Prediction of Dyspnoea Following Lung Resection for Cancer

Brian Daniel Lafferty
MBChB, FRCA, MAcadMEd

Submitted in fulfilment of the requirement for the Degree of Doctor of Medicine

College of Medical, Veterinary and Life Sciences
University of Glasgow

December 2021

Abstract

Lung cancer is the leading cause of cancer death in the UK, with a high incidence in Scotland. In suitable cases surgical resection is the first-choice treatment but is associated with high rates of post-operative morbidity and mortality.

Survival with a meaningful quality of life is important. Public engagement work by our research group has demonstrated that second only to "being alive and cancer free," exercise capacity was the main priority of patients. However, prediction of post-operative dyspnoea is often difficult and inaccurate. Conventional prediction relies on estimation of function and quantity of lung remaining following surgery. The British Thoracic Society and the National Institute of Clinical Excellence recommend pulmonary function testing and calculation of predicted post-operative FEV_1% (ppoFEV_1%) and DLCO% (ppoDLCO%), with <40% in either domain being considered 'high risk' for post-operative dyspnoea. Whilst these calculations correlate well with post-operative pulmonary function they are not well associated with functional outcomes. No effective method exists for identifying risk of, nor therapeutic strategies to prevent, post-operative dyspnoea.

The British Thoracic Society, The European Society of Thoracic Surgeons and the National Institute of Clinical Excellence highlight the need for studies concerning patient fitness and operative risk when assessing patient suitability lung resection. Furthermore, the James Lind Alliance identified *"improving recovery from surgery for elderly patients"* as a top 10 priority.

The aim of this thesis was to improve conventional prediction of post-operative dyspnoea. Pilot data from our research group demonstrated association between B-Type natriuretic peptide and both; post-operative cardiac dysfunction and post-operative dyspnoea. The author proposes a novel scoring tool incorporating B-Type natriuretic peptide alongside conventional measurements.

B-Type natriuretic peptide is a quantitative biomarker of myocardial dysfunction, identifying patients at risk of cardiopulmonary complications in a variety of surgeries. Current international guidelines recommend using B-Type natriuretic peptide to aid prognostication of peri-operative morbidity in high-risk patients prior to non-cardiac surgery, yet its potential role in peri-operative decision

making in lung resection is unclear. No previous work has compared B-Type natriuretic peptide to functional outcomes following lung resection.

The first investigation of this thesis (chapter 8) explores conventional risk prediction methods in a single site derivation population of 93 patients at the Golden Jubilee National Hospital. Results highlighted poor performance of conventional methods to predict post-operative dyspnoea, confirming the sole use of pulmonary function in this setting could be improved.

In response to these findings, *new* models were explored (Chapter 9). Univariate analysis identified risk predictors for candidates with and without post-operative dyspnoea. Variables with significance were used to derive new predictive models, incorporating B-Type natriuretic peptide. New models improved prediction within the internal dataset.

An external dataset from three other UK sites was used in an attempt to validate these new models (Chapter 10). Conventional and new models performed similarly within the external population, highlighting the challenge of creating a new scoring tool. Although B-Type natriuretic peptide did not improve risk prediction in either the internal or external dataset, the analysis highlighted the potential of other variables to predict post-operative dyspnoea, such as body mass index, diabetes status and pre-operative pain and quality of life scores.

Secondary analyses demonstrated post-operative B-Type natriuretic peptide was greater in those with increasing post-operative morbidity (>1 complication), those with new post-operative atrial fibrillation and those with pulmonary complications (Chapter 11). A positive association between post-operative BNP and length of hospital stay was also demonstrated. Lung function testing displayed an association with post-operative outcome when used as a continuous variable. There also existed an association between pre-operative quality of life, pre-operative performance status and pre-operative ASA which has not been shown before in this population. These positive findings could be useful in the pre-operative setting when planning surgery in a shared decision setting.

The work within this thesis confirms current risk prediction methods must be improved, but also highlights the challenges involved in creating scoring tools for

use in clinical practice. Future work in this area may involve low technology testing such as heart rate recovery, in addition to the independent predictors of post-operative dyspnoea discovered here, to improve prediction of dyspnoea following lung resection surgery.

Acknowledgements

Firstly, I would like to acknowledge my supervisors. Dr Ben Shelley and Dr Philip McCall for their support and encouragement to embark upon this research project. I am grateful to both for granting me this opportunity and their continual guidance and advice has made all of what I have achieved possible. Their passion for research and high standards inspired me and I have learned so much working with them both.

I wish to thank Dr Catherine Sinclair and the research nursing team at the Golden Jubilee National Hospital; Elaine Matthews, Joanne Kelly, Ruth McLaren, Elizabeth Boyd, Julie Buckley, Robert McDade, Alexandra Tewkesbury, Angela McGlade and Joanne Ford who supported the study and helped with recruitment when I was not able to be present.

I am indebted to the patients and staff at the Golden Jubilee National Hospital who were involved in this research project. Thank you to all who supported and assisted in recruitment of patients and taking blood samples for me.

I am also indebted to the patients and staff at the other collaborating centres who were involved in this research project and supported with recruitment; Aberdeen Royal Infirmary, Edinburgh Royal Infirmary and Royal Victoria Hospital Belfast.

Finally, this thesis is dedicated to my family without whom this would not have been possible. To my wife Jo-Ann, who has supported me and sacrificed much through a global pandemic to ensure I managed to complete this work on time. To my two daughters, Sarah and Emma, all of this work has been done with you in mind. I would also like to thank my parents for a lifetime of support and advice, without which I would not be where I am today.

Word Count

66,815 words (excluding Appendices)

Author's Declaration

The majority of the work detailed within this thesis was performed by me while I was employed as an Anaesthetic Clinical Research Fellow at the University of Glasgow, laterally being completed upon returning to Anaesthetic Speciality Training within the west of Scotland. It was carried out between February 2018 and July 2021.

The majority of recruitment was performed by me at the Golden Jubilee National Hospital for the derivation dataset, with the remainder of the recruitment carried out by individuals acknowledged.

Statistical advice was sought from Drs Martina Messow and Robin Young at the Robertson centre for biostatistics at the University of Glasgow, but all analyses detailed within this thesis were performed by me.

The remainder of the work described in this thesis was performed by myself and the writing of this thesis is my own work.

Brian Daniel Lafferty

December 2021

Table of Contents

ABSTRACT **2**

ACKNOWLEDGEMENTS **4**

WORD COUNT **6**

AUTHOR'S DECLARATION **7**

TABLE OF CONTENTS **8**

LIST OF TABLES **12**

LIST OF FIGURES **14**

GRANTS **16**

PUBLISHED ABSTRACTS **16**

ORAL PRESENTATIONS **16**

DEFINITIONS/ABBREVIATIONS **18**

1 LUNG CANCER **21**

1.1 LUNG CANCER: INTRODUCTION 21
1.1.1 Lung cancer mortality *21*
1.1.2 Risk factors for lung cancer *22*
1.2 CLASSIFICATION OF LUNG CANCER 22
1.2.1 Lung cancer management and surgery *22*
1.2.2 Post-operative mortality and morbidity *24*
1.2.3 Quality of life and patient reported outcome measurements *25*
1.2.4 Patient reported outcome measures *26*
1.2.5 Dyspnoea following lung resection *27*
1.2.6 Shared decision making *30*
1.2.7 Conclusion *31*

2 MEASURING DYSPNOEA AND QUALITY OF LIFE **32**

2.1 SCORING TOOLS TO ASSESS DYSPNOEA 32
2.1.1 Minimum Clinically Important Difference *33*
2.1.2 Visual Analogue Scale to measure dyspnoea *34*
2.1.3 Medical Research Council scale to measure dyspnoea *35*
2.1.4 University of California and Sand Diego Shortness of Breath Questionnaire to measure dyspnoea *37*
2.2 SCORING TOOLS TO ASSESS QUALITY OF LIFE 38
2.2.1 EQ-5DL quality of life scoring tool *38*
2.2.2 European Organisation For Research And Treatment Of Cancer quality of life questionnaire *40*
2.2.3 World Health Organisation disability schedule 2.0 ..... *42*
2.2.4 Hospital anxiety and depression scale *42*

3 PROPOSED MECHANISMS OF DYSPNOEA AND REDUCED FUNCTIONAL CAPACITY FOLLOWING LUNG RESECTION SURGERY **44**

3.1 REDUCED RESPIRATORY MUSCLE FUNCTION AND REDUCED PULMONARY FUNCTION 46
3.2 CARDIOVASCULAR DYSFUNCTION 48
3.2.1 Post-operative dysrhythmias *48*
3.2.2 Right ventricular dysfunction *49*
3.2.3 Myocardial injury following non-cardiac surgery *53*
3.2.4 Pulmonary embolism *54*
3.2.5 Shunting *55*
3.2.6 Conclusion *56*

4 OPERATIVE RISK ASSESSMENT IN LUNG CANCER PATIENTS **57**

4.1 METHODS: REVIEW OF THE LITERATURE 57
4.2 INTRODUCTION TO RISK ASSESSMENT IN THORACIC SURGERY: CONVENTIONAL RISK STRATIFICATION 59
4.2.1 Peri-operative death 60
4.2.2 Post-operative cardiac event 62
4.2.3 Dyspnoea 63
4.3 PULMONARY FUNCTION TESTING AND CALCULATION OF PREDICTED POST-OPERATIVE VALUES 64
4.3.1 Forced vital capacity/ Forced expiratory volume 65
4.3.2 Diffusion capacity 65
4.3.3 Predicted post-operative lung volumes 66
4.4 FEV_1 AND DLCO IN RISK PREDICTION 69
4.4.1 FEV_1 and DLCO to predict post-operative dyspnoea 69
4.4.2 Reconsideration of 40% cut off for risk prediction 77
4.4.3 FEV_1 and DLCO to predict post-operative mortality 82
4.4.4 FEV_1 and DLCO to predict post-operative pulmonary complications 94
4.5 FUNCTIONAL ASSESSMENT FOR RISK STRATIFICATION 103
4.5.1 Sub-maximal exercise testing 103
4.5.2 Stair Climbing 104
4.5.3 Six minute walk testing 109
4.5.4 Shuttle walk testing 112
4.5.5 Low technology testing: conclusion 115
4.5.6 Pre-operative cardiopulmonary exercise testing 115
4.6 OTHER PREDICTIVE MARKERS OF PERI-OPERATIVE MORBIDITY, MORTALITY AND FUNCTIONAL OUTCOME IN LUNG CANCER 122
4.6.1 Reduced pre-operative arterial oxygen content 122
4.6.2 Arterial oxygen desaturation 122
4.6.3 Pre-operative pulmonary artery pressure 123
4.6.4 Pre-operative hypercarbia 124
4.6.5 Other predictive markers of peri-operative morbidity, mortality and functional outcome in lung cancer: conclusion 125

5 THE USE OF BIOMARKERS IN PERI-OPERATIVE RISK PREDICTION 126

5.1 B-TYPE NATRIURETIC PEPTIDE 126
5.1.1 B-Type natriuretic peptide as a biomarker in non-thoracic surgery 127
5.1.2 B-Type natriuretic peptide as a predictor of functional outcome in non-thoracic surgery 131
5.1.3 B-Type natriuretic peptide in thoracic surgery 131
5.1.4 B-Type natriuretic peptide as a predictor in thoracic surgery 132
5.1.5 B-Type natriuretic peptide in pulmonary embolism and pulmonary hypertension 135

6 HYPOTHESES AND AIMS 139

7 GENERIC METHODS 141

7.1 ETHICAL APPROVAL 141
7.2 STUDY DESIGN 141
7.3 STUDY SETTING 141
7.4 PATIENT POPULATION 141
7.5 DEFINITIONS OF EXCLUSION CRITERIA 142
7.6 JUSTIFICATION OF INCLUSION/EXCLUSION CRITERIA 143
7.7 CONSENT 143
7.8 SITE INITIATION VISITS AND TRAINING 143
7.9 ANAESTHETIC PROTOCOL 144
7.10 DATA COLLECTION 144
7.11 BASELINE DEMOGRAPHIC DATA 144
7.12 SELF-REPORT OF DYSPNOEA TOLERANCE AND QUALITY OF LIFE 145
7.13 LABORATORY SAMPLING 146
7.14 INTRA-OPERATIVE CLINICAL DATA 147
7.15 POST-OPERATIVE CLINICAL DATA 148
7.16 SAMPLE SIZE AND POWER 152
7.17 DATA SYNTHESIS AND STATISTICS 152
7.18 PRIMARY OUTCOME ANALYSIS 154
7.19 SECONDARY OUTCOME ANALYSES 156
7.20 MISSING DATA 157

7.21 Calculation and interpretation of Brier scoring 158
7.22 Classification and interpretation of net reclassification improvement 159

8 PATIENT DEMOGRAPHICS & GENERIC RESULTS 160

8.1 Patient demographics and characteristics 160
8.2 Intra-operative data 162
8.3 Peri-operative dyspnoea 164
8.4 Quality of life and disability scoring 166
8.5 Peri-operative pain, anxiety and depression scores 171
8.6 Lost to follow Up 171
8.7 B-Type natriuretic peptide analysis 174

9 PRIMARY OUTCOME ANALYSIS: MODEL DERIVATION 175

9.1 Univariate analysis: demographics. 175
9.2 Model derivation and binary logistic regression analysis – *conventional practice* 180
9.3 Model 1: ppoFEV_1% (<40%) and ppoDLCO% (<40%) 180
9.4 Model 2: Age, gender, ppoFEV_1% (<40%) and ppoDLCO% (<40%) 181
9.5 Model 3: Age, gender, ppoFEV_1% (linear) and ppoDLCO% (linear) 181
9.6 Model 4: Age, gender, ppoFEV_1% (linear), ppoDLCO% (linear) and pre-operative BNP 182
9.7 Model Derivation: Summary of results (Models 1- 4) 183

10 IMPROVEMENT OF RISK PREDICTION MODEL 185

10.1 Model 5: Age, gender, ppoFEV_1% (linear), ppoDLCO% (linear) & next best variable. 185
10.2 Model 6: Age, gender, ppoFEV_1% (linear) & ppoDLCO% (linear), pre-operative BNP and next best variable 186
10.3 Model 7: Forwards regression for all variables 187
10.4 Model 8: Pre-operative BNP and forwards regression for other variable selection 188
10.5 Model Derivation: summary of results (models 5-8) 189
10.6 Comparison of models: summary 192
10.7 Worked example of net reclassification improvement (NRI): model 3 Vs model 7 193

11 EXTERNAL VALIDATION OF RISK PREDICTION MODELS 195

11.1 Patient demographics for validation dataset: comparison 196
11.2 Missing data 197
11.3 External validation 197
11.3.1 Model discrimination 198
11.3.2 Mean calibration 199
11.3.3 Model calibration 199
11.4 Changing sensitivity: model 7 205
11.5 Removal of missing data: model 3 206

12 SECONDARY OUTCOME ANALYSIS 207

12.1 Acute complications and length of hospital stay 207
12.1.1 Incidence of acute complications by primary outcome group 209
12.1.2 Post-operative high dependency, intensive care requirements and hospital mortality: primary outcome group 210
12.2 Intra-operative management by primary outcome group 210
12.3 Peri-operative change in B-Type natriuretic peptide 211
12.4 Association between B-Type natriuretic peptide and acute post-operative complications 213
12.4.1 Association between B-Type natriuretic peptide and new post-operative atrial fibrillation 214
12.4.2 Association between B-Type natriuretic peptide and post-operative cardiopulmonary complications 214
12.4.3 Association between peri-operative B-type natriuretic peptide and length of hospital stay 215
12.5 Post-operative quality of life and disability by primary outcome group 216
12.5.1 European organisation for the research and treatment of cancer quality of life questionnaire 216
12.6 Prediction of quality of life and MRC deterioration 218
12.6.1 Prediction of a deterioration in Medical Research Council score 219

12.6.2 Prediction of quality of life (EQ-5DL index value) 220
12.6.3 Prediction of quality of life (EORTC sumscore) 221
12.7 Concordance between scoring tools used to measure dyspnoea 222

13 SUMMARY OF RESULTS AND FINDINGS 224

14 DISCUSSION OF RESULTS AND FUTURE DIRECTION 229

14.1 Demographics and burden of dyspnoea 229
14.2 Peri-operative B-Type natriuretic peptide to improve risk prediction of post-operative dyspnoea 231
14.3 Model derivation and validation 233
14.4 External validation 240
14.5 Deterioration in MRC score 241
14.6 Peri-operative quality of life and disability 242
14.7 Prediction of quality of life using derived scoring tools 243
14.8 Secondary outcomes - complications 243
14.9 What does this research add to existing literature? 244
14.10 Strengths and limitations 247
14.11 Future directions 249
14.12 Interventions 251

15 APPENDICES 253

15.1 Appendix 1 254
15.2 Appendix 2 255
15.3 Appendix 3 257
15.4 Appendix 4 258
15.5 Appendix 5 265
15.6 Appendix 6 269
15.7 Appendix 7 270
15.8 Appendix 8 277
15.9 Appendix 9 278
15.10 Appendix 10 279

LIST OF REFERENCES 281

List of Tables

Table 1 - Modified Borg Scale (MBS) for dyspnoea 35

Table 2 - Medical Research Council dyspnoea scale. 36

Table 3 - Modified Medical Research Council dyspnoea scale. 37

Table 4 - Variables within *Thoracoscore*. 62

Table 5 - Active cardiac conditions. 63

Table 6 - Revised cardiac risk index 63

Table 7 – Post-hoc analysis by the author of extrapolated data 73

Table 8 - Studies examining FEV_1 and DLCO as predictors of post-operative dyspnoea following lung resection. 79

Table 9 - Studies examining FEV_1 and DLCO as predictors of post-operative mortality following lung resection 85

Table 10 – Studies examining FEV and DLCO as predictors of quality of life following lung resection. 91

Table 11 – Studies examining FEV_1 and DLCO as predictors of post-operative pulmonary complications following lung resection. 99

Table 12 - Risk prediction using pre-operative stair climbing in lung resection surgery – Brunelli et al 107

Table 13 – Post-operative BNP thresholds for incidence of mortality or non-fatal myocardial infarction 30 days after surgery. 130

Table 14- Interpretation of strength of association by correlation coefficients. 153

Table 15 - AUROCC interpretation 154

Table 16 - Pre-operative patient demographics 161

Table 17 - Intra-operative details 163

Table 18 - Intra-operative details 164

Table 19 - Peri-operative MRC and UCSD SOBQ scoring 165

Table 20 - Pre-operative and post-operative World Health Organisation performance status score and pre-operative and post-operative WHO DAS 2.0 score. 167

Table 21 - Peri-operative EQ-5DL score 169

Table 22 - Peri-operative EORTC QoL scores 171

Table 23 - Peri-operative HADS, BPI and VAS scores 171

Table 24 - Lost to follow up: demographic comparison 172

Table 25 - Lost to Follow up: Intra-operative comparison 173

Table 26 – Lost to follow up: post-operative ESTS complication comparison 173

Table 27 - Univariate analysis: demographics 175

Table 28 - Univariate analysis: dyspnoea and quality of life scores 177

Table 29 - Logistic regression using ppoFEV_1%<40% and ppoDLCO%<40% to predict dyspnoea at three months post-operatively (Model 1 derivation) 180

Table 30 - Logistic regression using ppoFEV_1%<40%, ppoDLCO%<40%, age and gender to predict dyspnoea at three months post-operatively (Model 2 derivation) 181

Table 31 - Logistic regression using ppoFEV_1% linear, ppoDLCO% linear, age and gender to predict dyspnoea at three months post-operatively (Model 3 derivation). 182

Table 32 - Logistic regression using ppoFEV_1% linear, ppoDLCO% linear, age, gender and pre-operative BNP to predict dyspnoea at three months post-operatively. (Model 4 derivation) 183

TABLE 33 - MODEL DERIVATION: SUMMARY OF RESULTS (MODELS 1-4) **184**

TABLE 34 - MODEL 1-4 SENSITIVITY AND SPECIFICITY SUMMARY **184**

TABLE 35 - LOGISTIC REGRESSION USING PPOFEV_1% (<40%), PPODLCO% (<40%), AGE, GENDER AND PRE-OPERATIVE EQ-5DL INDEX SCORE TO PREDICT DYSPNOEA AT THREE MONTHS POST-OPERATIVELY. (MODEL 5: DERIVATION). **185**

TABLE 36 - LOGISTIC REGRESSION USING PPOFEV_1% (<40%), PPODLCO% (<40%), AGE, GENDER AND PRE-OPERATIVE EQ-5DL INDEX SCORE TO PREDICT DYSPNOEA AT THREE MONTHS POST-OPERATIVELY. (MODEL 6: DERIVATION). **187**

TABLE 37 - LOGISTIC REGRESSION USING FORWARDS STEPWISE TO SELECT STRONGEST RISK PREDICTION MODEL FROM ALL VARIABLES WITH SIGNIFICANCE AT UNIVARIATE ANALYSIS TO PREDICT DYSPNOEA AT THREE MONTHS POST-OPERATIVELY. (MODEL 7: DERIVATION) **188**

TABLE 38 - PRE-OPERATIVE BNP AND LOGISTIC REGRESSION USING FORWARDS STEPWISE TO SELECT NEXT STRONGEST MODEL FROM ALL VARIABLES TO PREDICT DYSPNOEA AT THREE MONTHS POST-OPERATIVELY (MODEL 8: DERIVATION) **189**

TABLE 39 - MODEL DERIVATION: SUMMARY OF RESULTS (MODEL 5-8) **191**

TABLE 40 - MODEL 5-8 SENSITIVITY AND SPECIFICITY SUMMARY **191**

TABLE 41 - MODEL DERIVATION COMPARISON WITH MODEL 3 ..... **192**

TABLE 42 - NET RECLASSIFICATION IMPROVEMENT (NRI) (MODEL 7 VERSUS MODEL 3) **194**

TABLE 43 - DERIVATION AND VALIDATION MODEL COMPARISON **196**

TABLE 44 - MISSING DATA: SUMMARY **197**

TABLE 45 - EXTERNAL VALIDATION: SUMMARY **198**

TABLE 46 - EXTERNAL VALIDATION: MODEL DISCRIMINATION **198**

TABLE 47 - MEAN CALIBRATION **199**

TABLE 48 - CHANGING SENSITIVITY: *'HYPOTHESIS FREE'* MODEL 7 ..... **206**

TABLE 49 - MODEL 3: MISSING DATA REMOVED. **206**

TABLE 50 - POST-OPERATIVE COMPLICATIONS (NON-ESTS COMPLICATIONS) **207**

TABLE 51 - POST-OPERATIVE COMPLICATIONS: EUROPEAN SOCIETY OF THORACIC SURGEONS (ESTS) DEFINITIONS **208**

TABLE 52 - POST-OPERATIVE HIGH DEPENDENCY, INTENSIVE CARE REQUIREMENTS AND HOSPITAL MORTALITY **208**

TABLE 53 - UNIVARIATE ANALYSIS: POST-OPERATIVE COMPLICATIONS **209**

TABLE 54 – INCIDENCE OF ESTS DEFINED POST-OPERATIVE COMPLICATIONS BY PRIMARY OUTCOME GROUP **209**

TABLE 55 - POST-OPERATIVE HIGH DEPENDENCY, INTENSIVE CARE REQUIREMENTS AND HOSPITAL MORTALITY: PRIMARY OUTCOME GROUP **210**

TABLE 56 - INTRA-OPERATIVE MANAGEMENT BY PRIMARY OUTCOME **211**

TABLE 57 – PERI-OPERATIVE B-TYPE NATRIURETIC PEPTIDE COMPARISON BY PRIMARY OUTCOME GROUP **212**

TABLE 58 - MRC DETERIORATION: SUMMARY OF RESULTS **219**

TABLE 59 - MRC DETERIORATION: SENSITIVITY AND SPECIFICITY SUMMARY **220**

TABLE 60 - QUALITY OF LIFE PREDICTION: MCID IN EQ-5DL QUESTIONNAIRE **221**

TABLE 61 – QUALITY OF LIFE PREDICTION: MCID IN EORTC *SUMSCORE*. **221**

TABLE 62 – SUMMARY OF STUDIES EXAMINING STAIR CLIMBING AS A PREDICTOR IN LUNG CANCER SURGERY ASSESSMENT **254**

TABLE 63 – SUMMARY OF STUDIES EXAMINING SIX-MINUTE WALK TESTING AS A PREDICTOR IN LUNG CANCER SURGERY ASSESSMENT **255**

TABLE 64 – SUMMARY OF STUDIES EXAMINING SHUTTLE WALK TESTING AS A PREDICTOR IN LUNG CANCER SURGERY ASSESSMENT **257**

TABLE 65 – SUMMARY OF STUDIES EXAMINING PRE-OPERATIVE CPET TESTING AS A PREDICTOR IN LUNG CANCER SURGERY ASSESSMENT **258**

List of Figures

FIGURE 1 - SURGICAL LUNG RESECTION RATES IN THE UK FOR PRIMARY LUNG CANCER (1980-2015). 23

FIGURE 2 - PROPORTION OF PATIENTS COMPLAINING OF (C/O) SYMPTOMS. 28

FIGURE 3 – RELATIONSHIP BETWEEN CHANGE IN EXERCISE CAPACITY AND CHANGE IN FEV_1%. 45

FIGURE 4 – RELATIONSHIP BETWEEN ALTERATIONS IN MAXIMAL OXYGEN UPTAKE (VO_2-MAX%) AND ALTERATIONS IN FORCED VITAL CAPACITY (FVC% PRE-OPERATIVELY) FOLLOWING LUNG RESECTION, 46

FIGURE 5 – TAKEN FROM MCCALL ET AL.120 LEFT AND RIGHT VENTRICULAR EJECTION FRACTION OVER TIME (%). 50

FIGURE 6 – PERI-OPERATIVE PULMONARY VASCULAR RESISTANCE INDEX. 52

FIGURE 7 – PERI-OPERATIVE RIGHT VENTRICULAR EJECTION FRACTION. 52

FIGURE 8 – FLOW CHART OF LITERATURE REVIEW. 58

FIGURE 9 – RISK ASSESSMENT PATHWAY FOR LUNG RESECTION SURGERY, AS PER BTS GUIDELINES. 60

FIGURE 10 - RISK ASSESSMENT FOR POST-OPERATIVE DYSPNOEA. 64

FIGURE 11 - BRONCHOPULMONARY SEGMENTS. 67

FIGURE 12 – PREDICTED POST-OPERATIVE FEV1 IN PATIENTS EXPERIENCING AND NOT EXPERIENCING COMPLICATIONS. 72

FIGURE 13 - PREDICTED POST-OPERATIVE DLCO% IN PATIENTS EXPERIENCING AND NOT EXPERIENCING COMPLICATIONS 72

FIGURE 14 - COMPARISON OF DYSPNOEA SCORES IN LOW AND HIGH DLCO% PATIENT GROUPS. 75

FIGURE 15 – CLASS OF DYSPNOEA VERSUS PULMONARY FUNCTION. 76

FIGURE 16- QUALITY OF LIFE SIX MONTHS AFTER LUNG CANCER SURGERY BASED ON PRE-OPERATIVE DLCO CATEGORY (<45%, 45-75%, >75%) 90

FIGURE 17 - PREDICTED PROBABILITY OF DEVELOPING POST-OPERATIVE PULMONARY COMPLICATIONS (PPCS) BY THE POINT SCORE. 97

FIGURE 18 - CHARACTERISTICS OF PATIENTS FOR ASSIGNMENT INTO RISK GROUPS 120

FIGURE 19 - ASSOCIATION BETWEEN $RVEF_{POD2}$ AND BNP_{POD2}. 132

FIGURE 20 - COMPARISON OF BNP LEVELS BETWEEN PATIENTS WITH DETERIORATED AND UNCHANGED POST-OPERATIVE FUNCTIONAL CAPACITY. 134

FIGURE 21 – BNP LEVELS IN PATIENTS WITH RV DYSFUNCTION IN 50 PATIENTS WITH ACUTE PULMONARY EMBOLISM. 136

FIGURE 22 - KAPLIEN-MEIER SURVIVAL CURVES ACCORDING TO MEDIAN VALUE OF BASELINE AND FOLLOW-UP BNP IN PATIENTS WITH PH. 138

FIGURE 23 - STUDY RECRUITMENT CONSORT DIAGRAM. 160

FIGURE 24 - PERI-OPERATIVE MRC SCORE. 165

FIGURE 25 - PERI-OPERATIVE UCSD-SOBQ SCORE. 166

FIGURE 26 - PRE-OPERATIVE VS POST-OPERATIVE EQ-5DL VISUAL ANALOGUE SCALE SCORE. 168

FIGURE 27 - PRE-OPERATIVE AND POST-OPERATIVE EORTC *SUMSCORE*. 170

FIGURE 28 – B-TYPE NATRIURETIC PEPTIDE CHANGES OVER TIME. 174

FIGURE 29 - MAGNITUDE OF EQ-5DL VAS CHANGE VS PRIMARY OUTCOME. 179

FIGURE 30 - MAGNITUDE OF EQ-5DL INDEX CHANGE VS PRIMARY OUTCOME. 179

FIGURE 31 - RECEIVER OPERATOR CHARACTERISTIC CURVES: MODEL 3 AND MODEL 4 TO PREDICT MRC SCORE >2 AT THREE MONTHS POST-OPERATIVELY. 183

FIGURE 32 - RECEIVER OPERATOR CHARACTERISTIC CURVES: MODEL 5 AND MODEL 6 TO PREDICT MRC SCORE >2 AT THREE MONTHS POST-OPERATIVELY. 186

FIGURE 33 - RECEIVER OPERATOR CHARACTERISTIC CURVES: MODEL 7 AND MODEL 8 TO PREDICT MRC SCORE >2 AT THREE MONTHS POST-OPERATIVELY. 189

Figure 34 - External validation population CONSORT diagram. 195
Figure 35 – *'Conventional'* Model 3 calibration curve in external dataset. 200
Figure 36 – *'Conventional'* Model 3 calibration curve in external dataset. 201
Figure 37 – *'Next best variable'* Model 5 calibration curve in external dataset. 202
Figure 38 - *'Next best variable'* Model 5 calibration curve in external dataset 203
Figure 39 – *'Hypothesis free'* Model 7 calibration curve in external dataset 204
Figure 40 - *'Hypothesis free'* Model 7 calibration curve in external dataset 205
Figure 41 - Peak post-operative BNP (pg/ml) and primary outcome. 212
Figure 42 - Peak Post-operative BNP and >1 post-operative complication 213
Figure 43 - Pre-operative BNP and new atrial fibrillation 214
Figure 44 - Post-operative peak BNP (pg/ml) and pulmonary complications. 215
Figure 45 – Association between post-operative peak BNP and duration of hospital stay. 216
Figure 46 - Pre-operative EORTC sumscore by 3-month MRC score (primary outcome) group. 217
Figure 47 - Correlation between pre-operative and post-operative EORTC sumscore. 218
Figure 48 - Concordance between scoring tools used to measure dyspnoea following lung resection for cancer. 223

Grants

- National Institute for Academic Anaesthesia Research Grant Round 1, £26,500. Funded by Association of Anaesthetists of Great Britain and Ireland/Anaesthesia. April 2018.

Published Abstracts

- *BNP for prediction of outcome following lung resection surgery (PROFILES); an interim analysis*. Journal of Cardiothoracic and Vascular Anaesthesia Oct 2020 (Vol 34)

- *Quantifying breathlessness following lung resection- concordance between scoring tools*. Journal of Cardiothoracic and Vascular Anaesthesia Oct 2020 (Vol 34)

- *PROFILES; an interim analysis*. Scottish Society of Anaesthetists Annals 2021

- *PROFILES- Concordance between scoring tool*. Scottish Society of Anaesthetists Annals 2021

- *Prediction of dyspnoea following lung resection surgery: post-hoc analysis of PROFILES study*. Journal of Cardiothoracic and Vascular Anaesthesia Oct 2021 (Vol 35)

Oral Presentations

- *bnP for pRediction of Outcome Following Lung rEsection Surgery (PROFILES); an interim analysis*. St Mungo University of Glasgow research symposium. March 2020

- *Validation of days Alive and out of hospital as a patient centred outcome for lung resection surgery*. St Mungo University of Glasgow research symposium. McCreadie S, Lafferty B, Shelley B. March 2020

- *BNP for prediction of outcome following lung resection surgery (PROFILES); an interim analysis*. European Association Cardiothoracic Anaesthesia 2020

- *Quantifying breathlessness following lung resection – concordance between scoring tools*. European Association Cardiothoracic Anaesthesia 2020

- *PROFILES; an interim analysis*. Scottish Society of Anaesthetists Annual Spring Meeting 2021. Best oral presentation – Winner of Campbell Quaich

- *PROFILES- Concordance between scoring tools*. Shortlisted best abstract. Scottish Society of Anaesthetists 2021 Spring Meeting 2021

- *Prediction of dyspnoea following lung resection surgery: post-hoc analysis of PROFILES study*. European Association Cardiovascular and Thoracic anaesthesia and Intensive Care 2021

- *Prediction of Dyspnoea following lung resection surgery*. Shortlisted best abstract for oral presentation. Association for cardiothoracic anaesthesia and critical care annual conference, Glasgow 2021

- *Prediction of Dyspnoea following lung resection surgery*. Shortlisted best abstract for oral presentation. Glasgow Anaesthetic Research Club Annual meeting, Glasgow 2021. Winner best oral presentation

- *Quantifying breathlessness following lung resection – concordance between scoring tools*. Shortlisted best abstract for oral presentation. Glasgow Anaesthetic Research Club Annual meeting, Glasgow 2021

Definitions/Abbreviations

6MWT	Six-minute walk test
ABG	Arterial blood gas
ACC	American college of cardiology
ACCP	American college of chest physicians
AF	Atrial fibrillation
AHA	American heart association
AIMS	Anaesthetic intra-operative management system
ANOVA	Analysis of variance
ARDS	Acute respiratory distress syndrome
AKI	Acute kidney injury
ASA	American society of anaesthesiologists
ATS	American thoracic society
AUROCC	Area under the receiver operating characteristic curve
BA	Bland Altman
BMI	Body mass index
BNP	B-Type natriuretic peptide
BPI	Brief pain inventory score
BTS	British Thoracic Society
CI	Confidence interval
CIS	Clinical information system
COPD	Chronic obstructive pulmonary disease
CPET	Cardiopulmonary exercise testing
CTPA	CT pulmonary artery
DLCO%	Diffusion capacity of carbon monoxide(%)
DNA	Deoxyribonucleic acid
DVT	Deep vein thrombosis
ECG	Electrocardiogram
EDTA	Ethylenediaminetetraacetic acid
EORTC	European Organisation for Research and Treatment of Cancer quality of life questionnaire
ERS	European Respiratory Society
ESOS	European Society objective score
EQ-5DL	EuroQol 5-dimensional quality of life questionnaire
FEV_1	Forced expiratory volume in 1 second
FEV1%	Percentage predicted forced expiratory volume in 1 second
FVC	Forced vital capacity
GORD	Gastro-oesophageal reflux disease
GMC	General Medical Council
HADS	Hospital anxiety and depression score
ICHOM	International Consortium for Health Outcomes Measurement
ICU	Intensive care unit

IQR	Interquartile range
LOS	Length of surgery
LOA	Length of anaesthesia
LV	Left ventricle
MBS	Modified Borg scale
MCID	Minimum clinically important difference
MDT	Multi-disciplinary team
MEP	Maximal expiratory pressures
MET	Metabolic equivalents
MINS	Myocardial injury following non-cardiac surgery
MIP	Maximal inspiratory pressure
MITS	Minimally invasive thoracic surgery
MRC	Medical Research Council score
NEQAS	National external quality assessment
NICE	National Institute for Health and Care Excellence
NP	Natriuretic peptides
NPR-C	Natriuretic peptide clearance receptors
NPV	Negative predictive value
NRI	Net reclassification improvement index
NRI*e*	Net reclassification improvement (proportion of patients assigned higher risk category
NRI*ne*	Net reclassification improvement (proportion of patients assigned lower risk category)
NRS	Numerical rating scales
NSCLC	Non-small cell lung cancer
OLV	One lung ventilation
OR	Odds ratio
PAP	Pulmonary artery pressure
PCA	Patient controlled analgesia
PE	Pulmonary embolism
PFT	Pulmonary function test
PH	Pulmonary hypertension
POAF	Post-operative atrial fibrillation
POC	Post-operative complication
POD	Post-operative day
PPC	Post-operative pulmonary complications
PPO	Predicted post-operative
PPV	Positive predicted value
PROM	Patient reported outcome measures
PS	Performance status
PVR	Pulmonary vascular resistance
QoL	Quality of life
RATS	Robotically assisted thoracoscopic surgery
SC	Stair climbing

SCLC	Small cell lung cancer
SF-36	Short form 36 quality of life questionnaire
SS	Skill score
STS	Society of thoracic surgeons
SWT	Shuttle walk testing
TIVA	Total intravenous anaesthesia
UCSD-SOBQ	University of California and Sand Diego shortness of breath questionnaire
VAS	Visual analogue scales
VATS	Video assisted thoracoscopic surgery
Ve/VCO_2	Minute ventilation/carbon dioxide ratio (ventilatory efficiency)
VO_2 Max	Maximal oxygen uptake
V/Q	Ventilation/Perfusion ratio
Wmax	Maximal exercise intensity
WHO	World Health Organisation
WHO-DAS	World Health Organisation disability schedule
WHO-PS	World Health Organisation performance status

1 Lung Cancer

1.1 Lung cancer: Introduction

1.1.1 Lung cancer mortality

Lung cancer accounts for the largest proportion of cancer deaths in the UK with 35,300 deaths per year.[1] With more than 46,000 new cases each year (47.4 per 100,000 population in the UK), lung cancer is also the second most prevalent cancer type – in both males and females.[2] Outside the UK, lung cancer accounted for 20% of all cancer deaths in Europe in 2016 and 27% of all cancer deaths in the USA in 2015.[3]

In males, the incidence of lung cancer has been decreasing over the past decade, secondary to reduced smoking rate. Conversely, in females, the incidence of lung cancer is increasing due to a simultaneous upward trend in smoking rates.[4] This increase in females is faster than the decline in males, meaning an overall increase in total cases of 3% in the last decade. The incidence of lung cancer is highest in areas of deprivation, where a three-fold increase can be observed as a result of increased smoking rates.[2]

The prognosis of patients with lung cancer is very poor; 5 years after diagnosis, only 1 in 10 are still alive.[1, 2] This low survival can in part be explained by late presentation and often advanced stage at diagnosis; only 18% of people are able to be offered curative surgery.[1] There are usually few signs or symptoms in the early stages of the disease process, but patients eventually develop a combination of; persistent cough, haemoptysis, dyspnoea, lethargy and weight loss.

In a drive to improve mortality rates, the National Institute for Health and Care Excellence (NICE) have published a quality statement with numerous targets, these include; increasing public awareness, ensuring adults with suspected or confirmed lung cancer receive evidence based support to stop smoking, increased access to lung cancer clinical nurse specialists and appropriate early investigations to accurately determine diagnosis and stage.[2]

1.1.2 Risk factors for lung cancer

One in 13 UK males and 1 in 15 UK females will be diagnosed with lung cancer in their lifetime yet it is thought 79% of lung cancer in the UK is preventable. Like most cancers, risk of developing lung cancer is dependent upon many factors including age, genetics and lifestyle factors such as smoking.[5] Smoking is the leading cause of lung cancer, resulting in 7 of 10 lung cancer cases in the UK.[6] Other causes include; ionising radiation (5%), workplace exposure to organic dust (13%) and air pollution (8%).[1,6] Age contributes to the risk of developing lung cancer, reflecting cell DNA damage over time. Lung cancer risk is 82% higher in people whose siblings have been affected by lung cancer and 25-37% higher in people whose parents have had the disease. This association is independent of smoking highlighting the importance of genetic factors.[7]

1.2 Classification of lung cancer

Cancer that originates within the lung is called primary lung cancer, whereas metastases from another organ system is called secondary lung cancer. There are two main forms of primary lung cancer, with classification based on the microscopic appearance of tumour cells; small cell and non-small cell lung cancer (SCLC and NSCLC respectively). SCLC is less common than NSCLC, more aggressive then NSCLC and less amenable to surgical resection. NSCLC is the most prevalent type of lung cancer accounting for >85% of cases. NSCLC can be one of three types: squamous cell carcinoma, adenocarcinoma or large cell carcinoma.

1.2.1 Lung cancer management and surgery

Management of lung cancer can be broadly classified into surgical and non-surgical and ranges from palliation to curative surgery, with or without chemo-radiotherapy. Other interventions include smoking cessation and targeted immunotherapy agents, dependent on cancer classification. Often, the type of lung cancer diagnosed and its resectability are evaluated alongside the general health of the patient to determine which management will be offered. Risk stratification of general health includes examination of cardiac risk factors and tests of pulmonary function. Patient preference should also play a major role in treatment planning, some patients deciding not have certain types of management including potentially curative surgery.[8] Surgical treatment remains

the best curative option for early stage lung cancer, despite advances in non-surgical management.[9] Reported survival following surgery with curative intent for early stage lung cancer ranges widely from 45-80% at five years.[10]

Types of surgery/resection performed range from pneumonectomy (the entire lung being removed) to more conservative, lung sparing, options such as; wedge resection (where the tumour and a small amount of surrounding tissue is resected), segmentectomy (where an anatomical *segment* is removed*)* and lobectomy (where an anatomical lobe is removed).[11] Around 7500 lung resections took place in 2015 in the UK and these numbers have been increasing each year: having doubled since 2002, when only 3000 lung resections for primary lung cancer took place (Figure 1).[12] Despite this, resection numbers in the UK are low compared to other countries with similar healthcare systems;[8] the reasons for this are complex and multifactorial.

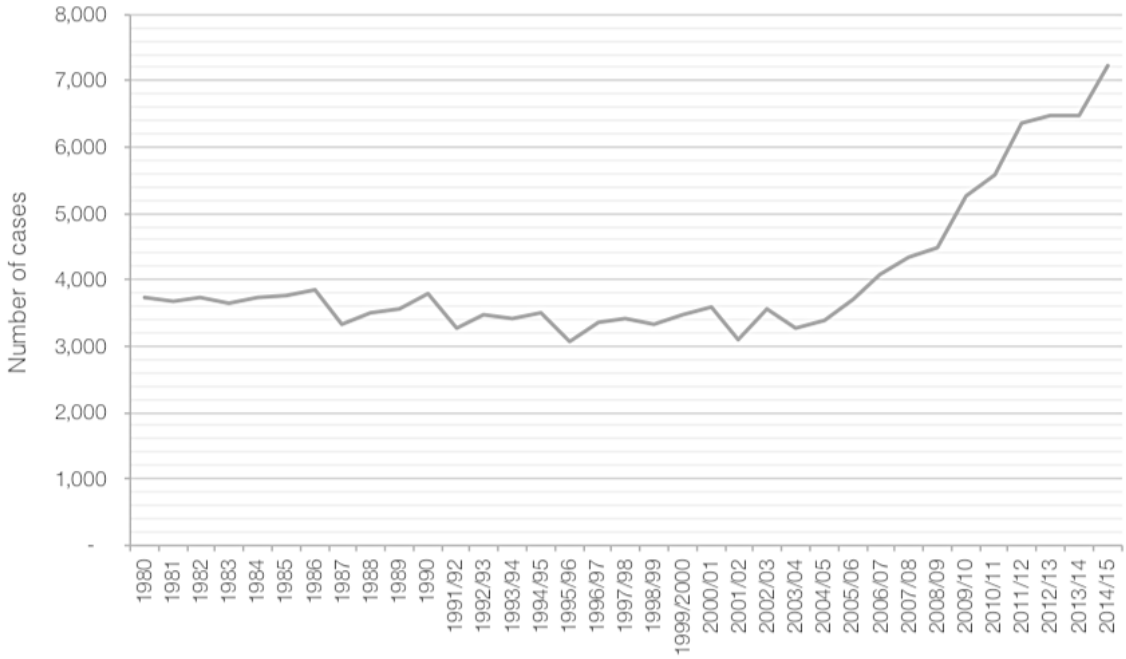

Figure 1 - Surgical lung resection rates in the UK for primary lung cancer (1980-2015).
Taken from *The UK Cardiothoracic Surgery Workforce report 2019.12*

There has been a drive by the department of health within the UK to increase resection rates to the levels of other developed nations and improve care for patients with lung cancer.[13] These interventions include; increasing surgeon numbers, restricting operations to high-volume centres, broadening the attitude

of referring clinicians to consider surgery as an equal option when the possible outcomes appear equal.[13] NICE advocate offering potentially curative treatment to patients if they accept the risks of post-operative dyspnoea and associated complications. Furthermore, surgical resection is increasingly being offered to older patients if they are prepared to accept the risks of surgery.

The development of minimally invasive surgical techniques has resulted in the avoidance of open surgery where possible. From being a small proportion of overall activity in the 1980's, video assisted thoracoscopic surgery (VATS) for lung cancer has increased considerably, now forming half of all UK cases. As a natural evolution to the VATS technique, a small number of UK centres are now performing robotically assisted thoracoscopic surgery (RATS).[8]

1.2.2 Post-operative mortality and morbidity

Patients presenting for lung resection surgery often have multiple co-morbidities, including cardiovascular and pulmonary disease and other associated medical conditions.[14] In addition, 90% of patients undergoing lung resection are also smokers, which is another risk factor for post-operative pulmonary complications. Consequently, lung resection patients have increased risk of post-operative morbidity and mortality. Despite improvements in perioperative care, mortality and morbidity associated with lung resection remain high.[15]

While most patients successfully having lung resection surgery are discharged from hospital, in hospital mortality remains approximately 1.7% in the UK.[15] A range of mortality rates have been reported for lobectomy with some authors observing a 90 day mortality as low as 2.1%. Within the UK, when looking at *all* lung resection patients, it has been observed in 10,991 patients who had surgery between 2004-2010 a 3% mortality rate within 30 days and 5.9% within 90 days of surgery.[16] Age is associated with early post-operative death. Other significant associations are performance status (PS), residual lung function, cancer stage and procedure type.[16]

The American College of Chest Physicians(ACCP) guidelines estimate the mortality risk to be 4% for lobectomy and 9% for pneumonectomy,[17] with the British Thoracic

Society (BTS) quoting similar figures.[18] In the elderly population (>70 years) these figures are higher as increasing age is associated with more co-morbidity.[6]

The Society of Thoracic Surgeons report the overall pulmonary complication risk within 30 days of lung resection to be around 13% and include complications such as; pneumonia, acute respiratory distress syndrome (ARDS), prolonged ventilatory support, atelectasis requiring bronchoscopy, pulmonary embolism, pulmonary oedema and reintubation requiring ventilation. The estimated risk of cardiac death or non-fatal myocardial infarction within the first 30 days of surgery is around 2-3%.[19] Historically, the incidence of other cardiac complications such as arrhythmias (atrial and ventricular) is reported as 15-25%, dependent on extent of lung resection.[20] Other recognised post-operative complications are stroke and acute kidney injury (AKI). The reported incidence of AKI is dependent on definition and varies between 5-10%.[21, 22]

1.2.3 Quality of life and patient reported outcome measurements

The World Health Organisation (WHO) define QoL as an:

> *'Individuals perception of their position in life in the context of their culture and value systems in which they live and in relation to their goals, expectations, standards and concerns.'*
>
> *WHO 1995* [23]

Lung cancer is associated with increased disruption to quality of life compared to other chronic disease and cancers.[18] It has been reported that those who go on to have surgical treatment of their lung cancer have a significant decrease in quality of life (QoL).[2, 24-26] Conventional parameters used to assess post-operative cardiorespiratory function do not correlate with quality of life reported by patients.[25] Patients following lung resection experience a shorter life expectancy and reduced QoL when compared to age-matched peers.[27]

The concept of QoL is subjective with individualised levels of satisfaction and well-being.[28] It is recognised that surgical management of lung cancer has significant impact on patient's QoL.[18] The reporting and use of QoL measurement in thoracic

surgery has improved but its use in clinical practice remains unclear and its value underestimated.[24]

> *'Lung function tests and exercise tests cannot be taken as sole surrogates for quality of life evaluation. A quality of life instrument should always be used.'*
>
> *BTS 2010* [18]

Survival has been traditionally used as an outcome measure. However, many patients do not regard immediate post-operative complications (including early mortality) as a reason not to have surgery: the prospect of physical disability and the risk of an impaired QoL after surgery can be a more important factor to aid decision making. For many, survival with limited QoL would be unnaceptable.[25, 29,30] Therefore, major international guidelines advocate *long-term function* should be considered before a decision to proceed with surgery is made. The long term goal of surgery should be to improve survival, with minimal decrease in QoL.[31]

Interest in functional assessment and QoL in lung resection patients started in the mid 1990's, with a recognition that the impact of surgery on these markers was not fully understood.[32] The potential benefit of surgery then started to be weighed against residual post-operative QoL, which until this point was difficult with such little data concerning patient reported QoL. Even now, this still represents a challenge to physicians and surgeons consenting patients for surgery.[18]

Improvements in diagnosis and management in recent years have changed the perspective of life expectancy and QoL.[24, 33] As life expectancy following lung resection increases, the ability to resume a normal lifestyle at conclusion of treatment becomes increasingly important.[30] However there are conflicting reports on the impact of surgery on QoL during the follow-up period.[34, 35]

1.2.4 Patient reported outcome measures

Patient reported outcome measures (PROMs) are defined as a health outcome, directly reported by the patient; these often incorporate a QoL assessment. This is in contrast to an outcome reported by someone else, usually a physician or nurse

reported outcome. Despite growing interest, routine collection of QoL PROMS is poorly performed.[24] No guidelines have been developed in the lung cancer setting about the best time to evaluate QoL after surgery.[24] Increased consensus is needed to ensure improved collaboration and standardisation between centres in collecting similar PROM's and QoL data. Scores to measure these outcomes such as EQ-5DL and EORTC QLQ-30 have been developed over the past 30 years in an attempt to improve standardisation, but few are routinely used in the lung cancer resection population.[36]

PROM collection has been demonstrated to enhance communication between patients and care providers.[37] In turn, this improves patient involvement in decision making. The International consortium for Health Outcomes Measurement (ICHOM) has identified a core set of outcomes and variables that can be collected for lung cancer patients internationally in routine clinical practice including; survival, complications within 6 months of surgery and patient reported QoL.[29] These are not specifically recommended for lung cancer resection patients. These core outcomes reflect the opinions of experts and patients' representatives globally and advocate the use of QoL instruments such as EORTC QLQ-C30 and EORTC QLQ-LC13 (Section 2.2) essential in the process of clinical care. In 2017, the UK Society of Thoracic Surgeons (STS) incorporated PROMs into its database for the first time, recognising this is a critical gap.[38, 39]

1.2.5 Dyspnoea following lung resection

Dyspnoea is a debilitating symptom following lung resection affecting QoL, functional status and psychological health/[40] Up to 30-50% of patients reporting long term *disabling* shortness of breath following surgery.[41] Dyspnoea is defined by the European Respiratory Society (ERS) and the American Thoracic Society (ATS) as:

> '*A subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity*'
>
> *ERS and ATS 1999* [42]

Dyspnoea is one of the most commonly reported complaints post-operatively in patients undergoing lung resection. Zieren et al reported dyspnoea as the most frequent and severe single complaints in 52 patients undergoing lung resection.[43] Twenty-one percent suffered from dyspnoea at rest at 12 months. Furthermore, patients following lobectomy suffered less frequently from dyspnoea than those undergoing pneumonectomy suggesting that the more lung parenchyma is lost the more at risk the patient is of long term post-operative dyspnoea. Over 20% represents a significant proportion of patients who may experience post-operative dyspnoea.

In 94 patients undergoing thoracotomy, Sarna et al observed the most common symptoms at 4 months post-operatively were dyspnoea (49%) and fatigue (57%). In many patients, these symptoms persisted for longer than 4 months, long into the post-operative recovery period.[44] In even earlier work, the same author observed dyspnoea in 142 patients undergoing lung resection surgery (Figure 2). Figure 2 displays the frequency of breathlessness following lung resection in this cohort with over 50% complaining of SOB when hurrying and 11% so disabled that they were unable to leave their house.

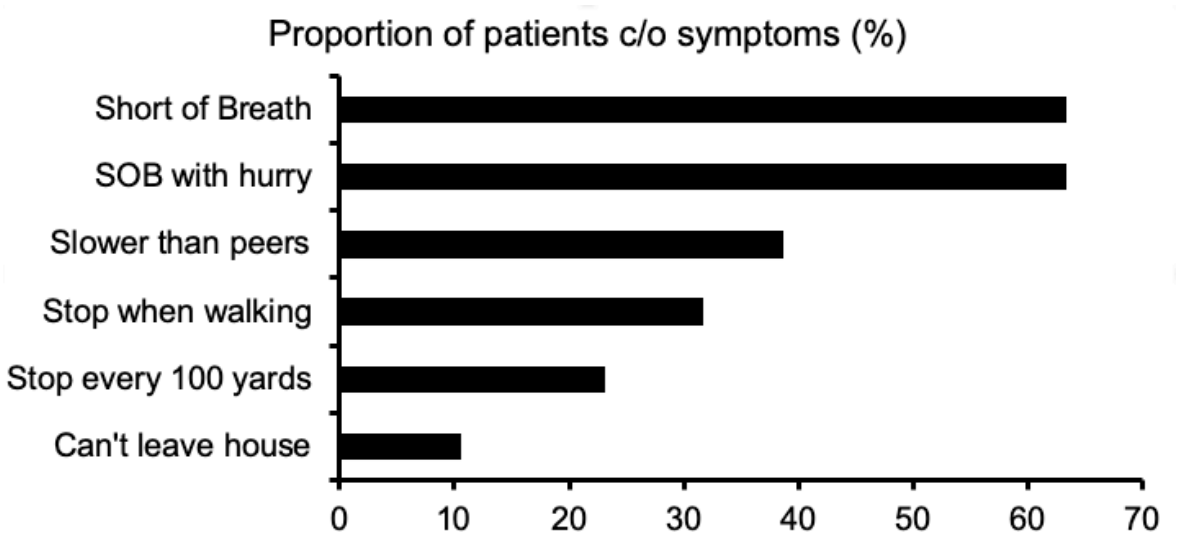

Figure 2 - Proportion of patients complaining of (c/o) symptoms.
Redrawn and adapted from Sarna et al 2004.[52] (n=142)

Similarly, in 117 patients, Dales et al observed moderate to severe dyspnoea among 31% of patients within 3 months of lung cancer resection for lung cancer.[45] Moderate to severe dyspnoea reported in 14% pre-operatively increased to 34% post-operatively, ($p<0.005$). Further studies examining long term post-operative

dyspnoea include; Feinstein et al [41] observed that dyspnoea is common 1 to 6 years after lung cancer in 342 patients undergoing resection and is associated with pre-operative dyspnoea, reduced diffusing capacity, depression and lack of physical activity. Balduyck et al[46] in a cohort of 100 patients undergoing lung resection observed a decrease in dyspnoea scale scores extending to the 12-month follow up point.

Myrdal et al[47] studied quality of life following lung resection including dyspnoea scores and observed a high incidence of breathlessness following lung resection. In 112 patients undergoing open surgery for lung cancer, the author observed breathlessness extending to 48 months following surgery. Interestingly, at the same cardiothoracic centre and in the same study the author compared these lung cancer patients to patients undergoing coronary artery bypass grafting (CABG) in the same timeframe. Breathlessness on physical exertion was more pronounced in patients with lung cancer than in the CABG patients, ($p<0.001$). Studies of CABG patients have shown *improvements* in physical function and breathlessness as early as 3 months post-procedure.[48] This illustrates the high levels of post-operative dyspnoea in the lung resection population not observed in other high risk populations.

Finally, in a pilot study of 25 patients focusing on functional capacity following lung resection, Young et al [49] (our research group) reported a difference in the distribution of dyspnoea scores over time; patients reported functional limitation and increased breathlessness using the Medical Research Council (MRC) dyspnoea scale at all post-operative timepoints, ($p=0.03$). Dyspnoea was measured at baseline, 2 months post-operatively and 1 year post-operatively. Ten patients had a deterioration in self-reported dyspnoea (40%).

This data serves to illustrate that breathlessness is prevalent following lung resection for cancer, having detrimental effects on post-operative quality of life and functional capacity. The reported post-operative dyspnoea extends beyond the immediate post-operative period, long into the recovery.

Not all patients end up with long term dyspnoea and prediction of those who will develop this post-operative disability is challenging. Dyspnoea is historically attributed to loss of lung parenchyma with reduced alveolar volume, however it

is increasingly becoming recognised that the pathophysiology of post-operative dyspnoea is complex, multifactorial and likely to involve cardiovascular mechanisms (Section 3).[50] Prediction of dyspnoea is conventionally performed using predicted post-operative forced expiratory volume in one second (FEV_1%) and predicted post-operative diffusion capacity of carbon monoxide (DLCO%) (chapter 4.2.3). Current guidelines acknowledge that prediction of disabling post-operative dyspnoea is important, difficult and could be improved.[18, 51, 52]

1.2.6 Shared decision making

Some patients would accept the risk of dyspnoea if they could be offered curative treatment. Conversely, many more patients survive the operation but are left with long term physical disability and reduced QoL which is intolerable.[53] In recent years, perspective has shifted from a more authoritarian patient pathway, with decision making dominated by the surgical team, to a scenario where the patient is more involved in the decision-making process.[51] This includes assessing the patients willingness to undertake surgery even if the risks, of dyspnoea for example, seem particularly high.[54] This is important as some patients may be ready to accept the short-term risk of immediate cardiopulmonary complications but not long-term risks of significant functional debility.[30] Like all surgery, the survival benefit must be weighed against the potential for a significant reduction in quality of life.[25]

Discussion of peri-operative risk should be based around *shared decision making*; patients should be involved in decisions about treatment and the specific risks they are prepared to accept should be explored. International societies, including NICE and BTS are uncertain how to include *shared decision making* into surgical decision-making algorithms. Although, not all patients wish to be involved with complex decision-making processes.[55]

In 2008, the General Medical Council (GMC) introduced a document advocating *shared decision making* and empowerment of the patient beyond the clinicians' recommendations.[56] The BTS are the first to include patient acceptance of risk as an integral part of risk assessment in the lung resection population.[18] This makes the role of QoL measurement and prediction of dyspnoea even more crucial to enable patients to have complete information about residual function and

outcome. Despite this, further investigation is required to establish how interventions affect QoL and ascertain the best method to measure this outcome.

1.2.7 Conclusion

Lung resection for cancer is common and with an ageing population a further increase in cases should be expected. Dyspnoea following lung resection is also common and its effect on post-operative QoL profound. The mechanisms driving post-operative dyspnoea has not been fully explained, but are likely multifactorial, including cardiovascular factors. Future work should attempt to fully understand these complex mechanisms and predict who is at risk of long-term disabling dyspnoea. If we could improve prediction of dyspnoea following lung resection this would not only enable improved shared decision making for surgery, but also facilitate targeted intervention and entry in trials aiming to ameliorate post-operative breathlessness.

2 Measuring dyspnoea and quality of life

This chapter explores some common patient self-reported questionnaires used to quantify dyspnoea and quality of life following lung resection surgery. The concept and importance of 'minimum clinically important differences' (MCID) when using self-reported questionnaires in clinical practice is also introduced.

2.1 Scoring tools to assess dyspnoea

There are numerous scoring tools to measure dyspnoea. Dyspnoea can affect many dimensions of a patient's life, reducing activity and causing distress and discomfort. Dyspnoea is subjective sensation, with patients experiencing different sensations with various intensity when attempting to describe and quantify. The American Thoracic Society (ATS) reiterates,

> *'Dyspnoea is symptom which can only be described and interpreted by the patient and therefore any assessment should be patient reported.'*
>
> *Parshall et al 2012* [57]

There are a variety of definitions of dyspnoea, from two words such as 'laboured breathing' up to whole paragraphs, but importantly no consensus exists. The ATS define dyspnoea as

> *'The subjective experience of breathing discomfort that is comprised of qualitatively distinct sensations that vary in intensity.'*
>
> *Parshall et al 2012* [57]

The use of tools to measure dyspnoea helps standardise the way in which this symptom is described. The two major reasons for measuring dyspnoea are to discriminate symptom severity between individuals and evaluate changes over time for a given individual. Psychophysical methods (relationship between a stimulus and a response) and clinical scales are used to asses dyspnoea which is a subjective sensation.[58]

Two types of uni-dimensional tool are used to measure dyspnoea; visual analogue scales (VAS) or numerical rating scales (NRS). These measure dyspnoea in general or on exercise and are often used to describe breathlessness in exercise tolerance tests. They are self-administered and quick to complete. Discussion in the chapter will be limited to the Visual Analogue scale (VAS), Modified Research Council (MRC) scale and the University of California and San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) (both are NRS) to measure dyspnoea.

2.1.1 Minimum Clinically Important Difference

Evaluation of health outcomes for patients has become increasingly important; subsequently the usage of self-reported questionnaires has increased. Interpretation of these outcome measures is challenging given the variety of questionnaires and scoring methods available. Unless the user is very familiar with a particular questionnaire it can be confusing to interpret meaningful change. The term minimal clinically important difference (MCID) was first described by Jaeschke et al[59] when they proposed *statistically significant* changes can occur using measurement tools which often do not have *clinical significance*. The MCID has thus been defined as;

> *'The smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patients management.'*
>
> *Jaeschke et al 2008 [59]*

This value may be larger than a statistically significant difference or change. Standardisation of patient reported outcome measures has improved the ability to determine care pathways that provide better results.[60] Some definitions also include a second construct, which would mandate that in addition to the minimal amount of patient change there must also be significant enough difference to alter patient management.

Varied definitions and inconsistent reporting of outcomes between trials researching similar topics make it difficult to draw comparisons, limiting the value

of each trial to improve overall patient experience or outcome.[61] This variability undermines systematic reviews and meta-analysis aiming to answer a specific research question. Two main issues were identified which cause this problem; which outcomes are selected and the criteria used to define them. In an attempt to improve this, the patient reported outcomes subgroup of the Standardised Endpoints in Perioperative medicine working group (StEP-COMPAC) has recommended the use of at least one patient reported outcome with an established MCID in every study.[62] The common goal being to define which measures should be used in future research and facilitate comparison between studies enabling robust evidence synthesis.[62]

Some limitations in defining the MCID exist which may be as a result of the patient's inability to understand the context of improvement; often reporting current state of health as a comparison against expectations or healthy peers. The MCID is not a universal fixed value and cannot be transferred across patient populations.[63] MCID can also be subject to recall bias and patient variation influencing reporting of change such as age, socioeconomic status and education. Several methods have been developed to calculate MCID's for scoring tools, but no clear consensus exists to select a *best* approach.[64]

2.1.2 Visual Analogue Scale to measure dyspnoea

A visual analogue scale is used to assess dyspnoea with the patient asked to provide a quantification of their dyspnoea by placing a mark on a horizontal or vertical line, usually 100mm in length, sometimes with descriptors or images at the extremes. The Modified Borg Scale (MBS) is the most widely used scale of this type to rate dyspnoea during exercise testing (Table 1).

Score	Difficulty of Breathing
0	Nothing at all
0.5	Very, very slight (just noticeable)
1	Very slight
2	Slight
3	Moderate
4	Somewhat severe
5	Severe
6	---
7	Very Severe
8	---
9	Very, very severe (almost maximal)
10	Maximal

Table 1 - Modified Borg Scale (MBS) for dyspnoea

The scale consists of a vertical line labelled 0-10 with descriptors of severity corresponding to specific numbers. Some of the numbers (6 and 8) do not have a description. The patient can choose the number or the verbal descriptor to quantify their dyspnoea. This style of grading dyspnoea allows for comparison between individuals, based on the assumption the verbal descriptors on the scale describe the same intensity for different subjects. Although the VAS can provide a dimensional measurement of severity of dyspnoea it does not consider the contributing factors. There also exists no criteria or guidelines to allow this type of scale to be used between different observers.[65] Inter-observer reliability has been quantified to support the use of the Borg scale in assessing exercise intensity with authors reporting test correlation coefficients ranging from 0.7 to 0.9.[66]

2.1.3 Medical Research Council scale to measure dyspnoea

Chosen as the primary outcome measure for the work presented in this thesis, the Medical Research Council (MRC) scale (Table 2) has been widely used since 1959 and is based on the exertional effort needed to perform specific tasks, resulting in dyspnoea. Use of the MRC scale is free but should be appropriately acknowledged by researchers. The MRC scale was derived from a coal mining population in Wales by Fletcher et al in the 1940s when studying respiratory problems at the pneumoconiosis unit, allowing a numerical value to be placed on each subjects exercise capacity.[67] Thus, allowing standardisation and comparison between patients/populations.[68] The MRC scale measures *perceived* respiratory disability and is simple to administer, allowing the patient to quantify the extent to which dyspnoea affects their mobility. All questions relate to everyday

activities, are easily understood by patients and can be scored in a few seconds. The scoring tool is usually self-administered, the patient selecting the best option to describe their dyspnoea but, with a slight change in question format, can be delivered by researchers or clinicians.

The MRC score does not quantify dyspnoea itself, but rather it quantifies the disability associated with breathlessness by identifying dyspnoea occurring when it should not (grades one and two) or quantifying exercise limitation (grades three to five). There is up to 98% agreement between observers recording of MRC dyspnoea score and strong association with lung function measurements.[69] While used extensively in the medical literature, the main limitation of the scoring tool is the broad grading; it may be insensitive in detecting small but important changes in dyspnoea levels.[70] There are no precise limits to several of the grades which may contribute to this insensitivity: an individual who can leave the house but walks less than 100 yards does not clearly fall into either grades four or five.[71] The MRC scoring tool is widely used to stratify risk in patient cohorts such as pulmonary rehabilitation in COPD[72] and often used to describe dyspnoea in patients with lung cancer at multi-disciplinary meetings. MRC grading can predict survival and is used to complement pulmonary function testing to describe disability in patients with COPD.[73, 74] NICE recommend use of the MRC dyspnoea scale in the diagnosis of COPD patients, a disease particularly prevalent in the lung cancer population.[75]

The MCID of the MRC scoring tool is widely accepted as being one, meaning any stepwise change represents a clinically important difference to patients.[76] However, the validity of this value is difficult to find and data is limited.[77]

Grade	Statement about perceived dyspnoea
1	Not troubled by breathlessness except on strenuous exercise
2	Short of breath when hurrying on level ground or up a slight hill
3	Walks slower than most people on the level stops after a mile or so or stops after 15 minutes walking at own pace.
4	I stop for breath after walking 100 yards or after a few minutes on the level ground
5	I am too breathless to leave the house, or breathless when dressing/undressing

Table 2 - Medical Research Council dyspnoea scale.
Redrawn from Stenton et al[71] - "The MRC breathlessness scale".

The scale has been 'modified' with more simplified statements and refers to 'people' instead of men but remains based on the same five stages of breathlessness due to exertion.[69] Confusingly, the original grades ranged from 1 to 5 while the modified version grades patients from 0 to 4 (Table 3). In its modified form the MRC scale has been used in more than just respiratory conditions, including disorders such as obesity.[78]

Grade	Statement about perceived dyspnoea
0	I only get breathless with strenuous exercise
1	I get short of breath when hurrying on level ground or walking up a slight hill
2	On level ground, I walk slower than people of the same age because of breathlessness or have to stop for breath when walking at my own pace on the level.
3	I stop for breath after walking 100 yards or after a few minutes on level ground
4	I am too breathless to leave the house or I am breathless when dressing

Table 3 - Modified Medical Research Council dyspnoea scale.
Redrawn from Williams et al 201779 "The MRC breathlessness Scale".

2.1.4 University of California and Sand Diego Shortness of Breath Questionnaire to measure dyspnoea

The University of California and San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) is a 24-item questionnaire commonly used and validated to measure dyspnoea with scores ranging from 0 to 120 (Appendix 9). The original version was developed by Archibald et al[80] in 1987 before being revised in 1998 by Eakin et al[81] to expand the rating scale and incorporate 3 new questions to the original 21-item questionnaire. These additional questions ask about fear of harm from over-exertion, limitations and fear caused by shortness of breath. Similar to the MRC scale, the questionnaire is self-administered.

The UCSD-SOBQ measures dyspnoea over the preceding week across a range of 21 activities of daily living on a six-point rating scale (0 = "not at all" to 5 = "maximal or unable to do because of breathlessness"). Since its development in 1987, the UCSD-SOBQ has undergone a number of revisions to clarify and expand the rating scale to minimise missing data. If patients do not perform the activity described

in any given question, they are asked to estimate the degree of shortness of breath anticipated.

In a group of 54 patients with a variety of respiratory conditions, Eakin et al[81] concluded the UCSD-SOBQ is a valuable tool in both clinical practice and research in patients with moderate to severe lung disease. The authors observed significant negative correlation to exercise tolerance (6-minute walk test), $r=-0.45$, $p<0.05$). Eakin et al[81] also reported excellent internal consistency (Cronbachs alpha >0.9).[82]

The MCID of the UCSD-SOBQ is generally accepted as a change of 5 units. This was originally proposed by Kupferberg et al in 2005 several years following its development.[83] Kupferberg studied 164 patients with moderate to severe COPD, simultaneously measuring dyspnoea using two further measures; chronic respiratory questionnaire and the transition dyspnoea index score. The MCID was evaluated by comparison of agreement between UCSD-SOBQ score and the other two scoring tools for dyspnoea. A change of 5 units being the MCID was confirmed by Ries et al in a retrospective review of published trials.[84]

2.2 Scoring tools to assess quality of life

Many scoring tools exist to quantify and assess quality of life in clinical practice. This section will discuss two tools used within this study; the EQ-5DL quality of life questionnaire and the European organisation for research and treatment of cancer quality of life questionnaire (EORTC).

2.2.1 EQ-5DL quality of life scoring tool

The current 5-level EQ-5DL was introduced in 2009 to improve upon the previous version of the questionnaire. The aim was to increase the sensitivity of the scoring tool and the revised version consists of two components; the EQ-5DL visual analogue scale and the EQ-5DL descriptive system.

2.2.1.1 EQ-5DL visual analogue scale

The EQ-5DL visual analogue scale (VAS) records the patients overall current health on a vertical visual analogue scale where the endpoints are labelled ‘the best health’ and ‘the worst health’ you can imagine. It provides a quantitative measure

of the patient's perception of their overall health.[85] While no validated MCID for the VAS exists for the lung cancer population, it has been proposed a deterioration of approximately seven units/percent would signify a clinically important change in QoL.[86] This is based on a retrospective analysis by Pickard et al[86] on 534 cancer patients (eleven different cancer types, including lung cancer) estimating MCID in EQ-5DL utility and VAS scores using an anchor-based technique. This author is the first to define this value and further work requires to be done before this becomes an accepted definition. Although, a change of seven units/percent appears adequate given published results in similar populations reporting comparable MCIDs.[87, 88]

2.2.1.2 EQ-5DL descriptive system

The EQ-5DL descriptive system comprises of five dimensions: self-care, pain/discomfort, mobility, usual activities and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, extreme problems. Health status is indicated by selecting the box next to appropriate statement for each dimension. The digits are combined into a five-digit number describing the patients' health state. This five digit-number is then converted into a single numerical value called a *summary health index*, which is adjusted to the population of whatever country the patient lives within. The *summary health index* is a continuum from zero to one – one represents 'best health' possible and zero represents 'dead'. However, health state scores less than zero are possible, ostensibly conferring a QoL 'worse than being dead'. Like the VAS component described in section 2.2.1.1, no official MCID for the *summary health index* value exists. This is surprising given the widespread use of this questionnaire. It has been proposed a deterioration of 0.18 units would represent a MCID in quality of life. This is based on work by Coretti et al[89] in a critical appraisal of 18 studies, the largest published paper and most commonly cited to date determining this value. Across these 18 studies, the author reported overall MCID ranged from 0.03 to 0.54 with a raw average across all studies of 0.18. Twelve of the studies were from musculoskeletal populations. Much debate still exists about the calculation and validity of MCID for this scoring tool. Further work is needed to confirm this value in the lung cancer population. However, a value of 0.18 units for the MCID for EQ-5DL *summary health index* value remains the most widely reported, with the best supporting evidence.

2.2.2 European Organisation For Research And Treatment Of Cancer quality of life questionnaire

The European Organisation For Research And Treatment Of Cancer quality of life questionnaire (EORTC) is a system for assessing the perceived QoL in cancer patients in clinical trials. It consists of a core questionnaire (QLQ- C30) and was released in 1993. A supplementary module exists for this scoring tool called the LC-13, which is designed specifically for the lung cancer population.[90]

2.2.2.1 EORTC QLQ-C30

The QLQ-C30 consists of 33 questions which are a combination of multi-item scales and single item measures – see Appendix 10 for an example questionnaire. The QLQ-C30 includes 5 functional scales, three symptom scales, a global health scale and six single items. Each of the multi-item scales includes a different set of items, such that no item appears in more than one scale. The scales and single item measures range in score from 0-100. A high scale score representing a higher response level. For functional and global health status a high response indicates a high level of functioning or high QoL whereas a high score for a symptom scale indicates a high symptom burden.

The principle for scoring these scales is to estimate the average of the items that contribute to the scale (*raw score*) and then use a linear transformation to standardise the raw score so that it ranges from 0 to 100: a higher score indicating a higher (better) level of functioning.

