on its own (Cronbach’s alpha = 0.53), but in combination with the core instrument attained acceptable
levels (Cronbach’s alpha = 0.71) (23). As with the core questionnaire test-retest reliability still needs to be evaluated. The QLQ-LC13 did display moderate clinical validity (23).

It should be emphasized that the validation studies of both the core instrument and the lung module were conducted in patients with inoperable lung cancer. Thus far there is no published research employing the QLQ-C30 as a primary outcome in the lung resection population. However, given the fact that lung resection patients suffer from the same or very similar pathologies (for the most part), it seems a logical extension to employ the EORTC instruments in a population receiving surgical treatment.

To date, only one study has attempted to measure "quality of survival" in patients who have undergone resection of bronchial carcinoma (51). Nou and Aberg (1980) employed a performance status measure (not a subjective health status measure) called the Carlens vitagram index (51). This measure focused on work capacity, mobility and physical symptomatology. It did not attempt to address important issues of psychological and social function (51,76). Furthermore, surgical practice has changed since the 1970s, when these data were collected. Nou and Aberg document numerous tumour debulking procedures and non-curative resections in their series (51). Still others underwent non-resectional (exploratory) thoracotomies (51). Such practices are relatively rare today. The significance of their study lies in the authors’ interest in examining the impact of surgery on patients’ return to normal function. One of Nou and Aberg’s primary outcome measures was in-hospital LOS, lending further credence to the present study (51).

A 1991 abstract describing a pilot project by Vance and colleagues used the SIP as a preoperative measure, and compared it with PFT measures as well as invasive procedures (126). Their sample size was small (n=16), so the results were inconclusive (126). Of interest, the authors found that SIP scores improved in the early postoperative period (126). The need for a larger prospective trial is evident.

Because lung resection candidates frequently suffer from varying degrees of COPD (usually secondary to smoking history), it is fair to ask whether these patients should be
evaluated using a health status measure previously employed in patients with chronic respiratory disease. Since postoperative ventilatory disability or death from respiratory failure are major concerns of clinicians, why not use a measure like the Chronic Respiratory Disease Questionnaire (CRQ), which has also been well-supported in the literature? We are interested in patients’ experiences of breathlessness, curtailed activity levels, sputum production, and other complaints common to both patients with COPD and lung cancer patients. However, only some of our patients have symptomatology of COPD, while a significant proportion do not. If we employed the CRQ, an inordinately large number of patients would have very low scores, and while emphysema and chronic bronchitis are eventually fatal, the rate of demise is over the course of decades, as opposed to a matter of months in some lung cancer patients. A comparative study of the CRQ and the EORTC instrument in a lung cancer population would be of significant interest to confirm or reject this hypothesis.

Summary of Literature Review:

Only one study examined subjective health status instruments as predictive measures of operative outcome following lung resection. The results of this pilot project, published in abstract form, were inconclusive (126). Such measures have been employed extensively in trials involving patients suffering from other forms of cancer (e.g. breast cancer, head and neck cancer) and in a setting of inoperable lung cancer. Because these measures are multidimensional in nature, they evaluate different features of health status in one instrument. They may present an effective and inexpensive method of preoperative evaluation. The instrument used in this study (the EORTC QLQ-C30 and the QLQ-LC13 lung modular supplement) has demonstrated reasonable internal consistency, validity, and responsiveness. Field testing of the instrument was performed in a population comparable to our own sample, namely, patients suffering from inoperable lung cancer. The EORTC
instrument is a logical choice for our investigation since the questions posed are readily applicable to lung resection patients. We chose a disease-specific measure because certain issues facing cancer patients are unique (possibility of adjuvant and neoadjuvant chemotherapy and radiation, potential for fatal outcome, cultural and personal perceptions of a cancer diagnosis, etc.). The measure chosen addresses issues important to lung cancer patients (dyspnea, cough, and pain, for example). Another advantage of the EORTC instrument is that it has shown cross-cultural relevance in field tests. Given the extraordinary cultural diversity of the research setting (Toronto), this is an important feature of the measure.

**RESEARCH HYPOTHESIS**

That multidimensional subjective health status measures (the global QL scale, the EF scale, and the SF scale of the EORTC QLQ-C30 instrument), will significantly enhance the ability to predict LOSOH following lung resection (via thoracotomy) when added to physiologic measures of pulmonary function (FEV1 and DLco) and functional exercise tolerance (6-minute walk test).
SPECIFIC AIMS OF THE STUDY

The aims of the study are three-fold:

Primary Goal:
1. To evaluate the ability of 3 types of preoperative measures to predict postoperative complications (LOSOH):
   a) PFTs: (specifically FEV1 and DLco)
      This is a measure of physiological impairment.
      It represents the current gold standard in preoperative testing for lung resection.
   
   b) Functional exercise test: (specifically 6-minute walk test)
      This is a measure of disability.
   
   c) Subjective health status instrument: (specifically the EORTC instrument)
      This is a measure of both disability and handicap

Secondary Goals:
2. To identify specific health status factors predictive of postoperative complications in patients undergoing thoracotomy for lung resection.

3. To determine the impact of thoracotomy/lung resection on health status and functional exercise ability in the first 3 months following surgery.
CHAPTER 3

METHODS

Study Design:
This was a single-centre prospective cohort study. The investigator collected the questionnaires (self-administered) and measured most of the walk tests. However, the investigator was blinded to the PFT results during the data collection phase to minimize rater bias during the study.

Subjects:
All patients being evaluated for possible single lung resection via thoracotomy incision at Mount Sinai Hospital (Toronto, Ontario, Canada) were eligible to participate in the study. Subject recruitment occurred between June 1995 and August 1996. Many subjects were recruited during their clinical assessment for surgical staging (usually bronchoscopy and mediastinoscopy), with their thoracotomy and lung resection being scheduled as a later separate operation based on the results of the staging procedure. Some subjects were booked for bronchoscopy and mediastinoscopy immediately prior to their thoracotomy/lung resection, and thus were recruited prior to this single staged procedure. This meant that numerous patients who were deemed to have unresectable tumour (based on mediastinoscopy results) were recruited and completed the preoperative assessments, but were withdrawn from the study after it was determined that they were unresectable. These unresectable subjects thereby served as a comparator group with the study sample proper, allowing the investigators to determine if these patients differed from the study subjects in their preoperative scores on the health status instrument, on the 6-minute walk test, and on pulmonary function testing.
Subjects with potentially resectable lung neoplasms were identified by the surgical staff of the Thoracic Surgery Division of Mount Sinai Hospital. Ethics approval for the study was obtained from the University of Toronto Review Committee for the Use of Human Subjects in June of 1995. Permission to approach patients was obtained from all three surgeons on the team. The study was fully explained to patients and written informed consent was obtained from all subjects (refer to Appendix B). Subjects were aware that if they proceeded to lung resection, the data collector would repeat the walk test and readminister the questionnaire prior to discharge from hospital, at their 1-month follow-up appointment with their surgeon, and again at their 3-month follow-up appointment.

**Inclusion Criteria:**

All subjects undergoing elective lung resection via unilateral thoracotomy (or undergoing staging in preparation for this procedure) were eligible to participate in the study. The extent of lung resection could range from wedge resection, to segmentectomy, to lobectomy, to ipsilateral bilobectomy, to pneumonectomy (3.67.85). Because cell type and extent of tumour invasion are difficult to determine definitively up until the time of thoracotomy, and because 30-day operative morbidity and mortality are typically related to the effects of the thoracotomy/excision, rather than to cell type, any patients presenting with lung neoplasms were considered potential candidates for the study. Patients with lung cancers or other neoplasms may require simple resections (simple thoracotomy and lung tissue excision) or extended procedures (entailing chest wall resection, intrapericardial dissection, or excision of a portion of the diaphragm). We also included those patients who had undergone lung resection previously (either for recurrent lung metastases or for recurrent primary lung cancer) in the study (157).

Subjects were required to speak and read English in order to complete the health status questionnaire. While the core questionnaire has been translated into 24 languages
(and the lung modular supplement is available in 9), the lack of an investigator to score and interpret these other translations precluded our using the other translations available (1-49). This is in accordance with the agreement signed by the principal investigator when the EORTC gave her permission to use their copyrighted instrument (24, 1-49). A copy of the questionnaire appears in Appendix A.

**Exclusion Criteria:**

Patients who underwent sternotomy and concurrent bilateral lung excisions were excluded from the study sample. This is based on incisional differences and significant differences in postoperative pulmonary mechanics. The single lung resection recipient has the advantage of relying on one "unaffected" lung, while those that undergo bilateral resections do not. Moreover, sternotomy entails a more stable incision site, with a less severe impact on chest wall mechanics (55, 57). Differences in postoperative pain severity have also been reported for this surgical approach (55).

Patients who received a thoracoscopic (VATS) approach were also excluded from the study based on incisional differences. Such procedures are not comparable to thoracotomy with respect to postoperative pain control, alteration of chest wall mechanics, and extent of resection (57). At our facility, VATS is used for the resection of isolated peripheral pulmonary nodules, open lung biopsy, pleural biopsy, surgical management of malignant pleural effusions, and so forth. It is not employed for significant lung resection for pulmonary malignancies (curative resections).

Patients undergoing concurrent bilateral thoracotomies and sternal transection (so-called clamshell approach) for bilateral lung resections were also ineligible for the study based on incisional differences. These patients are thought to be at a greater disadvantage in terms of postoperative pain control, requirements for overnight mechanical ventilation, etc., and so were not thought to be comparable to the study population in question. These patients are more prone to postoperative complications.
Figure 3.1

Preoperative
- Subject recruitment
- Preadmission clinic
- Subject recruitment
- Preadmission clinic

Appt. 1 wk.
- Hospital
- Discharge
- Surgery
- Preop. 1 wk.
- Follow-up
- Follow-up
- 3-month
- Appointments
- Appointments

Tests
- Final Tests
- 2nd Postop.
- 1st Postoperative

Tests
- 6-min. walk.
- Questionnaire
- 6-min. walk.
- Questionnaire
- 6-min. walk.
- Questionnaire
- 6-min. walk.
- Questionnaire

ordered
- not already done. PT's
- stair-climbing.
- 6-min. walk.
- ordered
- not already done. PT's
- stair-climbing.

Timeline for Subject Testing
**Procedure:**

Subjects were recruited from one of three sources at Mount Sinai Hospital: the thoracic surgery inpatient ward (upon admission to hospital), the outpatient preadmission unit (where patients receive the majority of their preoperative work-up in preparation for elective surgical admissions), or from the thoracic surgeons' outpatient clinics. Written informed consent was obtained from all subjects (Appendix B). Figure 3.1 depicts the timeline for the critical events involved in data collection for this study.

Once recruited, subjects underwent their routine preoperative assessments. Preoperative PFTs were performed in the laboratory (usually the Mount Sinai Hospital Pulmonary Function Laboratory, or in a comparable laboratory setting) employing American Thoracic Society (ATS) standards. PFTs are considered part of the routine evaluation of lung resection candidates at Mount Sinai Hospital, and in most other thoracic surgery programs.

In addition to pulmonary function testing, routine preoperative bloodwork, ABGs, and urinalysis were performed. Patients received preoperative chest radiographs, and full or partial metastatic workup (as indicated): CT scanning of the head and abdomen, total body bone scan, and abdominal ultrasound.

Preoperatively, patients were requested to complete the EORTC QLQ-C30 core questionnaire and its 13-item lung modular supplement (the QLQ-LC13). The instrument was self-administered, an approach which was determined to be reliable in previous clinical trials with the instrument (24). Most subjects were able to self-administer the questionnaire (148,150). 4 subjects were unable to fill out the questionnaire themselves. 2 owing to upper extremity disability and 2 secondary to visual impairments. In the case of visual impairment, the data collector read the questionnaire aloud to the subject, who dictated his/her response. In the case of upper extremity disability, the data collector sat
beside the subject, allowing them to read the items for themselves and dictating their responses to the data collector. Previous clinical trials by the EORTC showed that mode of administration (self- versus interviewer-administration) had no statistically significant impact on item responses (24).

At all times, the data collector avoided leading the subjects towards a particular answer, and subjects were reassured that there was "no right or wrong answer" and that they should answer the question as they alone interpreted it. To ease apprehension among the subjects, the data collectors made themselves available to provide clarification as necessary.

In addition to the 6-minute walk test, a stair-climbing measure (submaximal VO₂) of functional exercise tolerance was also performed. Only a single trial of this test was undertaken, as subjects frequently found this test difficult to tolerate. A timed, symptom-limited, self-paced stair-climb was undertaken (14,115,117). Because of our desire to monitor patients during the test (some patients become distressed during this test), the rater was required to follow the subject, and continuous pulse oximetry was recorded. This was essential to the safety of the test. Each flight consisted of 10 steps, with each step being 0.19 m in height. The stair climb was performed immediately following the first 6-minute walk test, according to the recommendation of Holden et al. (1992) (115).

Procedure for 6-minute Walk Testing:

a) Subject Testing:

At our facility (and at many others) lung resection candidates routinely undergo preoperative functional exercise testing (6-minute walk test and stair-climbing test). The distance walked (m) in 6 minutes is the measure of interest (14,115).

It was not feasible for the investigator to perform all walk tests on all subjects (given the large sample size, as well as the large number of comparator subjects, and the number of follow-up tests required). Therefore a second rater was trained by the
investigator to ensure inter-rater consistency. The results of intra- and inter-rater reliability testing are in Chapter 4.

We attempted to match the recommendations cited in the literature regarding standardized testing procedures for the 6-minute walk test (38-39.41.66.112.115). Access to patients preoperatively was limited, therefore only 2 measures of the 6-minute walk test were employed at baseline (not 3 trials). We compared the reproducibility of the 2 tests.

Preoperative testing took place in 1 of 3 hospital corridors (beside the surgeons' outpatient clinics, beside the outpatient preadmission unit, and on the thoracic surgical unit). Every attempt was made to perform testing under quiet conditions with a minimum of distractions and corridor traffic. We chose a course 33 m long (38.41), which was a little shorter (7 m) than the length of each ward. This mimics the distance many patients cover during ambulation postoperatively. Start and finish lines were placed at either end of the course and the walls were marked at quarter intervals to facilitate measurement.

We were familiar with the importance of standardized testing (38.41.66.114.129). Standardized instructions were provided to subjects, and no encouragement was offered during the tests (standardized encouragement protocol) (66). Subjects were required to perform the test independently and the rater remained at the start of the course throughout the test, so as not to affect the subjects own pace. Family members were advised that they were not permitted to walk with the patient, nor could they offer encouragement or converse with the subject, in an attempt to minimize the possibility of external influences affecting the subject's performance. In keeping with Guyatt's recommendations (1985), subjects were told to cover as much ground as they possibly could in the test period (38).

When the 6 minutes had elapsed, the tester called out 'Stop' and the subject stopped where they were in the corridor. The total distance covered in metres was measured and recorded. Chairs were placed at either end of the course, and subjects were informed that if they needed to rest they could. Subjects were advised that they could terminate the test for any
reason and should inform the tester should they develop any symptoms, such as chest pain, dyspnea, light-headedness, bronchospasm, or claudication pain.

Prior to the walk, the rater assessed heart rate (HR) and oxygen saturation via pulse oximetry with the subject at rest. The same Nellcor N-100 pulse oximeter with a finger probe was employed for all testing. Pulse oximetry was reassessed at the end of the test. Continuous monitoring was not possible as the portable oximeter used was rather weighty and this would have necessitated the subject carrying it throughout. This would have proved awkward for subjects and might have prevented them from giving an optimal effort. If patients experienced symptoms during the test or they appeared in any distress to the rater, the rater reapplied the pulse oximeter and took a reading. If desaturation or unusual HR response were evident, the rater terminated the test. If saturation and HR were within normal limits, and the subject wished to continue they were permitted to do so. Such intermittent testing was an essential safety feature of the testing procedure. Following the walk test, subjects were asked to rate their perceived exertion using a Borg scale. A copy of the scale appears in Appendix C.

It was our original intention to employ 2 practice sessions and take the third attempt as a baseline measure, in keeping with the recommendations of Guyatt, McGavin and others (38-39.41.66.114.129). After testing the first few subjects, it became apparent that this would not be feasible. These subjects undergo extensive diagnostic testing during their routine evaluation for lung resection (venous and arterial blood tests, PFTs, plain chest radiography, CT scanning, metastatic workup, bronchoscopy, mediastinoscopy, and tissue biopsies). At the time of designing the study, patients were routinely brought into hospital the afternoon of the day before surgery. This at least afforded the time to perform two tests with an acceptable rest period in between (2.5 to 3 hours) on that day. But after the first 2 months of data collection, hospital policy changed to a same day admission pattern for elective surgery. This left the investigator with less time to access subjects preoperatively. Subjects began to be recruited in the outpatient preadmission unit the week
prior to their surgery. It was not possible to test subjects twice during the course of the preadmission clinic, so the second test had to occur on the day of surgery. This often entailed testing patients at 6:30 a.m. as soon as they arrived at the hospital. Most morning cases were transferred to the operating room between 7:10 and 7:20. This did not afford the patients much time for testing. Bringing the patients in for a separate exercise test also was not possible, as many patients lived outside the immediate vicinity of the hospital.

We therefore compared the 6-minute walk distances recorded for patients who had undergone 2 testing sessions with the results of those who had undergone 3 sessions within the first few months of the trial. Based on early favourable results we decided to proceed with a policy of 2 sessions.

In addition to preoperative testing, patients also underwent postoperative evaluation, which entailed a 6-minute walk test. Patients were reassessed prior to discharge from hospital, and again at their scheduled outpatient follow-up appointments with their surgeon (at 1 month and 3 months postoperatively). The walk test at time of discharge was performed using the course on the thoracic surgery unit. The outpatient follow-up tests were usually performed in the surgeons’ outpatient clinics. Another change of policy during the course of the study further complicated postoperative testing. The surgeons at Mount Sinai Hospital used to follow all lung resection candidates in the Mount Sinai clinics. This winter, they began following many of their lung resection patients at the weekly Princess Margaret Hospital clinics. This made it more difficult for the investigator to perform follow-up testing. The Princess Margaret Hospital does not have any corridors adjacent to the clinics that measures 33 m in length. The investigator chose to use a course exactly half the distance of those used at Mount Sinai Hospital. Subjects were informed of the alteration in course length and all appeared to understand well.
b) Testing of Healthy Normals:

The high degree of reliability of the results obtained from preliminary subject testing (and the lack of improvement between the first and second trials) made us question whether the results obtained were related to the relatively high functional status and relative lack of preoperative symptomatology. They certainly differed from the results of patients with severe COPD (as reported in the literature), in that our subjects appeared to be more consistent. We decided to test a small group of young healthy normals \((n = 14)\) for comparison with our study sample. Only the study of Lipkin et al. (1986) mentions the results of 10 healthy normals \((n = 112)\). No other information is available in the current literature regarding the upper limits of this test. In addition, we were curious as to the consistency of results when subjects were tested under ideal conditions: no musculoskeletal pathology, no respiratory or cardiovascular disorders, subjects with a high degree of physical fitness, and who understood the principles of the testing procedure.

