Early Stage Non-Small Cell Lung Cancer's Treatment Journey: Patients' Perspectives

Cecilia Pompili

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Medicine

July 2019

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Publications

Chapter 1

Pompili C, Absolom K, Velikova G, Backhus L. Patients reported outcomes in thoracic surgery. J Thorac Dis. 2018 Feb;10(2):703-706.

This publication highlights the background of all the work of this thesis, describing the current status of the PROMS collection in thoracic surgery. It has been described in Chapter 1. I have done a broader review of the surgical literature and design of the manuscript. I prepared the manuscript for publication. Other authors contributed to the manuscript writing.

Chapter 1-2

Koller M, Hjermstad MJ, Tomaszewski KA, Tomaszewska IM, Hornslien K, Harle A, Arraras JI, Morag O, Pompili C, Ioannidis G, Georgiou M, Navarra C, Chie WC, Johnson CD, Himpel A, Schulz C, Bohrer T, Janssens A, Kulis D, Bottomley A; EORTC Quality of Life Group, EORTC Lung Cancer Group, and European Society of Thoracic Surgeons. An international study to revise the EORTC questionnaire for assessing quality of life in lung cancer patients. Ann Oncol. 2017 Nov 1;28(11):2874-2881.

This publication reports on publications is part of wider collaborative research I am involved in. The data reported in the publication is not a direct outcome of this PhD work but the findings helped guide the choice of measure in the LILAC prospective study. Some of this work is briefly described in Chapter 1 and 2 in particular. I had a role in the data collection (questionnaire-related patient interviews) and contributed to the manuscript. Other authors contributed to the data collection and analysis, and the preparation of the manuscript.

Chapter 3

Pompili C, Absolom K, Franks K, Velikova G. Are quality of life outcomes comparable following stereotactic radiotherapy and minimally invasive surgery for stage I lung cancer patients? J Thorac Dis. 2018 Dec;10(12): 7055-7063.

This publication reports on the results of the Best Evidence Topic. The results from this review are described in Chapter 3. I was responsible for data collection and analysis and the preparation of the manuscript. Other authors advised on the review topic, double checked findings and assisted with preparation of the manuscript.

Chapter 5

Pompili C, Velikova G, White J, Callister M, Robson J, Dixon S, Franks K, Brunelli A. Poor preoperative patient-reported quality of life is associated with complications following pulmonary lobectomy for lung cancer. Eur J Cardiothorac Surg. 2017 Mar 1;51(3):526-531.

This publication reports the findings of the association of quality of life and postoperative complications in surgical lung cancer patients. The results of this work are described in Chapter 4. I had a lead role in the study design, data collection and analysis

of this study. I prepared the manuscript for publication. Other authors contributed to study design, data collection and analysis and contributed to the manuscript.

Chapter 5

Pompili C, Shargall Y, Decaluwe H, Moons J, Chari M, Brunelli A. Risk-adjusted performance evaluation in three academic thoracic surgery units using the Eurolung risk models. Eur J Cardiothorac Surg. 2018 Jul 1;54(1):122-126.

This publication is part of my work within the ESTS database project and describes the preliminary results of the use of Eurolung Risk scores to stratify patients and compare units. Part of the results has described in Chapter 5. I had a lead role in the study design, data analysis of this study. I prepared the manuscript for publication and presented it an oral presentation during the ESTS Conference. Other authors contributed to study design and contributed to the manuscript.

Chapter 5

Pompili C, Koller M, Velikova G, Franks K, Absolom K, Callister M, Robson J, Imperatori A, Brunelli A. 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 Sep; 123:149-154

This publication describes how the EORTC Summary Score could be used to evaluate the Quality of Life in lung cancer surgical patients, and describes its sensitivity to detect changes in this population. I had a lead role in the study design, data collection and analysis. I prepared the manuscript for publication. Other authors contributed to study design, data collection and analysis and contributed to the manuscript.

Chapter 6

Pompili C, Franks KN, Brunelli A, Hussain YS, Holch P, Callister ME, Robson JM, Papagiannopoulos K, Velikova G. Patient reported outcomes following video assisted thoracoscopic (VATS) resection or stereotactic ablative body radiotherapy (SABR) for treatment of non-small cell lung cancer: protocol for an observational pilot study (LiLAC). J Thorac Dis. 2017 Aug;9(8): 2703-2713.

This publication gives a full report of the LILAC Protocol. This work is described in Chapter 7. I had a lead role in the study design, ethical approval and I prepared the manuscript for publication. Other authors contributed to study design and contributed to the manuscript.

Acknowledgements

This research has been carried out within the Patient Centred Outcomes Research (PCOR) group based at St James Institute of Oncology, Leeds, UK. Current members which have been involved in work described in this thesis include Prof Galina Velikova, Dr Penny Wright (PW), Dr Kate Absolom (KA), Marie Holmes (MH), Beverly Clayton (BC), Zoe Rogers (ZR), Lorraine Warrington (LW), Andrea Gibson (AG), Sarah Dickinson (SD), Florien Boele (FB), Robert Carter and Jeremy Dwyer. Past members who have been involved include Dr Trish Holch (TH), Emma Smyllie (ES) and Dr Simon Pini (SP).

Firstly, I would like to profoundly thank my excellent supervisory team. I would like to thank Prof Galina Velikova for giving me the opportunity to do this PhD, and for inspiring and mentoring me as a clinical academic. I am immensely grateful for all the support, advices and to have been able to understand deeply the challenges in pursuing this career. I am sincerely grateful to Dr Kate Absolom for her daily support in all the aspect of this PhD and to have guided me closely from the beginning in this academic field. I would like to thank Dr Kevin Franks for his valuable expertise, particularly as he made me also a better clinician in showing that lung cancer can be treated not only with surgery.

I would like to thank the entire PCOR team, because they made me feel supported not only for the recruitment but also helped my resilience especially in last month's when needed after overwhelming clinical days. I am eternally grateful to Sarah, Lorraine, Marie, Andrea and Bev for all their patience and help. Thank you also to Dr Penny Wright

who has always been next to me physically and mentally. The preparation of this thesis would never have been possible without their constructive suggestions, continual encouragement and assistance.

A huge thank you to Dr Trish Holch who has provided lots of encouragements over the years. A special remark goes to Prof Hilary Bekker, who helped in integrating the decision-making concepts in this work with professional enthusiasm and expertise.

I would like to express my gratitude to all of the patients and health care professionals who took part in this research, for giving up their time and for sharing their thoughts and experiences. In particular, thank you to Dr Matthew Callister and Dr Jonathan Robson for always been supportive with my research. A special mention to Sandra Dixon, thoracic CNS who practically supported all the studies described in this thesis.

Thanks to Francesca, to have given positivity to these years. Thank you to Alex, who immensely supported our family during this journey but most importantly who first showed me the importance and the value of research. And thanks to Edoardo, my son, who grew up with a surgeon mum doing a PhD. I hope this thesis, will represent for him an inspiration to never stop studying and learning in life. I dedicate this thesis to them.

Abstract

Background

Video assisted thoracoscopic surgery (VATS) is becoming the standard of care for early-stage non-small-cell lung cancer (NSCLC). Stereotactic Ablative Body Radiotherapy (SABR) is an accepted alternative for medically inoperable patients

Aims

The primary objectives of this thesis were to assess the role of Patients Reported Outcomes Measures (PROMs) during the first year following VATS or SABR treatment for early-stage NSCLC and the potential of Quality of Life (QOL) in predicting postoperative outcomes.

Mixed method

Preliminary studies

Prior to delivering a prospective longitudinal observational study, important practical and methodological issues were addressed. Analysis from a systematic review, found there is no sufficient standardized data describing the evolution of QOL after VATS or SABR. Analysis from a service evaluation project of 330 patients demonstrated it was feasible to collect PROMS data in surgical patients in clinical practice and that preoperative QOL information is associated with postoperative clinical outcomes. The EORTC SumSc was evaluated and found to be sensitive to detecting changes in the perioperative period. Electronic methods for the collection and presentation of PROMS data were implemented to improve clinical data capture from electronic health records (EHR).

Prospective study (Lilac study)

Interim analysis (n=225) from this prospective evaluation of QOL demonstrated that QOL did not change significantly for SABR patients although baseline scores were significantly worse. The surgical population experienced a worsening in respiratory symptoms by six-months. Qualitative analysis of patient interviews found the Lilac study was acceptable to patients and revealed low electronic PROMS completion rate might be due to inherent demographic differences in the patient population.

Conclusions

QOL has the potential to add important information in the decision-making process of early-stage NSCLC patients not otherwise captured by traditional objective parameters. Further analysis of the 12-month results and staff involvement using qualitative and quantitative assessments will provide insights into motivators and barriers of adoption in clinical practice.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AE	Adverse Event
BET	Best Evidence Topic
CNS	Clinical Nurse Specialist
COPD	Chronic Obstructive Pulmonary Disease
CSES	Chronic disease Self efficacy Scale
CTCAE	Common Terminology Criteria for Adverse Events
DPA	Data Protection Act
EORTC-QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire- Core 30
EPR	Electronic Patient Record
ePROMS	Electronic or web-based Patient Reported Outcome Measures
eRAPID	electronic patient self-Reporting of Adverse-events: Patient Information and aDvice
ESTS	European Society of Thoracic Surgeons,
FEV1	Forced Expiratory Volume in the First second
DLCO	Diffusing capacity of the lungs for carbon monoxide
GCP	Good Clinical Practice
GP	General Practitioner
HRQOL	Health Related Quality of Life
NHS	National Health Service
NSCLC	Non-small cell lung cancer
PCOR	Patient Centred Outcomes Research

PICOS	Population, Intervention, Comparator, Outcome, Study design
PPM	Patient Pathway Manager
PROS	Patient Reported outcomes
PROMS	Patient Reported Outcome Measures
PROSPERO	International prospective register of systematic reviews
QOL	Quality of Life
RCP	Royal College of Physicians
RCT	Randomised Controlled Trial
SABR	Stereotactic Ablative Body Radiotherapy
SDM	Shared Decision Making
VATS	Video Assisted Thoracoscopic Surgery

Chapter 1 Introduction

1.1 Lung Cancer Epidemiology

Worldwide, lung cancer remains the most common cancer diagnosed and greatest cause of cancer related death.

Lung cancer is the third most common cancer in the UK, accounting for 13% of all new cancer cases (2015) (1). In males in the UK, lung cancer is the second most common cancer (13% of all new male cancer cases). In females in the UK it is the second most common cancer (12% of all new female cancer cases). Most importantly, it accounts for 21% of all cancer deaths in this country.

In the 2005 NICE Lung Cancer Guidelines, deaths from lung cancer were the most common cause of cancer related deaths in men, and the second most common cause in women. However, lung cancer has since become the commonest cause of cancer related death in both sexes. This is reflecting a change in smoking habits among females: the peak prevalence of smoking in young women was only reached in the 1990's, and so the incidence of lung cancer amongst older women has only recently stabilised considering an average gap of 20 years between smoking and lung cancer diagnosis (2). The sharp decrease in the incidence of male lung cancer over the past two decades reflects the earlier decline in smoking prevalence among men and the combined increase in females. The UK incidence of lung cancer in men is similar to most of Western Europe and lower than most of Eastern Europe. The incidence in women is amongst the highest in the whole of the European Union (3).

However, recent published statistics have projected that lung cancer mortality rates are to fall by 21% in the UK between 2014 and 2035. This will implicitly include changes in cancer risk factors, diagnosis and treatment(1).

Lung cancer is uncommon in people younger than 40, with more than 75% of lung cancers diagnosed in people over the age of 65. More specifically a study funded by the British Lung Foundation break down the incidence by age: In 2012, only six people for every 100,000 had lung cancer among those aged 31–40, rising steadily to 23 per 100,000 among those aged 41–50, peaking at 631 per 100,000 among those aged 71–80, and 666 per 100,000 in those aged 81 and over(4).

1.2 Lung Cancer types

Lung cancers are classified into two types: small-cell lung cancers (SCLC) and non-small-cell lung cancers (NSCLC).

SCLC account for about 20 % of cases. SCLC is more likely than NSCLC to have spread by the time of diagnosis. SCLC is the most aggressive form of the disease, having greater potential to metastasise than other types of lung cancer. Nearly all patients (over 95%) diagnosed with SCLC are current or ex-smokers(5). The main treatment is systemic therapy with minimal option for surgery and radiotherapy. For this reason, this type of cancer was not considered for inclusion in this thesis which focused on surgical and stereotactic options.

NSCLC accounts for the other 80% of lung cancer cases. This is a heterogeneous cancer group which has recently been re-classified. In 2015, the World Health Organisation (WHO) Classification of Tumours of the Lung, Pleura, Thymus and Heart was published with updated nomenclature and sub-division. It clearly describes a trend towards a more

personalised medicine, where therapeutic decisions are based on the specific histologic and genetic characteristics of the patient's tumour (6). NSCLC comprises approximately 80–85% of all lung cancers with adenocarcinoma (ADC, approximately 40–50% of cases) and squamous cell carcinoma (SqCC, approximately 20–30% of cases) comprising the predominant histological sub-types of NSCLC (7). The main update on the sub-classification of these cancers was related to these two sub-types: the reclassification of "bronchioloalveolar carcinoma" into several different sub-types, including adenocarcinoma in situ, minimally invasive ADC, and ADC with lepidic growth pattern, (3) the sub-classification of SqCC into basaloid, keratinizing, and non-keratinizing types, and the requirement of immunohistochemical evidence of squamous differentiation in non-keratinizing SqCC(6).

This thesis focused on NSCLC because it is the most prevalent. Furthermore, the main treatment options focused on local cancer eradication (surgery and RT) whereas the treatment approach for SCLC is mainly systemic (chemotherapy).

The treatment options for NSCLC are described below.

1.3 NSCLC Staging and Treatment

In 2016, the Union for International Cancer Control/American Joint Committee on Cancer/International Association for the Study of Lung Cancer (UICC/AJCC/IASLC) Staging Classification published the 8th Edition of the International Staging of Thoracic Malignancies (8). The TNM staging details are described in Figure 1-1 and Figure 1-2.

Figure 1-1 TNM Staging 8th Edition

T / M	Subcategory	N0	N1	N2	N3
T1	T1a	IA1	IIB	IIIA	IIIB
	T1b	IA2	IIB	IIIA	IIIB
	T1c	IA3	IIB	IIIA	IIIB
T2	T2a	IB	IIB	IIIA	IIIB
	T2b	IIA	IIB	IIIA	IIIB
T3	T3	IIB	IIIA	IIIB	IIIC
T4	T4	IIIA	IIIA	IIIB	IIIC
M1	M1a	IVA	IVA	IVA	IVA
	M1b	IVA	IVA	IVA	IVA
	M1c	IVB	IVB	IVB	IVB

Figure 1-2 TNM Tumour Characteristics

TNM 8th - Primary tumor characteristics	
T_x	Tumor in sputum/bronchial washings but not be assessed in imaging or bronchoscopy
T_0	No evidence of tumor
T_{is}	Carcinoma in situ
T_1	≤ 3 cm surrounded by lung/visceral pleura, not involving main bronchus
$T_{1a(mi)}$	Minimally invasive carcinoma
T_{1a}	≤ 1 cm
T_{1b}	> 1 to ≤ 2 cm
T_{1c}	> 2 to ≤ 3 cm
T_2	> 3 to ≤ 5 cm *or* involvement of main bronchus without carina, regardless of distance from carina or invasion visceral pleural *or* atelectasis or post obstructive pneumonitis extending to hilum
T_{2a}	>3 to ≤4cm
T_{2b}	>4 to ≤5cm
T_3	>5 to ≤7cm in greatest dimension *or* tumor of any size that involves chest wall, pericardium, phrenic nerve *or* satellite nodules in the same lobe
T_4	> 7cm in greatest dimension *or* any tumor with invasion of mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, oesophagus, spine *or* separate tumor in different lobe of ipsilateral lung
N_1	Ipsilateral peribronchial and/or hilar nodes and intrapulmonary nodes
2	Ipsilateral mediastinal and/or subcarinal nodes
3	Contralateral mediastinal or hilar; ipsilateral/contralateral scalene/ supraclavicular
M_1	Distant metastasis
M_{1a}	Tumor in contralateral lung or pleural/pericardial nodule/malignant effusion
M_{1b}	Single extrathoracic metastasis, including single non-regional lymphnode
M_{1c}	Multiple extrathoracic metastases in one or more organs

TNM staging of lung cancer defines the extent of disease and consequently assigns prognosis and guides treatment. TNM staging has three components: the features/extent of the primary tumour (T), regional lymph node(s) involvement (N), and distant metastases (M). This edition re-categorizes the tumour size and other non-quantitative tumour descriptors (T), and further sub-classifies extra-thoracic metastases (M). The clinical nodal (N) classifier is unchanged as the earlier version correlates well with prognosis. The survival is inversely proportional to every centimetre increase in tumour size up to 7cm, where the same prognosis as T4 disease is reached.

The focus of this thesis is the clinical stage I which is defined by the previous table as a tumour less than 4cm in size, without atelectasis/pneumonitis and/or involvement of main bronchus, irrespective of distance to main carina, no visceral pleural involvement and no lymph node spread. The reason to focus on this stage, is because although surgical indications are extended to Stage II, SABR treatment is still limited to the Stage I with the exception of tumour size >4 but <5cm which is considered Stage IIa (9).

In 2014, the UK Office for National Statistics (ONS) reported that patients diagnosed with distant metastatic disease (stage IV) had a 1-year survival rate of just 15–19% compared with 81–85% for stage I (10). If identified at an early-stage, surgical resection of NSCLC offers a favourable prognosis, with five-year survival rates of 70–90% for small, localised tumours (stage I). However, most patients (approx. 75%) have advanced disease at the time of diagnosis (stage III/IV) and despite significant developments in the oncological management of late stage lung cancer over recent years, survival remains poor(11). From the most up to date guidelines, surgery is still reserved for stage I-II and selected cases of stage III NSCLC if the tumour is found to be resectable and the patient is able to tolerate surgery (12).

However, the last National Lung Cancer Audit in UK in 2016 , found that 37% of patients are alive at least one year after diagnosis, which is a significant improvement to the 31% diagnosed in 2010(13). The main finding was there has been a further increase in the number of patients receiving surgery with 17.5% of NCSLC patients diagnosed in 2016 received surgery compared to 16.7% of patients diagnosed in 2010.

If surgery is not an option, other treatments are available, from the most advanced radiation techniques like Stereotactic Ablative Body Radiotherapy (SABR) to chemotherapy and targeted therapy. With the advancement of genetics and biomarkers

testing, specific mutations have been identified to better target treatment for individual patients.

This thesis will focus on early stage NSCLC (I-II) as they are the most commonly treated with surgery or SABR.

1.4 Early stage Non-small cell lung cancer treatment

The diagnosis of stage I NSCLC is usually made in the absence of tumour-related symptoms. Poor baseline performance status and quality of life (QoL) is mostly caused by comorbidities such as chronic obstructive pulmonary disease (COPD)(14).

The current treatment strategy for NSCLC depends on clinical staging. Surgical resection is generally considered the treatment of choice in patients with stage I and II disease whose performance status allows for general anaesthesia and a lung resection(15, 16).

It is known that lung cancer patients commonly have multiple commodities with 54% presenting with three or more. These additional comorbidities are associated with poor survival outcome in lung cancer patients (17).

For early stage node negative lung cancer surgery is the current recommended treatment for fit patients(15) however, according to the National Lung Cancer Audit (NLCA) 2017, only 17.5% of all patients with lung cancer underwent surgery in 2016. It still shows great variations between centres not explained by case-mix adjustment. Adjusted surgical resection rates varied from 4.8% to 40.1% across the UK and 60 organisations failed to meet the audit standard of 17%(13). The treatments for Stage I NSCLC are investigated in this thesis.

1.4.1 Surgery

Surgical resection remains the standard of care for functionally operable early stage NSCLC (15). The principal aim of surgical treatment for NSCLC is to obtain a complete resection with negative margins. Lobectomy is the standard recommended operation for Stage I NSCLC. It consists of the surgical removal of a pulmonary lobe (upper, middle or lower) and accounts for more than half of lung resections in most specialized thoracic units. Bilobectomy is performed exclusively for right lung cancers (right upper and middle lobectomy, or right middle and lower lobectomy).

Pneumonectomy is the largest lung resection possible. It consists of the resection of a whole lung. It is mostly used in patients with central tumours, involvement of a mainstem bronchus, left or right pulmonary artery, and both superior and inferior pulmonary vein. It has limited use in the treatment of early-stage NSCLC.

Intra-operative staging is also crucial during the operation to decide on the extent of resection according to the intra-operative tumour (T) and nodal (N) status. At present, it is recommended that all patients considered for active treatment have a pre-operative positron emission tomography (PET)-CT. This has helped in defining standardised pathways and guide pre-operative nodal staging(18). This is particularly important as surgery, compared to other treatments, has the obvious advantage of pathological confirmation and further mediastinal staging.

Systematic nodal dissection is generally advocated to evaluate the hilar and mediastinal lymph nodes(19). It implies removal of, at least, three hilar and interlobar nodes and three mediastinal nodes from three stations in which the subcarinal is always included.

Modern advances in diagnostics and screening programs have contributed to an increase in the number of patients presenting with smaller cancers. Also, improvement

in peri-operative management has increased the rate of elderly and medically high-risk patients undergoing lung resection. These are the reasons why the extent of resection for early lung cancer also remains a matter of debate in our field.

Lobectomy remains the preferred operation and is associated with better survival and lower locoregional recurrence, but there is increased interest in the role of sublobar resections(15). Sublobar resections include anatomical (segmentectomies) and non-anatomical (wedge) resections. These may preserve more lung function compared to lobar resections, and may be especially useful in patients with marginal pulmonary function. Sublobar resections do have an important role in those select candidates. Patients with other comorbidities prohibitive for lobectomy should also be considered for a sublobar resection(20, 21). Oncologically however, there is a big debate and growing interest in these types of operations, which have extended surgery to previously unfit patients(22). The most important recommendation so far is related to the resection margins: sub-lobar resections should only be performed if there is confidence that adequate negative margins can be obtained(23).