Recently, the EORTC QoL group recommended the use of the QLQ-C30 *summary score* (*Sumscore*) to supplement the 15-outcome profile generated by the QLQ-C30. The *Sumscore* is a global representation of overall QoL and summarises all 15 scores and is arguably easier to interpret than individual scores in each domain. It has been observed the *Sumscore* is more sensitive to changes in a subjects QoL than the global health score and thus can be used as an easy to interpret patient reported outcome measurement.[91]

The *Sumscore* is calculated from the mean of 13 of the 15 QLQ-C30 scales - global quality of life scale and financial impact scale are not included. No validated MCID

exists in the lung cancer population for this score, although there exists general consensus about the value which authors have proposed.

One of the first studies to propose an MCID for the QLQ-C30 *Sumscore* was by Osoba et al, but in the breast cancer population.[92] Using an anchor-based approach and an alternative subjective significance questionnaire in 300 patients, the author concluded a change in 10 points in the *Sumscore* corresponded to a MCID in quality of life.

In the largest critical review to date by Fiteni et al, 18 studies were examined to determine an MCID for the QLQ-C30 *Sumscore* - incorporating the work by Osoba et al.[93] Fiteni also proposed this was found to be represented by a 10-point change in the QLQ-C30 *Sumscore* - all studies that reported an MCID for the QLQ-C30 confirmed a 10-point decrease to represent a meaningful change.[94-96] However, the meta-analysis demonstrated the challenges in agreeing an MCID for the QLQ-C30 *Sumscore* and the heterogeneity of measurement and analysis.

An MCID of 10 has generally been adopted into the lung cancer population, as described. Future work should aim to publish recommendations and confirm an MCID for the QLQ-C30 *Sumscore* in the lung cancer population.

2.2.2.2 EORTC QLQ-LC13

The EORTC study group has developed a supplemental disease specific modular system to complement the core QLQ-C30 questionnaire and assess disease specific QoL: the QLQ-LC13 is a lung cancer module. It contains a 13-item lung cancer specific questionnaire consisting of both multi-item and single-item measures of lung cancer associated symptoms; haemoptysis, dyspnoea, pain, coughing, sore mouth, peripheral neuropathy and hair loss. The QLQ-LC13 was validated in 1994 by the EORTC study group (Bergman et al[97]) in 17 countries. The questionnaire was found to discriminate clearly between patients differing in performance status and to be a clinically valid and useful tool to assess disease and treatment specific symptoms in lung cancer patients when combined with the core QLQ-C30 questionnaire. For over two decades its performance has been continually investigated and improved. There is however no data or evidence to support a

Sumscore equivalent or MCID for the LC-13 supplemental module of the QLQ-C30 questionnaire.

Since selecting the QLQ-LC13 module to be included in the work presented in this thesis, an updated module (QLQ-LC29) has been released in 2020, updating the LC-13 module since its development in 1994.[98]

2.2.3 World Health Organisation disability schedule 2.0

The World Health Organisation disability schedule 2.0 (WHO DAS 2.0) is a generic assessment instrument for health and disability used across multiple conditions. The assessment is short, easy to administer and applicable in both clinical and general population settings across cultures and in all adult populations. The questions cover 6 domains of functioning including cognition (understanding and communicating), mobility (moving and getting around), self-care (hygiene and eating), getting along (interacting with other people), life activities (domestic responsibilities) and participation (joining in community activities).[99] It contains 12-items for overall functioning scored on a Likert scale of zero to four, zero being *'no difficulty'* and four being *'extreme difficulty'*. The cumulative score is converted to a percentage: with the maximum possible score being 48 (100%). The WHO DAS 2.0 has not been specifically validated in the lung cancer population but remains a well-recognised tool to measure global disability. No validated MCID exists for the WHO DAS 2.0 disability tool, however Shulman et al 2020 proposed a change/decrease of 5% or more after surgery should be considered a clinically important change in disability [100] – also observing patients with a score <16% following surgery have an acceptable symptom state. Conversely, a score of >35% can be considered as having at least moderate disability.[100] The patient reported outcomes subgroup of the StEP-COMPAC initiative has recommended the use of WHO DAS 2.0 as the gold standard measure of functional status in clinical trials in the perioperative setting.[62]

2.2.4 Hospital anxiety and depression scale

The hospital anxiety and depression scale (HADS) was devised by Zigmond et al[101] in 1983 to measure anxiety and depression in the general population and has developed into a popular tool in clinical practice, which has also been validated

in the lung cancer population.[102, 103] The tool is simple to use and very few people have difficulty completing the questionnaire. Anxiety and depression are assessed together, recognising the two often co-exist. The questionnaire has seven questions for anxiety and seven for depression which are interspersed. Each component therefore has a maximum score of 21 and each component must be scored separately. Castelli et al[104] observed the HADS tool to be an effective screening questionnaire for depression in the lung cancer population. Physical symptoms are excluded from the scale, such as sleep disturbance or pain, due to potential confounding.

Through a systematic review of studies using the HADS questionnaire, Bjelland et al identified a cut-off score of 8 out of the possible maximum score of 21.[102] A score of ≥8 has a specificity of 0.78 and sensitivity of 0.9 for diagnosis of anxiety. For depression, a score ≥8 has a specificity of 0.79 and sensitivity of 0.83.[102]

2.2.4.1 Measuring Dyspnoea and Quality of Life: conclusion

It is important to ensure the tool selected to measure dyspnoea or quality of life has a validated MCID to enable a significant patient centred difference to be detected in addition to statistically significant results. In an attempt to standardise patient reported outcomes, the patient reported outcomes subgroup of the Standardising Endpoints in Perioperative medicine (StEP-COMPAC) initiative has recommended a list of outcomes, from which one should be selected for use in every study. All of the scoring tools described within this chapter were selected and used within this thesis to measure dyspnoea and quality of life. Their selection was based on their strengths, simplicity of use and proposed MCID values which enabled comparison of meaningful patient centred changes.

3 Proposed mechanisms of dyspnoea and reduced functional capacity following lung resection surgery

Over 40% of patients report long term disabling dyspnoea following lung resection which is not fully explained by changes in pulmonary function,[49] as will be discussed in this chapter. This may reduce the patient's ability to be physically active, reducing post-operative quality of life. The proposed mechanisms by which this occurs are likely to be multifactorial. Clinical guidelines advocate the use of predicted post-operative pulmonary function to calculate the risk of post-operative dyspnoea (Section 4.3.3). However, this has been shown to be poorly associated with changes in exercise capacity.[105]

Lung resection surgery may result in persistent reduction in post-operative pulmonary function of 10-40% which may contribute to long term global functional impairment.[105] This is defined by a reduction in FEV_1% and DLCO% from pre-operative values. This reduction in pulmonary function is multifactorial and due to removal of lung tissue and alteration in chest wall movement due to surgical incision.[106] Bolliger et al 1996 found patients undergoing lobectomy and pneumonectomy had significantly reduced pulmonary function test results and CPET results which persisted following surgery with increased dyspnoea particularly in those patients undergoing pneumonectomy.[107] This supports a hypothesis that increased dyspnoea is due to increased loss of lung parenchyma available for gas exchange. While lung function has been shown to decrease following lung resection, this does not completely explain the decline in functional capacity observed in some patients.

Pelletier et al[108] observed a change in FEV_1% was a poor predictor of change in functional capacity following lung resection (Figure 3). In 47 patients undergoing lung resection, FEV_1% predicted[A] accounted for only 30% of the variance in exercise capacity. Similarly, Larsen et al confirmed alteration in FEV_1 is a poor predictor of a deterioration in exercise capacity following pulmonary resection in 97 patients (Figure 4).[109] In some patients, exercise capacity increased or barely

[A] FEV1% predicted is defined as the FEV1% of the patient divided by the average FEV1% in the population for any person of similar age, sex and body composition. ppoFEV1% is defined as the predicted post-operative FEV1% as calculated by lung segment calculation (section 4.3.3).

changed despite a loss of up to 35% of ventilatory capacity (as measured by FVC). While the association between exercise capacity and FVC was significant, only a weak relationship existed with FVC predicted accounting for just 18% of the variance in functional capacity. There was no difference between the magnitude of lung lost and loss of functional capacity in those patients undergoing lobectomy.

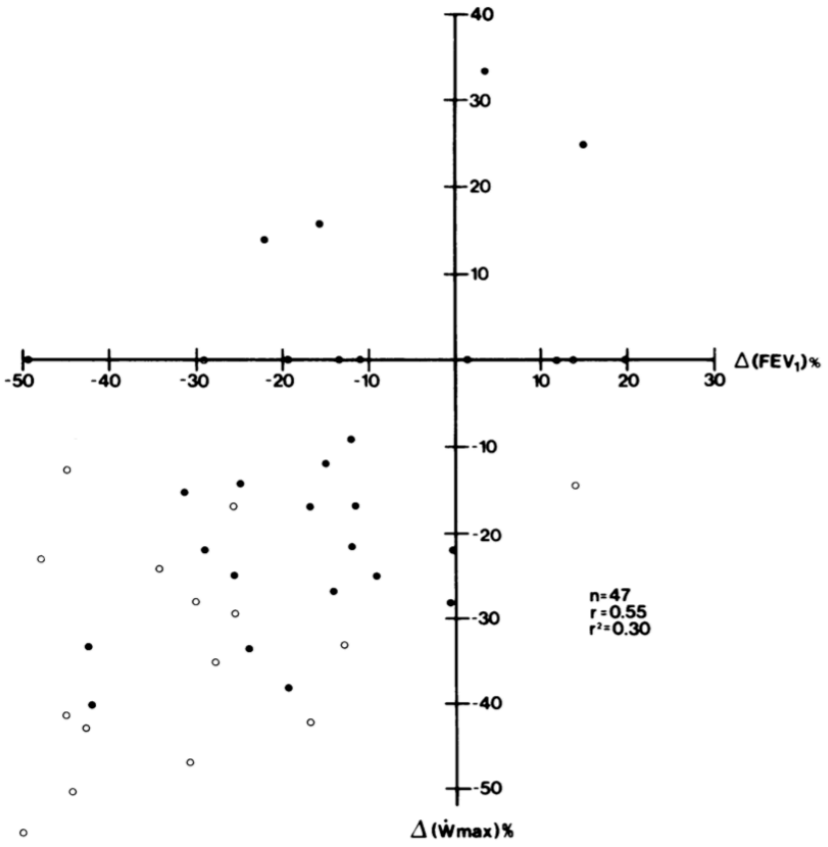

Figure 3 – Relationship between change in exercise capacity and change in FEV_1%.
Expressed as percentages of initial values. Wmax – maximal exercise intensity. Black dots = lobectomy. White dots = pneumonectomy, (n=47). Patients in left and right upper quadrants had an increase in FVC but still had a reduction in functional capacity. Taken from Pelletier et al[108]

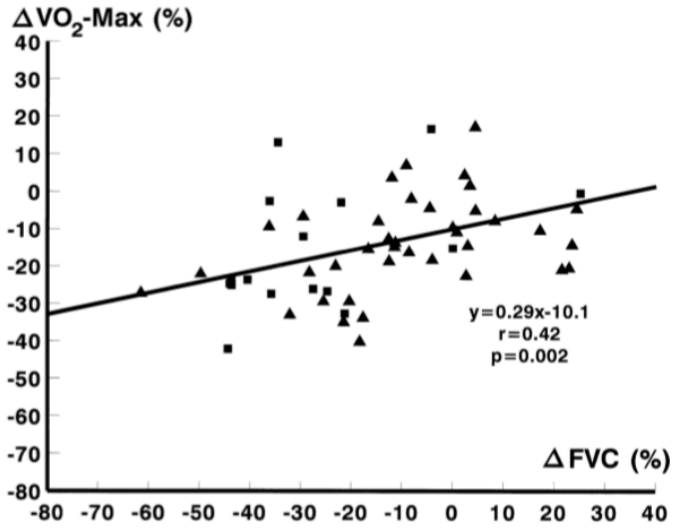

Figure 4 – Relationship between alterations in maximal oxygen uptake (VO_2-Max%) and alterations in forced vital capacity (FVC% pre-operatively) following lung resection,
Taken from Larsen et al[109] (n=97).

In 1998, Nezu et al[110] examined 82 patients undergoing lung resection for cancer to assess the effects of pulmonary resection on post-operative recovery and limitation of exercise capacity. The authors observed a reduction in FEV_1 and VO_2max at more than 6 months when compared to pre-operative values in both lobectomy and pneumonectomy patients. Compared to pre-operative values, the functional percentage decrease in FEV_1 at six months for lobectomy and pneumonectomy groups were 11% and 36%, and for VO_2max were 13% and 28%, respectively. Maximum heart rate and heart rate percentage decreased in both lobectomy and pneumonectomy groups, ($p<0.05$). Combined with no reduction in breathing reserve, this would suggest cardiac limitation as a contributing factor to the reduction in functional capacity following lung resection.

3.1 Reduced respiratory muscle function and reduced pulmonary function

A reduction in respiratory muscle strength following lung resection is suggested as contributing to post-operative dyspnoea, however there is much debate surrounding how much influence it has. It is well established that dysfunction of the respiratory muscles following any type of surgery (thoracic or not) may lead to a reduction in vital capacity, tidal volume and total lung capacity – secondary to muscle/nerve injury or chronic pain impairing ventilation causing insufficient cough. In turn, this can reduce functional residual capacity affecting the gas

exchange properties of the lung by increasing ventilation/perfusion (V/Q) mismatch.[111]

Respiratory muscle function can be affected by damage directly to the muscle, or the nerves supplying the muscle, due to surgical incision or indirectly by a change in the respiratory mechanics.[112] Chest wall distortion may reduce chest wall compliance and increase work of breathing with reduced efficiency of the respiratory muscles or reduced compliance of lung tissues. Respiratory muscle mass may also be deconditioned following major lung resection. A decrease in chest wall compliance observed after thoracotomy may increase morbidity and mortality in those with borderline pre-operative lung function.[112] Maximal inspiratory/expiratory pressures (MIP and MEP) are often used as markers of respiratory muscle function.

Chest wall damage may be a major determinant in the decrease in respiratory muscle strength seen after lung resection; a smaller decrease has been observed in respiratory muscle strength in patients undergoing video assisted thoracoscopic surgery (VATS) compared to open thoracotomy. Nomori et al[113] observed this in 81 patients undergoing lung resection measuring MIP and MEP pre-operatively and 12 weeks post-operatively; thirty-one patients within this group underwent VATS surgery and had better/increased post-operative MIP and MEP than those undergoing conventional thoracotomy, ($p<0.01$). Nomori observed a reduction in post-operative respiratory muscle function in those who were aged >70 years, ($p<0.01$). These results may support the hypothesis that increased post-operative dyspnoea is multifactorial and involves impaired respiratory mechanics in vulnerable patients (such as the older population with increasing burden of cardiovascular morbidity).

Brocki et al[114] recruited 80 patients from a single centre undergoing lung resection to evaluate respiratory muscle strength following surgery. MIP and MEP were used as markers of muscle function and measured pre-operatively, 2 weeks and 6 months post-operatively. Brocki found no change in pressures at all peri-operative time points and concluded:

'Respiratory muscle function is unlikely to be the sole cause of dyspnoea following lung resection…'

Brocki et al 2018[114]

3.2 Cardiovascular dysfunction

The inconsistent association between pulmonary function and post-operative functional capacity suggests other factors play a major role in long term dyspnoea. Several authors have suggested this may result from cardiac rather than pulmonary limitation.[110, 115] This section explores the potential cardiovascular factors contributing to post-operative dyspnoea and reduced functional capacity following lung resection surgery.

3.2.1 Post-operative dysrhythmias

Arrhythmia, in particular atrial fibrillation (AF), is the most common cardiac complication after thoracic surgery which may contribute to long term dyspnoea and reduced functional capacity with a reported incidence of *new onset* post-operative atrial fibrillation (POAF) of >10-20% following lung resection surgery. The pathophysiology of POAF is multifactorial, complex & poorly defined; involving interaction between triggering stimuli and sustaining processes acting on a vulnerable myocardium, predisposed to developing a tachyarrhythmia.[116] Risk factors include increasing age, male gender, electrolyte imbalance and infection. The extent of pulmonary resection is also associated with POAF; patients undergoing pneumonectomy having higher rates of POAF than those undergoing lobar resections.[116]

The prognostic significance of arrhythmias is difficult to quantify as it can be associated with or induced by other complications such as heart failure or pulmonary oedema. PAOF is often considered a benign operation-related problem which is transient however it can result in post-operative hypotension, stroke, myocardial infarction and increased duration of hospital stay.[117] While AF is a well-recognised cause of dyspnoea in the general population, little evidence exists to link new POAF to long term dyspnoea in the lung resection population.

Amar et al[118] studied 100 patients undergoing pulmonary resection without a history of AF or previous thoracic surgery and examined the effects of pre-defined risk factors on the incidence of AF. Echocardiograms were performed pre-operatively and post-operatively to evaluate cardiovascular dysfunction and estimate right ventricular systolic pressure (using tricuspid regurgitation jet velocity). POAF occurred in 18% of patients and echocardiography revealed elevation of right ventricular systolic pressure, when compared to those without POAF. Amar concluded, increased right heart pressure, but not fluid overload or right heart enlargement, *may* predispose to supraventricular tachycardias following lung resection.[118]

The only study found after a review of the literature to examine AF, lung function and patients with long term dyspnoea is Ariansen et al.[119] In this study the authors report reduced FEV_1 was associated with new onset POAF, with patients complaining of dyspnoea. The patients within the study with AF had reduced lung function compared to subjects in sinus rhythm. Furthermore, patients with AF were more likely to lie below 5th percentile of predicted FEV_1 ($p<0.05$) compared to control subjects. The dyspnoea frequency and severity scores correlated with VO_2 peak in AF patients ($r= -0.6$, $p<0.01$) and with FEV_1% in control subjects ($r= -0.3$, $p<0.05$). Ariansen concluded, dyspnoea was therefore related to exercise capacity rather than lung function in AF patients.

3.2.2 Right ventricular dysfunction

It has been hypothesised that right ventricular (RV) dysfunction occurs in *some* patients following lung resection and may contribute to long term functional outcome.[120] The incidence of post-operative RV dysfunction following lung resection is difficult to quantify with limited information available on RV adaptation following pulmonary resection – the impact of major lung resections on RV function has not been well investigated.[121] This is likely because assessment of RV function is difficult due to the RV's retrosternal position, complex geometry and marked load dependence.[122, 123]

Our research group has investigated peri-operative RV dysfunction using cardiovascular magnetic resonance (CMR) imaging. In 25 patients undergoing lung resection, McCall et al observed a decrease in right ventricular ejection fraction

(RVEF) persistent until 2 months post-operatively, (p=0.02).[120] Interestingly, no changes in left ventricular ejection fraction occurred over the same time frame (Figure 5). The mechanism proposed was a mismatch between afterload and contractility. An increase in pulsatile afterload, resulting from the operative pulmonary artery was also observed.[120] This work confirms similar work in this area which have also reported a reduction in RVEF post-operatively.

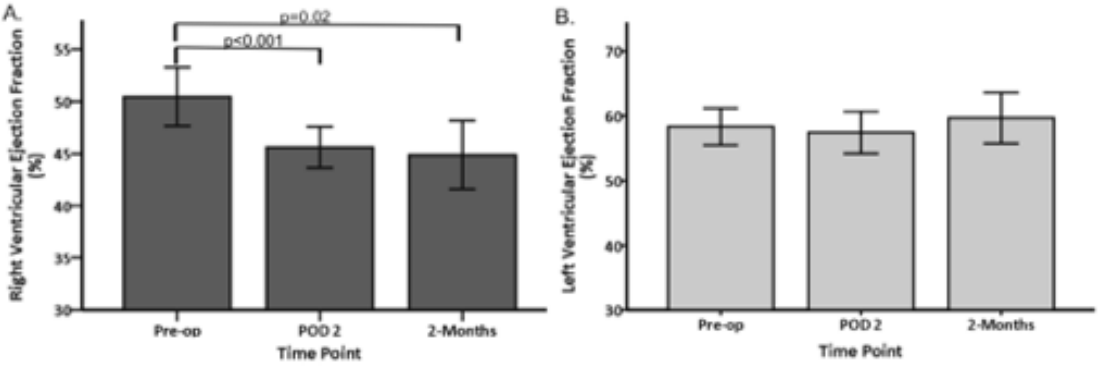

Figure 5 – Taken from McCall et al.120 Left and right ventricular ejection fraction over time (%).
A = Peri-operative right ventricular ejection fraction. **B** = Peri-operative left ventricular ejection fraction Pre-op = pre-operative. POD 2 = Post-operative day 2. (n=25)

Other studies have described a 15-25% relative reduction in RV ejection fraction following lung resection.[124, 125,126] Reed et al[126] observed RV dysfunction in 15 patients in the post-operative period, using pulmonary artery catheters. Right ventricular end-diastolic volume increased on post-operative day one and post-operative day two, ($p<0.05$). Furthermore, although pulmonary artery pressures were observed to be increased in the immediate post-operative period, pulmonary vascular resistance remained lower or unchanged from pre-operative levels; suggesting a rise in static afterload may not be the only contributing factor.

The aetiology of right ventricular dysfunction following lung resection has been the subject of much debate.[127] Apart from rare conditions such as pulmonary embolism, alterations in RV contractile performance and increases in RV afterload are *suggested* mechanisms of RV dysfunction following lung resection, but have yet to be proven. Increased RV afterload would seem an intuitive cause of RV dysfunction in this population. The degree of RV dysfunction in some instances is

related to the extent of surgery performed, with increased RV dysfunction in patients undergoing pneumonectomy instead of lobectomy.[121, 128, 129]

Limited work has explored the association between RV dysfunction and reduced long-term functional outcome following lung resection and this remains challenging.[127] In a small cohort of 35 patients, Foroulis et al observed that patients with an increased systolic pulmonary artery pressure at 6-months following lung resection had increased levels of dyspnoea (p=0.01) and an increase in post-operative complication rates, including arrhythmias.[129,124] Lewis et al observed that intra-operative RV dysfunction observed on echocardiography identified patients who would subsequently develop 'cardiorespiratory disability' long after surgery (defined as New York Heart Association (NYHA) heart failure class III/IV), suggesting the impact of peri-operative RV dysfunction can be sustained long into the post-operative period.[124] At the time of PA clamping, the mean RVEF for those patients in class I/II NYHF was 43% compared with 31% for those in NYHA class III/IV. This observation was made in a small cohort of just 20 patients, but may suggest that increased peri-operative afterload on a vulnerable RV has an impact on functional outcome. However, in larger populations these findings have not been reproduced.[124, 125]

In 1994, in a small cohort of 20 patients who had major lung resection, Okada et al[130] observed RV dysfunction up to three weeks post-operatively when compared to pre-operative values, ($p<0.05$). In this group, pulmonary vascular resistance index at rest initially increased then returned to baseline levels. However, during exercise, post-operative PVR levels increased markedly, ($p<0.05$) (Figure 6). Right ventricular ejection fraction was decreased post-operatively at rest and during exercise, compared to pre-operative levels, ($p<0.05$) (Figure 7). These results indicate post-operative RV dysfunction with a potential to have an effect on functional capacity and quality of life in some patients. Okada et al hypothesise that post-operative functional capacity is influenced by the degree of RV dysfunction in maintaining pulmonary blood flow, as the pulmonary artery pressure sets the limit for increasing flow with exercise. During exercise there is an increased oxygen demand which cannot be met due to RV dysfunction; the RV cannot eject increased cardiac output through a reduced pulmonary vascular bed (following resection). This is reduced contractile reserve. This supports that a change in afterload may be the main factor affecting RV pump performance, with

exercise-loading driving the RV to its limitations and exposing underlying dysfunction.

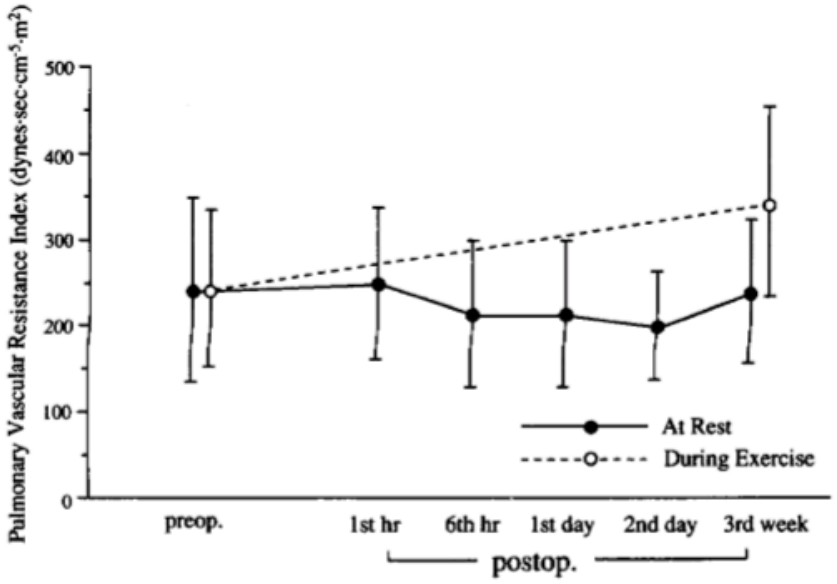

Figure 6 – Peri-operative pulmonary vascular resistance index.
Taken from Okada et al[130]**.** Increased PVR during exercise at three weeks post-operatively. (n=20).

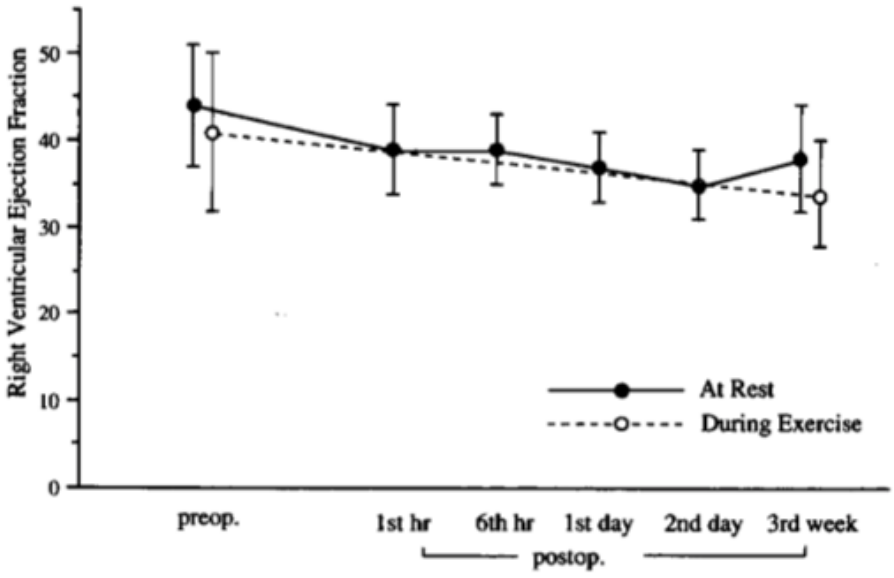

Figure 7 – Peri-operative right ventricular ejection fraction.
Taken from Okada et al[130]**.** Post-operative decreased RVEF with no return to baseline levels at rest or during exercise. (n=20).

Similar to the work by McCall et al described above in this section, Okada et al also observed peri-operative LV function was unchanged: indices of function such as cardiac index, arterial pressure and pulmonary capillary wedge pressure were no different to pre-operative values at 3 weeks post-operatively, ($p>0.05$).

In 1996, Okada et al again examined peri-operative RV function, this time exploring the RVEF in the pre-operative risk evaluation of candidates for pulmonary resection.[131] In a cohort of 18 patients, there existed no association between pre-operative RVEF (%) and either incidence of post-operative complications or length of hospital stay, ($p>0.05$). However, patients with a pre-operative decrease in RVEF *following exercis*e had a longer length of hospital stay, increased incidence of post-operative complications and simultaneous increase in PAP, ($p<0.05$). Patients with impaired pre-operative RV contractile reserve may be the group who struggle in the face of the increased afterload challenge of lung resection. This suggests the RV may have a role in the mechanism of functional deterioration experienced by some patients. The deterioration of the RVEF during the stress of exercise, when faced with a reduction in the pulmonary vascular bed following lung resection, may play a part in the decline of post-operative cardiopulmonary reserve. These results are similar to Okada and co-workers earlier paper discussed above in this section and would require further work to confirm the findings.

If peri-operative RV dysfunction following lung resection was associated with long term functional impairment, identification of or susceptibility to pre-operative RV dysfunction could improve prediction of post-operative dyspnoea. Targeted interventions could also attempt to decrease RV dysfunction in those most at risk, improving patient outcome. However, to date no study has attempted to undertake this.

3.2.3 Myocardial injury following non-cardiac surgery

Myocardial injury following non-cardiac surgery (MINS) is a relatively new concept defined as a 'prognostically relevant myocardial injury due to ischaemia occurring within 30 days after non-cardiac surgery'.[132] While no study demonstrates a clear link between MINS and post-operative dyspnoea in the lung resection population, it reasonable to hypothesise that peri-operative myocardial injury could lead to long-term heart failure with coinciding breathlessness. The VISION study (Vascular events in non-cardiac surgery patients cohort evaluation) was an international prospective study that estimated worldwide prevalence of MINS to be around 8% in patients undergoing major non-cardiac surgery, which was associated with 30 day mortality (adjusted hazards ratio, 3.87, 95% CI 2.96-5.08).[133] The POISE study

(Peri-operative ischaemia evaluation) also reported a MINS rate of 5.7% and suggested this will increase with time given the ageing population with increased co-morbidity.

In a study of 598 patients undergoing lung resection surgery, myocardial injury was observed in 1.2%; abnormal exercise testing and intraoperative hypotension being the strongest predictors for these events.[134] Herrington et al [135] also investigated myocardial injury (defined as myocardial ischaemia or infarction) following lung resection and observed the incidence of myocardial injury was low (0.13%) in those with no previous myocardial history and moderate (2.8% - 17%) in patients with previous history of infarction. There was also no association between anaesthetic technique, nor duration of procedure and peri-operative myocardial injury.

Post-operative myocardial injury is a strong predictor of mortality after non-cardiac surgery, therefore cardiac risk for lung resection must be assessed prior to surgery. The American college of cardiology and American heart association guideline remains the best method for cardiac risk assessment in non-cardiac surgery. This is discussed further in section 4.2.2.

3.2.4 Pulmonary embolism

Pulmonary embolism (PE), by occlusion of the pulmonary arterial bed may lead to life threatening reversible or irreversible myocardial injury & right ventricular failure. Similar to MINS detailed above, no study demonstrates a clear link between PE and long term post-operative dyspnoea in the lung resection population but it is again reasonable to hypothesise that peri-operative myocardial injury or right ventricular dysfunction as a result of massive PE could lead to long-term heart failure and breathlessness. PE following non-cardiac surgery and lung resection surgery is more common than traditionally thought. In a study of 66 patients undergoing elective intermediate to high risk non-cardiac surgery, Grobben et al[136] unexpectedly observed clinically silent PE in one third of patients with peri-operative myocardial injury (n=46) using post-operative computed tomography.[136] In this study, none of these Pes were clinically suspected. This suggests the prevalence of post-operative PE may be much higher than reported, as a large number may go undiagnosed.

The diagnosis of post-operative PE is difficult to confirm due to the lack of specific clinical manifestations; chest pain, dyspnoea, tachycardia and decreased oxygen saturations can often be mistaken for incisional pain, reduced blood volume and pulmonary atelectasis. Deep Vein Thrombosis (DVT) of upper of lower limbs are found in 90% of patients with PE, suggesting it is the result of DVT in most cases.

Post-operative DVT is more common in lung resection surgery but may lead to serious complications such as pulmonary hypertension after pulmonary embolism. One of the few studies to examine thromboembolism after lung resection is by Ziomek et al[137] which prospectively observed 77 patients for 30 days post-operatively. The incidence of thromboembolism was higher in bronchogenic carcinoma than in metastatic cancer or benign disease and also increased in frequency with increasing size of cancer and size of lung resection. The overall incidence was 26%, (19% post-operative); 4 patients having pulmonary embolism in which 1 fatality occurred.

Given the potential for irreversible myocardial injury and right ventricular failure it could be postulated this may result in long term dyspnoea following surgery, but to date there is no evidence to make this link.

3.2.5 Shunting

The *rare* development of an atrial shunt (right to left) through a patent foramen ovale (PFO) may be a cause of long term post-operative dyspnoea. The overall prevalence of PFO in the general population is 20-35% and given not all patients get a pre-operative echocardiogram prior to lung resection, it is possible some patients may present with post-operative shunt. Factors for the development of the shunt include; mediastinal shifting and rotation, compression of the atrium by pleural fluid, reversal of inter-atrial pressure gradient due to a decrease in right ventricular compliance, PE, MI and positive pressure ventilation. Clinical features include postural dependent (worse in upright position) and volume dependent (worse in dehydrated patients) dyspnoea. Treatment is by surgical repair or with the increasingly popular percutaneous device closure.[138]

While this is an interesting and rare cause of post-operative dyspnoea, it seems unlikely that this is a major contributor to long term shortness of breath following

lung resection. Only a few cases of right-to-left shunts following lung resection have been reported as case studies.[139]

3.2.6 Conclusion

The cause of dyspnoea following lung resection is likely multifactorial. While a decrease in lung parenchyma (and subsequent decrease in surface area available for gas exchange) would be an intuitive cause of dyspnoea, cardiac dysfunction is likely to play a role. Those patients undergoing more extensive resection such as pneumonectomy often demonstrate reduced exercise capacity limited primarily by cardiovascular dysfunction, regardless of other co-morbidities,[115] adding to the increasing body of evidence to support this theory.

4 Operative risk assessment in lung cancer patients

4.1 Methods: review of the literature

A database search was performed at the University of Glasgow in August 2018, in consultation with the library. The following strategy and key words were used using the *Embase* online database, 1946 to present with daily updates;

1. Pneumonectomy / (18605)
2. Thoracic Surgery / (26945)
3. Pneumonectom$.tw (7482)
4. (Lung adj4 resection$.tw) (9692)
5. 57obectomy$.tw (20154)
6. or/1-5 (63663)
7. "Quality of life"/ (379792)
8. exp dyspnoea
9. exp Exercise Test/ (63044)
10. Walk test/ (1183)
11. (QLQ-C30 or QLQC30 or EQ-5DL or EQ5DL).tw (6533)
12. (MRC or Modified Research Council Dyspn?ea Scale or Modified Research Council Dyspn?ea Score).tw (8910)
13. (Life Quality or Quality of life).tw (358517)
14. (walk test$ or Stair test$ or (stair? Adj2 climb$) or exercise test$).tw (37820)
15. (Dyspn?ea or shortness of breath or breathlessness).tw (78792)
16. forced expiratory volume/ (51543)
17. (forced expiratory volume or FEV or FEV1).tw (49950)
18. brain natriuretic peptide/ (24922)
19. (BNP or Brain Natriuretic Peptide).tw (25719)
20. exp lung function test/ (124859)
21. exp lung diffusion capacity/ (6136)
22 exp oxygen consumption (352274)
23. exercise test/ or exp cardiopulmonary exercise test/ or exp treadmill exercise/ (58246)
24. anaerobic threshold/ (3322)
25. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 (678904)
26. 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 (539350)
27. 6 and 25 and 26 (1574)
28. limit 27 to English language (1465)

Title review for relevance reduced the 1465 studies down to 537 for abstract review. Following review, 127 abstracts remained for full text review. References of all articles were reviewed for further relevant work, with 89 additional studies being identified (Figure 8).

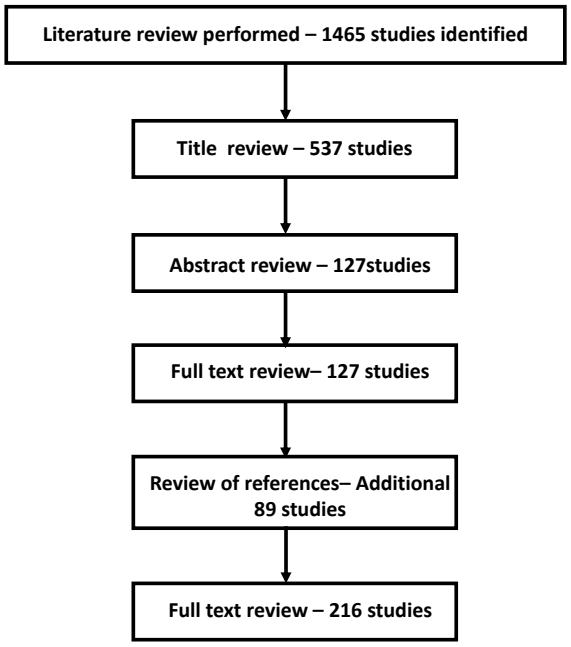

Figure 8 – Flow chart of Literature review.
Performed August 2018

Studies were included if they reported any variable used in the prediction of dyspnoea, complications, morbidity and mortality or quality of life following lung resection surgery. Articles examining *lung volume reduction surgery* were excluded; lung volume reduction is a palliative procedure used to treat severe emphysema or COPD. The work presented in this thesis will not include any studies reporting lung volume reduction as this represents a different population.

For discussion the studies have been divided into 9 groups;

1. Studies utilising *FEV_1 and DLCO* to predict post-operative dyspnoea
2. Studies utilising *FEV_1 and DLCO* to predict mortality
3. Studies utilising *FEV_1 and DLCO* to predict post-operative quality of life
4. Studies utilising *FEV_1 and DLCO* to predict post-operative pulmonary complications
5. Studies utilising *cardiopulmonary exercise testing* to predict post-operative dyspnoea, mortality, quality of life or pulmonary complications
6. Studies utilising *shuttle walk testing* to predict post-operative dyspnoea/mortality/quality of life or pulmonary complications

7. Studies utilising *stair climbing* to predict post-operative dyspnoea/mortality/quality of life or pulmonary complications

8. *Other* predictive markers of dyspnoea/mortality/quality of life or pulmonary complications such as arterial oxygen content, arterial oxygen desaturation, pre-operative pulmonary artery pressure, arterial carbon dioxide content and minute ventilation to carbon dioxide output ratio

9. *Biomarkers* to predict post-operative dyspnoea/mortality/quality of life or pulmonary complications such as B-Type natriuretic peptide

4.2 Introduction to risk assessment in thoracic surgery: conventional risk stratification

The British Thoracic Society (BTS) and the American College of Cardiology (ACC) adopt what is termed a 'tripartite' approach to quantify 'surgical risk' and facilitate calculation and assessment of individual outcomes to be discussed with the patient and by the multi-disciplinary team (MDT) (Figure 9).[18, 52] Within the guidelines, estimated risk of post-operative cardiac events, peri-operative death and post-operative dyspnoea must be considered before offering a patient surgery - the patient must be involved in the calculation, assessment and discussion of these risks in an attempt to predict individual outcomes. Ideally, the patient must accept the potential impact on quality of life before proceeding with surgery (shared decision making). This section will focus on the *tripartite* approach for pre-operative risk stratification in lung resection patients.

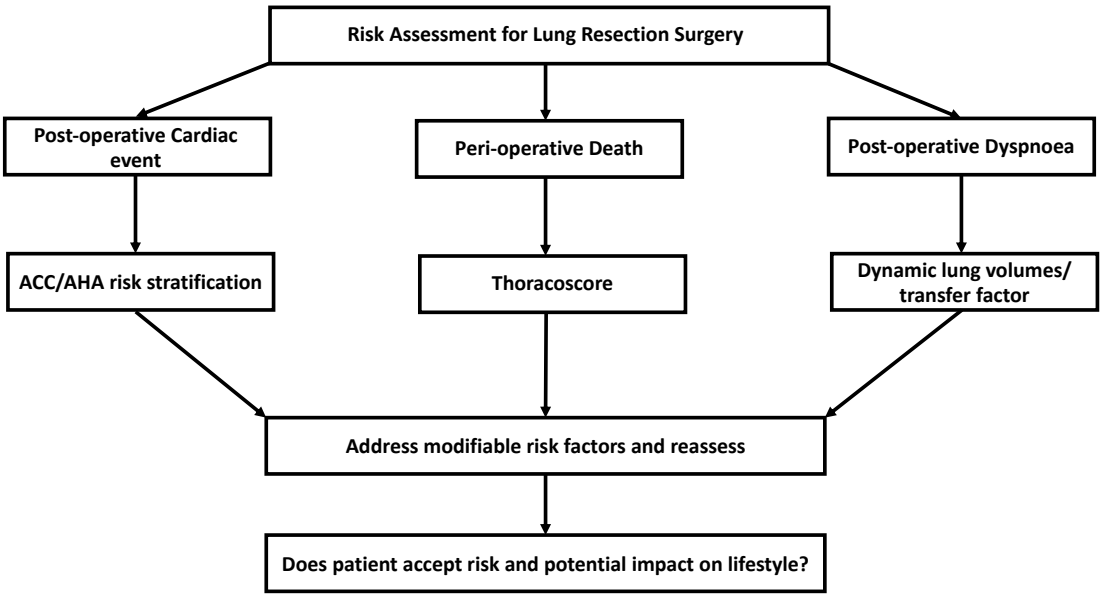

Figure 9 – Risk assessment pathway for lung resection surgery, as per BTS guidelines.
ACC = American College of Cardiology, AHA = American Heart Association. Redrawn from British Thoracic Society 2010 Risk Assessment for Lung Resection Surgery.[18]

4.2.1 Peri-operative death

Risk of in-hospital mortality for lung resection is important in the peri-operative decision-making process: the national average being 2.3% for lobectomy and 5.8% for pneumonectomy in the UK according to the latest BTS guidelines, 2010.[18] Validated across the world, *Thoracoscore* is currently the most widely used model to quantify this risk and was developed by the French Society of Cardiovascular Surgery in 2007.

Thoracoscore calculates in-hospital mortality risk and was constructed from 15,183 patients undergoing thoracic surgery. It has been validated with an area under the receiver operator characteristic curve (AUROCC) of 0.86.[140] It was developed and validated in France and then re-validated in a patient set in the United states of America with an AUROCC of 0.95.[141] *Thoracoscore* consists of nine variables (Table 4); age, sex, ASA score, performance status, dyspnoea score, priority of surgery, extent of surgery, malignancy diagnosis and a comorbidity score. An equation transforms these variables into a predicted in-hospital death rate.

$$Logit = -7.3737 + \text{Sum (calculated beta)}$$

$$Predicted\ in\ hospital\ death\ rate = \frac{\text{e(logit)}}{1 + \text{e(logit)}}$$

Equation 1 - Regression equation used to calculate *Thoracoscore*. Calculated beta value taken from sum of beta-coefficients, displayed in Table 4 below. Falcoz et al[142]

The European Society Objective Score (ESOS) is an alternative to *Thoracoscore* to predict in-hospital mortality with only two variables, ppoFEV_1 and age and was derived in 2005 from the online European thoracic surgery database of 3426 patients.[143] A 2012 study by Barua et al[144] of 290 patients demonstrated the superiority of the ESOS to *Thoracoscore*; its sensitivity 88% and specificity 67% was better than *Thoracoscore* sensitivity (67%) and specificity (53%). However, when this paper was further scrutinised, these results were in a single centre with a single operating surgeon. While *Thoracoscore* has been validated to predict in-hospital mortality, it has poor correlation with FEV_1 and peak VO_2 – suggesting they may not be useful for prediction of in-hospital mortality in the lung cancer population. Prior to *Thoracoscore*, no peri-operative scoring tool to quantify predicted mortality existed. Instead, careful consideration was given to increasing age and co-existing morbidities.[145] Most UK cardiothoracic centres tend to outperform, with the *Thoracoscore* overestimating mortality.[146]

Variable	Value	Beta-coefficient
Age (years)	<55 55-65 >65	- 0.7679 1.0073
Gender	Male Female	0.4505 -
ASA score	≤2 ≥3	- 0.6057
MRC dyspnoea score	≤2 ≥3	- 0.9075
WHO performance score	≤2 ≥3	- 0.9075
Comorbidities	1-2 >3	0.7447 0.9065
Priority for surgery	Elective Emergency	- 0.8443
Operation type	Pneumonectomy Other	1.2176 -
Diagnosis group	Benign Malignant	- 1.2423

Table 4 - Variables within *Thoracoscore*.
Adapted from Falcoz et al[142]. ASA = American Society of Anaesthesia score. MRC = Medical Research council. WHO = World Health Organisation. Co-morbidities included smoking addiction, history of cancer, hypertension, chronic obstructive pulmonary disease, Ischaemic heart disease, diabetes, peripheral vascular disease, alcoholism and obesity.

4.2.2 Post-operative cardiac event

The latest BTS guidelines 2010 advocate cardiovascular risk and morbidity should be assessed prior to surgery using the American College of Cardiology (ACC) and American Heart Association (AHA) 2007 guidelines. Combined incidence of cardiac death and non-fatal myocardial infarction is reported to be 1–5% in the lung resection population.[52] A full history, physical examination, assessment of functional status and resting ECG are prerequisites and must be performed in each patient. Any patient in which this identifies an *active cardiac condition* (Table 5) requires evaluation by a cardiologist and optimisation before considering surgery. If unexplained dyspnoea or a murmur are discovered, the patient should have an echocardiogram.

Condition	Example
Unstable coronary syndrome	Unstable/severe angina
Decompensated heart failure	NHYA class IV
Significant arrhythmias	3rd degree AV block Mobitz type II Supraventricular arrhythmias with ventricular rate >100bpm
Severe heart valve disease	Severe aortic stenosis: mean pressure >40mmHg or aortic valve area <1.0cm^2

Table 5 - Active cardiac conditions.
Redrawn from British Thoracic Society guidelines 2010.[18] NYHA – New York Heart Association classification, bpm = beats per minute

In those who *do not* have an active condition the *revised cardiac risk index* is used which is a validated model with AUROCC of 0.81 (Table 6).[147] The ACC/AHA suggest patients with less than two risk factors and reasonable functional capacity can proceed to surgery without further investigation. Patients unable to climb a flight of stairs and with poor cardiac function or greater than three risk factors should have further assessment to screen for reversible cardiac ischaemia, with tests such as stress testing or exercise thallium scanning.

Number of factors	Risk of major cardiac complication*
0	0.4%
1	1%
2	7%
>2	11%

Table 6 - Revised cardiac risk index
Redrawn from British Thoracic Society Guidelines 2010[18]
Risk factors: high risk surgery (includes *all* thoracic surgery), ischaemic heart disease, congestive cardiac failure, cerebrovascular disease, insulin therapy for diabetes, pre-operative serum creatinine >177 micro/mol/l. *Cardiac complications defined as MI, pulmonary oedema, ventricular fibrillation or primary cardiac arrest, complete heart block.

4.2.3 Dyspnoea

Dynamic lung volumes and transfer factor have been conventionally used to estimate risk in patients being considered for lung resection surgery (Figure 10) and will be discussed in more detail in section 4.4. Dyspnoea is a common complication following lung resection, with 30 - 50% of patients reporting disabling shortness of breath (section 1.2.5).[41] Patients who have a ppoFEV_1% AND/OR ppoDLCO% <40% are categorised moderate/high risk- the origins of this pathway will be explored further in section 4.4.1. *Functional assessment* should be

performed in this 'high risk' group to further stratify risk and determine the impact of surgery. Following this, moderate risk patients need to be informed of mild/moderate risk of post-operative dyspnoea (consider split function testing in this group if suspicion of ventilation perfusion mismatch to allow more accurate prediction). Patients in the high-risk group are at increased risk of post-operative ventilator dependence, and should be considered for lung parenchymal sparing surgery. Functional assessment is covered in section 4.4. Most authors acknowledge the importance of predicting post-operative disabling shortness of breath, but recognise the challenges, limitations and the need for improvement.[18, 51, 52]

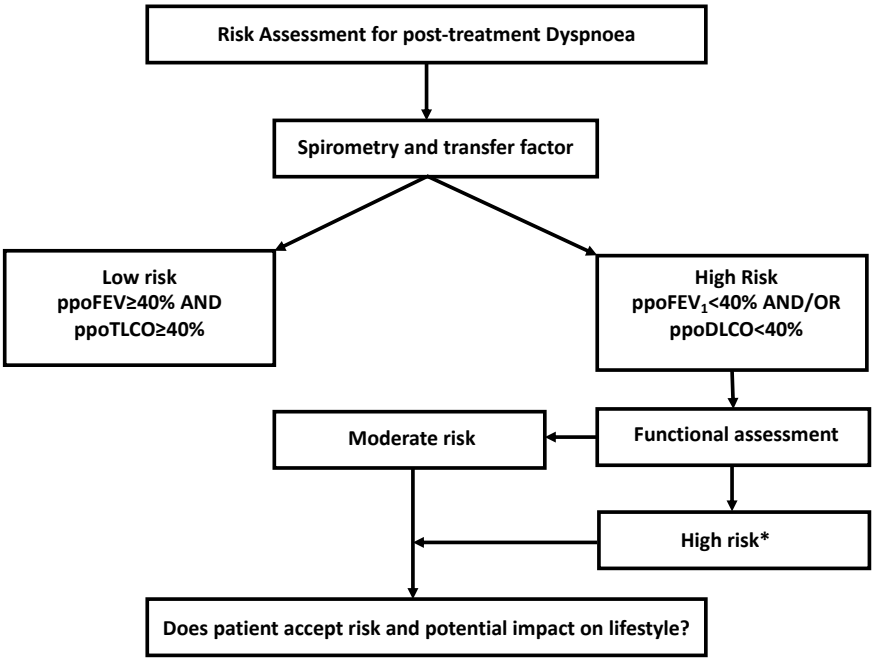

Figure 10 - Risk Assessment for post-operative dyspnoea.
Redrawn from British Thoracic Society Guidelines 2010.[18] ppoFEV$_1$ = Predicted post-operative forced expiratory volume, ppoDLCO = Predicted post-operative diffusion capacity.). * = Patients in high risk group are at increased risk of post-operative ventilator dependence

4.3 Pulmonary function testing and calculation of predicted post-operative values

As already described, prediction of post-operative dyspnoea following lung resection is centred around pulmonary function testing. This section will explore the values obtained from pulmonary function testing and their usefulness in the prediction of post-operative dyspnoea.

4.3.1 Forced vital capacity/ Forced expiratory volume

Forced vital capacity (FVC) is the total amount of air that can be exhaled in a single maximal exhalation. FEV is the amount of air a person can exhale during a forced vital capacity breath in a given timescale. The amount of air exhaled during the first second of this breath is often used and is known as the forced expiratory volume in one second (FEV_1). These values are obtained with spirometry and are repeated with and without bronchodilator therapy. The values generated are provided in litres (FEV_1), or as a percentage compared with the predicted value derived from population studies varying with gender, age, height and race (FEV_1%). A reduced FEV_1% signifies obstructive lung conditions such as asthma, bronchiectasis and chronic obstructive pulmonary disease (COPD). The ratio of FEV_1/FVC can be calculated, reflecting the amount of air exhaled in the first second divided by all of the air exhaled during a maximal exhalation. This is used to diagnose COPD: an FEV1% < 80 and a post bronchodilator FEV_1/FVC ratio of <0.7 confirming the diagnosis (As described by the Global Initiative for COPD).[148]

4.3.2 Diffusion capacity

Diffusion capacity is calculated from the quantity of carbon monoxide absorbed in a given unit of time and is measured in mmol $kPa^{-1}min^{-1}$. As with FEV_1, it can also be reported as a percentage derived from population studies (DLCO%). It provides a measure of alveolar capillary function and estimates how much oxygen is able to pass from the lungs into the blood stream. Most DLCO values are generated using a single breath technique and are reported along with the spirometry values, according to the standards set by the American Thoracic Society and European Respiratory Society in 2017.[149] Patients inhale a maximal breath of a test gas; a mixture of a tracer (commonly helium), 0.3% carbon monoxide, carbon dioxide, oxygen, and nitrogen. This vital capacity breath is held for 10 seconds and then exhaled into a sample bag after which the rate of diffusion of carbon monoxide can be estimated. The change in carbon monoxide concentration is then multiplied by the total lung capacity value to calculate the diffusing capacity. Interest in DLCO started in 1988 when Ferguson et al[150] suggested it could be a predictor of outcome in lung resection.

The implications of a low DLCO are well recognised, as conditions that interfere with the total functioning surface area of the alveolar-capillary interface are numerous: COPD (including emphysema and bronchitis), smoking, interstitial lung disease, heart failure and pulmonary vascular disease. The causes and clinical significance of a *high* DLCO are less well discussed and have been attributed to large lung volumes, obesity and asthma.[151] The normal range for DLCO is 80-120% of its predicted value for men and 76% - 120% for women.[152]

4.3.3 Predicted post-operative lung volumes

For patients undergoing lung resection, pre-operative measurements are used to predict post-operative values based on the extent of surgery anticipated.[18] The predicted post-operative value can be expressed as a percentage of normal values for the patients FEV_1 or in litres ($ppoFEV_1$% or $ppoFEV_1$, respectively).

Calculation of post-operative lung function is usually performed using segment counting, with both lungs composed of 19 segments; Left upper lobe = 5, Left lower lobe = 4, Right upper lobe = 3, Right middle = 2, right lower lobe = 5 (Figure 11).[18] The total number of unobstructed, functional segments is best obtained from imaging. The number of functional segments to remain following resection is calculated and then divided by the total number of functional segments pre-operatively, to provide a predicted post-operative value. This value is then multiplied by pre-operative pulmonary function to obtain predicted post-operative values of FEV1 and DLCO (Equation 2).

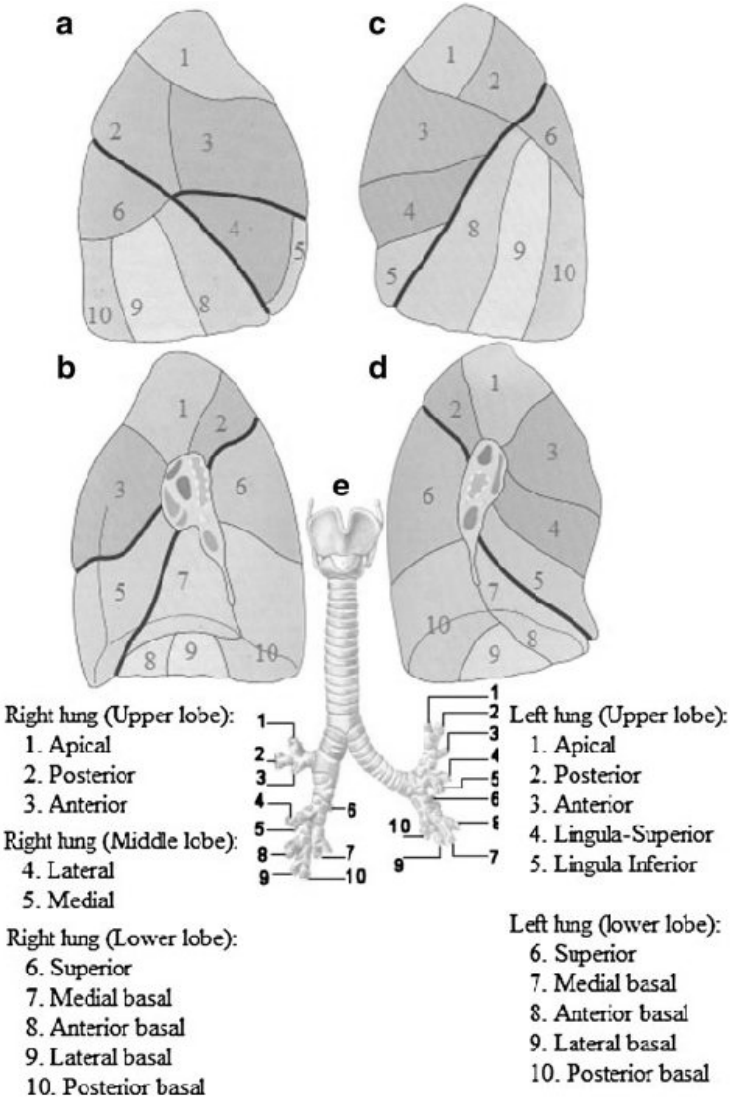

Figure 11 - Bronchopulmonary segments.
Total of 19 segments; Left upper lobe = 5, Left lower lobe = 4, Right upper lobe = 3, Right middle = 2, right lower lobe = 5. Taken from Kabir et al 2010.[153]

$$ppo\ value = R \times \left(\frac{Pre - operative\ value}{T}\right)$$

Equation 2 – Predicted post-operative calculation. T = total number of functioning segments prior to resection, R = number of functional segments following lung resection.

Predicted post-operative FEV_1 and ppoDLCO are strongly associated with measured values.[154] In 55 patients, Markos et al[154] confirmed this association in patients undergoing lung resection using quantitive perfusion lung scans (explored in greater detail in section 4.4.1). The predictions of post-operative function were associated with measured values for both FEV_1 ($r=0.86$, $p<0.001$) and DLCO ($r=0.8$, $p<0.001$). Interestingly, when this group is subdivided into whether lobectomy of pneumonectomy was performed, there was a significant but weaker association

between predicted and measured post-pneumonectomy FEV_1 values ($r=0.51$, $p<0.05$). Post-pneumonectomy measured and predicted DLCO values were not associated ($r=0.17$, $p>0.05$). In this group patients were more likely to have received aggressive radiotherapy; in the postpneumonectomy patients who *did not* receive radiotherapy there was better association between predicted and measured postoperative DLCO values ($r=0.84$, $p<0.01$). This suggests aggressive radiotherapy may have an effect on the actual post-operative values.

Predicted post-operative FEV_1 values may not be accurate in patients with obstructive disease and cannot be used alone for patient selection; Brunelli et al found that patients *with* COPD had lower losses of FEV_1 and DLCO compared with patients *without* COPD at 3 months after lobectomy for lung cancer.[155, 156] Conversely, using an absolute value for $ppoFEV_1$ (in contrast to $ppoFEV_1\%$) may prevent the elderly, those with small stature and females from having curative surgery. Expressing $ppoFEV_1$ as a percent predicted post-operative is desirable.[155]

In 376 patients undergoing lung resection, Ferguson et were unable to demonstrate association between ppoDLCO and $ppoFEV_1$. In this cohort ppoDLCO, but not $ppoFEV_1$ was a strong predictor of post-operative pulmonary complications (PPCs) and death.[157] This reinforces that FEV_1 and DLCO measure different aspects of lung function and should be considered independently when evaluating a patient for surgery.[156]

Diseased lung is not uniform and therefore any given lung segment can have varying function. Consequently, any patient with borderline $ppoFEV_1$ or ppoDLCO, following simple segment counting, or those with a bronchial lesion obstructing flow, may require further investigation to quantify regional contribution from the lung requiring surgery. This is best achieved with ventilation perfusion (V/Q scan) scanning. V/Q scanning is more accurate than the segment counting method described above; ventilation is measured using inhaled Xenon 133 and perfusion measured by technetium 99 labelled macro aggregates.[158]

Many studies have examined the role of $ppoFEV_1$ in risk prediction and patient selection for surgery. Initially, Olsen et al[159] observed patients with a $ppoFEV_1$ of as low as 0.8 litres could tolerate resection. This was in a very small study of only 13 patients in 1974.

FEV_1 (and FEV_1%) is reduced following surgery and can accurately be predicted.[105, 160] However, several prospective studies have questioned the ability of FEV_1% or $ppoFEV_1$% to predict dyspnoea, quality of life or PPC (explored further in Section 4.4).[25, 32, 161-163]

4.4 FEV_1 and DLCO in risk prediction

The use of FEV_1 and DLCO is well established to quantify the risk of post-operative dyspnoea in clinical practice within the UK, Europe and America. This section will look at the evidence and literature supporting the use of pulmonary function testing in this setting. The utility of FEV_1 and DLCO to predict mortality, post-operative quality of life and post-operative pulmonary complications will then be explored.

4.4.1 FEV_1 and DLCO to predict post-operative dyspnoea

While FEV_1 and DLCO are described extensively in the literature as independent predictors of post-operative *'outcome'*, little evidence exists to demonstrate utility as predictors of post-operative dyspnoea, despite guidelines advocating their use for this.[18] This is consistent across all current and major national guidelines by the British Thoracic Society (BTS), the National Institute for Health and Care Excellence (NICE), the European Respiratory Society (ERS), the European Society of Thoracic Surgeons (ESTS) and the American College of Chest Physicians (ACCP) [18, 51, 164, 165] is a similar stepwise approach to risk assessment. All guidelines acknowledge the difficulties of predicting post-operative dyspnoea. The stepwise approach begins with measuring FEV_1 and DLCO (and calculating $ppoFEV_1$% and ppoDLCO%) before addressing any potentially modifiable risk factors and reassessing. Spirometry is recommended as an initial screening test and as a gateway to more sophisticated testing for *high-risk* patients, such as CPET testing. [18, 166, 167] Guidelines traditionally recommend an optimum cut-off point of $ppoFEV_1$% and ppoDLCO% of 40% to determine if a patient is at high risk of dyspnoea. Following spirometry, patients with a calculated $ppoFEV_1$% *and* ppoDLCO% >40% are deemed low risk, so do not require further investigation. Patients with a calculated $ppoFEV_1$% <40% *and/or* a calculated ppoDLCO% <40% are described as being at moderate/high risk of post treatment dyspnoea and warrant further functional assessment. Predicted post-operative FEV_1% pre-dates

ppoDLCO% as a method of risk prediction of dyspnoea,[145] with ppoDLCO% first (if true) being suggested as useful in outcome prediction in 1963.[168] Table 8 summarises studies that explore DLCO and FEV_1 as predictors of post-operative dyspnoea.

The landmark study which helped derive the 40% 'high-risk' cut-off value was from original work conducted by Markos in 1989 from a single centre in 55 consecutive patients.[154] This is a prospective study with no power analysis performed for the primary outcome of ascertaining predictors of mortality and morbidity after lung resection. This study includes a limited sample size of patients and follows them through their lung resection surgery into the recovery period. Two patients were inoperable and six had a thoracotomy without resection, despite this they appear to be included in the final analysis. Cardiopulmonary complications occurred in 16 patients within 30 days.

One of the major criticisms of this important study is the lack of predictive statistical analysis.

> *'Using Wilcoxon ranking test, the ability of FEV1, DLCO and their predicted post-operative values to individually predict post-operative complications, death or respiratory failure were assessed.'*
>
> *Markos in 1989* [154]

Wilcoxon ranking test is a test of association, not prediction. Markos claimed DLCO% predicted post-operative complications but this was in the lobectomy group (n=32), with 76% correct prediction and no further prediction statistics or cut offs given. When *all* patients were combined (lobectomy and pneumonectomy), no variables were predictive of post-operative complications. Post-operative predicted FEV_1% was predictive of post-operative complications in all patients undergoing surgery (68% correct prediction, $p<0.05$). No odds ratios or confidence intervals are provided alongside this statement. Forty-seven patients with a ppoFEV_1% ≥40% had no complications. In patients with ppoDLCO% <40% (n=6), the mortality rate was 50%. In patients with ppoFEV_1 <40% (n=6), the mortality rate was also 50%. It is interesting why Markos claims these variables

can predict any post-operative outcome. Furthermore, any association shown may not be clinically significant given the low numbers of patients and high risk of error within this study, the results of which, have not been replicated. Despite measuring dyspnoea and quality of life, these are not further explored for association or prediction. Markos et al concluded a patient is suitable for surgery with a lower risk of complications if ppoDLCO% >40% and suggest when ppoFEV_1% and ppoDLCO% are both <35%, the predicted risks of surgery may be unacceptably high.[154] While association is shown between estimated and actual pulmonary function tests and between complications and pulmonary function tests,

> *'Logistic regression was not used to assess prediction of individual complications because of the small number of each'*
>
> *Markos in 1989* [154]

Markos et al provided a table with data for each patient allowing extrapolation and further analysis and to illustrate the association between those with and without post-operative complications and ppoFEV_1%/ppoDLCO% (Figure 12 and Figure 13). There was no difference between median ppoFEV_1% in patients with and without post-operative complications. There was a difference in median ppoDLCO% between patients with and without post-operative complications.

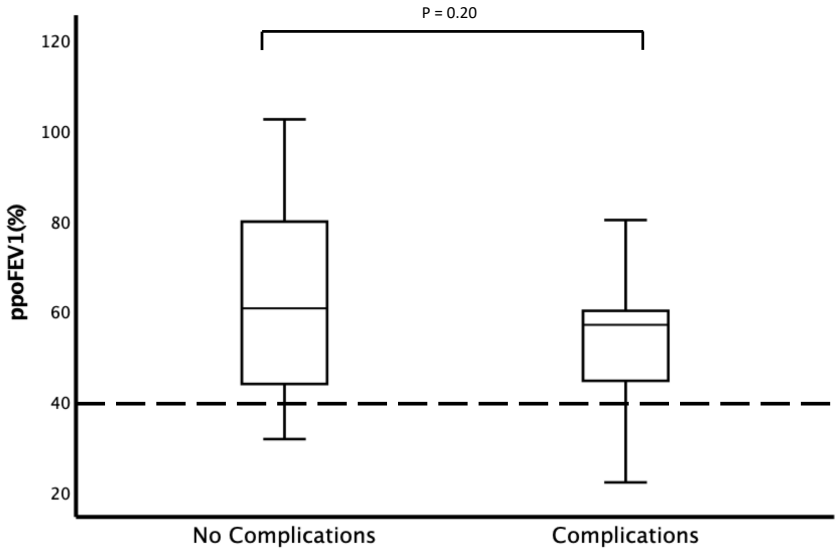

Figure 12 – Predicted post-operative FEV1 in patients experiencing and not experiencing complications.
Drawn from data provided within Markos et al.[154] Comparison of ppoFEV1% in those with (n=15) and without (n=32) complications. Patients with no resection or surgery performed excluded from analysis. Mann Whitney-U test. Dashed line represents current 40% cut off for risk prediction. ppoFEV1% = predicted post-operative forced expiratory volume.

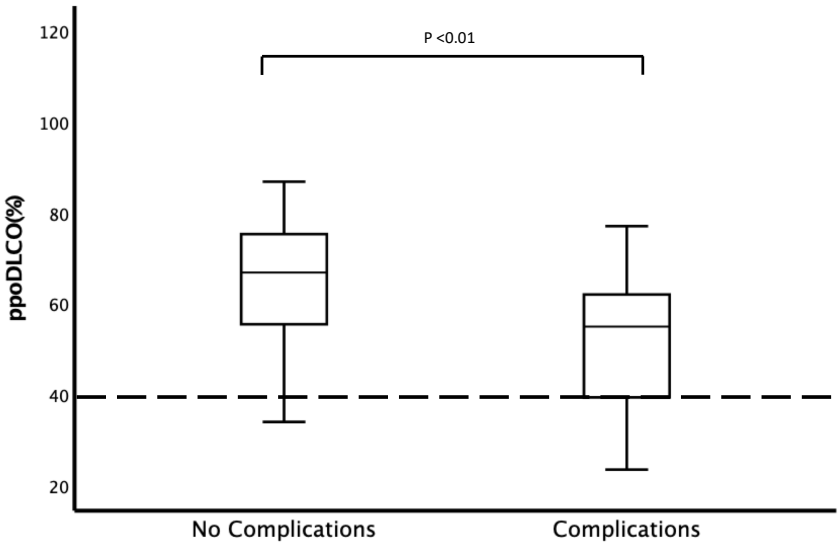

Figure 13 - Predicted post-operative DLCO% in patients experiencing and not experiencing complications
Drawn from data provided within Markos et al.[154] Comparison of ppoDLCO% in those with (n=15) and without (n=32) complications. Patients with no resection or surgery performed excluded from analysis. Mann Whitney-U test. Dashed line represents current 40% cut off for risk prediction. ppoDLCO% = predicted post-operative diffusing capacity of lung for carbon monoxide.

Using data provided on each patient within a table in the study, I performed further predictive analysis. A comparison between the incidence of complications in patients and either an FEV_1<40% *or* DLCO<40% and patients without

complications and FEV_1>40% *or* DLCO>40%, was not significant, (p=0.09, Chi-squared). The predictive values of the pre-operative lung function variables to predict post-operative complications are summarised in Table 7.

Variable	OR	Significance (p value)	Sensitivity (%)	Specificity (%)	AUROCC (95%CI)
ppoFEV_1<40%	0.88	0.30	20	91	0.55 (0.37 – 0.74)
ppoDLCO<40%	1.66	0.07	26	94	0.60 (0.42-0.79)
ppoFEV_1<40% *or* ppoDLCO<40%	1.25	0.1	33	88	0.60 (0.42 – 0.79)

Table 7 – Post-hoc analysis by the author of extrapolated data
Logistic regression using FEV_1% and DLCO% to predict post-operative complications. taken from Markos et al.[154] FEV_1 = Forced expiratory volume in one second, DLCO = Carbon monoxide diffusing capacity, OR = Odds ratio, ppo = Predicted post-operative, AUROCC = Area under the receiver operator characteristic curve

Table 7 displays the inability of ppoFEV_1<40% or ppoDLCO<40% to predict post-operative complications. The point of maximum sensitivity and specificity for ppoDLCO% falls at 33% with a sensitivity of 67% and specificity of 74%. The point of maximum sensitivity and specificity for ppoFEV_1% falls at 26% with a sensitivity of 93% and specificity of 40%. There is low confidence in these value as none are statistically significance.

Markos et al also used composite endpoints to increase 'positive event rate', combining respiratory and cardiac complications. It is surprising how the results from this paper have become so influential in shaping current guidelines for the prediction of *dyspnoea* when most of the analysis concentrated on *post-operative complications*. Furthermore, the author observed a low complication rate and performed no predictive statistics. Markos et al recorded pre-operative dyspnoea scores for patients but did not use this in any of the analysis. Yet in the commentary Markos et al claim;

> '*Dyspnoea grade remained stable from 3 to 12 months, suggesting that if post-operative survival is predicted no major respiratory disability should develop within at least 1 year of surgery.*'
>
> *Markos et al 1989* [154]

This is a strong conclusion to draw from a small study which was not designed to answer the specific question regarding prediction of long-term post-operative dyspnoea. Therefore, we should be cautious in accepting the predictive utility of DLCO or FEV_1 for either post-operative respiratory failure *or* long-term dyspnoea.

Another influential study seeking to determine whether low DLCO% is a predictor of high post-operative mortality and morbidity is Bousamra et al.[169] This is a comparatively large study of 325 patients in a single centre undergoing major pulmonary resections. Once again, no predictive analysis was performed despite the aim of the study "to determine whether low diffusion capacity is a predictor of high post-operative mortality and morbidity after major lung resection." In keeping with previous work this study demonstrated those with a low pre-operative DLCO% had an increased post-operative respiratory complication rate (18% Vs 10%, $p<0.05$) and were more breathless than prior to surgery, ($p<0.01$). Low DLCO% was defined as DLCO <60% (% predicted) for pneumonectomy or <50% (% predicted) for lobectomy.

Bousamra and colleagues assessed pre-operative and post-operative dyspnoea (6-month time point) using the Medical Research Council dyspnoea scale - post-operative dyspnoea was worse in patients with low DLCO, ($p<0.01$). Lung resection resulted in significant decline in the median baseline dyspnoea scores only amongst those patients with low pre-operative DLCO% (Figure 14). This was from just nine patients within the study, seven of which underwent more extensive resection along with simultaneous radiation therapy. The cut-off of ppoDLCO% of 60% is not in keeping with any recognised value or guideline and is not justified in the manuscript. Similar to Markos et al discussed above, this study displays association between pulmonary function and dyspnoea but does not address *prediction* of dyspnoea following lung resection surgery.

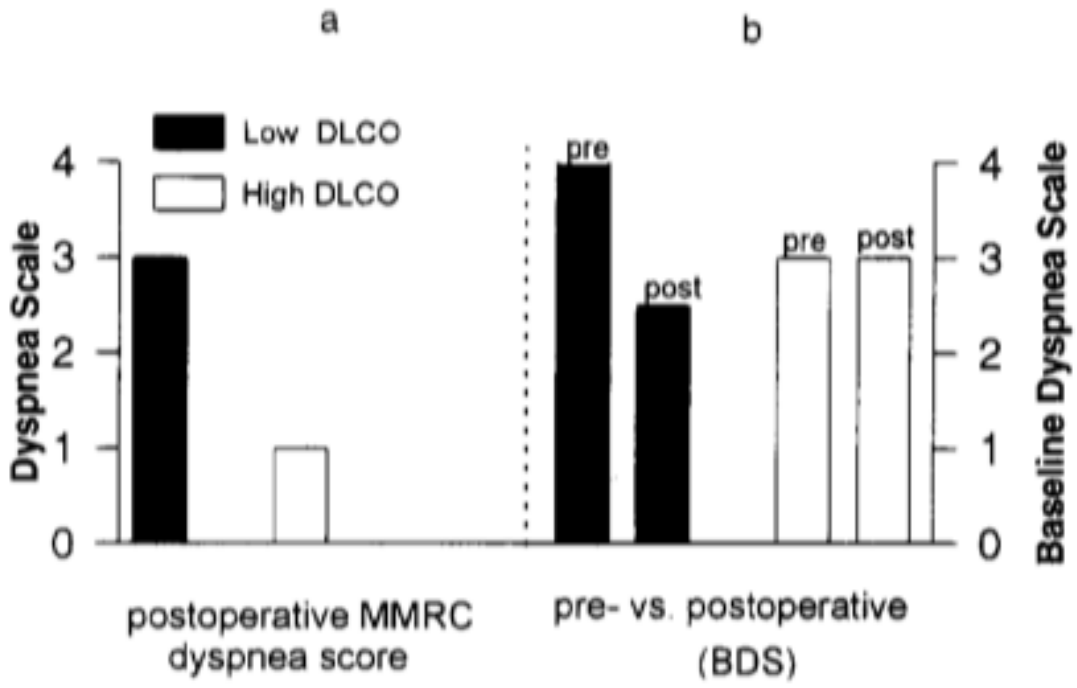

Figure 14 - Comparison of dyspnoea scores in low and high DLCO% patient groups.
a) modified medical research scores for patients with low and high diffusion capacity (DLCO%) 6 months after resection. Dyspnoea (High MRC score) is greater in the low DLCO% group than high DLCO% group, ($p<0.01$, Wilcoxon rank sum). Low DLCO% was defined as DLCO < 60% (% predicted) for pneumonectomy or <50% (% predicted) for lobectomy. b) Pre-operative and post-operative median baseline dyspnoea scores (BDS) for patients with low and high DLCO. Post-operative low DLCO% group dyspnoea score is less than pre-operative low DLCO% group dyspnoea score, ($p<0.01$, Wilcoxon rank sum). Taken from Bousamra et al.[169] (n=325)

One of the few studies to support FEV_1% or DLCO as predictors of post-operative dyspnoea is by Foroulis et al.[170] In a small study of 35 patients undergoing lung resection surgery for cancer, they attempted to identify parameters determining the clinical state of a patient following lung resection surgery. Dyspnoea on exertion was categorised into four stages; one being dyspnoea on heavy exertion and four being dyspnoea on minimal exertion. FEV_1% and FVC were lower in patients with class three and four dyspnoea, than patients with class one and two dyspnoea (Figure 15). This displays an association between pulmonary function and post-operative dyspnoea. Following multiple regression analysis, FEV_1% and FVC% were independent predictors of post-operative dyspnoea (FEV_1% OR = -0.45, 95%CI -0.002 – 0.001, $p<0.01$) (FVC% reduction OR = 0.42, 95%CI 0.09 - 0.43, $p<0.01$). While this *does* demonstrate pulmonary function as a predictor of post-operative dyspnoea, these findings were in a small number of patients (n=35) who all underwent pneumonectomy. Only 45% of the variance in dyspnoea could be

explained by pulmonary function and RVSP, (R^2 = 0.45). The author did not do any further AUROCC analysis or sensitivity/specificity analysis.