Subjects were recruited from the physical therapy department and were either physical therapists or physical therapy students under 35 years of age.

The normals were requested to perform the walk test on 2 occasions. The second test was always performed 5 days after the first. Subjects were instructed to wear the same footwear on both days. As with the study sample, pulse oximetry was performed before and after the test. At the end of the test, subjects were asked to rate their level of perceived exertion using the Borg scale, and to indicate the factor that prevented them from walking further (e.g., dyspnea, pain, fatigue, muscle soreness, light-headedness). Inter-rater reliability was assessed using the two raters who had evaluated the study sample. As with the sample ratings, the raters were blinded to the subjects' previous results.

Postoperative Evaluation:

Following their surgery, subjects were again evaluated prior to discharge home (either on the day of discharge from hospital, or the day prior to discharge). The subjects
were asked to complete the questionnaire, and the 6-minute walk test was repeated. The 2
data collectors who had performed the preoperative exercise testing administered the
postoperative evaluations. Routine outpatient follow-up appointments are scheduled with
the surgeon at approximately 1 month and 3 months postoperatively. The investigators
repeated the walk test and the health status evaluation during these regularly scheduled
outpatient appointments.

Besides the preoperative pulmonary function studies, the exercise tests, and the
health status measures, the postoperative in-hospital LOS and LOSOH was tracked. Any
readmissions to hospital (either at our own facility or at another) were documented.
LOSOH is the dependent variable of interest in this study. Other variables of interest
included sociodemographic data (including age, sex, marital status and family/social
support, occupation, mother tongue), preoperative body mass index (BMI), preoperative
ABG results, duration of operative procedure/anaesthesia, method of postoperative
analgesia, transfer to and LOS in the Intensive Care Unit, necessity of prolonged
mechanical ventilatory support postoperatively, specific types of morbidity experienced,
extent of resection, estimated intraoperative blood loss (EBL), and requirements for
supplemental home oxygen at discharge home. The sociodemographic data was collected
from patient subjective reports. The other data were collected by chart review. This data
was analysed using descriptive methods.

**Outcome Measures:**

**Postoperative LOSOH:**

This was the dependent variable of interest in our study. Postoperative in-hospital
LOS has been used in numerous studies and review articles as a reflection of postoperative
complications (14, 51, 95, 103, 115, 117, 170). Most studies evaluating operative outcome
consider morbidity and mortality rates within the first 30 postoperative days (3.6-
A number of studies examined in-hospital LOS as a global measure of outcome and as a reflection of relative operative risk. Even those that did not examine overall in-hospital LOS, examined length of admission to the ICU (in days) as a measure representative of postoperative complications.

The investigators wanted to employ a global outcome measure that could encompass all operative complications (including death). As mentioned above, some authors have attempted this by using in-hospital LOS postoperatively. But this necessitates a separate analysis for morbidity and for mortality (or even separate analyses for different classes of complications). Otherwise, a patient might die on the third postoperative day, yet receive a LOS score of only 3 days. Statistically this patient would appear to have an excellent outcome, but in fact he would have suffered the worst possible complication. Moreover, thoracic surgery patients may experience a wide array of complications (bronchopleural fistula, respiratory failure, pulmonary embolus, intractable vomiting, myocardial infarction, anaemia, electrolyte imbalance, stroke, etc.) of varying degrees of severity and of varying etiologies. The sequelae range from the relatively minor (prolonged air leak) to the moderately severe (pneumonia) to death (the worst possible outcome).

Because of this problem, the authors chose to employ LOSOH in the first 30 postoperative days as representing all possible complications. Therefore the patient who dies on postoperative day 3 will receive a score of 0, the worst possible score. A subject who appears to do well and is discharged home 5 days postoperatively only to be readmitted 4 days later for a further 2 weeks would receive a LOSOH of 11 days. If he had not been readmitted, his LOSOH would have been 25 days. Thus a lower score reflects a poorer outcome. This strategy did prove useful in that there were a number of subjects who experienced prolonged admission, some who required urgent readmissions, and one who died at postoperative day 2 (see Chapter 4, Results). A consideration of the independent variables employed follows.
PFTs:

Data concerning FEV\textsubscript{1} and DLco were collected by the investigator after data collection was complete, so as to minimize bias. This necessitated a review of the clinic and inpatient charts, and, on occasion, the PFT lab records. Whenever possible, patients underwent laboratory testing in pulmonary function laboratories employing strict ATS guidelines for testing (97-99). This was to ensure consistency of measurement. The laboratory at our facility applies ATS standards.

Functional Exercise Tests:

6-minute walk tests:

This topic has been considered in the previous section (Procedure), and in the literature review in Chapter 1. Patient instructions regarding the performance of this test appear in Appendix E.

Stair climbing tests:

Subjects were asked to perform a timed, symptom-limited stair-climb (as described in Holden, 1992 (115)). The stair-climb was performed in a hospital stairwell, ensuring consistent conditions of lighting, step height (0.19 m) and depth, flooring surface, railing configuration, numbering of floors, etc. Using a finger pulse oximeter (Nellcor, N-100), the tester took a reading of HR and oxygen saturation at the bottom of the flight to be climbed. She waited until a stabilized reading was obtained. In this test, the data collector walked with the patient in order that she could monitor the HR and oxygen saturation continuously. The rater made every effort to follow the subject’s pace and instructed the subject to climb at their own pace and that the rater would follow the subject. Subjects either stopped the test with the onset of symptoms (dyspnea or fatigue or pain) or were asked to stop based on HR response or desaturation. A HR increase of 35 to 40 bpm was
considered an indication for the tester to stop the test. It should be noted that most subjects
demonstrated excellent motivation for the test, but that most experienced the onset of
symptoms immediately upon stopping the climb as well as a rapid increase in HR after
stopping.

The formulae for calculating workload and VO₂ max achieved appear below (see
Holden et al., 1992; Olsen et al., 1991) (14, 115):

\[
\text{Work} = \text{height of step} \times \text{steps/minute} \times \text{weight (kg)} \times 0.1635
\]
\[
\text{VO}_2 = (5.8 \times \text{weight}) + 1.51 + 10.1 \times \text{work}
\]

The rater timed the subject with a stopwatch and counted the number of stairs
climbed. Immediately upon stopping the climb, the tester recorded the maximum HR and
lowest oxygen saturation detected. The subject was presented with the modified Borg RPE
(see Appendix C) and asked to rate their perceived exertion during the stair climb (they
were already familiar with the scale from the 6-minute walk test). All these measures were
recorded by the data collector.

*Health Status Instrument:*

Self-administration of the questionnaire was conducted preoperatively and
postoperatively. The rationale for self-administration was outlined previously. The
EORTC instrument takes approximately 10 minutes to complete.

Because this instrument represents a profile, and not a single index score (no
overall score), it was necessary to choose the scales to be used for the multiple regression
analysis. For methodological reasons (a limited number of independent variables), we
chose to limit the scales for use in the regression equation to the following: SF scale, EF
scale, and global QL scale. All are multi-item scales which have exhibited acceptable
reliability and validity in field tests of the QLQ-C30 (24, 123, 149). All the other scales
were examined using descriptive statistical techniques (in secondary analyses), but were not part of the primary analysis.

Other variables of interest:

Other measures were collected for the descriptive analysis, namely: BMI, extent of resection, duration of surgical procedure, postoperative analgesia technique (epidural vs. intravenous morphine vs. patient controlled analgesia), and so forth. These were collected by chart review following the data collection phase of the study.

Data Analysis:

Sample Size:

A sample size of 70 was estimated for the primary statistical analysis in this trial (multiple linear regression). This sample size estimation was based on the method described by Kleinbaum and Kupper (1978) and Norman and Streiner (1994) (78.54). whereby the “number of data be considerably more than the number of variables” (54). The sample size should be at least 10 times the number of variables included in the regression equation (54.78). Consultation with a statistician (Dr. S. Bull) confirmed this approach. In addition, we based our sample size on issues of feasibility. Historical data at our facility and the recent addition of a third surgeon to our staff meant that we expected to recruit approximately 70-80 subjects within a 12-month period (the time allotted for data collection). Moreover, previous experience during a related trial indicated that subject compliance would be high (19 patients, no drop-outs). Based on this information, we estimated a drop-out rate of 10% for the present study. Originally, 76 subjects were recruited, but 6 had to be withdrawn from the trial at the end of data collection, as they had received different surgeries from that stipulated in the inclusion criteria.
Primary Analysis:

The dependent variable for the primary data analysis is LOSOH in the first 30 days, measured in days. The independent variables of interest are FEV, (% predicted), DLco (% predicted), 6-minute walk distance (m), EF score (0-100), the SF score (0-100), and the global QL score (0-100) (149). All the questionnaire scales are presented as standard scores (149). The QLQ-C30 and its lung module (QLQ-LC13) do not have a single index score, so no overall score is available for analysis. However, the global QL score is a very broad measure of perceived health-related quality of life (149).

The individual items (questions) within the EORTC scales employed use Likert method scaling (21.149). One criticism frequently levelled at such health status items is that they represent ordinal level data (as distinct from interval level, which is a minimum requirement for adopting parametric statistical techniques) (68). But Streiner and Norman point out that under most circumstances “one can analyse data from rating scales as if they were interval without introducing severe bias.” (68, p. 29) This is a commonly adopted assumption in health status measurement (and a topic of considerable debate amongst researchers): fortunately, the experiences of previous researchers have demonstrated that these simple scoring methods are remarkably robust (149). All three of the scales employed from the EORTC instrument are multi-item scales whose scores are converted from summated raw scores to standard scores (0-100) using linear transformation statements (21.149). Streiner and Norman (1994) observe that such linear transformations preserve the normal distribution of the raw scores (68): because of this normalized distribution of scores, they are more likely to meet the assumptions of parametric statistics (68).

Multiple linear regression analysis was performed with the dependent and independent variables mentioned above, using a forward interactive stepwise approach. Prior to performing this multiple regression, univariate regression and correlation
(Pearson's r) were performed in order to explore these variables in depth (54.78,158-162). p values of 0.05 or less were considered significant.

*Secondary Analysis:*

Descriptive analysis of the secondary variables of interest (extent of resection, duration of anaesthesia, ABGs, BMI, and so forth) and their relation to the dependent variable (LOSOH) were performed (including graphical analysis and summary statistics). For the 6-minute walk test results, one-way repeated measures analysis of variance (ANOVA) was performed. Paired t-tests were used for comparing groups.

All statistical analysis was undertaken using the program SYSTAT for Macintosh (163).
CHAPTER 4

RESULTS

Of the 102 patients who met the eligibility criteria, 3 refused to participate (2 patients cited severe anxiety regarding the impending surgical procedure, and a third refused as she was unable to ambulate more than 60 m without severe dyspnea or sit unsupported during the interview). None of the 99 subjects recruited expressed reluctance to complete the preoperative testing protocol.

Table 4.1 lists the reasons for the investigator withdrawing subjects from the study. 29 subjects underwent preoperative workup only to be removed from the study: a summary of their preoperative data appears in the table. These data can be used for comparison with those of the study sample, listed in Table 4.3. The majority of subjects who did not undergo resection were deemed inoperable based on their tumour staging. 19 subjects who were deemed unresectable underwent preoperative work-up and these data appear in Table 4.1. One 86-year-old subject underwent initial workup for surgery, but declined surgery, despite potential resectability. This 86-year-old subject noted that she did not want an operation. “because my quality of life is really good right now, and I’d rather see how things go without the operation.”

9 subjects were withdrawn from the study as they underwent a surgical procedure different from that required by the entry criteria (thoracotomy and lung resection) (1 underwent exploratory sternotomy, 4 underwent VATS (thoracoscopic wedge resection) rather than full thoracotomy, and 4 underwent sternotomy and bilateral wedge resections. The results of the preoperative assessments for these patients appear in Table 4.1. An additional 27 subjects were identified who underwent thoracotomy and lung resection, but these patients were excluded from the study as they did not speak English (they did not
meet the inclusion criteria stipulated in Chapter 3). Thus their numbers do not appear in the analysis.

In all, 70 subjects met the eligibility criteria (Chapter 3). Table 4.2 provides demographic and operative information regarding the study sample. Simple resections refer to procedures that entail lung resection and thoracotomy alone. Extended resections refer to those patients who underwent not only thoracotomy and lung resection, but also concomitant chest wall resection, intrapericardial dissection, brachial plexus decompression, or additional incisions (thoracotomy plus sternotomy, thoracotomy plus laparotomy, etc.). Table 4.3 provides summary statistics of the independent variables of interest.

Table 4.1 Reasons for Removal from Study

<table>
<thead>
<tr>
<th>Reason for Exclusion</th>
<th>No</th>
<th>Mean FEV₁ (% pred.)</th>
<th>Mean DLCO (% pred.)</th>
<th>Mean QL score</th>
<th>Mean SF score</th>
<th>Mean EF score</th>
<th>Best Walk (m)</th>
<th>BMI (kg/m²)</th>
<th>Age (yrs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unresectable</td>
<td>19</td>
<td>80 ± 17.5</td>
<td>63 ± 14.2</td>
<td>57 ± 22</td>
<td>80 ± 28</td>
<td>69 ± 24</td>
<td>362.53 ± 119.70</td>
<td>23.8 ± 4.1</td>
<td>68 ± 10.9</td>
</tr>
<tr>
<td>Different Procedure</td>
<td>9</td>
<td>87 ± 25.7</td>
<td>81 ± 24.8</td>
<td>69 ± 13</td>
<td>87 ± 14</td>
<td>79 ± 22</td>
<td>419.27 ± 136.48</td>
<td>28.6 ± 7.7</td>
<td>56 ± 7.3</td>
</tr>
<tr>
<td>Declined Surgery</td>
<td>1</td>
<td>not tested</td>
<td>not tested</td>
<td>92</td>
<td>100</td>
<td>90</td>
<td>184.39 ± 26.3</td>
<td>26.3 ± 86</td>
<td>65</td>
</tr>
</tbody>
</table>
# Table 4.2 Characteristics of Sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. patients</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40</td>
<td>57.1</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>42.9</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>65 ± 11.2</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>29 - 83</td>
<td></td>
</tr>
<tr>
<td>Surgical Procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wedge resection</td>
<td>9</td>
<td>12.9</td>
</tr>
<tr>
<td>Lobectomy</td>
<td>37</td>
<td>52.8</td>
</tr>
<tr>
<td>Lobectomy + wedge(s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt; 2 lobes resected)</td>
<td>9</td>
<td>12.9</td>
</tr>
<tr>
<td>Bilobectomy</td>
<td>4</td>
<td>5.7</td>
</tr>
<tr>
<td>Pneumonectomy</td>
<td>11</td>
<td>15.7</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple</td>
<td>49</td>
<td>70</td>
</tr>
<tr>
<td>Extended</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>Tumour Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary lung cancer</td>
<td>55</td>
<td>78.6</td>
</tr>
<tr>
<td>Metastases</td>
<td>11</td>
<td>15.8</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>2</td>
<td>2.8</td>
</tr>
<tr>
<td>Benign lesions</td>
<td>2</td>
<td>2.8</td>
</tr>
<tr>
<td>LOSOH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>21.1 ± 5.7</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3: Study Sample: Summary Statistics of Primary Outcome Variables (n = 70)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV₁ (% pred.)</td>
<td>88.3 ± 29.0</td>
</tr>
<tr>
<td>Dlco (% pred.)</td>
<td>76.3 ± 17.9</td>
</tr>
<tr>
<td>QL Score (std. score)</td>
<td>65 ± 24.4</td>
</tr>
<tr>
<td>SF Score (std. score)</td>
<td>74 ± 24.7</td>
</tr>
<tr>
<td>EF Score (std. score)</td>
<td>72 ± 23.6</td>
</tr>
<tr>
<td>Best Walk (m)</td>
<td>390.79 ± 105.14</td>
</tr>
</tbody>
</table>

**Mean ± SD

Primary Analysis:

The dependent variable for the primary data analysis is LOSOH within the first 30 postoperative days. A higher score represents a shorter in-hospital convalescence and therefore a better outcome. A low LOSOH score is taken to represent complications. A death within the first 30 days receives a score of 0. The mean (SD) LOSOH for our sample was 21.1 (± 5.7) days. Table 4.4 catalogues all complications suffered by patients within the sample (25 complications in total, for an overall complication rate of 35.7%). Each complication occurred in a separate subject, with no subjects experiencing more than one complication.

Table 4.4 Complications Suffered (Entire Sample, n = 70)

<table>
<thead>
<tr>
<th>Complication</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>1</td>
</tr>
<tr>
<td>Respiratory Failure</td>
<td>3</td>
</tr>
<tr>
<td>Cardiac failure (prolonged ischaemia)</td>
<td>1</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>2</td>
</tr>
<tr>
<td>Atelectasis (req bronchoscopy)</td>
<td>2</td>
</tr>
<tr>
<td>Arrhythmias (Atrial fibrillation)</td>
<td>5</td>
</tr>
<tr>
<td>Dyspnea NYD</td>
<td>3</td>
</tr>
<tr>
<td>Infected pneumonectomy space</td>
<td>1</td>
</tr>
<tr>
<td>Pulmonary edema</td>
<td>1</td>
</tr>
<tr>
<td>Prolonged air leak</td>
<td>4</td>
</tr>
<tr>
<td>Pyloric stricture/ileus</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 4.1 Frequency Distributions of LOSOH for Entire Sample and Group A
The LOSOH variable demonstrates a bimodal distribution, indicating that the extreme outliers (4 subjects) constitute a separate subpopulation of the sample. The histogram of LOSOH for the sample (n=70) depicts this unusual distribution (Figure 4.1 top view). Skewness of the entire sample distribution was -2.58 SD units. The remaining subjects (n = 66; those who had a LOSOH > 5 days) show a more Gaussian distribution when compared with that of the entire sample (although the histogram in Figure 4.1 (bottom graph) is negatively skewed (-0.95 SD units)). The 4 furthest outliers within the larger sample cluster close to a LOSOH of 0 days, with the best outcome being a LOSOH of 5 days. Because of the bimodal distribution of the original sample, it was necessary to analyze the two subgroups separately. We searched for any factors that might account for the extremely poor LOSOH scores suffered by these 4 outliers (did all these patients require ICU admission postoperatively, did they all have the same procedure, were they characterized by some other unusual feature?). Unfortunately, we were unable to identify any factor which set them apart from the rest of the sample, except for their extraordinarily poor outcomes. We subsequently analyzed our results for the entire sample and for both subgroups separately. The subgroup with a LOSOH > 5 days will be referred to as Group A (n = 66), while those with a LOSOH ≤ 5 days will be labelled Group B (n = 4).