1.4.1.1 Surgical approach: Open Vs Minimally-invasive (VATS)

There is significant debate about the most effective approach for lung cancer. Minimally invasive techniques like video-assisted (VATS) and robot-assisted thoracoscopic surgery (RATS) have been shown to reduce post-operative complications and shorten hospitalisation.

Compared to the standard open lobectomy procedure, VATS is a minimally invasive procedure with lower complication rates and shorter hospital stays compared to the traditional approach (24, 25). The benefits of VATS are particularly evident in high-risk patients (as per ACCP guidelines(26)). VATS has gained increasing popularity over the

last few decades to diagnose and treat lung cancer. In higher risk patients, a more limited anatomical sub-lobar resection or wedge excision may be performed (27, 28). Although it has been said that the period of recovery post-operation is shorter after undergoing a VATS, little is known about how VATS patients fare in terms of their QOL compared to patients who have undergone traditional open resections (29-31). As such, more studies need to be undertaken to explore the post-operative impact of VATS and open lung resections on patients' daily lives.

1.4.2 SABR

Patients with potentially curable NSCLC, who may be at high operative risk due to comorbidity, have potential for achieving disease control and survival comparable to surgery with SABR (9, 32-35).

This approach has been developed thanks to major advances in technology, allowing much higher doses to be delivered safely and accurately over a significantly shorter treatment time and with lower toxicity. SABR is an outpatient treatment with three to eight treatments given according to tumour location using nationally agreed guidelines(35). Indeed, SABR is now a recognised standard of care in early peripherally located inoperable Stage I lung cancer. In the UK, SABR doses for NSCLC aim to deliver between 54 and 60 Gy in three to eight fractions on an alternate-day basis(35).

One of the indications of SABR is the peripheral position of the tumour: defined as needing to be >2 cm away in all directions from the central airways (trachea, carina and main bronchus up to the division of the second order bronchi). To confidently plan lung SABR, planning margins to account for tumour motion should be individualised to each patient. The most common method to achieve this is a four-dimensional CT planning scan, where the tumour motion is assessed through multiple phases of breathing. A

recent review demonstrated the combined reported a median overall survival (OS) from 15 studies was 38.44 months, with an average follow-up of 29.4 months(36). Furthermore, it has been shown that the treatment is tolerable with high rates of local control (~90%)(37-39). In terms of toxicity, Grade 5 were rare and Grade 1 and 2 toxicities were very common and reported in up to 100% of cases, especially Grades 1–2 fatigue(36).

Data suggest that outcomes are superior to those achieved with conventionally fractionated radiotherapy(40).

There remains equipoise regarding the effectiveness of SABR compared with surgical resection in higher-risk patients and medically operable patients, which hopefully will be addressed by randomised clinical trials. Initial retrospective and prospective investigations have not been able to definitively conclude that SABR outcomes for operable patients are equivalent to surgery.

Due to poor accrual, early randomised Phase III trials comparing surgery and SABR, the STARS (StereoTActic Radiotherapy vs. Surgery) and the ROSEL (Radiosurgery Or Surgery for operable stage non-small cell Lung cancer), had to close early after enrolling only 36 of the intended 1030 patients and 22 of the intended 960 patients (41). The SABRTooth trial (Stereotactic ablative radiotherapy with surgery in patients with peripheral stage I NSCLS considered higher risk of complications from surgical resection) was a pilot study investigating SABR versus surgery completed in 2017(42). Leeds was also the main site and Dr Franks is Chief Investigator of this NIHR funded project (Clinicaltrial.gov: NCT02629458). It aimed to assess the feasibility and acceptability of conducting a Phase III randomised controlled trial comparing SABR with surgery in patients considered at high risk of surgical complications. Despite recruiting at a higher rate than previous SABR versus surgery trials, the SABRTooth study failed to meet its recruitment target

demonstrating that at this stage conducting a large RCT in the UK was shown not to be feasible.

The preliminary results from the SABRTooth analysis have given important insights in developing a detailed knowledge of this treatment and its possible effect on the future of NSCLC care.

1.5 Patient selection process for treatment and Risk scores

1.5.1 Functional Algorithms

Lung resection is applicable only in cases where the tumour is deemed radically resectable and the patient is deemed operable, which means fit enough to tolerate surgery without experiencing major post-operative complications or deterioration of their QOL.

For practical reasons, published evidence on operability and functional assessment is often summarised in algorithms or flowcharts. Algorithms should be used as guides to standardise the pre-operative clinical practice minimising variations and inappropriate exclusions. They essentially categorise the patients in class of risk.

The two most recent functional algorithms are those proposed by the European Society of Thoracic Surgeons/European Respiratory Society (ERS-ESTS) joint task force on fitness for radical therapy (16) and the American College of Chest Physicians (ACCP) (26). Both algorithms emphasise the importance of a preliminary cardiologic evaluation. We usually refer to the ACCP as most recent one. In these algorithms, those patients with low cardiologic risk or with an optimised cardiologic treatment may proceed with the rest of functional workup.

The risk level has been defined according to the Thoracic Revised Cardiac Risk Index (ThRCRI) (43, 44). In order to calculate the ThRCRI of a lung resection candidate, four different factors (each of them having a specific weight for the final index) should be considered:

1. History of coronary artery disease, 1.5 points.

2. Cerebrovascular disease, 1.5 points.

3. Serum creatinine level greater than 2 mg/dl, 1 point.

4. Pneumonectomy, 1.5 points.

Summing the points of each factor, the patient's aggregate ThRCRI is obtained, which ranges from a minimum of 0 to a maximum of 5.5. This value defines the different risk classes predicting an incremental risk of cardiac morbidity:

• Class A: 0 points. Risk of cardiac complication: 1.5 %.

• Class B: 1–1.5 points. Risk of cardiac complication: 5.8 %.

• Class C: 2–2.5 points. Risk of cardiac complication: 19 %.

• Class D: 2.5 points. Risk of cardiac complication: 23 %.

In the functional cardiac algorithm patient at low risk is defined as ThRCRI<2. Both algorithms recommend measurement of the Forced Expiratory Volume in the First second (FEV1) and Diffusing Capacity for Carbon Monoxide (DLCO) in all patients. These two parameters must be expressed as percentage of predicted values (Figure 1-3 and Figure **1-4**).

Figure 1-3 ERS/ESTS Functional Algorithm for Surgical Candidates

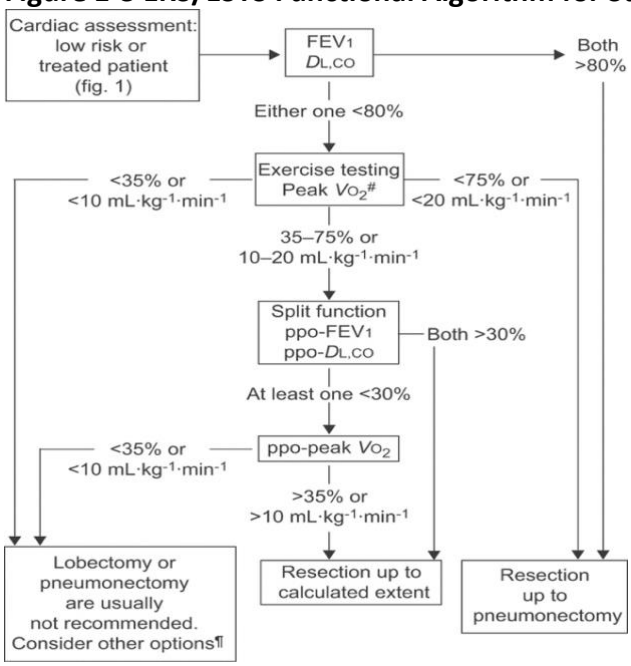
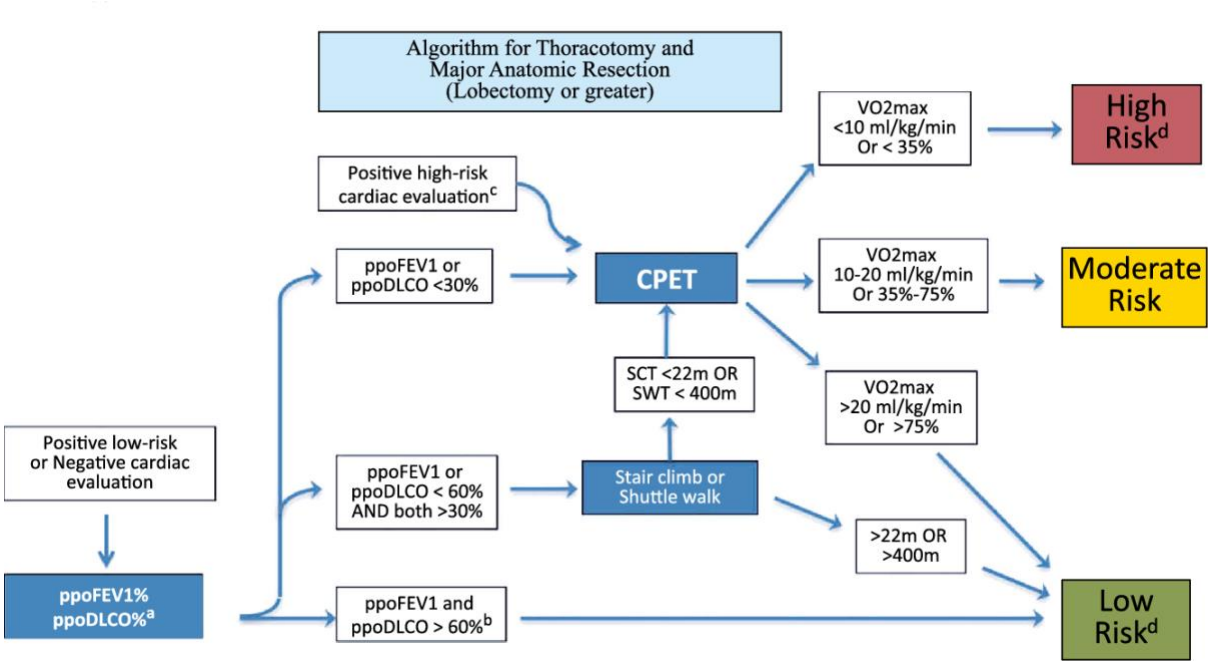


Figure 1-4 ACCP Functional Algorithm for Surgical Candidates



In the ACCP flowchart, patients deemed at low cardiologic risk and with both Predicted Post-operative (ppo)FEV1 and ppoDLCO greater than 60% are regarded at low risk for surgery (risk of mortality lower than 1%). Patients with either ppoFEV1 or ppoDLCO between 30% and 60% should perform a low technology exercise test as a screening

test. If the performance at low technology exercise test is satisfactory, patients are regarded at moderate risk (morbidity and mortality rates may vary according to the values of split lung functions, exercise tolerance and extent of resection). The cardiopulmonary exercise test is indicated only in cases where ppoFEV1 or ppoDLCO are lower than 30% or when the performance at stair climbing test or shuttle walk test is not satisfactory (i.e. altitude reached at stair climbing test <22m or a shuttle walk distance < 400m). As in the European algorithm, maximal oxygen uptake (peakVO2) values lower than 10 ml/kg/min or 35% predicted indicate high risk for major anatomic resection through thoracotomy (risk of mortality may be higher than 10% and considerable risk of severe cardiopulmonary morbidity and residual functional loss is expected). Conversely peak VO2 values greater than 20 ml/kg/min or 75% predicted, indicate low risk.

These algorithms identify classes of peri-operative risk. In the ACCP guidelines (see Table 1-1), low risk is defined as an expected risk of mortality below one percent. In patients with moderate risk morbidity and mortality, rates may vary according to the values of split lung functions, exercise tolerance and extent of resection. High risk is defined as a risk of mortality higher than 10%. Considerable risk of severe cardiopulmonary morbidity and residual functional loss is also expected in this category.

Table 1-1 Risk Classes according to the ACCP Guidelines

LOW RISK	MODERATE RISK	HIGH RISK
ppoFEV1 and ppoDLCO >60%	ppoFEV1 or ppoDLCO <60%	ppoFEV1 or ppoDLCO <60%
and	and	and
VO2max >20 ml/kg/min	VO2max 10-20 ml/kg/min	VO2max <10 ml/kg/min

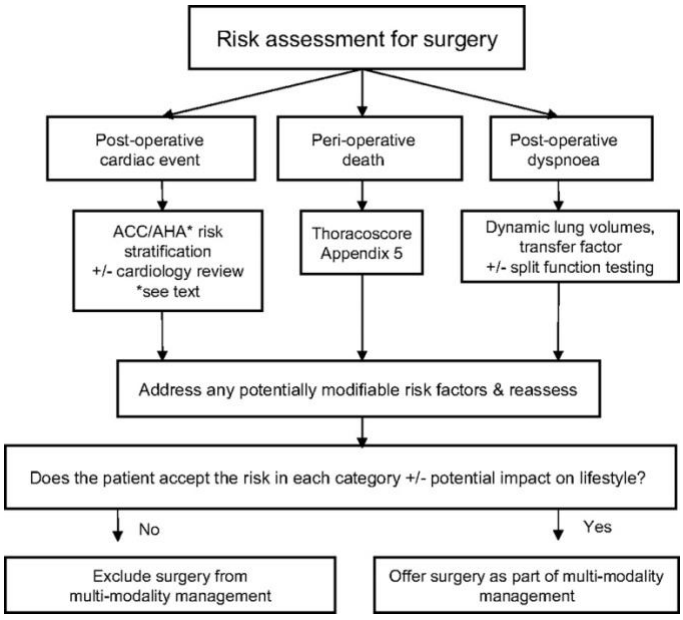
1.5.2 Risk scores

Scoring systems are often used in our specialty to predict the probability of selected outcomes in groups of patients, thus enabling risk stratification. The main limitation is the lack of accuracy on an individual patient basis. Scoring systems may be accurate in estimating the mortality rate in a certain group of patients but they fail in identifying which of the individual patients will die after the operation. Therefore, they cannot be used as a selection tool for surgery, but only for estimating the risk of morbidity and mortality, which can be informative during pre-operative counselling and tumour board discussion. Several comorbidity scores have been used in our specialty and different outcomes have been risk adjusted using risk models. The most frequently used objective indicators of surgical risk are morbidity and mortality. Mortality is usually defined as in-hospital mortality or one occurring within 30 days from operation if the patient was discharged from hospital. The definition of high surgical risk is critical for communicating with the patient the different treatment options especially because, if the doctor is focusing on more objective short-term outcomes, the patient will be interested in long-term perceived ones.

However, outcomes need risk-adjustment, to account for different case-mix and prevalence of risk factors to prevent risk-averse behaviours and misleading information. Functional operability has been the objective of many investigations and different organisations have published clinical guidelines to help thoracic surgeons and Multi-Disciplinary Teams (MDT) in stratifying the risk of a surgical oncological procedure. Risk models can be used to audit internal performance over a certain period of time (internal audit) or as instruments to allow a fair comparison between different centres (external audit).

Furthermore, risk scores have been recently implemented in the pre-operative evaluation flow-chart. The latest British Thoracic Surgery Guidelines (BTS) (45) which introduced the Thoracoscore in their physiological pre-operative workout as shown in Figure 1-5(46).

Figure 1-5 BTS Functional Algorithm for Surgical candidates



The first risk model developed from the European Society of Thoracic Surgeons (ESTS) database (called **European Society Objective Score** or ESOS) was published in 2005. The entire sample was split in a derivation (60%) and validation (40%) set. The model was

first derived from a population of 1753 patients undergoing any type of lung resections (from wedges to extended pneumonectomy) for lung tumours (47). The predictive model for in-hospital mortality included only 2 variables: age and predicted post-operative forced expiratory volume in one second (ppoFEV1).

For a given patient, the predicted risk of in-hospital death according to the model is given by the expression: exp ($logit_2$)/(1+($logit_2$)), where logit equation was the following: $logit_2$= -5.8858 + (0.501 * age)- (0.0218 *ppoFEV1%).

The ESTS has recently updated their risk adjusting morbidity and mortality models (48). These models are in use to calculate risk-adjusted outcomes as part of the Composite Performance Score (49, 50) applied to assess eligibility for the European Institutional Accreditation.

Two types of models were generated for each outcome: a logistic regression equation and an aggregate score (more user friendly). The variables were first selected by univariable analysis, then entered in a stepwise logistic regression analysis and finally validated by bootstrap resampling technique. The following categorical variables were associated with morbidity: sex, pneumonectomy, extended resections, diabetes, CAD (coronary artery disease), induction therapy, CKD (chronic kidney disease), CVD (cerebrovascular disease), and thoracotomy approach. Univariable analysis showed that the following variables were associated with mortality: age, ppoFEV1 (predicted post-operative forced expiratory volume in 1s), BMI (body mass index), ASA (American Society of Anaesthesiologists) and ppoDLCO, sex, CAD, CVD pneumonectomy, thoracotomy approach and extended resections were also associated with increased risk of mortality. The following logistic cardiopulmonary morbidity model was developed (Logistic EuroLung1) as seen in Figure 1-6.

Figure 1-6 EuroLung1

Logit= -2.465 + 0.497 X sex male + 0.026 X age + 0.231 X CAD + 0.371XCVD + 0.152 X CKD – 0.015 X ppoFEV1 + 0.514X extended resections

The aggregate score was developed by assigning proportionally weighted points to the regression coefficient (Table 1-2).

Table 1-2 Distribution of complications according to the EuroLung1 morbidity score

EuroLung 1	Morbidity Rate (%)
0-1	5.2
2-4	8.2
5-7	14.3
8-11	21.6
12-16	32.4
17-19	43.1

Similarly, a 30-day mortality regression model was developed. The following logistic mortality model was generated (Logistic EuroLung2) as in Figure 1-7 and aggregate score distributions are shown in Table 1-3.

Figure 1-7 EuroLung2

logit (mortality)= -5.029 + 0.903Xsex male + 0.044Xage + 0.264XCAD + 0.582XCVD - 0.064XBMI + 0.300Xextended resection + 0.929X pneumonectomy + 0.894Xthoracotomy – 0.009XppoFEV1

Table 1-3 Distribution of complications according to the EuroLung2 aggregate mortality score

EuroLung 2	Mortality Rate (%)
0-3	0.4
4-6	1.4
7-8	2.9
9-11	5.2
12-14	11.3
15-17	29.4

In comparison, the European Society Objective Score (ESOS) model published in 2005 showed a consistent underestimation of mortality when plotted in the current set of patients.

Other risk models have been developed and tested in this field. One of most implemented in clinical practice is the in-hospital mortality risk score derived from a total of 18,049 lung resections for NSCLS. This was entered into the French national database Epithor and called Thoracoscore (46). More recently, after the publication of a study demonstrating the underestimation of risk with in-hospital mortality, Powell et al. used the National Lung Cancer Audit (NLCA) linked to Hospital Episode Statistics (HES) to produce a new score to predict 90-day mortality after surgery in those with lung cancer (51, 52). At the moment the ESTS is launching an online calculator for the EuroLung scores, which is already available for the Thoracoscore for example. This will help the clinician in clinical practice to get those data directly in clinic and easily counselling with the patients.

1.5.3 Clinical Gestalt: Intuition, Information and Surgical Risk

Improved patient selection has been facilitated by identification of specific risk factors, generation of management guidelines, and development of specific risk scoring systems

(51, 53, 54). However, it has been demonstrated in lung cancer surgery that when a surgeon estimates the risk for a major lung resection, based on case-matched clinical vignettes, they are only moderately accurate(55). Judgment about the appropriateness of treatment has been evaluated in surgery and is influenced by a variety of factors (56), some of which are surgeon related, including training level, recent adverse patient outcomes, and prioritisation of risk versus benefit (57, 58). In addition to using objective measures of fitness, surgeons strongly rely on their clinical reasoning skills to assess risk in individual patients. In general, surgeons tend to overestimate the risk of complications in healthy patients and underestimate risks in sicker ones (59). The best we can do to balance information and intuition is to find out the right mix of conscious and unconscious analysis for each patient. The increased experience will then allow us to perform a more accurate rapid cognition analysis on an individual basis(60). Patient preferences, expectations and values must be factored with objective risk analysis, physician reasoning skills and prognosis of disease to assist in risk communication during patient counselling.

These concepts are far more difficult to estimate and risk-stratify since they involve a certain level of subjectivity. The first step to introduce the subjective parameter into an objective risk-assessment is to look at the only patient-reported outcome which has been recently implemented in oncological clinical practice: quality of life (QOL).

1.6 Patient Reported Outcomes Measures (PROMS) in Lung cancer setting and in patient selection setting

1.6.1 Evidence for impact of Lung cancer on patients' lives

Lung cancer patients have an increased tension-anxiety status and are generally psychologically depressed compared to the general population (61-63). This emotional status may influence their perception of the surgical risk and their decision to proceed to surgery. It has been shown that lung resection causes a transient reduction in many physical and mental domains, and that patients generally tend to return to pre-operative levels after three to six months (64). Several studies have shown that objective parameters, traditionally used to stratify the surgical risk (age, FEV1, DLCO, VO2max etc.) are not associated with patient reported residual QOL. In this regard, the peri-operative changes of quality of life in high-risk surgical candidates have been shown to be similar to their lower risk counterparts (65-67). The only objective parameter consistently associated with impairment in residual QOL is the extent of resection. In a recent synthesis of evidence, it was found that patients after pneumonectomy experience a large decline in physical functioning, mental health or cognitive function, and role limitation caused by physical problems, which persist even one year after surgery (64).

From the patient's perspective the decision to proceed to surgery depends mainly on the estimated risk of fixed long term outcomes, such as permanent debility, oxygen dependency, limitations in activities of daily living (68). Unfortunately, there is insufficient evidence in the literature to formulate a reliable prediction of such risks, and certainly research is most needed in this area of research. The traditional pre-operative

parameters of pulmonary function test (PFTs), age and comorbidities are correlated primarily with short-term morbidity and mortality, but not well correlated with long-term functional outcomes.