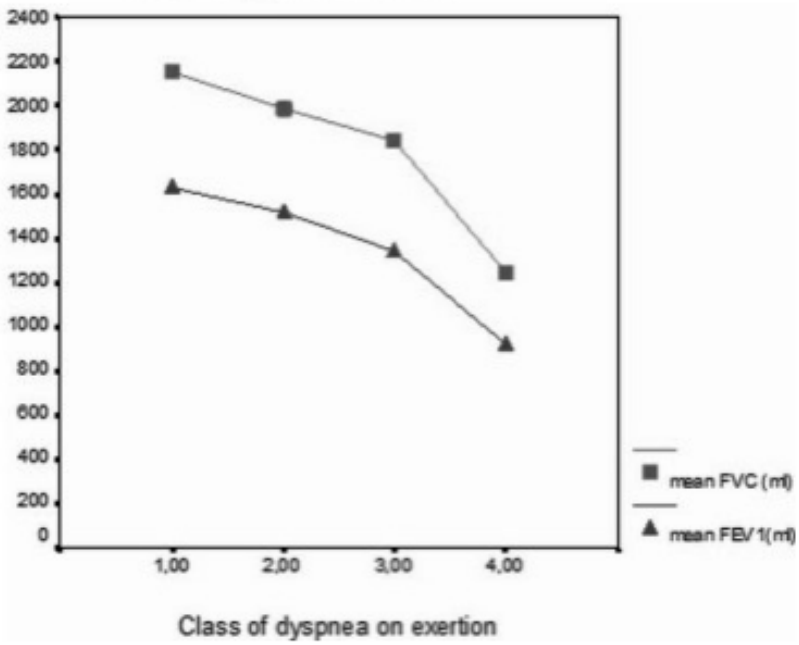

Figure 15 – Class of dyspnoea versus pulmonary function.
Changes in mean FVC and FEV_1 values (mls) from pre-operative values according to class of dyspnoea on exertion (y-axis measured in mls). Dyspnoea on exertion was categorised into 4 stages; 1 being dyspnoea on heavy exertion and 4 being dyspnoea on minimal exertion. Taken from Foroulis et al 2009.[170] (n=35)

There are other general limitations about current guidelines to predict post-operative dyspnoea. The majority of data used to shape these national guidelines failed to differentiate between open and video assisted thoracic surgery (VATS) procedures, although were published around the time VATS surgery was becoming more popular in the UK. The data was also from small single centre studies or were not externally validated.[107, 150, 171, 172]. The majority of work examining FEV_1 as a predictor of functional outcome has been on patients undergoing open thoracotomy.[34, 173, 174] This is no longer a true representation of current case mix as the latest figures from the activity and audit report 2018 by the Society for Cardiothoracic Surgery in Great Britain and Ireland demonstrate VATS resection now predominates, with 56% of all lung resections performed using 'minimal access'.[175] VATS procedures are associated less post-operative complications with equivalent long-term survival in lung resection candidates.[176]

Generally, within the guidelines those patients considered borderline for surgery, with a ppoFEV_1% *or* ppoDLCO% between 30 and 40%, supplementary cardio

pulmonary exercise testing (CPET) or other low technology testing such as six minute walk testing, stair climbing or shuttle walk testing is advocated.[18, 51, 165, 177] These supplementary tests will be explored later, in section 4.5. Importantly, all guidelines appear consistent in their message that if highlighted as being 'high risk', and the patient accepts the calculated risk and its potential impact on lifestyle, surgery should still be considered.[18]

There exists a continuum of risk and benefit along which each patient sits.[18] The BTS and NICE advocate multidisciplinary management on an individual patient basis: surgical intervention with a risk of dyspnoea is *not* best evaluated with a dichotomisation into those who can or cannot undergo potentially curative surgery. The BTS, NICE and American College of Chest Physicians (ACCP) are consistent and acknowledge this difficulty with the need for multidisciplinary management and decision making on an individual patient basis.[18, 51, 165]

4.4.2 Reconsideration of 40% cut off for risk prediction

Different lower limits of ppoFEV_1% and ppoDLCO% exist depending on which guidelines are consulted. This makes consideration for surgery difficult for patients falling within the 30 - 40% ppoFEV_1%/ppoDLCO% group, as many could tolerate potentially curative surgery.[18, 178]

The British Thoracic Society guidelines in 2001 suggested a cut-off limit of 40% ppoFEV_1% *and/or* ppoDLCO% to predict post-operative dyspnoea.[145] This 40% cut off value was shown to be poorly associated with QoL after surgery (discussed later in section 4.4.3.2).[172] This poor association between lung function and QoL resulted in the subsequent 2010 guideline development committee for the BTS questioning this as a lower limit acceptable for surgery.[18] However, despite these findings, the 2010 guidelines remained fixed at the 40% cut-off value (Section 4.4.1). It appears one of the main influences on not changing the 40% cut-off value is because of a study by Brunelli et al in 2007.[156] This study examined outcome following lung resection (mixture lobectomy and pneumonectomy) in 253 patients with ppoFEV_1% or ppoDLCO% ≤30% and displayed an overall mortality rate of only 4% (n=10). This paper favoured lowering the cut-off to ≤30% but as it was not powered to study post-operative dyspnoea it failed to impact the updated 2010 guidelines.[18]

Conversely, the ERS and ESTS clinical guidelines on fitness for radical therapy in lung cancer patients[177], the American College of Chest Physicians (ACCP) [165] and The NICE guidelines for lung resection (2019)[51] recommend 30% as a new cut-off value for risk stratification all having advocated 40% cut-off in previous editions. Another study by Brunelli et al[179] suggested ppoFEV_1% was *not* a reliable predictor of complications and those with a ppoFEV_1% <40% had a low mortality rate (4.8%). Brunelli hypothesised these findings were explained by the 'lung volume reduction effect' that can reduce the functional loss in patients with existing airflow limitation. In lung resection, removal of tissue can benefit patients by working in a similar fashion to lung volume reduction surgery, reducing the size of an overinflated lung and allowing expansion of the remaining often more functional lung. Given several studies have shown improvement in pulmonary function after lobectomy in lung cancer patients with severe COPD [105, 156, 180, 181], traditional operability criteria (<40% cut-off) based on pulmonary function has been questioned. This evidence, alongside improvements in peri-operative management and surgical techniques, were the main drivers for a change to the new reduced 30% cut-off.

Table 8 - Studies examining FEV_1 and DLCO as predictors of post-operative dyspnoea following lung resection.

Study	Population	Method	Outcome	Comment
Bousamra et al 1996 [169]	Retrospective n=325 Single Centre	Pre-op PFTs Low DLCO < 60% (% predicted) for pneumonectomy or <50% for lobectomy. MRC dyspnoea at pre-op and >6 months post-op Low DLCO group n=62. standard DLCO group n=263	Pulmonary resection resulted in a significant decline in the median baseline dyspnoea scores in low DLCO group, using MRC score. Low DLCO had increased respiratory complication rate. (18% Vs 9.5% p=0.05)	Association demonstrated, no predictive statistics provided. Extent of resection and radiation therapy may contribute to dyspnoea. *ppoDLCO* not used. Small sample size prevented full analysis of dyspnoea/low DLCO group 60% cut-off not conventional
Foroulis et al 2009 [170]	Prospective n= 35 Single Centre	Pre-op and 6 months; PFTs, Pre-operative and post-operative dyspnoea class; I = heavy exertion II = mod exertion III = mild exertion IV = min exertion Pneumonectomy n=35	Post-operative FEV_1%, FVC% and ppoFVC% were lower in patients with class III and IV dyspnoea than in patients with class I and II dyspnoea (p = 0.002 and p = 0.003) Predicted FEV_1 less than 1.4L connect with increased dyspnoea FEV_1% and FVC% were independent predictors of post-operative dyspnoea (FEV_1% OR=-0.45, 95%CI -0.002 – 0.001, p<0.01) (FVC% reduction OR=0.42, 95%CI 0.09-0.43, p<0.01)	Association between pulmonary function and post-operative dyspnoea Patients with no dyspnoea after surgery may have post-operative FEV_1% <40% FEV_1% and FVC% independent predictors of post-operative dyspnoea

Study	Population	Method	Outcome	Comment
Handy et al 2010 [182]	Prospective n=198 Single centre	Low Risk (LR) n=166 High Risk (HR) n=32 LR = ppoFEV_1 and DLCO >40% HR when either <40%	High Risk at 6 months: less FEV_1% decrease (-2.2% vs -10.1%; p=0.02) & greater dyspnoea MRC score (0.76 vs 0.31; p=0.03) No difference in absolute values or magnitude of change in quality of life for HR vs LR preoperative vs post-operative.	Selection bias -not many high-risk patients get operations. Association demonstrated, no predictive statistics provided.
Markos et al 1989 [154]	Prospective n=55 Single centre	Lung function and exercise capacity pre-op, 3 months and 12 months Pneumonectomy = 18 Lobectomy = 29 No resection = 6 Inoperable = 2	Post-hoc analysis done; ppoFEV_1<40% or ppoDLCO<40% to predict complications. OR 1.25, p=0.1 (AUROCC 0.60 95%CI, 0.42 – 0.79)	Predictive ability of dyspnoea based on prediction of complications. Association demonstrated, no predictive statistics provided (post-hoc analysis done) Post-hoc analysis observes no predictive ability of ppoFEV_1% or ppoDLCO% for complications, let alone dyspnoea

4.4.2.1 FEV and DLCO to predict dyspnoea: conclusion

Following a review of the literature and the evidence presented above it is reasonable to conclude that ppoFEV_1 and ppoDLCO are *not* reliable predictive markers of dyspnoea following lung resection. Very little evidence exists to support the use of pulmonary function to predict dyspnoea. To date, no robust, appropriately powered, large, multicentre studies have been conducted demonstrating FEV_1%, FEV_1, ppoFEV_1, ppoFEV_1%, ppoDLCO nor ppoDLCO can reliably predict dyspnoea following lung resection. Studies incorporate these values to predict complications or morbidity/mortality; dyspnoea being included as a *secondary* outcome or not considered at all. Yet, in all international guidelines, pulmonary function is advocated to predict long term dyspnoea. Furthermore, whilst some studies to date demonstrate association between pulmonary function and post-operative dyspnoea, they do not perform predictive analysis. It is therefore surprising that guidelines have been shaped around these results.

Increasingly, surgery is being performed with minimally invasive surgical techniques. Though unproven, less invasive surgical approaches may speed up recovery and perhaps reduce dyspnoea rates in the recovery period; smaller surgical wounds healing faster and allowing patients to mobilise earlier. Conversely, treatment for lung cancer is changing; patients with multiple co-morbidities previously refused intervention could be offered surgery, potentially causing dyspnoea rates to increase. Patients with lung function values less than recommended cut-off values *may* tolerate surgery and survive with reasonable quality of life. Despite this, national guidelines continue to use ppoFEV_1% and ppoDLCO% as risk assessment tools prior to surgery.[18, 51, 52, 177] The difficulties with pulmonary function to predict post-operative dyspnoea means new, validated, methods are needed. Identification of a sub-population at increased risk of dyspnoea could provide opportunity to intervene. Improved prediction of dyspnoea would allow targeted recruitment of patients into studies aiming to ameliorate post-operative dyspnoea following lung resection. Improved prediction would enable better quantification of the risk of dyspnoea prior to surgery, facilitating shared decision making.

4.4.3 FEV_1 and DLCO to predict post-operative mortality

FEV_1 and DLCO have been used in an attempt to predict mortality for many years.[145] Historically, a pre-operative FEV_1 of greater than 1.5 Litres for lobectomy and greater than 2 Litres for Pneumonectomy was considered suitable for surgery with an operative mortality < 5%.[183] This was based on data from the 1970's which has since been superseded by predicted post-operative percentage, rather than absolute values in litres.[18] Table 9 summaries studies examining the utility of FEV_1 and DLCO to predict post-operative mortality following lung resection. Recently FEV_1 and DLCO have been superseded by *Thoracoscore* (Section 4.2.1*)* to predict in-hospital mortality. Lung function does not feature in *Thoracoscore*, the most widely used scoring tool to quantify risk of in-hospital mortality following lung resection. At univariate analysis, Falcoz et al[142] observed a difference in pre-operative FEV_1% in those who were dead and alive until discharge. The value was dichotomised into < or >50% and reached statistical significance. However, following multivariate analysis, FEV_1% was not an independent predictor and does not feature within the scoring tool.

The study cited most commonly and which was one of the first to confirm DLCO and FEV_1 as predictors of mortality was Markos et al[154] which has previously described (section 4.4.1).[154] In 55 patients undergoing lung resection the author observed a 5% mortality rate at 12 months. Markos et al demonstrated the best predictor of mortality in this group was ppoFEV_1% ($p<0.01$, no OR given) in those that underwent pneumonectomy or ppoDLCO ($p<0.01$, no OR given) for the best overall predictor of mortality among all patients (lobectomy and pneumonectomy). Markos et al favoured the use of percentage of predicted value rather than absolute units to predict mortality. No power calculation was performed for the primary outcome of survival and with only 55 patients and three deaths, it is surprising how influential this paper has become. In addition, 15 patients were lost to follow up at 12 months.

In another influential study looking at the prediction of mortality, Brunelli et al[184] examined survival up to 12 years post-operatively in a prospective observational analysis of a database from a single centre, (n=296). At univariate analysis, climbing 18 meters or more on stair climb test, age, sex, FEV_1%, DLCO%, tumour stage, haemoglobin level and oncology group score were all associated with

survival and were included in the multivariable cox's proportional hazards model. This model demonstrated DLCO% (HR 0.98, p=0.02) was an independent predictor of survival. No cut-off value of DLCO% was described and this analysis did not include ppoDLCO which is now more widely used. It is not clear from the statistical methods if the reported hazard ratio was for each % change in DLCO - if so, any relative small difference may have a substantial impact on survival. If not, given the hazard ratio is close to the value of one, despite its statistical significance, this is unlikely to have any meaningful clinical predictive value.

As part of the English national lung cancer audit, Powell et al[16] retrospectively analysed 10,991 patients undergoing lung cancer surgery over a 6-year period from 2004. The aim of this study was to determine 30 and 90-day mortality following surgery, and develop a predictive scoring tool to estimate early post-operative mortality. They observed a 3% mortality rate within 30 days and 5.9% mortality rate at 90 days. The authors constructed a multivariable model including all factors associated with death at univariate analysis, before assessing the individual variable significance. Age, performance status, ppoFEV_1% (ppoFEV_1% <40% associated with death at 90 days (OR 2.32, 95% CI 1.23-4.38) and procedure type were all independently associated with post-operative death. DLCO was not measured in this population.

One of the few studies to examine long term survival and pulmonary function is by Ferguson et al[185]. In a retrospective analysis of data from a single centre in Chicago from 2014 of 854 patients the author demonstrated association between both ppoFEV_1%, ppoDLCO% and long-term survival. Data were collected from 1980 to 2006 and median survival following surgery was 42 months. Factors associated with long-term survival included age, cancer stage, performance status and history of previous myocardial infarction. Long-term survival was defined as up to 10 years from the time of surgery. Pre-operative FEV_1% (HR 1.04, 95% CI 1.00-1.09) and DLCO% (HR 1.04, 95% CI 1.00-1.08) were not associated with mortality (p>0.05), for each 10-point decrement. However, ppoFEV_1% (HR 1.06,95%CI 1.01-1.12, p=0.02) and ppoDLCO% (HR 1.06,95%CI 1.01-1.12, p=0.02) were independent predictors of mortality for each 10-point decrement. When pre-operative FEV_1% (HR 1.26, 95% CI 1.01-1.46, p=0.03) and DLCO% (HR 1.26, 95% CI 1.05-1.52, p=0.01) were coded as binary values (with a cut off of < or > 80%) for analysis they remained predictors of overall survival. It is surprising they selected this as a value

with no justification given as most guidelines use an alternative cut-off of 40%. Ferguson et al proposed predicted post-operative lung function was strongly associated with long term survival. However, while ppoFEV_1% and ppoDLCO% are associated with mortality these results may not be clinically significant for each 10-point decrement in lung function, as the HR is close to the value one.

Table 9 - Studies examining FEV_1 and DLCO as predictors of post-operative mortality following lung resection

Study	Population	Method	Outcome	Comment
Ferguson et al 1988 [150]	Retrospective n=237 Single centre	Lobectomy n=164 Pneumonectomy n=73	DLCO the strongest predictor of mortality ($p<0.01$) and was the sole predictor of post-operative complications ($p<0.05$). No further predictive stats Patients with a DLCO <60% had mortality of 20% and pulmonary complications of 40%	DLCO is an important predictor of post-operative morbidity following lung resection even in patients with normal FEV_1
Ferguson et al 2008 [186]	Retrospective n=1046 Single centre	Three outcomes identified; operative mortality, pulmonary morbidity, overall morbidity. Two groups created, with and without diagnosis of COPD	Overall mortality – 5.8% Pulmonary morbidity - 14.0% Overall morbidity – 31.4% Pulmonary morbidity (OR 0.84, 95% CI 0.75-0.94, $p=0.002$,) and operative mortality (OR 0.66, 95% CI 0.52-0.86, $p<0.001$) were related to ppoDLCO%	ppoDLCO% was the strongest predictor of operative mortality FEV_1% not related to mortality DLCO is a predictor of mortality regardless of FEV_1 values
Powell et al 2013 [16]	Retrospective n=10,991 Database review	2004 – 2010 English hospitals 30-day and 90-day mortality	30-day mortality - 3% 90-day mortality - 5.9% ppoFEV_1% <40% associated with death at 90 days (adjusted odds ratio 1.48, 95% CI 1.13 to 1.95)	Decreasing ppoFEV_1% <40% value associated with death within 90 days of surgery ($p<0.01$)
Markos et al 1989 [154]	Prospective n=55 Single centre	Lung function and exercise capacity pre-op, 3 months and 12 months. Pneumonectomy n=18 Lobectomy n=29 No resection n=6 Inoperable n=2	3-month survival n=55 (100%) 12-month survival n=40 (72%) 30-day mortality n=3 (5%) ppoFEV_1%($p<0.01$) & ppoDLCO($p<0.01$) best 'predictors' of mortality. Those with <40% ppoFEV_1% (n=6, range 22-37%) had 50% mortality	ppoFEV_1% best predictor of death for pneumonectomy group ppoDLCO best overall predictor of death Association demonstrated only, no predictive statistics provided

Study	Population	Method	Outcome	Comment
Brunelli et al 2012 [184]	Prospective Observational n=296 Single centre	Survival up to 12 years. Thoracotomy n=296	DLCO (HR,0.98; p=0.02), was an independent prognostic factor of survival FEV_1% (p=0.1) was not a significant independent predictor of survival	FEV_1% associated with survival but did not independently predict survival
Ferguson et al 2014 [185]	Retrospective n = 854 Single centre	Data evaluated for all-cause mortality Long term survival up to 10 years Lobectomy n=634 Bilobectomy n=68 Pneumonectomy n=152	Median survival - 42 months Pre-operative FEV_1% and DLCO% ((p>0.05) not associated with long term survival When FEV_1% and DLCO% dichotomised into <80% or >80% they became independent predictors of overall survival ppoFEV_1% (HR 1.06,95%CI 1.01-1.12, p=0.02) and ppoDLCO% (HR 1.06,95%CI 1.01-1.12, p=0.03) were independent predictors of mortality. (For each decrement of 10%)	ppoFEV_1% and ppoDLCO% associated with long-term survival and independent predictors of mortality

4.4.3.1 FEV_1 and DLCO to predict post-operative mortality: conclusion

FEV_1 and DLCO have been historically used to predict post-operative mortality but now *Thoracoscore* has been accepted as the gold standard in prediction of in-hospital mortality. Early studies to support FEV_1 and DLCO to predict mortality have small patient numbers and the results have not been replicated in larger studies. There has been evidence to suggest ppoFEV_1% may be able to predict long term survival in addition to peri-operative mortality. The difficulty in predicting mortality following lung resection is widely recognised; no tool will ever perfectly describe the risk of post-operative death.

4.4.3.2 FEV_1 and DLCO to predict post-operative quality of life

As described in section 1.2.3, lung cancer resection is associated with more decline in quality of life (QoL) compared to other cancers and conditions.[187] QoL following lung resection has had growing interest in recent years;

> *'Quality of life is as important as quantity of life.'*
>
> *Pompili et al 2011* [188]

Table 10 summarises studies to date that explore pulmonary function in the prediction of post-operative QoL. Similar to studies concerning prediction of mortality and dyspnoea, much of the evidence for lung function in prediction of QoL is based on association rather than prediction.

When assessing a patient's QoL, a validated instrument should be used with the aim to optimise care delivered to improve outcomes.[18, 51] Patients following lung resection for cancer experience a shorter life expectancy and reduced quality of life when compared to age-matched peers.[27] Despite this frequent reduction in QoL, only two questionnaires are validated in this population.[18] Clinicians or researchers rarely include QoL when attempting to predict risk and suitability for surgery.

There are many QoL measurement tools, meaning it is difficult to know which is best in any clinical setting. QoL assessment comprises broadly of two main domains; physical and mental components. For example, the ability to ambulate

is essential to perform activities of daily living, influencing both physical and mental wellbeing. This is why physical debility is perceived as more important than pulmonary complications by patients with lung cancer.[30]

In a prospective study of 156 patients, Brunelli et al[25], demonstrated that candidates for lung cancer resection had worse QoL at 1 month when compared to pre-operative values, ($p<0.0001$). There was no association between pre-operative FEV_1 or DLCO and QoL ($r<0.2$, $p=0.9$, and 0.8 respectively). Interestingly, this study also demonstrated those patients deemed 'high risk' (as defined by ppoFEV_1% and ppoDLCO% <40%) could safely undergo lung resection surgery and have reasonable quality of life scores post-operatively. These high-risk patients (n=12) had similar residual QoL scores at 3 months compared with their 'low risk' counterparts, ($p=0.3$). Brunelli et al used the Short Form 36 (SF-36) Item health survey quality of life tool [B], which is regarded as reliable tool to detect a clinically significant difference in physical and mental quality of life. One of the few criticisms of this study is the drop-out rate (16%) during the follow up period - these patients are likely to have had the *worst* functional outcome and be most affected by surgery. Failure to include them in the analysis may give a skewed impression of post-operative quality of life. No sensitivity analysis was performed in this study to assess the impact of the dropout cohort and no predictive statistics were performed. These results should therefore be interpreted with this in mind.

Ilonen et al[162], prospectively explored the ability of pre-operative pulmonary function to predict post-operative QoL outcomes using the 15-D health related QoL tool [C]. They examined QoL in 53 patients at 3, 12 and 24 months post-operatively. Decreased values of QoL were observed between patients with FEV_1%<70% (n=18) and >70% (n=30) in the breathing domain of the QoL score

[B] Short Form 36 Health Survey is a 36-item patient reported QoL questionnaire. It consists of 8 scaled scores which are weighted sums of the questions in their section. Each scale is transformed into a 0-100 scale. The lower the score the more disability. The higher the score the less disability. The eight sections are vitality, physical functioning, bodily pain, general health perception, physical role functioning, emotional role functioning, social role function and mental health. A version two was produced by developers with different wording and layout.[189]

[C] 15D Health related quality of life tool is a generic 15-dimensional standardised self-administering quality of life tool that can serve as both a profile and a single index score measure. It consists of the following dimensions; moving, seeing, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms. A single index number is generated via a formula to generate the utility score i.e. the 15D score over all 15 dimensions. The maximum score is 1 (no problems on any dimension) and the minimum score is 0 (deceased). A change of >0.03 is clinically important.[162]

($p<0.01$) and total score ($p<0.01$) at three months. No association existed between FEV_1 or DLCO and the QoL score pre-operatively or at three, 12 or 24 months. During the study, the author had notable drop-out rates; one patient at the three-month time point, six further patients at 12 months and another six patients at 48 months-leaving a total of 36 patients (67%) for the final analysis. This was recognised by the author as patients with advanced cancer generally do not complete and return surveys in the later stages of their disease, [190] therefore this would tend to overestimate QoL because respondents will be in better health. However, no sensitivity analysis was performed. Conversely, patients without disease progression may also not complete surveys, preferring to put treated cancer behind them. The pre-operative questionnaire in this study was performed only one day before surgery which may have influenced the results, with patients being optimistic about upcoming surgery. These results agree with Brunelli et al, (discussed above) that pre-operative pulmonary function was not predictive or associated with post-operative QoL scores.[25]

In 2011, Pompili et al[54] prospectively examined predictors of quality of life in 172 patients undergoing lobectomy using the SF-36 to measure emotional and physical components of QoL. Predicted post-operative FEV_1% was associated with *emotional decline* ($p<0.01$). Though DLCO and FEV_1 approached significance at univariate analysis, pre-operative pulmonary function was not independently associated with post-operative QoL.

Another well cited study investigating if pulmonary function in this setting was performed by Handy et al.[32] In a prospective study of 131 patients using the SF-36 questionnaire, the author demonstrated a low pre-operative ppoDLCO% was *associated* with poor post-operative QoL. Predicted post-operative DLCO% was divided into three categories; <45%, 45-75% and >75%. Using multivariate analysis of variance testing, the authors were able to show patients with a ppoDLCO <45% experienced significant post-operative decline in QoL (post-operative role functioning, physical and bodily pain), ($p<0.05$) (Figure 16).

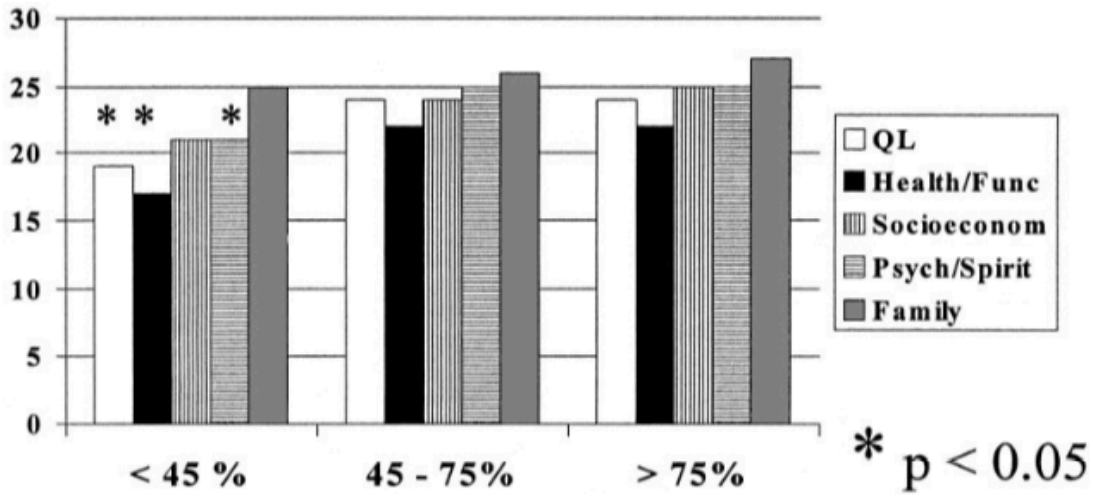

Figure 16- Quality of Life six months after lung cancer surgery based on pre-operative DLCO category (<45%, 45-75%, >75%).
Taken from Handy et al[32]. QL = Overall quality of life. Health/func = health and functioning subscale, Socioeconom = socioeconomic subscale, Psych/spirit= psychological/spiritual subscale, Family = family subscale. (n=131)

While there was a difference between the two groups, ppoDLCO% was not an independent predictor of QoL: no predictive statistics were performed and only association was displayed. Handy et al inappropriately conclude -

'DLCO, not FEV_1 predicts poor post-operative quality of life...'

Handy et al 2002 [32]

In 2005, Win et al[191] studied 110 patients undergoing lung resection in a single centre in the UK. QoL was measured using the EORTC pre-operatively and at one, three and six months post-operatively. QoL data were available for all 110 patients pre-operatively, 94 (85%) at one month and 83 (75%) at three and six months post-operatively. QoL deteriorated at one month after surgery (p=0.001) but had returned to pre-operative levels by three months, (p=0.93). In an exploratory analysis, ppoDLCO% was weakly associated with global QoL at 6-month time point (r=0.22, p=0.05). No other measures were correlated with QoL at six months post-operatively. However, this is once again a test of association and not prediction, much like the rest of the evidence explored within this chapter. Given only a weak association was demonstrated, Win et al appropriately conclude-

'Pre-operative percentage predicted DLCO was suggestive of a worse post-operative QoL at 6 months…'

Win et al 2005 [191]

Table 10 – Studies examining FEV and DLCO as predictors of quality of life following lung resection.

Study	Population	Method	Outcome	Comment
Brocki et al 2015 [192]	Prospective n=78 Single centre	Health related quality of life (SF-36) and peri-operative lung function Wedge resection n=12 Lobectomy n=54 Bi-lobectomy n=7 Pneumonectomy n=5 Thoracotomy n=60	Higher ppoFEV_1% associated with better physical functioning (SF-36) 3 weeks post-operatively (95% CI 0.04 -0.25, p=0.005).	High proportion of thoracotomy patients limits generalisability to current practice Association demonstrated, no predictive statistics provided
Brunelli et al 2007 [25]	Prospective n=156 Single centre	Health related quality of life (SF-36) and peri-operative lung function Pre-op,1 month and 3 months. Lobectomy n =144 Pneumonectomy n=12	QoL not associated with ppoFEV_1%, ppoDLCO% (r<0.2, p=0.9,0.8 respectively) ppoFEV_1 <40% or ppoDLCO < 40% defined as 'High risk' subgroup: 'High risk' patients had postoperative physical and emotional quality of life scores similar to those observed in younger and fitter patients	Functional capacity forms basis of QoL. No association between QoL and FEV_1 and DLCO.
Greillier et al 2007 [193]	Prospective n=94 Single centre	Peri-operative EORTC (QLQ-C30) and peri-operative lung function Lobectomy n=88 Pneumonectomy n=6	Pre=operative DLCO% associated with improved physical (r=0.357, p=0.03) and social wellbeing (r=0.387, p=0.02).	Weak association between DLCO% and post-operative QoL outcomes No association between FEV_1 and post-operative QoL outcomes. Association demonstrated, no predictive statistics provided

Study	Population	Method	Outcome	Comment
Handy et al 2002 [32]	Prospective n = 131 Single centre	Health related quality of life (SF-36) and peri-operative lung function Sub groups: DLCO<45%, 45-75% and >75%. Lobectomy n=102 Pneumonectomy n=10 Wedge n=5 Segmentectomy or no resection n=14	DLCO <45% subgroup significant difference with pre and post-operative QoL scores ($p<0.05$), for most but not all indices of QoL.	Association demonstrated, no predictive statistics provided
Ilonen et al 2010 [162]	Prospective n=53 Single centre	QoL (15D score) peri-operative lung function Pre-operative/3/12/24 months. Lobectomy n=49 Bi-lobectomy n=4	No association between pre-operative PFTs and post-operative QoL No association with between FEV_1 or DLCO and the QoL score at any of the time points pre or post operatively. ($p>0.05$, t-test).	The authors conclude pre-op FEV_1 & DLCO cannot be used to predict quality of life in patients undergoing lobectomy or bi-lobectomy
Pompili et al 2011 [54]	Prospective n=172 Single centre	Health related quality of life (SF-36) and peri-operative lung function Lobectomy n=160 pneumonectomy n=12	No association between $ppoFEV_1\%/FEV_1\%$/ DLCO% and ppoDLCO% and physical decline ($p>0.05$) No association between DLCO%/ppoDLCO%/FEV_1 and emotional decline ($p>0.05$) Multi-variate logistic regression demonstrated that $ppoFEV_1\%$ is an independent predictor of emotional decline ($r= -0.03$, $p=0.04$)	No pre-operative association found between pulmonary function and global post-operative QoL

Study	Population	Method	Outcome	Comment
Win et al 2005 [191]	Prospective n=110 Single centre	Peri-operative EORTC (QLQ-C30) and peri-operative lung function Pre-operative,1,3 and 6 months Pneumonectomy n=36 Lobectomy n=65 Other n=9	QoL at 6 months after surgery was associated with pre-op DLCO ($r = 0.22$, $p = 0.05$, no further statistics given)	Association demonstrated, no predictive statistics provided

4.4.3.3 FEV_1 and DLCO to predict post-operative quality of life: conclusion

It would seem intuitive that patients with decreased post-operative pulmonary function results may have reduced quality of life. Therefore, predicting QoL from pulmonary function tests would be logical. However, FEV_1 and DLCO do not correlate well with QoL reported by patients, as evidenced within this chapter. Pre-operative lung function in the lung cancer population is a poor predictor of patient's perceptions of daily activities and overall quality of life.[25] The evidence presented within this chapter displays a consistent message that there does exist a clear reduction in QoL following lung resection, however pulmonary function testing is not a reliable *predictor* of QoL deterioration.[25, 32, 34, 54, 162, 191, 193-196] One of the criticisms of these studies is a lack of standardised assessment tools to measure, quantify and compare QoL in the lung cancer resection population; this may demonstrate why international guidelines advocate further work in this area.[18, 51] The peri-operative use of simple QoL questionnaires should be routine. Improved prediction of post-operative QoL could aid planning of interventions such as supportive physical and psychological programs.[25]

4.4.4 FEV_1 and DLCO to predict post-operative pulmonary complications

Despite improved patient selection, along with better surgical and anaesthetic techniques, post-operative pulmonary complications (PPCs) occur in 10-20% of patients following thoracic surgery.[197] FEV_1 and DLCO have been identified as potential predictors of morbidity and complications following lung resection for cancer [198] but are not included in major international guidelines for this purpose.[18, 51] Conflicting information exists regarding which is the best pre-operative clinical, pulmonary function or laboratory data to use for post-operative risk estimation.[199] Since the 1980's, papers have been published addressing FEV_1 and DLCO as potentially being able to define the risk of PPC's. Table 11 summarises studies examining the utility of FEV_1 and DLCO as predictors of PPC's.

Predicted postoperative values for FEV_1 and DLCO have a stronger association with post-operative outcomes than pre-operative values, especially in patients with compromised respiratory function.[155, 200] The larger the amount of functioning lung that is resected the greater the physiological impairment.[185] Kearney et al[200] observed in 331 patients that a low ppoFEV_1 was the best indicator of patients at

high risk for complications and the only independent predictor of outcome - for each 0.2 litre (L) decline in ppoFEV_1 the odds ratio for complications was 1.46 (95% CI 1.2,1.8). A ppoFEV_1 of less than 1L was the best predictor of complications ($p<0.001$), whereas *pre-operative* FEV_1 was not predictive. As ppoFEV_1 declined, a significant association was found between reduced ppoFEV_1 value and increased complication rate ($p<0.001$). When adjusted for confounders, ppoFEV_1 was the only independent predictor of complications. Only 17 patients in this study had pre-operative FEV_1 values less than 1L, which is likely to have contributed to the overall low complication rate. Kearney et al concluded ppoFEV_1 was the best indicator of patients at risk of post-operative complications.

An early study by Dales et al[201] evaluated the ability of pre-operative physiological variables to predict PPCs. One hundred and seventeen consecutive patients undergoing thoracotomy for lung cancer were recruited from two teaching hospitals and surgery performed exclusively by four surgeons. Eighteen patients had pneumonectomy, 86 lobectomy, two wedge resections and 11 had thoracotomy without any lung resection. PPCs were experienced in 43 patients (37%)[D]. The incidence of complications increased with age; patients more than 75 years of age had twice as many complications as those younger than 50. The incidence of PPCs was higher in patients with an FEV_1% <60% than patients with an FEV_1% >60% (50% Vs 21% respectively), $p<0.05$. The absolute value of FEV_1 was not associated with PPCs (values not shown by author). No justification is given for the 60% cut-off but this may reflect local risk prediction guidelines in the recruiting centre at the time of the study. Despite the authors performing no predictive analysis it is surprising they concluded:

> *FEV_1 is one of the better indicators of PPCs'*
>
> *Dales et al [201]*

[D] PPCs defined in this study as: pneumonia, atelectasis prompting bronchoscopy, pulmonary embolism, type two respiratory failure at 24 hours, air leak requiring Intercostal chest drain for more than 7 days, bronchopleural fistula, empyema, chylothorax, haemothorax requiring intervention, tension pneumothorax, lobar gangrene, mechanical ventilation greater than 3 days and oxygen demand of greater than 60% at 24 hours post operatively.

One of the largest studies exploring pulmonary function to predict pulmonary complications is published by Amar et al[197] and retrospectively analysed a database from a single centre in New York, USA (n=956 with PPCs in 121 [12.7%]). PPCs were defined[E], similar to other studies in this area. The study had large numbers, a standardised anaesthetic regime, identical surgical approach and predefined risk factors before the data review commenced. In 93 patients, no DLCO measurements were available (10% of study population). There were differences in DLCO% and ppoDLCO% ($p<0.0001$ for both) between those with and without PPCs. However, there were no differences in FEV_1% and ppoFEV_1% between the two groups ($p=0.08$ and $p=0.07$ respectively). Logistic regression demonstrated ppoDLCO (Odds ratio [OR] for each 5% decrement: 1.13 [95%CI 1.06, 1.19], $p<0.001$) was an independent predictor of PPCs. Amar et al then created a simple additive model incorporating ppoDLCO% and pre-op chemotherapy use. These being the only independent predictive variables. A point score of two for history of chemotherapy, and one for each 5% decrement of ppoDLCO% less than 100% were allocated. The model discrimination was an area under the receiver operating characteristic curve (AUROCC) of 0.63 ($p=0.4$, No CI given by author).[197] Three groups were created corresponding to low (0-10 points), middle (11-13) and high risk (14-19) (Figure 17). These groupings were statistically significant between those with and without complications ($p<0.0001$). Despite showing statistical significance, the discriminative capability of 0.63 is weak and unlikely to be useful clinically (the value of AUROCC in risk prediction is discussed in detail in section 7.17). This prediction tool was not externally validated, instead it was internally validated using a *'jackknife'* technique. It is encouraging validation was performed- this is unusual in published work within this area. This study does display some potential of ppoDLCO% to predict PPCs but further work would be needed to enable clinical use.

[E] PPCs defined in this study as: atelectasis requiring bronchoscopy, respiratory failure requiring intensive care unit admission, pulmonary embolism, pneumonia (new pulmonary infiltrate with fever treated with IV antibiotics), acute lung injury (the worst being Acute Respiratory Distress Syndrome, ARDS) and the need for oxygen therapy at hospital discharge

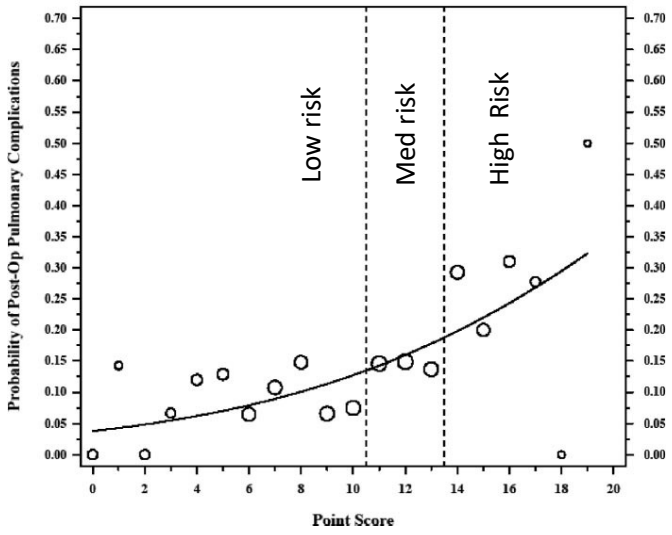

Figure 17 - Predicted probability of developing post-operative pulmonary complications (PPCs) by the point score.
Circles = observed probabilities of PPCs. Area of the circle corresponds to the number of patients with each score in the study cohort. Curve indicates predicted probability of PPCs generated from a spline smooth of the logistic regression model based on point score. Vertical dashed lines = cut offs of point scores of the three risk categories for PPCs. Taken from Amar et al [197]. (n=956)

Several times Brunelli et al[202, 173, 203] have examined the utility of pulmonary function to predict PPCs, from a single centre. These studies incorporated cardiac complications within 30 days into a composite outcome of cardio-respiratory complications and did not exclusively try to predict PPCs. Patients were prospectively and consecutively included from a clinical database. All patients had identical surgical approach and standardised anaesthetic technique with PPCs defined[F] in line with the EACTS/ESTS thoracic surgery database definitions. In two studies, Brunelli et al[202,203] explored ppoDLCO%/ ppoFEV_1% and *association* with post-operative complications. In the first, following multivariate logistic regression, only ppoFEV_1 (OR 0.98, 95% CI 0.97-0.99, for 1% decrement in ppoFEV_1) remained independently predictive of post-operative cardiopulmonary morbidity. While ppoFEV_1 was independently predictive, how useful this would be clinically is uncertain. In the second, the only independent predictor of cardiopulmonary complications was altitude reached at stair climb ($p<0.01$, OR -0.18, no CI given). Brunelli concludes therefore, FEV_1% and ppoFEV_1% were associated with, but not

[F] Complications defined were as follows; Respiratory failure requiring mechanical ventilation for more than 48 hours, pneumonia (chest x-ray infiltrates, increased white cell count, fever), atelectasis requiring bronchoscopy, adult respiratory distress syndrome, pulmonary oedema, pulmonary embolism, myocardial infarction, arrhythmias requiring medical treatment, cardiac failure with chest x-ray findings, acute renal failure and stroke.

predictors of, cardiopulmonary complications. This series of studies by Brunelli suggests the ability to predict PPCs using pulmonary function is poor.

In another large commonly referenced study exploring the utility of pulmonary function to predict PPC, Licker et al[204] examined 1239 patients undergoing lung resection by thoracotomy. This was a retrospective observational study of consecutive patients and the author reported FEV_1% <60% was an independent risk factor of respiratory complications following logistic regression (OR 2.7, 95%CI 1.3 -6.6). This threshold of 60% was defined from within this cohort using AUROCC analysis (0.66). The study included two hospitals which performed the surgery across 14 years. The anaesthetic technique used for all patients was well described and standardised along with respiratory complication definitions[GG] collected within the first 30 post-operative days. It must again be acknowledged that a 60% cut-off value is not used within guidelines but, in contrast to other studies within this chapter, on this occasion the authors justified the cut-off they selected as being derived as optional (Youdens) within the dataset. The authors concluded:

> *'An FEV_1% <60% was the main predictor of peri-operative respiratory morbidity and mortality'*
>
> *Licker et al 2006 [204]*

[GG] Respiratory complications defined included; prolonged chest drainage >7 days, haemothorax with a need for transfusion and chest drainage >800mls, reintubation with ventilatory failure, atelectasis with need for CPAP or bronchoscopy, pneumonia with temperature >38 degrees celsius and new infiltrates on chest x-ray, acute lung injury with diffuse pulmonary infiltrates and PaO2/Fio2 ratio <220mmHg and finally bronchopleural fistula with bronchoscopy and application of glue.

Table 11 – Studies examining FEV_1 and DLCO as predictors of post-operative pulmonary complications following lung resection.

Study	Population	Method	Outcome	Comment
Amar et al 2010 [197]	Retrospective n=956 Single centre	PPCs within 30 days Thoracotomy n=956 Database review	ppoDLCO (Odds ratio [OR] for each 5% decrement: 1.13 [95%CI 1.06, 1.19], p<0.001) was an independent predictor of PPCs FEV_1% was not a predictor of PPCs. Model incorporating ppoDLCO% and pre-op chemotherapy (only independent predictors) AUROCC 0.63 (p=0.4, no CI given)	93 (10%) patients did not have DLCO% measured Weak AUROCC value Displays potential of ppoDLCO as predictive variable but further work needed.
Bobbio et al 2009 [205]	Prospective n=73 Single centre	PPCs within 30 days Lobectomy n=64 Bi-lobectomy n=5 Segmentectomy n=4	FEV_1% independent predictor of PPCs (p=0.005, no OR/CI stated by author) DLCO associated (p=0.03) but not predictive of PPCs AUROCC of FEV_1% 0.70 (95%CI, 0.57—0.85)	FEV_1% was an independent predictor of PPCs FEV_1% displayed potential as predictor of PPCs
Brat et al 2016 [206]	Retrospective n=76 Multicentre	PPCs within 30 days Pneumonectomy n=17 Lobectomy n=47 Segmentectomy n=10	FEV_1% and DLCO% not significantly different at univariate analysis in those with and without PPCs (p = 0.90, p = 0.55)	FEV_1% and DLCO% not predictive of PPCs
Brunelli et al 2002 [173]	Prospective n=160 Single centre	PPCs within 30 days Thoracotomy n= 160	FEV_1% and DLCO% not independently associated with PPCs (altitude reached at stair climbing was only independent predictor of cardiopulmonary complications)	Regression model and values not displayed in paper.
Brunelli et al 2012[207]	Prospective n=225 Single centre	PPCs within 30 days Lobectomy n=197 Pneumonectomy n=28	FEV_1% and DLCO% not different in those with and without PPCs at univariate analysis	Pulmonary function not predictive of PPCs

Study	Population	Method	Outcome	Comment
Brunelli et al 2004 [203]	Prospective n=109 Single centre	PPCs within 30 days Patients >70 years Thoracotomy n=109	ppoFEV_1% different between groups ($p<0.01$) ppoDLCO% not different between groups ($p>0.05$) Following multivariate analysis: ppoFEV_1% and ppoDLCO% were not independent predictors of PPCs	Pulmonary function not predictive of PPCs
Brunelli et al 2008 [202]	Prospective n=536 Single centre	PPCs within 30 days Pre-op stair climbing and PFTs Database search	ppoFEV_1% was an independent predictors of PPCs ($p=0.004$), no further predictive statistics given. ppoDLCO not an independent predictor but associated with PPCs on univariate analysis ($p=0.03$)	ppoFEV_1 was an independent predictor of PPCs – no predictive statistics displayed. Database search only
Brutsche et al 2000[174]	Prospective n=125 Single centre	PPCs within 30 days Thoracotomy n=125	FEV_1% significant at univariate analysis ($p=0.09$) but not predictive an independent predictor of PPCs.	FEV_1% not an independent predictor of PPCs
Dales et al 1993 [201]	Prospective n=117 Single centre	PPCs during hospital admission Pneumonectomy n=18 Lobectomy n=86 Other n=13	PPCs in 21% in those with a normal FEV_1% compared with 50% among those with an FEV1% <60% ($p<0.05$). No predictive statistics	FEV_1% associated with but not predictive of PPCs No justification for 60% cut-off
Ferguson et al 2008 [186]	Retrospective n=1008 Single centre	PPCs within 30 days Lobectomy n=752 Bi-lobectomy n=83 Pneumonectomy n=173	Difference between ppoDLCO% and ppoFEV_1 at univariate analysis in those with and without PPCs, ($p<0.05$). ppoDLCO% independent predictor of PPCs (OR 0.728, 95%CI 0.565- 0.939, $p=0.015$)	ppoDLCO% independent predictor of PPCs in those with and without COPD ppoFEV_1 independent predictor in COPD cohort only

Study	Population	Method	Outcome	Comment
Kearney et al 1994 [200]	Prospective n=331 Single centre	Pre-defined PPCs during hospital admission Lobectomy n=150 Pneumonectomy n=46 Wedge n=135	$ppoFEV_1$ was an independent predictor of PPCs for each 0.2 litre(L) decline in ppoFEV1 the odds ratio for complications was 1.46 (95% CI 1.2,1.8)	$ppoFEV_1$ was an independent predictor of PPCs No further predictive statistics
Keeratichananont et al 2016 [208]	Prospective n = 78 Single centre	Pre-defined PPCs within 30 days No case breakdown	FEV_1% <60% an independent predictor of PPCs (HR 1.48, 95%CI 1.13-2.21, p=0.04) AUROCC of FEV_1% 0.93 (No p-value or CI reported)	FEV_1% was an independent predictor of PPCs Small patient cohort
Licker et al 2006 [204]	Retrospective n=1239 Multicentre	Pre-defined PPCs within 30 days Thoracotomy n=1239	FEV_1% <60% independent predictor of PPCs (OR 2.7, CI 1.3 -6.6).	FEV_1% was an independent predictor of PPCs
Matsuoka et al 2004 [163]	Prospective n=130 Single centre	Pre-defined PPCs, unclear duration. Lobectomy n=130	FEV_1% and $ppoFEV_1$% no different at univariate analysis in those with and without PPCs. No p-value given by author. Not taken forward to multivariate regression.	FEV_1% and $ppoFEV_1$% not independent predictors of PPCs
Markos et al 1989 [154]	Prospective n=55 Single centre	Pre-defined PPCs within 30 days Pneumonectomy n=18 Lobectomy n=29 Other n=8	$ppoFEV_1$% independent predictor of PPCs ($p<0.05$, no OR/CI given) DLCO, DLCO%, ppoDLCO and ppoDLCO% independent predictors of PPCs ($p<0.05$, no OR given)	Association demonstrated, no predictive statistics performed despite authors results Small cohort/single site

Study	Population	Method	Outcome	Comment
Kim et al 2016 [194]	Prospective n=343 Single Centre	Pre-defined PPCs during hospital admission Wedge n=20 Lobectomy n=302 Pneumonectomy n=21	DLCO% (OR:0.97, 95% CI 0.957 -0.991, p=0.003) independently associated with PPCs.	DLCO% independent predictor of PPCs FEV_1% was not an independent predictor of PPCs

4.4.4.1 FEV_1 and DLCO to predict post-operative pulmonary complications: conclusion

Many authors have attempted to address the topic of risk prediction (of post-operative pulmonary complications) for lung resection surgery using absolute and predicted values of FEV_1 and DLCO. Early studies demonstrated *association* between FEV_1/DLCO and PPCs but advocated the use of pulmonary function to *predict* post-operative pulmonary complications. In recent years however, the role of FEV_1 in predicting risk before lung resection has been questioned.[209] Further work has observed that pulmonary function poorly predicts PPCs, with most studies demonstrating no independent predictive value of FEV_1 or DLCO.[174, 207, 203, 163] Those studies which have observed independent predictive strength of either FEV_1% of DLCO% had poor discrimination with results unlikely to be useful in clinical practice. Others did not comment on discrimination with no predictive statistics, therefore little conclusion can be drawn from these results. Therefore, it is not surprising pulmonary function does not feature in national guidelines to aid prediction of PPC's, given the lack of high-quality evidence.

4.5 Functional assessment for risk stratification

This section explores the additional *functional assessments* recommended in those moderate/high risk patients outlined in current risk stratification guidelines (section 4.2). When patients are deemed high risk for lung resection surgery based on pulmonary function testing, submaximal exercise testing and cardiopulmonary exercise testing are recommended to assist the prediction of surgical outcome.[18] It is logical that exercise testing would be an ideal tool to evaluate fitness of a patient as during exercise, oxygen consumption, carbon dioxide production and cardiac output all increase. The level of work achieved reflects the ability of the heart and lungs to deliver oxygen to the tissues.[210]

4.5.1 Sub-maximal exercise testing

Assessment of exercise capacity in patients with lung cancer undergoing surgery can be supplemented with the use of low technology tests such as shuttle walks, six-minute walk tests and stair climbing evaluation.[18, 51] The 2010 BTS guidelines *do not* recommend these tests as a single or reliable method of patient evaluation.

Instead these low technology tests should be used to indicate the necessity for further pre-operative assessments of cardiopulmonary function.[18]

An international survey completed in 2009 by the European Society of Thoracic Surgeons observed that 65% of responders performed one pre-operative low technology test routinely in lung cancer resection patients to aid the shared decision-making process; 18% using them systematically in all patients being considered for surgery often as a surrogate to VO_2 measurement.[177]

4.5.2 Stair Climbing

Stair climbing (SC) has historically been used in the evaluation of a patient's fitness for surgery. Van Norstrand et al have been credited with studying this first in the lung cancer surgery population, discovering those unable to climb two flights of stairs, had a 50% mortality rate following pneumonectomy.[211] SC can be performed as a symptom limited test or submaximal test. Setting an objective target of 2 or 3 flights is the sub maximal way of distinguishing those fit for surgery or not. However, most studies on SC are carried out as symptoms limited testing; patients being instructed to climb the maximum number of stairs until exhaustion.[173] Results are expressed as number of steps climbed, vertical metres(m) achieved (altitude) or speed of ascent. Due to the variable height of a step across the world, results expressed in metres are favoured, allowing standardisation of results; steps varying from 0.15 to 0.17m.[212] Pate et al were the first to encourage this reporting process in a small study of 12 patients.[213] The British Thoracic Guidelines 2010 recognises the lack of standardisation amongst studies looking at stair climbing as predictor of outcome, whether that be survival, quality of life or complication risk.[18] Appendix 1 summarises studies using pre-operative stair climbing in the lung cancer population.

Equations allow the results of the SC test to be transformed into an estimated VO_2 peak.[214] The European Society Cardiology (ESC)/European Society Anaesthesia (ESA) guideline on non-cardiac surgery advise assessment of functional capacity based on metabolic equivalents or METS.[215] One MET is equivalent to basal metabolic rate. Climbing two flights of stairs demands 4 METS and strenuous sports can require upwards of 10 METS. Prior to thoracic surgery, the inability to climb two flights of stairs and achieve 4 METS is associated with an increased incidence

of post-operative cardiac events and increased mortality (relative risk 18.7; 95% CI 5.9 -59).[215, 216] Conversely, a poor functional status was not associated with an increased mortality after other non-cardiac surgery (relative risk 0.47, 95% CI 0.09 – 2.5).[216]

In a retrospective study with a small number of patients (n=54) undergoing surgical lung resection, Olsen et al 1991 were the first to suggest SC was associated with post-operative complications and a greater frequency of complications (n=26)[H]. [214] Regression analysis suggested a weak negative correlation between steps climbed and complications rate ($r = -0.3$, $p < 0.05$). The number of complications was greater in those seven patients unable to complete three flights than in those 47 who did (2.3 ± 2 vs 0.7 ± 1, $p<0.05$).

Olsen et al concluded:

> *'The study would need to be repeated in a larger cohort to have any clinical significance and they could not confirm or dispute the use of pre-operative SC to predict post-operative complications.'*
>
> *Olsen et al 1991* [214]

Brunelli et al [173] [210] [217] [184] [203] have since been the main authors exploring SC as a predictor of risk for lung resection surgery within the literature. Table 12 summaries the main studies and findings from their work. The common finding from all of Brunelli's work is that SC was repeatedly observed to be an independent predictor of *cardiopulmonary* complications. Pulmonary complications are never examined independently in this evidence, nor is the prediction or association of post-operative dyspnoea. All of the studies within the table also concentrate on the first 30 days following surgery and not long-term complications or functional limitations, which would be more beneficial for my research question.

[H] Complications were defined as; arrhythmias, atelectasis, pneumonia, pulmonary embolism, hypotension, myocardial infarction, respiratory failure, death.

4.5.2.1 Stair Climbing: conclusion

SC appears frequently within current guidelines used widely within clinical practice to guide risk prediction and decision-making processes. Despite evidence displaying performance at stair climbing is associated with (*and* predictive of) cardiopulmonary complications, few have examined the predictive ability of SC for post-operative dyspnoea. Arguably, SC provides little physiological information other than the height achieved, although one study has displayed association with results achieved at SC testing and CPET testing.[217] If future work can confirm these findings, SC may become a cheaper and more readily available alternative to CPET testing in risk stratification for lung cancer resection. The first large systematic review, published in 2020 assessing performance of SC test by Boujibar et al[218], supported this theory and indicated SC could be used first line to predict post-operative morbidity - all of the articles discussed in this chapter were included within this analysis.

Table 12 - Risk prediction using pre-operative stair climbing in lung resection surgery – Brunelli et al

Study	Population	Method	Outcome	Comment
Brunelli et al 2002 [173]	Prospective n=160 Single centre	Pre-operative stair climbing PPCs within 30 days Lobectomy n=111 Pneumonectomy n=28 Wedge n=21	Patients with complications climbed a lower altitude at the stair climbing test ($p < 0.01$, no other stats given) Patients with complications climbed less altitude than those who were complication free (14.96m (5.5) vs 20.60(4.62)), ($p<0.0001$)	Altitude reached at the stair climbing test was associated with but not an independent predictor of post-operative complications
Brunelli et al 2004 [203]	Prospective n=109 Single centre	Pre-operative stair climbing PPCs within 30 days Lobectomy n=70 (>70 years)	Following multivariate analysis: altitude achieved at pre-op stair climb height ($r=-0.18$, no OR or 95%CI given, $p<0.01$) was an independent predictor of PPCs	Stair climbing better than FEV_1 at predicting complications in patients >70 years
Brunelli et al 2008 [202]	Prospective n=536 Single centre	Pre-operative stair climbing CPCs within 30 days Lobectomy n=440 Pneumonectomy n=96	Height achieved at stair-climbing test (OR 0.95, 95% CI 0.91-0.99, $p=0.045$) was an independent predictor of cardiopulmonary complications	SC was an independent predictor of cardiopulmonary complications (significance borderline)

Brunelli et al 2008 [210]	Prospective n=640 Single centre	Pre-operative stair climbing CPCs within 30 days Lobectomy n=533 Pneumonectomy n=107	Patients with complications had lower values of SC, ($p<0.01$). SC was independent predictor of cardiopulmonary complications ($p=0.04$, OR 1.4 (95% CI 1.02-1.95). For the ability to achieve 12m of altitude - 13% PPV and 97% NPV for mortality and 40% PPV and 78% NPV for morbidity	Brunelli et al recommend the systematic use of symptom-limited stair climbing as a first line screening test in all candidates for lung resection as it is an independent predictor Proposed 12m and 22m cut off altitude, ($p=0.02$).
Brunelli et al 2012 [184]	Retrospective n=296 Single centre database analysis	Pre-operative stair climbing Survival – defined as interval between the surgical resection and death or last contact Lobectomy n=296	Five-year survival of patients who climbed more than 18 meters was longer than those who climbed less than 18 meters (77% vs 54%, $p<0.01$) Climbing more than 18m was an independent prediction factor associated with survival (HR 0.5, $p<0.01$, no 95% given)	Pre-operative stair climbing is an independent predictor of survival.

4.5.3 Six minute walk testing

Six minute walk testing (6MWT) was developed in 1963 by Balke[219] to assess functional capacity in chronic respiratory disease and heart failure. The American Thoracic Society (ATS) published guidelines in 2002 on the 6MWT for the objective evaluation of functional exercise capacity.[220] The 6MWT must be performed with two observers on a 30m course using a pulse oximeter to measure heart rate and oxygen saturation. Additionally, patients grade themselves on the Borg breathlessness scale. No cut off distance in the 6MWT has been quoted in the 2010 BTS guidelines to differentiate between high and low risk patients.[18, 51, 52] Unlike other low technology forms of exercise testing, the 6MWT assesses sub maximal functional capacity; with patients completing the protocol at their own intensity.[221] This is the main criticism of this form of testing and its ability to predict pulmonary complications.[18] Appendix 2 summarises studies using pre-operative SWT in the lung cancer population.

Bagg et al[222] were the first to use 6MWT to evaluate the lung resection population and they concluded it had limited use in pre-operative evaluation in a small study of 22 patients. Seven patients suffered post-operative ventilatory complications with no difference in distance walked between these patients and those with no complications.

In a single centre, Markos et al[154] confirmed the inability of the 6MWT to be used in surgical decision making. In 55 patients the author prospectively looked at 30-day complication rate.[l] Sixteen patients developed post-operative cardiopulmonary complications and three died. There was no association between distance at walk testing and post-operative complications, ($p>0.05$). This paper has been discussed extensively in previous sections (4.4.1). The BTS reference this paper when acknowledging the limitations of 6MWT before lung resection surgery.[18]

Holden et al[223] examined 6MWT, as a supplement to conventional spirometry testing, to predict outcome following lung resection surgery. In 16 patients the

[l] Complications defined in this study as: death, respiratory failure ($PaCO_2 > 45mmHg$ or $PaO_2 < 60mmHg$ in absence of oxygen supplementation), pneumonia, atelectasis, pulmonary embolism, myocardial infarction, arrhythmias requiring therapy and intensive care admission

author states distance achieved during 6MWT was predictive of 90-day mortality and travelling a distance exceeding 1,000 feet was predictive of survival with a sensitivity of 100%, a negative predictive value of 100%, and a positive predictive value of 85%, ($p<0.05$). These results are impressive, but given the cohort was only 16 patients, further prospective evaluation in larger populations would be required before routine use.

The usefulness of 6MWT may lie post-operatively to quantify exercise capacity, rather than be used pre-operatively to predict complications or outcome. Brocki et al[224] found association between performance on 6MWT and physical functioning component of the SF-36 QoL questionnaire following surgery in a prospective observational study of 78 patients for lung cancer surgery. The same author also studied 65 patients undergoing lung resection. Each patient underwent 6MWT pre-operatively, at two weeks post-operatively and again at 6 months. Two weeks following surgery patients had a decline in 6MWT results, ($p<0.01$) before recovering to baseline distance at 6 months.

One of the few studies to display the ability of the 6MWT to predict cardiopulmonary complications is by Pierce et al[225] in 1994, when he studied 52 patients presenting for lung resection in a single centre. Each patient performed a 'best of three' pre-operative 6MWT and cardiopulmonary complications were recorded up to 32 days post-operatively[J]. Only 6MWT was independently predictive of respiratory complications at "multivariate analysis" (no further stats given). Mean (SD) 6MWD was 501 (47) m in those with post-operative complications ($n=36$) vs. 556 (88) m in those without complications ($n=9$), ($p=0.03$).

Finally, in one of the largest studies to look at 6MWT and lung resection, Marjanski et al 2015[226] studied 253 patients undergoing lung resection in a single centre. Pre-operative 6MWT was performed in all patients with the aim of the study to evaluate the cut-off value of the distance achieved at 6MWT and identify those patients at increased risk of post-operative complications[K]. In a multivariate

[J] Respiratory complications were defined as chest infection (temperature >37.5 for >24 hours with chest x ray findings), atelectasis on chest x ray, pulmonary embolism, respiratory failure with intubation requirement or PaCO2 > 45mmHg at any time post-operatively and symptomatic dyspnoea at rest. Cardiac complications were categorised as acute myocardial infarction, cardiac failure and arrhythmias.

[K] Complications defined as; bleeding requiring transfusion, haematoma, transfusion > 2 units of blood, bronchial stump fistula, pneumothorax, wound infection, pleural empyema without fistula,

analysis, a 6MWT distance of <500m was found to be significantly associated with increased risk of cardiopulmonary complications (p<0.01, OR 2.6 95%CI 1.28 - 5.30). Sixty-five patients *could not perform* 6MWT (either due to frailty or other significant comorbidity and lower limb disability) had a similar 30-day mortality (p=0.2) but a higher 90-day mortality of 7.6% (5 from 65) (p<0.01, OR 21 95%CI 2.32 – 484.12). Marjanski et al appropriately concluded therefore, patients who walk <500m during the 6MWT before lobectomy have an increased risk of post-operative complications and prolonged hospital stay and suggested that 6MWT should be used in the basic initial routine pre-operative risk assessment and quantification for lung resection. A limitation of this study is most of the patients with a high peri-operative risk due to multiple co-morbidities could *not* undertake the 6MWT. While the 30-day mortality was the same in this small group, the 90-mortality was higher – suggesting an inability to undertake a 6MWT is in itself may be a marker of peri-operative mortality.

4.5.3.1 Six minute walk testing: conclusion

The 6MWT has been validated in other respiratory and cardiac conditions such as pulmonary hypertension, COPD and stroke. To date, no authors have studied the utility of the 6MWT to predict dyspnoea following lung resection for cancer. The strength of the 6MWT to predict morbidity and mortality in the lung cancer population is debatable.[18] Based on the inconsistency of evidence in the literature, the European Respiratory Society/ European Society of Thoracic surgeons [177] do not advocate the use of 6MWT in patient evaluation, a recommendation that has also been supported by the ACCP.[165] Similar to shuttle walk testing, the convenience of performing a simple low cost pre-operative test would be useful, however the 6MWT should not be a substitute for more sophisticated CPET testing.

---

prolonged air leak >7 days, urinary tract infection, other infections, renal insufficiency, chylothorax, return to theatre, paresis of recurrent laryngeal nerve, respiratory insufficiency requiring reintubation, atrial arrhythmia requiring treatment, ventricular tachycardia requiring treatment, atelectasis requiring bronchoscopy, prolonged mechanical ventilation, pneumonia, adult respiratory distress syndrome, tracheostomy, myocardial infarction, pulmonary embolism, sepsis, psychosis cerebral infarction, intra-operative death, death during hospitalisation and other complications.

4.5.4 Shuttle walk testing

Shuttle walk testing (SWT) measures the distance a patient is able to walk between two cones 10m apart, over a given timeframe. Speed is gradually increased until the patient can no longer complete any further crossings. This test is easy to perform and has advantages when compared with cardiopulmonary exercise testing, the gold standard for peri-operative evaluation. SWT can be used to screen 'high risk patients' (based on poor pre-operative pulmonary function test results) with a distance of 400m set as a guide. Patients may require further CPET evaluation prior to surgery. Less studies have concentrated on SWT to predict complications and outcome than other pre-operative tests such as CPET and FEV_1/DLCO, this will be explored later in the chapter. Appendix 3 summarises studies using pre-operative SWT in the lung cancer population.

The British Thoracic Society (BTS) acknowledge the limitations of SWT to predict outcome following a large prospective study by Win et al 2004 [227] where 111 patients underwent SWT prior to surgery. No difference in SWT distance was observed in those with *favourable* and *poor* surgical outcomes[L]. Sixty-nine patients had a favourable surgical outcome and 34 had a poor outcome – the remaining patients are not accounted for. Mean shuttle distance in the favourable outcome group was 419 metres and poor outcome group 388 metres, (p=0.6). Furthermore, there was no difference between shuttle walk distance in those who died (p=0.5) and those who had a short or prolonged hospital stay, (p=0.5). No definition is given regarding what short or prolonged stay is in this context. If pre-operative shuttle walk distance was <250 metres, 50% of female and 83% of male patients had poor outcome. If <400 metres, 27% of female and 43% of male patients had poor outcome. Shuttle walk distance was used as a continuous measure in this analysis but the author does discuss its use in categorical groups; if a patient walked <250 metres the chance of having a poor outcome was 66% which reduced to 44% when the walk distance was <300 metres. No further analysis is done with this categorical data and the authors conclude the absolute distance achieved on SWT does not predict outcome in lung cancer patients with borderline function. If a normal range of shuttle walk distances corrected for age and sex were to become

[L] Poor surgical outcome was defined as any of the following; post-operative death, myocardial infarction, heart failure, respiratory failure, septicaemia, pneumonia and significant cardiac arrhythmia

available, the role of the shuttle walk could be re-evaluated, considering age, sex and other variables may increase the predictive value.

The 2014 American College of Cardiology (ACC) and American Heart Association (AHA) guidelines for the evaluation of patients for non-cardiac surgery[52] is also cautious in recommending SWT as a predictor of postoperative morbidity and mortality. This is based on a study by Struthers et al[228] that described patients often having satisfactory CPET results with poor shuttle walk distances. The study concentrated on 50 patients from a single centre, none of which were patients undergoing thoracic surgery and instead had intra-abdominal surgery. Thirty-nine patients had a measured VO_2 peak of 15 ml/O2/kg/min or greater. There was an observed relationship between distance walked and measured VO_2 peak ($R^2 = 0.57$, $p<0.01$). Three hundred and sixty metres was selected as a cut-off as this represented similar cut-offs used to predict out come in oesophagectomy patients. Nineteen patients not able to walk 360m achieved the goal oxygen consumption of 15 ml/O2/kg/min or greater on CPET testing –representing 38% of this sample population. Thirty-two patients were 'low-risk' with both an AT of >11ml/O2/kg *and* VO_2 peak >15ml/O2/kg/min. Only 13 of these low-risk patients could be identified with a shuttle walk distance of 360m, (PPV 1.0, NPV 0.49). As discussed above, given a suggested distance of 400m is used as a screen for thoracic surgery in the current 2010 BTS guidelines this cut-off is suggested with caution.[18] The study was not powered to compare outcomes between patients.

The European Respiratory Society and Society of Thoracic Surgeons latest guideline in 2009 states:

> *'SWT distance underestimates exercise capacity at the lower range and was not found to discriminate between patients with and without complications. Thus, it should not be used alone to select patients for operation. It could be used as a screening test: patients walking less than 400m may have a peak VO2 <15ml/kg/min.'*
>
> *Brunelli et al 2009* [164]

SWT cannot be recommended as an independent assessment tool because two prospective studies failed to validate it as a predictor of outcome – Win et al 2004 [227] (discussed above in this section) and Win et al 2006.[229]

In another study of 125 prospectively recruited patients by Win et al[229] , SWT tended to underestimate exercise capacity at the lower range compared with peak oxygen consumption. Fifty-five patients achieved more than 400m on the SWT, all of whom had peak VO_2 >15/ml/kg/min. Seventy patients failed to achieve 400m on the test; in 22 of these patients VO_2 peak was <15/ml/kg/min. Seventeen patients achieved less than 250m, nine of whom had VO_2 peak >15/ml/kg/min. (250m was the previous recommended distance at SWT, below which the patient would not need to undergo further CPET testing as it was assumed VO_2 max would be less than 10ml/kg/min and 'high risk'.) There was moderate association between shuttle walk distance and VO_2 peak ($r=0.67$, $p<0.001$). Given that several patients who failed to walk 250m had peak oxygen consumption >15ml/kg/min, SWT *cannot* be used reliably as a single assessment tool. Win et al claimed SWT had a significant predictive value and concluded that if the (then) current 250m cut off was adhered to, then some patients would be needlessly excluded from surgery. A threshold of 400m in the shuttle walk test had a sensitivity of 77% and specificity of 54% for one-year survival. The mean (SD) shuttle distance in those who survived to one year was 428m (135) compared with 335m (122) in those who died. These findings had some implications for both the BTS and ACCP guidelines for patient selection- revising the 250m cut off for further investigation and surgery up to 400m. Though this study was planned with a single primary outcome defined and rigorous follow-up, meaning dropout rates were very low, importantly, no predictive statistics were performed.

4.5.4.1 Shuttle walk testing: conclusion

SWT is a simple test requiring little technology or training and would be useful if validated in the lung cancer population. While there exists an association between pre-operative SWT distance and post-operative outcome, few authors have examined its predictive value despite its popularity. Little work has been done to validate this as a tool to predict post-operative mortality, morbidity (including post-operative dyspnoea). SWT has been suggested as a screening tool in those

with borderline function who may require more sophisticated tests and until further research is performed, this will remain its primary role.

4.5.5 Low technology testing: conclusion

Pre-operative low technology testing is a simple method of highlighting major problems with the cardiovascular system which may preclude surgery.[18] They could be regarded as global tests, capable of uncovering deficits in the oxygen transport system[210], and are being used with increased frequency during pre-operative evaluation before lung resection.[210] Much work has been done to research and validate pre-operative low technology exercise tests in the lung cancer population, however none have explored the prediction of long term dyspnoea.[165]

Pre-operative SC is the most promising low technology test to assess exercise capacity and predict long term functional performance.[18, 51, 165] However, it is unlikely to replace CPET testing, particularly in high risk cases.[5] Low-technology tests are a gateway to CPET which remains the most reliable marker of a reduced aerobic reserve.[230]. Only high-risk patients (categorised pulmonary function test) with good performance in the SC, with no other risk factors should be considered to proceed directly to surgical intervention.[18, 51, 165] A recent high-quality meta-analysis published in 2020 has supported this theory and indicated SC could be used first line to predict post-operative morbidity - future guidelines may therefore reflect this.

The major limitation with low-technology testing is poor reproducibility. Attempts have been made to standardise procedures but notable methodological limitations still remain; the equipment used, hospital structure and instructions given to patients.