The 4 patients in Group B differed from the sample as a whole and from Group A in several respects. Because of this group's small number, it was not possible to perform regression and other parametric tests on this data. Descriptive analysis is the only feasible approach with this scenario. Table 4.5 summarizes the salient characteristics of Group B. Of the 4 patients in this group, 3 had a LOSOH of 0, with one death (at postoperative day 2). 2 subjects suffered prolonged respiratory failure and persistent infections (one patient became colonized with an antibiotic resistant bowel infection and on follow-up had not returned home in the 6 months following the operation). The one subject with a LOSOH of 5 days suffered massive cardiac complications and required emergency coronary angioplasty postoperatively. Of note, his past medical history included severe coronary
artery disease. The Group B subpopulation did not differ significantly from the sample as a whole in terms of operative procedure performed (either in amount of lung tissue resected, or in extent of procedure). 2 underwent lobectomy, 1 underwent bilobectomy, and 1 a lobectomy plus wedge resection. 2 underwent extended procedures, the other 2 received simple excisions. These 4 subjects reported satisfactory emotional and social function preoperatively. All 4 rated their global quality of life as good to excellent. In terms of preoperative PFTs, 2 patients had severely diminished FEV₁ scores, but the other 2 had normal values. We can say little about DLco values, since only 1 Group B patient had these values available for analysis. This patient suffered the one perioperative death (worst outcome), and he had a DLco of 90% predicted. 2 subjects had 6-minute walk distances over 350 m, while the other 2 subjects had distances of between 250 and 300 m. The Group B patients consisted of 2 men and 2 women. This subpopulation did differ significantly in terms of age. The Group B patients were significantly older than the sample as a whole (mean age 65.5 ± 11.2 years) and the Group A subjects (mean age 64.7 ± 10.9 years). Group B had a median age of 78 years, with a range of 74 to 83 years. The age ranges for the entire sample and for Group A were identical (29 to 83 years). No significant correlation between age and LOSOH was found in Group A, although analysis of the entire sample revealed a significant (albeit weak) negative correlation between these 2 variables (Pearson's r = -0.37, p=0.03). Figure 4.2 depicts this correlation. The removal of the Group B subjects from the sample nullified this relationship. No differences regarding gender and marital status were noted between Groups A and B. 2 of the 4 subjects in Group B required repeated Intensive Care Unit admissions (and mechanical ventilation) postoperatively, one for cardiac failure, the other for respiratory failure.
<table>
<thead>
<tr>
<th>Demographic Data</th>
<th>Subject B1</th>
<th>Subject B2</th>
<th>Subject B3</th>
<th>Subject B4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>74</td>
<td>82</td>
<td>83</td>
<td>78</td>
</tr>
<tr>
<td>Sex</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Married</td>
<td>Married</td>
<td>Widowed</td>
<td>Widowed</td>
</tr>
<tr>
<td>Living Arr.</td>
<td>Family</td>
<td>Family</td>
<td>Alone</td>
<td>Alone</td>
</tr>
<tr>
<td>LOSOH (d)</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PFTs: (% pred)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV₁</td>
<td>34</td>
<td>90</td>
<td>56</td>
<td>88</td>
</tr>
<tr>
<td>DLco</td>
<td>N/A</td>
<td>90</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>HRQL data:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF Score</td>
<td>100</td>
<td>100</td>
<td>83</td>
<td>100</td>
</tr>
<tr>
<td>SF Score</td>
<td>100</td>
<td>100</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Global QOL</td>
<td>92</td>
<td>100</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>Walk distance (m)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1</td>
<td>231.00</td>
<td>449.54</td>
<td>290.73</td>
<td>198.00</td>
</tr>
<tr>
<td>Trial 2</td>
<td>369.55</td>
<td>388.37</td>
<td>297.00</td>
<td>252.45</td>
</tr>
<tr>
<td>BMI</td>
<td>N/A</td>
<td>26.2</td>
<td>27.0</td>
<td>21.6</td>
</tr>
<tr>
<td>ABGs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PaCO₂</td>
<td>38</td>
<td>40</td>
<td>N/A</td>
<td>45</td>
</tr>
<tr>
<td>PaO₂</td>
<td>65</td>
<td>94</td>
<td>N/A</td>
<td>75</td>
</tr>
<tr>
<td>Stair climb Work (W)</td>
<td>84.5</td>
<td>149.8</td>
<td>76.7</td>
<td>N/T</td>
</tr>
</tbody>
</table>
Figure 4.2: Scatterplots of LOSOH vs. Age: Entire Sample (n=70) and Group A (n= 66)
Examination of Variables:
Dependent Variable:

As mentioned above, the dependent variable of interest has a bimodal distribution, making it difficult to analyze using parametric statistics (these assume a normal distribution). We attempted a number of approaches to analyze these data.

In cases of severely skewed distributions, transformations of the offending variable are frequently recommended (54, 162). Norman and Streiner (1994) point out that simply looking at the shape of the distribution does not exclude the possibility of attempting a variety of transformations (not simply the one most likely to assist based on statistical theory) (54). They recommend attempting different transformations until the most normal distribution is achieved. This was attempted with LOSOH in our analysis. Given the extreme left skew, the transformation recommended by most authors was \(-1/\log(\text{LOSOH})\) (54,162-163). This did nothing to improve the relative influence of these outlying values of the Group B subjects. The distribution remained skewed and essentially bimodal. The second transformation of choice (for a moderate negative skew) was \(-1/\sqrt{\text{LOSOH}}\) (54,162). Again, this yielded no benefit regarding the dependent variable’s distribution. Positive and negative reciprocal transformations were equally fruitless. For this reason, analysis proceeded examining the entire sample, and then as two groups separately. The two groups did not differ in terms of severity of pulmonary function compromise, walk distance achieved, and health status scores.

In examining LOSOH, we assume that this value represents postoperative complications. However, it is possible that factors other than complications could be influencing the distribution of the dependent variable. The Group B patients all experienced extreme postoperative complications (death, respiratory or cardiac failure). It was necessary to examine the LOSOH distribution of Group A to determine whether LOSOH was related to complications or was reflecting non-operative factors (for instance.
surgical preference, administrative influences, or a consistent pattern of pre-weekend discharges). Refer to Table 4.4 for a list of complications. 21 of these complications were suffered by subjects in Group A. We divided the Group A patients into 3 equal groups based on LOSOH (14 to 17 days, 18 to 22 days, and 23 to 26 days).

100% of patients with LOSOH of 14 to 17 days suffered complications (5 subjects in total with the following complications: 1 pulmonary edema, 1 postoperative ileus, 1 pyloric stricture, and 2 with protracted dyspnea). 3 of these 5 patients required readmission to hospital for management of these complications. 1 of these patients died in the third postoperative month (he remained in hospital for the subsequent 3 months). 1 was oxygen dependent 6 months postoperatively (she had been discharged home on room air) and one is presently receiving palliative care for distant recurrence.

63.2% of patients (12 of 19) experiencing a LOSOH between 18 and 22 days sustained complications. The 12 complications were: 2 severe atelectases requiring bronchoscopy, 2 pneumonias, 1 respiratory failure, 1 infected pneumonectomy space, 1 protracted dyspnea (etiology unknown, but suspicious for pulmonary embolus), 3 postoperative arrhythmias (atrial fibrillation), and 2 prolonged air leaks. The patient readmitted with respiratory failure required prolonged mechanical ventilation and died on postoperative day 40. The patient with the infected pneumonectomy space was readmitted both during the perioperative period and on 3 subsequent occasions. On later readmissions he was found to have an aggressive local recurrence and he died at approximately 3 months postoperatively.

9.5% of patients with a LOSOH of 23 to 26 days (4 of 42 patients) experienced operative complications. The complications sustained were: 2 arrhythmias (atrial fibrillation) and 2 prolonged air leaks. None of these 4 subjects required readmission to hospital and all reported being fully recovered at 3 months postoperatively.

A review of data gathered from the hospital records indicated that only 3 patients' discharge dates were delayed for non-medical reasons. 2 patients were kept in hospital
Figure 4.3 LOSOH vs. Surgeon (Sample and Group A)
because they lived alone and at long distances from the facility (one lived 60 miles away, the other more than 200 miles from our facility). 1 subject lived with her elderly husband and was discharged home on postoperative day 7 (rather than day 5) as she was concerned about her ability to cope at home (moreover, she lived over 400 miles away from the facility). The LOSOH for these patients ranged from 18 to 25 days. 2 additional patients were discharged home at 9 days postoperatively secondary to the medical team’s concern regarding postoperative monitoring; both of these patients had histories of severe coronary artery disease. Neither patient suffered a complication, but the team felt it was better to delay discharge in the interest of vigilant monitoring.

Finally, we documented differences in discharge based on attending surgeon. These results appear in Figure 4.3. No differences in discharge practice between the 3 surgeons were observed. It should be noted that surgeon 2 only had 9 patients participating in the study.

The complication profile, the discharge profile based on attending surgeon, and the chart review for non-medical reasons for delay of discharge indicate that LOSOH within the first 30 postoperative days is a measure which does represent postoperative complications.

*Independent Variables:*

Figures 4.4 to 4.9 (following pages) depict the frequency distributions of the independent variables in this study. The frequency distributions are given for the sample as a whole and for Group A. There appears to be little difference between the entire sample and Group A (in terms of summary statistics) with respect to the independent variables. The negatively skewed distributions observed for the variables EF score and SF score demonstrate a ceiling effect, indicating that the majority of patients coming into hospital reported high levels of social and emotional function preoperatively (21).
Figure 4.4 Graphs of Preoperative EF Scores (Sample and Group A)
Figure 4.5: Graphs of Proportion of SF Scores (Sample and Group A).

Histogram of SF Scores (Group A)
Histogram of Global QL Scores (Entire Sample)

Histogram of Global QL Scores (Group A)

Figure 4.6 Graphs of Preoperative Global QL Scores (Sample and Group A)
Histogram of FEV1 % predicted (Entire Sample)

Histogram of FEV1 % predicted (Group A)

Figure 4.7 Graphs of Preoperative FEV1 Scores (Sample and Group A)
Figure 4.8 Graphs of Preoperative DLco (Sample and Group A)
Histogram of Mean Walk Distance (m) (Entire Sample)

Histogram of Mean Walk Distance (m) (Group A)

Figure 4.9 Graphs of Preoperative 6-minute Walk Distance
Prior to performing the multiple linear regression analysis of the variables, we performed separate univariate linear regression analyses for each of the independent variables. This permitted exploration of the straight-line relationship between the various independent variables and LOSOH; however, it did not depict the interrelationships between these independent variables (78). Multiple regression allows us to explore intercorrelations among the independent variables (78).

Table 4.6 shows the simple regression analyses of the entire sample and of Group A. These analyses demonstrated significant differences between Group A and the entire sample, both in terms of Pearson's correlation coefficient \( (r) \) and \( R^2 \). We will first consider the regression analysis of the sample as a whole \( (n = 70) \). These values appear on the left of Table 4.6. Diagnostic tests of the residuals for the univariate regression analyses confirmed that the excluded subjects (Group B) were indeed outliers (160). Walk distance was the only independent variable that predicted a significant proportion of the variance in LOSOH, accounting for 9.8 % \( (p=0.009) \) of the variance of LOSOH. We employed the mean 6-minute walk distance achieved \( (\text{mean (SD)} = 374.97 \pm 102.74 \text{ m}) \). Both Trial 1 walk distance and mean walk distance showed a weak association with LOSOH, with Pearson's \( r \) values of 0.34 and 0.31 respectively \( (p < 0.05 \text{ for both correlations}) \). A detailed analysis of the walk test results appears later in this chapter.

Table 4.5 also contains the results of simple linear regression of the other independent variables of interest, but only one of these variables demonstrated a trend towards a significant (albeit weak) association with LOSOH, namely FEV\(_1\) \( (r = 0.25) \). However, the \( R^2 \) of FEV\(_1\) to LOSOH did not reach significance. As mentioned above, age demonstrated a rather weak (but significant) negative correlation with LOSOH \( (r = -0.37, p=0.003) \).
Table 4.6: Univariate Regression of Independent Variables with LOSOH

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Whole Sample ( (n = 70) )</th>
<th>Group A ( (n = 66) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mult R</td>
<td>R²</td>
</tr>
<tr>
<td>FEV₁</td>
<td>0.250</td>
<td>0.061</td>
</tr>
<tr>
<td>DLco</td>
<td>0.038</td>
<td>0.001</td>
</tr>
<tr>
<td>EF Score</td>
<td>0.174</td>
<td>0.030</td>
</tr>
<tr>
<td>SF Score</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Global QL score</td>
<td>0.040</td>
<td>0.002</td>
</tr>
<tr>
<td>Mean Walk Distance</td>
<td>0.313</td>
<td>0.098</td>
</tr>
<tr>
<td>Age</td>
<td>0.370</td>
<td>0.127</td>
</tr>
</tbody>
</table>

The corresponding \( R^2 \) was also statistically significant for the sample as a whole (\( R^2 = 0.127 \), \( p=0.003 \)).

When simple regression and correlation were performed with the subjects stratified into Group A (LOSOH > 5 days), a different picture emerged. Starting with the 6-minute walk test, the correlation of the mean walk distance with LOSOH rose to 0.38 (\( p<0.005 \)) and the \( R^2 \) was 0.144 (\( p<0.005 \)). However, a number of other variables also became more strongly associated with LOSOH with the removal of the Group B patients. The right column of Table 4.6 shows a moderate correlation between the global QL Score (QL) and LOSOH (Pearson’s \( r = 0.38 \), \( p<0.005 \)). Simple linear regression revealed that QL accounts for over 14 per cent of the variance in LOSOH (\( p<0.005 \)). Moreover, the SF score was also significantly correlated (albeit weakly) with the dependent variable (\( r = 0.28 \), \( p<0.05 \)), with an \( R^2 \) of only 0.077 (\( p<0.05 \)). EF scores on the health status instrument were not associated with curtailed LOSOH. Of interest is the absence of an association between the pulmonary function measures (FEV₁ and DLco) and postoperative LOSOH for Group A, in contrast with the entire sample. This relationship should be borne in mind when we proceed to the multiple regression analysis.
After performing this exploratory univariate analysis of the dependent and independent variables, multiple linear regression analyses were executed using the data for the entire sample and for Group A. These results are the focus of the following section. Given the paucity of associations seen in the separate measures of association between the dependent variable and each of the independent variables taken in sequence, it is legitimate to ask whether we should proceed with a multiple regression analysis. While a certain amount of information has been gained by performing these individual analyses with the variables of interest, more information can be gained from a multiple linear model (78). Even if none of the independent variables is predictive of LOSOH, the multiple regression analysis may provide some significant insights concerning the relationships among these potential predictors (for instance, multicollinearity) (54, 78).

Multiple Linear Regression:

The planned comparison for this study was a multiple linear regression, in order to evaluate the relative predictive ability of each of the independent variables in the equation

\[ Y (\text{LOSOH}) = b_0 + b_1 FEV_1 + b_2 DLco + b_3 EF + b_4 SF + b_5 QL + b_6 Walk \]  

(54)

While the relationship of interest is a linear one, the relative contribution of each predictor variable is not necessarily equally proportioned. In other words, there may be significant intercorrelation among the variables of interest (54). The best graphical depiction of a multiple linear regression is a plane. However, depending on the relative sign of the individual correlations examined, the orientation will shift (78). This is why it is impossible simply to extrapolate the overall relationship from the individual univariate analyses of the independent variables (78). While the F test will provide important information regarding the significance of the overall predictive power for all independent variables entered into the equation, important information is missing if we rely on this
statistic alone (78). "The partial F test assesses whether the addition of any specific independent variable, given the others already in the model, significantly contributes to the prediction of Y." (78, p. 141). These values have important implications for deciding which variables are entered into the equation, and which will be disregarded as irrelevant. Finally, a correlation matrix of the independent variables informs the investigator about the degree of overlap or redundancy among the predictor variables (avoidance of multicollinearity). This provides a further check on the number of variables that ought to be entered into the equation. The goal of multiple linear regression is to determine which predictor variables independently contribute to the variance in the dependent variable (161). One of the reasons much attention is devoted to examining the predictor variables and limiting the number entered into the regression equation relates to the principle of increasing probability of finding significance with an increasing numbers of tests: the probability of finding a significant predictor variable rises with the number of variables entered into the equation (54).

A number of assumptions are inherent to multiple linear regression. One of the primary assumptions is that the sample under consideration is assumed to follow a normal (or Gaussian) distribution (160). This is one of the primary arguments for stratifying the sample into Groups A and B, since Group A exhibits a more normal distribution than that of the entire sample. While the dependent variable (LOSOH) in Group A does not follow a perfectly normal curve, the distribution does lose its bimodal distribution (please refer again to Figure 4.1). Diagnostic tests of the residuals for the univariate regression analyses outlined in the previous section confirm that the excluded subjects are indeed outliers (160). Such diagnostic tests (e.g. Cook’s distance) indicate whether outliers are exercising undue influence on the total regression equation. Another important assumption inherent to multiple linear regression is that the error component (residuals) is randomly distributed. Again, various diagnostic tests of the residuals will reveal whether or not the error terms are randomly distributed, or are perhaps associated with another (confounding) factor.
In our sample \((n=70)\), residual diagnostics indicated that these 4 subjects were markedly affecting the distribution. Their subsequent removal (multiple regression for Group A) revealed that the Group A residuals had a near normal distribution. Graphical depictions of these diagnostic tests appear in Appendix D.

As mentioned above, stratification based on LOSOH into Groups A and B proved fruitful in terms of simple linear regression and correlation. The same strategy was therefore employed for the multiple regression analysis. Obviously, Group B was too small to allow a regression analysis, but Group A yielded some interesting results.

Multiple linear regression analysis was performed, using a forward interactive stepwise approach. We began with an analysis of the entire sample. No predictors were identified. Because age and gender were potentially confounding variables, these were added into the regression equation to see if any significant change transpired. Gender did not alter the results. The addition of the variable age identified \(\text{it as the only significant predictor of LOSOH.}\) No significant interaction effects were identified (e.g., QL x Age, etc.). Diagnostic tests of the regression identified the 4 Group B subjects as the only significant outliers. Backwards stepwise regression did not alter the results.