1.6.2 Use of Patient Reported Outcome Measures (PROMS) in clinical practice

Over the last few decades assessments of patients' QOL questionnaires have been broadly adopted as integral part of cancer clinical trials. Quality of life in fact can be regarded as a broad global term related to the patient's subjective evaluation of life as a whole. On the other hand, it could be a more specific multi-dimensional construct related to aspects assumed to be affected by health care interventions. The latter refers to the definition of Health-related Quality of Life (HRQOL). It covers subjective perceptions of the positive and negative aspects of a patients' symptoms and, importantly, disease symptoms and side effects of treatment (69). There is now general agreement that the minimum requirements for domains within HRQOL assessment tools are inclusion of physical, social and emotional functioning and disease- and treatment-related symptoms and side effects. Recently, spiritual and existential issues, sexuality and body image, and other dimensions have also been included(70).

Patient-Reported Outcomes (PROs) are defined as any data that are reported directly by the patient without an intermediary such as a family member or a healthcare professional (71, 72).

Patient Reported Outcome Measures (PROMS) are standardised and validated instruments to measure patients' perceptions of their health status and their HRQOL (73).

With advances in oncological treatments, clinicians are now interested not only in survival alone but also how these treatments are affecting the patients remaining life. In the cancer setting, PROMS have become an increasingly important tool for evaluating outcomes in different treatment populations and monitoring patients during cancer care.

The research community are now including PROs as standard data to capture patient's subjective experiences, more commonly as a secondary endpoint but recently also as a primary outcome (74). Also, the provision of PROMs data to clinicians has been demonstrated to increase clinician's awareness of issues and facilitate the identification, discussion and documentation of symptoms and HRQOL (75, 76). This helps the patient to be more involved in making difficult health-related decisions.

Over the years, however, some methodological issues have been raised related to the many challenges associated with implementing PROMS into clinical practice (77). The main problems are in fact the measures to choose, method of collection, frequency of completion, presentation to clinicians and training of clinicians. Also, as demonstrated by our particular population of elderly patients, it is mandatory to consider carefully the group of patients and the care settings where the PROMS will be implemented.

1.6.3 Evidence on HRQOL in the context of the early stage NSCLC treatments

Although guidelines refer to objective thresholds to estimate the surgical risk, the main concern for candidates to lung resection is not so much immediate mortality or complications, but permanent disability and loss of independence (68). However, QOL is not an easy concept to investigate in this group of patients. Furthermore, there is

scant evidence in the literature that can help the physician to counsel the patient about prediction of residual QOL.

An official NHS PROMS programme has covered four common elective surgical procedures since 2009 (78). Despite evidence of correlations between PROMS and survival (including our research (79)), the routine collection after NSCLC surgery is still sporadic. In fact, in the setting of oncology, the majority of available PROMS studies are predominantly reported by oncologists or palliative care teams (80).

A published survey among European thoracic surgeons, revealed a lack of standardised PROMS collection among this community with 88% of all surgeons currently not incorporating these outcomes into their clinical practices (81).

As SABR is a relatively recent technique implemented in lung cancer care, PROMS were reported only in few limited trials(82). The paucity of published data for early stage NSCLC will be considered in detail in Chapter 3.

1.6.4 The need of PROMS for a shared decision-making process

It is increasingly recognised that inclusion of validated PROMS assessments within clinical trials can provide important data for clinicians to inform treatment decision-making. Within the oncology clinical trial literature there are numerous examples of where clinical decision-making has been influenced by the outcomes of PROs assessments.

PROMS can inform important aspects that are being discussed in clinic with patients during difficult decision conversations. It has been reported that the majority of UK patients want to be involved in decisions regarding their health status (83). Increasing patient-doctor communication about treatment options is a priority for the NHS: the biggest predictor of legal complaints is not poor outcomes, but a combination of poor

outcomes with bad communication (84). Data about the perceived involvement of patients in the decision-making process or presence of any residual decisional conflicts may help to identify groups of patients requiring decision support. In this case it may be useful to include decision aids as part of process of care.

Although internationally accepted functional guidelines defining the risk of surgery exist, selection of the best radical treatment for borderline patients remains an area of debate. The advent of SABR for peripherally located NSCLC as a non-surgical radical treatment for patients medically unfit for surgery has further classified NSCLC patients into three broad categories:

1. Patients fit for surgery;

2. Patients clearly unfit for surgery and for whom SABR represents their ideal radical treatment;

3. Borderline patients at increased surgical risk for whom both surgery (with higher risk) and SABR may represent two ontologically acceptable treatments.

Borderline patient's decisions are influenced by both objective and subjective assessments. Currently decision for treatment is typically done based upon an objective assessment of the patient's fitness against short-term risk of morbidity/mortality (30 and 90 day) and does not consider other important longitudinal outcomes like post-treatment QOL. In the particular setting of early stage NSCLC, informing patients of their quality of life evolution after these two treatments, may help them the best treatment decision. These longer terms outcomes are likely to have a large influence on the patient's choice of treatment and currently this data is not known.

1.7 PROMS questionnaires in Lung Cancer settings

1.7.1 Quality of Life (QOL)

In the clinical setting, the majority of studies have used a combination of both a generic and disease-specific questionnaire. This combination enables assessment of general health domains like physical or social functioning using questionnaires (European Organisation for Research and Treatment of Cancer Questionnaire-Core 30 or EORTC-QLQ-C30) or Functional Assessment of Cancer Therapy (FACT-G)(85) as well as symptom-specific instruments, which are related to the disease or treatment.

In thoracic surgery, two types of QOL questionnaires are used: generic and lung cancer specific. Figure 1-8 shows the differences between the more broadly used questionnaires as reported in my previous published Scopus review (86).

Figure 1-8 QoL instruments in Lung Cancer studies

Generic Instruments	Cancer-specific Instruments	Lung cancer specific instruments
Short Form Health Survey (SF-36)	EORTC QoL Questionnaire Core 30 (EORTC QLQ-C30)	EORTC Lung Module (LC-13)
Ferrans and Powers QoL Index (QLI)	Functional Assessment of Cancer Therapy (FACT-G)	Functional Assessment of Cancer Therapy-Lung (FACT-L)
Nottingham Health Profile (NHP)	Functional Living Index (FLIC)	Lung Cancer symptom Scale (LCSS)
WHO QoL Instrument (WHOQOL-100)		

EORTC, the European Organization for Research and Treatment of Cancer; QoL, quality of life.

A feature of a generic tool is that it helps to compare our population with the healthy one. Intuitively, symptom changes cannot be investigated which are caused by specific treatments. One of the most used tools in this category is the Short Form 36 (SF-36)(87). Cancer-specific questionnaires study the effect of cancer and its treatment on QOL. The most widely used tool in oncology is the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30) (88). This questionnaire is supported by several specific complementary modules, with the aim to be more responsive to changes in different sub-groups. The EORTC LC13 questionnaire for lung cancer is useful to study specific lung symptoms e.g. cough, haemoptysis, chest pain and dyspnoea (89). However, so far, no specific validated questionnaire has been developed for the lung cancer surgical population.

That is why EORTC QOL Lung Cancer Group (which I am a member of) has recently finished the Phase IV project to revise the lung cancer module. This will cover all QOL aspects relevant in the newly available diagnostic and therapeutic options and it will highlight all QOL aspects that are relevant for patients with lung cancer but are missing

in the LC13 e.g. post-surgical symptoms. I used the EORTC-C30 questionnaire with the LC13 module as this is the most internationally validated instrument at the moment and covers most of the common symptoms applicable to both treatments.

1.7.2 Patient Satisfaction

Patient satisfaction reflects the perception of the customer about the level of quality of care received during the episode of hospitalisation. It is a model of non-clinical indicators of performance, which is not necessarily correlated with more common clinical outcomes(90). In terms of patient satisfaction, there are no specific Lung Cancer questionnaires. Satisfaction is an abstract and multi-dimensional concept, which is hard to be directly observed or measured, therefore should be evaluated using a variety of multi-item scales (91).

Measuring the satisfaction of patients requires validated tools. The EORTC IN-PATSAT 32 is a multi-dimensional questionnaire, adapted to measure patient satisfaction related to physician and hospital staff, as well as aspects of the organisation of care and services (92). I have used this questionnaire in previous research (93), however in this PhD I had two populations to investigate which are quite different. The EORTC IN-PATSAT 32 was not chosen as it has only been validated for inpatients and the SABR patients are day-case and therefore they would not have been able to rate satisfaction with the system and staff in the same way. The length of the questionnaire (32 items) was not suitable for a multi-questionnaire project.

The Patient Satisfaction Questionnaire Short Form (PSQ-18) was chosen as it is a cross-cultural validated survey for use in different settings. It is short, reducing patient burden of filling in multiple questionnaires and covers the most important aspects of hospital attendance. This was helpful as the surgical procedure was an inpatient treatment but

the SABR was an outpatient one. The PSQ-18 questionnaire is an 18-item self-administered survey including different scales reflecting the perceived level of satisfaction in relation to the care provided by doctors. The team behind this Likert scale questionnaire proposed seven dimensions of patient satisfaction directed toward their doctors. These are general satisfaction, technical quality, interpersonal manner, communication, financial aspects, time spent with doctor, and accessibility and convenience(94).

1.7.3 Shared Decision-Making

Shared decision-making (SDM) is a concept suited to situations where patients are given two or more medically reasonable options with no professional consensus (95). As described previously, although current treatment guidelines do not recommend SABR as first-line treatment for moderate risk patients with NSCLC, multiple observational studies have suggested therapeutic equipoise exists between SABR and surgery. The lack of long-term QOL data from these two treatments has highlighted the importance of understanding whether a truly informed “shared decision” is made, which is highly dependent on the interaction in the doctor/patient consultation.

In order to develop SDM guidelines or integrate this concept into the existing decisional algorithms, the last two decades have witnessed an increasing number of trials investigating the overall lack of concordance between physician and patient perceptions of the decisional context in many clinical areas including lung cancer management. The majority of these trials have shown that concerns and treatment strategies were insufficiently discussed between the patients and physicians. A recent systematic review recommended the choice for the most appropriate instrument to be best based on the instrument's content and characteristics such as the perspective they assess(96).

Although the area of SDM is relatively new, an abundance of research in which measures are developed and tested are available (97). As a result, while there are many different measures available, the degree to which the available measures are validated varies significantly. Many of the scales available have been validated in only a small number of studies.

The Decision Self-Efficacy Scale (DSE) measures self-confidence or belief in one's ability to make informed decisions and participate in shared decision-making with health professionals (98). It is an 11-item instrument with a five-point response scale ranging from 0 (not at all confident) to 4 (very confident). An example question is: *'I understand the information enough to be able to make a choice'.* Internal consistency has been evaluated in women considering hormone replacement therapy (alpha coefficient 0.89). It generates a unique score that is measuring the patient's efficacy.

O'Connor developed this questionnaire focusing on the social aspect of decision-making but accompanying this was a decision conflict scale, measuring the cognitive aspects of this using a conceptual framework (99, 100). The decision conflict is clinically more applicable to the conflicting choice between surgery and SABR especially for those borderline patients, where there is equipoise in terms of risks and benefit.

1.8 Electronic PROMS

Electronic methods for collecting PROMS (ePROMS) have the potential to facilitate the follow-up and communication of cancer patients: they are convenient for patients, increase data accuracy, reduce long-term costs and provide large datasets detailing patient experiences.

It is important to incorporate ePROMS to support patient care in routine practice and a recent publication demonstrated the clinical benefits associated with symptom self-reporting during cancer care (101).

In Leeds, the Patient-Centred Outcomes Research Group (PCOR- University of Leeds, LIMR and Leeds Cancer Centre) is an international leader in the implementation of ePROMS assessment in routine oncology practice (102) to support clinicians and inform patient care. A web-based system (QTool) has been developed in Leeds as a patient portal for entering patient self-reported symptoms, which is securely linked in real-time with electronic patient records (PPM).

eRAPID (electronic patient self-Reporting of Adverse-events: Patient Information and aDvice) was a National Institute for Health Research-funded programme, and developed a system for patients to self-report and manage adverse events (AE) online during and after cancer treatment in Leeds (103). It has demonstrated the feasibility of AE reporting from home and integration into Electronic Patient Records (EPR) for use in routine care (although in different cancer type populations (104)).

1.9 Leeds Setting

Leeds is a tertiary referral centre in West Yorkshire, with expertise in delivering both minimally invasive thoracic surgery and SABR.

1.9.1 Leeds VATS Programme

In 2014 200 curative lung resections were performed for stage I-II NSCLC (70% using VATS approach). Ten% of the surgical patients were octogenarians, 50% were older than 70, 30% had moderate-to-severe COPD and more than 10% had a history of ischemic heart disease.

1.9.2 The UK SABR Consortium and Leeds SABR programme

The UK SABR Consortium was formed in 2007 to facilitate the safe introduction of the technique using national guidelines. SABR was introduced to the Leeds Teaching Hospitals (LTHT) in May 2009. A recent publication has audited outcomes for the first series of lung cancer patients treated using SABR in the UK. As in other series (36) excellent local control (>95% at 3 years) and low rates of nodal relapse have been observed (105).

Between 2009 and 2015, the Leeds Cancer Centre treated over 500 patients, with local and national referrals. Leeds was the leading centre for a feasibility study of SABR versus surgery in patients considered at higher-risk of surgery (SABRTooth). The introduction of SABR in Leeds has led to a significant increase in overall radical treatment rates for patients with stage I lung cancer, without resulting in a sustained reduction in surgical resection rates (106).

1.10 Need

Lung cancer has the highest mortality of all cancers in the UK, and accounts for the largest single cause of premature death in Leeds(107). The latest analysis of lung cancer incidence rates report significant variation across the UK, with the highest rates in the North of England (13).

So far, there are insufficient data to identify clear-cut criteria for defining high-risk patients or to help clinicians and patients decide about the best treatment for early stage NSCLC. Furthermore, this data does not include Patient Reported Outcomes and do not account for individual patient preferences. There is a clear need for supporting

decision-making processes using PROMS as a key part of information provision to patients.

1.11 Preliminary studies and context of the thesis

The development of reliable automated systems for collection and analysis of quality of life data in cancer patients is a key requirement for the use of QOL instruments in clinical practice. Furthermore, evidence of the feasibility of collection of data (patient-reported and clinical ones) was pivotal information in planning this project.

This thesis benefited from the established collaboration of three units in Leeds, as described below.

1.11.1 Section of Patient Centred Outcomes Research

Electronic methods for patient reporting are acceptable to patients and provide better quality data (102). Research by the Section of PCOR (in over 2000 patients) has shown that patients can routinely complete PROMS on touch screen computers in clinic (75). Patients are also willingly to use PROMS from home via the internet (76, 108) and using mobile devices (109). With internet access at 82% of the UK population (Office National Statistics, 2010), using a web-based system to measure radiotherapy toxicity is attractive and may allow a more consistent method of monitoring late toxicity when patients do not routinely attend the hospital or are followed up by different specialty teams. Enabling patients to complete assessments at home may allow significant reductions in routine outpatient follow-up visits and reduce costs.

1.11.2 Clinical Oncology

Historically Phase III trials evaluating the effectiveness of SABR in radical treatment of lung cancer compared to surgery have been challenging and have failed to recruit. To address this, Leeds was recognized as the leading centre for a feasibility study of SABR versus surgery in patients considered at higher-risk of surgery (SABRTooth). The pilot project of this PhD will be complementary to SABRTooth, focusing on those patients not considered suitable for SABRTooth, i.e. patients that are fit for surgery or patients that are too high risk for surgery and receive SABR(42).

The SABRTooth trial (NCT02629458) fell short of recruitment and less than 50% of patients were randomised within the target study time. The reason for declining randomisation among eligible patients was predominantly their preference towards one or the other treatment with 60% of patients preferring SABR and 30% surgery (110).

This study showed a large scale RCT to address the fundamental question of which treatment is most cost-effective for borderline patients is not feasible in the United Kingdom.

Therefore, alternative approaches are needed to answer this key question in the face of an increasing proportion of elderly and co-morbid patients presenting with early stage lung cancer.

1.11.3 Thoracic Surgery

PROMS collection started in the Thoracic Surgery Unit in September 2014 with the administration of pre-operative surveys from consecutive patients submitted to pulmonary resections for lung cancer. A preliminary study evaluating the QoL trajectory of these patients was presented at the 2015 Meeting of the International Society of Quality of Life Research (ISOQOL)(111).

1.12 Overall Hypothesis and aims

The overall aim of this thesis is to assess the role of the Patient Reported Outcome Measures during radical treatment for early stage NSCLC (VATS resections and SABR).

The QOL profiles were compared to potential confounding treatment-related factors and patient-related factors, for example baseline patient characteristics, potentially enabling treatment selection modification for high-risk individuals in the future.

It is hypothesised that pre-treatment QOL will be a more important predictor of post-operative clinical outcomes than traditional objective parameters. Overall this research aims to explore the potential value of incorporating QOL scores into the preoperative selection process, to improve current risk scores. In addition, these QOL data will provide an important insight on patient experiences after radical treatment and will provide patient-centred data to discuss with patients before VATS resection and SABR.

I will test these hypotheses using mixed methodology and two datasets. First using existing data from a prospective audit-based cohort of surgical patients with QOL before and after surgery. Second, a prospective observational study measuring acute and longer-term PROMS over a one-year period in patients treated with SABR and VATS surgery using PROMS integrated into electronic patients record (EPR). I will also take the opportunity with the prospective study to explore patient efficacy in treatment decision making and satisfaction with care.

Specifically, I aim to explore:

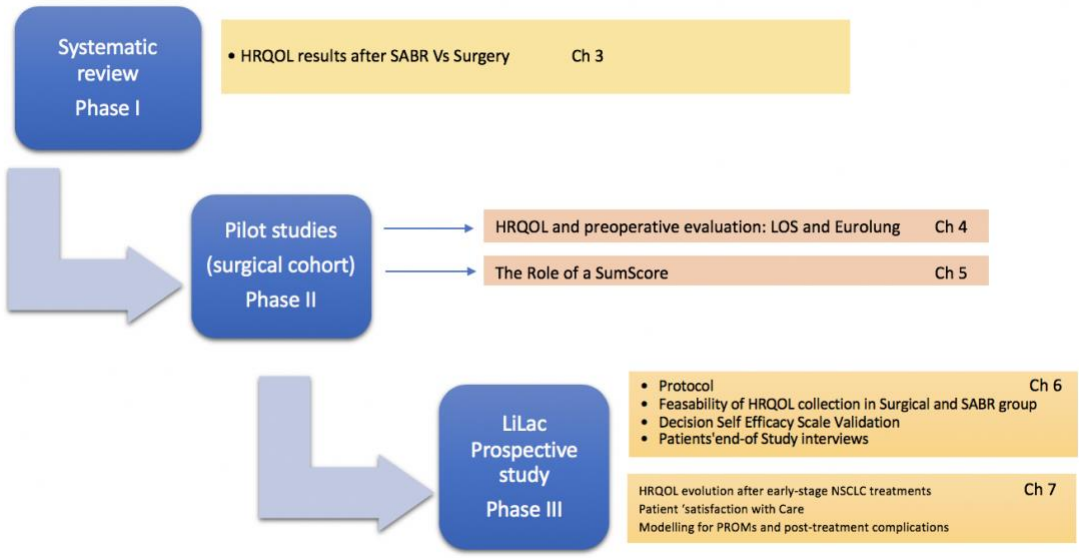
1) The actual value of the Quality of Life collection before and after Early Stage NSCLC treatments.
2) The potential of QoL in predicting post-operative outcomes after surgical treatments for Early Stages NSCLC.

3) The feasibility and acceptability of routine patient reported outcomes (PROMS) and ePROMS collection within clinical practice in patients treated for early-stage NSCLC (via patient interviews).
4) The QOL trajectory during the first year after VATS resection and SABR and the role of other PROMs (Patient Satisfaction with Care and Decision self-efficacy).
5) The relationship between PROMS and objective parameters in the two groups and their potential role in pre-operative risk-assessment.

1.13 Structure of the thesis

Figure 1-9 provides an overview of the PhD phases and the chapter integration into them.

Figure 1-9 Thesis Flow-chart



Chapter 2 describes the methodology used.

Chapter 3 describes a systematic review of the current evidence of quality of life reporting in early stage NSCLC radical treatments. **(Aim 1)**

Chapter 4 describes an exploratory quantitative work to investigate the association between quality of life, preoperative risk-score and post-operative morbidity and mortality after surgical treatment. **(Aim 2)**

Chapter 5 describes exploratory quantitative work to compare the performance of a QOL summary score with traditional QOOL scales and their sensitivity to change in a cohort of surgically treated NSCLC patients. **(Aim 2)**

Chapter 6 summarises the methodology of the prospective study on patient reported outcomes trajectory of patients treated with radical intent for early stage NSCLC- the LILAC study. Descriptive analysis of the socio-demographic and clinical data will be presented along with provisional data evaluating the feasibility and acceptability of the PROMs collection through consideration of recruitment and attrition rates, missing data and patient feedback. The qualitative work undertaken will further explore motivations and barriers for engagement with the system, and their perception of its impact on their care **(Aim 3).** It also validates and investigates the psychometric properties of the decision self-efficacy scale questionnaire in the NSCLC population. **(Aim 4).**

Chapter 7 describes the QOL evolution in the two groups **(Aim 4).** It also investigates the role of the Decision self-efficacy, QOL and other clinical-demographic characteristics on PROM (patient's satisfaction and 6 months quality of life outcomes) and clinical outcomes (complications) among patients treated with radical intent for early stage NSCLC **(Aim 5)** in order to explore potential of inclusion in pre-treatment risk assessment.

Chapter 8 summarises and discusses the work in the preceding chapters. Strengths and limitations are outlined and recommendations for future research are made.