4.5.6 Pre-operative cardiopulmonary exercise testing

Cardiopulmonary exercise testing (CPET) can be used to assist the prediction of surgical outcome following lung resection for cancer. This includes an assessment of exercise capacity. Formal CPET provides a global assessment of the response to exercise and increase workload involving the cardiovascular, pulmonary, haematopoietic, neurophysiological and skeletal muscle systems.[231] The British

Thoracic Society guidelines 2010[18] and NICE guidelines[51] (Lung cancer, diagnosis and management 2019) recommend CPET testing to measure peak oxygen consumption in patients with moderate to high risk of post-operative dyspnoea based on pre-operative pulmonary function. Following CPET testing, a VO_2 max <15ml/kg/min is suggested as high risk and nationally accepted as an appropriate cut-off. However, within the west of Scotland population, a paper by Forshaw et al[232] demonstrates that within an oesophageal resection cohort (a similar demographic to the lung cancer population), patients presenting for surgery had a mean VO_2 max less than 15ml/kg/min. Therefore, suggesting a less fit population within this region and perhaps a need for local validation before accepting national recommendations.

CPET is the final assessment of operative risk, falling at the end of the algorithm described in section 4. Appendix 4 summarises studies which concentrate on the value of pre-operative CPET testing in patients undergoing lung resection surgery. On one hand, the most important outcome when considering surgery for lung cancer is to survive the procedure – however the majority of current evidence does not fully address this because of underpowered studies. Without surgery, for many lung cancer patients the alternative outcome is death. In the other hand, there exists a risk of post-operative QoL becoming a trade-off for an increase in life expectancy.

The American Thoracic Society (ATS) and American College of Chest Physicians (ACCP) have published guidelines on how to perform and interpret cardiopulmonary exercise testing in adults.[231] CPET is a relatively non-invasive physiological overview to evaluate the submaximal and peak exercise capacity of an individual. CPET measures respiratory gas exchange including oxygen uptake (VO_2), carbon dioxide output (VCO_2) and minute ventilation (VE). Blood pressure, pulse oximetry and electrocardiography are also collected during a symptom limited progressive maximal exercise tolerance test. Two modes of exercise are commonly used to perform CPET testing: treadmill testing and cycle ergometer.

Normal VO_2 in a healthy adult at rest is about 3.5ml/minute per kilogram (250ml/minute), increasing during exercise to values 15 times higher (30-50ml/min/kg). VO_2 max is defined as the maximal oxygen uptake during exercise and is used interchangeably with VO_2 peak. It represents the maximal achievable

level of oxidative metabolism and is achieved when a plateau in VO_2 has occurred with increasing work rate. However, it is recognised in clinical testing a clear plateau may not be achieved before symptom limitation of exercise. Exercise limitation in patients with reduced VO_2max is complex and multifactorial, involving several components of the oxygen transport utilisation process.[231] BTS recommend clinicians use CPET testing and specifically VO_2 max for functional assessment in those with moderate to high risk of post-operative dyspnoea, with pre-operative values of 15ml/kg/min as a cut off for 'high risk' based on several historical studies.[18] This is because above these levels, no patient in this study experienced any adverse events.

A frequently referenced study to hypothesise preoperative exercise testing could predict post-operative complications was by Smith et al[233] who prospectively studied 16 patients undergoing lung resection for cancer at a single centre. All patients underwent cycle ergometry to determine VO_2 max in addition to routine lung spirometry. Post-operative complications were defined[M] as those falling within 30 days of surgery and included complications expected to occur as a result of poor cardiopulmonary reserve. Non-pulmonary complications or technical problems were excluded. Cardiopulmonary complications occurred in n=11 (50%). Those without complications had higher VO_2 max than those who had cardiopulmonary complications (22.4 ml/kg/min Vs 14.9 ml/kg/min, $p<0.01$). One patient with VO_2 max >20ml/kg/min had a complication and every (n=6) patients with VO_2 max <15ml/kg/min had a complication. Patients were classified 'high risk'(n=14) on the basis of preoperative spirometry results: FVC <50%, FEV_1 <2L or < 50% of FVC or DLCO <50% or ppoFEV1 <1L. Eight (57%) of these high-risk patients had post-operative cardiopulmonary complications and within this group, mean VO_2 max was greater in those with no post-operative complications (21ml/kg/min) Vs those with post-operative complications (14ml/kg/min, $p<0.01$). Smith et al concluded:

> *'The results of this study indicate a strong association between preoperative exercise capacity and the incidence of*

[M] Respiratory failure with PCO2 >45mmHg or the need for mechanical ventilation for >48hours postoperatively, myocardial infarction, cardiac arrhythmias requiring therapy, pneumonia with temperature >38 for >48 hours and purulent sputum/chest x-ray findings, lobar atelectasis, pulmonary embolus and death.

post-operative complications. Therefore, exercise testing is a useful adjunct in evaluation of operative risk for thoracotomy.'

Smith et al 1984 [233]

This is a well conducted study but patient numbers are low, with only 16 undergoing lung resection surgery. Therefore, this study displays a potential for CPET to be useful as a predictive test instead of a strong predictor, as the author advocates. Furthermore, only association has been observed as no predictive statistics have been carried out.

Another commonly referenced early study by Bechard et al[234] in 1997 also examined the role of CPET during pre-operative assessment of lung resection candidates. The study prospectively recruited 50 patients at a single centre with each patient undergoing preoperative exercise testing while the surgeons were blinded to the results. Ten patients had pneumonectomy, 28 patients had a lobectomy and 12 had wedge resections. Once again, complications were defined[N] as those falling within 30 days of surgery and only those complications expected to occur as a result of poor cardiopulmonary reserve were recorded. Eight patient had complications and those *without* complications had a higher VO_2 max (17ml/kg/min Vs 9.9ml/kg/min, $p<0.001$). Seven patients had a VO_2 max <10ml/kg/min and of these n=2 (29%) patients died and n=3 (43%) had associated morbidity. Twenty-eight patients had a VO_2 max ranging from 10-20mls/kg/min. Of these, n=0 patients died and n=3(11%) had associated morbidity. No patients with a VO_2 max >20ml/kg/min sustained any morbidity or death, ($p<0.001$). Authors concluded exercise testing is important in the pre-operative assessment of patients and a VO_2 max <10ml/kg/min is associated with significant morbidity and mortality. This is a small single centre study with limited numbers, but its strength lies in its simplicity and well conducted methodology: pre-defined complications, robust statistical analysis and single operating surgeon. However,

[N] Complications defined as : acute hypercarbia with PCO2>45mmHg, mechanical ventilation >48 hours, myocardial infarction, cardiac arrhythmias requiring treatment, pneumonia with temperature >38 degrees and sputum, pulmonary embolus diagnosis with imaging lobar atelectasis and death.

with only 8 patients developing complications it would be difficult to draw any meaningful conclusions exclusively from these results.

In a large meta-analysis, Benzo et al[235] aimed to determine if VO_2 max was different between those who developed post-operative cardiopulmonary complications and those who did not. Post-operative complications the author included were defined [O] and 14 studies including 955 patients were included in the final analysis. Following a review of the literature, Benzo and co-workers concluded that exercise capacity (expressed as VO_2 max) is lower in patients that develop complications after curative lung resection. All studies included in the analysis expressed data for VO_2 max in ml/kg/min while only 11 provided data for VO_2 max expressed as a % of predicted. The mean VO_2 max in ml/kg/min of 20ml/kg/min across all studies for non-complicated patients is consistent with the threshold proposed for patients with no risk of complications. The author concluded-

> *'Cardiopulmonary complications are important however surgery often still remains the best chance of a cure. Exercise capacity may represent a modifiable risk factor which could be improved with pulmonary rehabilitation to decrease the incidence of post-operative complications.*
>
> *Benzo et al 2007 [235]*

In one of the largest and best conducted prospective multicentre studies exploring the role of pre-operative CPET testing in lung cancer population to predict post-operative complication, Loewen et al[236] studied 346 patients. Patients were classified into low, high and very high risk based on pre-operative CPET and pulmonary function testing as described in Figure 18.

[O] Respiratory failure (Acute respiratory distress syndrome (ARDS), prolonged postoperative mechanical ventilation or reintubation), pneumonia, atelectasis requiring bronchoscopy, myocardial infarction, and arrhythmias requiring intravenous treatment.

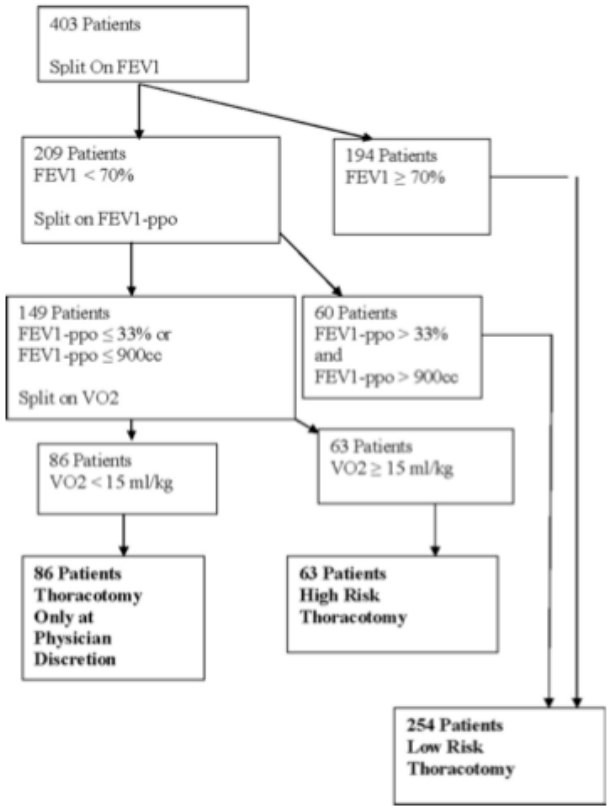

Figure 18 - Characteristics of patients for assignment into risk groups
Taken from Loewen et al[236]

The high-risk population (n=63) included patients with predictive post-operative FEV_1 <900 ml (or <33% of predicted) and maximal VO_2 of <15 ml/kg. The very high-risk population of patients had either a submaximal exercise test or both predictive post-operative FEV_1 <900 ml and maximal VO_2 <15 ml/kg. All high-risk and low-risk patients were to have surgery and the rest of the patients (very high-risk category) had surgery at the discretion of the treating physician (n=86). Measured outcomes included cardiorespiratory complications and surgical mortality within 30 days post-operatively. Overall mortality was n=15 (4%) and cardiorespiratory complications occurred in n=138 (40%) and were predefined by the author.[P] Patients with VO_2 max% <65% predicted were more likely to have cardiorespiratory complications (p<0.01, no test given) 95% CI (4.36,13.23) and

[P] Red blood cell transfusion, postoperative fever and duration, wound infection, empyema (absent fistula), empyema (fistula present), prolonged air leak, atelectasis, pneumonia, respiratory failure, dysrhythmia, myocardial infarction, deep vein thrombosis, pulmonary embolism and post-operative death.

poor outcome (p<0.04, no test given), defined as respiratory complications or death. The author concluded

> *'VO_2 was a significant predictor of operative mortality and morbidity. Patients with a low FEV_1 can tolerate lung resection with acceptable mortality and morbidity levels, provided peak VO_2 exceeded 15/ml/kg/min or 60% predicted: consistent with other work in this area'*
>
> *Loewen et al 2007* [236]

Generated from CPET testing, the ratio of minute ventilation to carbon dioxide output (also known as *ventilatory efficiency* - *Ve/VCO_2*) has been proposed by Brunelli et al as a technique to predict cardiopulmonary complications and death following lung resection surgery.[207] It has been used to stratify outcome in patients with heart failure[237] and is a novel technique in thoracic surgery patients, including the lung resection population. In 225 consecutive patients undergoing lobectomy, it was observed Ve/VO_2 was the strongest predictor of complications; 25 patients *with* complications had a higher mean (SD) Ve/VCO_2 ratio than those without (34.8 [5.5] vs 30.9 [6.1], p=0.001). This remains a novel approach in predicting morbidity and mortality following lung resection and much work is required to confirm and validate this finding.

4.5.6.1 Pre-operative cardiopulmonary exercise testing: conclusion

Patients may be willing to accept a higher rate of morbidity and mortality to have the chance of surgery and cure. [236] Pre-operative CPET testing is used to complement prediction of post-operative complications and is most beneficial in 'high risk' patients being considered for surgery when decision making is difficult.[231] Authors have proposed specific cut off points for VO_2 max that discriminate those that will or will not develop post-operative complications however they have failed to identify any pre-operative variables that can predict long term outcomes.[18] Furthermore, many of the studies used to define these cut offs have small sample sizes, leading to imprecise risk estimates.[238] While lower CPET values are generally associated with increased post-operative *complications*, no authors have studied the *predictive* value of CPET testing for long term

dyspnoea. Further work must also be done to determine if CPET testing is indeed an independent predictor of mortality (better than *Thoracoscore*).[238] Some authors have used composite endpoints to increase event rate and achieve significant results – for example grouping cardiac and pulmonary complications into 'cardiopulmonary' complications. Many patients may not consider immediate post-operative complications as sufficient cause to be refused curative surgery, but they may consider long term physical disability unacceptable.[30, 238] This is one of the key reasons the use of CPET testing has been challenged.

4.6 Other predictive markers of peri-operative morbidity, mortality and functional outcome in lung cancer

Thus far, conventional predictors of post-operative risk in lung resection surgery have been explored, including pulmonary function and functional assessment tests. This section introduces other potential variables that may have a role in prediction of risk for lung resection surgery.

4.6.1 Reduced pre-operative arterial oxygen content

Arterial blood gas measurements (ABG) have historically been included as part of routine pre-operative risk evaluation despite few studies addressing this topic.[165] Mittman et al observed increased operative risk if arterial oxygen tension fell below 6.7kpa at rest or during exercise- this was published in a review article and it is not clear what evidence there is to support this.[239] In 1982, Nagasaki et al[240] also observed that a reduced arterial oxygen concentration at rest was associated with increased incidence of complications. In 961 patients the author observed increased cardiorespiratory complications rate in those with an arterial blood gas concentration <8kpa pre-operatively. Given the clear lack of evidence, the predictive value of reduced oxygen content for functional operability and outcome is not readily used or featured in guidelines.

4.6.2 Arterial oxygen desaturation

Another historical marker of risk prediction is oxygen desaturation during the 6MWT. In 127 patients with lung cancer, Franczuk et al[241] published an abstract observing those with lower minimal oxygen saturation had increased peri-operative complication rate. Those with complications (n=41) had a lower

minimum oxygen saturation (92.2%) Vs those without (95.1%), (p=0.001). Despite Franczuk et al concluding these results display the potential value of pre-operative oxygen desaturation in assessment of lung resection candidates, arterial oxygen desaturation has never been included in risk prediction guidelines.

4.6.3 Pre-operative pulmonary artery pressure

Pulmonary hypertension is a long-standing risk factor for lung resection.[242] Pulmonary artery catheter insertion for measurement is invasive, leaving doppler echocardiography the only non-invasive assessment of pulmonary artery pressure. It has been suggested that the adequacy of the pulmonary vascular bed and circulation that determines post-operative exercise tolerance and functional capacity, rather than lung function capacity.[242]

Pre-operatively, Fee et al[243] measured pulmonary vascular resistance (PVR) using right heart catheterisation and observed higher PVR predicted mortality better than lung function tests alone. This study was performed in 45 male patients, only 30 of whom had lung resection surgery: it is not clear why. Of the 30 operations; 10 were biopsies, two segmenectomoies,11 lobectomies and 7 pneumonectomies. Post-operatively, 5 patients died of respiratory failure. This study is more than 30 years old and only a proportion of patients underwent lung resection. The applicability of the findings is therefore questionable. Fee et al conclude-

> *'PVR measurement is a physiological method of evaluating tolerance to lung resection, especially in those patients who have borderline lung function'*
>
> *Fee et al 1978* [243]

Despite this evidence that measurement of pulmonary artery pressure (PAP) may be helpful to predict function following lung resection, more recently in a larger study of 33 patients Pierce et al[244] found measurement was not better than pulmonary function testing. Pulmonary artery pressure was estimated in all patients, apart from six who had pulmonary artery catheters inserted, by doppler echocardiography. Cardiopulmonary complications and long-term survival data were collected. Baseline pulmonary artery pressure was no different in those who

survived at follow-up Vs those who died (35mmHg Vs 35mmHg, p=0.4). Forty-two patients had post-operative cardiopulmonary complication and pre-operative pulmonary artery pressure was not different between those with and without complications (36.3mmhg Vs 33.4mmHg, p=0.09). Pierce et al concluded, pre-operative PAP did not have a strong predictive power for mortality or peri-operative cardiopulmonary complications. The measurement of pre-operative pulmonary artery pressure does not feature in any national guidelines for the assessment of patients prior to lung resection surgery.

4.6.4 Pre-operative hypercarbia

Some evidence suggests that increased arterial carbon dioxide content (hypercarbia), secondary to poor ventilatory function, is associated with an increased peri-operative complication rate and reduced post-operative functional outcome.[165, 245] A pre-operative value of >6kpa has been observed to represent increased risk. In 1968, a 'resting hypercapnia' was initially described as a strong contraindication to pneumonectomy by Karliner et al, who examined 29 patients undergoing pneumonectomy.[246] This was in a single centre in a small number of patients.

More recently in 1994, Kearney et al observed a low cardiopulmonary complication rate (n=56, 17%) in a single centre with a consecutive series of 331 patients undergoing pulmonary resection.[200] There was no difference in cardiopulmonary complication rate in those with and without pre-operative hypercapnia; 4 of 30 patients with $PaC0_2$ >6kpa had complications Vs 50 of 285 patients that did not, (p>0.05). Hypercapnia was not an independent predictor of post-operative cardiopulmonary complications. One final study by Harople et al also observed a low cardiopulmonary complication rate in those patients with pre-operative hypercapnia.[247] In 883 consecutive patients undergoing lung resection surgery, 136 patients in this series had pneumonectomy and are analysed. No patient with a PaCO2 >6.6kpa had any major cardiopulmonary complications, (p<0.05).

Despite initial evidence suggesting pre-operative hypercapnia may contribute to post-operative morbidity, more recent and larger studies have observed this not to be true. To date, no studies have displayed that pre-operative hypercapnia is an independent predictor of post-operative morbidity or mortality. Therefore, it

is unsurprising pre-operative hypercapnia is not mentioned as a predictive marker in major international guidelines.

4.6.5 Other predictive markers of peri-operative morbidity, mortality and functional outcome in lung cancer: conclusion

The British Thoracic Society guidelines on management of patients with lung cancer does not include details of pre-operative arterial blood gas sampling to facilitate clinical decision making and estimate peri-operative risk or long term function outcome/dyspnoea.[18] Historically, pre-operative arterial blood gas sampling was used to determine patients at high risk for complications and mortality.[248] In the 2013 American college guidelines on the physiological evaluation of the patient with lung cancer being considered for surgery,[165] pre-operative arterial blood gas sampling is discussed under 'risk of perioperative morbidity and mortality'. Despite appearing in this national guideline, no recommendations advocate pre-operative arterial bloods gas results for risk prognostication and instead focus on lung function and CPET testing. Measurement of pulmonary artery pressure is an invasive procedure which has growing evidence demonstrating it is no more useful than lung function to predict outcome.

Considerable amounts of literature describe alternative methods to predict risk for lung resection surgery as described in this section. Despite this, routine methods in clinical practice and guidelines still incorporate lung function and CPET testing in those who are deemed to be at increased risk.

5 The use of biomarkers in peri-operative risk prediction

A biological marker or 'biomarker' is a measurable indicator of a condition or state detected using blood or urine. In 1998, the National Institute of Health Biomarkers consortium defined a biomarker as:

> *'A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention.'*
>
> *Biomarkers Definitions Working Group 2001* [249]

Biomarkers often have a clinical role in narrowing treatment decisions and are predictive, diagnostic or prognostic. The ideal biomarker would be highly specific, highly sensitive, easy to measure, rapid and inexpensive. Natriuretic Peptides (NP) are biomarkers of myocardial dysfunction widely used in clinical practice and increasingly being applied to the field of peri-operative risk prediction in a variety of clinical settings.

5.1 B-Type natriuretic peptide

Natriuretic peptides (NPs) are vasodilator hormones involved in the regulation of blood pressure and volume homeostasis.[250] Secreted from cardiac myocytes in response to myocardial stretch, B-Type natriuretic peptide (BNP) is a 32 amino acid polypeptide easily measured in plasma. In normal physiological conditions, the ventricular myocardium produces little BNP, but many pulmonary and cardiac conditions can stimulate increased production. Myocardial wall stress increasing with volume or pressure overload is the most significant pathway leading to BNP gene activation and secretion, though hypoxia and cardiac myocyte damage can also cause BNP release. BNP exists in two forms within plasma; active BNP and an inactive precursor, N-Terminal-pro BNP (NT-pro BNP). Although both forms are released in a 1:1 ratio, the level of NT-pro BNP in the circulation is higher than BNP (because of a decreased clearance rate of NT-pro BNP). NT-pro BNP is eliminated passively by skeletal muscle, the liver and the kidneys. BNP is cleared

via natriuretic peptide clearance receptors (NPR-C) and the neutral endopeptidase system, found mainly in the kidney and vascular endothelium. Elimination of both forms of BNP is affected by renal function, making its interpretation in renal failure more difficult.[251] BNP varies not just by renal function but also by age, gender, BMI, pre-existing heart failure and other co-morbidities.[252] In 1586 patients, Mueller et al[253] displayed the usefulness of BNP in the evaluation of acute dyspnoea. Importantly, despite how many factors can influence BNP levels, patients getting point of care BNP levels upon attendance to accident and emergency had less hospital admissions and reduced intensive care referrals, ($p<0.01$) - BNP appearing to aid clinical decision making in this setting.

BNP and NT-pro BNP are both used in clinical practice. However, NT-pro BNP has a longer half-life of 2 hours, making its interpretation better in some settings.[254] It has been observed BNP and NT-pro BNP simultaneously rise following lung resection surgery with BNP demonstrating an earlier peak in the post-operative period compared with NT-pro BNP.[49, 255, 256] The differing half-lives of BNP and NT-pro BNP and mode of degradation impact on their clinical utility.[255] It has been suggested that by peaking sooner, BNP may be more sensitive to an intra-operative myocardial insult in the lung resection population and lend itself to early identification of patients at increased risk of post-operative clinical deterioration and complications.[255] Being secreted from *both* ventricles, elevated BNP does not specifically reflect RV dysfunction but if it were elevated in the absence of LV dysfunction, this would signify RV dysfunction.[120] BNP has been shown to be associated with impaired RV function in a range of clinical settings. [257, 258]

5.1.1 B-Type natriuretic peptide as a biomarker in non-thoracic surgery

The European Society of Anaesthesiology (ESA) and the European Society of Cardiology (ESC) recommend the use of BNP for prognosis in patients at high risk of cardiac complications undergoing non-cardiac surgery.[215] The ESA and ESC also state, due to the evolving evidence in patients undergoing non-cardiac surgery, biomarkers cannot be proposed for routine use in all patients. Instead, BNP may be considered in high risk patients (METS ≤4 or with revised cardiac index value >1 for vascular surgery and >2 for non-vascular surgery).[215] An evidence gap has

been identified for the need of interventional studies considering BNP and other biomarkers in the peri-operative period.[215]

One of the most influential studies exploring BNP as a predictive biomarker is the *Measurement of Exercise Tolerance before Surgery* (METS) study: an international multicentre study including 25 hospitals and 1401 patients.[259] The authors prospectively explored the prediction of death or complications after major elective non-cardiac surgery using the biomarker NT pro-BNP. Patients were at least 40 years old, scheduled for major non-cardiac surgery and had at least one risk factor for cardiac complications[Q]. The primary outcome was death or MI within 30 days after surgery. Higher NT-pro-BNP concentrations predicted 30-day mortality or MI and 1-year mortality. When exploring prediction of 30-day mortality or MI, the authors created a baseline model composed of age, sex and revised cardiac index score. The predictive performance of this model improved with the addition of pre-operative NT-pro BNP (OR 1.78 CI 1.21- 2.62, p=0.003). Model discrimination increased from AUROCC of 0.70 to 0.71, with an NRI of 0.20, (p=0.02). The authors conclude-

> *'Natriuretic peptides should supersede subjective assessment for the estimation of peri-operative cardiac risk for major non-cardiac surgery'*
>
> *Wijeysundera et al 2018 [259]*

When exploring prediction of one-year mortality, Wijeysundera et al built another baseline model, this time only using patient's revised cardiac index score. With the addition of NT-pro BNP, the predictive strength of the model increased (OR 2.91 (CI 1.54 – 5.49, p=0.001). Model discrimination increased from an AUROCC of 0.65 to 0.72, with a net reclassification index (NRI) of 0.39. This is a large, well conducted study and these findings support the use of natriuretic peptides in peri-operative risk prediction strategies. However, further work would be required to define optimum cut-offs and other combinations of useful prognostic information that could be implemented in clinical practice.

[Q] History of heart failure, stroke, diabetes or coronary artery disease.

In another frequently sited large meta-analysis of 7 studies (>2800 patients) in 2009, Karthikeyan et al [260] examined the role of pre-operative BNP as an independent predictor of adverse cardiovascular outcomes within 30 days of non-cardiac surgery. Following robust search methodology, nine observational studies met eligibility. Pre-operative BNP measurement was an independent predictor of peri-operative cardiovascular events among studies considering an outcome of only mortality or MI (OR 44, 95%CI 7.6- 257, I^2 =51.6%). The wide confidence interval suggests further work may need to be done before BNP can reliably predict post-operative events. A further sub-analysis of only seven of the studies (two studies excluded; one because outcome was AF and the other because of the definition of MI did not match the others) demonstrated that an elevated pre-operative BNP remained predictive of cardiovascular outcomes at 30 days (OR 19.3, 95%CI 8.5 – 43.7). While the authors did not look beyond this initial 30-day post-operative period, this analysis highlights the predictive potential of BNP in non-cardiac surgery.

One study from within the meta-analysis by Karthikeyan et al (discussed above) was the first to explore BNP as a peri-operative biomarker of cardiac risk in patients undergoing non-cardiac surgery. In 1590 patients, Dernellis et al risk stratified patients using both the Goldman criteria [R] and BNP levels.[262] This is one of the largest and most influential studies concerning risk prediction using BNP. A level of ≥300pg/ml was considered to be high risk, with 81% of patients having a major adverse cardiac event. It is not obvious why a cut-off of 300pg/ml was selected. Dernellis et al also observed over 70 patients had *'potentially preventable cardiac events'*, which the addition of BNP may have alerted and decreased the mortality rate. The author concluded elevated BNP was an independent predictor for post-operative cardiac risk (OR 34, 95%CI 17 - 69, p<0.01).

[R] In 1977, Goldman and colleagues were the first to develop a pre-operative cardiac risk index in 1000 patients undergoing non-cardiac surgery. Nine variables were associated with increased risk for cardiac complications; pre-operative 3rd heart sound, MI within the last six months, more than five premature ventricular ectopic contractions, rhythm other than sinus on ECG, age > 70 years, intra-peritoneal/intra-thoracic or aortic operation, emergency operation, aortic stenosis and poor general medical condition. Each risk factor was assigned a point score and patients are stratified into four risk categories based on total points.[258]

Pre-operative BNP is a potential predictor of peri-operative cardiovascular complications in non-cardiac surgery; however, the significance of *post-operative* BNP is not as well explored. In a robust meta-analysis of 18 studies and >2000 patients undergoing non-cardiac surgery it was observed by Rodseth et al that the addition of *post-operative* BNP to a risk prediction model (which already contained *pre-operative* BNP), improved risk classification of mortality or non-fatal MI at both 30 days (NRI 20%, p<0.01) and 180 days (NRI= 0.11, p<0.01).[263] Post-operative BNP was also a strong independent predictor of death or non-fatal MI (primary outcome) at 30 days (OR 3.7, 95%CI 2.2 – 6.2, p<0.01) and 180 days (OR 2.2, 95% CI 1.9 – 2.7, p<0.01) following surgery.

Rodseth et al also demonstrated higher values of *post-operative* BNP were associated with an increased mortality and non-fatal MI within 30 days (primary outcome). For the primary outcome, a post-operative BNP level of 245pg/ml had an AUROCC 0.80 (95%CI 0.77 – 0.84) and independently predicted (OR 4.5, 95%CI 2.74 – 7.4, p<0.01) 30-day mortality or non-fatal MI. This level of 245 pg/ml was selected as it was the highest AUROCC discrimination point based on previous work identifying death or non-fatal MI.[264] The author explored BNP thresholds of 250pg/ml and >400pg/ml, as these cut-offs have been demonstrated to be useful in the detection of chronic heart failure in other research publications.[265] This was to examine if clinically useful groups could be created in the detection of cardiac complications (Table 13). Increased levels of post-operative BNP substantially increased the odds of the primary outcome.

BNP Value (pg/ml)	% (95% CI)	Adjusted Odds ratio (95% CI)
0-250	6.6 (4.7 – 9.2)	1
>250 – 400	15.7 (6.4 – 26.1)	2.5 (1.39 – 4.49)
>400	29.5 (20.7 – 37.8)	5.9 (3.71 – 9.26)

Table 13 – Post-operative BNP thresholds for incidence of mortality or non-fatal myocardial infarction 30 days after surgery.
Redrawn from Rodseth et al[264]. BNP = B-Type natriuretic peptide, CI = Confidence interval.

Detection of increased post-operative BNP levels *may* facilitate intervention or patient optimisation before these adverse events occur, with close post-operative monitoring. However, further studies are needed to determine the optimal post-operative time to measure BNP levels.

Adding BNP to peri-operative risk calculators appears to improve predictive strength, but its impact on patient outcomes has yet to be clarified.[254] The results of these two studies by Rodseth suggests post-operative BNP measurement may also provide prognostic information and be used to stratify cardiovascular risk following non-cardiac surgery.

5.1.2 B-Type natriuretic peptide as a predictor of functional outcome in non-thoracic surgery

The only population, other than lung resection patients, where peri-operative BNP is studied as a predictive biomarker of functional outcome has been in patients who have experienced ischaemic cardioembolic stroke. Rost et al[266] observed BNP predicts functional outcome in ischaemic stroke in a study of 569 patients. BNP quintiles were used for analysis and multivariate logistic regression was used to assess association between these quintiles of BNP and functional outcome. Elevated BNP (collected within 48 hours of hospital admission) decreased the odds of good functional outcome at 6 months (OR 0.64;95%CI 0.41-0.98) and increased the odds of death (OR 1.75;95% CI 1.36-2.24). Addition of BNP to multivariate models increased the predictive performance for functional outcome and mortality. However, this is not a comparable population to patients undergoing lung resection and the mechanisms driving elevated BNP may not be related.

5.1.3 B-Type natriuretic peptide in thoracic surgery

In 22 patients undergoing lung resection surgery, our research group has previously demonstrated moderate negative association between post-operative right ventricular ejection fraction (RVEF) (determined by cardiovascular magnetic resonance) and change in BNP levels in patients undergoing lung resection surgery.[267] There existed an association between post-operative BNP (day 2) and post-operative RVEF (day 2) ($r = -0.44$, $p=0.04$) (Figure 19). There also existed a moderate negative association between change in peri-operative BNP and change in RVEF (day 2- pre-op), ($r=-0.52$, $p=0.01$). This suggests those patients with the highest change in peri-operative BNP were those with the largest change in $RVEF_{CMR}$ at the same time point. In the same population, no association existed between change in peri-operative BNP levels (day 2) and change in LVEF (day 2), ($r = -0.29$, $p=0.19$) or post-operative BNP (day 2) and post-operative LVEF ($r=-0.06$,

p=0.78), suggesting BNP signal is being driven by the right ventricle and not the left ventricle.

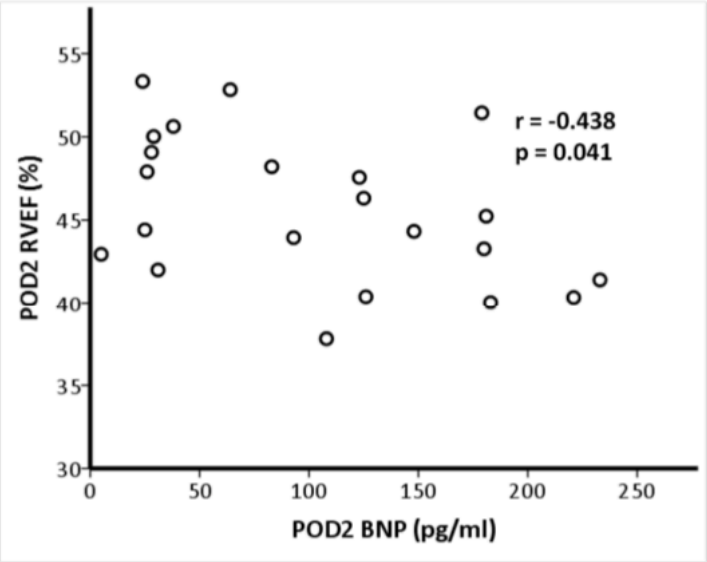

Figure 19 - Association between $RVEF_{POD2}$ and BNP_{POD2}.
Taken from McCall et al[268]. (n=22). RVEF = Right ventricular ejection fraction, BNP = B-Type natriuretic peptide, POD2 = Post-operative day two.

Additionally, McCall et al[267] also explored the prediction of post-operative right ventricular dysfunction (defined as right ventricular ejection fraction ≤45% at day 2) using BNP. Change (delta) in BNP (post-operative BNP day2 – pre-operative BNP) between those with RVEF>45% was lower than those with RVEF ≤45%, (p=0.02). Using AUROCC analysis, they demonstrated delta BNP was able to detect post-operative RVEF dysfunction (day 2) with an AUROCC of 0.78 (95%CI 0.58 – 0.99).

In summary, McCall et al[267] demonstrated consistent association between BNP and *RV function* on POD2. There was no change in LVEF over the duration of the study and no association between BNP and LVEF. This adds to the hypothesis that BNP is released in response to cardiovascular changes affecting the *right* and not the left ventricle, in patients undergoing lung resection.

5.1.4 B-Type natriuretic peptide as a predictor in thoracic surgery

There has been growing interest in BNP in the thoracic population, particularly in patients undergoing lung resection surgery. While peri-operative changes in BNP have been demonstrated to be useful in the prediction of post-operative morbidity and mortality.[269]

5.1.4.1 B-Type natriuretic peptide as a predictor of complications in thoracic surgery

In thoracic surgery, pre-operative BNP has been reported to be a predictor of post-operative atrial fibrillation.[269] In a meta-analysis, Simmers et al[270] combined 742 patients across five observational studies. The incidence of AF was 14.5% and an elevated pre-operative BNP was associated with post-operative AF (OR 3.13, 95% CI 1.38-7.12). The natriuretic peptide thresholds used varied between studies: two of the studies within this meta-analysis used BNP as a continuous variable and via post-hoc analysis identified optimal BNP threshold using the AUROCC of 160pg/ml NT-pro BNP and 30pg/ml BNP. The author concluded:

> *'Patients with an elevated pre-operative BNP level are at increased risk of post-operative AF and further work should be done to incorporate BNP into risk prediction modelling'*
>
> *Simmers et al 2015* [270]

5.1.4.2 B-Type natriuretic peptide as a predictor of dyspnoea in thoracic surgery

Little work has been done to explore the ability of BNP to predict *long term* functional outcome in any population, let alone in the lung resection population. Despite the evidence to support BNP as a potential predictor of post-operative complications, mortality and RV dysfunction, few have explored the role of BNP to predict long term functional capacity or dyspnoea following lung resection surgery.

In previous work from our research group, Young et al[49] observed in 27 patients that BNP levels were associated with subjective and objective markers of functional limitation following lung resection. BNP was a predictor of functional deterioration and showed a potential for use in risk stratification. Patients had BNP measured pre-operatively, post-operative day 1 and day 2 and at 2 months post-operatively. Functional assessments were based on 6-minute walk test and MRC dyspnoea scale scoring. Deterioration in functional capacity was defined as an increase in MRC score and/or a decrease in 6-minute walk distance. Seventeen patients (68%) demonstrated deteriorated functional capacity at 2 months. *Pre-*

operative BNP was higher in patients with a reduced 6-minute walk distance (p=0.01). There was a significant negative association between high pre-operative BNP levels and low 2-month post-operative 6-minute walk distances (r=-0.43, p=0.05). BNP was higher in the group showing functional deterioration compared to the group with no change at *all* peri-operative time points (p<0.01) (Figure 20). Importantly, pre-operative BNP was a predictor of functional deterioration at the 2-month time point with AUROCC of 0.82 (95%CI 0.65 – 0.99, p=0.01). A pre-operative BNP level of 46.5 pg/ml had a sensitivity of 58% and a specificity of 100% to predict a deterioration in functional capacity 2 months following lung resection surgery, giving a PPV of 100% and an NPV of 53%.

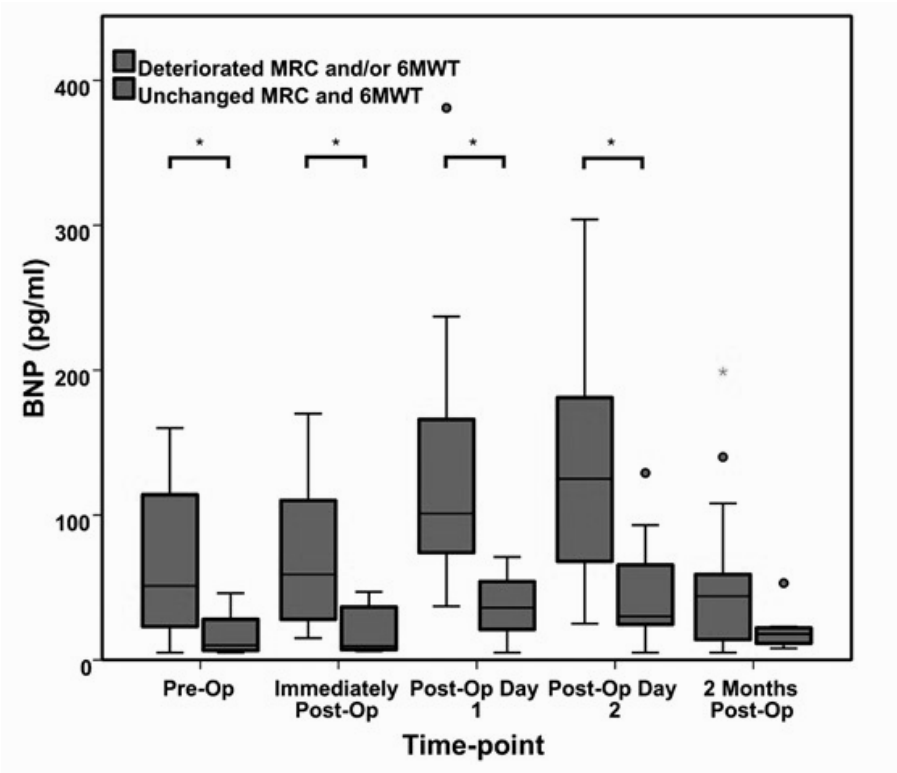

Figure 20 - Comparison of BNP levels between patients with deteriorated and unchanged post-operative functional capacity.
Change over time in patients showing functional deterioration, p<0.01. Change over time in patients with no functional deterioration, p<0.01. Comparisons between groups *= p<0.01. BNP = B-Type brain natriuretic peptide, MRC= Medical research council, 6MWT = 6-minute walk test, pre-op = pre-operative, post-op = post-operative. Taken from Young et al [49] (n=27)

Young et al[49] thus demonstrated the potential ability of BNP to predict functional outcome following lung resection in a small population from a single centre. The study described in this thesis was conceived to validate these findings in multiple tertiary cardiothoracic centres and determine if the addition of the biomarker BNP

to conventional pulmonary function based risk prediction could improve prediction of dyspnoea following lung resection for cancer.

5.1.5 B-Type natriuretic peptide in pulmonary embolism and pulmonary hypertension

It would be prudent to examine if any other analogous populations have evidence supporting the use of BNP as a predictive biomarker. Increased afterload is one of the proposed mechanisms driving RV dysfunction in the lung resection population, therefore it would be reasonable to explore the evidence of BNP as a biomarker in conditions which share this feature, such as pulmonary hypertension (PH) and pulmonary embolism (PE).

5.1.5.1 B-Type natriuretic peptide in pulmonary embolism

BNP has been found to be increased following PE with increasing serum levels associated with increased haemodynamic instability.[271] Furthermore, BNP has been found to have prognostic value in patients with pulmonary embolism, improving the prediction for the absence of major adverse cardiovascular events and a benign clinical course.[272] In 50 prospectively recruited patients, with confirmed PE by CTPA or echocardiogram, Kruger et al.[257] demonstrated BNP levels were increased in patients with PE who have RV dysfunction. Those without RV dysfunction had normal BNP levels in the absence of LV dysfunction. BNP elevation was highly predictive of RV dysfunction[S], but not of in-hospital mortality or complication rate. BNP discriminated patients with or without RV dysfunction (AUROCC 0.78 95% CI 0.64-0.92) (Figure 21). A BNP >90pg/ml (determined by Youdens index) was associated with a risk ratio of 28.4 (95% CI 3.22-251.12) for diagnosis of RV dysfunction. Sensitivity was relatively low at 64%, with a specificity of 94%.

[S] Diagnosis of RV dysfunction was made in the presence of any of these criteria: 1) Dilation of the right ventricle (diastolic diameter >30 mm) or a RV/LV end-diastolic diameter ratio >1 in the 4-chamber view. 2) Hypokinesis of the right ventricle. 3) Abnormal motion of the interventricular septum. 4) Tricuspid valve regurgitation (jet velocity >2.5 m/s).

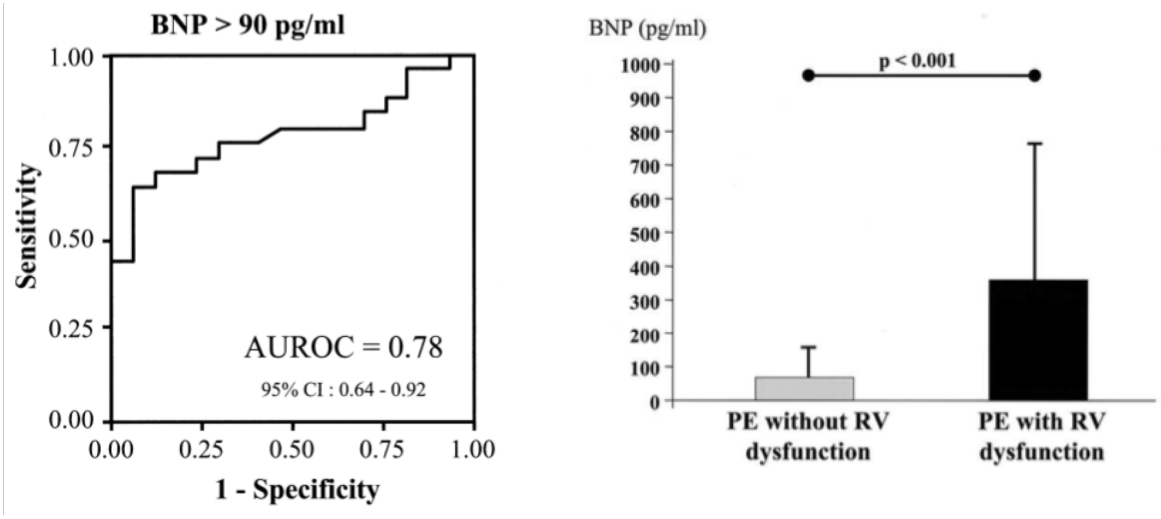

Figure 21 – BNP levels in patients with RV dysfunction in 50 patients with acute pulmonary embolism.
ROC curve of BNP for prediction of right ventricular dysfunction (left figure). BNP concentrations of patients with and without RV dysfunction (right figure). (n=50). Taken from Kruger et al.[257]

Confirming this finding in another study, Choi et al[273] observed BNP was a significant predictor of RV dysfunction during PE in 84 patients. The author measured NT-pro BNP and found a level of >620pg/ml was an independent predictor of RV dysfunction, after adjustment for baseline characteristics (OR 5.04 95%CI 2.35-9.59, p<0.01). The value of 620pg/ml was based on Youdens index, on AUROCC analysis. Other biomarkers such as troponin were measured but BNP displayed the best sensitivity and specificity to predict RV dysfunction. Choi et al concluded, simple measurement of biomarkers could be useful to predict the presence of RV dysfunction, especially in hospitals where echocardiography is unavailable.

Kline et al[274] compared eight biomarkers for prediction of RV dysfunction and defined adverse outcomes[T] six months after PE. The author observed only BNP and Troponin-I had significant prognostic use, with BNP being the best to predict mortality. This was a prospective study that recruited 152 patients with a complete dataset for analysis; 37 (24%, 95% CI 18-32%) patients had RV dysfunction at 6 months, diagnosed with echocardiography. BNP had an AUROCC 0.71 (95%CI 0.60-0.81) for predicting RV dysfunction/hypokinesis at 6-months. Overall

[T] Defined as; presence of dyspnoea at rest on more than one half of days or exercise intolerance based upon a six-minute walk test <330m at 6-month follow up.

mortality was 8.5% (n=13); mortality for those with a BNP >100pg/ml was higher than those with BNP<100pg/ml, (p=0.003).

5.1.5.2 B-Type natriuretic peptide in pulmonary hypertension

Patients with chronic pulmonary hypertension may also represent a clinically analogous population to patients following lung resection surgery with increased afterload, albeit over a longer time frame. Chronic RV dysfunction is associated with increased BNP levels, the degree of RV dysfunction being proportional and reflective of the plasma increase.[258] Right ventricular systolic dysfunction before treatment in pulmonary hypertension (baseline) predicts early death.[258] BNP is also predictive of decreased survival in this population..[275]

The 2016 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension advocate NT-pro BNP as an independent risk predictor in this population.[276]

> *'The biomarkers BNP and pro-BNP remain the only biomarkers consistently used in pulmonary hypertension, correlating with myocardial dysfunction and providing prognostic information at the time of diagnosis and during follow-up assessment'*
>
> *Galiè et al 2016* [276]

There is no clear advantage of using BNP versus NT pro BNP, although BNP appears to have slightly more association with pulmonary haemodynamics and is less affected by kidney function.[276] In PAH risk stratification, the ESC/ERS guidelines are as follows; low risk PH patients are those patients with a baseline BNP <50ng/L, intermediate risk patients have a BNP level 50-300 ng/L and high-risk patients are >300ng/L. The effectiveness of therapy is also monitored with levels of BNP, with a target normalising concentration.

One of the most widely referenced papers exploring biomarkers in PH is by Nagaya et al.[275] The author sought to assess the prognostic significance of plasma BNP in patients with PH. BNP was measured in 60 patients with PH at diagnosis, before any treatment had been commenced. BNP measurements were repeated in 53 patients with a mean follow-up period of three months. After multivariate

analysis, baseline plasma BNP levels were an independent predictor of mortality: patients with BNP levels >150pg/ml having a lower survival, ($p<0.05$) (Figure 22).

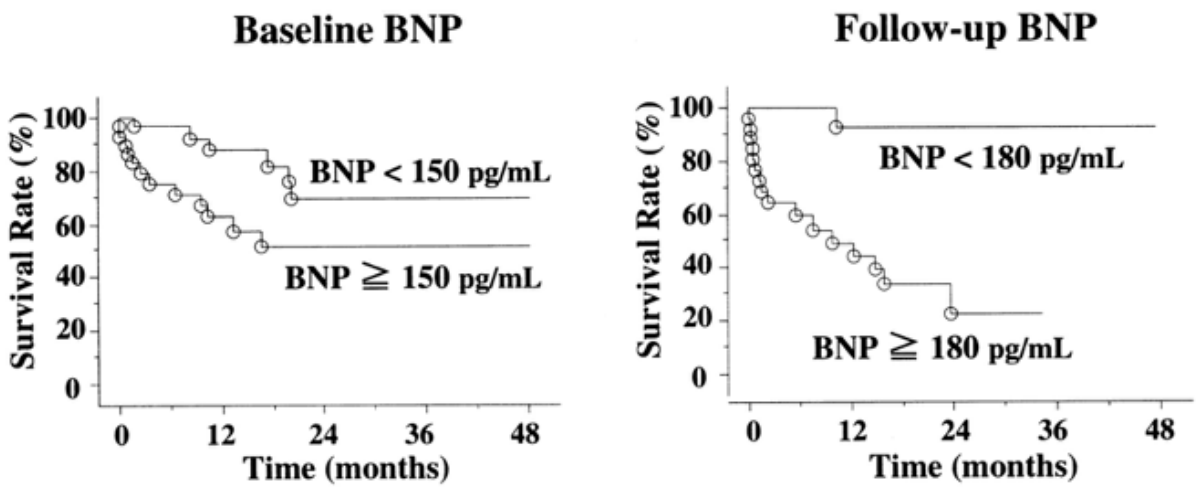

Figure 22 - Kaplien-Meier survival curves according to median value of baseline and follow-up BNP in patients with PH.
Patients with baseline BNP <150pg/ml had increased survival. At 3-month follow-up BNP <180pg/ml had increased survival. Taken from Nagaya et al[275]. (n=53). BNP = B-Type natriuretic peptide.

5.1.5.3 B-Type natriuretic peptide in thoracic surgery: conclusion

BNP has been explored in the non-thoracic population and is recommended to compliment pre-operative cardio-respiratory risk stratification in high risk patients.[215] BNP cut-off values which are most sensitive and specific for predicting outcome have been published within the non-thoracic population.[215] However, data on the use of pre-operative biomarkers are sparse and consequently universal pre-operative routine BNP measurement for risk stratification and to prevent cardiac events is not recommended. Instead, BNP measurements are considered to obtain independent prognostic information for peri-operative and late cardiac events in high risk patients.[215]

Our research group has demonstrated firstly a negative association between post-operative right ventricular ejection fraction (RVEF) and change in BNP levels in patients undergoing lung resection. Secondly, pre-operative BNP was an independent predictor of functional deterioration in a small cohort and showed a potential for use in risk stratification. Furthermore, some evidence exists that BNP is an independent predictor in conditions with analogous physiology of increased RV afterload, such as PH and PE. Despite its potential, there has been little work to explore BNP's role to predict long-term functional outcome in thoracic lung resection surgery.

6 Hypotheses and aims

Following lung resection surgery for cancer, survival with a meaningful quality of life and acceptable exercise capacity is extremely important to patients. However, it is generally recognised conventional prediction of post-operative functional capacity and dyspnoea is challenging, relying on pulmonary function testing and calculation of predicted post-operative FEV_1% (ppoFEV_1%) and DLCO% (ppoDLCO%), (Chapter 4). No effective method exists for identifying risk of, nor therapeutic strategies to prevent, post-operative dyspnoea.

The aim of this thesis was to improve conventional prediction of post-operative dyspnoea. Pilot data from our research group demonstrated association between B-Type natriuretic peptide and both; post-operative cardiac dysfunction and post-operative dyspnoea. The author proposes a novel scoring tool incorporating B-Type natriuretic peptide alongside conventional measurements. B-Type natriuretic peptide has been advocated by major international guidelines to improve prognostication of peri-operative morbidity in high-risk patients prior to non-cardiac surgery, yet its potential role in peri-operative decision making in lung resection is unclear. No previous work has compared B-Type natriuretic peptide to functional outcomes following lung resection.

The first investigation of this thesis (Chapter 8) examines conventional risk prediction methods in a single site derivation population at the Golden Jubilee National Hospital. It is anticipated results will confirm poor performance of conventional methods to predict post-operative dyspnoea and confirm the sole use of pulmonary function in this setting could be improved. The addition of B-Type natriuretic peptide to conventional methods will then be explored and the hypothesis is that this will improve prediction of post-operative dyspnoea. Furthermore, the use of pulmonary function as a continuous variable (without a 40% cut-off) will be analysed with the hypothesis this will also improve prediction.

New models will then be created (Chapter 9), examining variables that could improve the prediction of post-operative dyspnoea. Variables significant at univariate analysis will be used to derive new risk prediction models, incorporating B-Type natriuretic peptide. The hypothesis is that these *new* models will improve prediction of dyspnoea within the internal dataset.

An external dataset from three other UK sites will be used to validate these new models (Chapter 10). Again, the hypothesis is that the *new* models created will increase predictive strength within this external dataset and highlight the potential of variables other than pulmonary function in this setting.

Secondary analyses will examine association between peri-operative B-Type natriuretic peptide levels, post-operative morbidity and length of hospital stay. It is anticipated increasing peri-operative BNP levels will be associated with increased morbidity and length of hospital stay.

7 Generic Methods

This section outlines the methodology in conducting *PROFILES* study (bn**P** for p**R**ediction of **O**utcome **F**ollow**I**ng **L**ung r**E**section **S**urgery). The statistical analysis plan (SAP) was developed in conjunction with the Robertson centre for Biostatistics (University of Glasgow) prior to conducting analysis on the internal and external datasets.

7.1 Ethical approval

Ethical approval was obtained from the London Queen Square Research Ethics Committee (REC Ref: 18/LO/1563/AM01, approval date 13th September 2018). (Appendix 5)

7.2 Study design

Multicentre prospective observational cohort study of patients presenting for lung resection by lobectomy or pneumonectomy by either video assisted thoracoscopic surgery (VATS), robotic assisted thoracoscopic surgery (RATS) or thoracotomy.

7.3 Study setting

Multicentre study led by the Golden Jubilee National Hospital/West of Scotland Heart and Lung Centre (GJNH) which is a tertiary referral cardiothoracic surgery centre and is one of the largest thoracic surgery units in the UK. The other three collaborating centres are the Royal Victoria Hospital, Belfast (Regional Cardiothoracic Surgery Unit providing service across Northern Ireland), Aberdeen Royal Infirmary (North of Scotland Cardiothoracic Surgery Unit) and Edinburgh Royal Infirmary (East of Scotland Cardiothoracic Surgery Unit).

7.4 Patient population

Potential study participants were identified from the waiting lists of surgeons involved in the study. All patients at the GJNH were screened by the author. After agreement with the operating surgeon, those patients not meeting any exclusion criteria were approached by a member of the research team, provided with an information leaflet and given a verbal outline of the study. This would generally

happen at a pre-operative assessment clinic around a week before surgery. Approximately 65% of the patients recruited from the GJNH were consented by the author.

Patients were advised a designated member of the research team would see them once admitted to hospital prior to their operation. This provided time to discuss the study further and answer any questions they may have. If appropriate, informed consent was obtained and those patients providing informed consent went on to participate in the study.

Inclusion criteria were; provision of informed consent, age >16 years and planned elective lobectomy/pneumonectomy lung resection by VATS, RATS or thoracotomy. Exclusion criteria were; pregnancy, wedge/segmental/sub lobar lung resection, on-going participation in any investigational research which could undermine the scientific basis of the study, conditions that disproportionately increase BNP such as sepsis, cirrhosis, colon cancer, any intracranial pathologies (see definitions below, section 6.5) and Medical Research Council (MRC) dyspnoea score > 2 pre-operatively. Excluding pathologies which disproportionately increase BNP would also create a more representative cohort for analysis.

Given the time sensitive nature of cancer surgery, patients often progressed to hospital admission without attending pre-assessment clinic. This precluded identification and consent by the approach detailed above. In these situations, patient information leaflets were provided by post along with a cover letter inviting patients to read about the study prior to hospital admission. Patients were then approached for recruitment on hospital admission as detailed above.

7.5 Definitions of exclusion criteria

Sepsis - Any patient presenting with life threatening organ dysfunction due to dysregulated host response to infection as per NICE guidelines 2017.[277]

Colon Cancer – Any patient with active colonic cancer.

Cirrhosis – Patients with a previous diagnosis of cirrhosis.

Intracranial pathology – Acute stroke with new onset neurological deficit, intracranial cancer of any type or sub-arachnoid haemorrhage within the last year.

7.6 Justification of inclusion/exclusion criteria

Nationally, over half of all lung resections are performed by thoracotomy, with the remainder being performed by minimally invasive techniques (Minimally Invasive Thoracic Surgery (MITS)).[12] At the GJNH, one of the largest thoracic surgical unit in the UK, MITS now account for >80% of resections. Inclusion of both thoracotomy, RATS and VATS allows for broad participation in collaborating hospitals and ensures wide applicability of results outside the participating centres. Sub-lobar resections were excluded.

As an MRC score >2 was defined as the primary outcome of the study, all patients had, by definition to have MRC of 2 or less pre-operatively. This is in keeping with identifying patients at risk of developing disabling dyspnoea post-operatively – patients with MRC >2 arguably already have disabling dyspnoea at presentation. To reduce the risk of false positive results, any non-cardiac, medical conditions resulting in elevated BNP were excluded.

7.7 Consent

Following a face-to-face meeting where patients were given the opportunity to ask questions, written informed consent was obtained. This took place on admission to hospital, usually the day before surgery.

7.8 Site initiation visits and training

Prior to recruitment commencing at each of the three collaborating centres, site initiation visits and training sessions were completed. This comprised of two presentations covering the main points of the study and provided an opportunity for questions. All paperwork was made available and a thorough explanation of data collection was provided. At this visit, research staff underwent training on the *Abbott i-STAT* point of care BNP analysis system.

7.9 Anaesthetic protocol

The intention of this study was to determine the response to lung resection and was not designed to assess the impact of specific anaesthetic or surgical techniques. For that reason, anaesthetic, surgical and post-operative management was left at the discretion of each participants treating team. Given the potential interest in surgical approach (thoracotomy / VATS / RATS), anaesthetic technique (TIVA or volatile), analgesic technique (epidural, PVB or other) and intra-operative ventilatory conduct on post-operative outcomes; all these variables were collected as part of the case report form with the intention of analysing them as secondary outcomes.

7.10 Data collection

Data were collected during the participants hospital admission and at 3-month follow-up. At the GJNH, this was performed by the author and designated research nursing staff. At collaborating sites, this was performed by a member of the designated research team. If a participant was not returning to the Golden Jubilee at 3-months as part of the described sub-study, a postal questionnaire was sent to their home address. All anonymised data were collated and stored in a password protected database on a secure NHS computer. A patient identification list was retained in a secure location within the research department of whichever hospital the patient had been recruited from.

The data from each participant at each of the collaborative sites was sent in anonymised format to the GJNH for analysis. The research team at the GJNH did not have access to any other patient details from each of the participating centres. If a patient needed to be contacted for any reason, this was done by the local recruitment team. Data collection involved baseline demographic data, self-reporting exercise tolerance, laboratory sampling and post-operative clinical data.

7.11 Baseline demographic data

Patient demographics were collected at the time of recruitment. Patient data was recorded manually on dedicated case report forms by one of the research team

and then entered into an electronic spreadsheet (Excel for Mac, Microsoft®, Version 16, 2018), and reviewed for errors by the author. Case notes were reviewed and a face-to-face interview conducted to allow completion of the case report form – between a member of the research team and the patient. At collaborating sites, trained research staff followed the same process. Again, this was checked for errors by the author and any data queries resolved.

Pulmonary function tests were performed in all patients prior to attending consultation with their thoracic surgeon. Where these had not been performed prior to this appointment, they were carried out by respiratory physiologists according to standardised guidelines. Thoracoscore was calculated for each patient using an online calculator.[278] An explanation of this score and the parameters required to calculate the score are described in section 4.2.2. Co-morbidities of note included; smoking addiction, history of cancer, COPD, arterial hypertension, heart disease, diabetes mellitus, peripheral vascular disease, obesity, alcoholism and gastro oesophageal reflux disease. An American Society of Anaesthesiology (ASA) grade was obtained for each patient, along with medication history and BMI.

Once the type of resection was confirmed post-operatively, ppoFEV_1% and ppoDLCO% were calculated (see section 4.3.3 - Equation 1).

7.12 Self-report of dyspnoea tolerance and quality of life

Self-reported functional status was recorded by completion of a written questionnaire encompassing WHO performance status classification,[279] visual analogue pain scale,[280] health related quality of life scoring by EQ-5DL[281] and QLQ-C30[282] questionnaires and WHO disability assessment schedule.[283] Breathlessness was assessed using the MRC scale[71] and University of California and San Diego Shortness of Breath Questionnaire (UCSD-SOBQ).[81] Anxiety and depression scores were collected using the Hospital Anxiety and Depression Score (HADS) questionnaire.[284] Pain assessment was recorded using the Brief Pain Inventory (BPI)/Visual Analogue Scale (VAS).[285]

These scores were completed pre-operatively and three months post-operatively. For the three-month follow-up, two attempts were made at postal follow up and

if no response was obtained, a reminder phone call was performed and a third postal questionnaire issued. If there was still no response following these attempts, then this was abandoned and the patient deemed 'lost to follow-up'.

There are many scoring tools to measure QoL, dyspnoea and anxiety and depression in lung cancer population. Therefore, selecting appropriate questionnaires is difficult. The questionnaires selected by the author for use in the study are commonly used within the medical literature and are validated to measure dyspnoea, quality of life, performance status and anxiety/depression.

The author selected to use the methods listed above to measure dyspnoea for several reasons. The MRC scale is well-established tool to quantify breathlessness which has been used in previous studies within our research group. Its main strength lies in its simplicity, as discussed previously in section 2.1.3. Finally, the UCSD-SOBQ was selected as this provides more detail into changes in peri-operative dyspnoea within the study, and so may be better equipped to detect subtle peri-operative changes.

Measurement of quality of life can be challenging given its broad definition and the large range of questionnaires in existence. The author selected the following tools to assess quality of life for multiple reasons; pilot work used the same QoL questionnaires, the questionnaires are validated for lung cancer with well recognised, discrete minimally clinically important differences (MCID) and they are well established to measure quality of life within the lung cancer population. The questionnaires used to measure QoL are multidimensional and encompass physical, social, cognitive, emotional, work and role related responses and sometimes also include disease related symptoms.

7.13 Laboratory sampling

Prior to induction of anaesthesia, a baseline 3ml EDTA blood sample was collected. A further 3 ml sample was taken on the morning of post-operative days 2 and 3. Where possible, to avoid unnecessary venepuncture, blood samples were obtained from a radial arterial cannula (routine for major thoracic surgery in all participating centres) or collected along with bloods required for normal clinical care. Additional baseline measures for haemoglobin, albumin, creatinine,

estimated glomerular filtration rate and C-reactive protein were analysed. These form part of the usual bloods taken for clinical care.

B-Type Natriuretic Peptide was analysed using the *Abbott i-STAT point of care BNP system* (Abbott point of care UK Berkshire, UK). Quality control measures were undertaken according to the manufacturer's guidelines. All samples were analysed within 30 minutes of collection and subsequently disposed of. This protocol was followed across all collaborative sites. The immediate managing clinicians were blinded to BNP results unless preoperative levels were >100pg/ml. In this instance, a discussion with the treating clinical team took place and investigations, such as echocardiogram were requested on a case-by-case basis. This may be considered an intervention and a change from routine care as pre-operative BNP would not routinely be checked prior to surgery. Therefore, ethical approval was sought for this process and an explanation given to patients within the patient information leaflet. Any abnormal blood test results were reported to the patient's general practitioner and discussed with the surgical team. The 100pg/ml cut-off was selected in line with primary care thresholds for further cardiac investigations and echocardiogram.

External quality control for BNP was performed through UK National External quality assessment (NEQAS, https://ukneqas.org.uk). This involved monthly sample analysis using national externally validated samples. Clinicians were blinded to NEQAS results and entered the results of the high and low sample into an online database. Monthly reports ensured the *i-STAT* machine was functioning within acceptable limits.

7.14 Intra-operative clinical data

Intraoperative anaesthetic data were collected automatically and continuously for the duration of surgery using the 'RECALL Anaesthetic Intra-operative Management System (AIMS) electronic charting system (Informatics Clinical Information Systems Limited, Glasgow). Duration of anaesthesia was prospectively recorded in this system by the anaesthetic team. Alternatively, if RECALL was not used, the appearance of end tidal carbon dioxide ($ETCO_2$) or the point of anaesthetic induction by drug charting was used to indicate the start and end of anaesthesia. The end of surgery was recorded and if this did not take place, the

loss of $ETCO_2$ or stoppage of anaesthetic agent was used as a surrogate. Arrival to and exiting from theatre were not used, as these values are not reflective of anaesthetic or surgical time. Duration of one lung ventilation (OLV) was taken as the time recorded by the anaesthetist, or where this did not happen, from inspection of the tidal volume and airway pressures vs. time curves. Other data recorded included; analgesic technique performed (paravertebral injection/catheter, epidural, patient-controlled analgesia), anaesthetic technique utilised (total intravenous or volatile) and type of volatile agent used (if applicable).

As changes to the planned surgery could be made, operation performed was confirmed post-operatively. Changes to the planned procedure could result from intra-operative findings, indeed some patients had planned frozen section biopsies and depending on these results did not proceed to having an anatomical lung resection at all. Other data recorded included side of surgery and operation type (thoracotomy, video assisted or robotic assisted).

7.15 Post-operative clinical data

Post-operative data were collected on a daily basis by the author and members of the research team. Duration of high dependency unit (HDU) stay was automatically recorded by the ICU clinical information system (Centricity CIS; GE Healthcare, Buckinghamshire, UK) and displayed in hours. Post-operatively, delayed discharge from HDU were common; patients are routinely not discharged overnight from HDU for safety reasons and downstream wards can sometimes be pressured for bed spaces. Therefore, if patients moved from continuous monitoring to intermittent monitoring this was deemed reflective of a stepdown of care requirement from level 2 care to level 1 care and the time of cessation of continuous monitoring taken as the HDU discharge timepoint.[286]

Duration of hospital stay was from day of surgery until the day of discharge. This was a pragmatic decision, whilst most patients are admitted the day before surgery at the Golden Jubilee National Hospital, some are admitted the day of surgery. Conversely, some are admitted several days before their operation for geographical reasons, or to allow essential preoperative tests to be completed.

Other parameters collected included; development of new-onset atrial fibrillation (confirmed by 12 lead ECG and recorded in the notes by the clinical team) and treatment received, need for vasopressor/inotrope infusions and their duration, need for nasal high flow oxygen/non-invasive ventilation and duration, need for Intensive care unit (ICU) admission and duration. Hospital mortality and its time from surgery was also recorded.

Complications were recorded in line with the European Society of Thoracic Surgeons definitions and included[287];

1. Atelectasis confirmed by chest x-ray and documented in medical notes by medical staff.

2. Pneumonia defined according to the latest CDC criteria.[288] Two or more serial chest radiographs with at least one of the following: new or progressive and persistent infiltrate, consolidation, cavitation in addition to at least one of the following:

- Fever (>38°C) with no other recognised cause
- Leukopenia (<4000 WBC/mm3) or leukocytosis (>12,000 WBC/mm3)
- for adults >70 years old - altered mental status with no other recognised cause.

With at least two of the following:

- New onset of purulent sputum or change in character of sputum
- Increased respiratory secretions
- Increased suctioning requirements
- New onset or worsening cough
- Dyspnoea or tachypnoea
- Rales or bronchial breath sounds
- Deteriorating gas exchange (e.g. Hypoxia (reduced PaO_2/FiO_2 ratio), increased oxygen requirements or increased ventilator demand).

3. Acute Respiratory Distress Syndrome (ARDS). Onset within 1 week of known clinical insult or new/worsening respiratory symptoms, bilateral infiltrates not explained by effusions on chest radiograph or CT scan, respiratory failure not fully explained by cardiac failure or fluid overload. Severity grading; mild, moderate and severe based on PaO_2/FiO_2 ratio.

4. Pulmonary aspiration confirmed by chest x-ray or clinical suspicion and documented in notes by clinical team.

5. Pulmonary embolism confirmed by ventilation perfusion (V/Q) scan or CT pulmonary angiogram.

6. Atrial arrhythmia defined as new onset of atrial fibrillation/flutter (AF) requiring medical treatment or cardioversion not including recurrence of AF which had been present pre-operatively.

7. Ventricular arrhythmia defined as sustained ventricular tachycardia or ventricular fibrillation that has been clinically documented and treated by ablation therapy, implantable cardioverter defibrillator, permanent pacemaker, pharmacologic treatment or cardioversion.

8. Myocardial infarction (MI) evidenced by one of the following criteria:

- Transmural infarction diagnosed by the appearance of a new Q wave in two or more contiguous leads on ECG

- Sub endocardial infarction (non-Q wave) evidenced by clinical, angiographic electrocardiographic signs

- Laboratory isoenzyme evidence of myocardial necrosis with suspected acute coronary syndrome and high sensitivity cardiac troponin I level >5ng/L.[289]

9. Deep Venous Thrombosis (DVT), documented and confirmed by ultrasound scan.

10. New onset renal failure in the post-operative period according to one of the following criteria: Increase of serum creatinine to a minimum of 2-fold the pre-operative creatinine level or a new requirement for dialysis post-operatively.

11.Urinary retention confirmed by documentation in medical notes by clinicians looking after the patient.

12. Hypotension defined as occurrence of systolic blood pressure <90mmHg requiring fluids or drug therapy.

13. Neurological complication defined as occurrence of one of the following; central neurologic post-operative events not present pre-operatively, central neurologic deficit persisting post-operatively for more than 72 hours, transient neurologic deficit (transient ischemic attack or reversible ischemic neurological deficit) with recovery within 72 hours, new post-operative coma persisting at least 24 hours and caused by anoxic/ischemic and/or metabolic encephalopathy, thromboembolic event or cerebral bleed.

Post-operative clinical data and complications were recorded prospectively by the author and members of the research team on a daily basis until patient discharge. These data were corroborated with medical notes. The severity of all post-operative complications was graded using the Clavien-Dindo classification system. Complications were divided into:

Minor

- Grade 1 – without need for pharmacological treatment or other intervention
- Grade 2 – requiring pharmacological treatment or minor intervention only

Major

- Grade 3 – surgical radiological endoscopic or multitherapy required
- Grade 3a – intervention does not require general anaesthesia
- Grade 3b – intervention requires general anaesthesia
- Grade 4 – requires intensive care unit management and life support

- Grade 4a – single organ dysfunction
- Grade 4b – multi-organ dysfunction intervention required

Mortality

- Grade 5 - death of the patient

7.16 Sample size and power

Previous work by our group demonstrated BNP levels in patients undergoing lung resection was associated with deterioration in functional capacity, defined as; an increase in MRC dyspnoea score and/or a significant decrease in 6-minute walk test distance.

A power calculation was performed to allow testing of the hypothesis that a risk score incorporating BNP would improve the AUROCC for the predictive model, based on traditional lung function testing, from 0.7 to 0.8. According to the method of Hanley and McNeil and based on a 25% incidence of the outcome of interest, 156 patients would be required to demonstrate an improvement in AUROCC from 0.7 to 0.8 with 5% significance and 80% power. This was repeated for an improvement in AUROCC from 0.6 to 0.7 with a required sample size of 176 patients. Based on this more conservative figure and allowing for dropouts due to a 3-month mortality of 6% (reported in large national studies, but less in previous work by our group) and a further 25% dropout from patients not returning questionnaires, a sample size of 250 patients was to be recruited; 100 for the GJNH and 50 each from the other three centres. This power analysis was further validated by bootstrapping.[U]

7.17 Data synthesis and statistics

All statistical analyses were performed using SPSS for Mac, version 26 (IBM, Armonk, NY, USA). A p-value <0.05 was considered statistically significant and absolute p-values are reported. When SPSS produced an output of p=0.000 this

[U] Bootstrapping resampling is a statistical procedure that resamples a single dataset to create simulated samples.

was presented as p <0.0005. Unless indicated, no adjustments were made for multiple comparisons.

Data are presented as mean (SD) or median (IQR) as appropriate to distribution. Normal distribution was determined by visual inspection of data distribution and with use of the Shapiro-Wilk or Kolmogogorov-Smirnov tests of normality. Where possible, data were transformed to normality for analysis. If not possible, an appropriate non-parametric test was used as an alternative.

Changes over time were assessed using one-way repeated measures analysis of variance (ANOVA) with post-hoc comparisons using the paired t-test. Friedman's test with post-hoc comparisons using the Wilcoxon rank sum test was used to asses changes over time for non-parametric data. Comparisons of parametric data between unpaired groups was made using Students t-test or one-way analysis of variance as appropriate. The Mann-Whitney U-test or Kruskal-Wallis test was used to compare non-parametric data between unpaired groups. Chi-squared or Fishers exact test were used to compare categorical variables.

Associations between continuous variables were visually inspected and assessed using Pearson's or Spearman's correlation coefficient as appropriate. Strength of association between variables obtained from correlation coefficients were interpreted as displayed in Table 14.

Correlation coefficient (r)	Interpretation
0 – 0.19	very weak
0.20 – 0.39	weak
0.40 – 0.59	moderate
0.60 – 0.79	strong
0.80 – 1.00	very strong

Table 14- Interpretation of strength of association by correlation coefficients.
British Medical Journal – correlation and regression [290]

The ability of continuous variables to predict binary outcomes was determined using the area under the receiver operating characteristic curve (AUROCC) Table 15. Optimal cut-off points of continuous variables were selected at points of maximal sensitivity and specificity (Youden's index). Positive and negative predictive values were calculated (PPV & NPV, respectively).

AUROCC Value	Discrimination Value
0.5 – 0.69	Poor discrimination
0.7 – 0.79	Acceptable discrimination
0.8 – 0.89	Excellent discrimination
>0.9	Outstanding discrimination

Table 15 - AUROCC interpretation
Adapted from Hosmer et al[291]. AUROCC = Area under receiver operator curve characteristic value.

7.18 Primary outcome analysis

The primary outcome was defined as the proportion of patients with a Medical Research Council dyspnoea score > 2, at three months post-operatively. The univariate characteristics between patients who were breathless (MRC score >2) and not breathless (MRC score ≤2) at three months were compared.