**Table 4.7: Multiple Linear Regression Table With Age as Predictor**

(Entire Sample)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>272.087</td>
<td>1</td>
<td>272.087</td>
<td>9.548</td>
<td>0.003</td>
</tr>
<tr>
<td>Residual</td>
<td>1937.856</td>
<td>68</td>
<td>28.498</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>(t)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>32.82</td>
<td>3.823</td>
<td>8.584</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>-0.178</td>
<td>0.058</td>
<td>-3.090</td>
<td>0.003</td>
</tr>
</tbody>
</table>

**Mult. \(R = 0.351\)**

**\(R^2 = 0.123\)**

**Adj. \(R^2 = 0.110\)**

No. obs. = 70
As Group A was sufficiently large, we proceeded with a multiple linear regression analysis of this group. An interactive forward stepwise regression analysis of Group A revealed the following predictors of LOSOH: QL Score and Walk Distance (mean walk), accounting for 22.4 per cent of the variance in LOSOH, and an adjusted $R^2$ of 0.199 ($p<0.05$) (refer to Table 4.8). The standard error of the estimate was 2.50. This time, the addition of the variable age did not produce any significant effect on the analysis. A correlation matrix of all six predictor variables revealed that SF and QL scores were the most highly correlated ($r = 0.50, p=0.02$), indicating significant overlap between these two variables. Even if we included SF score first in the equation, the subsequent addition of QL cancelled most of the contribution of SF to LOSOH. This indicated that SF was acting redundantly in conjunction with QL.

Table 4.8:

a) Multiple Linear Regression Table of Group A

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>112.211</td>
<td>2</td>
<td>56.105</td>
<td>8.957</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>388.343</td>
<td>62</td>
<td>6.264</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>17.291</td>
<td>1.261</td>
<td>13.715</td>
<td>0.000</td>
</tr>
<tr>
<td>Global QL</td>
<td>0.030</td>
<td>0.014</td>
<td>2.156</td>
<td>0.035</td>
</tr>
<tr>
<td>Walk Dist.</td>
<td>0.008</td>
<td>0.003</td>
<td>2.457</td>
<td>0.017</td>
</tr>
</tbody>
</table>

b) Correlation Matrix of Predictor Variables (Group A Analysis):

<table>
<thead>
<tr>
<th></th>
<th>Constant</th>
<th>EF Score</th>
<th>SF Score</th>
<th>QL Score</th>
<th>FEV1</th>
<th>DLco</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF Score</td>
<td>-0.283</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF Score</td>
<td>-0.014</td>
<td>-0.196</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QL Score</td>
<td>0.045</td>
<td>-0.398</td>
<td>-0.504</td>
<td>-0.211</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1</td>
<td>-0.357</td>
<td>-0.113</td>
<td>-0.075</td>
<td>-0.211</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLco</td>
<td>-0.448</td>
<td>-0.065</td>
<td>-0.252</td>
<td>0.358</td>
<td>-0.217</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Walk</td>
<td>-0.306</td>
<td>0.039</td>
<td>0.089</td>
<td>-0.437</td>
<td>0.128</td>
<td>-0.374</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Of interest is the comparatively increased significance in DLco when QL score was entered into the regression equation first. Univariate regression had indicated no significant association between LOSOH and DLco in either Group A or the entire sample. For Group A, DLco showed a trend toward significance within the stepwise regression equation. However, in the multiple regression model, the FEV_1 loses the trend towards significance seen in univariate regression. In fact, its addition into the multivariate regression equation results in a dramatic decrease in adjusted $R^2$ (from 0.199 to 0.154, $p<0.05$). It should be noted that backward stepwise regression yielded similar results (interactive stepping allowed us to enter and remove variables based not solely on perceived/potential statistical significance, but also on clinical significance). Age did not alter the predictive ability of these 4 independent variables in the Group A analysis.

**Post hoc Tests:**

Following the multiple regression analysis outlined above, a number of *post hoc* tests were run on some additional variables of interest. We were concerned about our inability to perform 3 trials of the 6-minute walk test (as recommended in the COPD and CHF literature) and wished to evaluate the reliability of this measure when only 2 trials were available. Besides repeated measures ANOVA, we evaluated the test-retest reliability, as well as the inter-rater and intra-rater reliabilities of this functional exercise test. We wondered whether various preoperative measures not included in the primary analysis might show an association with operative outcome (for instance preoperative room air ABGs, body mass index (BMI), VO_2 and workload achieved during stair climbing). Of particular interest was whether any of the other scales comprising the questionnaire might be predictive of operative outcome. Administration of neoadjuvant chemotherapy and/or radiation therapy and its correlation with LOSOH was examined. Besides exploring such
preoperative measures, postoperative recovery of functional exercise capacity (6-minute walk distance) and health status (QLQ-C30 and QLQ-LC13) was tracked. Measurements were taken immediately prior to discharge home, and at the first two follow-up appointments in the outpatient thoracic surgery clinic (please refer to the Figure 3.1 in Chapter 3). Postoperative readmissions to hospital were likewise tracked during that period. The results of our secondary analyses appear below. We begin with a consideration of the 6-minute walk test.

**Additional Preoperative Measures Evaluated:**

**6-Minute Walk Test:**

Considerable attention was devoted to the reliability of the 6-minute walk test as employed in our study. We compared the test-retest reliability for patients who received both 2 and 3 repeated measures. As 2 raters were used, inter-rater reliability was also examined. The reliability coefficients for the 6-minute walk ratings within the study sample appear in Table 4.10. In addition to reliability testing performed with data obtained from the study sample, we elected to test a small subgroup of healthy normals (n = 14) to examine the upper limits of the test. These healthy normals served as a comparator group with the study sample. In addition, we wanted to know how similar our patients were to patients with documented COPD. We compared the results of our sample with those obtained from patients with documented COPD (data collected from another facility for another study). All of the COPD patients had undergone 3 tests to determine a baseline. The results of all these analyses appear below.

**Study Subjects:**

In all, 70 subjects met the eligibility criteria for the study. Of these subjects, 2 were excluded from the 6-minute walk results secondary to lower extremity orthopaedic pathology limiting their ability to participate in the walk test. Our interest lies in cardio-
pulmonary restrictions to distance walked, not any and all restrictions. Restrictions secondary to orthopaedic impairments render these subjects incomparable in terms of the goals of the test in the lung resection population. One of these subjects had a pre-existing above-knee amputation, the other a Kotz prosthesis of the lower extremity, both secondary to osteosarcoma. Therefore, the data on 6-minute walk tests appearing in this section will consist of that collected on the 68 remaining subjects.

Besides evaluating reliability, it was important to know the overall change that had occurred between the 2 trials (in those subjects receiving 2 tests), and the change experienced by the 6 subjects underwent 3 tests (as it is assumed that a training effect occurs over the course of the 3 trials, as described in the literature). We performed a repeated measures analysis of variance (ANOVA) on those who received 2 trials \( n = 50 \) and compared them with those few subjects who received 3 trials \( n = 6 \). The results of this analysis appear in Table 4.9. There was no significant improvement with repeated testing, either in a 3-trial model or in a 2-trial model. There was likewise no significant difference found between Walk 1 and Walk 2.

Besides evaluating the added benefit of a third trial, we also investigated various types of reliability (in essence, the reproducibility) of the 6-minute walk in the lung resection population. For the first two walks we examined the overall test-retest reliability \( n = 55 \), the intra-rater reliability of the primary rater (investigator) \( n = 32 \), and the inter-rater reliability (between the first and second rater, \( n = 23 \)). The results appear in Table 4.10 (following page). Following the Pearson's correlation coefficients, we also calculated the Intraclass Correlation Coefficient (ICC) for the walk tests. The ICC more accurately estimates test-retest and inter-rater reliability \( (54.164.172) \). These results also appear in Table 4.10.
Test-Retest Reliability of 6-minute Walk Test (2 Trials)

Figure 4.10
Intra-rater Reliability of 6-minute Walk (Sample)

![Graph showing Intra-rater Reliability of 6-minute Walk Test]

Inter-rater Reliability of 6-minute Walk (n = 23)

![Graph showing Inter-rater Reliability of 6-minute Walk Test]

Figures 4.11 and 4.12 Intra-rater and Inter-rater Reliability of 6-minute Walk Test
Table 4.9: One Way Repeated Measures ANOVA on Walk Tests
(Comparison of Dependent Variable Means of 3 trials to 2)

<table>
<thead>
<tr>
<th>Group</th>
<th>Walk 1</th>
<th>Walk 2</th>
<th>Walk 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 trials</td>
<td>326.15</td>
<td>344.13</td>
<td>324.77</td>
</tr>
<tr>
<td>(n = 6)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>1399.85</td>
<td>2</td>
<td>699.92</td>
<td>0.21</td>
<td>0.82</td>
</tr>
<tr>
<td>error</td>
<td>33898.48</td>
<td>10</td>
<td>3389.85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>DF</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>5903.46</td>
<td>1</td>
<td>5903.46</td>
<td>2.88</td>
<td>0.10</td>
</tr>
<tr>
<td>error</td>
<td>100556.40</td>
<td>49</td>
<td>2052.17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.10: Reliability Coefficients of the 6-minute Walk Test

<table>
<thead>
<tr>
<th>Reliability Test</th>
<th>Pearson's r</th>
<th>p value</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest reliability</td>
<td>0.78</td>
<td>&lt;0.001</td>
<td>0.48</td>
</tr>
<tr>
<td>Intra-rater reliability</td>
<td>0.89</td>
<td>&lt;0.001</td>
<td>0.36</td>
</tr>
<tr>
<td>Inter-rater reliability</td>
<td>0.74</td>
<td>&lt;0.001</td>
<td>0.46</td>
</tr>
</tbody>
</table>
First, let us address the issue of test-retest reliability (regardless of rater). The Pearson correlation coefficient shows a strong association between the scores on the first test and those on the second. Pearson’s r rises when a single rater is employed (of interest is the low ICC for intra-rater tests). When 2 raters are used, the lowest Pearson’s r is seen, however the ICC calculated was only slightly lower than for the overall test-retest statistic. The ICCs calculated for each form of reliability indicate that within-subjects variations are marked, and the strong associations found between trials 1 and 2 do not reflect a highly reproducible test. The strongest ICC was found for test-retest reliability. Figures 4.10 through 4.12 show the scatterplots of these correlations. A strong linear relationship between the 2 trials is observed for all tests.

Healthy Normals:

As mentioned in the Methodology section of this chapter, we wished to perform the 6-minute walk test under ideal conditions in a sample of young healthy normals who understood the test well. The results of this testing appear below. Please refer to Figure 4.13.

14 subjects were recruited for testing and they underwent 2 trials of the 6-minute walk. The mean age of the subjects was 25.8 (+ 2.8) years. The 2 raters used for this testing were the same 2 performing the testing on the lung resection candidates. All subjects were tested by 2 raters. A repeated measures ANOVA (essentially a paired t-test) was performed. We analyzed the results in terms of the factor ‘rater’ to see if any consistent differences occurred based on the order of rater. When Rater 1 tested first the dependent variable mean for Walk 1 is 698.49 and for Walk 2 is 698.49 (perfect
Test-retest Reliability of 6-minute Walk (Normals)

Figures 4.13
agreement). When Rater 2 rated second Walk 1 is 699.21 and Walk 2 is 699.02 (almost perfect agreement). ICCs calculated were 0.99 and 0.98 respectively.

The limiting factor reported by all healthy subjects was not dyspnea, but tibialis anterior cramping. These results indicate a highly consistent test, limited by the same factor in all subjects. The range for the test in normals is 530 m to 780 m, with the one subject scoring 530 m claiming not to have understood that a maximal effort was required (despite several repetitions of the instructions to this effect by each rater). All other subjects had values ranging from 640 m to 780 m. There were no differences in walk distance observed based on the height and weight of the subject (short subjects could walk virtually the same distances as tall ones).

Comparison of Sample Subjects with COPD Subjects:

Walk test data was made available to the investigator courtesy of Dr. Roger Goldstein (West Park Hospital). The data described 89 patients with COPD who had received 3 trials of the 6-minute walk test. The results of the repeated measures ANOVA of this data appears in Table 4.11. There was a significant difference (p<0.05) between trials 1 and 2 and between trials 3 and 1 but not between trials 2 and 3 (38.41). ICC calculated from the ANOVA table was 0.92.

FEV₁ data (% of predicted) was also available on these comparator subjects. 25 of these subjects had an FEV₁ of less than 60 % predicted. When the results of these comparator subjects and those of our study sample were compared (stratified based on FEV₁), similarities were found between the 2 sets of subjects. If the best trial for the walk test in the present study was employed, and the baseline measure (trial 3) was employed from the COPD comparator data, stratification based on FEV₁ demonstrated significant differences in walk test performance. Subjects with FEV₁ < 60 % predicted had mean (± SD) walk distances of 301.64 (± 101.99) m (COPD) and 336.10 (± 74.06) m (lung resection candidates). In contrast, those patients with FEV₁ > 60 % predicted in both
groups showed significantly higher mean (± SD) walk distances (397.61 (± 85.46) m for COPD patients, 388.66 (± 104.78) m for lung resection candidates respectively). For both groups taken as a whole (unstratified), the mean walk distances are different by only 24.64 m.

Table 4.12 depicts the relative change in repeated trials of the walk test for both the COPD comparators and the lung resection study sample. Only 9 subjects in the lung resection sample had an FEV₁ < 60 % predicted, and no statistically significant improvement in distance walked is seen between trials 1 and 2, although the trend seems to be towards higher scores on the second test. Among the COPD comparators, a statistically significant improvement is seen between trials 1 and 2 (p<0.05), but not between trials 2 and 3. In those lung resection candidates with an FEV₁ > 60 % predicted, the patients walked further than those with pulmonary function impairment, but there was no trend toward increased distances on the second trial. Among the COPD comparators with higher FEV₁ values, there is statistically significant improvement seen between trials 1 and 2 (p<0.05), but not between trials 2 and 3.

Finally, when both the lung resection and the COPD subjects are compared to the young healthy normals, both groups display considerably lower overall scores, as well as more variability among the results.

Table 4.11: Data of Comparator Subjects with COPD:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>391138.40</td>
<td>2</td>
<td>19569.20</td>
<td>17.6</td>
<td>0.000</td>
</tr>
<tr>
<td>error</td>
<td>195250.30</td>
<td>88</td>
<td>1109.40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.12: Walk Test Performance Compared with Lung Resection Candidates and COPD Patients (Stratified by FEV₁)

<table>
<thead>
<tr>
<th></th>
<th>FEV₁ &lt; 60 % predicted</th>
<th>FEV₁ &gt; 60 % predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Walk 1</td>
<td>Walk 2</td>
</tr>
<tr>
<td>COPD sample</td>
<td>280.52</td>
<td>305.64</td>
</tr>
<tr>
<td>Lung Resect. Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walk 1</td>
<td>291.28</td>
<td></td>
</tr>
</tbody>
</table>

Stair-climbing Test:

In addition to the 6-minute walk test, subjects were asked to perform a stair-climbing test (another measure of functional exercise tolerance). 56 patients performed stair climbing. 15 patients did not perform the test: 2 secondary to marked dyspnea on level walking (subjects did not feel comfortable proceeding with the stair climb). 2 secondary to severe myocardial impairment (previously documented). 4 secondary to impaired balance. 2 secondary to marked orthopaedic impairment (1 amputation. 1 recent reconstructive lower extremity surgery--both crutch-walkers), 3 secondary to lack of time to perform the test (OR calling for patient early). and 2 because they had large cumbersome intravenous pumps in situ. 1 patient declined to perform the test, saying that he lacked confidence in his ability. In those subjects who performed this test, the mean (± SD) VO₂ achieved was 27.6 (± 4.8) mL/kg/min., with a mean workload of 141.2 (± 53.7) W during the test. The mean number of stairs climbed was 49 (± 20) steps for the entire sample. The average rate of stair climbing was 62 (± 16) steps/min. Groups A and B did not differ significantly from the sample as a whole with respect to their stair climbing results. Of the 3 Group B subjects who performed the stair-climbing test, the patient with
the worst outcome (death) was consistently above the sample average in terms of all parameters outlined above. The other 2 subjects were consistently below. We performed univariate regression analysis with the entire sample and with Group A to see if any of the stair-climbing values (rate of climb, VO₂ or workload achieved) might predict LOSOH. It should be noted that these 3 values are inter-related. We analyzed them separately in order to determine which was the most predictive. Some authors have found the total number of stairs climbed (and the rate of the climb) to be predictive of postoperative complications (115), while others have not (14). It was of interest to determine which value had the best predictive ability. However, we recognize the inter-relatedness of the measures.

Correlation of the various stair-climbing measures with LOSOH was evaluated. The results appear in Table 4.13. The entire sample was evaluated first, and then Group A was separated out. For the sample as a whole, a prolonged LOSOH was positively (but weakly) associated with faster rate of climb (steps/min.) (Pearson’s r = 0.32, p=0.02). A prolonged LOSOH also showed a modest correlation with a higher level of work achieved during the test (r = 0.34, p=0.02). A weaker association was found with VO₂ estimated from stair climbing (r = 0.29, p=0.05).

In addition to the association of stair climbing measures with LOSOH, the association of stair climbing with other preoperative measures was evaluated. Table 4.13 contains these correlations as well. DLco showed a moderate positive correlation with work expended during stair climbing (in W) (r = 0.48, p=.006). The association of preoperative DLco and rate of stair climb was of similar strength (r =0.45, p=0.009). The association of DLco with total number of stairs climbed did not reach statistical significance. There was no association found between preoperative measures of DLco and estimated VO₂.

The other PFT value of interest in this study was FEV₁. It was weakly correlated with the amount of work performed during stair climbing (r = 0.35, p=0.04) and with VO₂ achieved (r = 0.33, p=0.06). FEV₁ was positively associated with rate of stair climb (r =
0.38, p=0.02), although the relationship was not strong.

The 6-minute distance demonstrated a strong association with all stair-climbing measures. Pearson’s correlations were as follows: rate of stair climb (r = 0.72, p<0.001), workload reached (r = 0.57, p<0.001), and estimated VO₂ (r = 0.63, p<0.001). A weaker correlation with total number of stairs climbed was seen (r = 0.41, p<0.001).