Chapter 2 Methodology

The purpose of this Chapter is to give an overview of the methodology of this thesis (more detailed descriptions of methodology are provided in individual Chapters). I describe the different methods employed and explore why each method has been chosen for the different phases of my thesis.

My thesis employs a mixed methods approach, which combines qualitative and quantitative methodological techniques within a series of connected study phases (112). Mixed methods studies aim to combine the strengths of quantitative methods, which generate numerical data, with qualitative approaches, which tend to generate non-numerical data using techniques such as semi-structured interviews, exploring a particular research question more comprehensively than it may be using either method in isolation. The value of a mixed methods approach to research in a healthcare setting is increasingly recognised, where questions are often multi-faceted and complex (112-114).

This approach allows researchers to explore diverse perspectives and uncover relationships that are present in complex and multifactorial research questions. The key word is 'mixed', as an essential component is data linkage, or integration at an appropriate stage in the research process of the project as a whole (114). A mixed methods design is crucial for answering research questions that neither quantitative nor qualitative methods could fully answer (115).

Mixed methods can be also used to obtain a better understanding of the relationship between these two types of data: qualitative and quantitative. Most importantly for this thesis, this methodological approach provides opportunities for participants (both patients and health professionals) to have a strong voice and share their experiences

across the research process, and they can facilitate future work and studies to help answer the same or similar research questions more accurately (116).

However, mixed methods have peculiar characteristics and differences within the same study. For example, the small sample sizes used in qualitative work may limit the generalizability of the findings. Quantitative analysis instead, aims to reduce confounding within the analysis and to be representative of a population and potentially generalizable to others. Both aspects need to be weighted and tailored to address each intervention. The Mixed Methods Appraisal Tool (MMAT) is a useful appraisal checklist for appraising published mixed methods research but could also be used as a design check-list when planning mixed methods studies (117).

2.1 Summary of research methods

Within this thesis I have employed the following methods:

- Synthesis of the evidence using systematic literature review methodology and structured literature reviews applied to a Best Evidence Topic Format (**Chapter 3**)
- Quantitative methods: prospective collection of clinical and quality of life data on a cohort of patients planned for surgery in the context of an audit project. Data analyses including descriptive statistics and regression analyses (**Chapter 4,5**)
- Quantitative methods: prospective, longitudinal study design with data analyses including descriptive statistics and regression analyses (**Chapters 6,7**)
- Psychometric analysis of a questionnaire (**Chapter 6**)

- Qualitative methods including semi-structured interviews analysed using thematic framework analysis on patient and staff feedback (**Chapter 6**)

As described in the thesis objectives in Chapter 1, this project was planned in three key phases (**Error! Reference source not found.**). The Local Research and Innovation Department has been involved since the beginning of each phase and Ethics approval was sought when required from the Research Ethics Committee: to note for Chapters 3, 4 and 5 no formal Ethical approval was requested. For Chapters 7 and 8 all HRA (Health Research Authority) procedures were followed in order to obtain approval from the NRES Yorkshire and the Humber-Leeds East Research Ethics Committee (REC Ref: 16/YH/0407).

The initial results of the systematic review described in Chapter 3 and the preliminary results of the pilot studies in Chapters 4 and 5 informed the analysis and development of the later phase. The qualitative work in Chapter 6 complements and informs the final quantitative work described in Chapter 7 and has also given insights for the prospective study amendments.

2.2 Phase 1: Systematic Review methods

I adopted an approach pioneered in emergency medicine, namely the Best Evidence Topic (BET). Clinicians select a clinical scenario from their daily practice that highlights an area of controversy. From this, a three-part question is generated and this is used to search Medline or other databases for relevant papers. Once the relevant papers are found, these papers are critically appraised using validated checklists and the results are summarized. Evidence-based clinical advice is reached following this process.

I decided to adopt this format as the topic chosen for this project does not have sufficiently robust evidence to allow a full systematic review. However, recent publications on clinical results of radiotherapy treatment have increased the number of trials attempting to compare these two treatments (Surgery Vs SABR).

BET articles are pragmatic reviews that were developed to teach the principles of evidence-based medicine (EBM) and to answer specific clinical questions faced in clinical practice, and for which meta-analysis with statistical pooling of available evidence was not feasible (118, 119).

Emergency medicine BETs were first published in the Journal of Accident and Emergency Medicine (now Emergency Medicine Journal) in 1998 and since 2000 have also been listed on the BestBET website (http://bestbets.org/). Other specialities rapidly adopted this methodology, especially surgical ones (cardiothoracic and paediatric for example (120)). BETs are emerging as a popular format by which clinical surgical questions can be addressed, especially those relating to gastrointestinal surgery, as demonstrated by recent reviews in the surgical area (121).

The particular problem facing surgical specialties is the evidence that does exist is frequently not of the highest quality and therefore most formal critical appraisal processes tend to discard the majority of our papers due to methodological flaws or poor design by their standards. In this specific topic area we acknowledged, from the beginning, the failure of trials in recruiting and randomizing patients between these two treatments. BETS are divided into five stages, based on the principles underlying all evidence-based medicine (Table 2-1).

Table 2-1 BET Phases

1	Asking the right question
2	Searching for the evidence
3	Appraising the evidence
4	Summarizing the evidence
5	Reviewing the evidence

As the main feature of a BET is to encourage assimilating the findings into immediate use in clinical practice, one of the crucial points in designing a synthesis of this type is to develop a focused research question. This ensures that each topic is rooted in clinical practice and will be of immediate value to clinicians. The question should be divided in three parts: patient characteristics, interventions and relevant outcomes. These points reflect the PICO frameworks in a more immediate reading format (122, 123). The next phase of this BET followed the structure of a formal systematic review: involving a detailed and comprehensive plan and search strategy derived a priori, with the goal of reducing bias by identifying, appraising and synthesizing all relevant studies on this topic. Details of this search will be described in detail in Chapter 3. Formal training was undertaken in York on Systematic Review and Critical Appraisal which has allowed me to systematically search and appraise all the papers related to this project. In order to achieve the aims of this BET, it was vital also to ensure that the search strategy had a high degree of sensitivity by looking at conference abstracts and expert opinion papers. This is due to the fact that comparison between Surgery and SABR for early-stage NSCLC is a recent debate which has been discussed frequently during International Conferences rather than in formal scientific evaluations. I was also supported by the University of Leeds Library in my initial searches, they reviewed and improved my strategy and search outcomes.

2.3 Phase 2: Service Evaluation Dataset for Pilot Studies

Data from a service evaluation I started in 2014 in Leeds Cancer Centre framed the main dataset of Phase II of my PhD. Coming from a background of PROMS research, I initially discussed with the clinical team in the Thoracic Department my previous experience in Italy regarding the implementation of QOL questionnaires in clinical practice. This led to the establishment of a QOL Survivorship clinic in my previous hospital. I presented these results and after discussion with all the clinical and research teams we decided to start routine collection of standardized QOL data in Leeds Thoracic Unit. Following on from this also routine collection of these data in other satellite Trusts during postoperative appointments was undertaken. The aim of this audit was not only to evaluate QOL changes after surgical resections, but also to stimulate discussion of important patient-reported parameters in outpatients clinic. Evaluation of the changes during the consultations however, was not possible due to staff shortage during the period of this audit, but it has increased the quality of care experienced by patients who report feeling completely supported when discussing with the Clinical Nurse Specialist (CNS) the results of their completed QOL survey. I also involved, at this stage, the Leeds Lung Cancer Support Group as PPI, to start involving patients in designing a robust study with PROMS collection. I have facilitated some meetings with them, presenting my previous QOL work, looking for their advice and feedback on this type of outcome based upon their lung cancer treatment journey. All pilot studies form part of this Quality of Life Audit which was approved by our R&I Department in 2015 as a Service Evaluation, so no Ethical Approval was sought.

My research aim in this phase was to preliminary test the association between PROMS (quality of life more specifically) and critical clinical variables in a local cohort of lung

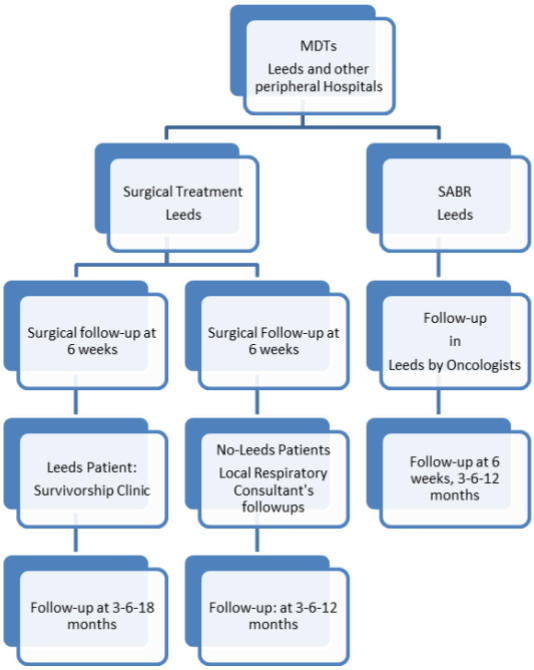
cancer patients treated with surgical resection in Leeds Cancer Centre. Most importantly, I wanted to investigate the relationship with those identified as important characteristics of the preoperative selection process for early stage NSCLC patients. I have explored this association via two cross-sectional studies (Chapter 4-5).

2.3.1 Descriptive and regression analyses to investigate the predictive value of QOL

Prospectively collecting clinical and quality of life data in the context of a service evaluation project has given me the opportunity to explore from this stage, the benefit, but also the challenges of this type of data collection in our particular field of oncology. Within the time frame of a PhD this study design provides the best method to estimate the association between preoperative quality of life data and postoperative clinical outcomes.

Patients who have had a surgical anatomical lung resection for NSCLC or suspected NSCLC, from April 2014 to September 2016 in the Leeds Cancer Centre were eligible. Patients were invited to complete a single PROMS assessment on paper, prior to their hospital admission for the operation. The patients were filling in the questionnaire during their outpatient appointment with the surgeon and returning it either immediately or on the admission day. The questionnaires were handled by the Lung Cancer Specialist Nurses (CNS) in clinic who returned to me for data input. The questionnaire was part of the Lung Cancer Specialist Nurse package of questionnaires including the validated Distress Thermometer given during the preoperative appointment and during the Survivorship postoperative clinic appointment at three months. Figure 2-1 shows the usual pathways of early stage NSCLC patients treated in Leeds with SABR or Surgical resections.

Figure 2-1 Leeds Lung Cancer post-surgical Pathways



The relationship between preoperative quality of life score and other potential confounding variables (postoperative complications and length of stay) was explored. Sample size calculations were performed in advance of recruitment into the clinical studies and are outlined in Chapters 4-5.

This collection of clinical and patient-reported data at two separate time-points does have a number of disadvantages (124). The study may be prone to non-response bias if the patients who chose to take part differ from the population of lung cancer patients as a whole, treated at Leeds Cancer Centre. However, this methodology allows an exploration of the prevalence of preoperative symptoms in a large-scale population and provides an opportunity to evaluate the relationship with preoperative QOL and other confounding factors. Furthermore, the post-treatment collection of QOL data time point was chosen reflecting current hospital practice of the first Survivorship clinic appointment being at three months. This may have increased return rate but of course excluded the important 90-day post-surgery period which is where most of the postoperative complications occur. For this reason, in the prospective LILAC study we

introduced the six-week time-point for QOL collection to better investigate the effects of immediate post-surgical complications on PROMS.

Auditing all consecutive patient candidates eligible for surgical resection for lung cancer is particularly suitable for estimating the prevalence of a symptom or disease in a population and the QOL immediately before lung cancer treatment had not been previously consistently explored in a large population like the one attending the Leeds Thoracic Unit.

2.3.2 Analysis of QOL change after surgery using summary score (SumSc)

Chapter 5 aims to validate the use of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC) Summary Score (125). The SumSc is a single higher-order factor model based on 27 of the 30 items of the QLQ-C30 (excluding Global QOL and Financial difficulties) that exhibits good model-data fit (125). The use of the QLQ-C30 summary score may supplement the 15-outcome profile generated by the QLQ-C30. The exact scoring algorithm for generating the QLQ-C30 summary score is available via the group's Web site: http://groups.eortc.be/qol.

The importance of finding a unique score for QOL is highlighted in Chapter 1 when introducing the concept of implementing QOL in the preoperative evaluation of the NSCLC patients. Also, as part of the European Society of Thoracic Surgeons Society (ESTS) Patient Centered Working Group, I have been promoting the implementation in the international ESTS Lung Cancer Registry of important PROMS like QOL. There is, however, a risk that patient-reported outcome (PRO) data may be under-represented in the big data 'revolution' (126).

QOL in Lung Cancer patients has been shown to drop consistently at three months after the operation (86). We decided to use the data collected at three months after the

operation, not only because the attrition rate was limited, but also because we were expecting the major effect on QOL at this time point. As described in the previous paragraph however, the Survivorship Clinic is a follow-up clinic run in Leeds by the Respiratory Physicians for all the early stage NSCLC treated and living in the Leeds area. Other patients are followed in the local satellite Trusts. This cannot rule out a selection bias in our results, but we have considered this and made some comparison with the entire population with the preoperative data.

We also compared the SumSc changes against the common EORTC QLQ C-30 scores in the whole population and also in a subgroup of patients.

2.4 Phase 3: Prospective Study (LILAC)

The prospective observational study aims to measure acute and mid-term QOL over a one-year period in patients treated with SABR and VATS surgery using PROMS integrated into electronic patients record (EPR) to explore:

1. The trajectory of QOL over the first year after VATS resection and SABR.
2. How patients engage with PROMS and ePROMS collection over the course of their first year after treatment and how their QOL is.
3. Patient's efficacy in making decisions and patient satisfaction with their care and how those aspects are related to each other.
4. Possible implementation issues through the patients and staff interviews.

This part of my thesis was an important opportunity to take a more formal approach in designing methods to capture longitudinal QOL data in two different treatment groups of lung cancer patients.

Following the experience of my previous studies and the audit described in this thesis, I identified the mixed methods approach as an ideal technique to prospectively collect PROMS and clinical data in these two groups of patients and explore via the direct patients' voice, possible ways to improve this collection and adoption in future clinical practice. After an initial quantitative phase of prospective PROMS collection, possible challenges or benefits are explored further with qualitative interviews to better understand how the personal experiences of individuals compare to the quantitative results. This kind of study tries to explain qualitatively how the quantitative mechanisms might work. Mixed methods give a voice not only to study participants but also to all the health care professionals involved in the data collection and interpretation to ensure that study findings are grounded in participants' experiences.

I have contributed substantially to the LILAC Grant application submitted to Yorkshire Cancer Research (YCR) with Professor Velikova as PI. Six months prior, I submitted a grant application on PROMS in lung cancer surgical patients to the same funders who had a special call that year. Unfortunately, it was rejected. From the feedback I received, I started building up the idea of this more comprehensive prospective evaluation, and after meeting with my primary supervisor Prof Velikova, we refined the project. From the experience of the previous audit I immediately realised the importance of involving all the clinical professionals around lung cancer care. I directly involved the other co-applicants (other surgeons, oncologist and chest physicians) in this grant as they played a key role in my previous experience of PROMS collection in this hospital. I worked on the application which was awarded to Professor Velikova following two stages and minor comments from the reviewers (YCR award L399).

The evaluation of PROMS in clinical practice is challenging. This is particularly true in the area of surgical fast-track and Enhanced Recovery After Surgery (ERAS®) policy where

the patient is admitted the same day of the operation (127). The main goal of the ERAS preoperative assessment is to identify patients at higher risk, to address modifiable risk factors, and to optimize organ function before surgery, so the patient could be in the best possible condition for the operation. Therefore, during the preoperative phase, attention is focused on the risk assessment and optimization of the patient's medical condition. This results in most patients having a single day when all these preoperative appointments occur, allowing minimal time for consideration of clinical trials or more generically, research.

Although observational study designs are uncontrolled (unlike a randomised controlled trial) this method can provide evidence of effectiveness and is often quicker and cheaper to perform.

In the past, several attempts have been made to compare SABR and VATS resection in a more consistent way. However, all these analyses are limited due to the quality of the retrospective data and, even with propensity matching, case selection and other significant factors which cannot be accounted for. Thus, a randomized trial including borderline patients and comparing the two treatments would be the ideal method to answer the clinical question whether one treatment is more cost-effective than the other. Indeed, randomized trials have been attempted in the past and all failed to recruit and closed prematurely (ROSEL (NCT00687986), STARS (NCT00840749), ASOSOG-RTOG (NCT01336894))(41, 128-130). Their target sample sizes were never achieved, and these trials have invariably shown that patient preferences play a crucial role in the decision-making process. Leeds has been the leading site of the SABRTooth trial (NCT02629458) (42). It was designed as a feasibility study having a primary endpoint as a "steady-state" recruitment rate of at least 3 patients per month. The study however fell short of recruitment and less than 50% of patients were randomized within the target study

timeframe. The reason for declining randomization amongst the eligible patients was predominantly their preference towards one or the other treatment, with 60% of patients preferring SABR and 30% surgery.

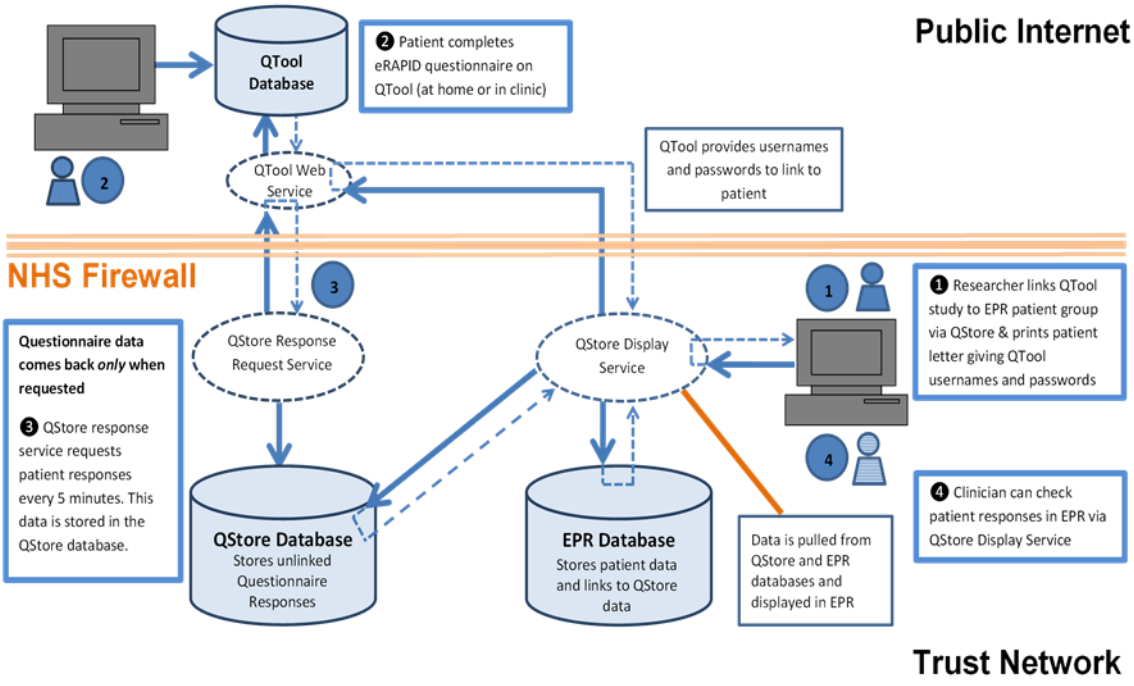
An alternative approach is to prospectively follow the treatment outcomes for those who received SABR or surgery, in an observational cohort, acknowledging however, these groups may not be directly comparable, with those who received surgery being 'fitter' than those who received SABR. We decided therefore to gain a better understanding of patient's preference, utilising patient-reported outcomes within our centre following those patients for one year after their treatment journey. This form of observational research provides a deep consistent knowledge of pre-treatment risk factors, patient preferences and post treatment quality of life trajectory among these two patient populations and will give further feedback via the qualitative stream informing future feasibility studies in this area.

2.4.1 ePROMS system at local centre

Previous research in our group using electronic PROMS (ePROMS) systems in patients treated with chemotherapy and radiotherapy has focused on use within a randomised clinical trial (RCT) (104). The PCOR group has a long-standing history in developing and piloting electronic methods for capturing PROMS in clinical practice. In particular, the hybrid system of incorporating ePROMS in the electronic patient record developed by the group has been a unique opportunity for me to adopt this approach in my research. An existing web-based questionnaire tool (QTool) previously commissioned by the PCOR group by a private software company (X-Lab) was further used to meet the needs of LILAC. QTool had previously been successfully used in a large-scale study to collect patient reported data from cancer survivors and link it with cancer registries (103, 131).

In order to facilitate the display of patient reports in individual EPRs, a link was created between QTool and the existing electronic health record system (PPM) used in the Leeds Teaching Hospitals Trust. This link was created via a service called QStore. The main challenge of this task was to maintain security of patient data within the EPR and work within the strict regulations of the network used by the NHS (Figure 2-2).