The original analysis plan was to construct a new risk prediction model for breathlessness following surgery incorporating patients recruited from all sites within the study, (n=250). The aim was then to compare conventional prediction of dyspnoea with and without the addition of pre-operative BNP across these 250 patients. Due to the COVID-19 global pandemic and a delay in patient recruitment, the analysis plan had to be changed during the study in order to facilitate timeous completion of this thesis. The recruitment of patients from the Golden Jubilee National Hospital (GJNH) finished much earlier than the other centres, which had all been more affected by the pandemic. Therefore, the analysis plan changed to recruit 125 patients from the GJNH and derive a new scoring tool, before validating with 125 patients from an external dataset – containing patients from the other three external sites. Recruitment of 125 patients from the GJNH was more than initially anticipated but was planned because the hospital was recruiting well, while the external centres were struggling. Sample size calculations for validation are usually obtained from a derivation cohort once outcome has been established but given the challenges of COVID, these targets were set based on circumstance. While the ethos of the original investigation was maintained, rather than analyse the entire dataset together the data was split to provide two equally sized derivation/validation datasets. The aim remained to test conventional prediction methods with and without pre-operative BNP.

All variables were compared in patients with and without dyspnoea at 3-months at univariate analysis, and were input to logistic regression to create new risk prediction models. All variables whose univariate tests resulted in a p-value < 0.10 were considered in the models. Eight models were derived.

Models 1-4 represented '*conventional'* risk prediction, first with ppoFEV_1% and ppoDLCO% dichotomised, then as continuous variables. The performance of these models was then explored with the addition of pre-operative BNP.

Models 5-8 were created to derive a '*new'* risk prediction model for post-operative dyspnoea. These models explored new variables (significant at univariate analysis) selected using forwards logistic regression. Once again, the additional predictive value of BNP was explored within these new models 5-8.

Area under the receiver operating characteristic curves (AUROCC) were computed as descriptive tools of the model's predictive capability. Optimal cut-off points for the prediction models were selected at points of maximal sensitivity and specificity (Youden's index). The Hosmer-Lemeshow goodness of fit test was used to examine the fit of the models. Patients were classified into high and low risk groups, using the risk prediction models.

The relative predictive strength of the models was evaluated and compared using;

- A comparison of AUROCC between models, using the statistical packages SPSS and Medstats online (pairwise comparison).

- Brier scoring (explained in section 7.21)

- Net reclassification indexing (explained in section 7.22).

After this comparison, the best models at internal derivation were then carried forward to external validation. Validation was assessed by firstly assessing model discrimination (AUROCC values). Next, calibration plots were drawn and visually inspected to compare predicted and observed probabilities. Finally, overall

model performance was described using sensitivity, specificity, positive predictive values and negative predictive values.

Calibration is the accuracy of estimates relating to the agreement between the estimated and observed number of events.[292] Calibration is important when aiming to use the model in clinical decision making, even when discrimination is moderate. Reporting on calibration is recommended by the Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines for prediction modelling studies.[293] Ideally to obtain a precise calibration curve it has been suggested a large sample size (> 200 patients *with* and 200 patients *without* the event) is needed, but not essential.[294] Calibration curves can still be drawn in smaller samples, however should be interpreted accordingly. This study did not have >400 patients, however calibration curves were still constructed to assess the degree of under/over estimation of risk.

Decile and Quintile calibration plots were drawn for each model to be external validated; a grouped calibration plot is a graphical analogue of the Hosmer-Lemeshow goodness of fit test for logistic regression models. Subjects were split into 10 groups using deciles of the predicted probability of the logistic model. The mean predicted probability for each group was calculated. The means of the empirical binary responses were calculated and plotted against the mean predicted probability of these 10 ordered pairs, displaying also the 95% confidence intervals. The mean of each group was connected with piecewise line segments and loess curves, to smooth the points for visual inspection. A diagonal line was also displayed on the calibration plots; to provide a reference marker reflecting theoretical perfect calibration.

Several authors dispute the use of 10 groups/deciles stating this is an arbitrary figure, using 5 or 15 could also be justified and there is no theory to guide the choice of splitting into deciles.[295]

7.19 Secondary outcome analyses

Peri-operative variables in patients with dyspnoea at 3 months (primary outcome) and those without were compared including;

- Acute complications, intra-operative details, duration of hospital stay and post-operative details such as duration of high dependency/intensive care stay and duration of hospital stay.

- In hospital, post-operative mortality.

Change in BNP over time was assessed by visual inspection of distributions at each peri-operative time point. Differences between pre-operative and post-operative (post-op day 2 and 3) levels were explored. Association between B-Type natriuretic peptide and acute post-operative complications, post-operative cardiopulmonary complications and duration of hospital stay was assessed.

A comparison was performed in quality of life and disability between those with and without the primary outcome using the EORTC and WHO DAS 2.0 scores.

The best new predictive models from model derivation (section 9 and section 10) were used to determine if they may also predict a deterioration in QoL, using the *EQ-5DL index score* and *EORTC Sumscore*. These models were created and predictive strength assessed using the same methodology as the models derived in section 9 and section 10.

Further analysis was performed to explore if any new model(s) could also predict a deterioration in breathlessness (an increase in MRC score of 1). This is in contrast to the primary outcome where MRC dyspnoea score was dichotomised into patients who scored >2 at three months post-operatively. The MCID of a deterioration in MRC score is an increase of 1. These models were created and predictive strength assessed using the same methodology as the models derived in section 9 and section 10.

7.20 Missing data

The derivation dataset was from patients recruited from GJNH (the base site) therefore missing data was <1% due to robust follow-up. Any missing data from the external dataset, when validating the models, was handled by overall and subgroup mean imputation from the derivation data, summarised in Table 44. Subgroups were determined by ppoFEV_1% and ppoDLCO% categories. For example,

whether a patient has a ppoFEV$_1$% of less than, or greater than, 40% would determine what value was imputed.

7.21 Calculation and interpretation of Brier scoring

Proposed in 1950 by Glenn W. Brier, Brier scoring is a function enabling the accuracy of prediction models to be measured. The score is applicable to outcomes which can be classified into mutually exclusive or discrete outcomes which must be binary or categorical in nature – it cannot be used for ordinal variables which can take on three or more variables.[296] If an event comes to pass, it is assigned a value of one. If an event does not occur, it is assigned a value of zero. Brier scores are bound between values of 0 and 1 – because all squared errors lie between 0 and 1 (the maximum error is $1^2 = 1$). Each individual probability is in the range of 0 to 1. Low values are desirable with perfect prediction getting a score of zero. A Brier score of 1 means perfect inaccuracy. A person who assigns a probability of 0.5 to every event would wind up with a Brier score of 0.25. The Brier score measures the mean square difference between the predicted probability assigned to a patient and the actual outcome observed, (Equation 3). The score becomes inadequate for very rare events as it does not discriminate between small changes that are significant for rare events.

$$BS = \frac{1}{N}\sum_{t=1}^{N}(ft - ot)2$$

Equation 3 - Calculation of Brier score. *ft* = probability that was forecast (from 0-1) for the tth event, *ot* = actual outcome (0 or 1) of the tth event, *N* = number of forecasting instances.

The brier score quantifies model performance but is not a relative metric and does not allow you to compare one model with another. To compare models the Brier skill score should be used, (Equation 4). A negative value means that the comparator is a poorer model than the baseline model. With a score of zero both models predict equally. A positive score reflects that the comparator model is superior to the baseline model. Therefore, unlike the Brier score, a higher brier skill score is desirable with a value of 1 being the best possible score.

$$BSS = 1 - \left(\frac{BS}{BS}of\ base\ model\right)$$

Equation 4 - Brier Skill Score. *BSS* = Brier Skill Score, *BS* = Brier score

7.22 Classification and interpretation of net reclassification improvement

Net reclassification improvement (NRI) is an index to quantify the performance of one risk prediction model in reclassifying patients either correctly or incorrectly, when compared to another model. NRI has two components, subjects *with* and *without* the event. Patients who have been correctly reclassified are assigned a +1 and those who were incorrectly reclassified are assigned a -1. Subjects not reclassified are assigned a 0. The assigned values are then summed in each group (those correctly or incorrectly classified) before being divided by the number of patients in that group. The sum of these two values is the NRI. The overall NRI cannot be interpreted as 'the net percentage reclassified' because of the weighting by the event rate: the overall NRI being the sum of 2 fractions with different denominators. Therefore, it should be presented as a unitless statistic.[297] An advantage is that NRI is easily understood by clinicians. A worked example of NRI is given in section 10.7 when comparing derived models.

8 Patient demographics & generic results

The results described here apply to the cohort of patients recruited from the Golden Jubilee National Hospital for derivation of the risk stratification tool(s) detailed in chapter 9.

8.1 Patient demographics and characteristics

From October 2018 to November 2019, 108 patients were recruited by the author and co-investigators (Dr Ben Shelley & Dr Philip McCall, consultant cardiothoracic anaesthetists) (Figure 23).

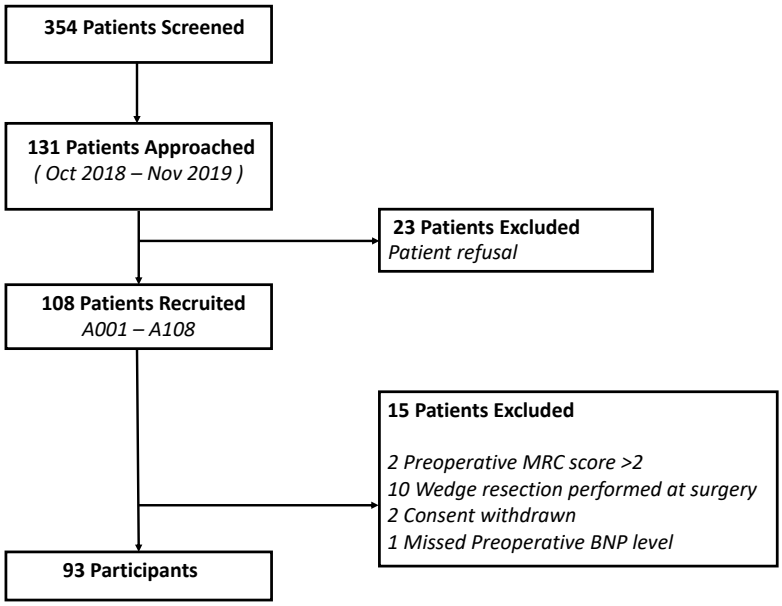

Figure 23 - Study Recruitment CONSORT diagram.
Golden Jubilee National Hospital. October 2018 - November 2019.

Fifteen patients were excluded from further participation in the study; two had a baseline MRC score >2 after being consented into the study and so met exclusion criteria, ten patients had wedge resections performed due to a change in surgical plan intra-operatively, two patients withdrew consent and no longer wanted to be part of the study and a single patient did not get pre-operative BNP measured (Figure 23).

Table 16 summarises patient demographics for those patients recruited into the study. As the main centre of recruitment for the study, missing data was <1% and

therefore was included in analysis, without imputation. The majority of patients had a known diagnosis of malignancy prior to surgery (71%) and were current or ex-smokers (82%). In keeping with a lung cancer population, most patients had a co-morbidity (74%). At pre-operative MDT, most patients were graded performance status one or two (86%).

Patient Demographics	n	Descriptive Statistics
Age (years)	93	66(11)
Male Sex	93	41(44%)
Known malignancy	93	66(71%)
Known metastasis	93	2(2%)
Perceived malignancy	93	27(29%)
Previous chemotherapy/radiotherapy	93	3(3%)
Alcohol consumption per week (units)	93	2(0,10)
Height (cm)	89	165(10)
Weight (kg)	89	76.2(17.6)
BMI (kg/m^2)	89	27.8(6.1)
Smoking Status		
Current or ex-smoker	93	76(82%)
Average pack years	93	30(7,42)
Average time since abstinence (years)	93	0 (0,12.0)
Lung Function		
FEV_1 (L)	91	2.07(1.78,2.80)
FEV_1 (% predicted)	92	86(18)
FVC (L)	88	3.20(0.94)
FVC (% predicted)	87	105(18)
FEV_1/FVC ratio (%)	90	71(11)
DLCO (ml/min/mmHg)	78	6.07(1.65)
DLCO (% predicted)	89	75(17)
Predicted post-operative FEV_1 (%)	89	70(17)
Predicted post-operative DLCO (%)	89	59(15)
Blood results		
Hb (g/L)	93	13.3 (1.5)
Albumin (g/L)	87	43(39,46)
Creatinine (micromole/L)	93	71(60,84)
eGFR <60 (ml/min)	93	12(13%)
CRP (mg/L)	92	4 (2,8)

Table 16 - Pre-operative patient demographics

Co-morbidities		Descriptive Statistics
Previous co-morbidity	93	69(74%)
Previous cancer	93	13(14%)
COPD	93	28(30%)
Hypertension	93	26(28%)
Heart disease	93	10(11%)
Diabetes	93	15(16%)
Peripheral Vascular disease	93	2(2%)
Obesity (BMI >30)	93	19(20%)
Alcoholism	93	4(4%)
GORD	93	18(19%)
Performance status (Clinician scored)		
0	93	30(33%)
1	93	50(53%)
2	93	11(12%)
3	93	2(2%)
4	93	0(0%)
Thoracoscore (%)	93	1.50 (0.90,2.30)
ASA		
I	93	10(11%)
II	93	53(57%)
III	93	29(31%)
IV	93	1(1%)

Table 16 continued - Pre-operative patient demographics.
FEV_1 = Forced expiratory volume in one second, FVC = Forced vital capacity, DLCO = Carbon monoxide diffusing capacity, BMI = Body mass index, Hb = Haemoglobin, eGFR = Estimated glomerular filtration rate, CRP = C-reactive Protein, COPD = Chronic obstructive pulmonary disease, GORD = Gastro oesophageal reflux disease, ASA = American Society of Anaesthesiology score. Values are number (%), mean (SD) or median (IQR). n represents number of patients with data available for each variable.

8.2 Intra-operative data

All 93 patients recruited into the study underwent surgery; 84 patients (90%) underwent lobectomy, three patients (3%) bi-lobectomy and six patients (7%) pneumonectomy. The majority of patients underwent MITS (n=60, (64%)). The mean (SD) ppoFEV_1% and mean ppoDLCO% were 70 (17)% and 59 (15)%, respectively (Tabl). Further operative data is summarised in and Table 17 Table 18.

Surgery Performed	n	Descriptive Statistics
Lobectomy	93	84(90%)
Bi-Lobectomy	93	3(3%)
Pneumonectomy	93	6(7%)
Operation type		
Open	93	33(36%)
Video assisted	93	52(56%)
Robotic assisted	93	8(8%)
Operation side		
Left	93	43(46%)
Right	93	50(54%)

Table 17 - Intra-operative details
Values are number (%), n represents number of patients with data available for each variable.

Total intravenous anaesthetic technique was used in the majority of patients (n=54, (58%)). Most patients received either a paravertebral catheter or paravertebral injection (n=70, (88%)). Intraoperative surgical time was not recorded by the anaesthetist or surgeon for 12 patients. Duration of anaesthetic was not recorded in three patients and one lung ventilation (OLV) time was not recorded for 17 patients.

Intra-operative Details	n	Descriptive Statistics
Technique		
Total intravenous	92	54(58%)
Volatile	92	38(42%)
Analgesia		
Epidural catheter	92	11(12%)
Paravertebral catheter	92	49(53%)
Paravertebral Injection	92	32(35%)
Morphine PCA	92	58(62%)
Fentanyl PCA	92	7(8%)
Oral opiates	92	24(26%)
Intercostal LA Infiltration	92	6(7%)
Surgery details		
Duration of surgery (mins)	81	149(113,190)
Duration of anaesthesia (mins)	90	171(144,227)
One lung ventilation time (mins)	76	126(94,154)

Table 18 - Intra-operative details
PCA = Patient controlled analgesia, LA = Local anaesthetic. Values are number (%) or median (IQR). n represents number of patients with data available for each variable. Mins = minutes.

8.3 Peri-operative dyspnoea

Seventy-five patients returned questionnaires at 3 months for analysis. Patients reported a clinically and statistically increase in dyspnoea following lung resection surgery, regardless of which scoring tool was used. Twenty-seven patients (36%) reported an MRC>2 at the 3-month time point (primary outcome). Median pre-operative University of California and San Diego Shortness of Breath Questionnaire (UCSD-SOBQ) score was ten. These are summarised in Table 19. The MCID score of the MRC score is 1; 46 (61%) patients reported a MCID change in MRC score. The MCID score of the UCSD-SOBQ is 5; 61 (81%) patients reported a MCID change in UCSD-SOBQ score.

MRC score pre-operative (patient scored)	Pre-operative (n=93)	Post-operative (n=75)	Significance (p-value)
1	37(40%)	9(12%)	**<0.01**[+]
2	56(60%)	39(52%)	
3	0(0%)	19(25%)	
4	0(0%)	2(3%)	
5	0(0%)	6(8%)	
	Pre-operative (n=93)	**Post-operative (n=74)**	
UCSD-SOBQ (0-120)	10(2,28)	24(11,56)	**<0.01**[#]

Table 19 - Peri-operative MRC and UCSD SOBQ scoring
Values are number (%) or median (IQR). n represents number of patients with data available for each variable. MRC = Medical Research Council dyspnoea scale, UCSD-SOBQ = University of California, San Diego shortness of breath questionnaire. # = Mann-Whitney U test, [+] = Pearson chi squared test. Significant results highlighted in bold.

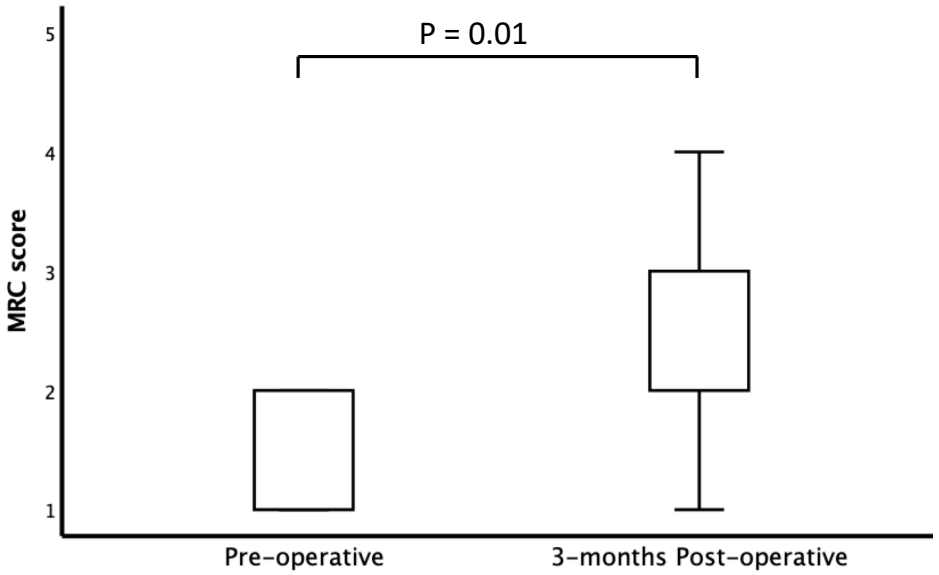

Figure 24 - Peri-operative MRC score.
MRC = Medical Research Council dyspnoea score. p=0.01, Pearson chi squared test. (n=75).

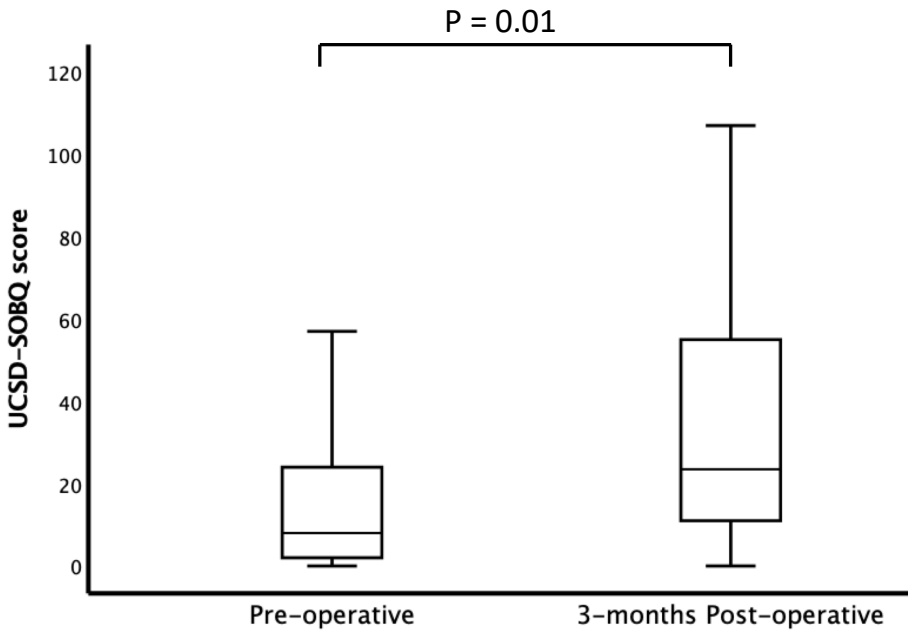

Figure 25 - Peri-operative UCSD-SOBQ score.
Mann-Whitney U test. UCSD-SOBQ = University of California, San Diego shortness of breath questionnaire **(**n=74). p=0.01

8.4 Quality of life and disability scoring

Most patients had a pre-operative World Health Organisation (WHO) performance status category of zero or one (n=83, 89%). There was an increase in performance status score post-operatively, reflecting a decrease in functional capacity. Median (IQR) *pre-operative* World Health Organisation disability assessment schedule 2.0 (WHO DAS 2.0) score was 13 (4, 33) % (Table 20). This increased post-operatively to 19 (10, 33)% and represents an increase in disability. Nineteen (25%) patients had a *post-operative* WHO DAS 2.0 score >16%, which is in keeping with mild to moderate disability.

Demographic	Pre-operative (n=93)	Post-operative (n=75)	Significance (p-value)
WHO PS score (patient scored)			**<0.01⁺**
0	36 (39%)	10 (13%)	
1	47 (50%)	42 (56%)	
2	8 (9%)	20 (27%)	
3	2 (2%)	3 (4%)	
4	0 (0%)	0 (0%)	
WHODAS 2.0 score (%)	13 (4,33)	19 (10,33)	**<0.01#**

Table 20 - Pre-operative and post-operative World Health Organisation performance status score and pre-operative and post-operative WHO DAS 2.0 score.
WHO PS = World Health Organisation performance status, WHODAS = World Health Organisation disability assessment schedule. Values are number (%) or median (IQR). n represents number of patients with data available for each variable. # = Mann-Whitney U test, ⁺ = Pearson chi squared test. Significant results highlighted in bold.

Table 21 summarises the EQ-5DL quality of life scoring tool domains; four out of five domains had deteriorated at the 3-month time point. *Both* EQ-5DL visual analogue scale and EQ-5DL index scoring had deteriorated by 3 months post-operatively. This represents a reduction in quality of life in this cohort. The only domain unchanged being anxiety and depression.

EQ-5DL (visual analogue scale) decreased post-operatively, ($p<0.005$) (Figure 26). EQ-5DL index score also decreased post-operatively ($p<0.01$, 95%CI 0.05 – 0.14). This represents an overall decline in post-operative quality of life. Thirty-three patients had a change in VAS score greater than the MCID of 7. Thirty-nine patients had a change in EQ-5DL index score greater than the MCID of 0.08. Both of these represent an overall decline in quality-of-life following surgery.

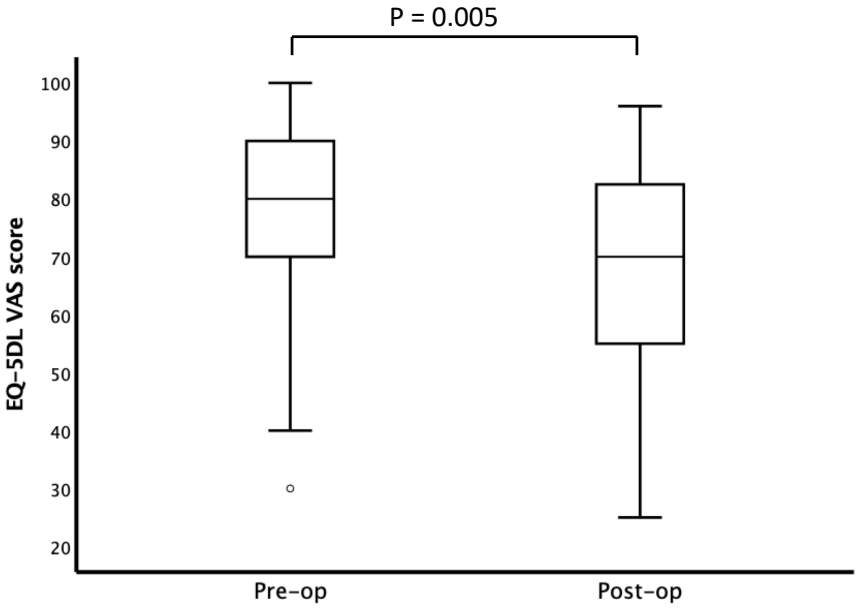

Figure 26 - Pre-operative Vs Post-operative EQ-5DL Visual Analogue Scale Score.
Pre-op = Pre-operative, Post-op = Post-operative, EQ-5DL = EuroQol 5DL, VAS = Visual Analogue Scale (n=75).

EQ-5DL Domain	Pre-op (n=93)	Post-op (n=75)	Significance (p-value)
EQ-5DL Mobility			**<0.01⁺**
1	46(49%)	28(37%)	
2	25(27%)	17(23%)	
3	20(22%)	25(33%)	
4	2(2%)	5(7%)	
5	0(0%)	0(0%)	
EQ-5DL self-care			**<0.01⁺**
1	79(84%)	49(66%)	
2	7(8%)	12(16%)	
3	7(8%)	13(17%)	
4	0(0%)	1(1%)	
5	0(0%)	0(0%)	
EQ-5DL usual activities			**<0.01⁺**
1	46(50%)	21(28%)	
2	31(33%)	25(33%)	
3	14(15%)	19(25%)	
4	2(2%)	8(11%)	
5	0(0%)	2(3%)	
EQ-5DL pain/discomfort			**<0.01⁺**
1	42(45%)	20(27%)	
2	23(25%)	31(41%)	
3	22(24%)	20(27%)	
4	5(5%)	4(5%)	
5	1(1%)	0(0%)	
EQ-5DL anxiety depression			0.9⁺
1	56(60%)	43(57%)	
2	20(22%)	21(28%)	
3	10(11%)	7(9%)	
4	6(6%)	4(5%)	
5	1(1%)	0(0%)	
EQ-5DL Index Value	0.77(0.65,0.88)	0.74(0.56,0.84)	**<0.01#**
EQ-5DL VAS	75(65,90)	70(55,85)	**<0.01#**

Table 21 - Peri-operative EQ-5DL score
Pre-op = Pre-operative, Post-op = Post-operative, EQ-5DL = Euro Qol 5 level 5-dimensional scoring tool, VAS = visual analogue scale. Values are number (%) or median (IQR). n represents number of patients available with data for each variable. # = Mann-Whitney U test, ⁺ = Pearson chi squared test. Significant results highlighted in bold.

All patients had a pre-operative European Quality of Life Questionnaire C30 completed with no missing data (Table 22). Two patients did not have sufficient

pre-operative or *post-operative* data to complete the EORTC LC-13 score and these were not included in analysis; this left 73 patients with paired data. Both EORTC domains *decreased* post-operatively reflecting an overall reduction in quality of life within this population. EORTC *sumscore* decreased *post-operatively*, p=0.005, Mann Whitney U test. (Figure 27). Thirty-one (41%) patients had a MCID decline in *sumscore* of 10 points post-operatively, representing a clinically significant decline in QoL.

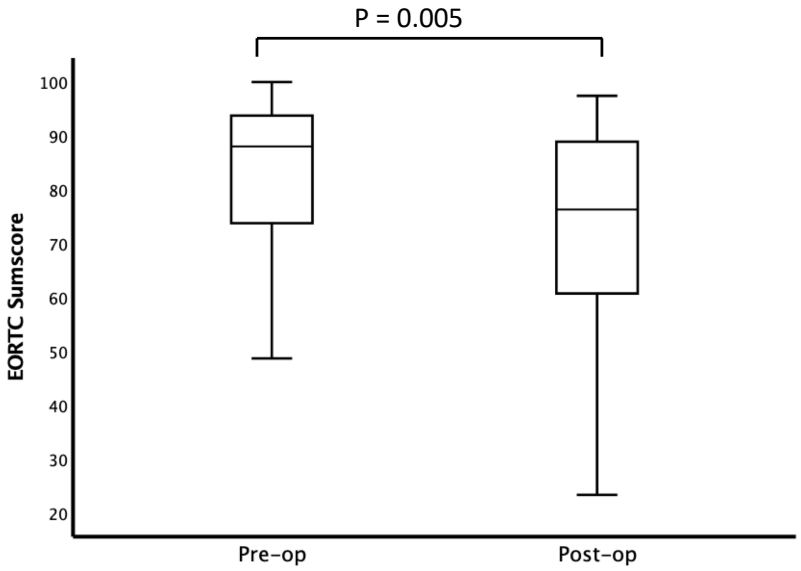

Figure 27 - Pre-operative and Post-operative EORTC *Sumscore*.
EORTC-QLQ = European organisation for the research and treatment of cancer quality of life questionnaire, Pre-op = Pre-operative, Post-op = Post-operative 3-month time point. p <0.01, Mann-Whitney U test.(n=75).

There was a decline in mean (SD) LC-13 *sumscore* from 88 (8) pre-operatively to 82 (12) post-operatively, (p<0.01). This represents a global *decline* in QoL. As described in section 2.2.2.2, the recommended MCID of the LC-13 *sumscore* is 2; Forty-nine (68%) patients had a decline in this domain again reflecting an overall decline in quality of life within this population.

EOTRC Domain	Pre-operative (n=93)	Post-operative (n=75)	Significance (p-value)
EORTC QLQ-30 *Sumscore* (%)	86(69,94)	76(60,89)	**<0.01#**
	Pre-operative (n=92)	**Post-operative (n=73)**	
EORTC LC-13 *Sumscore* (%)	88(83,94)	83(75,91)	**<0.01#**

Table 22 - Peri-operative EORTC QoL scores
Values are median (IQR). n represents number of patients with data available for each variable. LC-13 = EORTC Lung cancer module, EORTC-QLQ-30 = European organisation for the research and treatment of cancer quality of life questionnaire. n represents number of patients with data available for each variable. # = Mann-Whitney U test. Significant results highlighted in bold.

8.5 Peri-operative pain, anxiety and depression scores

Patients were increasingly depressed 3-months following surgery with an increase in HADS D score (Table 23). This was in contrast to no change in the EQ-5DL depression score above (Table 21). Similarly, patients reported increasing levels of pain post-operatively; BPI score, VAS average and worst pain scores all increased significantly.

Domain	Pre-operative (n=93)	Post-operative (n=73)	Significance (p-value)
HADS A score	5(2,9)	6(2,9)	0.4#
HADS D score	3(1,6)	5(1,8)	**<0.01#**
	Pre-operative (n=93)	**Post-operative (n=74)**	
BPI score (0-70)	5(0,21)	14(3,30)	**<0.01#**
VAS Average (0-10)	0(0,0)	1(0,3)	**<0.01#**
VAS Worst (0-10)	0(0,0)	1(0,3)	**<0.01#**

Table 23 - Peri-operative HADS, BPI and VAS scores
Values are median (IQR). n represents number of patients with data available for each variable. HADS A = Hospital Anxiety and Depression (Anxiety score), HADS D = Hospital Anxiety and Depression (Depression score), BPI = Brief pain inventory score, VAS = Pain visual analogue score# = Mann-Whitney U test. Significant results highlighted in bold.

8.6 Lost to follow Up

Eighteen patients did not return 3-month questionnaires and were excluded from the primary outcome analysis. To examine for potential sources of bias, the demographics of those who didn't return forms were compared to those who did (Table 24). The two groups appeared broadly similar and notably, there were no significant difference in pre-operative lung function, performance status or Thoracoscore between those who returned the 3-month questionnaires and those who did not.

Patient Demographics	3-month Lost to Follow up (n=18)	Overall (n= 93)	Significance (p-value)
Age (years)	65(8)	66(11)	p=0.86*
Gender (male)	11(61%)	41(44%)	p=0.68^{+}
Known malignancy	15(83%)	66(71%)	p=0.28^{+}
Known metastasis	0(0%)	2(2%)	p=0.53^{+}
Perceived malignancy	3(17%)	27(29%)	p=0.32^{+}
Previous chemotherapy/radiotherapy	0(0%)	3(3.2%)	p=0.54^{+}
Alcohol consumption per week (units)	0(0,10)	2(0,10)	p=0.80$^{\#}$
Smoking Status			
Current or ex-smoker	14(78%)	76(82%)	p=0.70^{+}
Average pack years	38(15,50)	30 (7,42)	p=0.43$^{\#}$
Average time since abstinence (years)	0(0,6)	0 (0,12.0)	p=0.24$^{\#}$
Lung Function			
FEV_1 (L)	1.92(1.66,2.83)	2.07(1.78,2.80)	p=0.79$^{\#}$
FEV_1 (% predicted)	90(16.51)	86 (18)	p=0.96*
FVC (L)	3.01(0.89)	3.20 (0.94)	p=0.40*
FVC (%)	104.00(16)	105(18)	p=0.56*
FEV_1/FVC ratio (%)	74(13)	71(11)	p=0.38*
DLCO (ml/min/mmHg)	6.58(2.18)	6.07(1.65)	p=0.27*
DLCO (% predicted)	82 (20)	75(17)	p=0.15*
ppoFEV_1 (%)	70(15)	70(17)	0.97*
ppoDLCO (%)	64(17)	59(15)	0.28*
Previous Comorbidity	15(83%)	69(74%)	p=0.41^{+}
Performance status (CS)			p=0.30^{+}
0	3(17%)	30(33%)	
1	14(78%)	50(53%)	
2	1(5%)	11(12%)	
3	0(0%)	2(2%)	
4	0(0%)	0(0%)	
Thoracoscore (%)	1.70(0.42,2.67)	1.50(0.90,2.30)	p=0.70$^{\#}$
ASA			p=0.16^{+}
I	0(%)	10(11%)	
II	8(44%)	53(57%)	
III	10(56%)	29(31%)	
IV	0(%)	1(1%)	

Table 24 - Lost to follow up: demographic comparison
Values are number (%), mean (SD) or median (IQR). n represents number of patients with data available for each variable. ppoFEV_1 = Predicted post-operative forced expiratory volume in one second, FVC = Forced vital capacity, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, CS = clinician scored. * = independent samples t-test, # = Mann-Whitney U test, $^{+}$ = Pearson Chi Squared test. Significant results highlighted in bold.

There was no difference in operation performed (p=0.88), operation type (p=0.43), predicted ppoFEV_1% (p=0.97) and predicted ppoDLCO% (p=0.28).

Surgery	3-month non-return (n=18)	Overall (n= 93)	Significance (p-value)
Operation performed			0.88[+]
Lobectomy Bi-Lobectomy Pneumonectomy	16(88%) 1(6%) 1(6%)	84(90%) 3(3%) 6(7%)	
Operation type			0.43[+]
Open Video Assisted Robotic Assisted	7(39%) 11(61%) 0(0%)	33(36%) 52(56%) 8(8%)	

Table 25 - Lost to Follow up: Intra-operative comparison
Values are number. n represents number of patients with data available for each variable. [+] = Pearson chi squared test.

Pulmonary complication rate as per European Society Thoracic Surgeons guidelines were lower in the initial group (16%) than in those lost to follow-up (39%), (p=0.03).

Post-Operative Complications	Lost to Follow up (n=18)	Overall (n= 93)	Significance (p-value)
Any ESTS complication	11(61%)	42(45%)	0.22[+]
Pulmonary complication	7(39%)	15(16%)	**0.03**[+]
Cardiac complications	8(44%)	31(33%)	0.37[+]
Cardiopulmonary complications	10(56%)	44(47%)	0.52[+]
Greater than one POC	5(28%)	15(16%)	0.24[+]

Table 26 – Lost to follow up: post-operative ESTS complication comparison
Values are number (%). n represents number of patients with data available for each variable. [+] = Pearson Chi Squared test. ESTS = European Society of Thoracic Surgeons, POC = Post-operative complication. Significant results highlighted in bold.

The remaining 75 patient dataset reflects the population that will be part of the following investigations. They are representative of the lung cancer population with typical demographics, similar operation type, complication rate and length of hospital stay. The number of patients lost to follow up was low and better than other studies within the lung cancer population.

8.7 B-Type natriuretic peptide analysis

BNP concentration was significantly different at each peri-operative time points (p<0.01, Friedman's test, Figure 28). BNP concentration was lower pre-operatively compared with post-operatively days two and three (p<0.01, Wilcoxon Rank Sum). There was no difference between day 2 and day 3 BNP concentration (p>0.99, Wilcoxon Rank Sum, Figure 28).

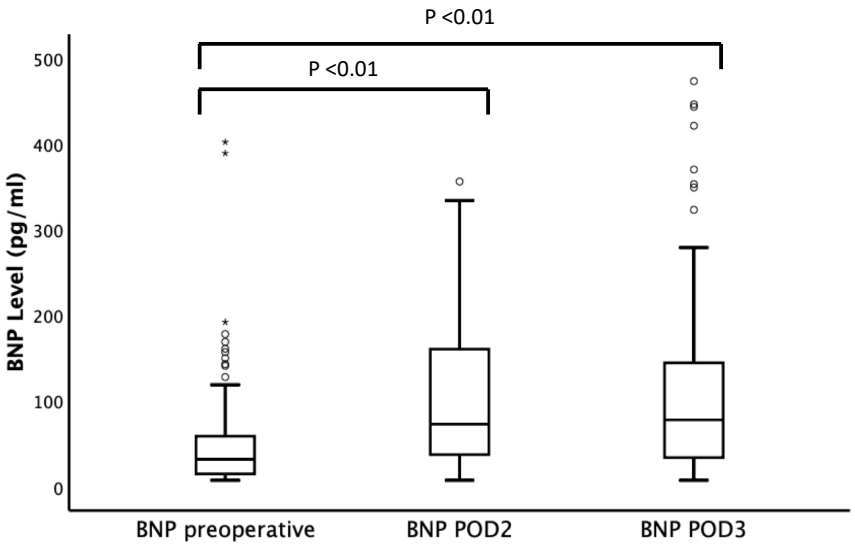

Figure 28 – B-Type natriuretic peptide changes over time.
Changes over time assessed with friedman's test (p=<0.01). Post-hoc comparisons with wilcoxon rank-sum test with p-values as illustrated. Y-axis censored at 500pg/ml, single outlier during study (Pre-operative 604pg/ml). BNP = B-Type natriuretic peptide; POD = Post-operative day. (n=93)

9 Primary outcome analysis: model derivation

This section explores the value of traditional risk prediction models derived in this population for prediction of MRC>2 at 3 months post-operatively (the primary outcome). Mean (SD) time to completion of 3-month survey follow-up was 95 days (30). Seventy-five patients had data available for analysis at the 3-month time point. Of these, 27 (36%) patients had an MRC >2 at 3 months. Missing data for this derivation dataset was <1%.

9.1 Univariate analysis: demographics.

Univariate analyses of those patient's with and without an MRC scoring >2 at 3 months is summarised in Table and Table . Those variables with significance in the range up to and including p=0.1 were subsequently considered at multivariate logistic regression analysis to derive a scoring tool to predict dyspnoea at 3 months following lung resection surgery. There was no difference in pre-operative BNP levels in the primary outcome groups, (p=0.79).

Patient Demographics	MRC≤2 (n=48)	MRC>2(n=27)	Difference (p-value)
Pre-operative BNP (pg/ml)	31(17,53)	28(7.5,65)	0.79#
Pre-operative MRC			**0.03+**
Pre-operative MRC 1 Pre-operative MRC 2	24(77%) 24(55%)	7(23%) 20(45%)	
Demographics			
Age (years)	65(13)	66(10)	0.78*
Gender (Male)	21(44%)	13(48%)	0.71+
Known malignancy	33(69%)	18(67%)	0.85+
Known metastasis	2(4%)	0(0%)	0.28+
Perceived malignancy	12(25%)	11(41%)	0.16+
Previous chemotherapy/radiotherapy	5(15%)	1(4%)	0.41+
Alcohol consumption per week (units)	2(0,12)	0(0,10)	0.67#
Height (cm)	165(10)	166(10)	0.98*
Weight (kg)	73.6(16.9)	81.6(15.0)	**0.05***
BMI (kg/m^2)	26.7(5.9)	29.8(5.7)	**0.03***

Table 27 - Univariate analysis: demographics

Smoking Status	MRC≤2 (n=48)	MRC>2(n=27)	Difference (p-value)
Current or ex-smoker	39(81%)	23(85%)	0.67+
Average pack years	28(7,40)	27(15,40)	0.55#
Average time since abstinence (years)	0.4(0,15)	0.5(0,18)	0.71#
Lung Function			
FEV_1 (L)	2.17(1.90,2.96)	1.83(1.69,2.50)	**0.06#**
FEV_1 (% predicted)	92(19)	84(18)	**0.07***
FVC (L)	3.44(0.92)	2.99(0.94)	**0.05***
FVC (%)	109(19)	100(17)	**0.04***
FEV_1/FVC ratio (%)	71(11)	70(11)	0.77*
DLCO (ml/min/mmHg)	5.82(1.43)	6.03(1.58)	0.59*
DLCO (% predicted)	74(15)	73(18)	0.93*
ppoFEV_1 (%)	74(16)	65(18)	**0.04***
ppoDLCO (%)	58(14)	56(15)	0.57*
Blood Results			
Hb (g/L)	133(14)	137(15)	0.20*
Albumin (g/L)	43(40,46)	43(39,45)	0.96#
Creatinine (micromole/L)	73(60,86)	69(58,83)	0.46#
eGFR (mL/min)	60(60,60)	60(60,60)	0.57#
CRP (mg/L)	3(2,8)	5(3,9)	0.43#
Co-morbidities			
Previous Co-morbidity	34(71%)	20(74%)	0.76+
Previous cancer	7(15%)	3(11%)	0.67+
COPD	13(27%)	9(33%)	0.57+
Hypertension	15(31%)	7(26%)	0.63+
Heart disease	4(8%)	3(11%)	0.69+
Diabetes	4(8%)	6(22%)	**0.09+**
Peripheral vascular disease	1(2%)	1(4%)	0.68+
Obesity (BMI >30)	8(17%)	7(26%)	0.33+
Alcoholism	2(4%)	0(0%)	0.28+
GORD	11(23%)	6(22%)	0.95+
Performance Status			0.18+
0	20(42%)	7(26%)	
1	22(46%)	14(52%)	
2	4(8%)	6(22%)	
3	2(4%)	0(0%)	
4	0(0%)	0(0%)	

Table 27 continued - Univariate analysis: demographics

Thoracoscore (%)	1.28(0.80,2.16)	1.57(1.02,2.30)	0.25#
ASA			**0.02+**
I II III IV	10(21%) 23(48%) 14(29%) 1(2%)	0(0%) 22(81%) 5(19%) 0(0%)	

Table 27 continued - Univariate analysis: demographics
Values are number (%), mean (SD) or median (IQR). n represents number of patients with data available for each variable. * = independent samples t-test, # = Mann-Whitney U test, + = Pearson Chi squared test. FEV_1 = Forced expiratory volume in one second, FVC = Forced vital capacity, DLCO = Carbon monoxide diffusing capacity, BMI = Body mass index, Hb = Haemoglobin, eGFR = estimated glomerular filtration rate, CRP = C-reactive protein, COPD = chronic obstructive pulmonary disease, GORD = gastro oesophageal reflux disease, ASA = American Society of Anaesthesiology Score, BNP = B-Type natriuretic peptide, ppo = predicted post-operative. Significant results highlighted in bold (significance being defined as those variables up to *p*=0.1).

Demographic	MRC≤2 (n=48)	MRC>2 (n=27)	Difference (p-value)
WHO score pre-operative (patient scored)			**<0.01+**
0 1 2 3 4	28(59%) 17(35%) 2(4%) 1(2%) 0(0%)	5(19%) 19(70%) 3(11%) 0(0%) 0(0%)	
Pre-op EQ-5DL mobility			**<0.01+**
1 2 3 4 5	32(67%) 10(21%) 6(12%) 0(0%) 0(0%)	8(30%) 7(26%) 10(37%) 2(7%) 0	
Pre-op EQ-5DL self-care			**<0.01+**
1 2 3 4 5	48(100%) 0(0%) 0(0%) 0(0%) 0(0%)	17(64%) 5(18%) 5(18%) 0(0%) 0(0%)	

Table 28 - Univariate analysis: dyspnoea and quality of life scores

Demographic	MRC≤2 (n=48)	MRC>2 (n=27)	Difference (p-value)
Pre-op EQ-5DL usual activities			**<0.01**$^{+}$
1	30(63%)	9(33%)	
2	15(31%)	10(37%)	
3	3(6%)	6(22%)	
4	0(0%)	2(8%)	
5	0(0%)	0(0%)	
Pre-op EQ-5DL pain/discomfort			0.13^{+}
1	28(58%)	10(37%)	
2	11(23%)	6(22%)	
3	9(19%)	10(37%)	
4	0(0%)	1(4%)	
5	0(0%)	0(0%)	
Pre-op EQ-5DL anxiety depression			**0.02**$^{+}$
1	32(67%)	12(44%)	
2	11(23%)	8(30%)	
3	1(2%)	6(22%)	
4	4(8%)	1(4%)	
5	0(0%)	0(0%)	
EQ-5DL index score	0.834(0.149)	0.693(0.191)	**<0.01***
Pre-op EQ-5DL VAS	80(75,90)	75(50,85)	**<0.01**$^{\#}$
Pre-op WHO DAS score (%)	6(2,19)	21(10,38)	**<0.01**$^{\#}$
Pre-op HADS A score	4(2,8)	7(2,10)	0.36$^{\#}$
Pre-op HADS D score	2(0.25,5)	4(2,8)	**<0.01**$^{\#}$
Pre-op BPI score (0-70)	2(0,7)	19(3,29)	**<0.01**$^{\#}$
Pre-op VAS average (0-10)	0(0,0)	0(0,1)	**<0.01**$^{\#}$
Pre-op VAS worst (0-10)	0(0,0)	0(0,1)	0.07$^{\#}$
Pre-op EORTC QOL Sumscore	90(80,96)	80(68,89)	**<0.01**$^{\#}$

Table 28 continued - Univariate analysis: dyspnoea and quality of life scores
Values are number (%) or median (IQR). n represents number of patients with data available for each variable. * = independent samples t-test, # = Mann-Whitney U test, $^{+}$ = Pearson chi squared test. WHO PS = World health organisation performance status, WHO DAS = World health organisation disability assessment schedule 2.0, FEV1 = Forced expiratory volume in one second, DLCO = Carbon monoxide diffusing capacity, HADS A = Hospital anxiety and depression (Anxiety) score, HADS D = Hospital anxiety and depression (Depression) score, BPI = Brief pain inventory score, VAS = Pain visual analogue score, Pre-op = Pre-operative, Post-op = Post-operative, EQ-5DL = Euro Qol 5 level 5 dimensional scoring tool, VAS = visual analogue scale, EORTC Lung cancer module, EORTC-QLQ = European organisation for the research and treatment of cancer quality of life questionnaire. Significant results highlighted in bold.

Patients with increased dyspnoea (MRC >2) at 3 months had reduced quality of life scoring post-operatively, with a greater deterioration in EQ-5DL VAS score, (p=0.01) (Figure 29). Of those patients who had a change in EQ-5DL VAS score

greater than the MCID of seven, 17 (52%) were breathless and had MRC >2 at 3-months, ($p<0.01$). Furthermore, patients with an MRC >2 at three months also had higher magnitude of change of EQ-5DL index score, ($p<0.01$) (Figure 30).

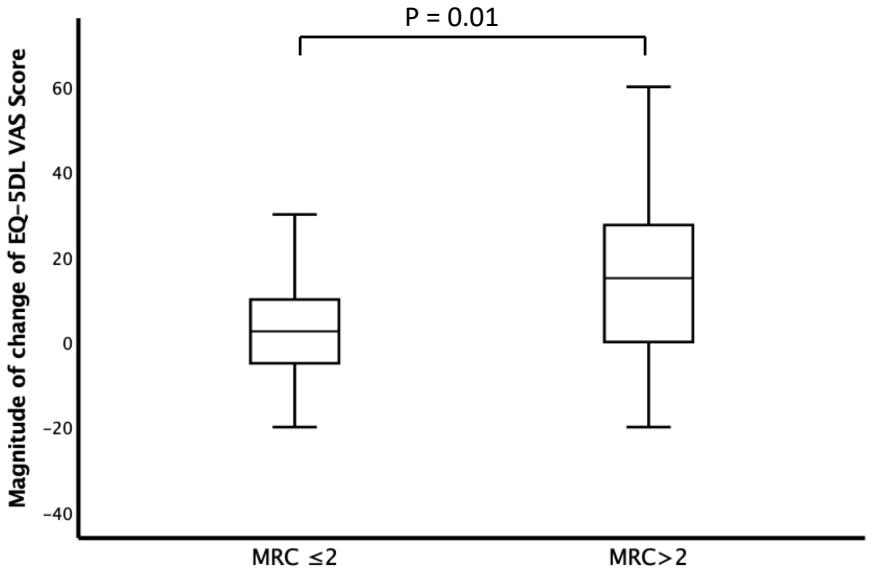

Figure 29 - Magnitude of EQ-5DL VAS change Vs primary outcome.
VAS = Visual analogue scale, MRC = Medical Research Council score at 3 months post-operatively, EQ-5DL = EuroQol 5DL (n=75)

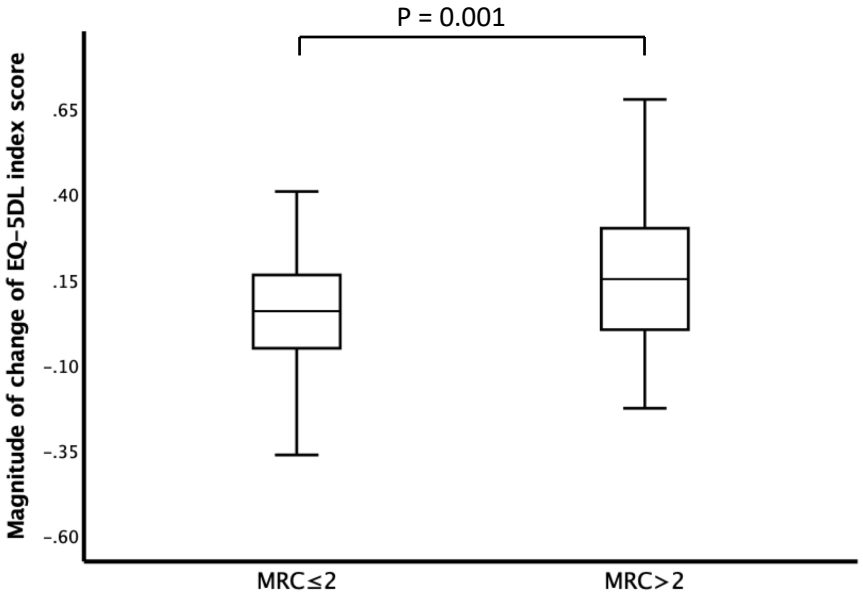

Figure 30 - Magnitude of EQ-5DL index change Vs primary outcome.
EQ-5DL = EuroQol 5DL, MRC = Medical Research Council score at 3 months post-operatively. (n=75).

As described in section 7.18, pre-operative BNP pg/ml (p=0.79), age (p=0.78), gender (p=0.71), ppoDLCO% (p=0.57) and ppoFEV_1% (p=0.04) were added into the models regardless of significance at univariate analysis. Other variables significant at univariate analysis to be taken forward into model creation included; pre-operative MRC (p=0.03), FEV_1(L) (p=0.06), FEV_1% (p=0.07), FVC Litres (p=0.05), FVC% (p=0.04), BMI (p=0.03), ASA (p=0.02), pre-operative WHO score (p=0.006), pre-operative EQ-5DL index score (p=0.001), pre-operative EQ-5DL VAS score (p=0.007), pre-operative WHO DAS 2.0 score (p=0.001), pre-operative HADS D score (p=0.029), pre-operative BPI score (p=0.001), pre-operative pain VAS average score (p=0.018) and diabetes (p=0.09).

9.2 Model derivation and binary logistic regression analysis – *conventional practice*

The four models in this section are intended to reflect current risk prediction methods. Models 1-4 were created with logistic regression. Linearity of the continuous variables with respect to the logit of the dependent variable was assessed via the Box-Tidwell procedure.[298] A Bonferroni correction was applied using all terms in the model.[299]

9.3 Model 1: ppoFEV_1% (<40%) and ppoDLCO% (<40%)

Model 1 incorporated ppoFEV_1 and ppoDLCO% alone, with both variables dichotomised at the conventional 'high risk' cut-offs of <40% (Table 29). Overall significance of the model was p=0.53. The model explained 1.7-2.4% of the variance in dyspnoea following lung resection and correctly classified 64% of cases. Sensitivity was 7.4% and specificity was 97.8%. (PPV = 67% & NPV =64%). Of the 2 predictor variables, neither was significant. The AUROCC was 0.52 (95% CI, 0.38 to 0.66), demonstrating no discrimination (Hosmer et al).[300]

Variable	Beta Co-efficient	Significance (p-value)	OR	95% CI for OR
ppoFEV_1% <40%	-1.07	0.42	0.34	0.03 – 4.68
ppoDLCO% <40%	-0.34	0.68	0.71	0.14 – 3.55
Constant	0.81	0.52	2.25	-

Table 29 - Logistic regression using ppoFEV_1%<40% and ppoDLCO%<40% to predict dyspnoea at three months post-operatively (Model 1 derivation)
FEV_1 = Forced expiratory volume in one second, DLCO = Carbon monoxide diffusing capacity, OR = Odds ratio, ppo = predicted post-operative.

9.4 Model 2: Age, gender, ppoFEV$_1$% (<40%) and ppoDLCO% (<40%)

Model 2 incorporates current risk prediction variables ppoFEV$_1$% and ppoDLCO% <40% (as in model 1) alongside age and gender; all variables were forced into the model (Table 30). Overall significance of the model was p=0.81 and the model explained 2-3% of the variance in dyspnoea following lung resection and correctly classified 64% of cases. Sensitivity was 7.4% and specificity was 97.8%. (PPV = 67% & NPV = 64%). Of the 4 predictor variables, none were significant. The model demonstrated no discrimination with AUROCC of 0.55 (95% CI, 0.41 to 0.69).[300]

Variable	Beta Co-efficient	Significance (p-value)	OR	95% CI for OR
Age	0.01	0.64	1.01	0.97 – 1.06
Gender	-0.08	0.88	0.92	0.34 – 2.50
ppoFEV$_1$% <40%	-1.06	0.43	0.35	0.03 – 4.84
ppoDLCO% <40%	-0.33	0.69	0.72	0.14 – 3.64
Constant	0.15	0.94	1.16	-

Table 30 - Logistic regression using ppoFEV$_1$%<40%, ppoDLCO%<40%, age and gender to predict dyspnoea at three months post-operatively (Model 2 derivation)
FEV$_1$ = Forced expiratory volume in one second, DLCO = Carbon monoxide diffusing capacity, OR = Odds ratio, ppo = predicted post-operative

9.5 Model 3: Age, gender, ppoFEV$_1$% (linear) and ppoDLCO% (linear)

Model 3 contains the same variables as model 2 but with ppoFEV$_1$% and ppoDLCO% retained as continuous variables, instead of the conventional binary cut off of <40% (Table 31). Overall significance of the model was p=0.14 and the model explained 9-13% of the variance in dyspnoea following lung resection and correctly classified 70.8% of cases. Sensitivity was 40.7% and specificity was 88.9%. (PPV = 68.8% and NPV =71.4%). Of the 4 predictor variables only one was significant, ppoFEV$_1$% (p=0.02, beta coefficient 0.95, (95% CI 0.91-0.99). The AUROCC was 0.68 (95% CI, 0.56 to 0.81) demonstrating poor discrimination.[300] Mean Brier score for this model was 0.22, (<0.44 and represents good but not perfect prediction of the model, as per section 7.21).

Variable	Beta Co-efficient	Significance (p-value)	OR	95% CI for OR
Age	0.02	0.38	1.02	0.97 – 1.07
Gender	0.48	0.41	1.62	0.52 – 5.03
ppoFEV$_1$%*	-0.05	**0.02**	0.95	0.91 – 0.99
ppoDLCO%*	0.03	0.28	1.03	0.98 – 1.07
Constant	-0.09	0.96	0.96	-

Table 31 - Logistic regression using ppoFEV$_1$% linear, ppoDLCO% linear, age and gender to predict dyspnoea at three months post-operatively (Model 3 derivation).
FEV$_1$ = Forced expiratory volume in one second, DLCO = Carbon monoxide diffusing capacity, ppo = predicted post-operative. Significant results highlighted in bold. *=Linear variable.

9.6 Model 4: Age, gender, ppoFEV$_1$% (linear), ppoDLCO% (linear) and pre-operative BNP

Model 4 contains the same variables as model 3 with the addition of pre-operative BNP – pre-operative BNP was forced into the model (Table 32). This was planned *a priori* to determine if pre-operative BNP increased predictive value to 'current practice' (Model 3).

Pre-operative BNP is treated as a continuous variable within the regression. Overall significance of the model was p=0.18. The model explained 1-1.4% of the variance in dyspnoea following lung resection and correctly classified 70.8% of cases. Sensitivity was 40.7% and specificity was 88.9%. (PPV = 68.8% and NPV =71.4%). Of the five predictor variables only one was significant, ppoFEV$_1$% (p=0.01, beta coefficient 0.94, (95% CI 0.90-0.99). The AUROCC was 0.69 (95% CI, 0.56 to 0.82), showing poor discrimination.[300] The mean Brier score for this model was 0.21 which is <0.44 and represents good but not perfect prediction of the model, as per section 7.21 . *Pre-operative BNP was not significant within the model* and did *not* improve the predictive strength of *model 3*.

The addition of pre-operative BNP to model 3 does *not* improve the model's predictive strength for dyspnoea at 3 months post-operatively (model 4). There is no discernible difference between AUROCC values for both models (p=0.69, pairwise comparisons of AUROCC) (Figure 31). There was no difference between sensitivity, specificity, PPV or NPV in prediction of dyspnoea at three months.

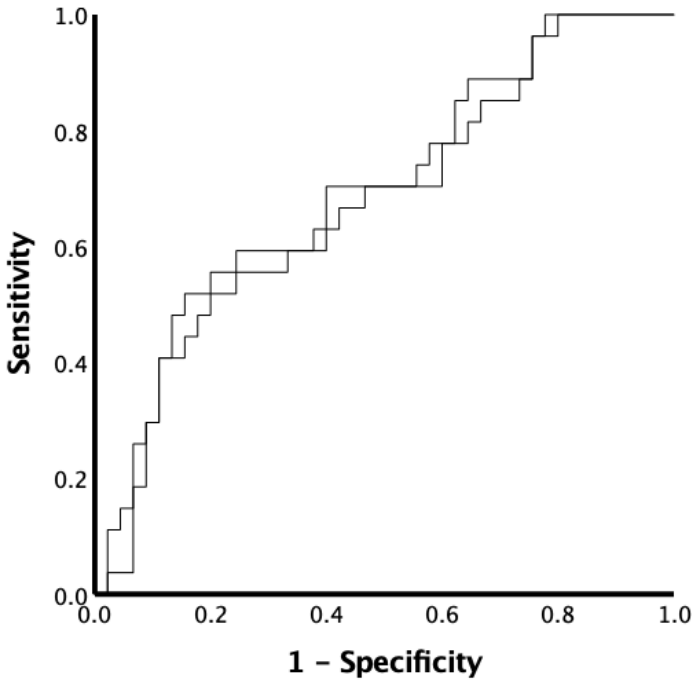

Figure 31 - Receiver operator characteristic curves: Model 3 and Model 4 to predict MRC score >2 at three months post-operatively.
AUROCC Model 3 = 0.68, AUROCC Model 4 = 0.69. Pairwise comparison of AUROCC curves p=0.69 (95% CI - 0.03 – 0.04). (n=75)

Variable	Beta Co-efficient	Significance (p-value)	OR	95% CI for OR
Age	0.03	0.28	1.03	0.98 – 1.09
Gender	0.52	0.38	1.68	0.54 – 2.27
ppoFEV$_1$%*	-0.06	**0.01**	0.94	0.90 – 0.99
ppoDLCO%*	0.03	0.20	1.03	0.98 – 1.09
Pre-op BNP (pg/ml)	-0.01	0.44	1.00	0.99 – 1.01
Constant	-0.34	0.87	0.71	-

Table 32 - Logistic regression using ppoFEV$_1$% linear, ppoDLCO% linear, age, gender and pre-operative BNP to predict dyspnoea at three months post-operatively. (Model 4 derivation)
ppoFEV$_1$ = Predicted post-operative forced expiratory volume in one second, DLCO = Predicted post-operative carbon monoxide diffusing capacity, Pre-op BNP = Pre-operative B-Type natriuretic peptide. Significant results highlighted in bold. *=Linear variable.

9.7 Model Derivation: Summary of results (Models 1- 4)

Summary of the results can be seen below in Table 33 and Table 34 for models 1-4. These models were intended to reflect current practice of prediction of dyspnoea following lung resection surgery and to assess the additive value of pre-operative BNP to the model. No models were significant predictors of post-operative dyspnoea at three months. ppoFEV$_1$% performed best when used as linear variables within these models; this appeared to have the biggest influence

on the predictive power of the models. Pre-operative BNP did *not* increase the predictive strength of the models.

Model	Variables within model	Variable significance (p-value)	Model significance (p-value)	AUROCC value (95% CI)	Mean Brier score
1	ppoFEV$_1$%<40% ppoDLCO%<40%	0.42 0.68	0.53	0.52 (0.38 - 0.66)	-
2	Age Gender ppoFEV$_1$<40% ppoDLCO<40%	0.64 0.88 0.43 0.69	0.81	0.55 (0.41 - 0.69)	-
3	Age Gender ppoFEV$_1$%***** ppoDLCO%*****	0.38 0.41 **0.02** 0.28	0.14	0.68 (0.56 - 0.81)	0.22
4	Age Gender ppoFEV$_1$%***** ppoDLCO%***** pre-op BNP	0.28 0.38 **0.01** 0.20 0.44	0.18	0.69 (0.56 - 0.82)	0.21

Table 33 - Model derivation: Summary of results (Models 1-4)
AUROCC = Area under receiver operator characteristic curve, Pre-op = Pre-operative, CI = Confidence interval, FEV$_1$ = Forced expiratory volume in one second, DLCO Carbon monoxide diffusing capacity, BNP = B-type natriuretic peptide, ppo = Predicted post-operative. Significant results highlighted in bold. * = linear variable.

Model	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Correct (%)
1	7.4	97.8	67.0	64.0	64
2	7.4	97.8	67.0	64.0	64
3	40.7	88.9	68.8	71.4	71
4	40.7	88.9	68.8	71.4	71

Table 34 - Model 1-4 Sensitivity and specificity summary
PPV = Positive predictive value, NPV = Negative predictive value.

10 Improvement of risk prediction model

Chapter 9 concentrated on *conventional* models that reflected current practice. No models within chapter 9 were strong predictors of dyspnoea at three months following lung resection. This chapter explores *new* models, both by adding variables to the 'conventional models' and by an unbiased approach to improve risk prediction. The additive value of pre-operative BNP is again explored.

10.1 Model 5: Age, gender, ppoFEV$_1$% (linear), ppoDLCO% (linear) & next best variable.

Model 3 (incorporating age, gender, ppoFEV$_1$% (linear), ppoDLCO% (linear) was used a baseline for model 5 (Table 35) along with the addition of the '*next best variable'* selected by forwards regression. This next best variable was pre-operative EQ-5DL index score. Overall significance of the model was p<0.01. The model explained 24-33% of the variance in dyspnoea following lung resection and correctly classified 72% of cases. Sensitivity was 48% and specificity was 87%. (PPV = 68% and NPV =74%). Of the predictor variables, 3 were significant ppoFEV$_1$% (p=0.01), ppoDLCO% (p=0.05) and EQ-5DL index score (p<0.01). The AUROCC was 0.81, showing excellent discrimination.[300] Mean brier score for this model was 0.18, confirming model 5 was a strong predictor of dyspnoea in this population.

Variable	Beta co-efficient	Significance (p-value)	OR	95% CI for OR
Age	0.04	0.17	1.04	0.98 – 1.10
Gender	0.90	0.19	2.47	0.64 – 9.56
ppoFEV$_1$%*	-0.07	**0.01**	0.94	0.89 – 0.99
ppoDLCO%*	0.06	**0.05**	1.06	1.00 – 1.12
Pre-op EQ-5DL index score	-6.42	**<0.01**	0.01	0.00 - 0.10
Constant	2.753	0.21	15.69	-

Table 35 - Logistic regression using ppoFEV$_1$% (<40%), ppoDLCO% (<40%), age, gender and pre-operative EQ-5DL index score to predict dyspnoea at three months post-operatively. (Model 5: derivation).

ppoFEV$_1$ = Predicted post-operative forced expiratory volume in one second, ppoDLCO% = Predicted post-operative carbon monoxide diffusing capacity, Pre-op EQ-5DL = Pre-operative Euro Qol 5 level 5-dimensional scoring tool, OR = odds ratio. Significant results highlighted in bold. * = Linear variable

10.2 Model 6: Age, gender, ppoFEV$_1$% (linear) & ppoDLCO% (linear), pre-operative BNP and next best variable

Model 4 (incorporating age, gender, ppoFEV$_1$% (linear), ppoDLCO% (linear) *and pre-operative BNP*) was used as a baseline for this model along with the addition of the next best variable, selected by forwards regression. This was to explore if pre-operative BNP could increase the strength of the model (Table 36). As in model 5, the next variable was pre-operative EQ-5DL index score. Overall significance of the model was $p<0.01$ and the model explained 27-37% of the variance in dyspnoea following lung resection, correctly classifying 77% of cases. Sensitivity was 56% and specificity was 82%, (PPV = 65% and NPV = 76%). Of the predictor variables, three were significant; ppoFEV$_1$% ($p<0.01$ (CI 95% (0.87-0.97), ppoDLCO% ($p=0.02$, CI 1.01-1.15) and EQ-5DL index value ($p<0.01$, CI 0.00-0.062). The AUROCC was 0.82 (95% CI, 0.73 to 0.92), showing excellent discrimination.[300] Mean brier score for model 6 was 0.17, confirming good predictive strength of the model. However, pre-operative BNP was *not* significant within the model and its addition did not discernibly increase the predictive strength when compared to model 5, ($p=0.52$) (Figure 32).

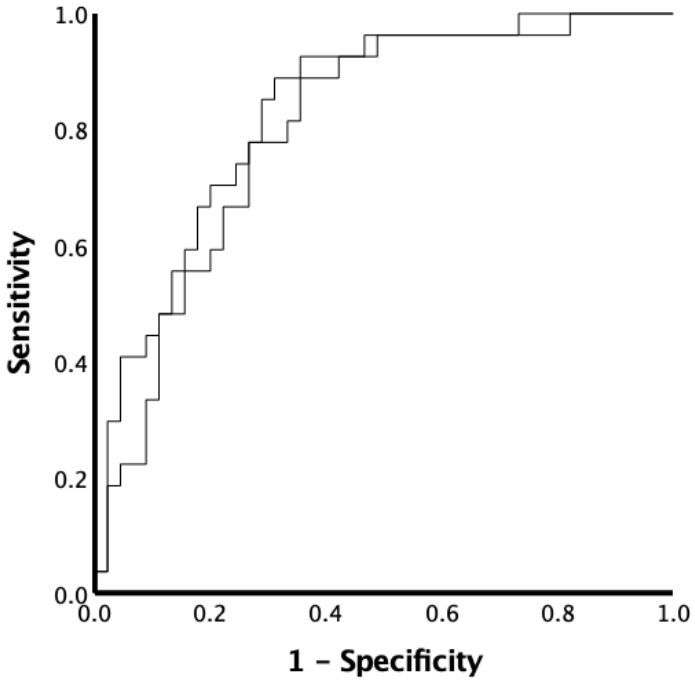

Figure 32 - Receiver operator characteristic curves: Model 5 and Model 6 to predict MRC score >2 at three months post-operatively.
AUROCC Model 5 = 0.81, AUROCC Model 6 = 0.82. Pairwise comparison of AUROCC curves, p=0.52 (95% CI -0.03 – 0.06). (n=75).

Variable	Beta co-efficient	Significance (p-value)	OR	95% CI for OR
Age	0.06	0.07	1.06	1.00 – 1.13
Gender	1.07	0.14	2.91	0.70 – 12.06
ppoFEV$_1$%*	-0.08	**<0.01**	0.92	0.87 – 0.97
ppoDLCO%*	0.07	**0.02**	1.08	1.01 – 1.15
Pre-op BNP	-0.01	0.13	0.99	0.98 – 0.06
Pre-op EQ-5DL (index score)	-7.28	**<0.01**	0.01	0.00 - 0.06
Constant	2.59	0.25	13.45	-

Table 36 - Logistic Regression using ppoFEV$_1$% (<40%), ppoDLCO% (<40%), age, gender and pre-operative EQ-5DL index score to predict dyspnoea at three months post-operatively. (Model 6: derivation).

Pre-op = pre-operative, ppoFEV$_1$% = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, EQ-5DL = Euro Qol 5 level 5-dimensional scoring tool, BNP = B-Type natriuretic peptide, OR = Odds ratio. Significant results highlighted in bold. *=Linear variable.

10.3 Model 7: Forwards regression for all variables

No variables were selected as a baseline for model 7 (Table 37) and using stepwise logistic regression the best variables were selected. The number of variables was not capped and only five were selected to be included in the model; BMI, ppoFEV$_1$%, pre-operative BPI score and diabetes status. Pre-operative BNP *was not selected* by regression to be in this model. *Overall* significance of the model was p<0.01 and it explained 29-39% of the variance in dyspnoea following lung resection, correctly classifying 74% of cases. Sensitivity was 58% and specificity was 85%. (PPV = 71% & NPV =76%). Of the predictor variables, ppoFEV$_1$% (p=0.04), BMI (p=0.06) and pre-operative BPI (p<0.01) were significant. The AUROCC was 0.83, (95% CI 0.74-0.92), showing excellent discrimination.[300] Mean Brier score for model 7 was 0.17, reflecting good predictive strength of the model.

Variable	Beta co-efficient	Significance (p-value)	OR	95% CI for OR
ppoFEV$_1$%*	-0.04	**0.04**	0.96	0.93 – 0.99
BMI	0.10	0.06	1.11	0.99 – 1.22
Pre-op BPI	0.06	**<0.01**	0.92	1.02 – 1.12
Diabetes	1.71	0.08	5.55	0.80 – 38.56
Constant	-1.51	0.41	0.22	-

Table 37 - Logistic Regression using forwards stepwise to select strongest risk prediction model from all variables with significance at univariate analysis to predict dyspnoea at three months post-operatively. (Model 7: Derivation)
ppoFEV$_1$% = Predicted post-operative forced expiratory volume in one second, BMI = body mass index, Pre-op BPI = Pre-operative brief pain inventory score, OR = odds ratio. Significant results highlighted in bold. *= Linear variable.

10.4 Model 8: Pre-operative BNP and forwards regression for other variable selection

Pre-operative BNP was forced into the model at baseline for model 8 (Table 38) and then forwards stepwise regression selected the next best variables. Addition of ppoFEV$_1$%, ppoDLCO%, BMI, pre-operative BPI score, pre-operative visual analogue pain score (average) and diabetes status all increased the predictive power of the model. Overall significance of the model was $p<0.01$ and the model explained 34-47% of the variance in dyspnoea following lung resection, correctly classifying 74% of cases. Sensitivity was 65% and specificity was 80%. (PPV = 68% & NPV =78%). Of the predictor variables, three were significant within the model; ppoFEV$_1$% ($p<0.01$), pre-op BPI score ($p<0.01$) and diabetes status ($p=0.02$). The AUROCC was 0.85 (95% CI 0.76-0.93), suggesting excellent discrimination.[300] Mean brier score for this model was 0.16 reflecting the strength of the model to predict dyspnoea. Pre-operative BNP *was not* significant within the model. When compared to model 7, the addition of pre-operative BNP in model 8 did not improve the predictive strength of the model, ($p=0.84$) (Figure 33).