Table 4.13 Pearson Correlations of Preoperative Stair Climbing Performance with Operative Outcome (LOSOH) and Other Preoperative Measures

<table>
<thead>
<tr>
<th>Stair climb measurement</th>
<th>Comparative measure</th>
<th>Entire Sample</th>
<th>Group A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p</td>
<td>r</td>
</tr>
<tr>
<td>Rate of climb (step/min.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOSOH</td>
<td>0.34</td>
<td>0.02</td>
<td>0.42</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.38</td>
<td>0.02</td>
<td>0.31</td>
</tr>
<tr>
<td>DLco</td>
<td>0.47</td>
<td>0.007</td>
<td>0.46</td>
</tr>
<tr>
<td>6-minute walk</td>
<td>0.72</td>
<td>&lt;0.001</td>
<td>0.69</td>
</tr>
<tr>
<td>Stair Work (W)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOSOH</td>
<td>0.38</td>
<td>0.01</td>
<td>0.51</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.38</td>
<td>0.02</td>
<td>0.32</td>
</tr>
<tr>
<td>DLco</td>
<td>0.48</td>
<td>0.006</td>
<td>0.48</td>
</tr>
<tr>
<td>6-minute walk</td>
<td>0.59</td>
<td>&lt;0.001</td>
<td>0.56</td>
</tr>
<tr>
<td>BMI</td>
<td>0.65</td>
<td>&lt;0.001</td>
<td>0.68</td>
</tr>
<tr>
<td>Estimated Stair VO₂ (mL/kg/min)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOSOH</td>
<td>0.37</td>
<td>0.01</td>
<td>0.46</td>
</tr>
<tr>
<td>FEV&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.37</td>
<td>0.03</td>
<td>0.29</td>
</tr>
<tr>
<td>DLco</td>
<td>0.45</td>
<td>0.01</td>
<td>0.44</td>
</tr>
<tr>
<td>6-minute walk</td>
<td>0.73</td>
<td>&lt;0.001</td>
<td>0.70</td>
</tr>
<tr>
<td>BMI</td>
<td>-----</td>
<td>NS</td>
<td>-----</td>
</tr>
</tbody>
</table>

Following this secondary analysis of the stair-climbing data and the positive univariate correlations seen in the entire sample as well as Group A, the multiple linear regression analysis was revisited. Rate, workload achieved and VO₂ were entered into the regression equation (as single additional measures, in an effort to avoid multicollinearity). For the entire sample (n=70), the 1 predictive stair climbing measure was workload achieved, with an R² of 0.130 and an adjusted R² of 0.112 (p=0.009). In the analysis of Group A (n=66) QL remained a strongly predictive variable, but the addition of stair work...
to QL accounted for over 36 per cent of the variance in LOSOH (p<0.001). The adjusted $R^2$ reached 0.334. These results appear in Table 4.14 below.

Table 4.14: Multiple Regression Using Stair Climbing Workload Achieved as a Predictor Variable (Entire Sample (a) and Group A (b)):

(a) Entire Sample:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Mult. R = 0.360</th>
<th>R² = 0.130</th>
<th>Adj. R² = 0.112</th>
<th>No. obs. = 51</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>209.103</td>
<td>1</td>
<td>209.103</td>
<td>7.290</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>1405.525</td>
<td>49</td>
<td>28.684</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>15.800</td>
<td>2.142</td>
<td>7.378</td>
<td>0.000</td>
</tr>
<tr>
<td>Stair work</td>
<td>0.039</td>
<td>0.015</td>
<td>2.700</td>
<td>0.009</td>
</tr>
</tbody>
</table>

(b) Group A:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>Mult. R = 0.602</th>
<th>R² = 0.362</th>
<th>Adj. R² = 0.334</th>
<th>No. obs. = 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>137.494</td>
<td>2</td>
<td>68.747</td>
<td>12.767</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>242.318</td>
<td>45</td>
<td>5.385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>16.579</td>
<td>1.207</td>
<td>13.733</td>
<td>0.000</td>
</tr>
<tr>
<td>QL Score</td>
<td>0.048</td>
<td>0.015</td>
<td>3.256</td>
<td>0.002</td>
</tr>
<tr>
<td>Stair work</td>
<td>0.019</td>
<td>0.007</td>
<td>2.865</td>
<td>0.006</td>
</tr>
</tbody>
</table>

ABGs, BMI and Smoking History:

Only 38 subjects had ABGs results available via chart review. 3 subjects had an elevated PaCO₂ (≥ 48 mmHg) documented. No association was found between preoperative PaO₂ or PaCO₂ and LOSOH postoperatively. This relationship held true for
both the sample as a whole and for Groups A and B. Examining the 4 Group B subjects, it was evident that there was no association between worsening ABG profile and the subsequent development of an adverse outcome. In fact, the subject who experienced the worst outcome (death) had ABGs well within normal limits. Moreover, in the whole sample, the subject with the most severe CO₂ retention (PaCO₂ = 52 mmHg) preoperatively experienced one of the best operative outcomes, with a LOSOH of 25 days.

Preoperative measures of BMI (kg/m²) were documented, using the height and weight data obtained from the pulmonary function test measures. BMI is a global measure of nutritional status. Obesity is characterized by a high value for BMI, and cachexia by low values. Extremes of either kind can affect respiratory status as well as functional mobility, and so could presumably affect operative outcome. BMI was found to follow a normal distribution for the study sample. The scatterplots depicted in Figure 4.14 (Views a and h, please refer to following page) demonstrate that there was no obvious association between LOSOH and BMI within the sample as a whole. Pearson correlation was not significant for the entire sample. In Group A, a weak positive correlation was found between LOSOH and BMI: r = 0.32, p=0.02. Examination of the BMI values for the subjects in Group B demonstrated no obvious abnormalities in these patients' nutritional statuses.

Previous smoking history was documented (measured in pack-years). The mean (± SD) smoking history was 35 (± 27) pack-years. Smoking history was not associated with prolonged in-hospital LOS either for the entire sample or in Group A. The 4 Group B subjects were not characterized by unusually heavy smoking histories. In fact, 2 of these subjects reported no smoking history, while the other 2 reported 20 and 27 pack-years of smoking respectively.
LOSOH vs. BMI (Sample)

LOSOH vs. BMI (Group A)

Figures 4.14 Scatterplots of LOSOH vs. BMI
Additional Scales of the Health Status Instrument:

Health status scales (both functional scales and symptom scales) other than the Global QL, EF and SF scales were administered with the EORTC instrument. Table 4.15 lists the statistically significant correlations found between these additional HRQL measures and LOSOH. Of interest are the significant associations found between LOSOH and the various functional scale scores in Group A. These correlations were masked in the entire sample secondary to the excessive influence of the 4 Group B subjects with respect to LOSOH. Note that the various symptom scales have negative relationships to LOSOH since higher scores on the symptom scales represent greater impairment. This is the opposite of the scores on the functional scales, where higher scores signal better function (149). The subjects in Group B generally had somewhat higher symptomatology scores on the preoperative HRQL scales than did the subjects in Group A (but these values were not statistically significant). 3 of 4 subjects in Group B reported sleep disturbance. All 4 subjects reported significant difficulty with coughing as measured on the lung module. 2 of these patients also reported hemoptysis. Table 4.15 indicates that reporting of hemoptysis was the one symptom scale that correlated significantly with operative outcome in the sample as a whole. No such correlation characterized Group A.

In addition, we performed simple correlation on PF score and 6-minute walk distance \((r = 0.55, p<0.001)\), confirming that the subjective rating correlates with the objective test. A scatterplot of PF score and 6-minute walk distance appears in Figure 4.15 (on the following page).
Scatterplot of Walk Distance vs. PF Score (Sample)

Figures 4.15 Scatterplot of Walk Distance vs. PF Score
Table 4.15  Correlations of LOSOH with Additional Preoperative Functional Scale Measurements taken with the QLQ-C30 and the QLQ-LC13 Modular Supplement

<table>
<thead>
<tr>
<th>Functional Scale</th>
<th>Entire Sample</th>
<th>Group A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core Questionnaire (QLQ-C30)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Function (PF)</td>
<td>0.12</td>
<td>0.39</td>
</tr>
<tr>
<td>Role Function (RF)</td>
<td>0.16</td>
<td>0.48</td>
</tr>
<tr>
<td>Fatigue (FA)</td>
<td>-0.17</td>
<td>-0.54</td>
</tr>
<tr>
<td>Nausea/vomiting (NV)</td>
<td>-0.22</td>
<td>-0.46</td>
</tr>
<tr>
<td>Pain (PA)</td>
<td>-0.07</td>
<td>-0.44</td>
</tr>
<tr>
<td>Sleep Quality (SL)</td>
<td>-0.15</td>
<td>-0.45</td>
</tr>
<tr>
<td>Appetite (AP)</td>
<td>0.18</td>
<td>-0.46</td>
</tr>
<tr>
<td>Constipation (CO)</td>
<td>-0.11</td>
<td>-0.46</td>
</tr>
<tr>
<td><strong>Lung Module (QLQ-LC13)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyspnea</td>
<td>-0.04</td>
<td>-0.25</td>
</tr>
<tr>
<td>Hemoptysis</td>
<td>-0.25</td>
<td>-0.19</td>
</tr>
<tr>
<td>Sore mouth</td>
<td>-0.17</td>
<td>-0.36</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>-0.14</td>
<td>-0.32</td>
</tr>
<tr>
<td>Neuropathy</td>
<td>-0.06</td>
<td>-0.38</td>
</tr>
</tbody>
</table>

*Tumour Status, Tumour Type, and Neoadjuvant Regimens:*

55 subjects in our series were diagnosed with primary lung cancer. 2 subjects were discovered to have benign lesions post-resection, but their lesions were suspicious for cancer preoperatively. 2 subjects were diagnosed with mesothelioma. The remaining 11 subjects had lung metastases (the majority with metastatic osteosarcoma): 2 of these subjects had isolated metastases from colon cancer. For those patients with primary lung cancer, we analyzed the association between LOSOH and stage of disease. These results
appear in Figure 4.16. Those with later stages of the disease showed slightly poorer outcomes, but these differences did not reach statistical significance. The majority of patients had Stage I or Stage II disease. All the Group B subjects had either Stage I or Stage II disease. Patients with Stage III disease demonstrated the greatest variability in LOSOH in Group A.

2 subjects in this study were diagnosed with mesothelioma preoperatively and both underwent extrapleural pneumonectomy. In this procedure, extensive resection is required. Both patients experienced slightly prolonged hospital stays (LOSOH 18 and 19 days respectively). While one of these patients was initially discharged from hospital at postoperative day 8, he was readmitted to hospital on postoperative day 26 with a suspected wound infection and rapid onset of respiratory failure. It quickly became apparent that the patient had suffered a local recurrence of rapid onset. The subject deteriorated rapidly and died on postoperative day 40. While this death was outside the 30 day perioperative period, this rapid succession of events is worth noting. I do not believe that it should be considered a perioperative complication, but rather, reflects the remarkably aggressive nature of this disease. In future investigations, it might be prudent to analyze cases of mesothelioma separately.

Because of these 2 subjects' relatively poor outcomes, an examination of their preoperative scores was warranted. Both patients performed preoperative exercise testing, both achieving more than 400 m on their 6-minute walk tests. Their stair-climbing scores were within normal limits. In terms of the health status measures, these subjects were characterized by poor global QL scores (8 and 42 respectively, with 8 being the second lowest score in the sample). 1 subject reported poor social function as well. Both subjects reported normal emotional function. Both subjects reported significant sleep disturbance, as well as greater than average symptomatology in terms of chest and arm pain, cough, and dyspnea. However, their symptom scores were no higher than those given by a number of patients suffering from lung cancer or other tumour types. One subject had an abnormal
Boxplot of LOSOH vs. Lung Cancer Stage (Sample)

Boxplot of LOSOH vs. Lung Cancer Stage (Group A)

Figures 4.16  LOSOH vs. Lung Cancer Stage
LOSOH (days) based on Administration of Neoadjuvant Therapy

Boxplot of Global QL by Administration of Neoadjuvant Therapy

Figures 4.17 and 4.18
FEV$_1$ score (67% of predicted). The DLco and BMI scores were within normal limits for both the subjects with mesothelioma.

Finally, administration of neoadjuvant therapy (preoperative chemotherapy and radiation therapy) was also evaluated. 7 patients received neoadjuvant therapy. These patients did not differ from the entire sample in terms of postoperative LOSOH. The results appear in Figure 4.17. None of the 4 patients in Group B received neoadjuvant therapy, so it does not appear to be a factor predisposing patients to extreme complications. By way of comparison, we examined the Global QL score based on whether patients received neoadjuvant therapy or not, and the patients who received neoadjuvant therapy did not rate their health status any lower than the patients who did not receive such therapy. Those who did not receive neoadjuvant treatments demonstrated much greater variability in terms of their HRQL than did those who had received this therapy. The results appear in Figure 4.18.

**Operative Factors Evaluated:**

A number of operative factors were evaluated to see if these demonstrated an association with postoperative outcome. Postoperative pain control was managed with epidural analgesia (morphine). Only two subjects received a different pain control modality from the standard approach: 1 subject received intercostal nerve blocks, the other intravenous patient controlled analgesia. Neither of these subjects experienced a different outcome from the remaining sample, nor did they belong to Group B.

Anaesthesia time was also tracked in our series. These data were collected from the operative notes. The mean (+ SD) anaesthesia time for our sample was 4.5 (+ 1.9) hours. Graphical depictions of these associations appear in Figure 4.19. The subjects in Group B were not characterized by prolonged administration of general anaesthesia. 2 of the subjects in Group A did have greatly prolonged anaesthesia times (14 and 10 hours...
Figures 4.19 Scatterplots of LOSOH vs. Anaesthesia Time
Boxplot of LOSOH vs. Amount of Lung Tissue Resected

Legend:
1 = <1 lobe
2 = lobectomy
3 = >1 lobe, <2 lobes
4 = bilobectomy
5 = pneumonectomy

Boxplot of LOSOH vs. Amount of Lung Tissue Resected (Group A)

Figures 4.20 LOSOH vs. Amount of Lung Tissue Resected
Boxplot of LOSOH vs. Extent of Resection (Sample)

1 = simple
2 = extended

Boxplot of LOSOH vs. Extent of Resection (Group A)

Figures 4.21 LOSOH vs. Extent of Resection (Simple or Extended)
and both required readmission within the first 3 to 4 postoperative weeks: one was readmitted to a peripheral facility with respiratory failure and underwent a protracted course of respiratory rehabilitation (still oxygen dependent at 4 months postoperatively; the patient was initially discharged from hospital on room air). The second subject was readmitted to our facility the third week following discharge home, with progressive weight loss, anorexia, and abdominal symptoms. His complications included pyloric stricture secondary to high doses of narcotic analgesics, as well as local recurrence. He eventually died 4 months postoperatively, having spent virtually no further time at home (beyond a few days). A few other subjects who did not have prolonged anaesthesia times also required readmission to hospital for either postoperative complications or aggressive local recurrence.

We also measured the association between amount of lung tissue resected and the extent of surgical resection (simple thoracotomy vs. extended procedures). These results appear in Figures 4.20 and 4.21. None of the 4 subjects in Group B received pneumonectomy. However, patients having bilobectomies experienced longer mean postoperative recoveries within the entire sample. 2 of the Group B subjects had extended procedures, and 2 did not.

Requirements for ICU admission were also monitored. 9 subjects required admission to the ICU postoperatively (within the first 24 hours). 2 additional subjects were transferred to the ICU preoperatively and postoperatively as part of a high-risk surgical study. The mean ICU admission lasted 6.7 (± 3.3) days for these patients. 2 of the 4 Group B patients were admitted to ICU postoperatively, these 2 having a protracted ICU LOS. The remaining patients were in ICU for 1 to 3 days postoperatively. The other 2 subjects in Group B were never admitted to ICU. Of interest, those subjects requiring ICU admission scored below average on most of the variables of interest. The one
exception was the EF score on the QLQ-C30 instrument. Its considerable ceiling effect was maintained. The mean values of all primary variables are listed in Table 4.16.

Table 4.16 Profile of Subjects requiring ICU admission

<table>
<thead>
<tr>
<th>Group</th>
<th>LOSOH (days)</th>
<th>EF Score</th>
<th>SF Score</th>
<th>QL Score</th>
<th>FEV1 (% pred)</th>
<th>DLco (% pred)</th>
<th>Walk dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICU (n=9)</td>
<td>14.7</td>
<td>84</td>
<td>70</td>
<td>63</td>
<td>66</td>
<td>76</td>
<td>289.67</td>
</tr>
</tbody>
</table>

Of particular note are the lower scores for total LOSOH, FEV₁, and DLco for those admitted to ICU postoperatively. Reasons for admission to the ICU differed. 2 patients required overnight mechanical ventilation following prolonged anaesthesia (OR time > 8 hours). 3 were transferred to the ICU for monitoring following very extensive procedures. 1 was transferred secondary to intraoperative complications. 2 others secondary to severe preoperative coronary artery disease, and one for increased risk of respiratory complications secondary to concomitant COPD. Surgeons are able to book their patients for elective ICU admission preoperatively, in anticipation of a difficult postoperative course, based on various preoperative measures (preop. ABGs, PFTS, past medical history, or clinical impression). 4 patients received such anticipatory transfers to ICU, only to be discharged from ICU the following day (ICU adm. less than 24 hrs.). Thus the association between impaired pulmonary function and ICU admission may represent surgeon choice based on anticipated risk.

Postoperative Recovery:

The results of the postoperative QL scores and the follow-up 6-minute walk distances appear graphically in Figures 4.22 through 4.27. Most subjects returned to their preoperative level of function on all measures by the 3 month evaluation.

The necessity for readmission to hospital (either our facility or others) was tracked. Any readmission to hospital within the first 30 days was added to the original hospital LOS.
to produce the LOSOH which was the dependent variable of the study. The 4 subjects who experienced extreme operative outcomes (those in Group B) have been discussed at some length already.

8 patients required readmission to hospital following initial postoperative stay within the first 30 days. The reasons for readmission appear in Table 4.17.

<table>
<thead>
<tr>
<th>Reason for Readmission within the First 30 days</th>
<th>No. patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilatory Insufficiency</td>
<td>3</td>
</tr>
<tr>
<td>Pulmonary Embolus (3 weeks postop)</td>
<td>1</td>
</tr>
<tr>
<td>Infection in pneumonectomy space</td>
<td>1</td>
</tr>
<tr>
<td>Pyloric stricture and weight loss</td>
<td>1</td>
</tr>
<tr>
<td>Aggressive local recurrence</td>
<td>1</td>
</tr>
<tr>
<td>Cardiac Arrhythmia (3 weeks postop)</td>
<td>1</td>
</tr>
</tbody>
</table>

The average length of readmission was 21 days, ranging from 2 days to 78 days. One additional patient was readmitted to hospital at 4 weeks postoperatively to undergo a second thoracotomy and lung resection for metastatic osteosarcoma. He had significant bilateral disease which required serial thoracotomies. He was tracked following the first thoracotomy. He was lost to follow-up following the second thoracotomy. Of interest, on both admissions he remained in hospital 5 days and 4 days respectively, experiencing no postoperative complications. Nor did he require further readmissions to hospital.