Figure 2-2 QTool/EPR system

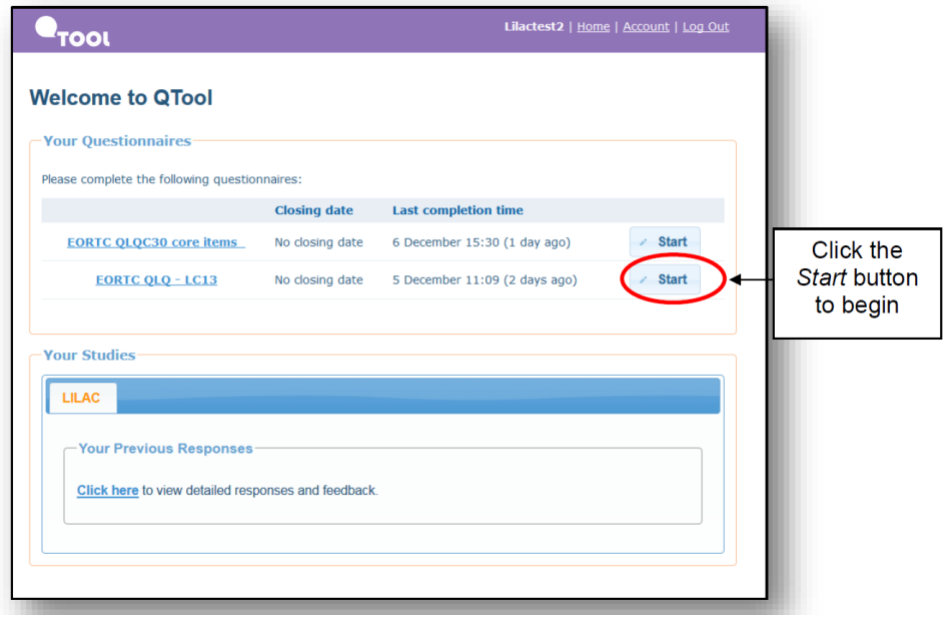


Patients enrolled in the LILAC study, were offered to complete the questionnaire on-line, but if not meeting the patient's need the option to fill them in on paper was also given.

I conducted a preliminary audit during the grant application phase, to assess internet use level among lung cancer patients in our setting in Leeds (through the surgical and oncological clinics). We surveyed 45 patients in total and 34 (75.6%) of the patients said they had access to the internet, though some of these were via a family member's device. On the day of approach in clinic, if the patient agreed to complete the LILAC study on line, they were given a brief demonstration of the system and provided with

an A5 postcard with their unique logon details, plus a detailed user manual with instructions on how to access the LILAC website and complete the questionnaires (Figure 2-3).

Figure 2-3 QTool online interface



Once a patient had completed a QOL questionnaire, this was immediately documented in their individual EPR, on the local system Patient Pathway Manager (PPM) (within a five-minute period). Clinicians use PPM to manage patient care during consultations and also during ward rounds. LILAC data were easily accessible on a tab within PPM, and accessed in a similar way to blood results. Clinicians were prompted to review patient data in routine consultations. Clinicians had the option to view patient data in graphical form (see Figure 2-4 for example) or can view the data in a table.

Figure 2-4 EPR QTool results

The reminder system has been used effectively within our organisation for previous studies. Invitations are sent after treatment at selected time points using either electronic, text message or paper methods.

At the end of their 12 month study period, a subset of participants per treatment group were purposively sampled by gender, age and method of completion and asked to take part in a semi-structured interview about their experiences of taking part in the LILAC study. A subset of health care providers were also purposively sampled by gender, age and speciality.

Patients were approached consecutively as they completed the study, with an aim to interview 5-10 patients overall from each treatment group. Interviews took place in a private room in the outpatient clinic at St James' University Hospital, Leeds or over the telephone if requested. Major details of the qualitative analysis will be given in Chapter 6.

The duration of follow-up for the prospective study was limited by the timeframe of the PhD and also determined, as it has been already demonstrated, that the main effect of surgery on QOL is within the first year. Chapters 6 and 7 provide baseline and an interim analysis of the study after 6-months follow-up. This early analysis aims to evaluate the acceptability and feasibility through assessment of recruitment rates, attrition, missing

data and early feedback in the form of questionnaires from patients. The final analysis, once follow up is completed in June 2019, will incorporate the 12 month collection point and will help in understanding the mid-term evolution of QOL. Early results on the trajectory of QOL are presented in Chapter 8. The sample size calculations for the prospective study were based on estimated numbers of patients treated each year in Leeds Cancer Centre for both treatments and also on the basis of the QOL results of the Pilot study. All elements of the LILAC study methodology will be described in detail in Chapter 6.

2.5 Summary

In summary, this thesis has a number of different methodological approaches to better describe the QOL evolution in early stage NSCLC.

The audit method was very valuable in testing the routine collection of QOL data in the clinical setting. The mixed methods approach gave me the opportunity to formally collect and analyse the data. It also helped in answering a broader range of research questions through the feedback of those actively involved with the research. The integration of quantitative and qualitative data in the form of a mixed methods study had the great benefit of enriching the analysis and findings of my thesis and gave us important aspects to address in future study design.

Chapter 3 PROMS assessment in early stage NSCLC: a systematic review

3.1 Introduction

This chapter aims to establish the most commonly reported quality of life outcomes after surgery and radical radiotherapy for early stage NSCLC using a modified version of systematic review and scoping review methods called Best Evidence Topic (BET). I was interested in exploring whether the current literature has evaluated these two treatments from the patient's point of view. For the purposes of this review I have decided to concentrate on quality of life rather than expand across more PROMS, not only as this is the most represented in the literature, but also as this is the main aspect the patients want to discuss when in clinic and during his/her treatment choice discussion.

Quality of life was also chosen as the main argument for a full review for this PhD for other reasons. Firstly, there has been considerable debate in treatment practice in lung cancer over the past few years with the increased use and improved results of radiotherapy regimens, arguing the reduced impact of these treatments on patients' daily lifestyle (34, 86, 132). Secondly, I am one of the co-authors on a full systematic review publication carried out in association with the EORTC QOL Group as part of a project reviewing and updating the EORTC Lung Cancer Module (89). I have also published a Scopus review in 2015 on the best available evidence published on quality of life after lung resection for cancer and aiming at identifying topics deserving further investigations (86).

3.2 Role and original contribution

I was responsible for all aspects of the planning, design and implementation of the review, with support from other researchers for double coding and data extraction. I have prepared the results for publication and this has been accepted for publication in the Journal of Thoracic Disease (133).

3.3 Aims and objectives

This systematic review has two main objectives:

1. To investigate the difference in terms of Quality of Life impact after two radical treatments of early stage NSCLC (VATS resection and SABR).
2. To consider the QOL questionnaires used for data collection and explore methodological issues (attrition rate, follow-up).

3.4 Systematic review methods

I adopted an approach pioneered in emergency medicine, namely the BET approach. Clinicians select a clinical scenario from their daily practice that is highlighted as an area of controversy. From this, a three-part question is generated and this is used to search Medline or other databases for relevant papers. Once the relevant papers are found, these papers are critically appraised using validated checklists and the results are summarized. A clinical bottom line explaining what should be discussed with patients is reached following this process.

This is a new and pragmatic approach to clinical review, especially when published evidence on the topic is very limited.

All electronic searches were conducted in Medline, PsycINFO and EMBASE and included studies from 1995 to present (10/2017). Studies using a prospective or cross-sectional design were included. Systematic review and meta-analysis studies were excluded. As this is a relatively new field of research, the systematic reviews published on PROMS in early stage NSCLC were including the same studies that were selected to be incorporated in my BET. Following one reviewer's suggestion, I have reviewed ad hoc the three systematic reviews published in the field making sure all the studies considered were included in this BET. Only one meta-analysis limited to surgical patients including two papers has been published on this topic. It was not included as the primary aim was to compare VATS and Open approach, which does not correspond to our research question.

The search strategy was restricted to studies that included QOL assessment for early-stages NSCLC radical treatments: VATS Surgery and SABR. We excluded studies with less than 20% of minimally invasive surgical approach (VATS) as this is now considered the standard approach for early stages NSCLC by most guidelines, for example the American College of Chest Physician (ACCP) guidelines (15).

Data has been extracted on trial design and QOL results to provide a quantitative assessment of the type of measures and results in use in clinical trials.

3.4.1 Protocol and registration

I have been in contact with the researchers at the York Centre for Dissemination in order to register this systematic review on the PROSPERO (International prospective register of systematic reviews) database. However, York researchers are reviewing the

guidelines aiming to include Best Evidence Topics in their database so this BET was not included on PROSPERO. There were no major deviations from the protocol.

3.4.2 Eligibility Criteria

The review question and eligibility criteria were developed and refined using PICOS (Population, Intervention, Comparator, Outcome, Study design) criteria outlined in Table 3-1. All relevant publications including published abstracts, protocols and qualitative studies were included. However, discussion papers, meta-analysis or systematic reviews were excluded. The three Systematic Reviews in this field were not relevant to our research question as they were not comparing QOL between the two groups. Only one small meta-analysis has been published in this area but its main objective was to compare the VATS and Open approach in the surgical cohort.

I reviewed the three papers and all studies were already considered in our final quantitative analysis.

Criteria were piloted with the help and support of the University of Leeds library researchers on a subset of 10 randomly selected papers and subsequently refined and clarified before running the final search.

Table 3-1 PICOS Criteria

PICOS	
Population	- Adults > 18, no upper age limit - Males and females - Worldwide - Early stage NSCLC (I-II) - Receiving radical cancer treatment
Intervention	- Radical surgery (including anatomical lung resection with systematic nodal dissection) - SABR treatment
Comparator	- The review included studies with both comparators including those with no comparator
Outcomes	- We aimed to collect where available, information on reason for not being submitted to radical surgery - We also aimed to collect information on any patient centred outcomes, including any quality of life measures
Study design	- The review was not restricted to RCTs, and observational studies with any quality of life evaluation data were included. Patients had to be ideally filling the quality of life questionnaires over time and there had to be at least one intended time point of postoperative assessment.

3.4.3 Information Source

Studies were identified from systematic searches of MEDLINE, EMBASE, PsychInfo, Web of Science, databases in October 2017. Searches were updated in February 2018. Reference lists of relevant publications were screened to identify papers not picked up by the electronic searches. In addition, citations of selected key papers were searched.

3.4.4 Search strategy

The search followed Centre for Reviews and Dissemination recommendations for undertaking systematic reviews and PRISMA guidelines (134). Only English language publications were included. A detailed example of the search strategy used for MEDLINE is outlined in Table 3-2**.** This search strategy was adapted for each of the databases.

Table 3-2 Search Strategy example

Database: Ovid MEDLINE
Neoplasms/ 1. "quality of life".mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (237685) 2. "patient reported outcome".mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (4270) 3. "eortc qlq".mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (2574) 4 "short form 36".mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (8159)

5 “functional assessment of cancer therapy”.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (1570)

6 “brief pain inventory”.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

7 NSCLC.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

8 “stage I”.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

9 exp lung neoplasm/ (138275)

10 “lung cancer”.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (94163)

11 (surg* and resect*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (145394)

12 lobectom*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (11896)

13 segmentectomy.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (1939)

14 “sleeve resection”.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (370)

15 (sbrt or sabr).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (2450)

16 “Stereotactic ablative radiotherapy “.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (303)

17 1 or 2 or 3 or 4 or 5 or 6 (243279)

18 7 or 8 or 9 or 10 (176028)

19 11 or 12 or 13 or 14 (153219)

20 15 or 16 (2518)

21 19 or 20 (155517)

22 17 or 18 or 21 (546389)

23 17 and 18 and 21 (505)

24 VATS.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (3001)

25 "video assisted thoracoscopic lobectomy".mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

3.4.5 Study selection

For initial screening, a decision for inclusion was made based on title and where available, abstract. This was carried out by myself using a cautious approach erring on the side of over inclusion. Following this, two reviewers (CP and ES) independently reviewed all relevant studies. In cases of disagreement the articles were revisited and reviewed collectively to reconcile differences and achieve consensus. I then extracted and analysed the data. Where there was insufficient information to decide, authors were contacted for further information. If no response was received within two weeks, a final decision was made based on available information.

3.4.6 Outcome measures examined

Four hundred and twenty-eight papers were found using the reported search. From these, 1 paper was identified that provided the best evidence to answer the question

and 16 papers were retained in order to provide supportive evidence for the research question.

3.4.7 Data extraction and type of information extracted

Studies including patients reporting on different aspects of QOL as a primary or secondary outcome were considered. PROMS were defined as any reports coming directly from the patient (72). Multi-dimensional PROMS (for example a measure covering different aspects of functioning such as physical, emotional or cognitive function) or single-item health outcomes were included if patient-reported.

Data was extracted into a predefined database for each study on (1) basic demographics (e.g. publication year, country, design); (2) clinical demographics (e.g. overall sample size, treatment regimens, primary endpoints); (3) clinical outcomes (e.g. survival measure(s) used, grade and percentage of toxicity reported) and (4) quality of life results for each time-point.

3.4.8 Quality assessment of studies and PRO reporting

Internal validity was assessed by applying the validated 16-item quality assessment tool (QATSDD) (135) as this can be applied to a methodologically diverse set of research articles. It was undertaken alongside data extraction. QATSDD was developed at the University of Leeds. It contains 16 reporting criteria scored on a scale from 0 to 3 (Not at all/Very slightly/Moderately/Complete). These criteria apply to quantitative and qualitative studies.

PROMS quality assessment was adapted from the recently published ISOQOL recommended standards (136).

3.4.9 Narrative Synthesis

A narrative synthesis was undertaken using the guidelines outlined by the Economic and Social Research Council (ESRC) (137). Microsoft excel was used to manage data. At the beginning, information from multiple publications relating to clinical outcomes were kept to assess whether clinical information could bias the PROMS results. In the final synthesis, that information was removed by the narrative text as the guidelines of BET were strict in terms of word limits.

3.5 Systematic review results

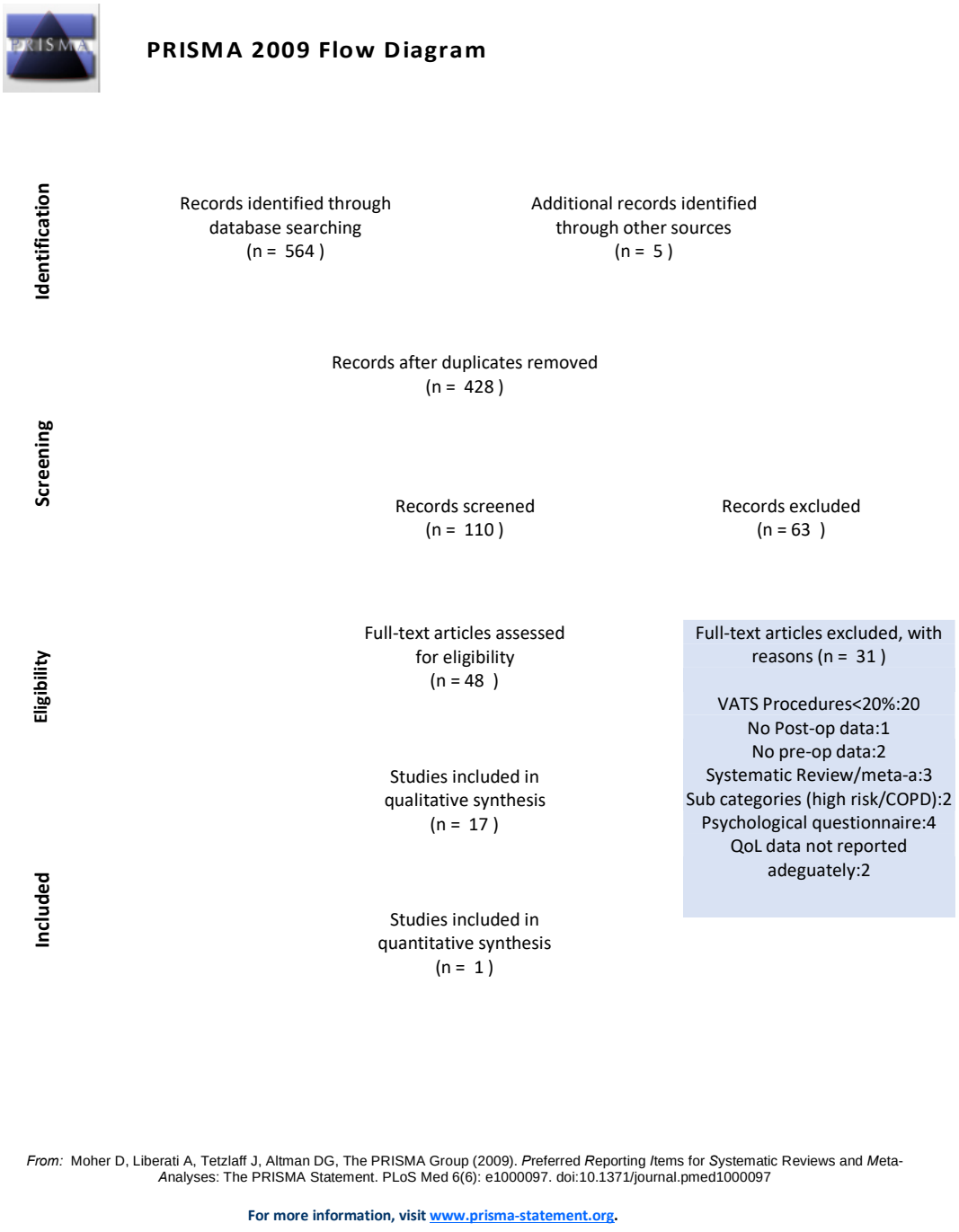
The search yielded 569 records (Figure 3-1). 428 records were screened after duplicates and articles published before 1995 were removed. 17 publications fulfilled the inclusion criteria. All seventeen studies appraised are summarised in

Table 3-3. The median duration of follow up for all studies was 16.6 months (range 3-60 months).

16 studies were identified, which separately investigated the effect of SABR or VATS lobectomy for early stages NSCLC on QOL. Only one RCT of 22 patients has been identified which directly compares the QOL outcomes of medically operable stage IA NSCLC patients treated with either SABR or surgery (34). They found similar results in most of the QOL scales, but also that SABR may have advantages in the global health and indirect cost of productivity loss. They have used validated instruments at baseline, and up to 24 months post-treatment: the European-Organization-for-Research-and Treatment-of-Cancer QOL Core questionnaire (EORTC QLQ-C30) and its lung cancer

supplement (LC-13). This paper is a secondary analysis of a non-blinded, phase 3 RCT of SABR versus surgery for stage IA NSCLC patients in The Netherlands which was prematurely terminated in 2010 due to slow accrual. Indirect costs of productivity loss were calculated with the Short Form Health and Labour Questionnaire (SF-HLQ, which includes work absences, reduced efficiency at work, and substitution for unpaid work) and were collected at the same time points (baseline, and then 3, 6, 12, 18, and 24 months post-treatment). They have correctly implemented the concept of clinical significance in PROMS study: a minimum threshold of a 10-point decrease (for global and functional scales) and increase (for symptom scales and items) were employed to constitute a clinically meaningful difference in scores (138). Time to deterioration (TTD) in PROMS was also calculated from the time of randomization to first appearance of a significant difference in PROMS scores. Patients without a documented clinically meaningful difference in PROMS were censored at the time of last assessment. One of the 11 surgical patients was submitted to a wedge resection (not anatomical lung resection) but included in the ITT analysis. Median follow-up was 42 months with death reporting not completely clear as defined as “due to comorbidity” in the surgical arm. In all comparisons, only global health status was found to be significantly worse on univariable cox proportional hazard modelling for surgical patients when compared to SABR (HR 0.19). However, the most important finding which can be found in the figures only, is that at 12 months only 5 surgical patients (50% of return rate) completed the questionnaires, 6 at 24 months and 5 at 30 months. With the small number of patients in the surgical follow-up it is not possible to make definitive conclusions regarding PROMS.

Figure 3-1 PRISMA flow chart



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Table 3-3 Search results

Author, date and country, Study type (level of evidence)	Patient group	Outcomes	Key results	Comments
RCT Trial				
Louie et al, Radiother Oncol, Netherlands (2015)(34) RCT (level 2b)	Secondary analysis of a non-blinded, phase 3 RCT of SABR versus surgery for stage IA NSCLC patients. 22 patients QOL Tools: -EORTC QLQ-C30 and LC-13 - EQ-5D FU: baseline, and then 3, 6, 12, 18, and 24 months	Time to Deterioration of at least 10-point decrease (global/functional scales) and of at least 10-point increase (symptom scales/items).	GH: Surgery 8, SABR 2 (HR 1 Vs 0.19, p=0.038). RF: Surgery 7, SABR 4 (HR 1 Vs 0.47, p=0.22) EF: Surgery 4, SABR 1 (HR 1 Vs 0.25, p=0.21)	Small sample size
SABR STUDIES				

Lagerwaard et al, JTO, Netherlands (2012) (139) Cohort Study (level 3)	382 patients QOL Tool: EORTC QLQ-C30 FU: Baseline, 3,6,12, 18 and 24M	*Baseline* GH PF RF Dyspnoea Fatigue Insomnia *Changes over time*	 62.9 ± 1.1 61.8 ± 1.1 63.5 ± 1.5, 47.1 ± 1.7, 37.4 ± 1.3 21.1 ± 1.6. PF decreased in 24M ($p < 0.01$) but not clinically significant (>than 10 points)	15.4% of patients refused surgery. Drop-out: 64% and 61% of patients, were unavailable for follow-up at 18 and 24 months

Mathieu et al, Pract Radiat Oncol, France (2015)(140) Cohort Study (level 3)	45 patients QOL Tool: QLQ-C30 and QLQ-LC13 FU: Baseline, 2,6,12,18,24,30,36M	Baseline-GH Baseline-PF Baseline-EF Social Functioning decline QLQ-LC13 coughing symptom	66 ± 20% 73 ± 22% 77 ± 26% Transient declines of 12 ± 29% 12M 11 ± 29% 24M Reduction 13 ± 17% 30M 13± 22% at 36M	16% of patients refused Surgery Data from patients who had disease recurrence were excluded. Collection of QOL data at the 2 and 3Y in 63% and 33%

Ubels et al, Radiation Oncology, Netherlands (2015)(141) Observational Study (level 3)	39 patients QOL Tool: EORTC QLQ-C30, EORTC QLQ-LC13 FU: BL, 3 weeks, 2, 4, 6, 9, 12, 15, 18, 21, 24M, then every 6M until 5 years.	GH PF, RF and Cognitive Functioning, Dyspnoea EF	Near the baseline in the first year. Then improve to decline again over the 5 years ($p < 0.0001$); significantly improved over time; fatigue deteriorated over time ($P = .05$); Deteriorated over time ($P = .006$). Improved significantly at 1 year compared to the baseline.	15% refused surgery At 5 years only 10 patients were still alive without progression and had filled the QOL survey

van der Voort van Zyp et al, Int. J. Radiation Oncology Biol. Phys, Netherlands (2010)(142) Observational Study (level 3)	39 patients QOL Tool: EORTC QLQ-C30, EORTC QLQ-LC13. FU: baseline, 3 weeks, 2, 4, 6,9,12M	GH, PF, RP, SP EF	No changes over time; Improvement over time (P =.02)	15% patients refused surgery Small sample size. The lack of >10-point changes suggests that there are no perceived changes in QOL scores.
Widder et al, Int. J. Radiation Oncology Biol. Phys, Netherlands (2011)(143) Cohort study (level 3)	Medically inoperable patients: 27 3D-CRT Vs 202 SABR QOL Tool: EORTC QLQ-C30 GH and PF + Dyspnoea LC13 FU: 3-6-12months	Dyspnoea PF GH	Increase by 3.2 (95% CI: 1.0–5.3; p < 0.01). Stable for all patients except for those with a high CCI. No significant changes	Different sample sizes (202 Vs 27). Comparison between techniques.