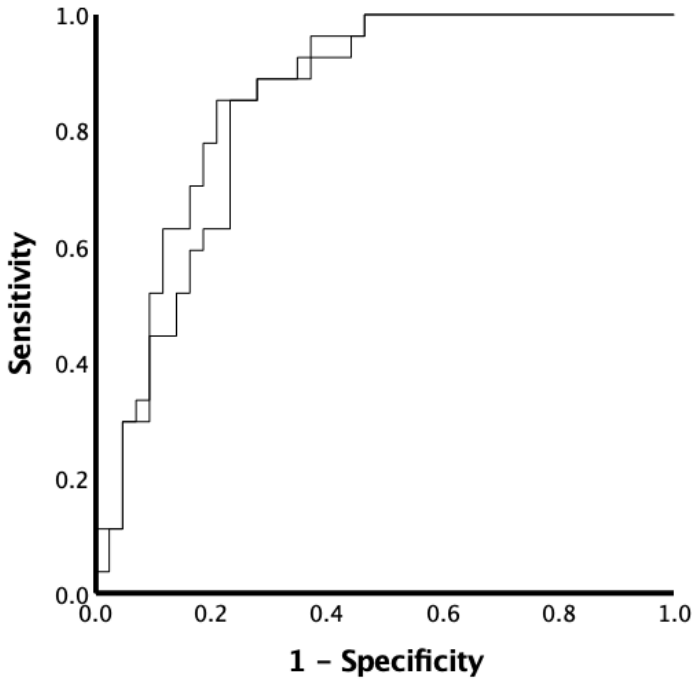

Figure 33 - Receiver operator characteristic curves: Model 7 and Model 8 to predict MRC score >2 at three months post-operatively.
AUROCC Model 7 = 0.83, AUROCC Model 8 = 0.85. p=0.84 (95% CI -0.05 – 0.06), pairwise comparison of AUROCC. (n=75).

Variable	Beta co-efficient	Significance (p-value)	OR	95% CI for OR
Pre-op BNP	-0.01	0.10	0.99	0.98 - 1.00
ppoFEV$_1$%*	-0.09	**<0.01**	0.91	0.85 - 0.98
ppoDLCO%*	0.07	0.07	1.08	0.99 - 1.17
BMI	0.08	0.20	1.08	0.96 - 1.21
Pre-op BPI score	0.09	**<0.01**	1.10	1.03 - 1.17
Pre-op VAS (average)	-0.36	0.10	0.70	0.46 - 1.07
Diabetes status	2.57	**0.02**	13.02	1.44 - 117.99
Constant	-1.29	0.52	0.28	-

Table 38 - Pre-operative BNP and logistic regression using forwards stepwise to select next strongest model from all variables to predict dyspnoea at three months post-operatively (Model 8: derivation)
Pre-op = pre-operative, ppoFEV$_1$% = Predicted post-operative forced expiratory volume in one second, BMI = body mass index, BPI = Brief pain inventory score, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BNP = B-Type natriuretic peptide, VAS = visual analogue pain score, OR = Odds ratio. Significant results highlighted in bold. *= Linear variable.

10.5 Model Derivation: summary of results (models 5-8)

Summary of the results can be seen below in Table 39 and Table 40 for models 5-8. These models were intended to improve current practice of prediction of dyspnoea following lung resection surgery. All models were significant predictors of post-operative dyspnoea at three months and ppoFEV$_1$% and ppoDLCO% performed best when used as continuous variables within these models. Pre-

operative quality of life score assessed by EQ-5DL appears to be a strong predictor of dyspnoea. Other QoL variables were incorporated into the models as predictors alongside BMI and diabetes status. There is no difference in the predictive strength of model 5 and 6, (p=0.52, pairwise comparison) (Figure 32). There is also no difference in the predictive strength of model 7 and 8, (p=0.84, pairwise comparison) (Figure 33). Therefore, similar to models 1-4 derived in chapter 9, pre-operative BNP did *not* increase the predictive strength of the models.

Model	Variables within model	Variable significance (p-value)	Model significance (p-value)	AUROCC Value (95% CI)	Mean Brier score
5	Age Gender ppoFEV$_1$% ***** ppoDLCO% ***** EQ-5DL index	0.17 0.19 **0.01** **0.05** **0.01**	**<0.01**	0.81 (0.71 -0.91)	0.18
6	Age Gender ppoFEV$_1$% ***** ppoDLCO% ***** Pre-op BNP EQ-5DL index	0.67 0.14 **0.01** **0.02** 0.13 **0.01**	**<0.01**	0.82 (0.73 -0.92)	0.17
7	ppoFEV$_1$% ***** BMI Pre-op BPI Diabetes	**0.04** 0.06 **0.01** 0.08	**<0.01**	0.83 (0.74 -0.92)	0.17
8	ppoFEV$_1$% ***** Pre-op BPI Diabetes Pre-op BNP ***** ppoDLCO%***** BMI Pre-op EQ-5DL VAS	**0.01** **0.01** **0.02** 0.10 0.07 0.20 0.10	**<0.01**	0.85 (0.76 -0.93)	0.16

Table 39 - Model derivation: summary of results (Model 5-8)
AUROCC = Area under receiver operator characteristic curve, Pre-op = pre-operative, CI = Confidence Interval, ppoFEV1 = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BNP = B-Type natriuretic peptide, BMI = body mass index, BPI = Brief pain inventory score, VAS = visual analogue pain score, EQ-5DL = Euro Qol 5 level 5-dimensional scoring tool. Significant results highlighted in bold.
*= Linear variable.

Model	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Correct (%)
5	48	87	68	74	72
6	56	82	65	76	77
7	58	85	71	76	74
8	65	80	68	78	74

Table 40 - Model 5-8 sensitivity and specificity summary
PPV = Positive predictive value, NPV = Negative predictive value.

10.6 Comparison of models: summary

Model 3 (*age, gender, ppoFEV$_1$% (linear) and ppoDLCO% (linear)*) is representative of what is currently used to assess patients prior to lung resection surgery (*conventional model*). The addition of pre-operative BNP into this model did not increase its predictive strength.

A comparison was performed between model 3 (*conventional*) and model 5 (*next best model; age, gender, ppoFEV$_1$% (linear), ppoDLCO% (linear) and next best variable pre-operative EQ-5DL index score*). No discernible increase in predictive strength existed between these models at AUROCC comparison, (p=0.06). The Brier skill score confirms that model 5 is a stronger predictive model than model 3, (0.20) - a positive score reflects that the comparator model is superior to the baseline model (section 7.21).

A comparison was also performed between model 3 (*conventional*) and model 7 (*hypothesis free model; ppoFEV$_1$% (linear), BMI, pre-operative BPI score and diabetes status*) (Table 41). Model 7 significantly improved model discrimination upon comparison of AUROCC, (p=0.03). The NRI value (0.13) was also low when comparing reclassification between them. The Brier skill score confirms that model 7 is a stronger predictive model than model 3, (0.21) - a positive score reflects that the comparator model is superior to the baseline model (section 7.21). Notably, *only model 7 significantly improved model discrimination when compared to model 3*.

	Model 5	**Model 7**
Difference in AUROCC	p = 0.06	**p = 0.03**
NRI	0.13	0.26
Brier SS	0.20	0.21

Table 41 - Model derivation comparison with model 3
Significant results highlighted in bold. SS= Skill score, NRI = Net reclassification improvement score. AUROCC = Area under receiver operator characteristic curve. Bold values denote significance.

10.7 Worked example of net reclassification improvement (NRI): model 3 Vs model 7

NRI quantifies how well a new model correctly *reclassifies* patients, i.e. does the new model improve the predictive strength of the previous model, (section 7.22). A worked example of the NRI calculation, comparing models 3 and 7, is summarised in Table 42; model 3 being the original (*conventional*) risk prediction model and model 7 the best performing risk prediction model (*hypothesis free model*).

Patients are categorised according to the baseline model where some are correctly classified, but a proportion will be incorrectly classified. NRI then systematically examines the reclassification of patients with the new model compared to the baseline. Many patients will not change category, but a proportion will be re-classified.

In the example (Table 42), 14 patients were correctly classified as high risk by both models and 28 patients were correctly classified as low risk by both models. Seven patients were incorrectly classified by both tests. Compared to model 3 (*conventional*) ,14 patients were correctly reclassified and seven patients were incorrectly classified by model 7 (*hypothesis free model)*.

Improved classification
No change
Worse classification

Patients with primary outcome n=27		
	Risk classification using model 3	
Risk classification using model 7	High risk	Low risk
High	14	9
Low	2	2

Improved classification	9
No change	16
Worse	2

Patients without primary outcome n=43		
	Risk classification using model 3	
Risk classification using model 7	High risk	Low risk
High	5	5
Low	5	28

Improved classification	5
No change	33
Worse	5

Table 42 - Net reclassification improvement (NRI) (Model 7 versus Model 3)
Primary outcome = Patients with a predicted Medical Research Score >2 at three months post-operatively. In patients with the primary outcome, n=7 (0.26) patients were correctly reclassified. In patients without the primary outcome, no patients were correctly reclassified. NRI = Net reclassification improvement is the sum of the correctly reclassified patients with and without the primary outcome, (NRI = 0.26).

11 External validation of risk prediction models

Validation was performed with an external dataset of patients (n=85) from three cardiothoracic centres within the UK (Section 7.18). The external validation CONSORT diagram is displayed in Figure 34. Only models 3, 5 and 7, as discussed in section 10.6, were selected to be externally validated based on performance at model derivation. Model 3 is the *'conventional model'* reflecting current practice. Model 5 builds on model 3 and selects the *'next best variable'*, in addition to current practice. Finally, model 7 is selected as a '*hypothesis free'* model, without any variables as a baseline. Pre-operative BNP did *not* improve the predictive strength of any model; therefore, no attempt was made to externally validate models 4,6 and 8. Similarly, models 1 and 2 had poor predictive strength, even within the derivation dataset, meaning no attempt at external validation was made for these models either.

Similar to the derivation population, five patients had a *starting* MRC score >2 after being consented into the study and therefore were excluded. Six patients had wedge resections performed due to a change in surgical plan intra-operatively and 14 were lost to follow up at three months post-operatively, mainly due to the COVID-19 pandemic (Figure 34).

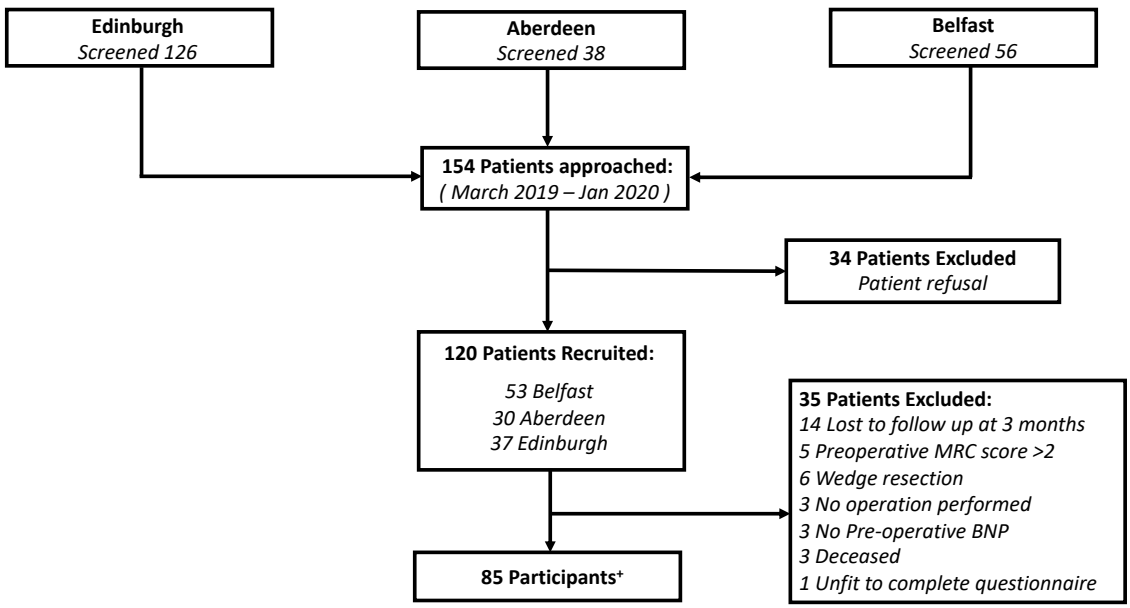

Figure 34 - External validation population CONSORT diagram.
(n=85) + = 85 patients with data available for analysis at 3 months post-operatively.

11.1 Patient demographics for validation dataset: comparison

A comparison was performed between the derivation and validation populations to determine if any baseline differences existed (Table 43). Only those variables *within* the models created in chapters 9 and 10 were compared.

Patient Demographic	Derivation population (n= 75)	Validation population (n=85)	Group difference (p-value)
Primary outcome (MRC)	27(36%)	22(26%)	0.17^{+}
Age (years)	66 (12)	68(9)	0.24*****
Gender (Male)	34(45%)	37(44%)	0.80^{+}
ppoFEV$_1$ (%)	70(17)	72(18)	0.50*****
ppoDLCO (%)	58(14)	62(17)	0.09*****
Pre-op BNP (pg/L)	48(60)	58(75)	0.33*****
Pre-op EQ-5DL (index value)	0.78(0.18)	0.86(0.17)	**<0.01***
Diabetes status (Y)	10(13%)	3(4%)	**0.02**$^{+}$
BMI (kg/m^2)	28(6)	26(5)	0.10*****
Pre-op BPI score	10(13)	7(12)	0.13*****
Pre-op VAS score	0.5(1.5)	0.6(1.5)	0.50*****

Table 43 - Derivation and validation model comparison
Values are number (%), mean (SD). n represents number of patients with data available for each variable. ppoFEV1 = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted Post-Operative Carbon monoxide diffusing capacity, Pre-op BNP = Pre-operative B-Type natriuretic peptide, BMI = body mass index, Pre-op VAS = Pre-operative visual analogue score, Pre-op BPI = Pre-operative brief pain inventory score, EQ-5DL = Euro Qol 5 level 5-dimensional scoring tool, * = independent samples t-test, $^{+}$ = Pearson chi squared test. Significant results highlighted in bold.

Apart from a difference in pre-operative EQ-5DL index scoring and diabetes status, there were no differences between the derivation and validation cohorts. A non-significant difference of ppoDLCO% existed between the two groups. Models selected to be validated contained at least one of these elements. Therefore, given the difference in populations it may be hypothesised that external validation discrimination will not be as strong as the internal derivation dataset. As pre-operative BNP did *not* improve the predictive strength of any model no attempt was made to externally validate any model containing this.

11.2 Missing data

Missing data was handled as described in methods (7.20) and summarised in Table 44. These values were imputed from the *derivation* dataset. Any missing data when validating the models was handled by overall and subgroup mean imputation from the derivation data. Subgroups were determined by ppoFEV_1% and ppoDLCO% categories. For example, whether a patient has a ppoFEV_1% of less than or greater than 40% would determine what value was imputed.

Variable missing	Missing (%)	Data used
ppoDLCO%	2	If ppoFEV_1<40% then ppoDLCO% = 39%
ppoDLCO%	20	If ppoFEV_1>40% then ppoDLCO% = 70%
ppoFEV_1%	1	If ppoDLCO%>40% then ppoFEV_1% = 73%
ppoFEV_1%	0	If ppoDLCO%<40% then ppoFEV_1% = 49%
BMI (kg/m^2)	1	27.9
Pre-op BPI score	2	10
Pre-op VAS score	2	0.5

Table 44 - Missing data: summary
ppoFEV_1 = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BMI = body mass index, BPI = brief pain inventory score, VAS = visual analogue pain score, Pre-op = pre-operative.

11.3 External validation

Eighty-five patients were used to validate models 3 (*conventional*) ,5 (*next best variable*) and 7 (*hypothesis free model*), as summarised in Table 45. The models derived (using derived beta co-efficients and intercepts) were run in the validation dataset. The beta co-efficients and intercepts were generated by and obtained from the models at internal derivation.

Model	Variables	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Brier score
3	Age Gender ppoFEV$_1$%* ppoDLCO%*	55	68	38	81	65	0.20
5	Age Gender ppoFEV$_1$%* ppoDLCO%* EQ-5DL-index	45	71	36	79	64	0.19
7	ppoFEV$_1$% * BMI Pre-op BPI Diabetes	50	73	39	81	67	0.20

Table 45 - External validation: summary
Model 3 = *'conventional',* Model 5 = '*next best variable',* Model 7 = '*hypothesis free'*, PPV= Positive predictive value, NPV = Negative predictive value. ppoFEV$_1$ = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BMI = body mass index, BPI = brief pain inventory score, VAS = visual analogue pain score, Pre-op = pre-operative. * = linear variable

11.3.1 Model discrimination

Model discrimination is summarised in Table 46 for models 3,5 and 7. The AUROCC value for *all* models falls between 0.6 - 0.7 and can be described as having poor discrimination, according to Hosmer et al.[300]

Model	Variables	AUROCC	95% CI
3	Age Gender ppoFEV$_1$%* ppoDLCO%*	0.68	0.55 – 0.80
5	Age Gender ppoFEV$_1$%* ppoDLCO%* EQ-5DL-index	0.66	0.52 – 0.80
7	ppoFEV$_1$%* BMI Pre-op BPI Diabetes	0.62	0.48 – 0.77

Table 46 - External Validation: model discrimination
Model 3 = *'conventional',* Model 5 = '*next best variable',* Model 7 = '*hypothesis free',* AUROCC = Area under receiver operator curve characteristic, CI = Confidence interval. ppoFEV$_1$ = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BMI = body mass index, BPI = brief pain inventory score, VAS = visual analogue pain score, Pre-op = pre-operative. * = linear variable

11.3.2 Mean calibration

Mean calibration is summarised in Table 47 for models 3,5 and 7. The prevalence of post-operative dyspnoea (defined as MRC>2) in the external group was n=22 (26%). The average estimated risk given by all models was above this, which indicates there is a tendency for all models to overestimate risk.[292]

Model	Mean Predicted Probability (%)	Overall event rate (%)
3	40	26
5	31	26
7	29	26

Table 47 - Mean calibration

11.3.3 Model calibration

Calibration curves were drawn for model 3 in the external dataset. Figure 35 displays results of calibration when the cohort was divided into deciles;10 groups of 10%. Overall, poor calibration can be observed; in the 2nd, 3rd and 5th deciles patients predicted probability tended to be underestimated. After the 5th decile, overestimation occurs, with the mean values of 8th, 9th and 10th deciles far from the perfect calibration line.

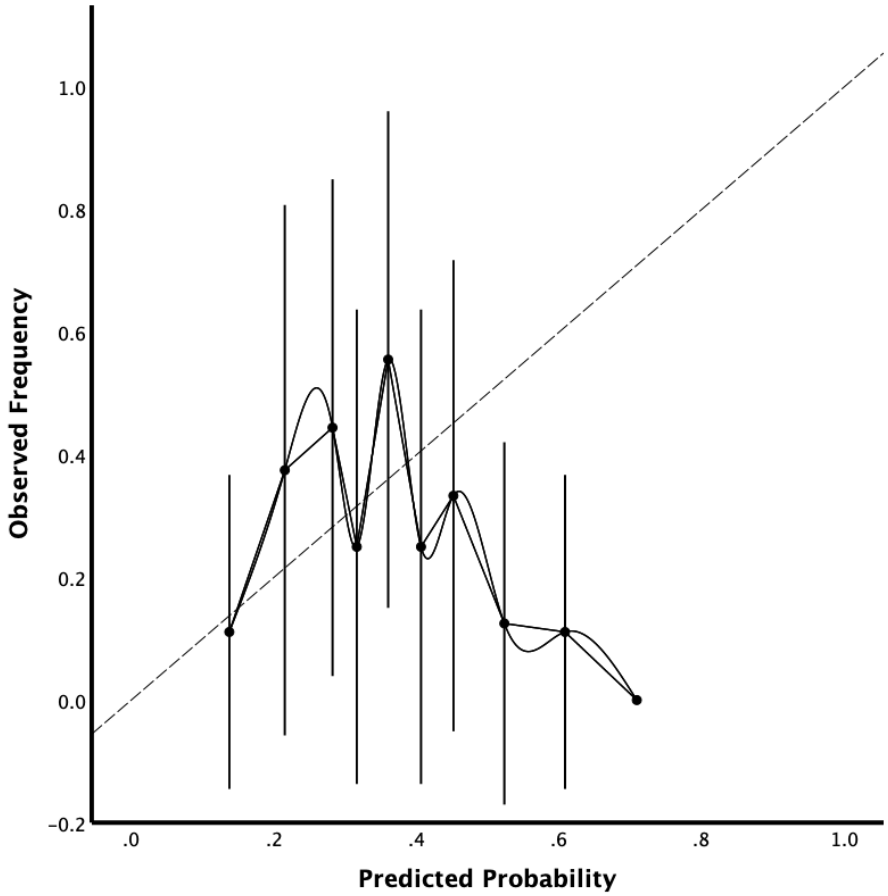

Figure 35 – *'Conventional'* Model 3 calibration curve in external dataset.
Grouped in deciles for analysis. Mean displayed with 95% CI for each group. Curved line joining means values represents spline curve. Straight dashed line represents perfect prediction. (n=85).

Figure 36 displays model 3 calibration, but with data grouped into quintiles. At visual inspection, this makes it easier to observe model 3 predicted probability well at lower values – the first 3 quintiles (60%). The mean values sit close to line of perfect calibration before overestimating at higher values.

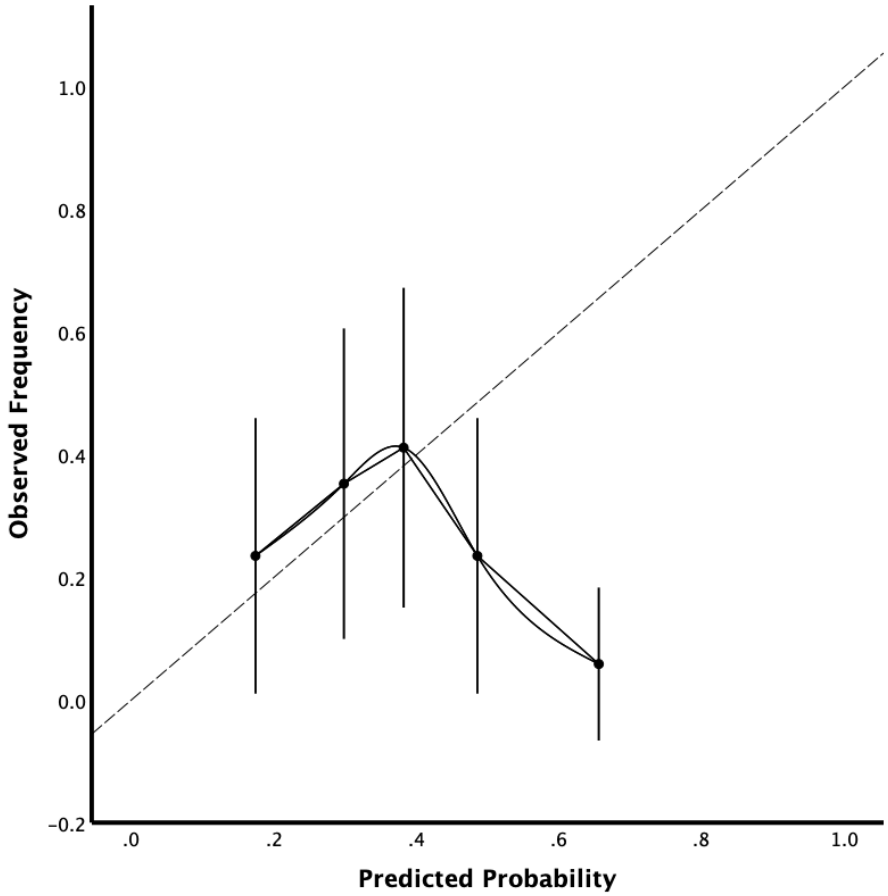

Figure 36 – *'Conventional'* Model 3 calibration curve in external dataset.
Grouped in quintiles for analysis. Mean displayed with 95% CI for each group. Curved line joining mean values represents spline curve. Straight dashed line represents perfect prediction. (n=85).

Calibration curves were also drawn for model 5; both in deciles and quintiles (Figure 37 and Figure 38, respectively). Poor calibration is observed with underestimation at low values of predicted probability and overestimation at high values of predicted probability.

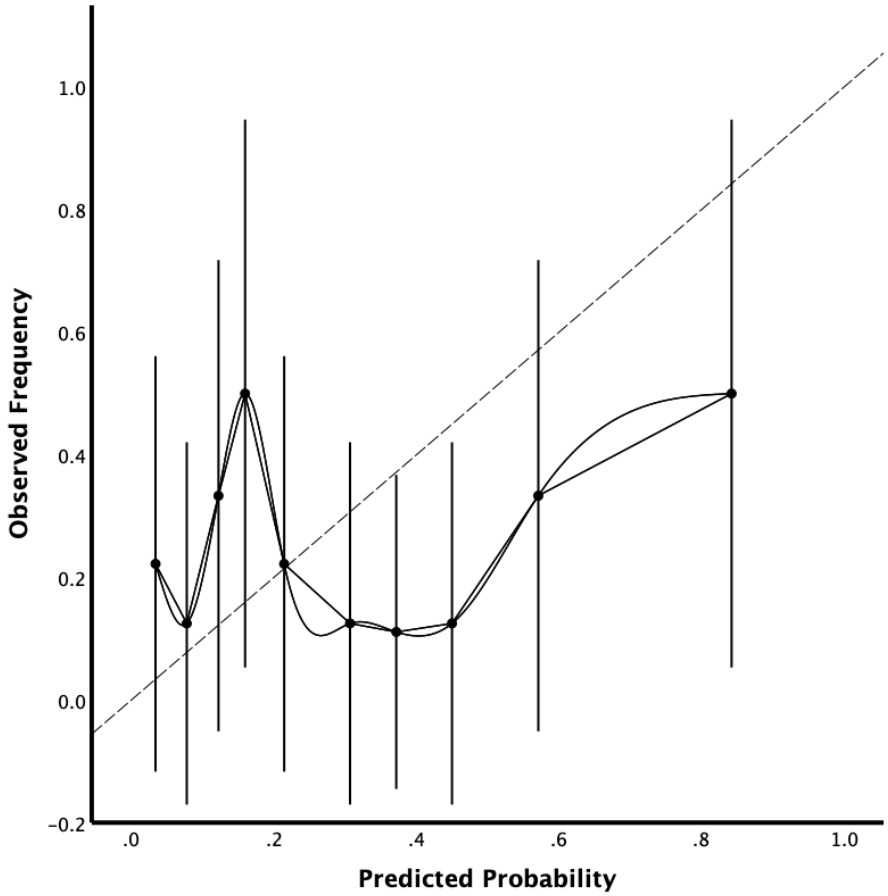

Figure 37 – *'Next best variable'* Model 5 calibration curve in external dataset.
Grouped in deciles for analysis. Mean displayed with 95% CI for each group. Curved line joining means values represents spline curve. Straight dashed line represents perfect prediction. **(**n=85).

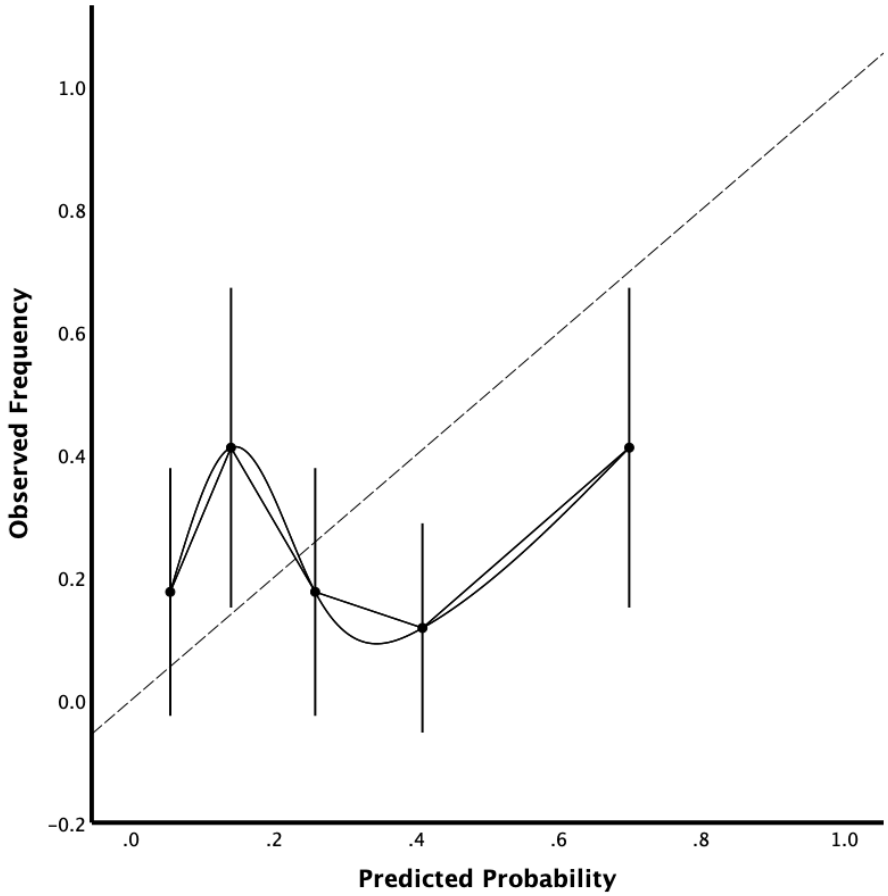

Figure 38 - *'Next best variable'* Model 5 calibration curve in external dataset
Grouped in quintiles for analysis. Mean displayed with 95% CI for each group. Curved line joining mean values represents spline curve. Straight dashed line represents perfect prediction. (n=85).

Finally, curves were drawn for *new* model 7 to observe calibration, Figure 39 and Figure 40. Of all models, model 7 displays the best calibration, mean values sit close to and all 95% CI values cross the *perfect prediction line*. There existed no significant difference when AUROCC values were compared for the 3 models. Therefore, since model 7 displays the best calibration in the external dataset this has been selected as the best model and will be the model in which changing sensitivity will be explored.

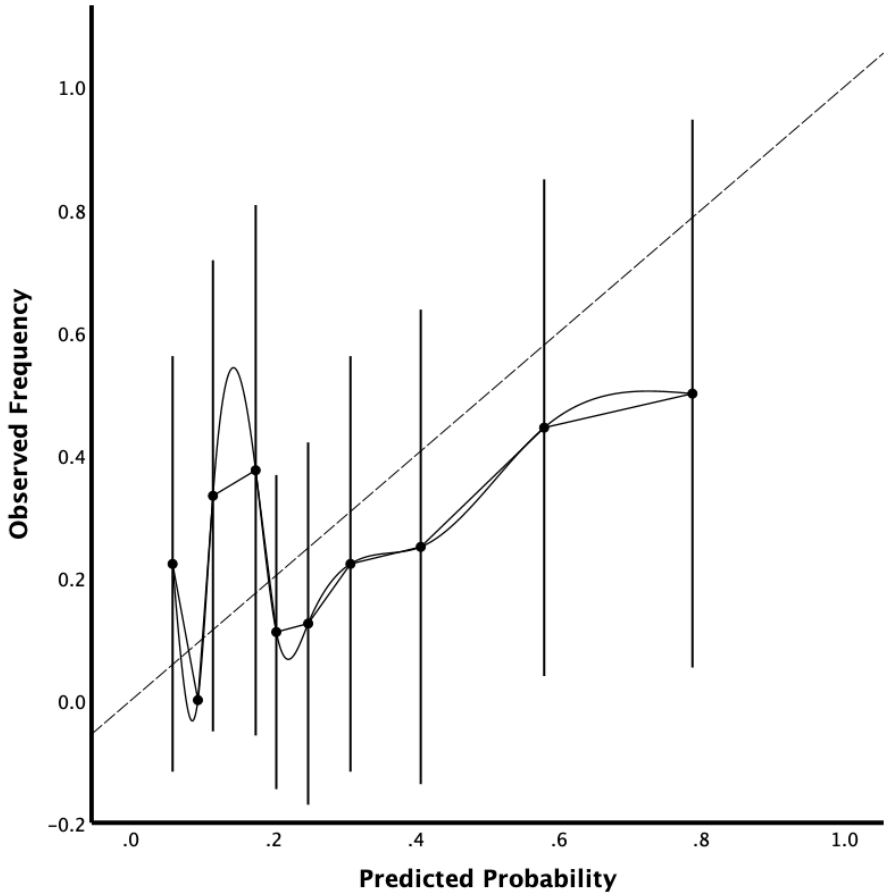

Figure 39 – *'Hypothesis free'* Model 7 calibration curve in external dataset
Grouped in deciles for analysis. Mean displayed with 95% CI for each group. Curved line joining means values represents spline curve. Straight dashed line represents perfect prediction. (n=85).

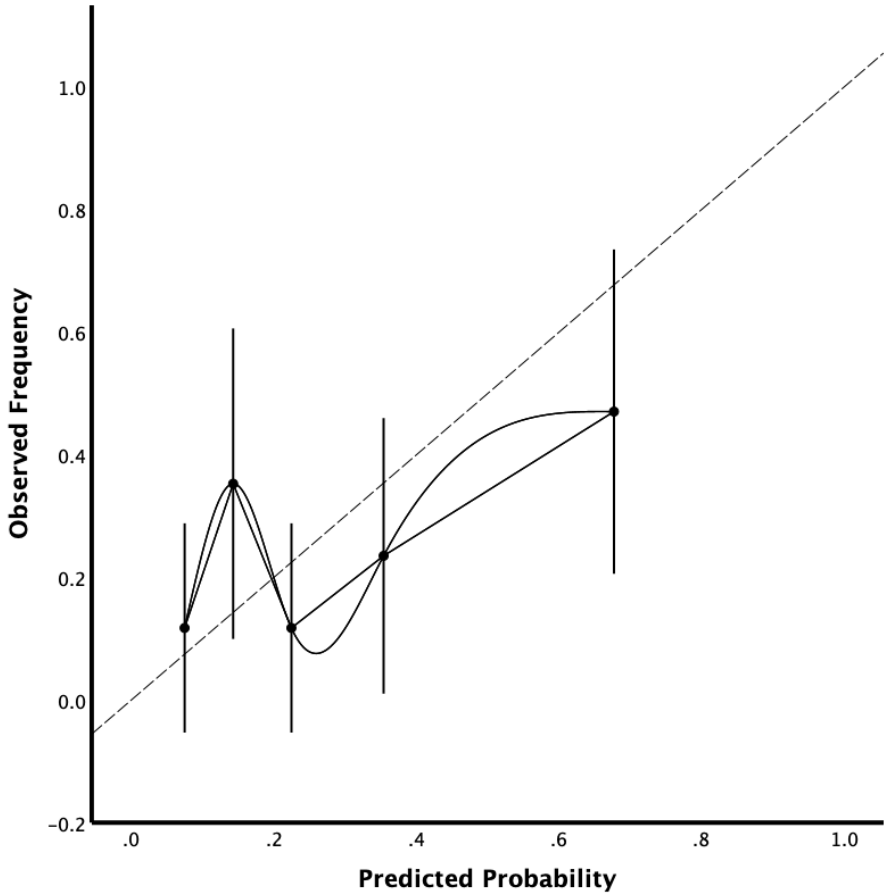

Figure 40 - *'Hypothesis free'* Model 7 calibration curve in external dataset
Grouped in quintiles for analysis. Mean displayed with 95% CI for each group. Curved line joining mean values represents spline curve. Straight dashed line represents perfect prediction. (n=85).

11.4 Changing sensitivity: model 7

The strength of a model lies in its ability to improve clinical decision making. The aim is to improve *conventional* prediction of the risk of post-operative dyspnoea and allow identification of a sub-population for targeted recruitment (prognostic enrichment) to interventional studies seeking to mitigate this risk. For prognostic enrichment, models should have high sensitivity and high NPV, targeting those who would benefit most from low-risk interventions. This will ensure recruitment of most patients at risk of post-operative dyspnoea whilst necessarily accepting a higher rate of false positives. Given the proposed intervention this scoring tool would be used for, this would be an acceptable risk. To that end, the cut-off of probability of the best model ('hypothesis free' model 7) was varied between high/low risk patients to optimise improved sensitivity and identify those most appropriate for recruitment into future studies. Table 48 summarises the results of varying probability to prioritise sensitivity of model 7 in the internal and external datasets. As expected, the model performed well in the internal dataset,

with a sensitivity and NPV >90%; this would work well for prediction into a prognostic enrichment study as described. However, model 7 could not achieve 90% sensitivity within the external dataset when probability was adjusted. The highest sensitivity that could be achieved was 59%, to maintain a NPV of 82%.

Model	Dataset	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Correct (%)	Brier Score	n
7	External	59	64	82	36	62	0.19	85
7	Internal	93	65	94	60	75	0.17	75

Table 48 - Changing sensitivity: *'hypothesis free'* model 7
PPV = Positive predictive value, NPV = Negative predictive value, n= number of patients with full dataset available for analysis. Model 7 contains - ppoFEV_1%, BMI, Pre-operative BPI, Diabetes status.

11.5 Removal of missing data: model 3

The performance of model 3 was assessed again, this time with no imputation of missing data (n=67). This was an ad-hoc analysis to explore if imputation of missing data could have contributed to overfitting of the models in the internal dataset and therefore perform poorly in the external dataset. Summary of results are displayed in Table 49 and it can be observed that removal of imputed data did not materially change the overall performance of the model.

Test	Missing data removed (%)	Missing data imputed (%)
Sensitivity	55	55
Specificity	70	68
PPV	44	38
NPV	78	81

Table 49 - Model 3: Missing data removed.
PPV = Positive predictive value, NPV = Negative predictive value.

12 Secondary outcome analysis

12.1 Acute complications and length of hospital stay

This section explores acute complications and length of hospital stay and the association between them and BNP and post-operative dyspnoea. Table 50 summarises the incidence of the pre-defined complications described in section 7.15, the most common post-operative complication being atrial fibrillation n=10(11%). These patients were from the base centre (GJNH) and therefore there was little missing data (<1%). As described in methods (section 7.19), the ESTS have a pre-defined list of post-operative complications, the incidence of which is summarised in Table 51. Forty-five percent of patients (n=42) had a least one ESTS complication with 47% (n=44) having a cardiopulmonary complication following surgery. All patients within the study had 'minor' complications, as defined by the clavien-dindo classification described in section 7.15. Nausea and vomiting were the most commonly observed ESTS complications in 28% and 16% of patients, respectively.

Post-operative data	n	Descriptive Statistics
Incidence AF	93	10 (11%)
Treatment of AF	93	10 (11%)
Vasopressor administration	93	8 (9%)
Duration of vasopressor treatment (hours)	8	11 (5)
Inotrope administration	93	5 (5%)
Duration of inotrope treatment (hours)	5	9 (4)
High flow nasal cannula	93	1 (1%)
Duration of high flow nasal cannula (hours)	1	5 (0)
NIV	93	3 (3%)
NIV duration (hours)	3	10 (10,39)

Table 50 - Post-operative complications (non-ESTS complications)
Values are number (%), mean (SD) or median (IQR). n represents number of patients with data available for each variable. NIV = Non-invasive ventilation, AF = Atrial fibrillation, ESTS = European society of thoracic surgeons.

Post-operative complications	n	Descriptive statistics
Nausea	93	26 (28%)
Vomiting	93	15 (16%)
Atelectasis	93	2 (2%)
Pneumonia	93	13 (14%)
ARDS	93	0 (0%)
Aspiration pneumonia	93	0 (0%)
Pulmonary embolism	93	0 (0%)
Atrial arrhythmias	93	10 (11%)
Ventricular arrhythmias	93	0 (0%)
Myocardial infarction	93	0 (0%)
Deep vein thrombosis	93	0 (0%)
Renal failure	93	0 (0%)
Urinary retention	93	6 (7%)
Hypotension	93	22 (24%)
Neurological complication	93	1 (1%)
Any ESTS complication	93	42(45%)
Pulmonary complication: ESTS	93	15(16%)
Cardiac complications: ESTS	93	31(33%)
Cardiopulmonary complications: ESTS	93	44(47%)
Greater than one POC	93	15(16%)

Table 51 - Post-operative complications: European Society of Thoracic Surgeons (ESTS) definitions

Values are number (%). n represents number of patients with data available for each variable. ESTS = European Society of Thoracic Surgeons, POC = Post-operative complication, ARDS = Acute respiratory distress syndrome.

Table 52 summarises post-operative high dependency/intensive care requirements, hospital mortality and length of hospital stay. No patients from this study died before hospital discharge. Only one patient required ICU care following surgery for 75 hours for non-invasive ventilation and high flow nasal cannula oxygen delivery. Most patients at GJNH have HDU admission for first 24 hours following surgery, median duration of stay in hours in this cohort being 23(20,26) hours. Median length of hospital stay in days was 6(5,9).

Post-operative LOS	n	Descriptive Statistics
HDU stay (hours)	93	23 (20,26)
Hospital stay (days)	93	6 (5,9)
ICU admission	93	1 (1%)
ICU duration (hours)	1	75 (0)
Hospital mortality	93	0 (0%)

Table 52 - Post-operative high dependency, intensive care requirements and hospital mortality

Values are number (%), mean (SD) or median (IQR). n represents number of patients with data available for each variable. LOS = Length of stay, HDU = High dependency unit, ICU = Intensive care unit.

12.1.1 Incidence of acute complications by primary outcome group

A comparison was made between pre-defined post-operative complications between those with and without dyspnoea (MRC>2) at three months (Table 53). There was a higher incidence of new post-operative atrial fibrillation in the group who did not have post-operative dyspnoea (n=8(17%)), p=0.02**,** Pearson chi-squared). There was also an increased incidence of vasopressor requirement in those with post-operative dyspnoea (n=4(15%)), p=0.03**,** Pearson chi-squared). No patients required high flow nasal cannula post-operatively from either group. No patient required ICU admission from either group.

Post-op Complication	MRC≤2 (n=48)	MRC>2 (n=27)	Difference (p-value)
Incidence of AF	9(19%)	1(4%)	0.07[+]
Treatment of AF	8(17%)	0(0%)	**0.02**[+]
Vasopressor administration	1(2%)	4(15%)	**0.03**[+]
Duration of vasopressor treatment (hours)	9(9,9)	8(6,15)	1.00[#]
Inotrope administration	2(4%)	0(0%)	0.28[+]
Duration of Inotrope treatment (hours)	10(10,10)	-	-
Incidence of NIV	2(4%)	0(0%)	0.28[+]
NIV duration (hours)	9.5(0.7)	-	-

Table 53 - Univariate analysis: post-operative complications
Values are number (%), mean (SD) or median (IQR). n represents number of patients with data available for each variable. Post-op = Post-operative. Post-op = post-operative, AF = atrial fibrillation, HFNC = high flow nasal cannula, NIV = non-invasive ventilation. # = Mann-Whitney U test, [+] = Pearson chi squared test. Significant results highlighted in bold.

A comparison was made between ESTS post-operative complication rate in those with and without dyspnoea at three-month time point (Table 54). There were no differences between the two groups.

Complication	MRC≤2 (n=48)	MRC>2 (n=27)	Difference (p-value)
Any ESTS complication	23(48%)	8(30%)	0.12[+]
Pulmonary complication ESTS	7(15%)	1(4%)	0.14[+]
Cardiac complications ESTS	17(35%)	6(22%)	0.23[+]
Cardiopulmonary complications ESTS	25(52%)	9(33%)	0.11[+]
Greater than one POC	9(19%)	1(4%)	0.06[+]

Table 54 – Incidence of ESTS defined post-operative complications by primary outcome group
Values are number (%). n represents number of patients with data available for each variable. ESTS = European Society of Thoracic Surgeons, POC = Post-operative complication. [+] = Pearson chi squared test

12.1.2 Post-operative high dependency, intensive care requirements and hospital mortality: primary outcome group

A comparison was made between HDU/ICU requirements, length of hospital stay and hospital mortality in those with and without dyspnoea at 3-months; there were no differences between the groups (Table 55). No patient from either group had an ICU admission or hospital mortality.

Post-Operative LOS	MRC≤2 (n=48)	MRC>2 (n=27)	Difference (p-value)
HDU stay (hours)	23 (19,26)	22 (19,25)	0.83
Hospital stay (days)	5.5 (4,8)	5 (4,7)	0.72

Table 55 - Post-operative high dependency, intensive care requirements and hospital mortality: primary outcome group
Values are median (IQR). n represents number of patients with data available for each variable. LOS = Length of stay, HDU = High dependency unit, ICU = Intensive care unit.

12.2 Intra-operative management by primary outcome group

A comparison was performed of intra-operative details between those with and without dyspnoea at three-months (Table 56) to determine if anaesthetic technique, analgesic technique, surgery performed or operation type had any influence on dyspnoea at 3-months. There were no differences between the two groups upon comparison.

Intraoperative details	MRC≤2	Descriptive Statistics	MRC>2	Descriptive Statistics	Difference (p-value)
Technique					0.87+
Total intravenous	47	27(57%)	27	15(56%)	
Volatile	47	20(43%)	27	12(44%)	
Analgesia					
Epidural Catheter	47	7(15%)	27	1(4%)	0.13+
PV Catheter	47	22(47%)	27	17(63%)	0.18+
PV Injection	47	16(34%)	27	10(37%)	0.80+
Morphine PCA	47	29(62%)	27	17(63%)	0.91+
Fentanyl PCA	47	4(9%)	27	1(4%)	0.43+
Oral Opiates	47	15(32%)	27	4(15%)	0.11+
IC LA Infiltration	46	3(7%)	27	2(7%)	0.88+
Operative details					0.46+
Lobectomy	48	45(94%)	27	23(85%)	
Bi-Lobectomy	48	1(2%)	27	1(4%)	
Pneumonectomy	48	2(4%)	27	3(11%)	
LOS (minutes)	40	142(122,194)	26	138(102,184)	0.42#
LOA (minutes)	45	170(144,219)	27	159(137,216)	0.50#
OLV (minutes)	38	125(92,146)	23	105(80,162)	0.32#
Operation type					0.41+
Open	48	19(40%)	27	7(26%)	
Video Assisted	48	25(52%)	27	16(59%)	
Robotic Assisted	48	4(8%)	27	4(15%)	
Operation side					0.07+
Left	48	18(38%)	27	16(59%)	
Right	48	30(63%)	27	11(41%)	

Table 56 - Intra-operative management by primary outcome
Values are number (%), mean (SD) or median (IQR). n represents number of patients available for each variable. * = independent samples t-test, # = Mann-Whitney U test, + = Pearson chi squared test. PCA = Patient controlled analgesia, LA = Local anaesthetic, PV = paravertebral, IC = intercostal, LOS = length of surgery, LOA = Length of anaesthesia, OLV = One lung ventilation.

12.3 Peri-operative change in B-Type natriuretic peptide

Based on the previous work described in section 5.1.4 by Young et al demonstrating increased BNP in patients who experience reduced functional capacity at all peri-operative timepoints, it was hypothesised the change in BNP from pre-operative levels to post-operative levels *may* be predictive of future shortness of breath. Although less useful as a predictive marker given surgery has already taken place, this may give some insight into the mechanisms underlying post-operative dyspnoea. Median *peak* post-operative BNP (pg/ml) was 87 (46,182)

and median peri-operative change in BNP (pg/ml) was 45(1,116). Table 57 summarises the difference in peri-operative BNP levels in those with and without MRC >2 at 3 months.

Pre-operative BNP levels increased post operatively day 2 (p<0.01) and day 3, (p<0.01). Post-operative peak BNP levels were also higher across all patients recruited at GJNH, (p<0.01). However, peak post-operative BNP was no different in those with and without the primary outcome, (Figure 41).

Those with the biggest change or increase in peri-operative BNP, may potentially become the most breathless at 3 months. However, change in peri-operative BNP level was no different in those who were breathless post-operatively, with an MRC >2 (primary outcome), or not. (n=74, p=0.2).

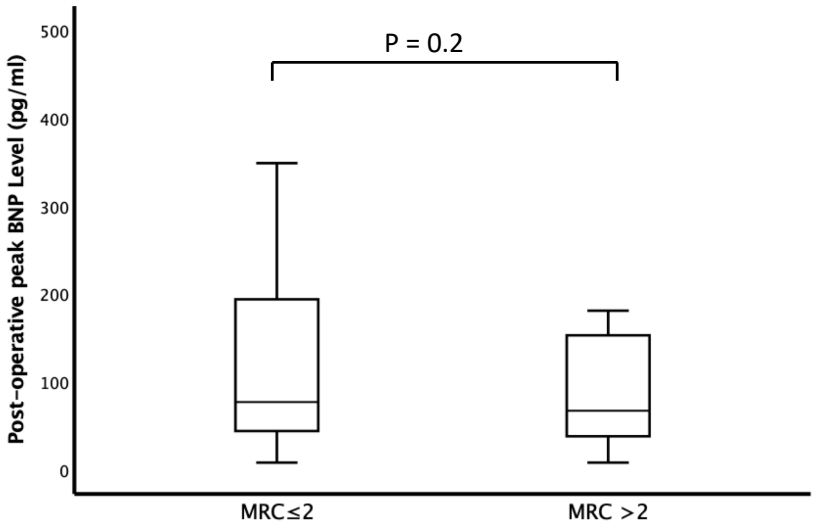

Figure 41 - Peak post-operative BNP (pg/ml) and primary outcome.
MRC = Medical research council dyspnoea score, BNP = B-Type natriuretic peptide. **(**n=74)

Post-operative data	MRC<2		MRC>2		(p-value)
Timepoint	**n**	**BNP (pg/L)**	**n**	**BNP (pg/L)**	
Pre-op	48	31(17,53)	27	28(7.5,65)	0.785#
Post-op day 2	47	67(37,141)	27	48(18,116)	0.28#
Post-op day 3	42	65(30,159)	25	55(28,113)	0.36#
Post-op peak	47	75(43,175)	27	66(26,134)	0.24#

Table 57 – Peri-operative B-Type natriuretic peptide comparison by primary outcome group
Values are median (IQR). n represents number of patients available for each variable, # = Mann-Whitney U test. Pre-op = Pre-operative, post-op = Post-operative, BNP = B-Type natriuretic peptide.

There was no difference in peri-operative BNP levels between those with and without the primary outcome.

12.4 Association between B-Type natriuretic peptide and acute post-operative complications

Association was sought between *pre-operative* BNP and those with and without post-operative complications. No difference in median [IQR] pre-operative BNP was observed between those with and without complications, (32pg/ml (24,62) compared with 30pg/ml (8,59) respectively, p>0,05, Mann-Whitney U). Median *pre-operative* BNP level for those with one less than one complication and those with *more than one* complication was also no different (31pg/ml (8,59) compared with 37pg/ml (25,92) respectively, p=0.26, Mann-Whitney U).

Median [IQR] *post-operative peak* BNP level was the same in those with (97 (57,199) and without (74 (37,165) a post-operative complication (p=0.08, Mann-Whitney U). However, median *post-operative peak* BNP level was higher in those with more than one post-operative complication (p=0.04, Mann-Whitney-U). (Figure 42).

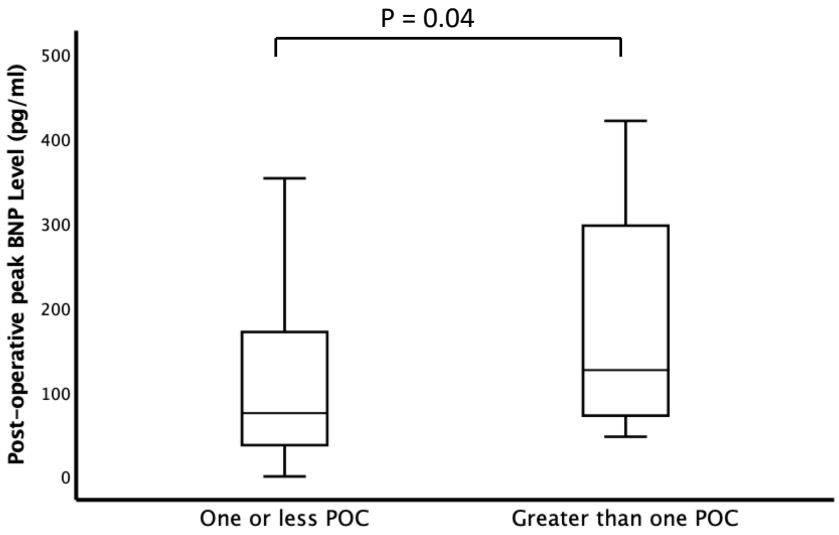

Figure 42 - Peak Post-operative BNP and >1 post-operative complication
POC = post-operative complication. BNP = B-type natriuretic peptide. (n=93).

12.4.1 Association between B-Type natriuretic peptide and new post-operative atrial fibrillation

Pre-operative BNP level was higher in those with new post-operative atrial fibrillation (AF), (p =0.04, Mann Whitney-U) (Figure 43). Median [IQR] *post-operative peak* BNP level was the same in those with and without new post-operative AF (131pg/ml (68,383) compared with 87pg/ml (41,177) respectively, p = 0.12, Mann-Whitney U).

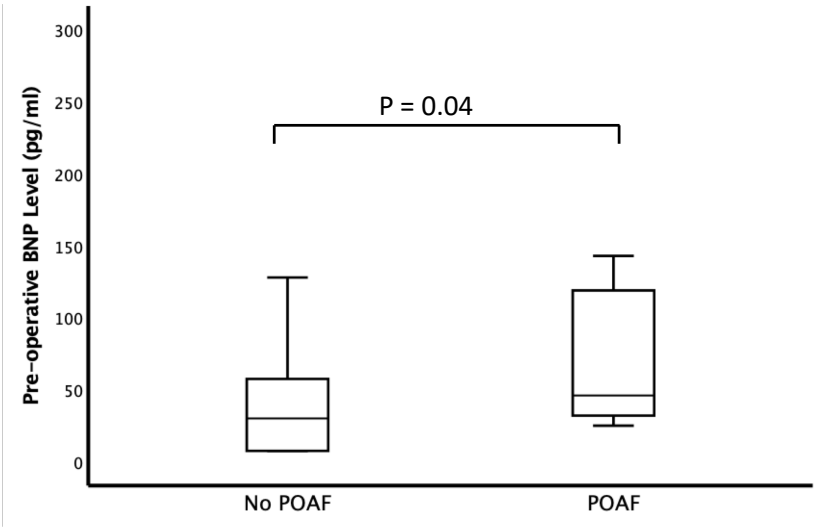

Figure 43 - Pre-operative BNP and new atrial fibrillation
POAF = Post-operative atrial fibrillation. (n=10 patients developed new post-operative AF), BNP = B-type natriuretic peptide. (n=93).

12.4.2 Association between B-Type natriuretic peptide and post-operative cardiopulmonary complications

Association was sought between those with and without cardiopulmonary complications as per ESTS association guidelines, defined in methods chapter 6. Median [IQR] *pre-operative* BNP level was the same for those with and without post-operative cardiopulmonary complications, (30pg/ml (8,59) compared with 32pg/ml (21,53), p= 0.56, Mann Whitney-U).

Median [IQR] *post-operative peak* BNP was higher in those with pulmonary complications than without (152pg/ml (94,336) compared with 75pg/ml (42,174) respectively, p<0.01, Mann Whitney-U) (Figure 44). Median *pre-operative* BNP was the same in those with and without pulmonary complications, p>0.05 (Mann-Whitney-U, data not shown).

Median *pre-operative & post-operative peak* BNP were the same in those with and without cardiac complications, p>0.07 for both, (Mann Whitney-U, data not shown).

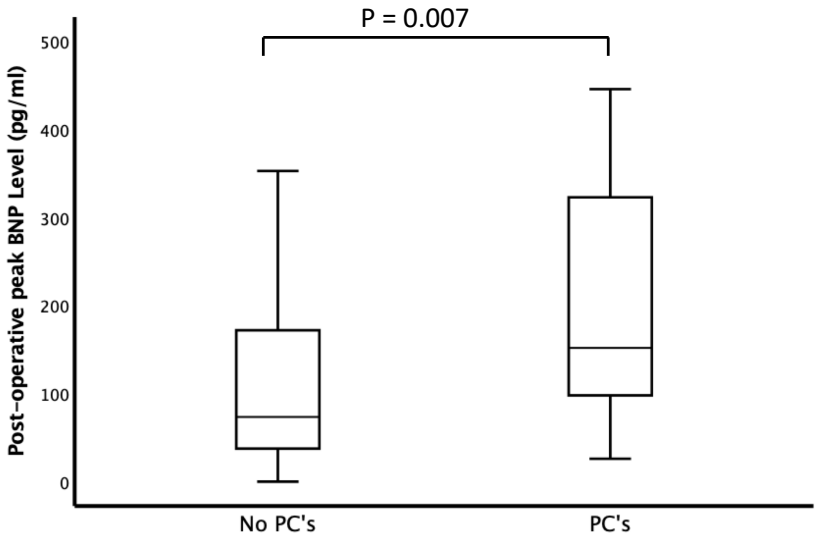

Figure 44 - Post-operative peak BNP (pg/ml) and pulmonary complications.
PC = pulmonary complications, BNP = B-Type natriuretic peptide. (n= 15 patients with pulmonary complications.) (n=93)

12.4.3 Association between peri-operative B-type natriuretic peptide and length of hospital stay

Given those who encounter post-operative complications may require prolonged care, an association between peri-operative BNP and length of hospital stay was sought. There was no association between *pre-operative* BNP and length of hospital stay in days, (r =0.10 p=0.32, spearman's rank test). However, there was a positive correlation between *post-operative peak* BNP and length of hospital stay in days (r= 0.27, p=0.01, Figure 45).

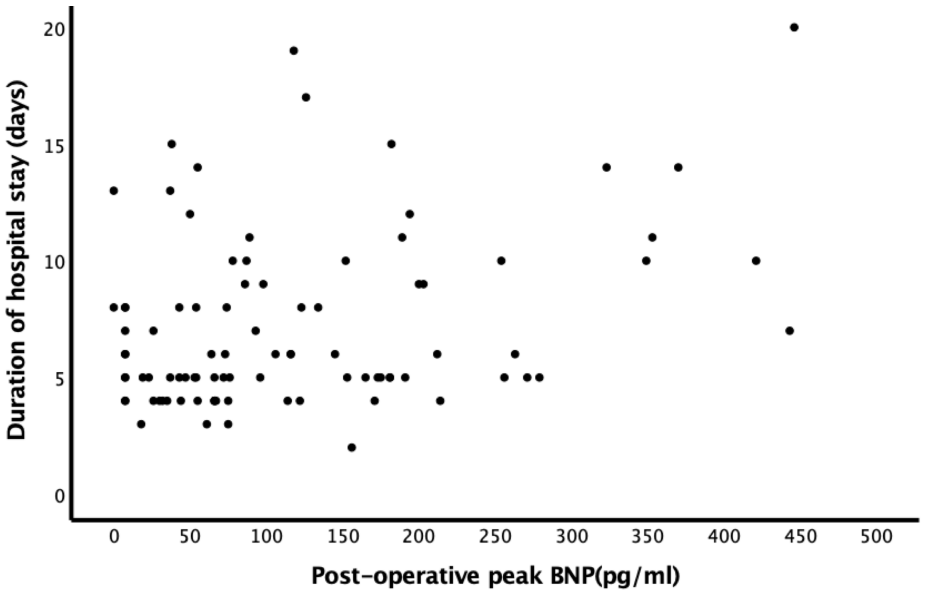

Figure 45 – Association between post-operative peak BNP and duration of hospital stay. (r=0.27, p=0.01, spearman's rank order). BNP = B-Type natriuretic peptide. (n= 91).

12.5 Post-operative quality of life and disability by primary outcome group

12.5.1 European organisation for the research and treatment of cancer quality of life questionnaire

A comparison was performed between EORTC QoL scores and WHO DAS 2.0 scores in those with and without the primary outcome. This comparison was to establish if those who become short of breath report differing *pre-operative* quality of life/disability scores to those who do not have dyspnoea at three months.

Pre-operative EORTC QoL score was lower in those patients who had an MRC>2 at 3-month time point, ($p<0.01$, Mann Whitney-U, Figure 46). This represents a *lower pre-operative* quality of life in those who would go on to report increasing dyspnoea at three months *post-operatively*. Furthermore, 31(41%) patients had a MCID decline in EORTC *sumscore* of >10 points post-operatively. Of these patients, n=17(55%) had a primary outcome of MRC>2 at 3 months post-operatively, ($p<0.01$, chi-squared test).

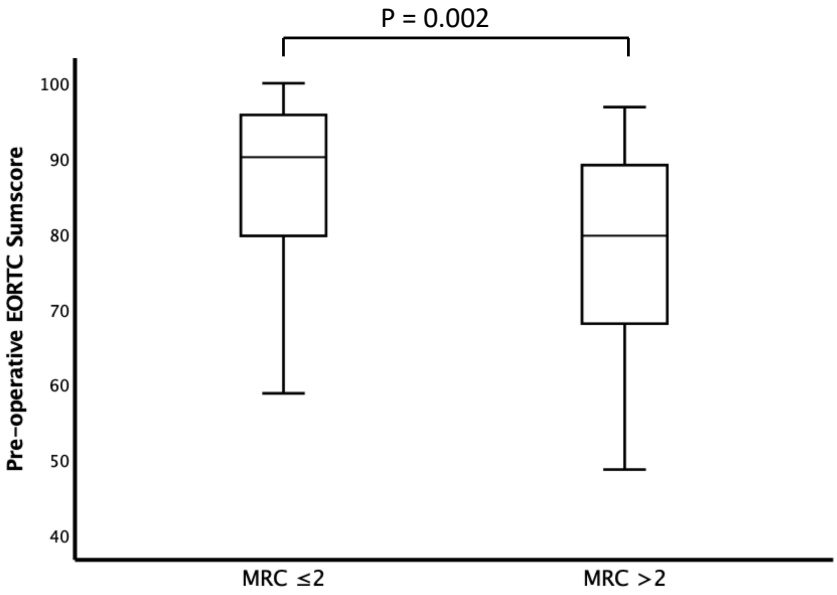

Figure 46 - Pre-operative EORTC sumscore by 3-month MRC score (primary outcome) group. MRC = Medical Research Council dyspnoea score, EORTC = European organisation for the research and treatment of cancer. (p<0.01, Mann Whitney U). (n=75).

A positive correlation existed between *pre-operative* and *post-operative* EORTC *sumscore* suggesting those who reported a poor quality of life pre-operatively were likely to report a poor quality of life following surgery, (Figure 47). Similarly, those with a high quality of life pre-operatively were likely to have a higher quality of life post-operatively.

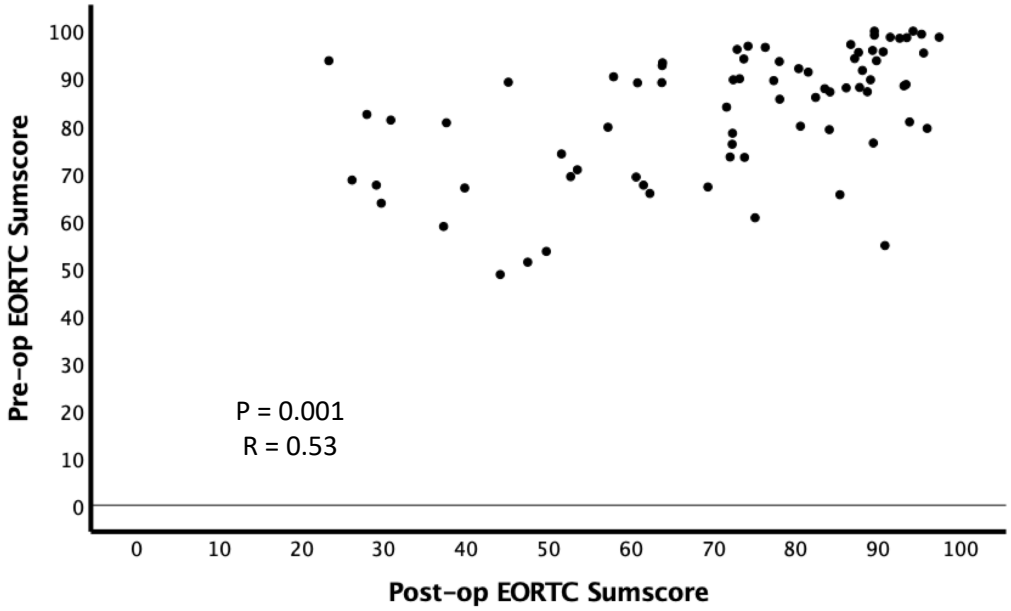

Figure 47 - Correlation between pre-operative and post-operative EORTC sumscore.
EORTC = European organisation for the research and treatment of cancer. Pre-op = Pre-operative, Post-op = Post-operative 3-month time point. (n=75).

As previously mentioned LC-13 lung cancer component of the EORTC questionnaire is not included in *sumscore* calculations and is considered separately in this the analysis. Forty-nine (68%) patients had a decrease in LC-13 score post-operatively compared to pre-operative levels and of these n=18(37%) had a primary outcome of MRC score >2 at 3 months, (p=0.8, chi-squared test).

The WHO DAS 2.0 disability questionnaire has an MCID increase of 5%. Of those patients with a primary outcome of MRC >2 (n=27), 23 (85%) had an MCID increase of >5% in WHO DAS 2.0 score, (p<0.01, chi-squared test). This suggests those with increased dyspnoea at 3 months were more likely to also report increasing disability post-operatively.

12.6 Prediction of quality of life and MRC deterioration

Further analysis was performed to observe if any of the derived risk prediction model(s) could also predict a deterioration in breathlessness (an increase in MRC of one). This is in contrast to the primary outcome, where MRC score was dichotomised into patients who scored >2 at three months post-operatively. The MCID of a deterioration in MRC score is an increase of one. The model(s) were created using the same methodology described in model derivation in chapter 9;

Model 3 is the *'conventional'* model using the same variables as before, Model 5 is the '*next best variable'* model building on model 3 using the next best variable selected at logistic regression and finally '*hypothesis free'* model 7, with no variables used as a baseline.

Given post-operative dyspnoea has an influence on quality of life, *'conventional'* Model 3 and '*hypothesis free' model* 7 were carried forward from derivation/external validation chapter 11 to determine if they are also predictive of a deterioration in QoL, using the *EQ-5DL index score* and *EORTC Sumscore*. As model 5 contained *pre-operative EQ-5DL index score* (a quality of life score) it was not included in quality of life prediction.

12.6.1 Prediction of a deterioration in Medical Research Council score

From 75 patients with paired data for analysis at 3 months, 49(65%) had a MCID deterioration (increase) in MRC score. Results of this regression analysis are summarised in Table 58 and Table 59.

Model	Variables within model	Variable significance (p-value)	Model significance (p-value)	AUROCC (CI 95%)
3	Age Gender ppoFEV$_1$%* ppoDLCO%*	**0.04** 0.73 0.09 0.14	**0.03**	0.70 (0.58-0.82)
5	Age Gender ppoFEV$_1$%* ppoDLCO%* Diabetes	**0.03** 0.96 0.08 **0.05** 0.09	**0.01**	0.72 (0.60-0.84)
7	Age BMI	**0.01** **0.04**	**<0.01**	0.73 (0.62-0.85)

Table 58 - MRC deterioration: summary of results
Model 3 = *'conventional',* Model 5 = '*next best variable',* Model 7 = '*hypothesis free',* AUROCC = Area under receiver operator characteristic curve, Pre-op = Pre-operative, CI = Confidence interval, ppoFEV$_1$ = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BNP = B-Type natriuretic peptide. Significant results in bold. * = Linear variable.

Model	Correct (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
3	67	80	46	70	59
5	70	83	48	72	63
7	65	77	44	69	55

Table 59 - MRC Deterioration: sensitivity and specificity summary
Model 3 = *'conventional',* Model 5 = '*next best variable',* Model 7 = '*hypothesis free'.* PPV = Positive predictive value, NPV = Negative predictive value.

According to the Hosmer et al[300], *'conventional'* model 3 displayed acceptable discrimination in predicting a deterioration in shortness of breath, (p=0.03). Model 3 represents conventional risk prediction, using age, gender, ppoFEV_1% and ppoDLCO%, with age being the most significant value.

'*Next best variable'* model 5 displayed acceptable discrimination in predicting a deterioration in shortness of breath, (p=0.01). Diabetes status was selected as the next best variable to improve the predictive strength of age, gender, ppoFEV_1% and ppoDLCO%. Age and ppoDLCO% were both significant variables within the model.

Finally, '*hypothesis free'* model 7 displayed acceptable discrimination in predicting a deterioration in shortness of breath at 3 months post-operatively. Model 7 represents a new risk prediction model, containing only age and BMI to predict a deterioration in shortness of breath at 3 months. Both age and BMI are significant within the model.

12.6.2 Prediction of quality of life (EQ-5DL index value)

Table 60 summaries the results of QoL (EQ-5DL index value) prediction using models three and seven. Seventy-five patients had paired data for analysis in the internal dataset; of these, thirty-nine (52%) patients had a MCID deterioration in QoL using the EQ-5DL questionnaire (index value).

Model	Variables within model	Variable significance (p-value)	Model significance (p-value)	AUROCC (95% CI)
3	Age Gender ppoFEV$_1$%***** ppoDLCO%*****	0.55 0.47 0.80 0.32	0.7	0.59 (0.45-0.72)
7	ppoFEV$_1$% BMI Pre-op BPI Diabetes	0.9 0.5 0.23 0.76	0.7	0.61 (0.48-0.74)

Table 60 - Quality of life prediction: MCID in EQ-5DL questionnaire
Model 3 = *'conventional',* Model 7 = '*hypothesis free',* AUROCC = Area under receiver operator characteristic curve, Pre-op = Pre-operative, CI = Confidence interval, ppoFEV$_1$ = Post-operative predicted forced expiratory volume in one second, ppoDLCO = Post-operative predicted carbon monoxide diffusing capacity, BNP =B-Type natriuretic peptide, BMI = Body mass index, BPI = brief pain inventory pain score. * = Linear variable

According to the Hosmer et al[300], model 3 and model 7 displayed poor discrimination when predicting quality of life at 3 months. Neither of the models were significant at analysis and neither were any of the variables within either of the models.

12.6.3 Prediction of quality of life (EORTC sumscore)

Table 61 summaries the results of QoL (EORTC *sumscore*) prediction using models three and seven. Seventy-five patients had paired data for analysis in the internal dataset; of these, thirty-one (41%) patients had a MCID decline in QoL using the *EORTC* questionnaire (*sumscore)* of >10 points post-operatively.

Model	Variables within Model	Variable significance (p-value)	Model significance (p-value)	AUROCC (95% CI)
3	Age Gender ppoFEV$_1$%***** ppoDLCO%*****	0.37 0.82 0.11 0.66	0.10	0.67 (0.55-0.80)
7	ppoFEV$_1$% BMI Pre-op BPI Diabetes	<0.01 0.52 0.36 0.72	0.08	0.7 (0.57-0.82)

Table 61 – Quality of life prediction: MCID in EORTC *sumscore*.
Model 3 = *'conventional',* Model 7 = '*hypothesis free',* AUROCC = Area under receiver operator characteristic curve, Pre-op = Pre-operative, CI = Confidence interval, ppoFEV$_1$ = Predicted post-operative forced expiratory volume in one second, ppoDLCO = Predicted post-operative carbon monoxide diffusing capacity, BNP = B-Type natriuretic peptide, BPI = brief pain inventory pain score, *=Linear variable

According to the Hosmer et al[300], '*conventional'* model 3 displayed poor discrimination when predicting quality of life at 3 months. Model 3 was not significant and none of the variables within the model were significant. '*Hypothesis free'* model 7 displayed acceptable discrimination with an AUROCC of 0.7, however the model was not significant, (p=0.08). Interestingly, ppoFEV_1% was a significant variable within the model which *may* signify an association between lower ppoFEV_1% and reduced post-operative quality of life.

12.7 Concordance between scoring tools used to measure dyspnoea

Seventy-five patients had data for analysis of concordance between scoring tools used to measure dyspnoea. Patients in this study reported breathless following surgery at 3 months post-operatively using the MRC scale, (section 8.3). No validated scoring tool (nor MCID) exist to quantify dyspnoea in the lung cancer population despite strong evidence this population have a high incidence of breathlessness.[301] Another scoring tool used to measure dyspnoea is the University of California and San Diego shortness of breath questionnaire, outlined in section 2.1.4. Using the UCSD-SOBQ it was also observed patients had increasing dyspnoea following surgery, (section 8.3).

The aim of this secondary analysis was to compare the tools used to measure dyspnoea and assess the concordance between them. A four-quadrant plot was created and direction of change analysis performed for concordance. Concordance was evaluated using the Landis and Koch[302] grading system; 0-20% (No agreement), 21-40% (fair agreement),41-60 (moderate), 61-80% (substantial), 81-100% (almost perfect). The MCID of the MRC scoring tool is one and the MCID of the UCSD-SOBQ is five.

Forty-six (61%) patients reported an MCID change in MRC score post-operatively compared to 61 patients (81%) reporting a change in UCSD-SOBQ post-operatively, (p=0.02, chi-squared test). Concordance between these scoring tools was 59% (Figure 48), reflecting moderate concordance. Arguably, the broad grading categories of the MRC scoring tool is a strength of its use, however this may omit clinically important change which may be captured with the use of an alternative scoring tool.

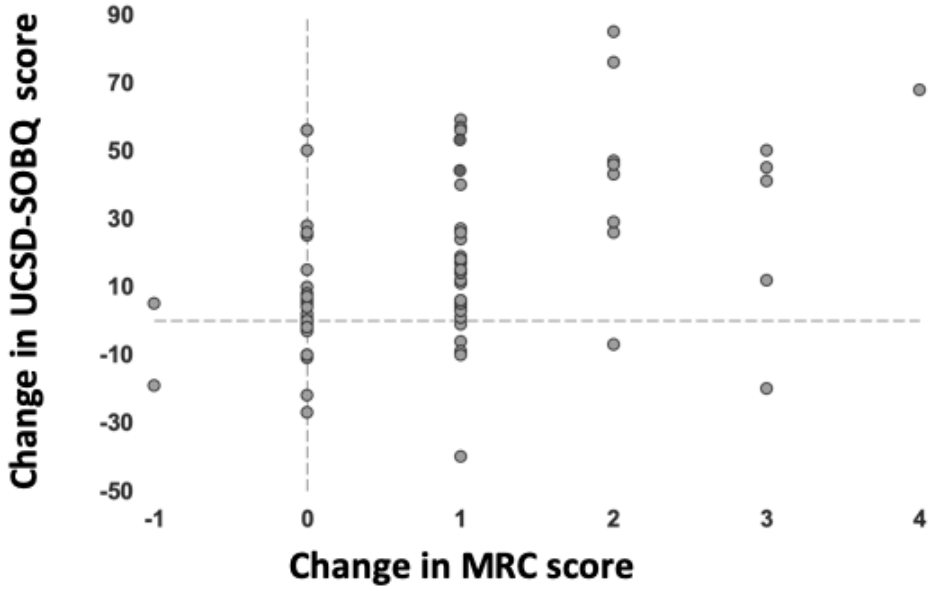

Figure 48 - Concordance between scoring tools used to measure dyspnoea following lung resection for cancer.
Positive values denote a deterioration in dyspnoea. Negative Values denote an improvement in dyspnoea. MRC = Medical Research Council dyspnoea scoring tool. UCSD-SOBQ = University of California and San Diego shortness of breath questionnaire. The MCID of the MRC scoring tool is one and the MCID of the UCSD-SOBQ is five. Concordance was 59% (n=75).