6 patients died following the initial 30 day perioperative period. 3 of these subjects were those who required readmission to our facility and experienced protracted hospital stays prior to death. Another 2 had recurrent disease and expired approximately 6 months postoperatively, but did not require readmission to our facility. They received palliative care at other facilities. The sixth death was unexpected, with the patient failing to attend his first outpatient follow-up appointment because of feeling unwell. The following week, we were informed of this patient's death. Of note, 3
Recovery Profile EF Scores over 3 Months

Recovery Profile of SF Scores over 3 Months

Figures 4.22 and 4.23
The graphs depicted in these pages of recovery during the first 3 months postoperatively are based on 4 measurement periods (pre-op, at discharge from hospital, at 1 month and at 3 months postoperatively). The overall changes for the relative scores for each graph appear below:

**E.F.**
- 3% change in score between pre-op and discharge
  - discharge and 1 month: 1% (p=NS)
  - 1 month and 3 months: <1% (p=NS)

**S.F.**
- 19% change in score between pre-op and discharge
  - discharge and 1 month: 1% (p=NS)
  - 1 month and 3 months: <1% (p=NS)

**Global QL:**
- 24% change in score between pre-op and discharge
  - discharge and 1 month: 12% (p=0.002)
  - 1 month and 3 months: 13% (p=NS)

**R.F.**
- 43% change in score between pre-op and discharge
  - discharge and 1 month: 15% (p=0.02)
  - 1 month and 3 months: 11% (p=NS)

**P.F.**
- 28% change in score between pre-op and discharge
  - discharge and 1 month: 9% (p=NS)
  - 1 month and 3 months: 10% (p=NS)

**Walk:**
- change in walk score between pre-op and discharge
  - discharge and 1 month: 183.44 m, p<0.001
  - 1 month and 3 months: 127.22 m, p=0.001

Recovery Profile of QL Scores over 3 Months

![Graph showing quality of life scores over time]

Figures 4.24
Recovery Profile of 6-minute walk Distance over 3 months

Test Time (1 = pre; 4 = 3 mo.)
Profile of Recovery of PF Scores over 3 months

Recovery Profile of RF Scores over 3 Months

Figures 4.26 and 4.27
subjects who had late deaths had a LOSOH > 20 days, while three had a LOSOH < 20 days.

**Summary:**

In summary, the dependent variable in this study had a bimodal distribution, necessitating analysis of the sample as a whole, as well as of Groups A and B. Examination of the subjects' LOSOH with Group A indicates that this measure does represent a progression of complication rates. Using multiple regression analysis, the only predictors of postoperative LOSOH in the *entire sample* were age and workload achieved during stair climbing. Univariate regression of the independent variables indicated that mean walk distance was also associated with LOSOH. The extreme outcomes suffered by the Group B patients are difficult to predict, whether by traditional physiologic measures, or by functional or health status ones. In those experiencing a more standard postoperative course (Group A), however, global QL rating, 6-minute walk distance, and (in a secondary analysis) workload performed during stair climbing, are the best predictors of outcome. DLco shows a trend toward predictive ability in the primary analysis of the Group A patients.

We evaluated numerous pre- and postoperative outcome measures in an effort to describe our sample further. Various subscales measured with the health status instrument preoperatively were associated with operative outcome. The administration of neoadjuvant therapy was not associated with reduced LOSOH. The 6-minute walk test results indicate that the first and second trials are highly associated with one another, but an examination of the ICC indicates that reliability was relatively poor in testing with our sample subjects. Healthy normals exhibited highly reliable results.
Postoperatively, a number of patients were readmitted to hospital. 6 patients died during or after long-term follow-up. However, most patients have returned to their preoperative level of function/health status within the first 3 postoperative months.
CHAPTER 5
DISCUSSION

This was a prospective cohort study evaluating the relative utility of 6 preoperative measures in the prediction of LOSOH following lung resection via thoracotomy. This surgical population is at significant risk for developing a wide array of postsurgical complications. A review of the literature reveals that the preoperative measures traditionally employed to evaluate risk (PFTs, ABGs, etc.) are not adequate for predicting who will tolerate this procedure and who will not. We hypothesized that a multidimensional health status measure would significantly enhance our ability to predict prolonged postoperative LOS, when added to the measures of physiologic impairment currently relied upon (namely, FEV₁ and DLco), and a measure of functional disability (the 6-minute walk test). A secondary goal of the study was to describe the postoperative recovery of health status and functional exercise ability in the first 3 postoperative months.

Because of the difficulty inherent in defining postoperative complications (91), coupled with our desire to employ a single outcome measure incorporating a number of different complications, we adopted LOSOH within the first 30 days as our dependent variable. While death is a widely recognized major complication (albeit a relatively rare one), non-fatal complications prove particularly difficult to define. There appear to be almost as many definitions of operative morbidity as there are researchers. Many studies evaluating operative complication rates use a host of different outcome measures to represent complications. Or they include all complications suffered and analyze separately for each complication (or class of complication: pulmonary, cardiovascular, and so forth). Or they look at the overall rate of morbidity. Clinical researchers require outcome measures which encompass a wide array of potential complications that have easily definable endpoints.
Some studies have employed postoperative LOS as a measure of operative outcome, either on its own or in combination with other measures (14,51,95,103,115,117,170-171). It has been shown to be correlated with postoperative complications in these studies. This entity is relatively easy to measure and define (and is widely recognized), but it has some important weaknesses as a measure representative of complication. It fails to account for early operative mortality (a patient dying on postoperative day 2 would appear to have had an excellent outcome if LOS alone was employed) or readmission to hospital. Moreover, it can be argued that LOS may more accurately reflect idiosyncrasies of a surgeon’s preference/practice — or a particular hospital’s policy — rather than the patient’s state of health (these limitations will be considered in more detail in a later section of this chapter).

Given the shortcomings of LOS, we elected to use what is in essence a reciprocal of this measure, namely LOSOH within the first 30 days. This measure has the advantage of encompassing morbidity and mortality within a single score. Thus patients who die early on in their operative course receive a score of 0. A patient who has a protracted recovery and is not discharged until the 26th postoperative day would receive a score of 4. We were cognizant that the same criticism of LOS (namely that it might represent non-medical factors governing postoperative discharge) might be leveled at LOSOH. Therefore we examined our LOSOH data to determine the degree to which it reflected complication rate, and the degree to which it was measuring some other factor. Our analysis indicates that LOSOH was indeed highly correlated with relative complication rate and relative severity of postoperative complications. As such, it may be considered representative of complications within the sample (with higher incidences of complications for shorter LOSOH, and a general increase in the severity of complications seen with decreasing LOSOH). This indicates that LOSOH has construct validity.

It could be argued that we should have employed the measure LOS and simply had a maximum (worst possible) score of 30, and a best possible score of 0. Operative deaths
could automatically be assigned a score of 30. Readmissions could have been tracked and LOS recalculated accordingly. LOS is certainly a measure very familiar to surgical investigators, and this adapted approach seems sensible. However, this is not the way in which LOS has traditionally been used in the literature. The use of different terminology distinguishes LOSOH from the more traditional usage of LOS.

LOSOH was the dependent variable of interest in our study. Unfortunately, this variable had a bimodal distribution (not a Gaussian one), making the planned parametric statistical analysis (multiple linear regression) difficult. We found that extreme complications (such as those suffered by the Group B subjects) were not predicted by any of the 6 independent variables employed. However, the addition of the variable age into the regression equation revealed it to be the predictor of LOSOH for the sample as a whole. Although all patients who developed extreme complications were over the age of 73, it should be noted that there were some patients (n = 9) in Group A who were over the age of 73 who did well postoperatively (LOSOH \( \geq 23 \) days). This significant association with age disappeared once the 4 extreme outliers (Group B) were excluded from the sample. Nevertheless, this is the only factor which seems to distinguish the Group B patients from those in Group A. As such, age does seem to pose some increased risk for patients undergoing lung resection. However, we would caution clinicians that age alone is not sufficient reason for denying an elderly patient a potentially curative resection. Our results demonstrate that elderly patients can and do enjoy successful operative outcomes.

In those subjects who suffered less extreme outcomes (Group A), certain measures included in the primary analysis were positively associated with better operative outcome (prolonged LOSOH): the global QL scale of the QLQ-C30 questionnaire, as well as the the 6-minute walk test showed some limited predictive ability. DLco showed a trend toward significant association with LOSOH. Of interest, the variable considered by many (1.9.64.81.102.128) to be the 'gold standard' (FEV\(_1\)) did not show significant predictive ability (although this could have been the result of Type II error) (54). We hypothesized
that preoperative health status measures would enhance our ability to predict operative outcome. In fact, the global QL scale employed was the single best predictor of postoperative complications in the Group A patients. The 6-minute walk distance was also a significant predictor of postoperative LOSOH.

Various post hoc tests were used to evaluate the predictive ability of additional preoperative measures of interest (those not included in the primary analysis). Such post hoc tests are valuable for exploring the data further and for hypothesis generation for future investigations, but they cannot be employed to prove or disprove our hypothesis. Stair climbing proved to be significantly predictive of postoperative LOSOH in both the sample as a whole and in the Group A subjects. In combination with the global QL scale in the Group A multiple regression analysis, these 2 variables accounted for over 36% of the variance in LOSOH. Other preoperative measures (such as ABGs, BMI, tumour stage, and smoking history) were not associated with poorer operative results. The administration of neoadjuvant therapy was not associated with operative outcome, although this could be the result of type II error (54.158).

Operative factors such as duration of anaesthesia, extent of resection, and amount of lung resected were also evaluated in this study. Extent of resection (simple thoracotomy versus extended procedures) had a minor effect on LOSOH, with extended resections being characterized by longer average postoperative admissions than simple resections. However, the 4 patients who suffered the worst outcomes were evenly divided as to extent of resection. The amount of lung tissue resected had little effect on operative outcome. Patients receiving pneumonectomy in our study did not have an increased incidence of complications, contradicting the results of some authors (1.3.81-82, 93). Our results are in accord with a recent prospective analysis by Deslauriers et al. (1994), who found no increased incidence of complications among patients receiving pneumonectomy (86).

This study also examined postoperative recovery following lung resection, with particular emphasis on postoperative health status and functional exercise recovery.
Overall, most patients had regained their preoperative health status and exercise ability by the third month postoperatively.

A detailed consideration of these results appears below.

**Primary Analysis:**

Multiple linear regression was employed to evaluate the following predictor variables: PFTs (FEV₁ and DLco), functional exercise testing (6-minute walk distance), and 3 of the health status scales of the EORTC QLQ-C30 questionnaire (SF, EF, and global QL scales). Postoperative complications (the dependent variable) were represented by the surrogate measure, LOSOH (in days), which was the dependent variable of interest.

The first obstacle encountered regarding this analysis was the skewed (essentially bimodal) distribution of the dependent variable. Nevertheless, we attempted to perform multiple regression on the entire sample at the outset. Exploration of other options, such as stratifying the sample on a variety of factors (requirements for prolonged mechanical ventilation postoperatively, operative type, etc.), or employing non-parametric forms of regression, or logistic regression yielded virtually the same distribution as in the original analysis. Therefore we adopted the approach of first analyzing the sample as a whole, then as 2 separate groups, based on this bimodal distribution. The 2 separate groups were: Group A (those who had a less extreme outcome, and whose distribution was essentially Gaussian) and Group B (those who experienced extreme -- even catastrophic -- complications). These groups were then analyzed separately.

In the multiple regression analysis, using the sample as a whole, none of the 6 primary independent variables was predictive of LOSOH. Diagnostic testing of this regression analysis confirmed that the 4 Group B subjects were indeed outliers that were exercising undue influence. When age was added to the regression equation, it became the sole predictor of LOSOH. This is a significant finding in our study and is in keeping with
a number of other investigators in the field. A consideration of the Group A and B analyses appears below.

Because Group B consisted of only 4 subjects, statistical analysis was not possible. and so a descriptive approach was adopted. Age was the only factor identified which seemed to predispose these patients to the dismal outcomes they sustained. The fact that 9 equally elderly Group A patients experienced satisfactory outcomes has already been mentioned above. The Group B subjects were no different from Group A in terms of their FEV₁, their health status scores, or their 6-minute walk distances. At best, we can state that increased age is associated with an increased risk of postoperative complications following lung resection via thoracotomy. but that it should not be used in isolation to prohibit elderly candidates from receiving potentially curative procedures. Of interest is the fact that controversy exists in the literature as to whether increased age poses increased risk for complications: some authors report that it does (1.81.86), while others report that it does not (82.93). Given the paucity of measures predicting LOSOH. we can only conclude (like many previous researchers) that major, life-threatening complications remain extraordinarily difficult to predict (170). Specifically, intraoperative complications or the presence of ‘silent’ underlying disease (e.g. silent coronary artery disease) may be unrelated to subjective and objective measures of health status and daily function. as employed in the present study.

Multiple regression analysis with the Group A data revealed 2 variables predictive of LOSOH: global QL score and mean 6-minute walk distance. Global QL was the most predictive measure. with walk distance adding significantly to the predictive ability of the equation. 2 additional variables showed trends toward significant predictive ability: SF score and DLco. Because of marked collinearity between the global QL score and the SF scores on the health status instrument. inclusion of one diminished the subsequent predictive ability of the other. No added benefit accrued from adding SF score into the regression equation. As global QL score is thought to encompass multiple dimensions of
HRQL, it makes intuitive sense that it is the most appropriate predictive measure to include in the analysis. Our finding that DLco showed a trend towards significance is supported by the literature (9,12-13,16,39,60,64,95,104). Of surprise was its improved association with LOSOH in the multiple regression equation, given the poor correlation found between LOSOH and DLco in the univariate regression analysis (both for Group A and for the sample taken as a whole). Diagnostics run on the regression equation did not indicate significant collinearity or skewing of the distribution of the DLco variable, so it is difficult to offer an explanation of its new predictive capacity seen in the multiple regression analysis of Group A. Of note, DLco was quite strongly associated with stair climbing measures (work and VO₂), both in the whole sample and in Group A. Some authors have noted that DLco is highly predictive of complications after lung resection. DLco evaluates the gas exchange function of the lung and may be predictive of ability to exercise (as well as predicting the patient’s response to other forms of physiologic stress) (13.71). In this study, the correlation matrix of the regression coefficients revealed that DLco was also moderately correlated with 6-minute walk distance (r = 0.37).

Of particular concern is the large degree of variance in LOSOH unaccounted for in the multiple regression analysis of Group A: almost 80% of the variance in LOSOH remains unexplained by our model. There are additional unidentified factors that are producing the majority of the variance seen in LOSOH. It is possible that these are factors relating to underlying physiological impairment (for instance, elevated pulmonary vascular resistance) (126). Our patients do not routinely receive preoperative ventilation-perfusion scans, but some authors argue that this test is predictive in high-risk candidates (13). If patients lose highly functional lung tissue during the resection, perhaps this would account for prolonged or unsatisfactory recovery from surgery. Perhaps silent myocardial dysfunction could be detected by preoperative echocardiography or cardiac catheterizations. Besides preoperative measures of physiologic impairment, perhaps measures of physical function (usual activity level, regularity of exercise, etc.) would be predictive of operative
outcome. Alternatively, the most salient psychosocial determinants of operative outcome have not been evaluated in this study. Factors such as locus of control, previous experience with cancer diagnoses (either their own or those of family members), preoperative expectations regarding functional status and quality of life postoperatively, or even confidence in the attending surgeon, might influence operative outcome. One of these examples will be considered briefly. With respect to patients' expectations of post-treatment quality of life, Haut et al. (1991) evaluated cancer patients' expectations of nausea and vomiting prior to the administration of chemotherapy: they found that pretreatment expectations were significantly predictive of symptoms post-treatment (125). Perhaps a similar trial could be undertaken with the lung resection population.

Besides considering the independent variables and their limited ability to predict LOSOH, we must consider some of the inherent limitations of this dependent variable. LOSOH was chosen because of its ability to encompass operative mortality and morbidity within a single measure. However, there are additional factors that might influence postoperative LOSOH which should be recognized. The most obvious are pressure to vacate beds (as imposed by the hospital administration) and discharge date as related to day of the week (i.e. discharge pattern around the weekend). During the course of the study, an outbreak of an antibiotic resistant gastrointestinal infection necessitated confinement of patients within their rooms and a push for early discharge of all patients not colonized by this bacterium (Vancomycin resistant enterococcus). 2 subjects were affected by this situation, with LOSOH 21 to 23 days (well within normal range). These subjects do not appear to have been forced out of hospital sooner than normal, based on retrospective chart review. They were stable and their wounds were well-healed. Neither subject was colonized with this particular bacterium. Analysis of the data collection sheets was used to determine if a disproportionate number of patients were discharged prior to the weekend. The results of this analysis appear graphically (for the entire sample) in Figure 5.1.
Histogram of Day of Discharge from Hospital (n=70)

Legend
1 = Mon.
2 = Tues.
3 = Wed.
4 = Thurs.
5 = Fri.
6 = Sat.
7 = Sun.

Figure 5.1: Discharge date by Day-of-week
There is a slightly increased incidences of discharges on Tuesdays and Sundays. The reasons for this pattern are unclear.

Individual surgeon preference for particular discharge patterns might also influence LOSOH. If one surgeon is more conservative in his/her practice than another, discrepancies in LOSOH might be noted. We examined our data for differences among the 3 surgeons, and found no differences in LOSOH based on surgeon. However, this must always be monitored if LOSOH is to be employed as an outcome measure in future studies.

Other factors that could influence the LOSOH construct are socioeconomic status, patient access to community resources, family supports, and so forth. If a patient is of a low socioeconomic status, he would be unable to hire privately funded home care services which might prevent him being readmitted to hospital. Access to private resources might also result in earlier discharge from hospital, as the patient could have more supports in the home (than can be offered by the publicly funded home care system). A supportive family situation might impact on discharge and readmission patterns in a similar way. Finally, if patients live in remote areas, with fewer community-based health services, this might also affect LOSOH. All these factors warrant further study. Unfortunately, we did not track this information formally in our study.

Our analysis of the Group A and B patients lends credence to the construct validity of the LOSOH measure. Those experiencing truncated LOSOH had more complications overall, with the worst LOSOH being correlated with the most severe complications. This is an exciting finding, but this construct will have to undergo further validation in future studies.