Ferrero et al, Lung Cancer, Italy (2015)(144) Cohort Study (level 3)	30 patients with inoperable Stage I NSCLC QOL Tool: Lung Cancer Symptoms Scale (LCSS) FU: Baseline, 1.5, 4.5, 7.5, 10.5M	Fatigue (baselines V s135D)	29 Vs 39,8 p=0.05 No other significant changes	Small sample size.
Jain et al, Radiat and Oncol, UK (2013)(145) RCT (level 2b)	54 patients with NSCLC <5cm. Comparing two groups: Group 1: 4 days of SABR Group 2: 11 days of SABR QOL Tool: EORTC QLQ-C30 and LC-13 FU: discharge, 1 and 4M	*(Group 1Vs Group 2)* PF RF Dyspnoea % of patients with a clinically meaningful worsening (>10points)	 BS: 79 Vs 68,6 4M: 71,3 Vs 69,9 BS: 93,8 Vs 71,6 4M: 83,3 Vs 77,3 BS: 25,9 Vs 44,4* 4M: 38,5 Vs 26,7 -Dyspnoea 1M: 44.4% vs 15.4% * 4M: 38.5% vs 12.0% * -PF 4M: 46.2% vs 16%*	Small sample size and limited follow-up.
Videtic et al, Support Care	22 patients	Global scores difference 1-12M	109 versus 112.	4.8% patients

Cancer, US (2013)(146) Cohort study (level 3)	QOL Tool: FACT-L and UCSD SOBQ (University of California at San Diego Medical Centre-Pulmonary Rehabilitation Pro-gram Shortness-of-Breath Questionnaire) FU: Baseline, 3,6,9 and 12, months			refused surgery. Limited sample size A non-significant 9-point drop in mean UCSD SOBQ dyspnoea scores.
Sun et al, J Community Support Oncol, US (2014) (147) Cohort study (level 3)	Observational study on 19 patients treated with SABR QOL Tool: FACT-L, Memorial Symptom Assessment Scale (MSAS) and FACIT-Sp-12 FU: Baseline, 6 and 12 weeks	QOL Emotional domains	No detrimental changes in QOL scores over time; Improvement in nervousness and worry scores over time but no significant change in overall emotional functioning	Small sample size
VATS STUDIES				
Bendixen et al, Lancet Oncol, Denmark (2017)(148)	RCT VATS Vs anterolateral thoracotomy	GH	VATS Baseline: 73,2 Open Baseline: 73,3 VATS 4W: 67,5	All the differences were only in

RCT (level 2a)	201 patients QOL Tools: EORTC QLQ-C30 and EQ-5D FU: Baseline, 2, 4, 8, 12, 26, and 52 weeks	PF EF	Open 4W: 64,8 VATS 52W: 77,2 Open 52W: 74,1 VATS Baseline: 88,6 Open Baseline: 88,4 VATS 4W: 83,9 Open 4W: 75,8* VATS 52W: 86,1 Open 52W: 82,9 VATS Baseline: 77,5 Open Baseline: 77,4 VATS 52W: 90 Open 52W: 83,03* EQ5D only significant differences were in self-care and anxiety.	few time points. They did not use the Lung Cancer module of the EORTC QLQ-C30.
Burfeind et al, J Thorac Cardiovasc Surg, US (2007)(65) Cohort study (level 3)	422 patients submitted to lobectomy. QOL comparison group 1: <70 years and group 2: >=70 years QOL Tool: EORTC QLQ-C30+2scales of LC13	*(group 1 Vs 2)* PF EF	Baseline: 83,7 Vs 81 3months:77,9 Vs 73,9 12months: 81,9 Vs 78 Baseline:74,1 Vs 78,9 3months:74,2 Vs 77,2 12months: 78,5 Vs 82,4	Retrospecti ve analysis. The most commonly missed survey time point was the 3-month survey with

	FU: 3-6-12months	GH	Baseline:18,3 Vs 16,8 3months: 33,4 Vs 26,1 12months: 22,2 Vs 17,6	28% of group 1 and 38% of group 2.
Handy et al, Eur J Cardiothorac Surg, US (2010)(149) Retrospective study (level 3)	241 patients submitted to Lobectomy (OPEN: 192 Vs VATS: 49). QOL Tool: Short Form 36 Health Survey (SF-36) and Ferrans and Powers quality-of-life index (QLI) FU: baseline and 6months	*Difference from baseline to 6M (Open Vs VATS)* PF GH Bodily pain Role Physical MH Energy	-11.6 -1.4 (p: 0.042). -3.3 Vs 4.8 (p:0.010) -4,4 Vs 9,6 (p:0.020) -18,6 Vs 12 (p:0.002) - 0.5 Vs 4.2 (p:0.38) -3.6 Vs 5.3 (p:0.054)	Limited follow-up (6 months)
Khullar et al. Ann Thorac Surg, US (2017) (150) Cohort study (level 3)	127 patients QOL Tool: 7 fixed-length PROMIS instruments FU: baseline, 1 and 6months	PF Pain intensity, interference, fatigue, and sleep impairment	Significantly lower (worse) at 1M visit than at baseline. All significantly higher (worse) at the 1M. No difference identified at 6M.	Short follow-up Only 70 VATS lobectomies included.

		Anxiety/fear and depression	significantly improved after the operation	
Rizk et al, Ann Thorac Surg, US (2014)(29) Cohort study (level 3)	206 Stage I NSCLC patients (74 VATS Vs 132 Thoracotomy) QOL Tool: SF-36 Physical component summary [PCS] and mental component summary [MCS] FU: baseline, 2 weeks, 4,8 and 12months	MCS PCS Pain	Baseline: 42,4 Vs 43,5 4M: 43,6 Vs 44,9 (p:0.036) 12M: 47,2 Vs 49 (p:0.08) Baseline: 48,9 Vs 50,3 4M: 45,7 Vs 45,5 (p:0.86) 12M: 48.1 Vs 48 (p:0.93) BPI: no statistical difference between two groups	Only 59% patients completed all the surveys.
Fagundes et al, J Thorac Cardiovasc Surg, US (2015)(151) Cohort Study (level 3)	60 stage I-II NSCLC patients treated with open and VATS lobectomy QOL Tool: MD Anderson Symptom Inventory (MDASI) FU: baseline, 3 and 5 days after surgery, and weekly for 3M	Moderate to severe symptoms	Day3: 51.6% for pain, 59.7% for fatigue, 54.8% for drowsiness, 33.9% for shortness of breath, and 56.5% for disturbed sleep. 3months: all symptoms had improved to better than preoperative	No objective measures affecting duration of hospital stay

Li et al, Chest, China (2002)(31) Cross-sectional study (level 3)	51 patients with NSCLC following resection, comparing VATS with thoracotomy QOL Tool: EORTC QLQ-C30, EORTC QLQ-LC13, Self-developed module. FU: 33.5M (VATS) and 39.4M (open)		Fatigue (74–92%), Coughing (75–82%) Dyspnoea (75–85%) Pain (67–71%)	Additional non-validated surgery-related questions. One-off survey.

*FACT-L: Functional Assessment of Cancer Therapy- Lung questionnaire; FACIT-Sp-12: Functional Assessment of Chronic Illness Therapy-Spirituality Tool; GH: General Health; RF: Role Functioning; EF: Emotional Functioning; HR: Hazard Ratio; PF: Physical Functioning; *: statistically significant; CCI: Carlson Comorbidity Index; MCS: Mental Composite Score; PCS: Physical Composite Score; MH: Mental Health*

3.5.1 SABR STUDIES

Out of nine studies evaluating the impact of SABR on QOL, only five studies specified the percentage of patients who refused surgery. In all other studies patients who had SABR treatment were patients considered medically inoperable and therefore are assumed to have more medical comorbidities and/or poorer cardio-pulmonary function than patients undergoing surgery.

Lagerwaard et al. (139) conducted the largest study on 382 patients over a period of 24 months. Physical functioning was the only QOL domain to statistically significantly worsen, though by less than the clinical meaningful significance of 10 points (138). Physical functioning in fact decreased by more than 10 points in 26% of patients, remained stable in 53%, and had improved in 22% after one year.

Mathieu (140) reported favourable long-term QOL and pulmonary function in 45 patients treated with SABR with a follow-up longer than 3 years. They also reported a QLQ-LC30 emotional score improvement at 36 months. However, the exclusion of patients with recurrent disease may have affected the QOL results.

Ubels et al. (141) prospectively studied QOL in 39 inoperable patients for 5 years. Although the emotional functioning scores improved significantly, dyspnoea slowly worsened 2 years after SABR. The trajectory of the global health showed that it remained near the baseline value during the first year, improved at 18 months and then significantly declined to the baseline value during the next years.

One of the first studies to explore QOL after SABR treatment was from van der Voort van Zyp et al. (142). The only significant change observed in 39 patients was an improvement in emotional functioning.

Widder et al. (143) looked prospectively at longitudinal changes of QOL parameters after SABR (202 patients) or three-dimensional conformal radiotherapy (27 patients). They found that global QOL and physical functioning were stable at any follow-up within the first year. They also reported a statistically significant increase in dyspnoea, although the observed changes were not clinically significant.

Ferrero et al. (144) study of 30 patients is the only one to report a clinically and statistically significant increase in fatigue after 135 days.

Jain et al. (145) reported dyspnoea, fatigue and coughing to be worse at baseline in 54 patients treated with SABR over 11 days compared to 4 days of treatment. However, more patients treated on 4 consecutive days experienced a clinically meaningful increase in dyspnoea at 1 and 4 months after treatment.

Videtic et al. (146) conducted a small prospective study in 22 patients which did not find any statistical difference after 12 months in terms of QOL. They reported however, a 9-

point drop from baseline to 12-week scores on the patients' UCSD dyspnoea questionnaire, approaching clinical significance of 10 points.

Sun et al. (147) showed that QOL was not seriously impacted in a small cohort of 19 early-stage lung cancer patients after 12 months of follow-up. The functional domain had the lowest score of all the subscales measured with the Functional Assessment of Cancer Therapy-Lung (FACT-L).

3.5.2 VATS Surgical STUDIES

The surgical studies investigating specifically the effect of minimally-invasive anatomical lung resection (studies with more than 20% VATS) on QOL were characterized by small sample sizes and limited longitudinal assessments. Five out of 7 studies' primary aim is in fact the direct comparison between different surgical access (open versus thoracoscopic).

Bendixen et al. (148) conducted the first RCT on 201 patients describing the trajectory of pain and QOL of open versus VATS lobectomies for cancer. With a follow-up of 52 weeks, they found QOL in the VATS group was significantly better than that of an age-matched cohort from the Danish population. After two weeks the worst levels of QOL were observed and then QOL gradually improved over 52 weeks.

Burfied et al. (65) showed within 422 patients submitted to lobectomies that QOL worsened at 3 months. However, at 6 and 12 months, all domains had returned to baseline except physical functioning, which remained below baseline in patients older than 70 years. Emotional functioning improved postoperatively in older and younger patients. Handy et al. (149) reported that compared with preoperative values, 6-months after resection, 49 VATS patients were not significantly different in physical function,

role physical, role-emotional, social function, mental health or energy. Postoperative categories of bodily pain and general health were significantly improved over preoperative values in the VATS group.

Most recently Khullar et al. (150), in the first attempt to implement PROMS into US national databases, evaluated 127 patients with the National Institutes of Health Patient-Reported Outcome Measurement Information System (PROMIS) platform. They confirmed a significant worsening in pain, fatigue, and sleep scores and a decrease in physical function 1 month after the operation. By 6 months, these had generally improved toward baseline. Anxiety/fear and depression both significantly improved after the operation. In 2014, Ritz et al. (29) prospectively compared 74 VATS and 132 open lobectomies. In both groups, QOL scores improved throughout the 12 months, and pain scores approached baseline levels by 4 months.

Fagundes et al. (151), conducted an interesting investigation on weekly symptom assessments in 60 surgical stage I patients from the third postoperative day to 3 months. All symptoms (except fatigue) returned to preoperative levels by the end of the first month and fatigue remained the most persistent symptom during the study.

Li et al. (31) included surgery-related questions in their retrospective study and found that 51 lung cancer patients following surgical treatment without recurrence had good QOL and high levels of functioning after a mean of 33.5 months follow-up, with no significant differences between the VATS and open groups.

3.6 Conclusions

3.6.1 Objective 1

This review demonstrated that there is not enough evidence to answer our research question regarding the differences in terms of impact on QOL after radical treatment of early stage NSCLC. In fact, only one small RCT was identified that provided evidence addressing the specific question proposed in our BET. It found that global health status was significantly worse for surgical patients when compared to SABR patients but only 22 patients were enrolled and 13 patients completed the questionnaires at 30 months.

As we did not find an answer to the specific question, we kept sixteen studies providing supporting evidence, but not directly comparing QOL between the two treatments. These studies assessed a total 832 SABR patients and 686 receiving anatomical VATS resection, and confirmed that in general physical components of QOL decrease immediately after treatment up to 3-months, returning to baseline after 1 year. Emotional functioning often supersedes the pre-operative values across treatments.

3.6.2 Objective 2

Regarding the methodological research questions, we found that there is no concordance in the choice of PROMS in the early stage NSCLC field. The most commonly used questionnaires were the EORTC QLQ C-30 and its lung cancer module LC-13. This was mainly expected in the SABR group as this is a questionnaire developed and validated primarily among patients receiving systemic and radiotherapy treatments. Three of the seven surgical papers have adopted them.

The remaining SABR papers adopted the FACT-L. In the surgical cohort there was less consistency (two used the SF-36, one the MD Anderson Symptoms Inventory (MDASI)

and one PROMIS scales). It was appreciated however, that most of these studies used lung-specific questionnaires, recognizing the importance of detecting specific respiratory symptoms in patients with lung cancer.

Most of the studies collected PROMS on paper. There was only one study offering electronic completion (150). Eligible patients were enrolled on the PROMIS Website and then completed the PROMS survey in clinic by using a touch-screen tablet device. The clinical research nurse was available for assistance as needed. They reported very good survey completion rates: 100% (127 of 127) for both the baseline survey and the initial post-operative survey and 90% (97 of 108) for the 6-month postoperative survey. However, they stated that patients unable to make their follow-up clinic appointments had their surveys administered over the telephone by clinical research nurses. They did not state how many completions were made by the patient themselves or by telephone via the research staff.

One study used a repeated computer/telephone interactive voice response (IVR)-administered symptom rating scale after patients were discharged from the hospital (151).

The follow-up time-points were very heterogeneous. In the surgical cohort the average was 12 months. In the 9 SABR studies the average follow-up was 19 months, however one study following the patients for 60 months had only 14 patients alive after the third year (150).

Table 3-4 and Table 3-5 show the information regarding the questionnaires compliance rates in both groups. The main finding is that studies were mostly done 5-8 years ago with most patients in service evaluation studies.

In fact, especially in the SABR group, most of the collection procedures indicate the patient was filling in the questionnaires immediately before the doctor appointment.

This may have increased the completion rate. Furthermore, the different sample sizes of the studies involved (from 19 to 382 patients) and different time-points of collection, has restricted the generalizability of these results.

Table 3-4 Surgical Studies compliance rates

	Baseline	3M	6M	12M	Comments
Bendixen (102 VATS patients)	58%			76%	32% of the total completed all the questionnaires
Burfeind (262 patients)		67%		72%	
Handy (49 patients)					Not reported
Khullar (127 patients)	100%		90%		PROMIS assessment through telephone calls from the nurses
Rizk (132 patients)	100%			60%	No details of death or withdrawals during the study
Fagundes (29 patients)					Not reported
Li (25 patients)					Not reported

Table 3-5 SABR studies Compliance Rate

	Baseline	3M	6M	12M	24M	Comments
Langerwaard (382 patients)		76%	62%	59%	39%	
Mathieu (45 patients)					63%	
Ubels (39 patients)				95% (20 pt)	100% (19 pt)	
Van der Voort van Zyp (39 patients)		90%	100 %	95% (19 pat)		
Widder (202 patients)		96%	74%	71%	44%	Questionnaires given before doctor's appointment
Ferrero (30 patients)						74% at 4M with 22 patients returning questionnaires
Jain (54 patients)		92% (25 pt per group)				
Videtic (22 patients)		81%		76%		
Sun (19 patients)						Not reported

3.6.3 Strengths and limitations

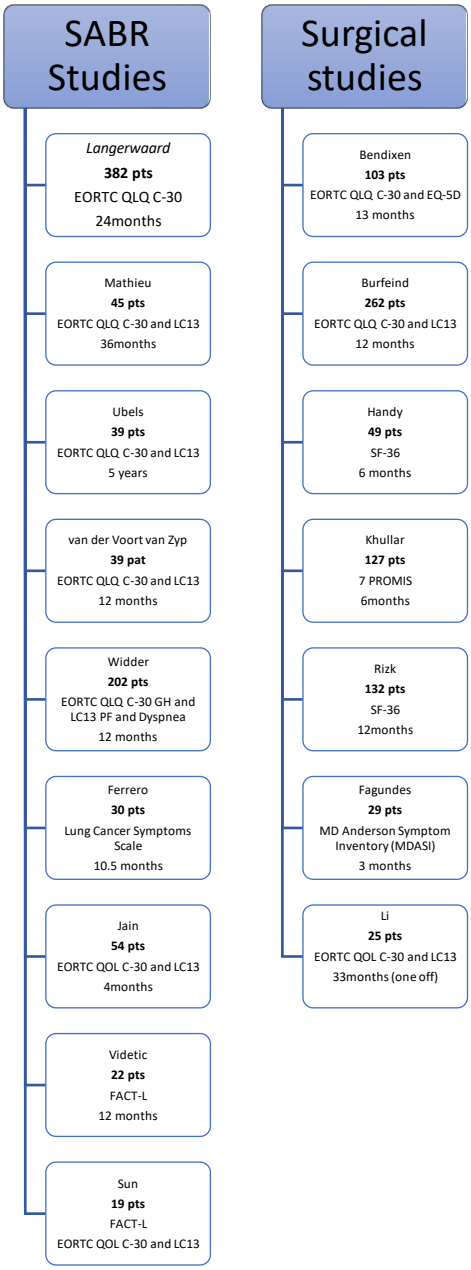
One of the main strengths of this review was that we have performed a systematic review using a strict, recognised, albeit pragmatic, methodological approach.

To my knowledge, this is the first review in this field to identify and detail all available QOL results after early stage lung cancer treatments, in addition to evidence relating to the type of questionnaires used in this field. As shown in Figure 3-2 the sample size of the studies included is very limited when considering the number of patients meeting

the eligibility criteria in each centre (SABR technique and VATS resections). Furthermore, although there are studies following the patients for more than one year, unfortunately the attrition rate is very high, evident in the Ubels paper where only 10 patients completed the questionnaire at 5 years (141).

We decided to include only English written papers for funding and time-related issues. We cannot rule out that the inclusion of different languages may have affected our results.

Figure 3-2 Supportive Studies with author, patients' number, instrument, latest follow-up



In order to meet the aims of this review, many publications were included which had limited information available and did not meet the precise inclusion criteria (what we defined as "supporting information"). However, this was necessary in order to answer the research questions and evaluate all evidence available.

3.7 Discussion

This Best Evidence Topic indicates that there is a clear paucity of evidence in QOL evaluation for these treatment modalities for early stage NSCLC. Only one small RCT (N=22 patients) in fact was identified that provided evidence addressing the specific question reporting that global health status deteriorates in more of the surgical patients compared to SABR.

Sixteen studies provided supporting evidence but did not directly compare QOL between the two treatments. The overall impression from these studies which assessed a total 832 SABR patients and 686 receiving anatomical VATS resections, is that physical components of QOL decrease immediately after treatment up to 3 months, returning to baseline after 1 year. Emotional functioning often supersedes the pre-operative values across treatments. Results from trials like the SABRTooth (42), STABLE-MATES (NCT01622621) (152) and VALOR (Veterans Affairs Lung-Cancer-Surgery or Stereotactic-Radiotherapy) (153) will give us information necessary to compare in a standardized way these two treatments in a comparable group of patients.

It was not possible to draw definitive conclusions on the adopted methodologies for the quality of life assessment as the sample size and study design, including collection time-points, as they were often too different.

We were able to demonstrate, however, that the most commonly used QOL instrument was a cancer-specific one, the EORTC QOL C-30, which in a few studies has been administered with the Lung Cancer specific module. This has enabled the following of cancer-related symptoms during the entire follow-up period. The latter was particularly important in the SABR studies where most of the questions were relevant to radiotherapy treatment. These data are confirming the limits of using the first version

of LC-13 in surgical NSCLC patients. This issue was one of the reasons informing the rationale for the project within the EORTC QOL group to update the Lung Cancer Module (154). I was leading the surgical group of researchers to implement, test and pilot surgical-related questions within this questionnaire.