13 Summary of results and findings

Research within this thesis focussed on prediction of dyspnoea following lung resection for cancer, the main findings and results of which are summarised below.

In totality, this study (PROFILES) confirms dyspnoea is common following lung resection, with 27 of 75 (36%) patients in the internal dataset reporting increased shortness of breath postoperatively. The study also establishes that conventional prediction of dyspnoea following lung resection using lung function is poor and highlights the challenges associated with creating a new scoring tool to improve prediction. Contrary to findings observed in a smaller pilot dataset, the addition of the biomarker B-Type natriuretic peptide did not improve the prediction of post-operative dyspnoea. New derived scoring tools performed well within the internal dataset but failed to improve prediction within the external dataset. The biomarker B-Type natriuretic peptide may however have a role in the prediction of post-operative complications or duration of hospital stay. The derived models may also be able to improve the prediction of post-operative quality of life but would require further work to confirm and validate.

The main findings and results of each result chapter will now be discussed;

Chapter 8

Ninety-three patients were recruited into the study from the Golden Jubilee National hospital to test conventional methods and derive a new scoring tool(s) for the prediction of post-operative dyspnoea. Pre-operative demographics would suggest our cohort of patients are similar to other studies exploring risk prediction for lung resection surgery; most were elderly were current or ex-smokers and had significant co-morbidities. The majority of patients reported good functional capacity prior to surgery with most patient's performance status 0 or 1 (86%) or ASA I/II (68%) – this was observed because of the inclusion criteria, not because these findings reflect the surgical population in general. Most patients within the study had minimally invasive surgery (64%), in keeping with current trends within the UK. Seventy-five patients had data for analysis at the 3-month post-operative time point, 27 (36%) of which reported the primary outcome of MRC score >2

reflecting post-operative dyspnoea. Similar to other studies, patients generally reported increasing shortness of breath following surgery; 61% of patients reported an increased MRC score and 81% reported an increased UCSD-SOBQ score post-operatively. Patients also reported an increased level of disability and reduced performance status post-operatively; one quarter of patients had a WHO DAS 2.0 score of >16%, representing moderate disability after surgery and 11% of patients reported a WHO performance status of >2 pre-operatively, increasing to 31% post-operatively. Quality of life, measured using the EQ-5DL and EORTC QLQ-30, also decreased following surgery – with 41% of patients having a peri-operative clinically significant decline in EORTC *sumscore*. Patients also tended to be more depressed following surgery, with HADS D scores increasing from three to five. There existed a significant increase in pain scores at three months post-operatively with BPI scores increasing from a median of five up to fourteen. There was no discernible difference in patient demographics when comparing those lost to follow-up at 3-months (19%) and patients with data for analysis. Finally, BNP was significantly increased post-operatively, peaking on day two and day three, ($p<0.01$).

Chapter 9

This chapter explored the strength of conventional risk prediction models, which are based on estimation of predicted post-operative pulmonary function. .

Conventional risk prediction dichotomises patients using ppoFEV_1<40% and ppoDLCO<40% to identify those who would have post-operative dyspnoea, this method performed poorly in our cohort, (AUROCC 0.52). When age and gender were included in this model, prediction did not improve, (AUROCC 0.55). In model three, when ppo$FEV_1$and ppoDLCO were used as linear variables, this improved prediction to an almost acceptable degree of discrimination, (AUROCC 0.68).

Pre-operative BNP levels were no different between patients with an MRC≤2 vs MRC >2 at 3-months, ($p=0.8$). The addition of pre-operative BNP did not significantly improve the predictive strength of model three (AUROCC 0.69), ($p=0.69$). In summary, no models were strong predictors of dyspnoea at three months.

Chapter 10

This chapter explored and derived new models to improve the prediction of dyspnoea following lung resection. All variables significant at univariate analysis were considered for entry into any new model derived. Despite pre-operative BNP not being significant at univariate analysis (p=0.8) this was included in logistic regression, in keeping with the aim of the thesis.

Using forwards logistic regression, the 'next-best variable' was added to the best conventional model 3 (age, gender, ppoFEV_1/ppoDLCO). The pre-operative quality of life scoring tool EQ-5DL was selected as the next best variable and was significant within this model 5, ($p<0.01$). Model 5 had excellent discrimination and improved conventional risk prediction within the internal dataset, (AUROCC 0.81). When pre-operative BNP was incorporated into model 5, model discrimination did not significantly improve (p=0.5) (AUROCC 0.82) and pre-operative BNP was not significant within the model, (p=0.13).

When all variables were considered, *without* a-priori inclusion of conventional lung function (model 7); ppoFEV_1%, BMI, pre-operative brief pain inventory score and diabetes were all strong predictors of post-operative dyspnoea, (AUROCC 0.83) - ppoFEV_1% was significant within this model, (p=0.04). Pre-operative BNP was not selected to be in this model. Finally, when pre-operative BNP was used as a baseline (model 8) and forwards regression was used to select the next best variables (ppoFEV_1/ppoDLCO, BMI, pre-operative BPI score, pre-operative VAS score and diabetes status), this achieved excellent discrimination (AUROCC 0.85) but did not significantly improve the prediction of model 7, (p=0.8). BNP was not significant within the model, (p=0.1).

Chapter 11

External validation was performed in 85 patients recruited from three other tertiary cardiothoracic centres within Scotland and Northern Ireland. Patient demographics of the patients in whom the models would be applied and be externally validated (3,5 and 7) were compared with no discernible difference. The best conventional model (model 3 – ppoFEV_1% (linear), ppoDLCO% (linear), age and gender) had discrimination approaching acceptable, within the external

dataset, (AUROCC 0.68). Model 5 (age, gender, ppoFEV_1% (linear), ppoDLCO% and EQ-5DL index value) also had poor discrimination within the external dataset, (AUROCC 0.66). Finally, the *hypothesis free* model 7 (ppoFEV_1% (linear), BMI, pre-operative BPI and diabetes status) also performed poorly within the external dataset, (AUROCC 0.62). All models were not statistically significant and had poor discrimination.

Mean calibration suggested *all* models tended to overestimate risk. Model 3 calibration revealed the model tended to overpredict risk in the top two quintiles (40%) of predicted values. Model 5 under predicted at lower values and over predicted at higher values. Model 7 displayed best model calibration; all mean values approaching perfect prediction and all 95% confidence intervals crossing the line of perfect prediction.

Model 7 cut-off was modified to optimise the model for high sensitivity and high negative predictive value – desirable qualities in prognostic enrichment studies. The model performed well in the internal dataset achieving a sensitivity of 93% and NPV of 94%. However, the best sensitivity that could be achieved with model 7 in the external dataset was 59%, with a NPV of 82%.

Chapter 12

Chapter 12 describes the secondary outcome analyses, including acute post-operative complications. Sixteen percent of patients experienced a post-operative pulmonary complication (as defined by the ESTS) and had higher median post-operative peak BNP, ($p<0.01$). Thirty-three percent of patients had a post-operative cardiac complication, also as defined by the ESTS. Post-operative peak BNP level was higher in patients experiencing *any* post-operative complication, as defined by the ESTS, ($p=0.04$). The incidence of new post-operative atrial fibrillation was 11% in our cohort, with increased pre-operative and post-operative peak BNP levels, ($p=0.04$).

There was no difference in peri-operative BNP levels (at all time points) between those with and without the primary outcome, ($p>0.2$). The mean duration of hospital stay was 6 days following lung resection. A positive correlation existed between post-operative BNP and length of hospital stay, ($r = 0.27$, $p= 0.01$),

suggesting patients with raised post-operative BNP levels tended to spend longer in hospital.

Patients had increased dyspnoea scores postoperatively regardless of which scoring tool was used to quantify this, ($p<0.05$). Sixty-one percent of patients reported an MCID change in MRC score post-operatively Vs 81% in UCSD-SOBQ, ($p=0.02$). Pre-operative quality of life was lower in patients who had long-term dyspnoea; *pre-operative* EORTC QoL score was lower in those patients who had an MRC>2 at 3-month time point, ($p<0.01$). Patients with MRC>2 at 3-months also experienced more disability post-operatively with higher WHO DAS 2.0 scores, ($p<0.01$).

Derived '*conventional'* model 3 and '*hypothesis free'* model 7 were tested to determine if they were predictive of a post-operative deterioration in quality of life using the EQ-5DL questionnaire, displaying poor discrimination and significance, ($p>0.05$ and AUROCC<0.65 for both). Derived models 3 and 7 were also tested to determine if they were predictive of a post-operative deterioration in quality of life using the EORTC questionnaire. Model 3 had poor discrimination (AUROCC 0.67) and was not significant ($p>0.05$). However, model 5 displayed acceptable discrimination (AUROCC 0.70) but failed to reach statistical significance, ($p=0.08$).

Models 3, 5 and 7 all displayed acceptable discrimination when predicting a deterioration in MRC score (AUROCC >0.7) and were all significant, ($p<0.05$). Model 5 had the best sensitivity (83%) and NPV (63%).

14 Discussion of results and future direction

This chapter will discuss the results of the study, exploring their strengths and limitations.

14.1 Demographics and burden of dyspnoea

The return rate of questionnaires within the study to allow analysis of the primary outcome was 77%, which is in keeping with other studies within our research group. This is a good return rate compared to other published studies which have a similar follow up period.[120] The power analysis for the study was performed based on a 25% non-return rate.

Patients with lung cancer often have multiple underlying co-morbidities, including chronic obstructive pulmonary disease. The demographics of the population in our study was broadly similar to other work in this field. Most patients tending to be current or ex-smokers (82%) and have previous co-morbidity (74%). Eighty-eight percent of patients within the study also had an ASA of either two or three, representative of the degree of co-morbidity that exists within the lung cancer population.

The mean predicted post-operative FEV_1% (70%) and mean predicted post-operative DLCO% (59%) fell above the recommended 40% for operative lung resection – suggesting patients within the study have been appropriately listed for surgery, using current major guidelines. Eighty-eight percent of patients had a baseline performance status of zero or one, again reflecting appropriate selection for surgery, with good pre-operative function. However, some patients with a poor baseline lung function and performance status were intentionally excluded. This would mean the population within this study was different to the overall population undergoing surgery at the GJNH during this period – with some speculative patients sitting outside the guidelines not being included within the study.

Ninety-three percent of patients within the study had a lobectomy or bi-lobectomy, with only a small proportion of patients undergoing complete pneumonectomy (7%). With the reported proportion of patients undergoing

pneumonectomy within the UK falling from 40% in 1981 to 5% in 2015, these figures are similar to other published work within this field.[303] Sixty-four percent of patients underwent minimally invasive thoracic surgery – the majority (55%) of lung cancer surgery now being performed using minimal access approach within Europe.[304]

When the demographics of those patients with and without an MRC score greater than two at the three-month time point were compared, there existed some differences; BMI and weight were higher in patients with increased post-operative dyspnoea scores. This may reflect the fact that increasing weight is associated with reduced baseline functional capacity and a traumatic surgical insult may cause these particular patients to experience increased incidence of shortness of breath post-operatively. Furthermore, it is established that increased BMI is associated with increased incidence and mortality of lung cancer.[305] Increased BMI also increases the risk of peri-operative complications in the lung cancer population [306], which may also contribute to increasing peri-operative dyspnoea.

Patients with increased MRC score and greater dyspnoea burden at three months post-operatively also had lower pre-operative FEV1 (L), FVC (%) and calculated ppoFEV_1% values, showing a possible association between lung function and post-operative dyspnoea levels. Importantly, ppoFEV_1 was an independent predictor of post-operative dyspnoea within *'hypothesis free'* model 7.

ASA scoring was higher in patients reporting increased dyspnoea at three-months, displaying a possible connection between increased co-morbidity and increasing dyspnoea scores. It was surprising that no difference existed in pre-operative performance status between patients who would develop post-operative dyspnoea and those who would not, given performance status is assessed routinely at pre-operative assessment and at MDT prior to listing for surgery. Pre-operative performance status plays a role in the risk prediction of mortality, as discussed in section 4.2.1. Generally, the demographic findings between the two groups that were different were all biologically plausible.

As discussed in section 1.2.4, dyspnoea is a debilitating symptom experienced by roughly half of all patients undergoing lung resection, affecting both functional and psychological health.[40,41] In this study, patients reported an increase in

dyspnoea following surgery regardless of which scoring tool was used to measure shortness of breath. Between 60-80% of patients within the study reported an increase in dyspnoea, similar to other studies reporting post-operative dyspnoea levels.[44,307,45,41]

Within the study, a greater proportion of patients (74%) who reported dyspnoea at three months had a pre-operative MRC dyspnoea score of two, instead of one. This is perhaps intuitive, as any patient with a pre-operative MRC dyspnoea score of one has further to deteriorate following surgery before reporting a post-operative MRC dyspnoea score of greater than two. However, despite reported pre-operative MRC dyspnoea scoring being significantly different between those with and without the primary outcome (MRC score greater than two at 3-month time point), this variable did not get selected at logistic regression to feature in the derived risk prediction scoring tools. To date, no study has reported baseline breathlessness as a risk factor for further deterioration.

Twenty-seven patients (36%) reported the primary outcome of an MRC score greater than two at three months post-operatively. As up to 30-50% of patients report long-term disabling dyspnoea following lung resection surgery[41], approximately one third in this cohort is encouraging - patients undergoing surgery at our centre tending to sit on the lower end of this spectrum. Another study within our research group has reported 40% of patients reporting a post-operative deterioration in dyspnoea.[49] Our results are consistent with this finding. This may be considered as selection bias, but this is not a negative given these are the patients intended to be captured within the study.

14.2 Peri-operative B-Type natriuretic peptide to improve risk prediction of post-operative dyspnoea

The primary aim of this study was to explore whether BNP could improve the prediction of dyspnoea. An increase in BNP levels was observed following lung resection surgery ($p<0.01$), similar to pilot work within our research group.

The results of this study suggest that BNP does not improve the prediction of dyspnoea following lung resection. At univariate analysis, pre-operative BNP level was no different in those with and without dyspnoea at three months. When BNP

was forced into model derivation, it was not significant within the model and did not improve the predictive strength.

There are a number of possible explanations why BNP did not improve prediction of post-operative dyspnoea within our cohort. Pilot work, which this study was based upon, observed an association between increased dyspnoea or reduced functional capacity (decreased distance at 6-minute walk test) and the biomarker BNP in a small cohort of 27 patients. BNP also improved the prediction of post-operative dyspnoea. Functional capacity assessment and/or 6-minute walking testing was not performed within this thesis. This was a pragmatic decision made when faced with recruitment of a large number of patients across multiple hospital sites. The results of the pilot study may also have been by chance (small numbers). Different inclusion and exclusion criteria existed between the two studies – patients with a high pre-operative MRC were intentionally excluded from this thesis. This suggests BNP may have a role in patients falling at the other end of the spectrum – with increasing dyspnoea, reduced performance status and increasing co-morbidity.

Another potential confounder is that the hypothesised underlying mechanism by which patients become short of breath post-operatively, may be incorrect. The concept of post-operative dyspnoea having a cardiovascular component may be wrong – therefore the addition of BNP was unlikely to improve predictive strength. Post-operative dyspnoea is not fully understood and likely to be multifactorial or include other factors not considered within this study. If post-operative dyspnoea does have a cardiovascular component, BNP may be too blunt a tool to detect these changes and subtle interplay of cardiovascular function and afterload.

According to the manufacturer's guidelines, measurement of BNP using *Abbott i-STAT point of care BNP* devices requires accurate timing of sample analysis. Point of care devices were selected to maintain consistency across all hospital sites. A record of quality control was maintained and the devices were externally quality controlled by NEQAS, who have no conflict of interest in the results of this study.

Other studies have identified specific cut-offs for BNP to predict post-operative outcomes, such as onset of new atrial fibrillation.[308] Creation of peri-operative BNP cut-offs may have been incorporated within our study to improve risk

prediction. However, while creating cut-offs makes the usability of risk prediction models favourable, information is often lost when continuous data is not handled in a linear fashion[309] – BNP was used as a continuous variable within the derived models within this study.

14.3 Model derivation and validation

Models 1-4 were designed to be reflective of current clinical risk prediction of post-operative dyspnoea. Model 1 and 2 reflect current risk prediction guidelines, with ppoFEV_1% and ppoDLCO% dichotomised above and below 40%. Age and gender were included in the models (as these should always be considered when estimating clinical risk) in addition to lung function, but failed to improve the predictive performance. Given age and gender feature in the prediction of 'in hospital mortality' with the use of *Thoracoscore*, it was unexpected they were not more influential within these scoring tools. However, age was used within *Thoracoscore* with cut-offs which was different to this study where age was handled as a continuous variable.

'*Conventional'* model 3 included ppoFEV_1% and ppoDLCO% as linear variables, alongside age and gender, and was selected to be taken forward to external validation. Model 4 was '*conventional'* model 3 with the addition of pre-operative BNP, which did not improve the prediction of post-operative dyspnoea. BNP was not significant at univariate analysis and was not therefore expected to improve the prediction of model 3.

Models 5-8 were derived to try and improve risk prediction of post-operative dyspnoea and explore the predictive strength of BNP. Models 5 and 7 were selected as the best models to progress to external validation.

When models 3,5 and 7 were compared within the internal dataset, '*hypothesis free'* model 7 had the best predictive discrimination, with an AUROCC of 0.83 (95%CI 0.74-0.92). Furthermore, when *conventional* model 3 was compared to models 5 and model 7, only model 7 had a significant increase in risk prediction, ($p=0.03$). Model 7 contained: ppoFEV_1 as a linear variable, BMI, Pre-operative BPI score and diabetes status. All of these variables would be easy to collect or calculate in a clinical setting. It is encouraging to see lung function was still be

selected into the model, albeit being used as a linear variable without cut-offs. Lung function has traditionally been used for risk prediction in this population and though despite having a questionable evidence base, lung function can't not have a role in post-operative dyspnoea. Validating any novel scoring tool without the inclusion of lung function would be very challenging as many clinicians would be reluctant to use a risk prediction calculator that did not incorporate ppoFEV_1% and/or ppoDLCO%.

It may seem intuitive that BMI should affect dyspnoea grade as body habitus can have a large influence on exercise tolerance and functional capacity. In our cohort, those with dyspnoea at three-months had a higher BMI (29.8) when compared to those without (26.7), (p=0.03). Despite this being statistically significantly different, clinically this modest difference in BMI is unlikely to represent any meaningful difference. Furthermore, this is at odds with most of the lung cancer literature, where patients with a low BMI have poorer outcome.[310] Most patients with advanced lung cancer are cachectic with a low BMI, post diagnosis weight loss potentially being a prognostic factor for survival and global quality of life.[311] Our patients have been selected for having good functional capacity and this may explain the higher than normal BMI in both groups - the cancer being recognised early with minimal progression of disease.

Pre-operative pain score and diabetes status are less obvious variables to be selected into the model. An increased pre-operative pain score having an influence on post-operative dyspnoea could be because these patients may have chronic pain due to the lung cancer disease process or other significant co-morbidity. These patients may therefore be vulnerable to any surgical insult causing a worsening of long term-symptoms with the chest wall becoming affected – chronic pain is a complex disease process, much of which remains unexplained. Pre-operative diabetes status may be linked to increased post-operative dyspnoea with the increased levels of cardiovascular co-morbidity observed in patients with poorly controlled diabetes. Diabetes could also be simply serving as a surrogate of otherwise unmeasured factors such as; all cardiovascular risk factors and the metabolic syndrome. Diabetes has also been associated with an increase in prevalence of respiratory symptoms as compared to the general population.[312] The mechanism behind this is unclear but thought to involve a faster lung ageing process. One half of patients with COPD have co-existing metabolic syndrome

which may explain why both of these factors have been selected; as the two conditions involve low grade inflammation, they have recently been collectively referred to as 'chronic systemic inflammatory syndrome'.[313, 314] It has been demonstrated that the risk of some cancers is higher in the diabetic population, but this is not the case for lung cancer.[315] However, it has been suggested that pre-existing diabetes is associated with a worse survival rate in those with newly diagnosed lung cancer, particularly females.[316]

While the derived models performed well within the internal validation dataset, the results were not reproducible in the external dataset. This may have occurred for a number of reasons. With many different statistical methods available to create a risk prediction model perhaps selection of an alternative technique may have had more positive findings - forwards logistic regression was used to derive the risk prediction models. Linear regression (using a continuous primary outcome), cox regression (considering the risk of dyspnoea with increasing post-operative time) and machine learning are all other options which may have had different results. Machine learning is a relatively new method of risk prediction in medical research – situations which are suited to machine learning are those where researchers are only interested in classification of conditions without wishing to draw conclusions from individual variables.[317] While machine learning may be useful, a very large data set is commonly needed.

Models developed using data with few positive events compared with the number of variables often underperform when applied to a new patient cohort.[318] The main reason for this is 'model overfitting' with overfitted models tending to underestimate the probability of an event in low risk patients and overestimate in high risk patients – the smaller the positive event ratio, the more overfitted the model will be.[319] In the current study twenty-seven patients had a 'positive event' of an MRC dyspnoea score of >2 at three months post-operatively. There exists no firm rule for the number of positive events per variable ratio when creating a scoring tool, but it has been suggested by some authors that is should be at least ten – for example a dataset should contain at least 70 events to fit a risk model with seven regression co-efficients.[319] Other authors have shown adequate model performance when the positive events to variable ratio falls short of this.[320] Model 3 had four variables, based on which, ideally should have 40 positive events within the dataset, falling short of this 'rule-of-thumb'. Given the pattern we observed

in all models (3,5 and 7) was one of underestimation at low risk patients and an even more pronounced overestimation in high risk patients, it seems reasonable model overfitting is the most obvious reason for poor external validation results.

In an attempt to reduce model overfitting, candidate predictors were carefully selected from univariate analysis based on the number of positive events. Given model overfitting is the most likely cause for poor results at external validation, two further strategies could be used to minimise or remove the underestimation (model overfitting); increasing the sample size and simultaneously increasing the positive event rate is the most obvious solution. Alternatively, more complex methods can alleviate the problem of overfitting, such as 'shrinkage' methods. 'Shrinkage' moves the regression coefficients towards zero, moving poorly calibrated predicted risks towards the average risk. This is commonly done by shrinking the regression coefficients by a factor, for example 30%. 'Penalised regression' is the most common shrinkage approach when the positive event rate is low – placing a constraint on the values of the regression co-efficient. Several penalised methods have been described that use different constraints[321], which are complex and have not been carried out within the risk prediction models described within this thesis.

The absence of an important predictor within the models is another obvious reason why the risk prediction models may not have performed as well as expected. The most obvious suggestion is that intra-operative variables not considered for inclusion into our risk prediction models may be independent predictors. This was intentional, as the purpose of this study was to improve prediction of dyspnoea prior to surgery. However, some intra-operative events or variables may by the key to understanding the mechanism of dyspnoea and explaining the difference between the internal and external centres within the study. Given a possible explanation for post-operative dyspnoea is RV dysfunction (Section 3.2.2), intra-operative factors include; duration of anaesthesia, single lung ventilation time, main pulmonary artery clamping occurrence/duration and size of surgical lung resection. All of these may have the potential to trigger RV dysfunction in a vulnerable ventricle (with recognition the mechanism leading to RV is complex and not fully understood).

Difference in methods of measurement is another common reason why risk prediction models can fail.[322] In this study, every attempt was made to standardise and therefore minimise any differences in measurement. Selection of well validated, simple and easy to use questionnaires was another step taken to decrease variability in completion. Arguably, selection of the MRC dyspnoea scoring tool may not have been the best to quantify or detect long-term dyspnoea and the University of California and San Diego Shortness of Breath Questionnaire may have been better – with a more granular and detailed account of the impact of surgery on this population and the post-operative dyspnoea experienced. However, previous work from within our research group (including the pilot work upon which this study was based) and patient public involvement all supported use of the MRC dyspnoea score.

A change in clinical setting results in a different case mix which can commonly affect the usability of prognostic models.[323] That said, all centres within the study were tertiary cardiothoracic centres performing large numbers of lung resections per annum with equally qualified surgeons/anaesthetists. Patient demographics were compared between centres with no discernible difference, making this an unlikely cause. While all centres within the study are tertiary cardiothoracic centres, the Golden Jubilee National Hospital is a higher volume centre than the rest – with a higher number of more complex patients undergoing surgery. With surgeons performing an increased number of operations, it is plausible this is driving improved outcome within this centre with patients experiencing less post-operative dyspnoea – for example reduced operative time because of very experienced surgeons and anaesthetists. Alternatively, some centres within the study may be risk averse in their selection of patients undergoing surgery – those patients who have borderline lung function not being considered for surgical lung resection. Finally, the other centres within the study may be excellent at patient selection, considering some other factor when making decisions about operative intervention, such as age or type and severity of other co-morbidities which may have an effect on functional outcome. Although, with most centres within the UK strictly following national guidelines, this seems unlikely.

Usability of a new prognostic model also is a major factor that can limit its clinical application. A novel prognostic model must also be acceptable to clinicians and be easy to use. Unambiguous definitions of predictors and reproducible

measurements using methods readily available are crucial in the success of a new prognostic model.[323] All models created within this study are intuitive, with easily obtained and reproducible variables. Also, the models are not complex involving too many variables, which can lead to error, frustration and reduced usability.

When a prognostic model performs poorly in external validation, adjusting the model using the new data can be considered to improve its performance. [323] This process is called *updating*. There are several methods to update risk prediction models; each with strengths and weaknesses.[324] Intercept recalibration is the simplest form, using a recalibration factor to correct the average of all predictions. However, as the predictions of a logistic regression model are not linear, this is not a statistically adequate method and does not improve discrimination. Logistic recalibration can also be used in small data sets, re-estimating the model intercept and calibration slope. This also does not however improve model discrimination. A third method is model revision: all individual risk factor effects are updated. However, this approach requires an extensive dataset, more than in PROFILES study.

A final simple method well described in the literature to update risk prediction models is to add new risk factors. This would generally apply and work best in models which have been in existence for many years, when over time new variables have been discovered to be valuable. Given the models within this thesis are new risk prediction models, this method of updating would not improve risk prediction. Furthermore, if any variable were of help they would have been selected as part of the regression analysis.

Another possible approach that may improve the prediction of post-operative dyspnoea would be to merge all internal and external data, before deriving further risk prediction models. New models would then be based on both the development and validated data, improving stability and generalisability, and increasing power. As the methodology was finalised before the commencement of this study (section 6), this has not been undertaken. Furthermore, internal validation before proceeding to external validation is a well-recognised technique to create risk prediction models and is a strength of the current work- [325] combining the data would not be in keeping with this approach.

An *impact study* may be undertaken to quantify whether the use of the prognostic model in daily practice improves decision making and patient outcome. This is something that could be done in the future. Prognostic models do not require an impact study to be carried out before use in clinical practice. Instead, this will be determined by the acceptable rate of false positive and false negative predictions and the consequences these erroneous results have for patient management and outcome.[323] If an impact study is not possible, an intermediate step using decision modelling techniques or *Markov chain* can be helpful.[326] Markov chains are used not only in medicine to improve risk prediction models but in many other professions where a scoring tool is used to forecast results.[327] The analysis helps evaluate potential consequences of using the model, subsequent therapeutic decisions and eventual patient outcome. If a Markov model did not observe any improved outcome a formal impact study would then be required.

Determining if the predictive accuracy of a new model in a population is adequate is also a matter of clinical judgement and depends upon what alternatives are available.[323] From the literature, it is unclear if experience and clinical judgement are as good as risk prediction models or guidelines to predict long term dyspnoea in lung cancer. No study from the review of the literature created a risk prediction scoring tool with clinical judgement built into the model. Similarly, no study has tested a new risk prediction model against clinical judgement or alongside clinical judgement to explore if predictive strength is improved with/without a novel scoring tool.

This study has confirmed prediction of dyspnoea following lung resection is challenging. It has also illustrated the difficulty of creating a new risk prediction model. Creating any risk prediction model is very complex and this becomes even more intricate when the mechanism behind post-operative dyspnoea is not fully understood. Dyspnoea is a very subjective symptom, meaning different things to different patients: making prediction of long-term dyspnoea even more complicated. In time, with future work in this area, hopefully prediction of dyspnoea will improve.

14.4 External validation

When the models were calibrated within the external dataset there were a few discussion points. *Conventional* model 3 predicted the first 70% of patients with reasonable degree of accuracy with all decile points lying close to the 'perfect prediction' line (section 11.3.3, Figure 35). The final 30% of patients were below the line of 'perfect prediction', diverging in this direction – patients at highest observed risk being decreasingly predicted at a lower risk value. This suggests a factor(s) which leads to increased predicted risk is actually proving to be protective. When the calibration curve was drawn using quintiles, this protective trend is increasingly obvious, (Figure 36). The external centres are perhaps selecting and risk stratifying patients using a different technique to the GJNH base centre – although this seems unlikely given most centres within the UK follow the discussed pre-operative risk assessment strategy (Section 4). Patient demographics were compared with no differences observed between the variables measured. There could however be some other variable that could account for this 'protection' in the high-risk patients. External centres may be conservative, offering surgery to patients who are low risk, with those patients who fall below the 40% value being assessed further using functional techniques before being turned away for surgery. At the Golden Jubilee National Hospital few patients (if any) are offered pre-operative CPET testing, to aid the operative decision-making process. In the derivation dataset, worsening lung function was associated with increased risk of dyspnoea. However, if the external centres were only selecting a minority of the very best of patients with poor lung function in the validation dataset, then poor lung function could in fact paradoxically be a marker of good patient outcome. This hypothesis is supported by the observation that a greater proportion of patients with poor lung function underwent surgical lung resection in the derivation dataset - eleven patients (15%) had ppoFEV1% and/or ppoDLCO% falling below 40% value who underwent surgery at the Golden Jubilee National Hospital. External sites had less (n=5,6%) patients with ppoFEV1% and/or ppoDLCO% falling below 40% who underwent lung resection. Small numbers of patients at the extremes of poor lung function may also be contributing to less accurate prediction by the models.

Model 5 (*conventional model 3 and next best variable (EQ-5DL)*) also underpredicted risk within 60% of the highest risk patients within the external

dataset (Figure 38). Despite the top 3 quintiles falling below the line of perfect prediction, the lines did not diverge in the highest risk patients, as observed in *conventional* model 3 (discussed above). The EQ-5DL QoL questionnaire was very influential within this model and was most significant predictor, ($p<0.01$). Patients from the external sites may have a higher global quality of life and this would explain the underestimation of risk displayed within model 5. The additional of the EQ-5DL QoL also prevented the divergence of lines (protective effect at highest risk) observed at external validation of model 3.

Upon visual inspection, model 7 (*'hypothesis free'*) appeared to display the best calibration within the external dataset (Figure 40). The highest risk patients (top 3 quintiles) remain underestimated, similar to model 3 and model 5, but sit much closer to the line of perfect prediction, with the line connecting the quintiles being 'non-divergent'.

Finally, age and gender were used in the baseline, *conventional* model in a linear fashion but perhaps are not as influential for risk prediction as believed. Other risk prediction model use cut-offs for age, such as *Thoracoscore* (Section 4.2.1), with patients falling above 65 years at increased risk. Being male increased risk within '*conventional'* model 3. However, age and gender not being selected in '*hypothesis free'* model 7 may suggest these factors are not as important as believed or were not handled in the correct fashion to maximise potential predictive value.

14.5 Deterioration in MRC score

As the minimally clinical important difference (MCID) of the MRC scoring tool is a change of one (i.e. a change that is meaningful to the patient), further analysis was performed to test the derived model's strength to predict a *deterioration* of post-operative MRC score, (Section 12.6.1). Following review of the literature, this is the first study to undertake such an analysis and have positive findings.

Conventional models (including age, gender, ppoFEV_1% linear and ppoDLCO% linear) were significant (all $p<0.05$) with acceptable discrimination (all AUROCC >0.7) for determining a *deterioration* of peri-operative MRC score. These were improved results compared to when the conventional models were used to predict

a post-operative MRC score of greater than two at three months (primary outcome). This ad-hoc analysis shows a potential for 'MRC deterioration' to be used in future studies as the primary outcome - predicting those who are most vulnerable to become *more* breathless post-operatively.

Arguably however, a deterioration in dyspnoea illustrated by moving from an MRC score of one to an MRC score of two may not represent someone who has 'disabling' post-operative dyspnoea, which is the cohort of patients prognostic enrichment aims to capture.

14.6 Peri-operative quality of life and disability

It is well established that patient reported quality of life (QoL) is reduced following lung resection surgery, more so than other chronic illnesses and cancers.[18] The patients within our study also reported a decline in QoL three months post-operatively, regardless of which quality of life scoring tool was used. Global QoL, measured using the EQ-5DL index score, displayed a reduction in overall quality of life. When each EQ-5DL domain was individually scrutinised, they all displayed a deterioration in QoL apart from the domain '*anxiety and depression',* which remained unchanged. Arguably, the anxiety and depression levels should have reduced following lung resection within our population given many patients have been offered the opportunity of definitive treatment.

For post-operative anxiety and depression scores to remain unchanged from pre-operative scores, perhaps suggests ongoing anxiety and depression from already higher than average levels. When anxiety and depression was analysed further using the 'HADS' scoring tool, it would seem patients indeed remain as anxious following their operation but experienced increased depression scores. Increased post-operative depression may reflect an increased drop observed in reported post-operative QoL due to increasing post-operative dyspnoea. There was also reduced performance status reported by patients post-operatively which could also account for the increased levels of depression. Patients had increased levels of pain post-operatively as well which may contribute to the increasing post-operative depression experienced by this cohort. The increased post-operative depression is likely to be multi-factorial in origin and may also incorporate some

other factor we did not explore- such as frailty, chemotherapy, radiotherapy, recurrence and unfavourable pathology results.

At univariate analysis, patients who had an MRC>2 at three months reported higher levels of disability and reduced quality of life – this was to be expected. This was a global decline in quality of life, across all domains, and in both QoL scoring tools. An understanding of the mechanism driving reduced post-operative QoL may allow us to decrease or stop the decline in QoL and improve patient satisfaction.

14.7 Prediction of quality of life using derived scoring tools

In section 11.5.1, '*conventional'* model 3 and '*next best variable'* model 5 were used to try and predict quality of life assessed by EQ-5DL (*index value*) and EORTC (*sumscore*). As dyspnoea can influence post-operative quality of life[97], it seemed reasonable to perform this analysis. Positive results would also offer additional validity to the derived models if they could predict other factors (such as QoL), perceived to be associated with dyspnoea. The scoring tools generally had poor discrimination which failed to reach clinical significance. Of note, '*hypothesis free'* model 7 was able to predict a significant reduction in quality of life (EORTC scoring tools – *sumscore*) with an AUROCC of 0.7 (95%CI 0.57-0.80), ppoFEV_1% being independently predictive within this model, ($p<0.01$). While this falls short of what is often deemed to be clinically useful,[300] this displays potential for future exploration. The variables within the model (ppoFEV_1%, BMI, pre-operative pain and diabetes status) are all variables which may reasonably contribute to global quality of life.

Few studies have explored the prediction of QoL following lung resection. Those that have, reported ppoFEV_1% and peri-operative pain levels as components and influencers of post-operative QoL.[54]

14.8 Secondary outcomes - complications

The incidence of new post-operative atrial fibrillation within our study was 11%, which is lower than the reported national (UK) incidence of 20-50%.[308] Sixteen percent of patients developed post-operative pulmonary complications within our

cohort, which falls within other reported ranges of 14%-40%.[328, 329] The incidence of cardiopulmonary complications within our cohort was 47%, which is higher than the reported 18.5% in the European Society of Thoracic Surgery (ESTS) registry.[304] However, when observing patients more closely in a study environment, it is likely that more complications will be captured than when performing a routine clinical audit.

There was no mortality in our group which is lower than 2.6% nationally, reported in the ESTS registry.[304] This could however be easily explained due to the small sample size. Furthermore, no patient within the study developed renal failure, again lower than 7.2% reported by the ESTS.[304] This could be as a result of small sample size or because of the selected definition of renal failure, as per ESTS: new onset renal failure with an increase in serum creatinine to two times the baseline or new requirement for dialysis post-operatively. The incidence of peri-operative renal failure varies according to the sensitivity of the definition used[330] and other studies may use definitions with less of an increase in creatinine required.

Median post-operative peak BNP level was higher in those with more than one complication within this study. This may be because the majority of patients experiencing complications had cardiopulmonary complications and therefore a greater increase in BNP post-operatively, may indicate a cardiac mechanism contributing to this.[304] Pre-operative BNP was higher in those with new post-operative atrial fibrillation and a positive correlation also existed between post-operative peak BNP and length of hospital stay. These observations highlight the potential of BNP as a predictive biomarker in lung resection surgery. As discussed in section 5, BNP is used in other high-risk non-cardiac surgical groups to determine cardiac risk, it would therefore seem reasonable to explore its predictive strength in the lung cancer population. Within these guidelines, other biomarkers are also referenced, such as high sensitivity troponin, which was not measured within this study.

14.9 What does this research add to existing literature?

The findings within this thesis add to the existing literature surrounding the prediction of dyspnoea in the lung resection population. This work has

demonstrated the need for, and challenges in creating, a validated scoring tool to predict dyspnoea in patients with lung cancer undergoing lung resection surgery. To date, most work has observed association between lung function, exercise testing and morbidity and mortality following lung resection, with few authors having attempted to predict long term dyspnoea. The study adds to the increasing body of evidence that dyspnoea is both common following lung resection and has an impact on quality of life. Regardless of which scoring tool was used to measure dyspnoea, a deterioration was observed post-operatively. Disability and QoL deteriorated following surgery, in all domains, apart from anxiety and depression – this may secondary to potentially curative surgery being performed.

The study adds to the evidence that the 40% predicted post-operative cut-off values for pulmonary function, used in guidelines to stratify risk, are poor predictors of dyspnoea. Forty percent may not be the optimal cut-off and the results of this thesis support the use of pre-operative lung function as a continuous variable to predict post-operative dyspnoea. This study is also one of the first studies to include age and gender into risk prediction models for lung cancer resection. It is one of the first studies to look for other predictors of dyspnoea other than lung function: the 'best' model 7 included BMI, pre-operative pain levels and diabetes status in addition to ppoFEV_1%. Pre-operative pain levels being very significant within the model. This was a new finding and is the first study to report a link between post-operative dyspnoea and pre-operative pain levels. Furthermore, the post-operative outcomes detailing dyspnoea, quality of life and performance status could be invaluable to inform patients prior to surgery in the shared decision-making process. Lung function testing does have an association with post-operative outcomes when used as a continuous variable. There existed an association between increasing post-operative dyspnoea burden and pre-operative quality of life, performance status and ASA score which has not been shown before. These findings may be an invaluable tool in the pre-operative shared decision-making process.

When sensitivity was changed within '*hypothesis free'* model 7 to achieve a high sensitivity and high NPV in the external dataset it was not possible to achieve 90% sensitivity, however a NPV of 82% would be acceptable for a low-risk intervention: if high-risk was predicted by the score this could be used as inclusion criteria in

future studies. This is the first study to publish a clinically useful scoring tool to quantify risk of long-term dyspnoea.

The best *new* model 7 performed well in the internal derivation dataset and expectedly less well in the external validation dataset, achieving an AUROCC value of 0.62 (95% CI 0.48 – 0.77). The AUROCC for prognostic models is typically between 0.60 and 0.85.[331] Values for AUROCC are generally criticised for an inability to detect meaningful differences.[331] Therefore, Brier skill score and net reclassification indexing were both used and confirmed the discriminative abilities of the new models.

By comparison, '*conventional'* model 3 performed poorly within the internal derivation dataset with similar values when applied to the external validation dataset, achieving an AUROCC value of 0.68 (95% CI 0.55 – 0.80) – however this was better than model 7 detailed above. Variables within this dataset included age, gender, ppoFEV_1% and ppoDLCO% which are routinely collected during conventional risk prediction of dyspnoea. Post-operative predicted FEV_1% and ppoDLCO% were used as linear variables within this model, instead of the traditional cut off values of above or below 40% deemed high risk and warranting further investigation.

Improved prediction of post-operative dyspnoea and implementation of novel scoring tool could benefit patients by:

- Screening for entry into a prognostic enrichment study aiming to ameliorate post-operative dyspnoea.

- Better informing them when making decisions about choosing to undergo surgery.

- Providing access to potentially curative surgery in those previously deemed too high-risk.

- Allowing development of novel targeted preventative therapies in those at most risk.

Finally, an accurate prognostic model is of no benefit if it is not generalisable or doesn't change behaviours.[323] All variables for consideration in the models are easy to collect and would not utilise additional resource. They are basic measurements which are already routinely collected at pre-operative assessment. As this scoring tool will not be used in clinical practice (nor is it designed to be so used) it is not possible to determine if it would change clinical decision-making behaviours.

14.10 Strengths and limitations

A strength of this study is that it is the only reported study exploring the predictive value of the biomarker BNP for long term dyspnoea following lung resection surgery. The study is one of the largest reported datasets looking at the prediction of dyspnoea following lung resection, incorporating a large number of patients across multiple cardiothoracic centres within the UK. Despite pre-operative BNP not improving current risk prediction strategies, this study confirms association between BNP and both the incidence of post-operative complications and hospital stay. Following lung resection, peri-operative BNP levels have previously been associated with post-operative complications, including AF.[154, 269, 270, 332, 333]

The use of two scoring tools to measure peri-operative dyspnoea was a strength of the study. By using two tools which have different strengths and weakness' we could confirm with more certainty the findings about dyspnoea burden in this population. The MRC grading system has several advantages; its ease of use and broad grading categories make its use intuitive with both clinicians and patients. In addition, it is one of the few dyspnoea scoring tools with a valid MCID. The broad grading category however is also a potential weakness of the MRC scoring tool and therefore the UCSD-SOBQ was also selected to quantify pre- and post-operative dyspnoea within the study. The UCSD-SOBQ has a valid MCID but is not as easy to complete, with substantially more questions and grading categories. Despite their popularity, neither the MRC score or UCSD-SOBQ are validated specifically in the lung cancer population.

The study observed pulmonary function is *associated* with, but not *predictive* of long-term dyspnoea following lung resection surgery. This is similar to other studies investigating the prediction of post-operative dyspnoea in this

population.[169, 170, 182] The analysis contained within this thesis displayed the questionable evidence base of the *predictive* strength of post-operative dyspnoea from pre-operative lung function, confirming at best an association with post-operative dyspnoea. By using lung function as a linear variable, prediction of post-operative dyspnoea burden may be able to be improved.

As part of model validation, calibration was also investigated in line with recommendations. Other studies looking at risk prediction often do not comment on this aspect of model validation and this sets this work apart, again adding to the strength of the findings.

The models created within this study have variables which are easy to obtain in clinical practice. It would not involve any extra resource or much time to implement, which is paramount when creating a new clinical risk prediction model: if too much effort is involved for little gain, medical staff will be reluctant to adopt the model. [323]

As detailed in chapter 6, missing data was minimal within the internal dataset (GJNH), despite the size of the study. Missing data was more prevalent within the external data set and some patient data was excluded as a result. This is unavoidable in clinical research and by excluding some patients from the final analysis balanced the two groups in size. A sensitivity analysis was performed and this demonstrated that missing data had no effect on the overall outcome.

The analysis plan had to be modified during the study for the purpose of this thesis, prior to any results being viewed or analysed. Due to time constraints and a delay in getting data returned from external sites, derivation of the scoring tools took place using data only from the patients recruited from the Golden Jubilee National Hospital. The original analysis plan aimed to test the additional utility of BNP to existing conventional pre-operative risk prediction methods across all centres.

The analysis plan changed to recruit 125 patients and derive a new scoring tool before validating with 125 patients from an external dataset, (Section 6). Derivation and validating a scoring tool within the same study was a strength rather than a limitation. No other published work concerning the prediction of

dyspnoea following lung resection has attempted to derive and validated a scoring tool within a single study.

The global COVID-19 pandemic resulted in a lower-than-expected return rate of questionnaires at the 3-month time point at this stage of the study (the end). This was for two reasons; follow-up and reminder letters were not sent routinely as they had been before and patients were less likely to engage with the study during the pandemic, likely due to increased pressures in their social/personal life. Consequently, less patients than anticipated were in each group. However, with approximately one third of patients reporting the primary outcome and having a 'positive event' this was sufficient when performing regression analysis.

Arguably, the analysis could be repeated to incorporate all patients within the study (internal and external) to derive a further scoring tool to predict post-operative dyspnoea. However, given how convincingly BNP did not improve risk prediction it would be unlikely to change the primary finding of this study.

14.11 Future directions

The study has served to emphasise the burden of dyspnoea following lung resection, regardless of which scoring tool was selected. Based on the results of this study, the 'best' scoring model developed to predict dyspnoea (model 7) may be strong enough to allow entry into further low risk interventional studies, (Section 14.12). However, the model would need to be improved to become useful in clinical practice - influencing decision making process' for lung resection and aiding the surgical consent process.

A review of the literature showed a strong association between low technology exercise testing and post-operative morbidity and mortality and therefore this may be incorporated into any future model derivation. Recently there has been interest in heart rate recovery in risk prediction studies [334] and pilot work within our research group has shown some promise. Sub-maximal exercise testing with heart rate recovery may be another variable that would strengthen prediction of dyspnoea following lung resection. This should be a focus of future work.

Our analysis focused on dyspnoea at the 3-month time point following lung resection. The trajectory of post-operative dyspnoea is not well defined and therefore this time point may not be optimal- perhaps we should instead seek to measure dyspnoea at the 1-month or 1-year time point following surgery. A study to determine and gain better understanding of the development and time-frame of post-operative dyspnoea would be beneficial.

As highlighted in the literature review, pulmonary function has been used for many years to predict post-operative complications with varied success. While we observed association between BNP and post-operative complications (AF and pulmonary complications) further work would be required to demonstrate any predictive value. Equally, prediction of hospital stay would be useful to plan service provision, but much work would be required to successfully use BNP in this setting. The potential role of BNP in this population is not yet known and requires further exploration. Given it is hypothesised dyspnoea following lung resection may be driven in part by cardiovascular dysfunction (in particular RV dysfunction), a future study should seek to better understand, explore and define the mechanism.

One aspect of the study, not yet discussed, is a comparison of the MRC scores clinicians awarded patients at MDT discussion, versus what patients self-graded graded. As these values often underpin clinical decision making, this comparison would be useful to determine if clinicians are over or underscoring patients when assessing and planning the best treatment modality. This could be a secondary outcome of any future study in this area.

If resource was not finite, it would have been useful to include measurement of other biomarkers such as NT-pro BNP and cardiac troponins. These biomarkers may have increased predictive strength of long-term dyspnoea and may be considered when planning future studies.

This thesis has utilised traditional data analysis tools and statistics in an attempt to improve the prediction of dyspnoea following lung resection surgery for cancer. Artificial intelligence and machine learning are becoming increasingly popular techniques to discover insights, find new patterns and discover relationships within data. The large clinical dataset created within this project, including many

outcome measurements with biological markers, would be suitable for future analysis using artificial intelligence and machine learning. This approach may unveil the key to accurate prediction of post-operative dyspnoea in this population and improve the shared decision-making process.

14.12 Interventions

The purpose of deriving new risk prediction models in this study was to facilitate prognostic enrichment and entry of patients into low-risk interventional studies, aiming to mitigate the risk of dyspnoea. Future low-risk interventions may include;

- **Pre-operative** - Pre-operative optimisation or 'pre-habilitation' in a selection of patients to reduce body weight, increase cessation of smoking, optimise medications and offer psychological support. Given pre-operative pain was predictive of post-operative dyspnoea within this study, consideration of pre-operative optimisation and referral to acute/chronic pain team may be advantageous.

- **Intra-operative** - Consideration of intra-operative factors that may be contributing to post-operative dyspnoea. Enhanced monitoring and particular attention to reduced anaesthetic times/OLV times/PA clamp times, which are known to be harmful.[335]

- **Post-operative** - Increased 'enhanced recovery' care in HDU or other high acuity area with extra monitoring and physiotherapy in the initial period following surgery. Early post-operative echocardiogram to assess cardiovascular status and optimise fluid balance. Routine administration of high flow nasal cannula in the post-operative period- this has been shown to be beneficial in the lung cancer population in reducing length of hospital stay,[336] but its effects on post-operative dyspnoea have yet to be confirmed.

Combined, the above peri-operative interventions may lead to marginal gains to reduce the burden of post-operative dyspnoea.

If dyspnoea is indeed contributing to post-operative quality of life, a discharge package of care could be implemented with appropriate psychological support and follow up. This would ensure patients most vulnerable to a reduced post-operative quality of life or long-term dyspnoea can be identified early with intervention to improve patient outcome. If dyspnoea is inevitable, an offer of psychological intervention to council patients about expecting post-operative shortness of breath and provide techniques to manage this symptom, may improve global quality of life in some patients.

15 Appendices

Appendix 1 – (**Summary table)** *Stair climbing test as a predictor in lung cancer surgery assessment*

Appendix 2 – (**Summary table)** *Six-minute walk test as a predictor in lung cancer surgery assessment*

Appendix 3 (**Summary table)** *Shuttle walk test as a predictor in lung cancer surgery assessment*

Appendix 4 (**Summary table)** *Pre-operative CPET testing for lung resection*

Appendix 5 Ethical approval for study to explore prediction of dyspnoea in lung cancer population

Appendix 6 Grant confirmation – AAGBI/NIAA

Appendix 7 Patient information leaflet V4 with ethical approval (Golden Jubilee National Hospital)

Appendix 8 Blank consent form for PROFILES study (Golden Jubilee National Hospital)

Appendix 9 University of California and San Diego Shortness of Breath Questionnaire

Appendix 10 European Organisation for research and treatment of cancer Quality of life questionnaire

15.1 Appendix 1

Table 62 – Summary of studies examining stair climbing as a predictor in lung cancer surgery assessment

Study	Population	Method	Outcome	Comment
Olsen et al 1991 [214]	Retrospective n = 54 Single centre	30-day complications Maximum 5 flights Step height 0.174m	Negative correlation between steps climbed pre-operatively and post-operative intubation time (r= -0.35), hospital days (r= -0.28) and total number of complications sustained (r= -0.30), (all p<0.05). No further stats provided.	Association demonstrated, no predictive statistics provided
Holden et al 1992 [223]	Prospective n = 16 Single centre	Stair climbing performed in all patients 90-day complications and mortality	Climbing > 44 steps had a PPV 91% and NPV 80% for 90-day mortality. No other predictive stats	Small patient numbers Association demonstrated, minimal predictive statistics provided
Pate et al 1996 [213]	Prospective n = 12 Single centre	30-day complications and mortality Only borderline patients included into study – defined as FEV1 <2L	No statistical analysis Results described; Ten patients climbed more than 3 flights One patient climbed two flights One patient failed the test and developed post-operative respiratory complications	Small patient numbers No statistical comparison due to small cohort

15.2 Appendix 2

Table 63 – Summary of studies examining six-minute walk testing as a predictor in lung cancer surgery assessment

Study	Population	Method	Outcome	Comment
Holden et al 1992 [223]	Prospective n = 16 Single centre	Pre- op 6MWT all patients 90-day complications and mortality	6MWT distance of >1000 feet predictive of survival longer than 90 days (sensitivity 100%, PPV 85%, NPV 100%). Patients with no complications had longer 6MWT distance ($p<0.05$). (no values given)	Small patient numbers Association demonstrated with minimal predictive statistics provided
Markos et al 1989 [154]	Prospective n = 55 Single centre	12-minute walk test Pneumonectomy n=18 Lobectomy n=29 No resection n=6 Inoperable n=2	No complications group mean walk distance 1,018m SD 282m Complications group mean walk distance 905m SD 163m ($p>0.05$)	No difference in walk test in patients with and without complications. No predictive statistics
Pierce et al 1994 [225]	Prospective n= 52 Single centre	Pre-op 6MWT all patients Cardiopulmonary complications and mortality within 32 days Pneumonectomy n=11 Lobectomy n=29 Wedge n=12	6MWT independently predicted respiratory failure. No results from regression displayed. (Patients with complications had shorted distance at 6MWT, p=0.03)	6MWT is independently predictive of respiratory failure. No results from regression displayed.

Study	Population	Method	Outcome	Comment
Marjanski et al 2015[226]	Retrospective n = 253 Single centre	Pre-op 6MWT all patients In hospital cardiopulmonary complications Lobectomy n=253	6MWT distance < 500m was associated with post-operative cardiopulmonary complications (p<0.01) 6MWT distance < 500m independent predictor of cardiopulmonary complications (OR 2.50, 95% CI 1.28-5.30 p<0.01)	6MWT independent predictor of cardiopulmonary complications

6MWT = six minute walk test

15.3 Appendix 3

Table 64 – Summary of studies examining shuttle walk testing as a predictor in lung cancer surgery assessment

Study	Population	Method	Outcome	Comment
Fennelly et al 2017 [337]	Retrospective n = 101 Single centre	Pre-op SWT (400m Vs 250m walk tested cut-offs) 30-day complications Lobectomy n=89 Pneumonectomy n=12	SWT distance less than 400m independent predictor of cardiopulmonary complications (OR 4.3 95% CI 1.4-12.7, p<0.01)	SWT distance <400m independent predictor of post-operative cardiopulmonary complications 400m distance used in guidelines
Win et al 2004 [227]	Prospective n = 103 Single centre	Pre-op SWT Duration of hospital stay, post-operative complications and mortality rates Lobectomy n=57 Pneumonectomy n=37 Bi-lobectomy n=6 Wedge or no resection n=11	Good outcome n=69 (mean shuttle 419m) (No post-operative complications) Poor outcome n=34 (mean shuttle 388m) (Post-op death or major complications) No difference between two groups (p = 0.6)	No difference between groups No predictive stats performed as no significance at univariate analysis.

SWT = Shuttle walk test

15.4 Appendix 4

Table 65 – Summary of studies examining pre-operative CPET testing as a predictor in lung cancer surgery assessment

Study	Population	Method	Outcome	Comment
Bayram et al 2007[338]	Prospective n=55 Single centre	Pre-op CPET 30-day complications Two groups with pre-operative VO_2 max - 15ml/kg/min cut-off Lobectomy n=31 Bi lobectomy n=6 Pneumonectomy n=18	More pulmonary complications in those with VO_2 max <15ml/kg/min (8.8 v 16 mls/kg/min, p=0.05)	Small patient numbers Association demonstrated with no predictive statistics provided despite title of paper
Beccaria et al 2001 [53]	Prospective n=62 Single centre	Pre-op CPET Pneumonectomy n=14 Lobectomy n=48	Post-operative complications n=9 Two patients with predicted post-operative VO_2max > 10ml/kg/min became oxygen dependent	Small patient numbers No analysis –narrative only

Study	Population	Method	Outcome	Comment
Bechard et al 1987[234]	Prospective n=50 Single centre	Pre-op CPET - surgeon blinded Pneumonectomy n=10 Lobectomy n=28 Wedge resection n=12	Patients without complications had a higher VO_2max (17ml/kg/min Vs 9.9ml/kg/min, p<0.001) n=7 patients had VO_2 max <10ml/kg/min. n=2(29%) died and n=3(43%) had morbidity. n=28 patients had VO_2 max 10-20mls/kg/min. n=0 died and n=3(11%) had morbidity. No patients with VO_2 max >20ml/kg/min sustained any morbidity or death (p<0.001)	<10ml/kg/min is associated with significant morbidity and mortality. Association demonstrated with no predictive statistics provided
Bobbio et al 2009[339]	Prospective n=73 Single centre	Pre-op CPET Post-operative cardiopulmonary complication Lobectomy n=64 Bi-lobectomy n=5 Segmentectomy n=4	VO_2 max in those with and without pulmonary complications 19.7 Vs 16.9 ml/kg/min, p=0.04 VO_2 max to predict pulmonary complications - AUROCC 0.69 (95% CI 0.57-0.85) At 15ml/kg/min, sensitivity was 85% and specificity was 32% for pulmonary complications. At 20ml/kg/min Sensitivity was 36% and specificity was 90%.	Association demonstrated but VO_2 max is not an independent predictor of pulmonary complications

Study	Population	Method	Outcome	Comment
Bolliger et al 1995[340]	Prospective n=80 Single centre	Pre-op CPET Two groups; A – No complications B – Post-operative complications within 30 days Lobectomy n=45 Pneumonectomy n=21 Segmentectomy n=14	VO_2max% predicted and VO_2max ml/kg/min were greater in group A with no complications (p=0.0001 & p<0.0002 respectively) VO_2max% predicted 86% at regression. No further statistics given	VO_2max% predicted is an independent predictor of post-operative complications
Brat et al 2016[341]	Retrospective n=76 Multicentre	Pre-op CPET 30-day pulmonary complications Pneumonectomy n=17 Lobectomy n=47 Segmentectomy n=10	VO_2 max (ml/kg/min) was not associated with respiratory complications (p=0.15) Not significant at regression	CPET testing not an independent predictor of complications
Brunelli et al 2009[342]	Prospective n=204 Single centre	Pre-op CPET 30-day complications Lobectomy n=177 Pneumonectomy n=27	VO_2 max (ml/kg/min) associated with pulmonary complications (p=0.015) but not cardiovascular complications (p=0.3) VO_2 max an independent predictor of pulmonary complications (CI (0.77,0.99), p=0.04) No further stats provided	Author concludes: *VO_2 max >20mls/kg/min is a safe cut off value; no mortality occurred and only 3.5% of patients observed morbidity"* VO_2 max is an independent predictor of pulmonary complications
Brunelli et al 2012[207]	Prospective n=225 Single centre	Pre-op CPET 30-days pulmonary complications Lobectomy n=197 Pneumonectomy n=28	VO_2 max ml/kg/min not associated with post-operative pulmonary complications or mortality, p=0.5 at univariate analysis	VO_2 max is not an independent predictor of pulmonary complications or mortality

Study	Population	Method	Outcome	Comment
Brunelli et al 2014[343]	Prospective n=157 Retrospective database analysis	Pre-op CPET Long term survival (5-year) Lobectomy or Segmentectomy n=157	Survivals of patients with pre-op VO_2max >60% longer: 73% vs 40%, p<0.01). VO_2max above 60% (p=0.001, hazard ratio 2.4) was an independent predictor of survival.	VO_2max (%) is an independent predictor of survival. (best cut-off 60%)
Brutsche et al 2000[174]	Prospective n=125 Single centre	Pre-op CPET 30-day complications Pneumonectomy n=33 Bi-lobectomy n=9 Lobectomy n=68 Wedge n=15	VO_2 max/kg body weight (% predicted) was an independent predictor of complications (OR -0.05, SEM 0.014, p<0.01, no further stats provided)	VO_2max is an independent predictor of post-operative complications
Larsen et al 1997[344]	Prospective n=97 Single centre	Patients divided into two groups; - patients with cardiopulmonary *and* technical complications or death within 30 days. - patients with cardiopulmonary complications only Lobectomy n=52 Pneumonectomy n=27 No resection n=18	VO_2 max (ml/kg/min) is an independent predictor of any complication, p<0.01. For cardiopulmonary complications alone, VO_2 max was not an independent predictor, p>0.05 VO_2 max(ml/kg/min) (48 months) predictor of long-term survival, p<0.01. No further stats provided	VO_2 max an independent predictor of any complication Recommended VO_2max cut off (% predicted) for cardiopulmonary death = 50% predicted

Study	Population	Method	Outcome	Comment
Loewen et al 2007 [236]	Prospective n= 346 Multicentre	Pre-op CPET 30-day cardiopulmonary complications or mortality Lobectomy n=213 Pneumonectomy n=53 Wedge n=73 No resection n=7	Patients with complications had lower VO_2max%, p<0.01. Patients with VO_2max% <65% predicted had poor outcome (p<0.01, no test given), defined as respiratory complications or death	Multicentre study with robust methodology
Markos et al 1989[154]	Prospective n=55 Single centre	Pre-op CPET Pneumonectomy n= 18 Lobectomy n= 29 No resection n= 6 Inoperable n= 2	Complications in 4 of 12 patients with pre-operative VO_2max% <20ml/kg/min and in 3 of 4 patients with VO_2max% > 20/ml/kg/min.	Author concludes a relationship between decreasing VO_2max% and increasing complication rate following lobectomy exists Association demonstrated with no predictive statistics provided
Matsuoka et al 2004 [163]	Retrospective n=130 Single centre	Pre-op CPET 30-day complications Lobectomy n=130	Patients with complications had a lower pre-op VO_2max (ml/kg/min) (p=0.01)	Small number of patients with complications (n=9) Association demonstrated with no predictive statistics provided
Nagamatsu et al 2014[345]	Retrospective n=315 Single centre	Pre-op CPET Lobectomy n=291 Bi-Lobectomy n=10 Pneumonectomy n=14	No difference in ppVO_2max/m2 in those with and without complications (p=0.07, no test stated). No regression performed.	ppVO_2max not an independent predictor of complications.

Study	Population	Method	Outcome	Comment
Nagamatsu et al 2004[346]	Prospective n=211 Single centre	Pre-op CPET Limited complication definition – tracheostomy, ventilation >2 days, daily bronchoscopic lavage >7 days and arrhythmias >3 days Lobectomy n=166 Bi-lobectomy n=21 Pneumonectomy n=24	VO_2% max higher ($p<0.01$) and anaerobic threshold higher ($p<0.01$), in those without complications VO_2 max% not a predictor of post-operative CP complications ($r=-0.0719$, $p=0.08$).	VO_2% max associated with but not predictive of CP complications
Pompili et al 2013[347]	Prospective n=221 Single centre	Pre-op CPET and 3-month post-op QoL measured Two groups based on VO_2 max (< > 15 ml/kg/min) Lobectomy n=204 Pneumonectomy n=17	VO_2max Vs post-operative QoL (3 months) SF-36 survey to measure QoL. Mental Component Score (MCS) and Physical Component Score (PCS). No difference or change in post-operative QoL between high or low VO_2 max groups. (PCS score 27% Vs 21%, $p=0.3$ and MCS score 67% Vs 70%, $p=0.6$, no test given).	No association demonstrated between pre-op VO_2max and post-operative QoL (3 months) No predictive statistics displayed
Smith et al 1984[233]	Prospective n=22 Single centre	Pre-op CPET 30-day cardiopulmonary complications Wedge resection n=1 Lobectomy n=12 Pneumonectomy n=4 No resection n=5	Those without CP complications had higher VO_2max than those without (22.4 ml/kg/min Vs 14.9 ml/kg/min, $p<0.01$) Every patient (n=6) with VO_2max <15ml/kg/min had complications.	Association demonstrated between pre-op VO_2max and post-op complications No predictive statistics performed Small study, n= 16 had lung resection.

Study	Population	Method	Outcome	Comment
Villani et al 2004[348]	Prospective n=150 Single centre	Pre-op CPET 30-day cardiopulmonary complications Pneumonectomy n=150	Patients with complications had lower VO_2max (ml/kg/min and % predicted) than those without ($p<0.05$) 3 of 4 patients who died had VO_2max <50% predicted ($p<0.05$)	Association demonstrated with no predictive statistics provided
Wang et al 1999[198]	Prospective n=40 Single centre	Pre-op CPET 30-day pulmonary complications Lobectomy n=29 Bi-lobectomy n=2 Wedge resection n=9	Those with and without complications had a similar mean VO_2max value (ml/kg/min), 16.3 Vs 17.9 ($p=0.27$) Cut off values of 12.5, 15,17.5 and 20ml/kg/min *not predictive* of pulmonary complications, $p>0.05$ for all, chi-squared. No further stats	VO_2max (ml/kg/min) not associated with or independent predictor of post-op pulmonary complications
Wang et al 2000[349]	Prospective n=65 Single centre	Pre-op CPET In hospital cardiopulmonary complications Lobectomy n=32 Bi-lobectomy n=2 Wedge resection n=10 Pneumonectomy n=10 No resection n=11	Patients with complications had lower VO_2max% predicted ($p<0.01$) and lower VO_2 max ml/kg/min ($p<0.01$) AUROCC to predict overall complications using VO_2max (ml/kg/min) was 0.86 and the best cut off point was 15 ml/kg/min, with a sensitivity of 58% and specificity of 89%. No further stats	Association demonstrated between pre-op VO_2max% and in hospital complications VO_2max was a predictor of post-operative complications

15.5 Appendix 5

Ethical approval for study to explore prediction of dyspnoea in lung cancer population.

NHS
Health Research Authority

London - Queen Square Research Ethics Committee
HRA NRES Centre Manchester
Barlow House
3rd Floor
4 Minshull Street
Manchester
M1 3DZ

Telephone: 0207 104 8019

13 September 2018

Dr Ben Shelley
Academic Unit of Anaesthesia, Pain & Critical Care Medicine
University of Glasgow
Level 2 New Lister Building, 10-16 Alexandra Parade
G31 2ER

Dear Dr Shelley

Study title:	**PROFILES: bnP for prediction of Outcome Following Lung resection Surgery**
REC reference:	**18/LO/1563**
Protocol number:	**N/A**
IRAS project ID:	**251030**

Thank you for responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved by the sub-committee.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact please contact hra.studyregistration@nhs.net outlining the reasons for your request.

Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.
You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to

facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA and HCRW Approval (England and Wales)/ NHS permission for research is available in the Integrated Research Application System, at www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" above).

Approved documents

The documents reviewed and approved by the Committee are:

Document	*Version*	*Date*
GP/consultant information sheets or letters [GP Letter]	1	31 July 2018
IRAS Application Form [IRAS_Form_17082018]		17 August 2018
Letter from funder [Grant Award Letter]		27 July 2018

Email: nrescommittee.london-queensquare@nhs.net

Enclosures: *"After ethical review – guidance for researchers" [SL-AR2]*

Copy to: *Dr Brian Lafferty*

Dr Catherine Sinclair , NWTC Board

15.6 Appendix 6

Grant confirmation – AAGBI/NIAA

15.7 Appendix 7

Patient information leaflet version 4 with ethical approval

NHS
Health Research Authority
London - Queen Square Research Ethics Committee
HRA NRES Centre Manchester
Barlow House
3rd Floor
4 Minshull Street
Manchester
M1 3DZ

Tel: 0207 104 8019

02 November 2018

Dr Brian Lafferty
Clinical Research Fellow
Golden Jubilee National Waiting Times Hospital
Agamemnon Street, Clydebank
G814DY

Dear Dr Lafferty

Study title:	**PROFILES: bnP for prediction of Outcome Following Lung resection Surgery**
REC reference:	**18/LO/1563**
Protocol number:	**N/A**
Amendment number:	**1 27/9/18**
Amendment date:	**27 September 2018**
IRAS project ID:	**251030**

The above amendment was reviewed by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

The Committee found no ethical issues with this amendment.

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Notice of Substantial Amendment (non-CTIMP)	1 27/9/18	27 September 2018
Participant information sheet (PIS) [V4PILSGJNH clean]	4	24 September 2018
Participant information sheet (PIS) [V4PILSGJNH tracked]	4	24 September 2018
Participant information sheet (PIS) [V4PILSothercentres clean]	4	24 September 2018
Participant information sheet (PIS) [V4PILSothercentres tracked]	4	24 September 2018

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

Working with NHS Care Organisations

Golden Jubilee National Hospital
NHS National Waiting Times Centre

Chairman Stewart MacKinnon
Chief Executive Jill Young

Beardmore Street
Clydebank G81 4HX
Scotland
Telephone 0141 951 5000
Fax 0141 951 5500

NHS SCOTLAND

PATIENT INFORMATION SHEET – Golden Jubilee National Hospital

PROFILES

'bnP for pRediction of Outcome FollowIng Lung rEsection Surgery'

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it would involve for you. One of our team will go through the information sheet with you and answer any questions you have. We suggest this should take about 15 minutes. You should understand enough about the risks and benefits to be able to make an informed decision.

Talk to others about the study if you wish.

Ask us if there is anything that is not clear.

What is the purpose of the study?

Lung cancer is the second most common type of cancer in the UK and the leading cause of cancer related death. Surgery to remove the tumour and the surrounding lung (lung resection) often provides the best chance of cure. Frequently, patients are smokers with related lung or heart problems increasing the risks associated with surgery. Whilst surgery for lung cancer is considered the best chance of 'cure', patients may suffer long term breathlessness, lowering quality of life. This is important; public engagement work we have performed demonstrates repeatedly that second only to "being alive and cancer free" exercise capacity is the main priority of post-operative patients.

In a previous study our research group showed that the function of the right side of the heart (the right heart) is decreased following lung resection. The decrease in right heart (the part that supplies blood to the lungs) function was associated with a prolonged stay in the high dependency unit and blood markers indicating damage to the heart. The process by which the damage occurs is poorly understood

Prediction of breathlessness is difficult and not solely caused by lung removal but also from decreased performance of the heart. Although the surgery does not directly involve the heart, it is thought the damage is caused indirectly by the surgery and by removal of part of the lung. Current methods for predicting the risk of breathlessness after surgery are inaccurate. Some patients are refused surgery based on these methods yet may have had successful surgery. Furthermore, no specific treatment exists for patients considered to be at increased risk of breathlessness.

By identifying patients at risk of breathlessness, we believe an opportunity exists to intervene. A small study we completed (a 'pilot study') suggests measuring a hormone called 'BNP' (B type-natriuretic peptide, released by the heart) will improve prediction of post-operative breathlessness.

Why have you been chosen?

You have been chosen because you are about to undergo lung surgery. We intend to include 250 patients in total in the study across four hospitals in Scotland and Ireland.

Do you have to take part?

It is up to you to decide whether or not to join the study. We will describe the study and go through this information sheet. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time, without giving a reason. This would not affect the standard of care you receive.

What would happen if you take part?

A doctor or nurse from the research team would take you through this information sheet and obtain your permission to participate in this study in writing.

If you decide to take part you will be asked some questions about your health and asked to fill in a short questionnaire.

The main part of this study involves additional blood samples collected to measure the level of the hormone called BNP. One blood test will be taken before your operation and two additional tests afterwards on the 2nd and 3rd days after your operation.

We will also ask you to fill in a short questionnaire before your operation and afterwards at 3 months and one year.

If you are having your operation at the Golden Jubilee Hospital you will also be asked to complete a walking test before and after your operation as well.

What will you have to do?

Blood sampling

For an operation like this it is routine to have a fine plastic cannula (tube) placed into one of the arteries at your wrist. Where possible we will aim to take the blood samples from this line and so no extra needles will be involved. The first blood sample will be collected just before the start of your operation. Further blood samples will be collected each morning on days 2 and 3 after your operation with other routine bloods. We will take 3ml (less than a teaspoon full) of blood each time.

Questionnaire completion

You will be invited to complete a questionnaire before your operation and again at three months and one year after surgery. It should take no longer than 15 minutes to complete each time. You will be able to return the questionnaire by post if needed.

Walking test – Golden Jubilee Hospital

To give us a better understand of exercise and functional capacity we will ask those having surgery at the Golden Jubilee to undertake a six minute walking test before their operation and again at three months post operatively. This involves simply walking as far as you can in the six minute time period and having your heart rate measured simultaneously. You will be screened in accordance with national guidelines to see if it safe for you to take part in this section of the study and if you feel unwell or wish to stop the test you may do so at any time. This is an optional extra part of the study and you can still take part in the main study even if you do not wish to do the six minute walk test.

Will my care be affected by taking part in this study?

There is no drug or procedure being tested. The drugs you will receive and procedures that take place at the time of your operation are not affected by taking part in the study.

As your care is not affected by the study, there are no alternatives for treatment.

What are the possible disadvantages and risks of taking part?

Exercise walk testing will pose minimal risk to you. Exercise testing will be guided by your ability and you may stop at any stage. If you experience any chest pain, tightness or discomfort and/or excessive breathlessness we will ask you to stop exercising immediately and we will not ask you to perform any further exercise. You will not be allowed to perform this test if you have had chest pains in previous month.

Occasionally research studies using blood testing reveal significant unexpected abnormalities which require medical follow-up, either for further investigation or (more rarely) treatment. The tests we are doing are for research, but we review them carefully to avoid missing any such abnormality. If any significant unexpected abnormality is found we will inform your surgeon and send the report to your GP, who will be able to take it further with you.

Your lung operation and anaesthetic have risks and side effects and these will be explained to you when you consent for it. The care you receive will not change because you are taking part in the study.

What are the side effects of any treatment received when taking part?

There are no specific side effects of taking part in the study. The drugs you will receive and procedures that take place at the time of your operation are not affected by taking part in the study.

Your lung operation and anaesthetic have risks and side effects and these will be explained to you when you consent for it.

What are the possible benefits of taking part?

There is no direct benefit to you personally, over and above that of the operation itself. We hope the results of this study will provide information that helps the medical profession develop future treatments for other patients undergoing lung surgery.