Our study is certainly not the first to conclude that it remains difficult to predict operative complications following thoracotomy and lung resection. Other authors have come to the same conclusion (81-83,86,126). Future research should investigate some of these unknown sources of variance.
The fact that we were only able to identify predictors in Group A (but not for the sample as a whole) has significant implications regarding our ability to predict postoperative LOSOHI following lung resection. If we believe that health status factors (in particular measures of psychosocial function) are important to operative outcome (in terms of recuperation, motivation to engage in postoperative rehabilitation, discharge home, and return to normal roles), it is logical that this would apply to those who are within the 'normal' range of operative recovery. If patients are depressed, if they feel that their quality of life is poor prior to surgery, this may translate into a reluctance to participate in their postoperative rehabilitation programs (including poor compliance with self-directed regimens of deep breathing and coughing exercises, delays in early ambulation, a reluctance to push themselves to improve their exercise tolerance, delayed weaning from supplemental oxygen, fears regarding discharge from hospital, and the like). Such treatments entail a significant degree of patient motivation and compliance. Thus preoperative QL score (and SF score) could correlate with less extreme outcomes in terms of LOSOHI, but would not necessarily be predictive of such severe outcomes as fatal myocardial infarction, life-threatening arrhythmias, major intraoperative complications, and so forth. An interesting area of investigation would be the cascade that might result from an initial mild atelectasis progressing through pneumonia and respiratory failure (progression from minor to major complications secondary to lack of motivation/compliance with postoperative treatment regimens). Another example of such a cascade of effects might be a refusal to ambulate/mobilize, followed by pulmonary embolus. Such outcomes warrant further study and would require large samples and a different study design.
Secondary Analyses:

Numerous secondary analyses were run on the data. As mentioned previously, such post hoc tests can only be used for hypothesis generation; no definitive conclusions can be drawn from these analyses. One goal of the study was to describe postoperative recovery from single lung resection using the health status instrument and the 6-minute walk test. Analysis of the data relating to postoperative recovery revealed that most subjects had returned to their preoperative levels of health status and exercise ability by the third postoperative month. Not unexpectedly, subjects reported their lowest health status scores during the first follow-up session (while still in hospital). Those subjective measures relating to physical ability to perform normal roles and common ADLs demonstrated the most extreme decline during postoperative recovery, with a return to preoperative levels of function at 3 months. The psychosocial function scales (emotional and social function scales of the QLQ-C30) showed the least change between the pre- and postoperative periods. Of interest is the minor improvement seen in the EF scores at 3 months following resection, when compared with preoperative scores. Most patients reported being very anxious prior to surgery. Those patients who reported being very distressed and anxious postoperatively however were more likely to decline completing the questionnaire (n = 3), so it is possible that those who completed the questionnaire postoperatively were experiencing less distress than those who did not (a potential source of bias).

In terms of objective postoperative measures (namely, the 6-minute walk test), subjects demonstrated their poorest functional exercise performance at the first follow-up session (immediately prior to discharge from hospital), with steady improvements occurring at 1 month and 3 months postoperatively. In general, subjects were still reporting relatively poor function at the 1 month interval. Dramatic improvements seem to occur within the second and third months of recovery. We lack sufficient data regarding those patients who underwent postoperative chemotherapy to draw any conclusions regarding its
impact on postoperative recovery in the first few months. This would be of interest in future studies.

The subjective global QL ratings showed moderate declines during the postoperative period, mid-way between the relatively minor changes reported in terms of psychosocial function, and the dramatic changes reported in terms of physical function. It is possible that this moderate recovery profile reflects the combined effects of physical dysfunction and 'normal' psychosocial function during the early postoperative phases. Since health status is a multidimensional construct, encompassing physical, social, and psychological functioning, the global QL scale may reflect the cumulative influence of all 3 dimensions (20-26,36,45-48,50,68,118,120-121).

Pearson product moment correlations between the subjective ratings of PF and RF, and the objective 6-minute walk measure at each time point (preoperatively and postoperatively at discharge, 1 month and 3 months) revealed moderate to strong associations between the subjective and objective measures. These associations suggest that the QLQ-C30 measure has concurrent validity. The health status instrument could be a useful measure for evaluating postoperative recovery in a clinical setting. It is sensitive to changes in physical functional status (changes in distance ambulated in 6-minutes), and the original validation studies by the EORTC also demonstrated that it was sensitive to changes in physical status as evaluated by the Eastern Cooperative Oncology Group (ECOG) performance status scale (24).

One of the difficulties experienced with postoperative follow-up in our study was ensuring that subjects were not missed secondary to scheduling changes in the clinic, missed appointments, or changes in follow-up policy. During the course of the study, the surgeons changed their follow-up policy, whereby many patients had their appointments scheduled at a different facility. The investigator was not informed of these changes immediately, so measurements on some subjects were missed because of this policy change. At the other facility, no corridor was as long as those in Mount Sinai Hospital
(namely, 33.0 m). The second facility's corridors were typically half the length or shorter. We were forced to test patients on a course 16.5 m long. This change was explained to subjects, with all other factors being kept constant (namely, standardized (absence of) encouragement, standardized instructions to the patient, the same two raters as preoperatively, and so forth). In all, 6 subjects underwent postoperative testing under these conditions. No differences (median for 6 subjects = 330.00 m, median for remainder of sample tested at original facility = 346.00 m. \( p = \text{N.S.} \)) were found in terms of these patients' postoperative scores relative to those followed at the original facility.

While subjects were typically eager to complete the preoperative measures, some patients reported that emotional and/or physical distress prevented them from undergoing testing postoperatively. While postoperative testing was usually performed prior to the patient seeing the surgeon (in the interests of consistency), on some occasions this was not possible. Depending on what the surgeon had told them in terms of pathological results, prognosis, recommendations for adjuvant chemotherapy, etc., subjects might be either unusually euphoric and motivated to perform, or significantly distressed and worried. Either reaction may affect postoperative scores (of either subjective or objective measures). A few subjects refused to undergo testing at follow-up after receiving upsetting news from the surgeon. Most of these subjects declined testing that day secondary to distress and anxiety. However, if they returned for further follow-up within the appropriate time frame, most of these patients agreed to subsequent testing. This points to the necessity of consistent timing in relation to the surgical consultation when performing postoperative follow-up tests, although the practicality of this is questionable.

Operative factors that seem to affect postoperative LOSOH are extent of resection and duration of anaesthesia. Patients who had extended resections demonstrated longer hospital admissions overall, when compared with those who received simple lung resection via thoracotomy alone. This is consistent with the physiological effects thought to accompany chest wall excision. Patients receiving extended chest wall resections may
experience an even greater impairment of normal ventilatory mechanics than those receiving simple thoracotomy. Busch et al. (1994), Wahi et al. (1989), and others have reported an increased incidence of postoperative complications in those undergoing chest wall resection at the time of lung resection (82, 105). It is sometimes suggested that these patients experience greater postoperative pain as well, but this has yet to be studied. Patients who receive intrapericardial dissection may be more prone to postoperative cardiac complications, such as atrial fibrillation. Again, this makes sense if we believe that myocardial irritability can predispose patients to arrhythmias postoperatively. The recent series by Amar and colleagues (1995) examined factors predisposing patients to postoperative tachydysrhythmias (93). He noted that extrapleural pneumonectomy is characterized by a greater incidence of supraventricular tachycardia (93). This is the only extended resection he describes: however, this type of resection frequently entails an intrapericardial dissection. The 2 patients with mesothelioma in our series underwent extrapleural pneumonectomy and intrapericardial dissection. 1 of these subjects experienced postoperative atrial fibrillation, and had a LOSOH of 19 days.

We also examined the amount of lung tissue resected to see if this was associated with postoperative complications. There was no significant difference seen in terms of mean LOSOH experienced; however, patients receiving bilobectomies showed considerably more variability in terms of LOSOH than did other patients. The Group B patients were distributed between the bilobectomy, lobectomy, and intermediate category (> 1 lobe resected). Of interest is the relatively good outcome experienced by those patients who underwent pneumonectomy in our study. Most authors report that pneumonectomy is characterized by a higher complication rate (1.3.81-82,93). However, Deslauriers (1994) reported no increased complication rate among pneumonectomy patients in his prospective study (86). None of the Group B patients in our study received pneumonectomies.

Various additional preoperative measures were evaluated which were considered likely to be associated with LOSOH. No correlation was found between postoperative
LOSOH and either hypercarbia or hypoxemia as documented on ABGs. However, it is possible that this lack of association is secondary to small numbers (lack of statistical power), as only 38 subjects had results available on the chart (information retrieval from the hospital and clinic charts being a major obstacle in this study). A few subjects demonstrated abnormal preoperative values, yet none of these subjects experienced significant postoperative complications. Indeed, some subjects with abnormal values experienced exceptionally good operative outcomes. This supports the findings of Kearney et al. (1994) and Busch et al. (1994) who all found that preoperative hypercarbia and hypoxemia did not predict operative morbidity and mortality (81-82).

Nutritional status (represented by BMI) was not associated with operative outcome. We hypothesized that severe malnutrition (or conversely, morbid obesity) might be associated with a delayed recovery or an increased incidence of postoperative complications. Several authors have documented a positive association between the two variables (12.82.127). We found no such association, but the study was underpowered to address this issue. None of the subjects in this study was morbidly obese. A few subjects had low BMI scores, but these subjects did not demonstrate a trend towards prolonged recovery periods. All 4 Group B subjects were within normal limits for BMI. Of interest in future studies would be to document relative preoperative weight loss (of unexplained origin). This was not recorded in the current study, but may be a marker of more advanced stages of neoplastic disease.

Another measure that we evaluated — and of particular interest to lung cancer researchers — is smoking history. We documented preoperative smoking history (in pack-years) and compared this with postoperative LOSOH. No association was found between these two variables, either in the entire sample or in Groups A and B. This finding is in keeping with the findings of Busch and colleagues (82). Other authors have found a positive correlation between smoking history and operative outcome (81.128).
**Functional Exercise Testing:**

The 6-minute walk test is considered an acceptable alternative to complex laboratory exercise tests of VO₂max (38-40.66, 112, 114, 129). It is a reliable and inexpensive screening tool, which is well-tolerated by most patients. We employed the walk test as a primary measure in our study. We encountered a significant problem in terms of the feasibility of repeating the test 3 times preoperatively, as advocated in the literature (38-39, 42, 61, 66, 112, 114, 129). Because of these difficulties, we decided to investigate the reliability of the test in this population, in particular the reliability of a 2-trial model. A review of the literature revealed that most of the validation studies of this test were performed in patients with severe COPD and CHF. There are no trials published specifically evaluating the upper limits of this test in normals, although Lipkin (1986) mentions 10 normal comparators (112). Early on in the study, we noticed no significant improvement between the first and second trials. Because this was so different from the published accounts, we wondered whether our population had a different reliability profile than those appearing in the COPD literature (perhaps secondary to less impaired pulmonary function and less symptomatology).

We evaluated test-retest reliability with the study subjects, as well as those who were excluded from the study based on stage of disease (those deemed inoperable). In addition, we evaluated test-retest reliability among healthy young normals to determine the upper limits of the test. We sought to determine whether the reliability profile of this test is a function of pathology, rather than a function of the intrinsic properties of the test.

With respect to the lung resection candidates evaluated, we found that the 2 trials of the 6-minute walk demonstrated strong correlations. The 7 subjects who underwent 3 trials demonstrated no significant improvement between the different trials. Intra-rater and inter-rater Pearson’s correlation coefficients were also acceptable. However, calculation of the ICC for each form of reliability revealed that our results were not nearly as consistent as we might assume from Pearson’s r values. There was a significant degree of intra-subject
variability. We were curious to find that the lowest ICC was found for the intra-rater testing. This might be secondary to some feature intrinsic to this population, or it could be related to the testing procedure itself. Because of the difficulty in accessing subjects prior to surgery, the raters were forced to test some patients the morning of their surgery. Specifically, the investigator (Rater 1, on whom the intra-rater reliability was evaluated) rated a disproportionate number of patients (for trial 2) between 6:30 and 7:00 a.m. on the morning of surgery. These patients may have been particularly anxious during their second test. They would not have eaten since the previous night, and often complained of having slept poorly. These factors might account for the large drop seen in some subjects' walk distances between trials 1 and 2, and therefore would account for the low ICC statistics. Because of the low ICC values, it seems unwise to recommend only 2 preoperative trials to establish a baseline measure. 3 trials would be prudent. It would be valuable to perform further testing, changing the time of the second test to see if an improved ICC would result.

Walk distance showed a significant association with postoperative LOSOH in the primary analysis. This test evaluates physical function, but also entails significant motivation on the part of the patient. Moreover, the walk test mimicks one of the primary components of postoperative rehabilitation following lung resection, namely early mobilization/ambulation. If patients show excellent functional performance status preoperatively, it is possible that this relates to their motivation and functional ability to engage in postoperative rehabilitation. Conversely, if patients have difficulty performing this simple test preoperatively, it is suggests that they may have more difficulty undergoing early mobilization postoperatively. Decreased activity levels postoperatively (in many surgical populations) are associated with increased pulmonary and cardiovascular complications (56).

The normative testing demonstrates that the 6-minute walk test is a highly reliable (reproducible) test in young healthy normals, and suggests that the variability seen in various patient populations is more a reflection of disease/symptomatology (or other
factors, such as test timing) rather than an inherent feature of the test itself. The upper limits for the test in normals range from approximately 650 m to 750 m. The limiting factor is tibialis anterior muscle fatigue and cramping. These results are borne out by unpublished reports of testing on normals carried out by Gibbons and colleagues in Montreal (personal communication, publication pending). As mentioned above, Lipkin et al. (1986) found similar results in their testing of 10 normal comparators (112). We realize that our numbers of normal subjects is small and plan to conduct further testing in the future, both on young healthy normals, and older healthy normals.

The results which we obtained both in a lung resection population and in normals (coupled with the data on COPD patients received from the rehabilitation facility) suggest that the reliability profile of the test may range along a continuum, with COPD and CHF patients exhibiting a steep learning curve with the test, and normals demonstrating little (if any) learning curve. This may or may not reflect the role of symptom severity and warrants further study. It seems plausible that lung resection candidates should lie somewhere in the middle, with varying degrees of concomitant COPD (albeit less severe than in the COPD populations evaluated in the literature) and symptomatology. As disease symptom severity increases (namely dyspnea, paroxysms of coughing, bronchospasm), patients (either with COPD or with lung neoplasms) may underachieve on initial trials of the test until they feel confident that they can perform the test safely following a number of practice runs.

As mentioned at the outset, some lung resection candidates have varying degrees of COPD, but many are completely asymptomatic and exhibit few (if any) symptoms of airflow limitation. For these reasons, we expected their mean walk distances to be higher than those achieved by patients with moderate to severe COPD. However, our results demonstrate that this was not the case. One possible explanation might be the difference in encouragement protocol between the 2 facilities. The rehabilitation facility used a testing protocol which included maximal standardized encouragement. Our facility employed a
standardized protocol of no encouragement. Discussions with staff at the rehabilitation facility indicated that other testing parameters were essentially the same (as outlined in the literature). Guyatt et al. (1984) noted that encouragement could increase the distance covered by up to 30 m in settings of COPD and CHF (66). It would be of interest to perform further testing with lung resection candidates and healthy normals to determine if this might influence the mean distances achieved.

Stair climbing is likewise a simple, cost-effective test for evaluating preoperative exercise tolerance, and may be used to estimate VO₂ (albeit roughly) (14,115). It has been employed by Holden et al. (1991) and Olsen et al. (1990), demonstrating significant predictive ability (14,115), but requires further validation. Variations on stair-climbing (such as desaturation during a climb of two flights of stairs) have been employed in other series (117). We adopted the approach of Holden and Olsen and calculated the estimated VO₂ (mL/kg/min) and found that this measure was indeed predictive of postoperative LOSOH, as was workload achieved (Watts), and rate of stair climb (steps/minute). The high correlations seen with the 6-minute walk test indicate that it is indeed measuring functional exercise tolerance, lending it concurrent validity (77). Both these tests measure functional exercise ability and performance is contingent upon a number of factors, such as normal daily activity levels, motivation, pulmonary function, and overall cardiopulmonary fitness (115). Pulmonary function measures (specifically FEV₁ and DLco) are also moderately correlated with the stair-climbing test, indicating that they are not testing the exact same entity, but lends credence to the assumption that pulmonary function is important to exercise tolerance (14,115). Again, this adds to the concurrent validity of these measures (77). This is an interesting finding, given previous studies evaluating 6-minute walk distance and PFTs which suggest relatively poor correlations between FEV₁ and distance walked (31,38,42).

There are a number of difficulties inherent to the performance of the stair-climbing test. While subjects are asked to climb as many stairs as they can (up to a maximum of
subject effort is highly variable. For safety reasons the rater climbs with the subject, monitoring HR and oxygen saturation. There is a mild but clinically significant lag between the reading of maximum HR (and minimum oxygen saturation) attained following the cessation of climbing. There was also a lag in onset of symptomatology (either extreme fatigue or acute onset of dyspnea). Rater concerns regarding patients' potential for significant distress during the test may result in the rater stopping the climb before the patient has made a sufficient effort. As such, the subjects' values of VO₂ and workload achieved may underestiamte their potential to perform on the test. Neither Holden (1992) nor Olsen (1991) performed 2 or 3 repeated measures of the test to determine its reliability, although it did demonstrate strong correlations with laboratory measures of VO₂ max. We followed their testing protocol, but perhaps we should have tried a repeated measures approach as we did with the level walking test. Subject compliance with a repeated measures protocol may be questionable (as patients reported this to be a more difficult test, and we were concerned about overloading subjects with tests preoperatively).

Finally, it is questionable whether a stair-climbing test is valid in all lung resection candidates. Level walking is an activity commonly performed by most patients. However, not all patients have stairs in their homes, and it may not be an activity in which they regularly engage. As such, it may not represent their normal level of function. However, its specificity is better than cycle ergometry for most elderly patients as subjects tend to attain higher VO₂ levels during stair-climbing than they do during cycle ergometry tests of VO₂ (79, 138).

**Limitations of the Study:**

3 patients refused to participate in the study secondary to extreme preoperative anxiety and physical dysfunction. This indicates that some (including endstage) patients may be (a) unable to participate in even minimally invasive trials and (b) may be so compromised by their disease that they are unable to complete questionnaires (and consent
forms) even with assistance available. It is of concern to the investigator that patients experiencing such severe dysfunction are unable to participate, as the sample does not represent those with all possible degrees of dysfunction (the sickest or the most anxious patients are unable to participate, so this represents a selection bias inherent to the study). However the 97% agreement offered by the remaining subjects was encouraging and indicates that, overall, such testing is quite acceptable to the majority of patients in the preoperative setting.

Problems with follow-up were encountered, with some patients again citing emotional and physical distress as reasons for foregoing follow-up testing on some occasions. Of interest, most patients agreed to complete testing at subsequent appointments. Those subjects who developed recurrences during the course of the follow-up period were most likely to refuse further testing of any kind. The second most likely set of patients to refuse follow-up testing were those who had recently received recommendations for adjuvant therapy and/or those who had been told that they had positive resection margins. This means that our protocol obtained incomplete follow-up evaluation of some of the most distressed patients, so the conclusions that we can draw from the follow-up data are limited.