The lack of definitive evidence in this field promoted the next step in this thesis. I understand better the value to pilot the routine collection of quality of life data in early stage NSCLC in clinical practice to investigate advantages and challenges in real-world settings (chapter 4 and 5).

Furthermore, this review gave me the opportunity to critically appraise the most common QOL instruments used in both these clinical groups, which were then administered in the prospective study.

Planning an observational prospective study with 250 patients of both treatment arms, gave me a unique opportunity to adopt a more quantitatively and qualitatively formal approach in the same setting. This work is described in the Chapters 6 and 7.

Chapter 4 Risk prediction-the role of preoperative quality of life

4.1 Background and overview of the chapter

In our preliminary work from the service evaluation project, we have been able to show that the subjective perception of a poor global health status is associated with the occurrence of postoperative cardiopulmonary morbidity after pulmonary lobectomy (155). This finding supports the adoption of a holistic approach during the surgical shared-decision-making process. In fact, the patient perceptions and values should be considered when counselling with patients and in the entire risk stratification process to tailor cancer treatment.

Nevertheless, risk stratification is aimed to find a measure of risk for a specific patient who is having a specific operation. It can be defined as the ability to predict the outcomes from a given intervention based on the pre-existing risks, i.e. less fit patients are expected to have worse outcomes for a given intervention than more fit patients. Our previous published work from this database focusing on the correlation between preoperative quality of life and objective clinical outcomes such as the postoperative complications, this chapter describes some of the exploratory analysis from the same cohort of surgically-treated lung cancer patients. I was particularly interested in focusing on the possible relationship between the quality of life and other clinical outcomes (like length of in hospital stay) and also with the most recent preoperative mortality and morbidity risk-adjusted classes. This analysis will support the ultimate aim of this thesis of better investigating the role of quality of life in the preoperative decision-tree and patients shared decision-making.

To better understand the context of the functional preoperative assessment in lung cancer surgery, I have briefly introduced the most important risk scores in thoracic surgery below.

4.1.1 Thoracic Surgery Risk Scores

Multiple systems have been developed over the last three decades though none can ever perfectly predict the patients' risk of death and serious complications. While in different surgical specialties there has been an extensive clinical use of risk stratification (like for example Parsonnet and EuroSCORE in cardiac surgery)(156), the application of scores in thoracic surgery is recent. So far, one of the commonly used systems in thoracic surgery has been the Thoracoscore(46). The risk scores allow surgeons to work out the risk of dying or experiencing postoperative events from particular operations, as long as the patients' medical conditions have been accurately identified. These scores can also help counsel patients properly patients during the process of gaining informed consent, help organizations with allocation of resources and aid with the assessment of the overall quality of care.

Objective outcome endpoints have been the topic of several publications and risk models have been published to assist the surgeon in stratifying the risk of morbidity and mortality based on physiologic characteristics of the patients and the extent of operation. The operative risk is usually presented in the form of proportions or percentages (i.e. 2% mortality risk). On the other hand, post-operative residual quality of life and independence are much more difficult if not impossible to predict, because they are subjective measures perceived by the individual patient within their own perceptual framework including social, cultural, emotional, health expectations, demographic, reference group etc. Furthermore, for so-called "high-risk" patients, the

impact of a slight deterioration in QOL for a patient with poorer QOL/more comorbidities before surgery could be more clinically significant and therefore a small numerical drop could have a much greater impact in QOL.

Currently there is no validated risk model or predictive equation that can reliably estimate the decline in physical or emotional components of QOL after lung resection for cancer. Permanent disability and loss of independence remains the main concern of surgical candidates, even more than immediate mortality or complications(157). These concepts are far more difficult to estimate and risk-stratify compared to objective outcomes such as mortality since they involve a high degree of subjectivity and because objective factors traditionally used to estimate mortality or morbidity have failed to show an association with physical or mental domains of QOL(158).

Furthermore, scoring systems are often used in our specialty to predict the probability of selected outcomes in groups of patients, thus enabling risk stratification. The main limitation is the lack of accuracy on an individual patient basis. Scoring systems may be accurate in estimating the mortality rate in a certain group of patients but they fail in identifying which of the patients will die after the operation. Therefore, they cannot be used as a selection tool for surgery, but only for estimating the risk of morbidity and mortality, which can be informative during preoperative counselling ideally leading to a more informed patient decision.

4.1.2 Charlson Comorbidity Index

The Charlson Comorbidity Index (CCI) is a score that quantifies the risk of mortality in relation to the multiple pathological conditions of a patient. It was originally derived in 1987 by Charlson and colleagues (159). They examined the impact of several baseline comorbidities on the mortality rate of 559 medical patients during their first year after

the admission. Seventeen conditions were found to correlate with mortality. Each of these conditions were weighted according to their independent contribution to the risk of death (expressed as a single numeric value). The sum of the individual pathological condition values indicated the CCI of a patient. In the original population the mortality risk progression related to the CCI was: CCI=0 - risk=8%, CCI=1 - risk=25%, CCI=2 - risk=48%, CCI>3 - risk=59%. Since its publication, the CCI has been widely used for the stratification of risk, and for measuring the burden of comorbidities on many groups of patients in the medical, surgical and intensive care fields. We have used the CCI in clinical practice as a useful comorbidity factor for research and audit purpose.

4.1.3 Physiological and Operative Severity Score for the enumeration of Mortality and morbidity (POSSUM)

The physiological and operative severity score for the enumeration of mortality and morbidity (POSSUM) was originally proposed by Copeland et al in the early 90's as a scoring system for auditing the quality of care in general surgery(160). Obtained as the sum of a Physiological Score (evaluating 12 baseline characteristics of the patient) and an Operative Severity Score (evaluating 6 operative factors), the POSSUM indicated a proportionally higher risk of morbidity and mortality associated with higher cumulative score values.

4.1.4 Thoracoscore

Based on the data collected within the French Society of Thoracic and Cardio-Vascular Surgery database (Epithor), the Thoracoscore is a risk index for predicting the in-hospital mortality after various thoracic surgery procedures (lobectomy: 24.1%, pneumonectomy: 6%, wedge: 43.4%, mediastinoscopy or other mediastinal surgery:

26.1%) (46). It was derived from the analysis of 10,122 patients (mortality rate: 2.1%), and its effectiveness was tested on a second group of 5,061 patients (mortality rate: 2.4%). Multivariable analysis identified several factors associated with mortality and included in the model for the prediction of in-hospital death: age, sex, dyspnea score, American Society of Anesthesiologists score, performance status classification, priority of surgery, diagnosis group, procedure class, and comorbidity score. The odds ratios of each of these nine factors were utilized in the prediction of mortality risk. On the other hand, different studies highlighted the limitations of the Thoracoscore when applied to specific subgroups of patients(161, 162).

4.1.5 ESOS

In 2005 Berrisford and colleagues, on behalf of the Audit and guidelines committee of the European Society of Thoracic Surgeons and the European Association of Cardiothoracic Surgeons, developed a model to predict the in-hospital mortality in patients undergoing their first lung resection(47). The analysis was conducted on data from 3,426 patients (wedge/segmentectomies 26%, lobectomies 59%, pneumonectomies 14%, lung volume reduction surgery 1%) collected in a European database gathered from 27 Thoracic Surgery Units from 14 different countries. The model to predict the in-hospital death was based on age, dyspnea score, American Society of Anesthesiologists score and type of resection.

4.1.6 Eurolung Risk scores

As an active member of the European Society of Thoracic Surgeons (ESTS), I have always participated in the initiative of the Society for more than 10 years. One of the most important is the European International Database for which I have contributed cases

since my work in Italy. The first version of the European Society of Thoracic Surgeons (ESTS) Database was created in 2001 as a standalone computer database. Following this, several units across Europe joined the project by applying via a web page linked to the ESTS web site and received a code enabling them to download and install the database. Each thoracic surgical unit could then export encrypted data, automatically attached to an email, and submitted to a central database. There was no fixed harvesting period, and units could submit their data any time they wished, providing more than 95% of fields were complete and valid. Since its inception, the online ESTS database continues to be a completely free database for all ESTS members. The main objective of the ESTS database is monitoring quality of care with the ultimate purpose of standardizing and improving the outcome of general thoracic surgery across Europe. To this purpose, several risk-adjusted models and composite performance scores have been produced to be used as instruments of clinical audit. Furthermore, from the ESTS database, the Composite Performance Score (CPS) for lung surgery has been developed by developing standardized outcome and process indicators covering all temporal domains (preoperative, intraoperative, and postoperative) of the index operation—lung cancer surgery(50). All the selected process indicators were evidence based according to existing guidelines. The ASOS risk scores were used to identify the two outcome indicators (risk-adjusted morbidity and mortality)(47, 49). However, these outcome models were developed from a sample of a few thousands of lung resection patients. During the following years, the ESTS Database has continued to grow and it seemed appropriate to update the old risk models with more reliable models to increase their representativeness and generalizability. Therefore, the ESTS Database Committee recently published new cardiopulmonary and mortality models called *Eurolung1* and *Eurolung2*, respectively and based on 47,960 anatomic lung resections registered in the

ESTS database from July 2007 to August 2015. The Eurolung1 and Eurolung2 models were used to predict risk-adjusted cardiopulmonary morbidity and 30-day mortality rates in each centre(48). These models have been recently developed by the ESTS Database Committee and have shown to be very accurate in the original population.

Another important characteristic of the unit participating to this ESTS database is to have the possibility-through an official application- to use data from the database for ESTS projects. I applied in 2014 for a project to generate a prolonged air leak risk score, which I have presented to the American Association of Thoracic Surgeons (AATS) as an oral paper and published in the Journal of Thoracic and Cardio-vascular Surgery (163). This research experience, has given me the opportunity to learn the importance and the pitfalls of "big data" and the chance to look in detail at the work done in our field from this database.

For this reason, I particularly looked at the area of research that I have developed in lung cancer surgery (PROMS and preoperative assessment) and I found the development process of the two new risk scores really interesting.

Given my interest in PROMS and preoperative assessment, I evaluated the ESTS Eurolung risk models for morbidity and mortality as instruments of internal audit of performance in a small number of thoracic surgery centres. The rationale was to provide an example of applicability of these risk models for institutional quality monitoring initiatives but also to test them in terms of inter-centre comparison. Through the ESTS I have had the chance to ask the involvement of two units which I have visited in the past as visiting physician for limited time: The Department of Surgery, St. Joseph's Healthcare, McMaster University, Hamilton, Canada (Prof Shargall) and the Department of Thoracic Surgery, University Hospitals Leuven, Leuven, Belgium (Dr Decaluwe).

I collected and analysed data from more than 2000 patients (2014–2016) from these centres who underwent anatomical lung resections. The Eurolung1 and Eurolung2 models were used to predict risk-adjusted cardiopulmonary morbidity and 30-day mortality rates. Observed and risk-adjusted outcomes were compared within each centre and I initially presented the results of this analysis at the ESTS Annual Conference in Copenhagen and published the full paper in the European Journal of Thoracic and Cardiovascular Surgery (53).

During this project however, I focused the attention on the lack of PROMS either in the development or subsequent phases of these preoperative thoracic surgery risk scores. Several studies have already shown that objective parameters, traditionally used to stratify the surgical risk (age, FEV1, DLCO, VO2max etc.) are not in fact associated with patient reported residual QOL (14, 62, 158). This information, along with the patient's perspective on decision to proceed to surgery depending mainly on the estimated risk of fixed long-term outcomes, such as permanent debility, oxygen dependency, limitations in activities of daily living(68, 157), has put the fundamentals for these exploratory analysis in our database.

4.2 Role and original contribution

I have had a key role in the planning, development and implementation of this service evaluation project. I wrote the protocol and chose the quality of life questionnaire. Specifically, I discussed the project and involved all the key worker in the care of lung cancer patients in the Thoracic Unit: the Clinical Nurse Specialist, consultants and all the outpatient's clinic staff. I also supervised two ESREP (Extended Student-led Research or Evaluation Project) 4th Year Medical Students involved in the data collection.

I managed the recruitment and supervised the follow-up of patients, with support from other members of the clinical team. I analysed the data.

4.3 Hypothesis

Current preoperative functional algorithms don't include PROMS. We previously demonstrated that QOL may be an objective parameter in predicting postoperative complications. Therefore, we hypothesized that QOL can predict other important clinical outcomes. The aim is to demonstrate the importance of inserting PROMS in preoperative risk stratification process through the scores, to become more accurate and practical by introducing factors not otherwise captured by traditional parameters.

Furthermore, to explore the role of QOL in the preoperative selections of surgical candidates, we hypothesize that the classes of patients identified by the Eurolung risk scores will identify patients with different QOL.

4.4 Aims

The objectives of this analysis are therefore to assess whether QOL has a role in predicting postoperative outcomes like morbidity with the completed dataset of surgical patients and also whether the most up-to-date risk scores were correlated to preoperative quality of life scores of patients' candidates to pulmonary lung resection.

4.5 Methods

Both analyses were done on the completed dataset. The complete methodology has been described in Ch 2.

The first analysis has been presented as oral presentation at the SCTS Annual Meeting in London from 10 to 12 March 2019. The VATS-only analysis has been accepted as a Poster at the upcoming ESTS Meeting in Dublin (June 2019).

4.5.1 Ethical consideration

This study took place within the Lung Oncology Service at St James's University Hospital, Leeds. The Leeds Teaching Hospitals Trust Research & Innovation department approved the project as service evaluation and approval from the local research ethics committee was not required. However procedures were undertaken in line with the DPA (Data Protection Act-(164)) and GCP guidelines (165).

4.5.2 Study design

4.5.2.1 Patient sample and eligibility

This is a retrospective analysis performed on a prospectively maintained database. Three hundred and thirty consecutive patients undergoing anatomical pulmonary resection for suspected or confirmed lung cancer (April 2014-September 2016) completed a preoperative QOL questionnaire which were analysed. All cancer patients were discussed at multidisciplinary tumour board meetings. Operability exclusion criteria were in accordance with current guidelines(26). Exploratory analysis on association of QOL and postoperative clinical outcomes were undertaken. All the recruitment processes and the QOL assessment have been described in Ch2.

4.5.2.2 Analysis

The clinical endpoint to be correlated to QOL was the incidence of postoperative events (length of stay and mortality) occurring within 30 days from the operation or over a longer period if the patient was still in hospital. Cardiopulmonary complications were

defined according to the joint STS-ESTS standard definitions and included the following: respiratory failure requiring mechanical ventilation longer than 24 hours or re-intubation, Acute Respiratory Distress Syndrome (ARDS), pulmonary embolism, pulmonary oedema, pneumonia, atelectasis requiring bronchoscopy, atrial fibrillation needing medical treatment or cardioversion, acute myocardial ischemia, acute cardiac failure, stroke, acute kidney failure(166). Several baseline and surgical variables, including the EORTC QLQC30 scales were tested for a possible association using univariable analysis. Normal distribution of the variables was tested by using the Shapiro Wilk normality test. Numeric variables with normal distribution were compared by using the unpaired Student's t test, while those without normal distribution were compared by the Mann Whitney test. Categorical variables were compared using the Chi square or the exact Fisher's tests as appropriate. I have run the statistics for this project and I have used STATA 14.

In addition to the EORTC quality of life scales, the following variable were tested: age, sex, body mass index (BMI), forced expiratory volume in one second expressed as percentage of normal for age sex and height (FEV1%), carbon monoxide lung diffusion capacity expressed as percentage of normal for age sex and height, history of coronary artery disease (CAD), cerebrovascular disease (CVD), diabetes or chronic kidney disease (CKD), Eastern Cooperative Oncology Group performance score (ECOG). Those variables resulting significant ($p<0.05$) at univariable analysis were then used as independent predictors in a stepwise logistic regression with backward elimination. Variables with $p<0.05$ were retained in the final model and their reliability tested by bootstrap analysis with 1000 samples. In the bootstrap analysis, repeated samples with the same number of subjects as the original database were generated with replacement and the logistic

regression repeated in each of these simulated samples. The variables occurring in more than 50% of the samples were judged stable and retained in the final model(167, 168).

As there is no clear evidence in terms of the effect on QOL of the surgical approach (minimally-invasive and open) on quality of life for lung cancer patients (29, 148, 149, 169), we have decided to perform a sub analysis including only VATS operations.

4.5.3 Results

30-day mortality occurred in 11 patients (3.3%). 90-day mortality occurred in 14 patients (4.2%). Cardiopulmonary complications within 30 days or during hospitalization occurred in 75 patients (23%). Median postoperative hospital stay was 5 days (IQR 3-7). Prolonged postoperative stay (POS) was defined as a stay longer than 7 days (upper quartile). 72 (22%) patients remained in hospital for more than 7 days after surgery.

The following Table 4-1 shows the results of the comparison of baseline and surgical characteristics between patients with and without a prolonged stay.

Table 4-1 Characteristics of patients included in the study

	No POS (n=258)	POS (n=72)	p-value
Age (mean, SD)	67.7 (9.4)	69.7 (10.2)	0.12
Gender male (n,%)	118 (46%)	39 (54%)	0.61
BMI (mean, SD)	27.4 (5.0)	26.0 (5.2)	0.067
FEV1% (mean, SD)	90.0 (21.5)	81.1 (25.2)	0.003
DLCO% (mean, SD)	74.2 (18.8)	64.7 (16.9)	<0.001
CAD (n,%)	17 (6.6%)	6 (8.3%)	0.20
CVD (n,%)	8 (3.1%)	9 (13%)	0.004
Diabetes (n,%)	24 (9.3%)	13 (18%)	0.037
Open surgery (n,%)	55 (21%)	21 (29%)	0.16
Pneumonectomy (n,%)	22 (8.5%)	8 (11%)	0.49

POS=prolonged hospital stay

The following table (Table 4-2) shows the results of the comparison of preoperative EORTC QOL scales between patients with and without a prolonged stay.

Table 4-2 EORTC scores in patients with and without POS

	No POS (n=258)	POS (n=72)	p-value
GHS	69.0 (21.4)	58.4 (28.6)	0.007
PF	85.5 (17.0)	78.0 (21.6)	0.005
RF	85.0 (24.8)	74.3 (31.1)	0.003
EF	74.3 (25.5)	76.9 (29.9)	0.20
CF	87.0 (20.0)	83.8 (23.9)	0.41
SF	87.9 (23.5)	83.8 (30.1)	0.63
FA	20.7 (22.3)	30.7 (31.6)	0.027
NV	3.0 (11.0)	3.9 (17.4)	0.96
PA	13.8 (24.6)	21.8 (31.2)	0.040
DY	25.6 (27.3)	27.8 (30.1)	0.50
SL	30.6 (33.1)	27.8 (37.1)	0.34
AP	13.6 (25.7)	20.4 (33.4)	0.13
Co	8.1 (20.5)	10.2 (26.0)	0.71
Di	5.3 (17.0)	5.1 (16.5)	0.79
Fi	8.9 (23.4)	10.2 (27.8)	0.98

Results are expressed as means and standard deviations. GH: global health scale; PF: physical functioning, RF: role functioning; CF: cognitive functioning; EF: emotional functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnea; SL: insomnia; AP: appetite loss; Co; constipation; Di: diarrhoea; Fi: financial impact

GH, PF, RF, FA and PA scales were all significantly different between the two groups and could potentially be entered in the regression equation. However, all these variables were highly correlated between each other (correlation coefficient greater than 0.5). The variable with the lowest p value (RF) was selected to avoid problems of multicollinearity.

This variable along with BMI, FEV1, DLCO CVD and Diabetes was entered in the regression analysis. The following table (Table 4-3) shows the results of the stepwise logistic regression analysis. Only the variables retained in the final model are shown.

Table 4-3 Regression analysis results

	Coefficient	SE	P value
FEV1	-0.015	0.007	0.022
DLCO	-0.02	0.009	0.011
CVD	1.68	0.55	0.002
RF	-0.12	0.005	0.010
Intercept	2.45	0.80	

RF remained significantly associated with prolonged hospital stay after adjusting the analysis for other potential confounders.

The RF lower interquartile in this population was 67. We categorized the patients in two groups according to this value. 100 patients had RF lower than 67. 31 of them experienced a prolonged hospital stay 31/100=30% (vs. 18% of those with a higher value, p=0.008). This represents 43% of all patients with a prolonged hospital stay (31/72=43%).

4.5.3.1 VATS only analysis

We have limited this sub-analysis to patients operated with a VATS approach: 250 consecutive patients submitted to VATS lobectomies (N=233) or segmentectomies (N=17) over the same period.

30-day cardiopulmonary morbidity and mortality rates were 22% and 2.4%. Median length of stay was 4 days (IQR 3-7). 51 (20%) patients remained in hospital longer than 7 days after surgery (upper quartile). General health (GH) (p=0.019), Physical Function (PF)(p=0.015) and Role Functioning (RF)(p=0.016) scales were all significantly worse in patients with prolonged stay and highly correlated between each other. Physical Functioning had the lowest p-value at univariate analysis and was selected for logistic regression analysis to avoid problems of multicollinearity. Logistic regression showed that PF was an independent factor (p=0.018) significantly associated with prolonged stay along with low DLCO (p=0.008), history of stroke (p=0.005) and low FEV1 (p=0.07). 45 patients had PF<73 (lower quartile value). 31/100=30% of them experienced prolonged hospital stay (vs. 18% of those with higher PF, p=0.042).

4.6 Exploratory analysis on preoperative QOL and Eurolung Risk classes

4.6.1 Analysis

For this analysis we have used all the population of the project (330 patients). The Eurolung2 aggregate score was calculated for each patient. Eurolung2 is the risk adjusting model for 30-day mortality developed from the ESTS database. The aggregate score is calculated assigning weighted points to the following variables: ppoFEV1<70%

1 point, history of coronary artery disease 1 point, extended resections 1 point, age>65 years 2 points, previous stroke 2 points, male sex 3 points, thoracotomy (as opposed to VATS) 3 points, pneumonectomy 3 points, BMI<18.5 kg/m^2 3 points.