What happens when the research stops?

After the blood tests and questionnaire at 1 year after your operation there will be no further research requirements.

What if there is a problem?

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. Please write to Mrs Paula McPhail – address given at the bottom of the leaflet.

Would participation in this study be kept confidential?

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence.

What happens if you don't want to carry on with the study?

You can withdraw from the study at any time. If you withdraw from the study, no more samples will be collected and we will destroy all your identifiable samples already taken, but we will need to use the data collected up to your withdrawal.

What if there is a problem?

We don't expect any problems in this low risk observational study. In the event that something does go wrong however, and you are harmed during the research due to someone's negligence then you may have grounds for a legal action for compensation against the hospital or health board but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you.

If you suffer any complications after your surgery it is possible you may require admission to the intensive care unit; if so you may be sedated and so unable to make decisions about your care. If this occurs we still plan to collect information about you for the study and would continue to collect blood tests.

Will participation in the study be kept confidential?

All information which is collected about you during the course of the research will be kept strictly confidential, and any information about you which leaves the hospital will have your name and address removed so that you cannot be recognised.

The necessary information will be taken from your medical records, from the hospital computers and the analysis of blood samples by members of the research team. All data will be anonymised by removing all names and addresses as soon as possible after collection, so that individuals cannot be recognised.

The anonymised data will be stored securely with a designated custodian. The study team will be the only people with direct access; they will analyse the data for the purposes of completing this project. It may also be used for further scientific work in the future; however this would require permission to be granted from an independent Research Ethics Committee. The data may also be reviewed by regulatory authorities responsible for monitoring the quality of research. Your study data may be held and processed on secure, password protected computers.

Anonymised data will be held for a maximum of 10 years before being destroyed securely. Your personal details (for example your name and address) will be stored separately for 3 years so that we are able to send you a copy of the study results. After this time these details will be destroyed securely.

Your GP will also be informed that you took part in the study.

What will happen to any samples you give?

This study involves taking blood samples which are in addition to those taken for normal patient care. The samples will be labelled by unique study number only – all data that is identifiable (eg names and addresses) will be removed.

Samples will be processed in the laboratories at the hospital and then be disposed of. Analysis is taking place in specialised laboratories in an alternative location from the routine blood tests you will undergo.

What will happen to the results of the research study?

The study is estimated to take 24 months commencing in October 2018. It is hoped to publish the results in 2020. We will send a summary of the study results to you at the end of the study. Individual patients will not be identified in any report / publication of the study.

Who is organising and funding the research?

The study is being organised by a group of doctors led by Dr Brian Lafferty and Dr Ben Shelley who work at both the University of Glasgow and Golden Jubilee National Hospital in Scotland. The study is funded by the National Institute of Academic Anaesthesia. Doctors and nurses conducting the research are not being paid specifically for including you in this study. The study is sponsored by the NHS National Waiting Times Centre Board.

Who has reviewed the study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study will be undertaken with their approval.

This study is being funded by the National Institute of Academic Anaesthesia whose expert panel has also reviewed the study as part of the funding process.

Further information and contact details

The NHS National Waiting Times Centre Board is the sponsor for this study based in the United Kingdom. We will be using information from you and your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The NHS National Waiting Times Centre Board will keep identifiable information about you for 10 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information by contacting Mrs Sharon Stott, Information Governance Manager, National Waiting Times Centre Board, Golden Jubilee National Hospital, Beardmore Street, Clydebank, Scotland, G81 4HX.

The NHS will collect information from you and your medical records for this research study in accordance with our instructions.

The NHS will use your name, CHI number and contact details to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Individuals from the NHS National Waiting Times Centre Board and regulatory organisations may look at your medical and research records to check the accuracy of the research study. The only people in the NHS National Waiting Times Centre Board who will have access to information that identifies you will be people who need to contact you to arrange appointments or complete questionnaires or audit the data collection process. The people who analyse the information will not be able to identify you and will not be able to find out your name, CHI number or contact details.

The NHS will keep identifiable information about you from this study for 10 years after the study has finished.

If you have any questions concerning the study, please ask the person presenting this form to you, or contact Dr Brian Lafferty on 07527 709 790.

If you have any complaints about any aspect of the way you have been approached or treated during the course of this study, you should write to **Mrs Paula McPhail, Legal and Feedback Co-ordinator/Complaints, National Waiting Times Centre Board, Golden Jubilee National Hospital, Beardmore Street, Clydebank, Scotland, G81 4HX**. You may also contact the independent Patient Advice and Support Service through your local citizens advice bureau.

15.8 Appendix 8

Blank consent form for PROFILES study – Golden Jubilee National Hospital

Golden Jubilee National Hospital
NHS National Waiting Times Centre

Chairman Stewart MacKinnon
Chief Executive Jill Young

Beardmore Street
Clydebank G81 4HX
Scotland
Telephone 0141 951 5000
Fax 0141 951 5500

NHS SCOTLAND

Patient Identification Number for this study: _____

CONSENT FORM

Title: PROFILES '**bn**P for p**R**ediction of **O**utcome **F**ollow**I**ng **L**ung r**E**section **S**urgery'

Name of Researcher: _____ Please initial box

1. I confirm that I have read and understand the information sheet dated..... (version.....) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐
3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the University of Glasgow, from regulatory authorities or from the NHS Trust. I give permission for these individuals to have access to my records when relevant to the research. ☐
4. I agree to my GP being informed of my participation in the study. ☐
5. I agree to take part in the above study. ☐

_____	_____	_____
Name of Patient	Date	Signature
_____	_____	_____
Name of person taking consent	Date	Signature

When completed: 1 for participant; 1 for researcher site file; 1 (original) to be kept in medical notes.

15.9 Appendix 9

This section displays the University of California and San Diego Shortness of Breath Questionnaire.

Instructions: For each activity listed below, please rate your breathlessness on a scale between zero and five where 0 is not at all breathless and 5 is maximally breathless or too breathless to do the activity. If the activity is one which you do not perform, please give your best estimate of breathlessness. Your responses should be for an 'average' day during the past week. **Please respond to all items by circling 0 – 5 as per box below.**

0	**Not at all**
1	**---**
2	**---**
3	**---**
4	**Severely**
5	**Maximally or unable to do because of breathlessness**

How short of breath do you get:						
At rest	0	1	2	3	4	5
Walking on a level at your own pace	0	1	2	3	4	5
Walking on a level with others your age	0	1	2	3	4	5
Walking up a hill	0	1	2	3	4	5
Walking upstairs	0	1	2	3	4	5
While eating	0	1	2	3	4	5
Standing from a chair	0	1	2	3	4	5
Brushing teeth	0	1	2	3	4	5
Shaving/brushing hair	0	1	2	3	4	5
Showering/ bathing	0	1	2	3	4	5
Dressing	0	1	2	3	4	5
Picking up and straightening	0	1	2	3	4	5
Doing dishes	0	1	2	3	4	5
Sweeping vacuuming	0	1	2	3	4	5
Making bed	0	1	2	3	4	5
Shopping	0	1	2	3	4	5
Doing laundry	0	1	2	3	4	5
Washing car	0	1	2	3	4	5
Mowing lawn	0	1	2	3	4	5
Watering lawn	0	1	2	3	4	5
Sexual activities	0	1	2	3	4	5
How much do these limit you in your daily life?						
Shortness of breath	0	1	2	3	4	5
Fear of 'hurting myself' by overexerting	0	1	2	3	4	5
Fear of shortness of breath	0	1	2	3	4	5

15.10 Appendix 10

European Organisation for research and treatment of cancer Quality of life questionnaire (EORTC) used within this study.

Please answer the following questions. If you do not perform a task please take your best estimate. Please respond to all questions by circling answers 1-4 as per box below.

1 – Not at all
2 - A little
3 - Quite a bit
4 - Very much

Question				
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or suitcase?	1	2	3	4
2. Do you have any trouble taking a long walk?	1	2	3	4
3. Do you have any trouble taking a short walk outside of the house?	1	2	3	4
4. Do you need to stay in bed or a chair during the day?	1	2	3	4
5. Do you need help with eating, dressing, washing yourself or using the toilet?	1	2	3	4
During the past week?				
6. Were you limited in doing either your work or other daily activities?	1	2	3	4
7. Were you limited in pursuing your hobbies or other leisure time activities?	1	2	3	4
8. Were you short of breath?	1	2	3	4
9. Have you had pain?	1	2	3	4
10. Did you need to rest?	1	2	3	4
11. Have you had trouble sleeping?	1	2	3	4
12. Have you felt weak?	1	2	3	4
13. Have you lacked appetite?	1	2	3	4
14. Have you felt nauseated?	1	2	3	4

15. Have you vomited?	1	2	3	4			
16. Have you been constipated?	1	2	3	4			
During the past week:							
17. Have you had diarrhoea	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 concentrating on things, like reading 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
30. How would you rate your overall quality of life during the past week?	1	2	3	4	5	6	7
During the past week:							
31. How much did you cough?	1	2	3	4			
32. Did you cough up blood?	1	2	3	4			

33. Were you short of breath when you rested?	1	2	3	4
34. Were you short of breath when you walked?	1	2	3	4
35. Were you short of breath when you climbed stairs?	1	2	3	4
36. Have you had a sore mouth or tongue?	1	2	3	4
37. Have you had trouble swallowing?	1	2	3	4
38. Have you had tingling hands or feet?	1	2	3	4
39. Have you had hair loss?	1	2	3	4
40. Have you had pain in your chest?	1	2	3	4
41. Have you had pain in your arm or shoulder?	1	2	3	4
42. Have you had pain in other parts of your body? (If so, where?)	1 2 3 4			
43. Did you take medication for pain? (yes or no? please circle)	1.Yes		2.No	
44. If so, how much did it help?	1	2	3	4

List of References

1 Lung cancer statistics: Cancer Research UK. 2015. Available from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer

2 Lung cancer in adults | Guidance and guidelines | NICE. 2018. Available from https://www.nice.org.uk/guidance/qs17/chapter/Introduction-and-overview

3 Jyoti Malhotra MM, Eva Negri , Carlo La Vecchia and Paolo Boffetta. Risk factors for lung cancer worldwide. European Respiratory Journal 2016; 889-902

4 Organisation WH. WHO launches new report on global tobacco use trends. 19th Dec 2019. Available from https://www.who.int/news/item/19-12-2019-who-launches-new-report-on-global-tobacco-use-trends (accessed August 2020 2020)

5 (NHS) NHS. Lung cancer:overview. Available from https://www.nhs.uk/conditions/lung-cancer/

6 Brown KF, Rumgay H, Dunlop C, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *Br J Cancer* 2018; **118**: 1130-41

7 Cote ML, Liu M, Bonassi S, et al. Increased risk of lung cancer in individuals with a family history of the disease: a pooled analysis from the International Lung Cancer Consortium. *Eur J Cancer* 2012; **48**: 1957-68

8 surgeons. Soc. National Thoracic surgery activity & outcomes report: The society for Cardiothoracic in Great Britain and Ireland 2019. 2019

9 (SIGN) HISSIGN. Management of lung cancer - a national clinical guideline. 2014

10 Strand TE, Rostad H, Møller B, Norstein J. Survival after resection for primary lung cancer: a population based study of 3211 resected patients. *Thorax* 2006; **61**: 710-5

11 Posther KE, Harpole Jr DH. The surgical management of lung cancer. *Cancer Investigation* 2006; **24**: 56-67

12 Ireland SACatSfCSiGBa. UK Cardiothoracic Surgery
Workforce report 2019 (SCTS). 2019

13 Sethi T, Lim E, Peake M, et al. Improving care for patients with lung cancer in the UK. *Thorax* 2013; **68**: 1181-5

14 Aldrich JM, Gropper MA. Editorial: Can we predict pulmonary complications after thoracic surgery? *Anesthesia and Analgesia* 2010; **110**: 1261-3

15 Ireland TSfCSiGB. The Society for Cardiothoracic Surgery in Great Britain & Ireland
Third National Thoracic Surgery Activity & Outcomes Report 2018. 2018

16 Powell HA, Tata LJ, Baldwin DR, Stanley RA, Khakwani A, Hubbard RB. Early mortality after surgical resection for lung cancer: an analysis of the English National Lung cancer audit. *Thorax* 2013; **68**: 826-34

17 Colice GL, Shafazand S, Griffin JP, Keenan R, Bolliger CT. Physiologic evaluation of the patient with lung cancer being considered for resectional surgery: ACCP evidenced-based clinical practice guidelines (2nd edition). *Chest* 2007; **132**: 161s-77s

18 Society BT. BTS Guidelines on the Radical Management of Patients with Lung Cancer | British Thoracic Society | Better lung health for all. *Thorax* 2010; **65**

19 Brunelli A, Varela G, Salati M, et al. Recalibration of the Revised Cardiac Risk Index in Lung Resection Candidates. *Annals of Thoracic Surgery* 2010; **90**: 199-203

20 Stougård J. Cardiac arrhythmias following pneumonectomy. *Thorax* 1969; **24**: 568-72

21 Cardinale D, Cosentino N, Moltrasio M, et al. Acute kidney injury after lung cancer surgery: Incidence and clinical relevance, predictors, and role of N-terminal pro B-type natriuretic peptide. *Lung Cancer* 2018; **123**: 155-9

22 Ishikawa S, Griesdale DE, Lohser J. Acute kidney injury after lung resection surgery: incidence and perioperative risk factors. *Anesth Analg* 2012; **114**: 1256-62

23 Organisation TWH. Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med., 1995 Nov; 1403-9.

24 Pompili C. Quality of life after lung resection for lung cancer. *Journal of Thoracic Disease* 2015; **7**: S138-S44

25 Brunelli A, Socci L, Refai M, Salati M, Xiume F, Sabbatini A. Quality of Life Before and After Major Lung Resection for Lung Cancer: A Prospective Follow-Up Analysis. *Annals of Thoracic Surgery* 2007; **84**: 410-6

26 Hirpara DH, Kidane B. Early Worsening of Quality of Life after Treatment of Stage I Lung Cancer. *Ann Am Thorac Soc* 2020; **17**: 935-6

27 Cannon J, Win T. Long-Term Quality of Life After Lung Resection. *Thoracic Surgery Clinics* 2008; **18**: 81-91

28 Bredda Saad IA, Botega NJ, Contrera Toro IF. Evaluation of quality of life of patients submitted to pulmonary resection due to neoplasia. *Jornal Brasileiro de Pneumologia* 2006; **32**: 10-5

29 Mak KS, van Bommel AC, Stowell C, et al. Defining a standard set of patient-centred outcomes for lung cancer. *Eur Respir J* 2016; **48**: 852-60

30 Cykert S, Kissling G, Hansen CJ. Patient preferences regarding possible outcomes of lung resection: what outcomes should preoperative evaluations target? *Chest* 2000; **117**: 1551-9

31 Denton TA, Matloff JM. Optimizing cardiothoracic surgery information for a managed care environment. *Ann Thorac Surg* 1995; **60**: 1522-5

32 Handy Jr JR, Asaph JW, Skokan L, et al. What happens to patients undergoing lung cancer surgery? Outcomes and quality of life before and after surgery. *Chest* 2002; **122**: 21-30

33 Treasure T, Holmes L. Measuring the quality of life. *J Heart Valve Dis* 1995; **4**: 337-8

34 Brocki BC, Andreasen JJ, Westerdahl E, De Souza DSR. Improvement in physical performance and health-related quality of life one year after radical

operation for primary or secondary lung cancer. *European Respiratory Journal Conference: European Respiratory Society Annual Congress* 2015; **46**

35 Möller A, Sartipy U. Predictors of postoperative quality of life after surgery for lung cancer. *J Thorac Oncol* 2012; **7**: 406-11

36 al NEe. Patient reported outcome measures in practice. British Medical Journal, 2015; 7818

37 Pompili C, Absolom K, Velikova G, Backhus L. Patients reported outcomes in thoracic surgery. *J Thorac Dis* 2018; **10**: 703-6

38 Khullar OV, Fernandez FG. Patient-Reported Outcomes in Thoracic Surgery. *Thoracic Surgery Clinics* 2017; **27**: 279-90

39 Khullar OV, Rajaei MH, Force SD, et al. Pilot Study to Integrate Patient Reported Outcomes After Lung Cancer Operations Into The Society of Thoracic Surgeons Database. *Ann Thorac Surg* 2017; **104**: 245-53

40 Laviolette L, Laveneziana P, Faculty ERS. Dyspnoea: a multidimensional and multidisciplinary approach. *Eur Respir J* 2014; **43**: 1750-62

41 Feinstein MB, Krebs P, Coups EJ, et al. Current dyspnea among long-term survivors of early-stage non-small cell lung cancer. *Journal of Thoracic Oncology* 2010; **5**: 1221-6

42 Society AT. Dyspnea. Mechanisms, assessment and management: a consensus statement. Am J Respir Crit Care Med, 1999; 321 – 40 .

43 Zieren HU, Muller JM, Hamberger U, Pichlmaier H. Quality of life after surgical therapy of bronchogenic carcinoma. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 1996; **10**: 233-7

44 Sarna L, Cooley ME, Brown JK, Chernecky C, Elashoff D, Kotlerman J. Symptom severity 1 to 4 months after thoracotomy for lung cancer. *Am J Crit Care* 2008; **17**: 455-67; quiz 68

45 Dales RE, Bélanger R, Shamji FM, Leech J, Crépeau A, Sachs HJ. Quality-of-life following thoracotomy for lung cancer. *J Clin Epidemiol* 1994; **47**: 1443-9

46 Balduyck B, Hendriks J, Lauwers P, Van Schil P. Quality of life evolution after lung cancer surgery: A prospective study in 100 patients. *Lung Cancer* 2007; **56**: 423-31

47 Myrdal G, Valtysdottir S, Lambe M, Ståhle E. Quality of life following lung cancer surgery. *Thorax* 2003; **58**: 194-7

48 Sjöland H, Caidahl K, Wiklund I, et al. Impact of coronary artery bypass grafting on various aspects of quality of life. *Eur J Cardiothorac Surg* 1997; **12**: 612-9

49 Young DJ, McCall PJ, Kirk A, Macfie A, Kinsella J, Shelley BG. B-type natriuretic peptide predicts deterioration in functional capacity following lung resection. *Interact Cardiovasc Thorac Surg* 2019; **28**: 945-52

50 BATES DV, MCILROY MB. Respiratory function after pneumonectomy. *Thorax* 1956; **11**: 303-11

51 NICE. Lung Cancer: diagnosis and management (NG122). National Institute for Healthcare and Excellence, 2019

52 Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J Am Coll Cardiol* 2014; **64**: e77-137

53 Beccaria M, Corsico A, Fulgoni P, et al. Lung cancer resection: The prediction of postsurgical outcomes should include long-term functional results. *Chest* 2001; **120**: 37-42

54 Pompili C, Brunelli A, Xiume F, Refai M, Salati M, Sabbatini A. Predictors of postoperative decline in quality of life after major lung resections. *European Journal of Cardio-thoracic Surgery* 2011; **39**: 732-7

55 Hotta K, Kiura K, Takigawa N, et al. Desire for information and involvement in treatment decisions: lung cancer patients' preferences and their physicians' perceptions: results from Okayama Lung Cancer Study Group Trial 0705. *J Thorac Oncol* 2010; **5**: 1668-72

56 General Medical Council U. Consent: Patients and Doctors making decisions together. 2008

57 Parshall MB, Schwartzstein RM, Adams L, et al. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med* 2012; **185**: 435-52

58 Ambrosino S. Dyspnoea and its measurement: CME article - educational aims. Breathe (ERS journals): European Respiratory Journal, 2004; 101-7

59 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**: 407-15

60 Cook CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *J Man Manip Ther* 2008; **16**: E82-3

61 Myles PS, Grocott MP, Boney O, Moonesinghe SR, Group C-S. Standardizing end points in perioperative trials: towards a core and extended outcome set. *Br J Anaesth* 2016; **116**: 586-9

62 Moonesinghe SR, Jackson AIR, Boney O, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine initiative: patient-centred outcomes. *Br J Anaesth* 2019; **123**: 664-70

63 Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther* 2012; **20**: 160-6

64 Wells G, Anderson J, Beaton D, et al. Minimal clinically important difference module: summary, recommendations, and research agenda. *J Rheumatol* 2001; **28**: 452-4

65 Ambrosino N, Scano G. Measurement and treatment of dyspnoea. *Respiratory Medicine* 2001; **95**: 539-47

66 Skinner JS, Hutsler R, Bergsteinová V, Buskirk ER. The validity and reliability of a rating scale of perceived exertion. *Med Sci Sports* 1973; **5**: 94-6

67 FLETCHER CM. The clinical diagnosis of pulmonary emphysema; an experimental study. *Proc R Soc Med* 1952; **45**: 577-84

68 CM F. Standardised questionnaire on respiratory symptoms: a statement prepared and approved by the MRC Committee on the Aetiology of Chronic Bronchitis (MRC breathlessness score). British Medical Journal, 1960

69 Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest* 1988; **93**: 580-6

70 *Mahler J. Clinical measurement of dyspnea. Mahler DA, ed. Dyspnea. New York, Marcel Dekker Inc*, 1998; *pp. 149-98*

71 Stenton C. The MRC breathlessness scale. *Occup Med (Lond)* 2008; **58**: 226-7

72 Wedzicha JA, Bestall JC, Garrod R, Garnham R, Paul EA, Jones PW. Randomized controlled trial of pulmonary rehabilitation in severe chronic obstructive pulmonary disease patients, stratified with the MRC dyspnoea scale. *Eur Respir J* 1998; **12**: 363-9

73 Nishimura K, Izumi T, Tsukino M, Oga T. Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. *Chest* 2002; **121**: 1434-40

74 Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999; **54**: 581-6

75 NICE. Chronic obstructive pulmonary disease in over 16s: diagnosis and management. 2010. Available from https://www.nice.org.uk/guidance/cg101/chapter/1- Guidance (accessed July 2021 2021)

76 De Torres JP, Pinto-Plata V, Ingenito E, et al. Power of outcome measurements to detect clinically significant changes in pulmonary rehabilitation of patients with COPD. *Chest* 2002; **121**: 1092-8

77 Cazzola. Table of Minimum Clinically Important Differences (MCID) (Cazzola 2015b). 2015. Available from https://copdx.org.au/copd-x-plan/appendices/appendix-5table-of-minimum-clinically-important-differences-mcid/ (accessed July 2021 2021)

78 Launois C, Barbe C, Bertin E, et al. The modified Medical Research Council scale for the assessment of dyspnea in daily living in obesity: a pilot study. *BMC Pulm Med* 2012; **12**: 61

79 Williams N. The MRC breathlessness scale. *Occup Med (Lond)* 2017; **67**: 496-7

80 Archibald CJ GT. Degree of objectively measured impairment and perceived shortness of breath with activities of daily living in patients with chronic obstructive pulmonary disease. Can J Rehab 1987, 1987; 45-54

81 Eakin EG, Resnikoff PM, Prewitt LM, Ries AL, Kaplan RM. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998; **113**: 619-24

82 Bujang MA, Omar ED, Baharum NA. A Review on Sample Size Determination for Cronbach's Alpha Test: A Simple Guide for Researchers. *Malays J Med Sci* 2018; **25**: 85-99

83 Kupferberg DH, Kaplan RM, Slymen DJ, Ries AL. Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. *J Cardiopulm Rehabil* 2005; **25**: 370-7

84 Ries AL. Minimally clinically important difference for the UCSD Shortness of Breath Questionnaire, Borg Scale, and Visual Analog Scale. *COPD* 2005; **2**: 105-10

85 EuroQoL. Available from https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/ (accessed February 2021)

86 Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes* 2007; **5**: 70

87 Zanini A, Aiello M, Adamo D, et al. Estimation of minimal clinically important difference in EQ-5D visual analog scale score after pulmonary rehabilitation in subjects with COPD. *Respir Care* 2015; **60**: 88-95

88 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005; **14**: 1523-32

89 Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. *Expert Rev Pharmacoecon Outcomes Res* 2014; **14**: 221-33

90 EORTC manual for scoring QLQ C30 and LC-13 questionnaire. Available from https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf (accessed February 2021)

91 Pompili C, Koller M, Velikova G, et al. EORTC QLQ-C30 summary score reliably detects changes in QoL three months after anatomic lung resection for Non-Small Cell Lung Cancer (NSCLC). *Lung Cancer* 2018; **123**: 149-54

92 Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; **16**: 139-44

93 Fiteni F, Anota A, Westeel V, Bonnetain F. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. *BMC Cancer* 2016; **16**: 122

94 Wu YL, Lee JS, Thongprasert S, et al. Intercalated combination of chemotherapy and erlotinib for patients with advanced stage non-small-cell lung cancer (FASTACT-2): a randomised, double-blind trial. *Lancet Oncol* 2013; **14**: 777-86

95 Yang JC, Hirsh V, Schuler M, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013; **31**: 3342-50

96 Fløtten Ø, Grønberg BH, Bremnes R, et al. Vinorelbine and gemcitabine vs vinorelbine and carboplatin as first-line treatment of advanced NSCLC. A phase III randomised controlled trial by the Norwegian Lung Cancer Study Group. *Br J Cancer* 2012; **107**: 442-7

97 Bergman B AN, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. . Eur J Cancer 1994; 635-42.

98 Koller M, Hjermstad MJ, Tomaszewski KA, et al. An international study to revise the EORTC questionnaire for assessing quality of life in lung cancer patients. *Ann Oncol* 2017; **28**: 2874-81

99 World Health Organisation W. WHO DAS 2.0 schedule. Available from https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health/who-disability-assessment-schedule (accessed October 2020)

100 Shulman MA, Kasza J, Myles PS. Defining the Minimal Clinically Important Difference and Patient-acceptable Symptom State Score for Disability Assessment in Surgical Patients. *Anesthesiology* 2020; **132**: 1362-70

101 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**: 361-70

102 Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002; **52**: 69-77

103 Schellekens MPJ, van den Hurk DGM, Prins JB, Molema J, van der Drift MA, Speckens AEM. The suitability of the Hospital Anxiety and Depression Scale, Distress Thermometer and other instruments to screen for psychiatric disorders in both lung cancer patients and their partners. *J Affect Disord* 2016; **203**: 176-83

104 Castelli L, Binaschi L, Caldera P, Torta R. Depression in lung cancer patients: is the HADS an effective screening tool? *Support Care Cancer* 2009; **17**: 1129-32

105 Brunelli A, Xiume F, Refai M, et al. Evaluation of expiratory volume, diffusion capacity, and exercise tolerance following major lung resection: A prospective follow-up analysis. *Chest* 2007; **131**: 141-7

106 Miserocchi G, Beretta E, Rivolta I. Respiratory Mechanics and Fluid Dynamics After Lung Resection Surgery. *Thoracic Surgery Clinics* 2010; **20**: 345-57

107 Bolliger CT, Jordan P, Soler M, et al. Pulmonary function and exercise capacity after lung resection. *European Respiratory Journal* 1996; **9**: 415-21

108 Pelletier C, Lapointe L, LeBlanc P. Effects of lung resection on pulmonary function and exercise capacity. *Thorax* 1990; **45**: 497-502

109 K RL, Svendsen UG, Milman N, Brenoe J, Petersen BN. Cardiopulmonary function at rest and during exercise after resection for bronchial carcinoma. *Annals of Thoracic Surgery* 1997; **64**: 960-4

110 Nezu K, Kushibe K, Tojo T, Takahama M, Kitamura S. Recovery and limitation of exercise capacity after lung resection for lung cancer. *Chest* 1998; **113**: 1511-6

111 Siafakas NM, Mitrouska I, Bouros D, Georgopoulos D. Surgery and the respiratory muscles. *Thorax* 1999; **54**: 458-65

112 Karlson KE, Seltzer B, Lee S, Gliedman ML. Influence of thoracotomy on pulmonary mechanics: association of increased work of breathing during anesthesia and postoperative pulmonary complications. *Ann Surg* 1965; **162**: 973-80

113 Nomori H, Horio H, Fuyuno G, Kobayashi R, Yashima H. Respiratory muscle strength after lung resection with special reference to age and procedures of thoracotomy. *Eur J Cardiothorac Surg* 1996; **10**: 352-8

114 Brocki BC, Westerdahl E, Langer D, Souza DSR, Andreasen JJ. Decrease in pulmonary function and oxygenation after lung resection. *ERS Monograph* 2018; **4 (1) (no pagination)**

115 Vainshelboim B, Fox BD, Saute M, et al. Limitations in exercise and functional capacity in long-term postpneumonectomy patients. *Journal of Cardiopulmonary Rehabilitation and Prevention* 2015; **35**: 56-64

116 Garner M, Routledge T, King JE, et al. New-onset atrial fibrillation after anatomic lung resection: predictive factors, treatment and follow-up in a UK thoracic centre. *Interact Cardiovasc Thorac Surg* 2017; **24**: 260-4

117 Bagheri R, Yousefi Y, Rezai R, Azemonfar V, Keshtan FG. Atrial fibrillation after lung surgery: incidence, underlying factors, and predictors. *Kardiochir Torakochirurgia Pol* 2019; **16**: 53-6

118 Amar D, Roistacher N, Burt M, Reinsel RA, Ginsberg RJ, Wilson RS. Clinical and echocardiographic correlates of symptomatic tachydysrhythmias after noncardiac thoracic surgery. *Chest* 1995; **108**: 349-54

119 Ariansen I, Edvardsen E, Borchsenius F, Abdelnoor M, Tveit A, Gjesdal K. Lung function and dyspnea in patients with permanent atrial fibrillation. *Eur J Intern Med* 2011; **22**: 466-70

120 McCall PJ, Arthur A, Glass A, et al. The right ventricular response to lung resection. *J Thorac Cardiovasc Surg* 2019; **158**: 556-65.e5

121 Wang Z, Yuan J, Chu W, Kou Y, Zhang X. Evaluation of left and right ventricular myocardial function after lung resection using speckle tracking echocardiography. *Medicine (Baltimore)* 2016; **95**: e4290

122 McLure LE, Peacock AJ. Cardiac magnetic resonance imaging for the assessment of the heart and pulmonary circulation in pulmonary hypertension. *Eur Respir J* 2009; **33**: 1454-66

123 Haddad F, Hunt SA, Rosenthal DN, Murphy DJ. Right ventricular function in cardiovascular disease, part I: Anatomy, physiology, aging, and functional assessment of the right ventricle. *Circulation* 2008; **117**: 1436-48

124 Lewis JW, Bastanfar M, Gabriel F, Mascha E. Right heart function and prediction of respiratory morbidity in patients undergoing pneumonectomy with moderately severe cardiopulmonary dysfunction. *J Thorac Cardiovasc Surg* 1994; **108**: 169-75

125 Matyal R, Mahmood F, Hess P, et al. Right ventricular echocardiographic predictors of postoperative supraventricular arrhythmias after thoracic surgery: a pilot study. *Ann Thorac Surg* 2010; **90**: 1080-6

126 Reed CE, Spinale FG, Crawford FA. Effect of pulmonary resection on right ventricular function. *Ann Thorac Surg* 1992; **53**: 578-82

127 Ferguson MK. Assessment of operative risk for pneumonectomy. *Chest Surg Clin N Am* 1999; **9**: 339-51

128 Kowalewski J, Brocki M, Dryjański T, Kaproń K, Barcikowski S. Right ventricular morphology and function after pulmonary resection. *Eur J Cardiothorac Surg* 1999; **15**: 444-8

129 Foroulis CN, Kotoulas CS, Kakouros S, et al. Study on the late effect of pneumonectomy on right heart pressures using Doppler echocardiography. *European Journal of Cardio-thoracic Surgery* 2004; **26**: 508-14

130 Okada M, Ota T, Matsuda H, Okada K, Ishii N. Right ventricular dysfunction after major pulmonary resection. *J Thorac Cardiovasc Surg* 1994; **108**: 503-11

131 Okada M, Ishii N, Yamashita C, et al. Right ventricular ejection fraction in the preoperative risk evaluation of candidates for pulmonary resection. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**: 364-70

132 Khan J, Alonso-Coello P, Devereaux PJ. Myocardial injury after noncardiac surgery. *Curr Opin Cardiol* 2014; **29**: 307-11

133 Botto F, Alonso-Coello P, Chan MT, et al. Myocardial injury after noncardiac surgery: a large, international, prospective cohort study establishing diagnostic criteria, characteristics, predictors, and 30-day outcomes. *Anesthesiology* 2014; **120**: 564-78

134 Cardinale D, Martinoni A, Cipolla CM, et al. Atrial fibrillation after operation for lung cancer: clinical and prognostic significance. *Ann Thorac Surg* 1999; **68**: 1827-31

135 Herrington CS, Shumway SJ. Myocardial ischemia and infarction postthoracotomy. *Chest Surgery Clinics of North America* 1998; **8**: 495-502

136 Grobben RB, van Waes JAR, Leiner T, et al. Unexpected Cardiac Computed Tomography Findings in Patients With Postoperative Myocardial Injury. *Anesth Analg* 2018; **126**: 1462-8

137 Ziomek S, Read RC, Tobler HG, et al. Thromboembolism in patients undergoing thoracotomy. *Ann Thorac Surg* 1993; **56**: 223-6; discussion 7

138 Marini C, Miniati M, Ambrosino N, et al. Dyspnoea and hypoxaemia after lung surgery: The role of interatrial right-to-left shunt. *European Respiratory Journal* 2006; **28**: 174-81

139 Marini C, Miniati M, Pratali L, et al. Interatrial right-to-left shunt after lung surgery: Diagnostic value of perfusion lung scanning. *American Journal of the Medical Sciences* 2004; **328**: 180-4

140 Falcoz PE, Conti M, Brouchet L, et al. The Thoracic Surgery Scoring System (Thoracoscore): risk model for in-hospital death in 15,183 patients requiring thoracic surgery. *J Thorac Cardiovasc Surg* 2007; **133**: 325-32

141 Chamogeorgakis T, Toumpoulis I, Tomos P, et al. External validation of the modified Thoracoscore in a new thoracic surgery program: prediction of in-hospital mortality. *Interact Cardiovasc Thorac Surg* 2009; **9**: 463-6

142 Falcoz PE, Conti M, Brouchet L, et al. The Thoracic Surgery Scoring System (Thoracoscore): Risk model for in-hospital death in 15,183 patients requiring thoracic surgery. *Journal of Thoracic and Cardiovascular Surgery* 2007; **133**: 325-32

143 Berrisford R, Brunelli A, Rocco G, Treasure T, Utley M. The European Thoracic Surgery Database project: Modelling the risk of in-hospital death following lung resection. *European Journal of Cardio-thoracic Surgery* 2005; **28**: 306-11

144 Barua A, Handagala SD, Socci L, et al. Accuracy of two scoring systems for risk stratification in thoracic surgery. *Interact Cardiovasc Thorac Surg* 2012; **14**: 556-9

145 Society BT, Party SoCSoGBaIW. BTS guidelines: guidelines on the selection of patients with lung cancer for surgery. *Thorax* 2001; **56**: 89-108

146 Sharkey A, Ariyaratnam P, Anikin V, et al. Thoracoscore and European Society Objective Score Fail to Predict Mortality in the UK. *World J Oncol* 2015; **6**: 270-5

147 Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999; **100**: 1043-9

148 Disease. GIfCOP. POCKET GUIDE TO COPD
DIAGNOSIS, MANAGEMENT, AND PREVENTION
A Guide for Health Care Professionals. 2020 Global Initiative for Chronic Obstructive Lung Disease, Inc, 2020

149 Graham BL BV, Burgos F. 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur Respir Journal*, 2017; 49

150 Ferguson MK, Little L, Rizzo L, et al. Diffusing capacity predicts morbidity and mortality after pulmonary resection. *J Thorac Cardiovasc Surg* 1988; **96**: 894-900

151 Saydain G, Beck KC, Decker PA, Cowl CT, Scanlon PD. Clinical significance of elevated diffusing capacity. *Chest* 2004; **125**: 446-52

152 Johnston JD TW. A stepwise approach to the interpretation of pulmonary function tests. *Am Fam Physician*, 2014; 359-66

153 Kabir NA, Harvey SB, Rosenfeld A. Design and fabrication of pulmonary embolism phantom for planar and SPECT V/Q imaging quality assurance. *Australas Phys Eng Sci Med* 2010; **33**: 271-7

154 Markos J, Mullan BP, Hillman DR, et al. Preoperative assessment as a predictor of mortality and morbidity after lung resection. *Am Rev Respir Dis* 1989; **139**: 902-10

155 Brunelli A, Rocco G. Spirometry: predicting risk and outcome. *Thorac Surg Clin* 2008; **18**: 1-8

156 Brunelli A, Refai M, Salati M, Xiume F, Sabbatini A. Predicted versus observed FEV1 and DLCO after major lung resection: a prospective evaluation at different postoperative periods. *Ann Thorac Surg* 2007; **83**: 1134-9

157 Ferguson MK, Reeder LB, Mick R. Optimizing selection of patients for major lung resection. *J Thorac Cardiovasc Surg* 1995; **109**: 275-81; discussion 81-3

158 Win T, Laroche CM, Groves AM, et al. Use of quantitative lung scintigraphy to predict postoperative pulmonary function in lung cancer patients undergoing lobectomy. *Annals of Thoracic Surgery* 2004; **78**: 1215-8

159 Olsen GN, Block AJ, Tobias JA. Prediction of postpneumonectomy pulmonary function using quantitative macroaggregate lung scanning. *Chest* 1974; **66**: 13-6

160 Brocki BC, Westerdahl E, Langer D, Souza DSR, Andreasen JJ. Decrease in pulmonary function and oxygenation after lung resection. *ERJ Open Research* 2018; **4 (1) (no pagination)**

161 Handy JR, Conner T, Grunkemeier GL, Wu YX. Functional outcomes and quality of life after high-risk vs low-risk pulmonary lobectomy for lung cancer. *Chest* 2010; **Conference**: CHEST 2010 Annual Meeting. Vancouver, BC Canada. Conference Publication: (var.pagings). 138 (4) (no pagination)

162 Ilonen IK, Rasanen JV, Knuuttila A, et al. Quality of life following lobectomy or bilobectomy for non-small cell lung cancer, a two-year prospective follow-up study. *Lung Cancer* 2010; **70**: 347-51

163 Matsuoka H, Nishio W, Sakamoto T, Harada H, Tsubota N. Prediction of morbidity after lung resection with risk factors using treadmill exercise test. *European Journal of Cardio-thoracic Surgery* 2004; **26**: 480-2

164 Brunelli A, Charloux A, Bolliger CT, et al. The European Respiratory Society and European Society of Thoracic Surgeons clinical guidelines for evaluating fitness for radical treatment (surgery and chemoradiotherapy) in patients with lung cancer. *European Journal of Cardio-thoracic Surgery* 2009; **36**: 181-4

165 Brunelli A, Kim AW, Berger KI, Addrizzo-Harris DJ. Physiologic evaluation of the patient with lung cancer being considered for resectional surgery: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest* 2013; **143**: e166S-e90S

166 Bolliger CT, Perruchoud AP. Functional evaluation of the lung resection candidate. *European Respiratory Journal* 1998; **11**: 198-212

167 Brunelli A, Fianchini A. Stair climbing test in lung resection candidates with low predicted postoperative FEV<inf>1</inf> [5]. *Chest* 1179; **124**

168 CANDER L. PHYSIOLOGIC ASSESSMENT AND MANAGEMENT OF THE PREOPERATIVE PATIENT WITH PULMONARY EMPHYSEMA. *Am J Cardiol* 1963; **12**: 324-6

169 Bousamra IM, Presberg KW, Chammas JH, et al. Early and late morbidity in patients undergoing pulmonary resection with low diffusion capacity. *Annals of Thoracic Surgery* 1996; **62**: 968-75

170 Foroulis CN, Lioulias AG, Papakonstantinou C. Simple laboratory parameters which can determine the clinical state of patients after pneumonectomy for lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2009; **4**: 55-61

171 Wang J, Olak J, Ferguson MK. Diffusing capacity predicts operative mortality but not long-term survival after resection for lung cancer. *J Thorac Cardiovasc Surg* 1999; **117**: 581-6; discussion 6-7

172 Win T, Jackson A, Sharples L, et al. Cardiopulmonary exercise tests and lung cancer surgical outcome. *Chest* 2005; **127**: 1159-65

173 Brunelli A, Refai MA, Monteverde M, Borri A, Salati M, Fianchini A. Stair climbing test predicts cardiopulmonary complications after lung resection. *Chest* 2002; **121**: 1106-10

174 Brutsche MH, Spiliopoulos A, Bolliger CT, Licker M, Frey JG, Tschopp JM. Exercise capacity and extent of resection as predictors of surgical risk in lung cancer. *European Respiratory Journal* 2000; **15**: 828-32

175 Doug W, Peter W, Robin K. The Society for cardiothoracic surgery in Great Britain and Ireland. Third National Thoracic Surgery Activity and Outcomes Report. 2018

176 Ezer N, Kale M, Sigel K, et al. Outcomes after Video-assisted Thoracoscopic Lobectomy versus Open Lobectomy for Early-Stage Lung Cancer in Older Adults. *Ann Am Thorac Soc* 2018; **15**: 76-82

177 Brunelli A, Charloux A, Bolliger CT, et al. ERS/ESTS clinical guidelines on fitness for radical therapy in lung cancer patients (surgery and chemo-radiotherapy). *European Respiratory Journal* 2009; **34**: 17-41

178 Nakagawa T, Chiba N, Saito M, Sakaguchi Y, Ishikawa S. Preoperative physiologic assessment of lung cancer surgery using the modified ACCP guidelines. *Journal of Thoracic Oncology* 2013; **2)**: S1114

179 Brunelli A, Al Refai M, Monteverde M, Sabbatini A, Xiume F, Fianchini A. Predictors of early morbidity after major lung resection in patients with and without airflow limitation. *Annals of Thoracic Surgery* 2002; **74**: 999-1003

180 Carretta A, Zannini P, Puglisi A, et al. Improvement of pulmonary function after lobectomy for non-small cell lung cancer in emphysematous patients. *European Journal of Cardio-thoracic Surgery* 1999; **15**: 602-7

181 Korst RJ, Ginsberg RJ, Ailawadi M, et al. Lobectomy improves ventilatory function in selected patients with severe COPD. *Annals of Thoracic Surgery* 1998; **66**: 898-902

182 Handy JR, Conner T, Grunkemeier GL, Wu YX. Functional outcomes and quality of life after high-risk vs low-risk pulmonary lobectomy for lung cancer. *Chest Conference: CHEST* 2010; **138**

183 Boushy SF BD, North LB. Clinical course related to preoperative and postoperative pulmonary function in patients with bronchogenic carcinoma. *Chest* 1971; **59(4)**: 383-91

184 Brunelli A, Pompili C, Berardi R, et al. Performance at preoperative stair-climbing test is associated with prognosis after pulmonary resection in stage i non-small cell lung cancer. *Annals of Thoracic Surgery* 2012; **93**: 1796-800

185 Ferguson MK, Watson S, Johnson E, Vigneswaran WT. Predicted postoperative lung function is associated with all-cause long-term mortality after major lung resection for cancer. *European Journal of Cardio-thoracic Surgery* 2014; **45**: 660-4

186 Ferguson MK, Vigneswaran WT. Diffusing capacity predicts morbidity after lung resection in patients without obstructive lung disease. *Ann Thorac Surg* 2008; **85**: 1158-64; discussion 64-5

187 Yabroff KR, McNeel TS, Waldron WR, et al. Health limitations and quality of life associated with cancer and other chronic diseases by phase of care. *Med Care* 2007; **45**: 629-37

188 Pompili C, Brunelli A, Xiume F, Refai M, Salati M, Sabbatini A. Predictors of postoperative decline in quality of life after major lung resections. *European Journal of Cardio thoracic Surgery* 2011; **39**: 732-7

189 Saad IAB, Botega NJ, Toro IFC. Predictors of quality-of-life improvement following pulmonary resection due to lung cancer. *Sao Paulo Medical Journal* 2007; **125**: 46-9

190 DL F. *Design and Analysis of Quality of Life Studies in Clinical Trials*. Chapman & Hall, 2002

191 Win T, Sharples L, Wells FC, Ritchie AJ, Munday H, Laroche CM. Effect of lung cancer surgery on quality of life. *Thorax* 2005; **60**: 234-8

192 Brocki BC, Westerdahl E, Andreasen JJ, Souza DSR. Improvements in physical performance and health-related quality of life one year after radical operation for lung cancer. *Cancer Treatment Communications* 2015; **4**: 65-74

193 Greillier L, Thomas P, Loundou A, et al. Pulmonary function tests as a predictor of quantitative and qualitative outcomes after thoracic surgery for lung cancer. *Clinical Lung Cancer* 2007; **8**: 554-61

194 Kim ES, Kim YT, Kang CH, et al. Prevalence of and risk factors for postoperative pulmonary complications after lung cancer surgery in patients with early-stage COPD. *International Journal of COPD* 2016; **11**: 1317-26

195 Nagamatsu Y, Iwasaki Y, Hayashida R, et al. Factors related to an early restoration of exercise capacity after major lung resection. *Surgery Today* 2011; **41**: 1228-33

196 Pompili C, Brunelli A, Xiume F, et al. Prospective external convergence evaluation of two different quality-of-life instruments in lung resection patients. *European Journal of Cardio thoracic Surgery* 2011; **40**: 99-105

197 Amar D, Munoz D, Shi W, Zhang H, Thaler HT. A clinical prediction rule for pulmonary complications after thoracic surgery for primary lung cancer. *Anesthesia and Analgesia* 2010; **110**: 1343-8

198 Wang J, Olak J, Ultmann RE, Ferguson MK. Assessment of pulmonary complications after lung resection. *Annals of Thoracic Surgery* 1999; **67**: 1444-7

199 F S, S B, J H, et al. Pulmonary complications following lung resection: a comprehensive analysis of incidence and possible risk factors. *Chest* 2000; **118**: 1263-70

200 Kearney DJ, Lee TH, Reilly JJ, DeCamp MM, Sugarbaker DJ. Assessment of operative risk in patients undergoing lung resection. Importance of predicted pulmonary function. *Chest* 1994; **105**: 753-9

201 Dales RE, Dionne G, Leech JA, Lunau M, Schweitzer I. Preoperative prediction of pulmonary complications following thoracic surgery. *Chest* 1993; **104**: 155-9

202 Brunelli A, Refai M, Xiume F, et al. Oxygen desaturation during maximal stair-climbing test and postoperative complications after major lung resections. *European Journal of Cardio-thoracic Surgery* 2008; **33**: 77-82

203 Brunelli A, Monteverde M, Al Refai M, Fianchini A. Stair climbing test as a predictor of cardiopulmonary complications after pulmonary lobectomy in the elderly. *Annals of Thoracic Surgery* 2004; **77**: 266-70

204 Licker MJ, Widikker I, Robert J, et al. Operative Mortality and Respiratory Complications After Lung Resection for Cancer: Impact of Chronic Obstructive Pulmonary Disease and Time Trends. *Annals of Thoracic Surgery* 2006; **81**: 1830-7

205 Bobbio A, Chetta A, Internullo E, et al. Exercise capacity assessment in patients undergoing lung resection. *European Journal of Cardio thoracic Surgery* 2009; **35**: 419-22

206 Brat K, Tothova Z, Merta Z, et al. Resting End-Tidal Carbon Dioxide Predicts Respiratory Complications in Patients Undergoing Thoracic Surgical Procedures. *Annals of Thoracic Surgery* 1725; **102**: 1725-30

207 Brunelli A, Belardinelli R, Pompili C, et al. Minute ventilation-to-carbon dioxide output (Ve/Vco<inf>2</inf>) slope is the strongest predictor of respiratory complications and death after pulmonary resection. *Annals of Thoracic Surgery* 2012; **93**: 1802-6

208 Keeratichananont W, Thanadetsuntorn C, Keeratichananont S. Value of preoperative 6-minute walk test for predicting postoperative pulmonary complications. *Therapeutic Advances in Respiratory Disease* 2016; **10**: 18-25

209 Salati M, Brunelli A. Risk Stratification in Lung Resection. *Curr Surg Rep* 2016; **4**: 37

210 Brunelli A, Refai M, Xiume F, et al. Performance at Symptom-Limited Stair-Climbing Test is Associated With Increased Cardiopulmonary Complications, Mortality, and Costs After Major Lung Resection. *Annals of Thoracic Surgery* 2008; **86**: 240-8

211 Van Norstrand D KM, Humphrey EW. Preresectional evaluation of risk from pneumonectomy. Surg Gynecol Obstet, 1968; 306-12

212 Brunelli A, Pompili C, Salati M. Low-technology exercise test in the preoperative evaluation of lung resection candidates. *Monaldi Archives for Chest Disease Pulmonary Series* 2010; **73**: 72-8

213 Pate P, Tenholder MF, Griffin JP, Eastridge CE, Weiman DS. Preoperative assessment of the high-risk patient for lung resection. *Annals of Thoracic Surgery* 1996; **61**: 1494-500

214 Olsen GN BJ, Weiman DS. Stair climbing as an exercise test to predict the postoperative com- plications of lung resection: two years experience. Chest, 1991; 587-90

215 Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). *Eur J Anaesthesiol* 2014; **31**: 517-73

216 Biccard BM. Relationship between the inability to climb two flights of stairs and outcome after major non-cardiac surgery: Implications for the pre-operative assessment of functional capacity. *Anaesthesia* 2005; **60**: 588-93

217 Brunelli A, Xiume F, Refai M, et al. Peak oxygen consumption measured during the stair-climbing test in lung resection candidates. *Respiration* 2010; **80**: 207-11

218 Boujibar F, Gillibert A, Gravier FE, et al. Performance at stair-climbing test is associated with postoperative complications after lung resection: a systematic review and meta-analysis. *Thorax* 2020; **75**: 791-7

219 BALKE B. A SIMPLE FIELD TEST FOR THE ASSESSMENT OF PHYSICAL FITNESS. REP 63-6. *Rep Civ Aeromed Res Inst US* 1963: 1-8

220 Society AT. ATS Statement: Guidelines for the Six-Minute Walk Test. Am J Respir Crit Care Med, 2002; pp 111–7

221 Brunelli A, Pompili C, Salati M. Low-technology exercise test in the preoperative evaluation of lung resection candidates. *Monaldi Archives for Chest Disease - Pulmonary Series* 2010; **73**: 72-8

222 Bagg LR. The 12-min walking distance; its use in the pre-operative assessment of patients with bronchial carcinoma before lung resection. *Respiration* 1984; **46**: 342-5

223 Holden DA, Rice TW, Stelmach K, Meeker DP. Exercise testing, 6-min walk, and stair climb in the evaluation of patients at high risk for pulmonary resection. *Chest* 1992; **102**: 1774-9

224 Brocki BC, Andreasen JJ, Westerdahl E, De Souza DSR. Improvement in physical performance and health-related quality of life one year after radical operation for primary or secondary lung cancer. *European Respiratory Journal* 2015; **Conference**: European Respiratory Society Annual Congress 2015. Amsterdam Netherlands. Conference Publication: (var.pagings). 46 (SUPPL. 59) (no pagination)

225 RJ P, JM C, K S, CE B. Preoperative risk evaluation for lung cancer resection: predicted postoperative product as a predictor of surgical mortality.

. Am J Respir Crit Care Med, 1994; 947-55

226 Marjanski T, Wnuk D, Bosakowski D, Szmuda T, Sawicka W, Rzyman W. Patients who do not reach a distance of 500 m during the 6-min walk test have an increased risk of postoperative complications and prolonged hospital stay after lobectomy. *European Journal of Cardio-thoracic Surgery* 2015; **47**: e213-e9

227 Win T, Jackson A, Groves AM, et al. Relationship of shuttle walk test and lung cancer surgical outcome. *Eur J Cardiothorac Surg* 2004; **26**: 1216-9

228 Struthers R, Erasmus P, Holmes K, Warman P, Collingwood A, Sneyd JR. Assessing fitness for surgery: a comparison of questionnaire, incremental shuttle walk, and cardiopulmonary exercise testing in general surgical patients. *Br J Anaesth* 2008; **101**: 774-80

229 Win T, Jackson A, Groves AM, Sharples LD, Charman SC, Laroche CM. Comparison of shuttle walk with measured peak oxygen consumption in patients with operable lung cancer. *Thorax* 2006; **61**: 57-60

230 Arena R, Sietsema KE. Cardiopulmonary exercise testing in the clinical evaluation of patients with heart and lung disease. *Circulation* 2011; **123**: 668-80

231 Society/ AT, Physicians ACoC. ATS/ACCP Statement on Cardiopulmonary Exercise Testing. American Journal of Respiratory and Critical Care medicine, 2001

232 Moyes LH, McCaffer CJ, Carter RC, Fullarton GM, Mackay CK, Forshaw MJ. Cardiopulmonary exercise testing as a predictor of complications in oesophagogastric cancer surgery. *Ann R Coll Surg Engl* 2013; **95**: 125-30

233 Smith TP, Kinasewitz GT, Tucker WY, Spillers WP, George RB. Exercise capacity as a predictor of post-thoracotomy morbidity. *Am Rev Respir Dis* 1984; **129**: 730-4

234 Bechard W. Assessment of exercise oxygen consumption as preoperative criterion for lung resection. Ann Thorac Surg 1987;44(4):344–9. [PubMed: 3662680]. Ann Thorac Surg 1987; 344–9.

235 Benzo R, Kelley GA, Recchi L, Hofman A, Sciurba F. Complications of lung resection and exercise capacity: A meta-analysis. *Respiratory Medicine* 2007; **101**: 1790-7

236 Loewen GM, Watson D, Kohman L, et al. Preoperative exercise Vo2 measurement for lung resection candidates: Results of cancer and leukemia group B protocol 9238. *Journal of Thoracic Oncology* 2007; **2**: 619-25

237 Arena R, Myers J, Williams MA, et al. Assessment of functional capacity in clinical and research settings: a scientific statement from the American Heart Association Committee on Exercise, Rehabilitation, and Prevention of the Council on Clinical Cardiology and the Council on Cardiovascular Nursing. *Circulation* 2007; **116**: 329-43

238 Lim E, Beckles M, Warburton C, Baldwin D. Cardiopulmonary exercise testing for the selection of patients undergoing surgery for lung cancer: friend or foe? *Thorax* 2010; **65**: 847-9

239 Mittman C, Bruderman I. Lung cancer: to operate or not? *Am Rev Respir Dis* 1977; **116**: 477-96

240 Nagasaki F, Flehinger BJ, Martini N. Complications of surgery in the treatment of carcinoma of the lung. *Chest* 1982; **82**: 25-9

241 Franczuk M, Rudzinski P, Usiekniewicz J, et al. Six minute walk test (6MWT) in patients with diagnosis of lung cancer-Clinical value in physiological evaluation candidates for surgical treatment. *European Respiratory Journal* 2012; **Conference**: European Respiratory Society Annual Congress 2012. Vienna Austria. Conference Publication: (var.pagings). 40 (SUPPL. 56) (no pagination)

242 RAMS JJ, HARRISON RW, FRY WA, MOULDER PV, ADAMS WE. Operative pulmonary artery pressure measurements as a guide to postoperative management and prognosis following pneumonectomy. *Dis Chest* 1962; **41**: 85-90

243 Fee HJ, Holmes EC, Gewirtz HS, Ramming KP, Alexander JM. Role of pulmonary vascular resistance measurements in preoperative evaluation of candidates for pulmonary resection. *J Thorac Cardiovasc Surg* 1978; **75**: 519-24

244 Pierce RJ, Sharpe K, Johns J, Thompson B. Pulmonary artery pressure and blood flow as predictors of outcome from lung cancer resection. *Respirology (Carlton, Vic)* 2005; **10**: 620-8

245 Tisi GM. Preoperative evaluation of pulmonary function. Validity, indications, and benefits. *Am Rev Respir Dis* 1979; **119**: 293-310

246 Karliner JS, Coomaraswamy R, Williams MH. Relationship between preoperative pulmonary function studies and prognosis of patients undergoing pneumonectomy for carcinoma of the lung. *Dis Chest* 1968; **54**: 112-8

247 Harpole DH, Liptay MJ, DeCamp MM, Mentzer SJ, Swanson SJ, Sugarbaker DJ. Prospective analysis of pneumonectomy: risk factors for major morbidity and cardiac dysrhythmias. *Ann Thorac Surg* 1996; **61**: 977-82

248 Reilly JJ, Mentzer SJ, Sugarbaker DJ. Preoperative assessment of patients undergoing pulmonary resection. *Chest* 1993; **103**: 342S-5S

249 Group. BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69**: 89-95

250 Tayama K, Takamori S, Mitsuoka M, et al. Natriuretic peptides after pulmonary resection. *Annals of Thoracic Surgery* 2002; **73**: 1582-6

251 Çavuşoğlu Y, Alper AT, Altay H, et al. Natriuretic Peptides in Clinical Practice. *Anatol J Cardiol* 2019; **21**: 1-40

252 Tsai SH, Lin YY, Chu SJ, Hsu CW, Cheng SM. Interpretation and use of natriuretic peptides in non-congestive heart failure settings. *Yonsei Med J* 2010; **51**: 151-63

253 Mueller C, Scholer A, Laule-Kilian K, et al. Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea. *N Engl J Med* 2004; **350**: 647-54

254 Malhotra AK, Ramakrishna H. N-terminal pro B type natriuretic peptide in high cardiovascular-risk patients for noncardiac surgery: What is the current prognostic evidence? *Ann Card Anaesth* 2016; **19**: 314-20

255 Emma Murphy AG, Philip McCall, Ben Shelley. Relative kinetics of B-type natriuretic peptide and N-terminal pro B-type natriuretic peptide after lung resection. British Journal of Anaesthesia, 2019

256 Gurgo AM, Ciccone AM, D'Andrilli A, et al. Plasma NT-proBNP levels and the risk of atrial fibrillation after major lung resection. *Minerva Cardioangiologica* 2008; **56**: 581-5

257 Krüger S, Graf J, Merx MW, et al. Brain natriuretic peptide predicts right heart failure in patients with acute pulmonary embolism. *Am Heart J* 2004; **147**: 60-5

258 Blyth KG, Groenning BA, Mark PB, et al. NT-proBNP can be used to detect right ventricular systolic dysfunction in pulmonary hypertension. *Eur Respir J* 2007; **29**: 737-44

259 Wijeysundera DN, Pearse RM, Shulman MA, et al. Assessment of functional capacity before major non-cardiac surgery: an international, prospective cohort study. *Lancet* 2018; **391**: 2631-40

260 Karthikeyan G, Moncur RA, Levine O, et al. Is a pre-operative brain natriuretic peptide or N-terminal pro-B-type natriuretic peptide measurement an independent predictor of adverse cardiovascular outcomes within 30 days of noncardiac surgery? A systematic review and meta-analysis of observational studies. *J Am Coll Cardiol* 2009; **54**: 1599-606

261 Goldman L, Caldera DL, Nussbaum SR, et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. *N Engl J Med* 1977; **297**: 845-50

262 Dernellis J, Panaretou M. Assessment of cardiac risk before non-cardiac surgery: brain natriuretic peptide in 1590 patients. *Heart* 2006; **92**: 1645-50

263 Rodseth RN, Biccard BM, Le Manach Y, et al. The prognostic value of pre-operative and post-operative B-type natriuretic peptides in patients undergoing noncardiac surgery: B-type natriuretic peptide and N-terminal fragment of pro-B-type natriuretic peptide: a systematic review and individual patient data meta-analysis. *J Am Coll Cardiol* 2014; **63**: 170-80

264 Rodseth RN, Biccard BM, Chu R, et al. Postoperative B-type natriuretic peptide for prediction of major cardiac events in patients undergoing noncardiac surgery: Systematic review and individual patient meta-analysis. *Anesthesiology* 2013; **119**: 271-83

265 Januzzi JL, Chen-Tournoux AA, Moe G. Amino-terminal pro-B-type natriuretic peptide testing for the diagnosis or exclusion of heart failure in patients with acute symptoms. *Am J Cardiol* 2008; **101**: 29-38

266 Rost NS, Biffi A, Cloonan L, et al. Brain natriuretic peptide predicts functional outcome in ischemic stroke. *Stroke* 2012; **43**: 441-5

267 McCall PJ. The right ventricular response to lung resection. . University of Glasgow, 2018

268 McCall P LB, Shelley B. Don't forget the right ventricle. Journal of Cardiothoracic and Vascular Anaesthesia, Aug 2020; 2283-4.

269 Nojiri T, Maeda H, Takeuchi Y, et al. Predictive value of B-type natriuretic peptide for postoperative atrial fibrillation following pulmonary resection for lung cancer. *European Journal of Cardio-thoracic Surgery* 2010; **37**: 787-91

270 Simmers D, Potgieter D, Ryan L, Fahrner R, Rodseth RN. The use of preoperative B-type natriuretic peptide as a predictor of atrial fibrillation after thoracic surgery: Systematic review and meta-analysis. *Journal of Cardiothoracic and Vascular Anesthesia* 2015; **29**: 389-95

271 Pruszczyk P, Kostrubiec M, Bochowicz A, et al. N-terminal pro-brain natriuretic peptide in patients with acute pulmonary embolism. *Eur Respir J* 2003; **22**: 649-53

272 Kucher N, Printzen G, Goldhaber SZ. Prognostic role of brain natriuretic peptide in acute pulmonary embolism. *Circulation* 2003; **107**: 2545-7

273 Choi HS, Kim KH, Yoon HJ, et al. Usefulness of cardiac biomarkers in the prediction of right ventricular dysfunction before echocardiography in acute pulmonary embolism. *J Cardiol* 2012; **60**: 508-13

274 Kline JA, Zeitouni R, Marchick MR, Hernandez-Nino J, Rose GA. Comparison of 8 biomarkers for prediction of right ventricular hypokinesis 6 months after submassive pulmonary embolism. *Am Heart J* 2008; **156**: 308-14

275 Nagaya N, Nishikimi T, Uematsu M, et al. Plasma brain natriuretic peptide as a prognostic indicator in patients with primary pulmonary hypertension. *Circulation* 2000; **102**: 865-70

276 Galiè N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Heart J* 2016; **37**: 67-119

277 Excellence NIfHaC. Sepsis. 2017

278 System) TTTSS. Available from https://sfar.org/scores2/thoracoscore2.php

279 Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982; **5**: 649-55

280 Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken)* 2011; **63 Suppl 11**: S240-52

281 Grutters JP, Joore MA, Wiegman EM, et al. Health-related quality of life in patients surviving non-small cell lung cancer. *Thorax* 2010; **65**: 903-7

282 Pompili C, Koller M, Velikova G, et al. EORTC QLQ-C30 summary score reliably detects changes in QoL three months after anatomic lung resection for Non-Small Cell Lung Cancer (NSCLC). *Lung Cancer* 2018; **123**

283 Shulman MA, Myles PS, Chan MTV, McIlroy DR, Wallace S, Ponsford J. Measurement of Disability-free Survival after Surgery. *Anesthesiology: The Journal of the American Society of Anesthesiologists* 2018; **122**: 524-36

284 Park S, Kang CH, Hwang Y, et al. Risk factors for postoperative anxiety and depression after surgical treatment for lung cancer. *European Journal of Cardio-thoracic Surgery* 2016; **49**: e16-e21

285 Kumar SP. Utilization of brief pain inventory as an assessment tool for pain in patients with cancer: a focused review. *Indian J Palliat Care* 2011; **17**: 108-15

286 Society IC. ICS 2009, Levels of Critical Care for Adult Patients. 2009

287 Fernandez FG, Falcoz PE, Kozower BD, Salati M, Wright CD, Brunelli A. The Society of Thoracic Surgeons and the European Society of Thoracic Surgeons general thoracic surgery databases: joint standardization of variable definitions and terminology. *Ann Thorac Surg* 2015; **99**: 368-76

288 Kalil AC, Metersky ML, Klompas M, et al. Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis* 2016; **63**: e61-e111

289 Chapman AR, Adamson PD, Mills NL. Assessment and classification of patients with myocardial injury and infarction in clinical practice. *Heart* 2017; **103**: 10-8

290 Journal BM. Correlation and regression. Available from http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression (accessed 4th October 2019)

291 DQ H, S L. *Applied Logistic Regression*. John Wiley and Sons, New York, 2000

292 Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, initiative TGEdtapmotS. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; **17**: 230

293 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**: W1-73

294 Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; **74**: 167-76

295 Allison P. Why I Don't Trust the Hosmer-Lemeshow test for Logistic Regression. Available from https://statisticalhorizons.com/hosmer-lemeshow (accessed June 2021 2021)

296 Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol* 2010; **63**: 938-9; author reply 9

297 Leening MJ, Vedder MM, Witteman JC, Pencina MJ, Steyerberg EW. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. *Ann Intern Med* 2014; **160**: 122-31

298 Box GEP, & Tidwell, P. W. Transformation of the independent variables. *Technometrics*, 1962; 531-50

299 Tabachnick BG, & Fidell, L. S. *Using multivariate statistics (6th ed.)*. Harlow, England: Pearson., 2014

300 Hosmer DW, Jr., Lemeshow, S., & Sturdivant, R. X. *Applied logistic regression (3rd ed.)*. 3rd Edn.: Wiley., 2013

301 Tan LE, Razak AM, Lim CS. Association of chronic obstructive pulmonary disease and postresection lung cancer survival: A systematic review and meta-analysis. *Journal of Investigative Medicine* 2017; **65**: 342-52

302 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159-74

303 Kulason K, Nouchi R, Hoshikawa Y, Noda M, Okada Y, Kawashima R. The beneficial effects of cognitive training with simple calculation and reading aloud (SCRA) in the elderly postoperative population: A pilot randomized controlled trial. *Frontiers in Aging Neuroscience* 2018; **10 (MAR) (no pagination)**

304 Salati M, Brunelli A, Decaluwe H, et al. Report from the European Society of Thoracic Surgeons Database 2017: patterns of care and perioperative outcomes of surgery for malignant lung neoplasm. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* 2017; **52**: 1041-8

305 Reeves GK, Pirie K, Beral V, et al. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007; **335**: 1134

306 Williams T, Gulack BC, Kim S, Fernandez FG, Ferguson MK. Operative Risk for Major Lung Resection Increases at Extremes of Body Mass Index. *Ann Thorac Surg* 2017; **103**: 296-302

307 Sarna L, Evangelista L, Tashkin D, et al. Impact of respiratory symptoms and pulmonary function on quality of life of long-term survivors of non-small cell lung cancer. *Chest* 2004; **125**: 439-45

308 Amar D, Zhang H, Tan KS, Piening D, Rusch VW, Jones DR. A brain natriuretic peptide-based prediction model for atrial fibrillation after thoracic surgery: Development and internal validation. *J Thorac Cardiovasc Surg* 2019; **157**: 2493-9.e1

309 Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006; **332**: 1080

310 Wang C, Guo M, Zhang N, Wang G. Association of body mass index and outcomes following lobectomy for non-small-cell lung cancer. *World J Surg Oncol* 2018; **16**: 90

311 Mytelka DS, Li L, Benoit K. Post-diagnosis weight loss as a prognostic factor in non-small cell lung cancer. *J Cachexia Sarcopenia Muscle* 2018; **9**: 86-92

312 De Santi F, Zoppini G, Locatelli F, et al. Type 2 diabetes is associated with an increased prevalence of respiratory symptoms as compared to the general population. *BMC Pulm Med* 2017; **17**: 101

313 Lee CT, Mao IC, Lin CH, Lin SH, Hsieh MC. Chronic obstructive pulmonary disease: a risk factor for type 2 diabetes: a nationwide population-based study. *Eur J Clin Invest* 2013; **43**: 1113-9

314 Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet* 2007; **370**: 797-9

315 Hall GC, Roberts CM, Boulis M, Mo J, MacRae KD. Diabetes and the risk of lung cancer. *Diabetes Care* 2005; **28**: 590-4

316 Luo J, Hendryx M, Qi L, Ho GY, Margolis KL. Pre-existing diabetes and lung cancer prognosis. *Br J Cancer* 2016; **115**: 76-9

317 Sundström J, Schön TB. Machine Learning in Risk Prediction. *Hypertension* 2020; **75**: 1165-6

318 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128-38

319 Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ* 2015; **351**: h3868

320 Puhr R, Heinze G, Nold M, Lusa L, Geroldinger A. Firth's logistic regression with rare events: accurate effect estimates and predictions? *Stat Med* 2017; **36**: 2302-17

321 Ambler G, Seaman S, Omar RZ. An evaluation of penalised survival methods for developing prognostic models with rare events. *Stat Med* 2012; **31**: 1150-61

322 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338**: b605

323 Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009; **338**: b606

324 Siregar S, Nieboer D, Versteegh MIM, Steyerberg EW, Takkenberg JJM. Methods for updating a risk prediction model for cardiac surgery: a statistical primer. *Interact Cardiovasc Thorac Surg* 2019; **28**: 333-8

325 Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model. *Eur J Cardiothorac Surg* 2018; **54**: 203-8

326 Beck JR, Pauker SG. The Markov process in medical prognosis. *Med Decis Making* 1983; **3**: 419-58

327 Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993; **13**: 322-38

328 Im Y, Park HY, Shin S, et al. Prevalence of and risk factors for pulmonary complications after curative resection in otherwise healthy elderly patients with early stage lung cancer. *Respir Res* 2019; **20**: 136

329 Agostini PJ, Lugg ST, Adams K, et al. Risk factors and short-term outcomes of postoperative pulmonary complications after VATS lobectomy. *J Cardiothorac Surg* 2018; **13**: 28

330 Prowle JR, Forni LG, Bell M, et al. Postoperative acute kidney injury in adult non-cardiac surgery: joint consensus report of the Acute Disease Quality Initiative and PeriOperative Quality Initiative. *Nat Rev Nephrol* 2021; **17**: 605-18

331 Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009; **338**: b604

332 Nojiri T, Inoue M, Yamamoto K, et al. B-type natriuretic peptide as a predictor of postoperative cardiopulmonary complications in elderly patients undergoing pulmonary resection for lung cancer. *Annals of Thoracic Surgery* 2011; **92**: 1051-5

333 Nojiri T, Inoue M, Shintani Y, et al. B-type natriuretic peptide-guided risk assessment for postoperative complications in lung cancer surgery. *World J Surg* 2015; **39**: 1092-8

334 Qiu S, Cai X, Sun Z, et al. Heart Rate Recovery and Risk of Cardiovascular Events and All-Cause Mortality: A Meta-Analysis of Prospective Cohort Studies. *J Am Heart Assoc* 2017; **6**

335 Lederman D, Easwar J, Feldman J, Shapiro V. Anesthetic considerations for lung resection: preoperative assessment, intraoperative challenges and postoperative analgesia. *Ann Transl Med* 2019; **7**: 356

336 Ansari BM, Hogan MP, Collier TJ, et al. A Randomized Controlled Trial of High-Flow Nasal Oxygen (Optiflow) as Part of an Enhanced Recovery Program After Lung Resection Surgery. *Ann Thorac Surg* 2016; **101**: 459-64

337 Fennelly J, Potter L, Pompili C, Brunelli A. Performance in the shuttle walk test is associated with cardiopulmonary complications after lung resections. *Journal of Thoracic Disease* 2017; **9**: 789-95

338 Bayram AS, Candan T, Gebitekin C. Preoperative maximal exercise oxygen consumption test predicts postoperative pulmonary morbidity following major lung resection. *Respirology* 2007; **12**: 505-10

339 Bobbio A, Chetta A, Internullo E, et al. Exercise capacity assessment in patients undergoing lung resection. *European Journal of Cardio-thoracic Surgery* 2009; **35**: 419-22

340 Bolliger CT, Jordan P, Solèr M, et al. Exercise capacity as a predictor of postoperative complications in lung resection candidates. *Am J Respir Crit Care Med* 1995; **151**: 1472-80

341 Brat K, Tothova Z, Merta Z, et al. Resting End-Tidal Carbon Dioxide Predicts Respiratory Complications in Patients Undergoing Thoracic Surgical Procedures. *Annals of Thoracic Surgery* 2016; **102**: 1725-30

342 Brunelli A, Belardinelli R, Refai M, et al. Peak oxygen consumption during cardiopulmonary exercise test improves risk stratification in candidates to major lung resection. *Chest* 2009; **135**: 1260-7

343 Brunelli A, Pompili C, Salati M, et al. Preoperative maximum oxygen consumption is associated with prognosis after pulmonary resection in stage i non-small cell lung cancer. *Annals of Thoracic Surgery* 2014; **98**: 238-42

344 Larsen KR, Lund JO, Svendsen UG, Milman N, Petersen BN. Prediction of post-operative cardiopulmonary function using perfusion scintigraphy in patients with bronchogenic carcinoma. *Clinical Physiology* 1997; **17**: 257-67

345 Nagamatsu Y, Sueyoshi S, Tsubuku T, Kawasaki M, Akagi Y. Predicting postoperative exercise capacity after major lung resection. *Surgery Today* 2014; **45**: 1501-8

346 Nagamatsu Y, Shima I, Hayashi A, Yamana H, Shirouzo K, Ishitake T. Preoperative Spirometry Versus Expired Gas Analysis during Exercise Testing as Predictors of Cardiopulmonary Complications after Lung Resection. *Surgery Today* 2004; **34**: 107-10

347 Pompili C, Brunelli A. Quality of life after lung resection is not associated with functional objective measures. *European Respiratory Journal* 2013; **42**: 283-5

348 Villani F, Busia A. Preoperative evaluation of patients submitted to pneumonectomy for lung carcinoma: Role of exercise testing. *Tumori* 2004; **90**: 405-9

349 Wang JS, Abboud RT, Evans KG, Finley RJ, Graham BL. Role of CO diffusing capacity during exercise in the preoperative evaluation for lung resection. *American Journal of Respiratory and Critical Care Medicine* 2000; **162**: 1435-44