The subjects recruited for our study included those with primary lung cancer, those with pulmonary metastases, and a few (2) with benign lesions (although these lesions were suspicious for lung cancer preoperatively). We did not attempt to compare health status as rated by those with primary disease with those experiencing metastatic disease. This would indeed be of interest. It is possible that those experiencing recurrences or metastases would rate their global QL and SF and EF scores significantly lower than those with primary (and newly diagnosed disease). In addition, the former groups might have considerably different expectations of surgery than the latter one. It makes intuitive sense that patients undergoing multiple surgeries should respond differently, both in terms of subjective ratings of health status and even in terms of LOSOH. The heterogeneity of the subjects in
our sample may account for the considerable variance unaccounted for in our study. It is certainly a topic which should be investigated further.

Another major limitation to the study is the skewed distribution of the dependent variable. I have discussed this limitation at considerable length already. We had expected more of a continuum of postoperative LOSOH. It is obvious from the analysis above that it is very difficult to predict the extreme outcomes suffered by the Group B subjects. Perhaps a multi-centre trial with a larger sample size might avoid this problem in the future, or perhaps separate analyses of those who suffered extreme complications would be more useful (i.e. separate consideration of morbidity and mortality).

An obvious concern of the investigators is the validity of the dependent variable chosen for our study, namely LOSOH. Many of our concerns have been addressed in previous sections. However, we must reiterate the need for further study to truly validate this promising surrogate measure of complications. It is highly correlated with postoperative complications (as well as with their relative severity), but considerably more work needs to be done before we can definitively declare this to be a global measure of postoperative complication rate. Attention to issues of health-care utilization patterns, days of normal activity, relative loss of premorbid function, and rapidity of disease recurrence should all be addressed in future studies.

Information retrieval proved an obstacle to data collection during this study. In an effort to avoid bias, I was blinded to certain preoperative measures (namely PFT and ABG results) during the data collection phase. This necessitated a delay in accessing this information from the patients' charts or from the clinical laboratories involved until after data collection was complete. This meant a lag-time in accessing the data for those patients recruited at the beginning of the study for up to 14 months. Upon completing the data collection phase, we attempted to access this information from the appropriate sources. It became apparent that some PFT and ABG data had been lost from patient files (in various settings), or not placed on permanent record. This loss of data occurred despite informing
all senior surgical residents of the study protocol, of double checking with patients that these tests had been ordered, of arranging appointments for the appropriate tests for patients, etc. In future studies, it would be better to employ an independent data manager to collect this information at the time of testing, with the investigator remaining blinded to the results. Unfortunately, we did not anticipate this problem during this trial, nor did we have funding to employ a research assistant or data manager.

**Implications and Future Directions:**

As mentioned above, a large amount of the variance in LOSOH remained unaccounted for in our analysis. Additional, as yet, unidentified factors are influencing operative outcome. It was the goal at the outset of this study to evaluate the relative predictive ability of new measures such as the QLQ-C30 questionnaire, specifically its measures of preoperative psychosocial function. The results of this study demonstrate that it is difficult to predict extreme complications, such as those suffered by the 4 Group B subjects. Subjective measures of health status did not predict these patients' extraordinarily poor outcomes. In those patients who experienced less extreme sequelae, the global QL scale was the preoperative measure most predictive of LOSOH. When combined with a functional exercise test (either the 6-minute walk or stair climbing), a significant proportion of the variance in LOSOH is accounted for. We did not find a significant association between LOSOH and FEV\(_1\). We do not recommend abandoning preoperative evaluation of this measure, but reliance on this measure alone for determining operability seems unwarranted. DLco showed more promise. We would recommend that surgeons continue to use these physiologic measures, but should incorporate both preoperative health status measurement and functional exercise testing into their
preoperative lexicon. While these measures may not identify those at risk of catastrophic complications, it will help them identify those at increased risk for prolonged hospitalization or early readmission. At the moment, no preoperative measure available appears predictive enough that it could preclude a candidate from receiving a potentially curative resection.

While the EORTC health status instrument shows some promise, certain limitations must be noted. The EF and SF scales showed significant ceiling effects, which makes us question the validity of employing these particular scales as preoperative measures (21). Scales such as these lack discriminant validity in that they fail to differentiate between those subjects who have a prolonged postoperative LOS from those who do not (68). The global QL score did not demonstrate such a ceiling effect, making it more suitable for preoperative evaluation. Those who rate their global QL highly are more likely to have a shorter perioperative recovery. Given the significant associations found between LOSOH and the role functioning, sleep quality, fatigue, and general pain scales in the Group A patients, it would be interesting to explore these further. This should be a focus of future studies. It might be valuable to employ a similar study design to the present one, substituting the above 4 measures for the 3 health status ones which we employed (with appropriate corrections to sample size).

Comparisons of the symptom scale scores for those who were deemed inoperable based on stage of disease with those subjects who proceeded to resection, suggest that the EORTC core questionnaire and its lung modular supplement can discriminate between subjects with different stages of lung cancer. Subjects who were unresectable were evaluated prior to their staging procedure (they were unaware of their disease stage at the time of testing). Subjects with stage III and IV cancer reported higher levels of distress and symptomatology than those with less advanced stages of disease. The instrument is therefore responsive to different stages of lung cancer in the preoperative setting. In previous studies validating the instrument in a setting of inoperable disease, the instrument
showed responsiveness to changes in clinical status over time, as well as the ability to discriminate between patients with objectively rated differences in clinical status (23.24).

With respect to health status measurement, it would be interesting to explore a number of other instruments to see if any of these might be more predictive of operative outcome than the EORTC instrument employed in this study. The FACT-L is a newer measure that is still undergoing validation. It would be interesting to examine it in a preoperative as well as a postoperative setting.

Another potential application is to use health status measurement over a longer period of time and to follow patients for the first year postoperatively (not just the first 3 months). This would be feasible given that most patients are monitored every 3 months by their surgeons for the first year postoperatively. This would allow us to track patterns of longterm functional recovery. Longer follow-up would allow us to track potential later recurrences of tumours.

A secondary preoperative measure that demonstrated considerable promise as a potential predictor was the stair-climbing test. Stair-climbing rate, workload achieved and submaximal VO₂ were correlated with LOSOH, both for the sample as a whole and for the Group A subjects. Workload (W) was the most predictive of these 3 interrelated measures and is worth investigating further. A prospective study evaluating related parameters of this test (relative desaturation, change in heart rate and respiratory rate, subjective ratings of dyspnea, and the influence of accompaniment/absence of a tester) is warranted. Cataloguing the specific types of postoperative complications (e.g. pulmonary, cardiac, other complications) predicted by this test would be valuable.

Of particular interest would be a trial investigating the role of various clinical interventions and their impact on operative outcome (including postoperative health status). Since we know that these measures (HRQL instruments and functional exercise tests) are sensitive to clinical change over time in a lung resection population, perhaps it could be used as an outcome measure to compare the outcomes of different forms of surgery (VATS
vs. thoracotomy; sternotomy vs. thoracotomy), as well as different postoperative rehabilitation protocols.

As a preoperative screening tool, perhaps clinicians could employ these measures (health status indicators and functional exercise tests) to identify patients who might benefit from a preoperative program of rehabilitation, education, and/or psychosocial support.

**Conclusions:**

This type of study — incorporating preoperative measures of health status, functional disability, and physiological impairment has not been undertaken before in the lung resection population. The 2 best predictors of postoperative LOSOH in our study were the global QL scale of the QLQ-C30 questionnaire and the 6-minute walk test. These 2 measures demonstrate promise as predictive measurement tools.

To our knowledge, LOSOH (as a measure representing postoperative complications) has not been used previously in surgical trials. LOSOH demonstrated construct validity in our sample. It has the potential to encompass a wide array of complications (including death) within a single measure. It also accounts for readmission to hospital within the first 30 postoperative days. However, while it correlates with postoperative complications and their relative severity, there are numerous possible confounding factors that could impact on LOSOH score. Future studies will be required in order to validate this surrogate measure.

Our study indicates that it remains extremely difficult to predict severe postoperative complications in subjects undergoing single lung resection. The current literature available on this topic supports this assertion. Our study used preoperative administration of a multidimensional health status instrument to see if it could assist in predicting prolonged hospital admission postoperatively. Such measures have not been used as predictive
measures before in this patient population. Besides the health status instrument, 6-minute walk distance and the current 'gold standard' (PFTs) were also evaluated preoperatively.

Neither the current gold standard (FEV₁ and DLco), nor measures of physical function (6-minute walk test), nor health status measures were able to predict extreme operative outcomes as measured by death or truncated LOSOH (the Group B patients in this study). Other physiologic measures commonly employed (for instance ABGs) were no more helpful in predicting the worst outcomes. Age was the only variable predictive of exceptionally poor LOSOH. However, equally elderly patients enjoyed good operative results in our study, so we do not recommend denying these patients potentially curative procedures based on age alone.

For those patients experiencing a more standard postoperative course (Group A), the QLQ-C30’s global QL scale was significantly predictive of LOSOH, as was the 6-minute walk test. These were the 2 best predictors of operative outcome in our study. The SF scale and DLco also demonstrated trends toward prediction in these patients. However, the majority of the variance contributing to postoperative LOSOH remains unaccounted for.

A secondary preoperative measure evaluated was the stair-climbing test (as an estimate of submaximal workload and oxygen consumption). This measure was positively correlated with operative outcome in the sample as a whole and for those who experienced more typical operative outcomes. This measure should be evaluated in more depth in the future.

Postoperative recovery following lung resection was evaluated using the subjective health status instrument and the level walking test. Patients experience significant declines in terms of physical function and quality of life in the immediate postoperative period, but most appear to return to their preoperative level of function by 3 months postoperatively. Both the 6-minute walk test and the health status questionnaire are useful evaluative measurements to employ postoperatively. Interestingly, emotional and social function
improve considerably beyond preoperative values during the first 3 postoperative months. Many more investigations are required before we can safely predict operative outcome in this challenging patient population.
APPENDIX A

EORTC Health Status Instruments

Please note that these instruments are copyrighted to the EORTC and have been reproduced here with the express permission of that organization. Permission to employ these instruments and to reproduce them must be sought from the EORTC, Brussels, Belgium.
EORTC QLQ-C30 (version 2.0.)

We are interested in some things about you and your health. Please answer all of the questions yourself by circling the number that best applies to you. There are no "right" or "wrong" answers. The information that you provide will remain strictly confidential.

Please fill in your initials: [ ] [ ] [ ]
Your birthdate (Day, Month, Year): [ ] [ ] [ ]
Today's date (Day, Month, Year): [ ] [ ] [ ]

1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase? [ ] [ ]
2. Do you have any trouble taking a long walk? [ ] [ ]
3. Do you have any trouble taking a short walk outside of the house? [ ] [ ]
4. Do you have to stay in a bed or a chair for most of the day? [ ] [ ]
5. Do you need help with eating, dressing, washing yourself or using the toilet? [ ] [ ]

During the past week:

6. Were you limited in doing either your work or other daily activities? [ ] [ ] [ ] [ ]
7. Were you limited in pursuing your hobbies or other leisure time activities? [ ] [ ] [ ] [ ]
8. Were you short of breath? [ ] [ ] [ ] [ ]
9. Have you had pain? [ ] [ ] [ ] [ ]
10. Did you need to rest? [ ] [ ] [ ] [ ]
11. Have you had trouble sleeping? [ ] [ ] [ ] [ ]
12. Have you felt weak? [ ] [ ] [ ] [ ]
13. Have you lacked appetite? [ ] [ ] [ ] [ ]
14. Have you felt nauseated? [ ] [ ] [ ] [ ]
15. Have you vomited? [ ] [ ] [ ] [ ]

Please go on to the next page
During the past week:

16. Have you been constipated?  
1 2 3 4
17. Have you had diarrhea?  
1 2 3 4
18. Were you tired?  
1 2 3 4
19. Did pain interfere with your daily activities?  
1 2 3 4
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?  
1 2 3 4
21. Did you feel tense?  
1 2 3 4
22. Did you worry?  
1 2 3 4
23. Did you feel irritable?  
1 2 3 4
24. Did you feel depressed?  
1 2 3 4
25. Have you had difficulty remembering things?  
1 2 3 4
26. Has your physical condition or medical treatment interfered with your family life?  
1 2 3 4
27. Has your physical condition or medical treatment interfered with your social activities?  
1 2 3 4
28. Has your physical condition or medical treatment caused you financial difficulties?  
1 2 3 4

For the following questions please circle the number between 1 and 7 that best applies to you

29. How would you rate your overall health during the past week?

1 2 3 4 5 6 7

Very poor

Excellent

30. How would you rate your overall quality of life during the past week?

1 2 3 4 5 6 7

Very poor

Excellent

* Copyright 1995 EORTC Study Group on Quality of Life. All rights reserved. Version 2.0
Patients sometimes report that they have the following symptoms. Please indicate the extent to which you have experienced these symptoms during the past week.

<table>
<thead>
<tr>
<th>During the past week:</th>
<th>Not at All</th>
<th>A Little</th>
<th>Quite a Bit</th>
<th>Very Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>31. How much did you cough?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>32. Did you cough blood?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>33. Were you short of breath when you rested?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>34. Were you short of breath when you walked?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>35. Were you short of breath when you climbed stairs?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>36. Have you had a sore mouth or tongue?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>37. Have you had trouble swallowing?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>38. Have you had tingling hands or feet?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>39. Have you had hair loss?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>40. Have you had pain in your chest?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>41. Have you had pain in your arm or shoulder?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>42. Have you had pain in other parts of your body?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

   If yes, where .................................................................

43. Did you take any medicine for pain?

   1    No  2    Yes

   If yes, how much did it help?  

   1    2    3    4

© QLQ-C30-LC13 Copyright 1994 EORTC Study Group on Quality of life. All rights reserved
APPENDIX B

Consent Form
APPENDIX B

Title of Research Project: Health Status Measures for Predicting Postoperative Complications Following Lung Resection

Researchers: Janet Parsons, B.Sc.(PT), M.Sc. student
University of Toronto, Institute of Medical Science; physiotherapist, Mount Sinai Hospital

Supervisor: Dr. Arthur Slutsky, Head of Respiratory Medicine Division, Institute of Medical Science, University of Toronto; Respirologist, Mount Sinai Hospital

I, ____________________________, have been invited to participate in a study by the above-named researchers. The study will examine the ability of a quality of life questionnaire to predict length of stay in the hospital after surgery. This questionnaire has been used before with other types of patients, and has been shown to be an effective measurement tool.

I will be asked to complete the questionnaire before the surgery, and again at 5 days and 3 weeks after the operation (during my regularly scheduled follow-up appointment with my surgeon). I will also be asked to perform a walk test (level walking for 6 minutes) before my surgery, and will repeat the test at 5 days and 3 weeks following the surgery.

This testing will help the researchers understand how patients' views regarding their health, as well as general physical function impact on the results of surgery. The testing will not alter the surgery I undergo, and it will not influence the types of treatment I will receive after the operation.

I am aware that data gathered during this study may be used in research reports, but that all personal information
and my identity will be kept strictly confidential. I understand that enrollment in the study is completely voluntary and that I am free to withdraw at any time. I realize that no money will be paid for taking part in the study.

If I have any questions regarding any aspect of this study, I may contact either Janet Parsons from the Physiotherapy Department (586-5035) or Dr. Arthur Slutsky of the Respirology Department (586-5298), or my surgeon (Dr. Michael Johnston/Dr. Alan Casson/Dr. Gail Darling) (586-8512), all of Mount Sinai Hospital.

______________________________  ________________________________
Subject signature                    Investigator signature

______________________________  ________________________________
Witness signature                  Date
APPENDIX C

Modified Borg Scale (RPE)
<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Nothing at all</td>
</tr>
<tr>
<td>0.5</td>
<td>Very, very light</td>
</tr>
<tr>
<td>1</td>
<td>Very Light</td>
</tr>
<tr>
<td>2</td>
<td>Light</td>
</tr>
<tr>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>4</td>
<td>Somewhat hard</td>
</tr>
<tr>
<td>5</td>
<td>Heavy</td>
</tr>
<tr>
<td>7</td>
<td>Very Heavy</td>
</tr>
<tr>
<td>10</td>
<td>Very, very heavy (almost maximal)</td>
</tr>
</tbody>
</table>
APPENDIX D

Diagnostics of Multiple Regression Residuals
Probability Plot of Residuals (Group A)

Plot of Studentized Residuals vs. Estimated Values (Group A)
Plot of Cook's Distance vs. Estimated Values (Grp. A)
Probability Plot of Residuals (Case 11 Excluded)

Plot of Studentized Residuals vs. Estimated Values
Plot of Cook's Distance vs. Estimated Values (Case 11 Excluded)
APPENDIX E

Instructions to Patients for Performance of the
6-minute Walk Test
Instructions to Patients for Performance of
the 6-minute Walk Test

This test is used to evaluate your ability to exercise and your overall heart and lung function. All patients having thoracic surgery at Mount Sinai Hospital undergo this form of testing as part of their routine preoperative workup.

We want to see how far you can walk in 6-minutes. The goal of the test is for you to cover as much ground as you can. You should walk as far and as fast as you can during the 6 minutes. At the end of the test, you should feel that you would be unable to cover more ground than you did during the 6-minute time period. When the 6-minutes are up, I will yell 'stop'. I want you to stop where you are. I will come to that spot to meet you.

If you need to, you can stop and rest during the test. If you become short of breath or excessively fatigued during the test, you can stop. When you feel more comfortable you may start walking again. There are chairs placed at either end of the course. These can be used for resting if needed. Or you can rest in the standing position, if you prefer.

I cannot walk with you, since most people tend to fall into pace with the person they are walking with. I do not want to influence the pace at which you are walking. I will stand at the end of the course timing you. No other persons are permitted to walk with you during the test.

I cannot give you any encouragement during the test, and I am not allowed to tell you how well you are doing. I can tell you how many minutes you have been walking (how much time remains in the test). You should try not to talk more than necessary during the test. I do want you to tell me if you become short of breath, develop any chest pain or chest tightness, or become dizzy or light-headed during the test.

You may stop the test at any time if you wish.

At the beginning and end of the test, I will measure you oxygen level and your heart rate, using this small probe (pulse oximeter). If you become short of breath, etc., during the test I may check your oxygen level and heart rate part-way through. At the end of the test I will also ask you to tell me how hard you felt you were working (exerting yourself) during the test (patient shown Borg RPE—see Appendix C). I will also ask you to tell me what was the main thing (reason) that prevented you from walking further (e.g. shortness of breath, leg pain/cramping, fatigue, etc.).
References