The points were summed for each patient and patients were grouped in 3 classes of risk according to their scores. The classes were created based on the incremental risk of mortality and to ensure balanced numerosity of the samples. The values of preoperative QOL scales were assessed within each Eurolung class to verify whether there was any interclass difference. ANOVA test was used to assess differences of QOL scales between classes of aggregate Eurolung. Post-hoc Tukey test was performed to assess he individual intergroup differences. 30-day mortality occurred in 12 patients (3.6%) Cardiopulmonary complications within 30 days or during hospitalization occurred in 75 patients (23%). Patients were divided in to 3 categories of aggregate Eurolung2 with incremental risk of 30-day mortality: class 1 (score 0-3) 165 patients, mortality rate 0%, class 2 (score 4-6), 95 patients, mortality rate 4.2%, class 3 (score>6) 70 patients, mortality 11.4%. Differences of EORTC QOL scales were tested between classes of aggregate Eurolung using ANOVA test (Figure 4-1).

Figure 4-1 QOL p value difference among the 3 Eurolung classes

QoL scale tests
GHS p=0.08
PF p=0.12
RF p=0.008 (class 1 vs. 3 p=0.020, class 2 vs. 3 p=0.009)
EF p=0.005 (class 1 vs 2 p=0.004)
CF p=0.91
SF p=0.09
FA p=0.026 (class 1 vs 3 p=0.031)
NV p=0.95
PA p=0.45
DY p=0.0006 (class 1 vs 3 p<0.0001, class 2 vs. 3 p=0.010)
SL p=0.029 (class 1 vs. 2 p=0.045)
AP p=0.015 (class 1 vs 3 p=0.033, class 2 vs 3 p=0.018)
Co p=0.95
Di p=0.30
Fi p=0.37

4.7 Discussion

It is clear from these data that quality of life assessed preoperatively is associated with the postoperative clinical outcomes of patients. The main results of this service evaluation are that the preoperative patient-reported Role Functioning was associated with prolonged postoperative hospital stay. Baseline QOL status should be taken into consideration to implement psycho-social supportive programmes in the context of

enhanced recovery after surgery. In fact, it can be explained that Role Functioning is capturing some aspect of the patients' life not otherwise included in the current objective risk scores.

When we excluded the 88 patients operated by open approach the results were slightly different but the same domains remained associated with prolonged hospital stay. A poor-level of preoperative patient-reported Physical Functioning was associated with prolonged postoperative hospital stay. Baseline QOL status should be taken into consideration to set appropriate patient expectations during preoperative counselling and to implement perioperative supportive programmes in the context of enhanced recovery after surgery. In this sub-analysis the PF has been selected as the one with the lowest p value, but Role Functioning was the second most significant.

The Role Functioning and Physical Functioning showed a significant association with a prolonged hospital stay. This information is particularly useful during pre-operative discussion with patients, when the doctors and nurses are setting the expectations of their immediate postoperative period. In this era of fast-track policies, where there is a focus on reducing the length of stay after the surgical treatment, it may be useful to tailor this expectation to the individual.

The Role Functioning in particular, is a core construct of QOL which comprises aspects of occupational and social roles relevant for patients in all treatment phases as well as for survivors. This QOL domain may have included social aspects that are affecting the hospital stay regardless of ERAS programme or clinical conditions. Predicting them may be useful to plan and tailor social and psychological interventions to help patient recovery and reduce unnecessary hospital cost.

The Eurolung/QOL exploratory analysis confirmed that the QOL domains associated with the prediction of postoperative mortality are less related to the objective clinical

parameters. Role Functioning and Emotional Functioning were found significantly different in those classes identified by the Eurolung 2. Patients considered for their clinical characteristics at higher risk of death after the operation by the Eurolung risk scores, have a Role and Emotional function lower than those with less probability of death following surgery. Surprisingly Physical Functioning and General Health Status were not different in those Eurolung classes.

This information confirms that some aspects of the patient-reported quality of life can capture aspects of patient's life not easily accessible with the data we have from the patient's past medical history or objective parameters e.g. pulmonary function tests.

The routine collection of these aspects of the patients in clinical practice may help define a new risk score which will be more tailored in defining high risk patients.

4.8 Limitations

Prolonged hospital stay is defined as those falling in the upper quartile in our cohort of patients. As this value may vary in other populations and generalisability would need to be tested with external validation.

We have included all the anatomical lung resections in our analysis. VATS-only has been further limited to lobectomies and segmentectomies. We cannot rule out that the inclusion of other non-anatomical resections would have shown different effect on the prediction of clinical outcomes.

The clinical impact of this analysis may have been limited by the preoperative-only assessment. It will be important with the prospective study to explore the effect of the treatment on the changes of QOL.

4.9 Conclusions

This chapter has provided important information to the role of patient-reported quality of life in the context of preoperative selection of patients for NSCLC surgical resection. In a cohort of 330 patients submitted to lung resection in a single centre, there appears to be a role of preoperative QOL in predicting an important postoperative outcome e.g. the length of inpatient stay.

We have also demonstrated that when the analysis is limited to VATS resection the results showed that the Role Functioning remained associated with the length of stay, pointing the interest to the ERAS programme which are tailored around the minimally invasive surgery. If patients report low scores in the Role Functioning preoperatively, they stay in hospital longer regardless of the clinical outcomes. In practice, we are faced with social and family problems around those patients, who are directly asking to stay in hospital longer as they do not feel “safe” at home or because they are not well supported by carers. The identification of those patients at risk in the preoperative period, with the quality of life may help in planning social intervention at home or with volunteers for those patients reducing their hospitals stay and ultimately reduce healthcare cost.

Furthermore, this data confirms the superiority of QOL compared to some other objective parameters in predicting most commonly used clinical outcomes.

The latter aspect has been also investigated by our second analysis investigating the preoperative quality of life in those patients considered at high risk of mortality by the Eurolung Scores.

The Role and Emotional functioning showed significantly lower values in patients at higher risk of surgical mortality. The other more physical parameters, which we were

expecting to be different did not show significant changes. General health, which is a comprehensive picture of the patient's health status, was not different in patients classified at high risk of death after the same lung cancer operation. We also confirmed the relative importance of the environment and the effect of the health conditions at work and social life. This concept in fact, captured by the role functioning is a novel concept in thoracic surgery and difficult to capture with objective data. However, in the context of health outcome research, it has been defined that the focus of this domain needs to be included (170, 171). The collection of the role functioning data is also complicated by the wide definition of roles and by fluctuations in role participation across the patient's life (172).

This data has highlighted the importance of looking at specific domains in the quality of life questionnaires and to include the length of stay in our future analysis as correlated to the preoperative quality of life.

Chapter 5 EORTC Summary Score– a study to validate a unique quality of life score in surgically treated early-stage NSCLC patients

5.1 Background and overview of the chapter

QOL is increasingly used as an end point in clinical trials. In diseases with a poor prognosis such as metastatic cancer, quality of life may be of major concern. However, clinicians are still reluctant to accept quality of life as an end point equivalent to more "objective" end points such as size of the tumour, complications or disease-free survival in patients with cancer.

Chapter 4 showed some preliminary data on how the patients reported quality of life may be correlated to some important preoperative and postoperative outcomes. Chapter 3 showed the lack of consistent evidence of quality of life collection for patients with early stage non-small cell lung cancer. The work described in this Chapter aimed to further explore the potential role of quality of life collection to improve patient experience throughout the surgical treatment period and potentially during radiotherapy treatment. Quality of life in our speciality is still not an outcome routinely collected as I demonstrated in the survey that I published in 2015 where only a minority of the European Surgeons declared to assess QOL(81). The global domain of QOL is a concept of particular interest especially from the patient's point of view as this is the "quality of life", which they are often referring to during the surgical consultation. However, it must explore all the components of the patient's life, correctly weighted.

Our hypothesis came also from the difficulty that we have noticed in some elderly patients answering to the questions related to their generic health status.

In particular, the exploratory analysis of a newly recommended summary score, if confirming our hypothesis, will be considered as a more sensitive QOL score and better suitable predictor in our specific clinical practice.

Barriers to the acceptance of quality of life routine data collection in surgery may include difficulties in both the understanding of the underlying concepts as well as in the interpretation of the results. In my experience I also found that a difficulty in finding a single score including all the different aspect of a patient's quality of life, may have not simplified its adoption in clinical practice.

Surgical multi-institutional databases have been advocated to improve NSCLC care across countries and International Societies have agreed to standardize definitions and variables (166). Nevertheless, patient-reported outcomes (PROS) are still missing from these databases, leaving this evidence to small institutional studies. One reason for this paucity of data about PROs is the lack of recommendation on the instruments to use or the presence of multiple scores generated by each survey, each associated with several and different dimensions of QOL. It appears clear that in the era of granular multi-institutional databases it is desirable to use a single score which ideally indicates the aggregate status of Health-Related-Quality-of-Life.

Among the community of thoracic surgeons, the adoption of the European Society of Thoracic Surgeons (ESTS) database has also standardized the data collection for assessment by risk-adjusted outcome and/or process indicators, which allows the comparison of institutional performance against European benchmarks. The ESTS Database was founded in 2001 by the ESTS database Committee with the aim to develop risk-adjusted instruments for assessing the performance of thoracic surgery units across

Europe (47, 49). As I have discussed in the previous chapters, the updated ESTS morbidity and mortality models have been used to define risk-adjust outcome indicators for auditing quality of care and to counsel patients about their surgical risk (48). However, in the ESTS database there is no mention of any patient-reported outcomes. This has also limited the possibility of including them into risk scores analysis or considering the patient's voice as quality indices of performance.

5.1.1 EORTC Summary Score (SumSc)

Recently, the EORTC Quality of Life Group recommended the use of the QLQ-C30 summary score to supplement the 15-outcome profile generated by the QLQ-C30. The scoring algorithm for generating the QLQ-C30 summary score is available via the group's web site, http://groups.eortc.be/qol.

Hinz et al. and Nordin et al. were the first to investigate the QLQ-C30 summary score. Hinz et al. (173) used a total score derived from summing up all 30 items of the questionnaire and two separate summary scores based on the sum of all items of the functioning domains and of the symptom domains, respectively. In 2001, Nordin et al. (174) investigated the known-groups validity of the two-item global quality of life scale and three alternative scoring algorithms for the QLQ-C30 based on (1) the 15 QLQ-C30 scale means; (2) the sum of all individual QLQ-C30 items (except for the item on financial problems); and (3) the sum of the scales assessing physical function, emotional function, quality of life, fatigue, nausea/vomiting, pain, appetite, and diarrhoea. For all proposed summary measures, change was categorised in one way or the other into improved, unchanged, and worse. The three alternative scoring approaches performed considerably better than the original, two-item quality of life scale. Overall, this study documented that the QLQ- C30 global quality of life scale may not be particularly well

suited for detecting changes between patient groups and/or changes over time, which is what we originally considered from our previous work in the thoracic patients.

More recently, Gundy et al. (175) used structural equation modelling to test seven alternative higher order measurement models for the QLQ-C30. All the models exhibited a moderate-to-good model-data fit. The model that showed the best statistical fit was a two-factor model of physical and mental health. This is conceptually similar to the SF-36 Health Survey component scores, and the factor structure of the PROMIS domain mapping project(87, 176). All the models tested in the paper which we have used as reference in fact, Financial Difficulties score is omitted. This is in line with the approach followed by Gundy et al. In the final accepted model of SumSc, prior to calculating the mean, the symptom scales need to be reversed to obtain a uniform direction of all scales. The summary score should only be calculated if all of the required 13 scale scores are available (174).

The EORTC QLQ C-30 SumSc has been tested in a large existing dataset for validity and responsiveness to change over time. The EORTC Quality of Life Group steering committee tested the SumSc comparing it to the individual QLQ-C30 scales using pre-treatment QLQ-C30 data and conducting a confirmatory factor analysis. They also tested the seven HRQOL higher order measurement models proposed by Gundy et al(125, 175). The use of a summary score describing sensibly the HRQOL in clinical practice, will facilitate the routine collection of PROs and consequently their adoption in applied health research. Preliminary data of a joint Society of Thoracic Surgeons (STS) and ESTS lung cancer surveillance survey, have shown that HRQOL collection is still underrepresented across the two continents five years after surgery (177). The EORTC SumSc is calculated from the mean of 13 of the 15 QLQ-C30 scores (the Global Quality of Life score and the Financial Impact score are excluded).

I was involved in this project along with my primary supervisor Prof Velikova (who was in the group of the researchers who firstly developed the SumSc). I was collaborating for years with Prof Koller from the University of Regensburg, Germany with whom I have worked in the project of updating the EORTC Lung Cancer Specific questionnaire(89, 154). His knowledge in lung cancer quality of life data and analysis has given invaluable insights into this project and to my interpretation of the results.

The results of this analysis have been presented at the 16th Annual British Thoracic Oncology Society (BTOG) Conference 2018, in Dublin as Poster and then published as an original article in Lung Cancer Journal(178).

5.2 Hypothesis

The hypothesis is that the EORTC SumSc is more sensitive than the EORTC standard General Health (GH) scale in reporting QOL before and after surgical resection for NSCLC. If the hypothesis is confirmed by our results, it will present a unique QOL index for clinical practice, to better inform patients about postoperative outcomes.

5.3 Aims

This chapter aims to compare the performance of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core30 (EORTC QLQ-C30) Summary Score with the traditional General Health (GH) scale preoperatively and explore its sensitivity to detect post-operative change.

5.3.1 Role and original contribution

As this project involves the same patients as the previous chapter, my role and contribution are very similar. I analysed the data and wrote the abstract submitted and

accepted as a poster presentation at the BTOG Conference in Dublin (January 2018). The manuscript has been published in the Lung Cancer Journal(178).

5.4 Methods

A prospective study was undertaken, collecting data from a sample of 326 patients who had undergone anatomic pulmonary resection for the treatment of lung cancer at the Leeds Cancer Centre, UK from April 2014 to September 2016 and had completed preoperative Health-Related Quality of Life assessment. We have a difference of 4 patients compared to the previous chapter as those questionnaires were found in a separate office only after this analysis. 66 patients completed the questionnaire at three months after treatment.

This number represents 47% of all patients operated on during the same period in our unit. The small proportion of patients filling the QOL data postoperatively is due to fact that only people living in the local Leeds metropolitan area received the follow-ups at the Leeds Cancer Centre whereas other patients were followed up in other satellite hospitals in the region. No statistical difference was noted in most of the clinical characteristics of patients completing the postoperative survey and those who didn't (Table 5-1), except for a higher proportion of female patients and patients operated through thoracotomy.

Table 5-1 Characteristics of patients with and without postoperative QOL completion.

Variable	Pts with pre-op QoL only assessment (n=326)	Pts with both pre and post op QoL assessments (N=66)	P value
AGE (mean, SD)	68.3 (40.9)	67.5 (28.2)	0.67
SEX (males, n, %)	134 (51.5)	20 (30.3)	0.002
BMI (mean, SD)	26.9 (4.9)	27.3 (5.5)	0.60
DLCO% (mean, SD)	72.7 (18.8)	70.6 (18.1)	0.44
FEV1% (mean, SD)	87.9 (21.6)	90.9 (23.7)	0.32
CAD (n,%)	21 (8)	1 (1.5)	0.058
CVD (n,%)	13 (5)	3 (4.5)	1
Pneumonectomies (n,%)	23 (8.8)	5 (8.1)	1
Thoracotomies (vs. VATS, n,%)	56 (21.5)	5 (7.5)	0.008

Patients were selected for operation according to current functional guidelines and after discussion at multidisciplinary tumour board meeting (26).

All patients were operated on by Board certified thoracic surgeons either through a muscle-sparing thoracotomy (n=61) or video-assisted thoracoscopic approach (VATS; n=265) depending upon the surgical indications (stage, size and location of the tumour). All patients had a systematic mediastinal lymph node dissection along with the lung cancer resection.

Postoperative care followed standardised pathways of care and included as early as possible mobilisation and oral food intake, intense chest physiotherapy and

rehabilitation, deep venous thrombosis prophylaxis and chest pain control using a combination of patient-controlled analgesia and paravertebral infusion of local anaesthetic.

5.4.1 Ethical Considerations

This study took place at St James's University Hospital, Leeds. The Leeds Teaching Hospitals Trust Research & Innovation department approved the project as service evaluation and therefore approval from the local research ethics committee was not required. However procedures were undertaken in line with the DPA (Data Protection Act) (164) and GCP guidelines (165).

QOL was assessed by administering the questionnaire in a clinic environment for self-completion. The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), Version 3, around 2 weeks before the operation and at 3 months after the operation during our Survivorship Clinic.

5.4.2 Analysis

5.4.2.1 Descriptive analysis

Quality of life scores were calculated according to the scoring manual [10]. The summary score was calculated according to Giesinger et al. [8], which led to a single score ranging from 0 (worst) to 100 (best). Descriptive statistics included counts and percentages, medians and inter-quartile ranges. Normality of distribution of numeric variables (including the QOL scores) was assessed by the Shapiro Wilk test. The following baseline and surgical variables were screened for a possible association with HRQOL scores: age, sex, forced expiratory volume in 1 second (FEV1) expressed in percentage of predicted value, Eastern Cooperative Oncology Group performance score (PS), diffusing capacity

of the lung for carbon monoxide (DLCO) expressed in percentage of predicted value. These variables have been associated with an increased risk of postoperative complications and consequently they are the most frequently factors investigated in lung cancer surgical patients (26, 86, 179).

A known-groups comparison [by gender, age (>70), PS>1, FEV1>70% and DLCO>70%] using the Wilcoxon matched pairs rank-sum test was also conducted.

Before-after differences of QoL scales at 3 months were calculated using the Wilcoxon matched pairs rank-sum test. In addition, the importance of the perioperative changes in QoL scales was measured by calculating the effect size (mean change of the variable divided by its baseline standard deviation)(180). Effect sizes of 0.2, 0.5 and 0.8 indicate a small, medium and large difference, respectively. The sign before the effect size indicates the direction of the difference (a positive sign means that the preoperative value is greater than the postoperative one). Between groups calculations used the independent t-test for numeric variables with normal distribution or of the Wilcoxon rank-sum test for those without normal distribution. Categorical variables were compared by using the Chi-square test or the Fisher's exact test (in case of 10 or fewer variables in at least one of the cells). All analyses were exploratory in nature, thus significant p-values ($p < 0.05$) should not be interpreted as confirming a priori hypothesis.

5.5 Results

The characteristics of the 326 patients included in the study are shown in Table 5-1.

Table 5-1 also shows the comparison between the patients with both preoperative and postoperative HRQOL assessments and those who have completed only the

preoperative questionnaire. In the group of patients who completed the postoperative questionnaire, we found 3-fold higher proportion of operations performed through thoracotomy (p=0.008) and a greater proportion of females (p=0.002).

5.5.1.1 Baseline HRQOL and analysis of SumSc

Table 5-2 shows the preoperative values of the individual QoL scores.

Table 5-2 Preoperative individual quality of life values

EORTC QLQC30 scales (median, IQR)	
GH	66.7 (58.3-83.3)
PF	86.7 (73.3-100)
RF	100 (66.7-100)
EF	75 (66.7-91.7)
CF	100 (83.3-100)
SF	100 (83.3-100)
FA	22.2 (0-33.3)
NV	0 (0-0)
PA	0 (0-16.7)
DY	33.3 (0-33.3)
SL	33.3 (0-66.7)
AP	0 (0-33.3)
Co	0 (0-0)
Di	0 (0-0)
Fi	0 (0-0)
SumSc	87.4 (77.2-93.6)

The median preoperative SumSc in the entire population was 87.4 (IQR 77.2-93.6).

One hundred and forty-seven patients (45%) were older than 70 years of age and compared to younger patients they had a higher (better) preoperative SumSc (90.6, IQR 79.9-94.9 vs. 85.3, IQR 74.0-92.5; p=0.003).

Twenty-four patients (7.4%) had a preoperative performance score (PS) greater than 1 (2 or 3). The Eastern Cooperative Oncology Group (ECOG) Performance Score measures how the disease impacts a patient's daily living abilities rating this from 0 (fully active) to 5 (death)(181).

The following tables show the results of the known group comparisons expressed in terms of means and standard deviations. Asterisks indicate statistical significance. A statistical difference was detected in the SumSc of older people compare to the younger counterparts (Table 5-3). The same comparison was not different when QOL was scored by the GH. No difference in all the domains except EF was found comparing male and female patients (Table 5-4). Both SumSc and GH were statistically different when the grouping was defined by the PS (

Table 5-5). SumSc failed to find difference when the groups were defined by their FEV1 values (Table 5-6)(more than 70%). Both general QOL scores detected statistical difference in patients with DLCO less or more than 70%, highlighting the importance of this preoperative parameter often associated to post-operative outcomes (Table 5-7)(182).

Table 5-3 QOL Comparison by Age

Baseline QoL scales	Older than 70Y (n=148)	Younger than 70Y (n=178)	p values
SumSc	85.9 (12.9)	80.8 (16.3)	0.003*
GH	70.6 (17.9)	67.3 (20.9)	0.25
PF	83.5 (16.3)	84.2 (19.1)	0.15
RF	85.8 (22.3)	80.1 (28.3)	0.14
EF	78.8 (22.5)	69.3 (26.0)	0.0005*
SF	89.3 (20.8)	83.8 (25.8)	0.039*
CF	88.7 (18.4)	83.1 (20.3)	0.002*
DY	23.2 (25.1)	27.9 (28.6)	0.17

Table 5-4 QOL Comparison by gender

Baseline QoL scales	Males (n=154)	Females (n=172)	p values
SumSc	82.9 (16.4)	83.2 (13.8)	0.38
GH	69.0 (20.8)	68.5 (18.6)	0.66
PF	84.2 (19.0)	83.6 (16.9)	0.22
RF	81.9 (26.8)	83.4 (25.0)	0.73
EF	76.2 (24.8)	71.3 (24.8)	0.038*
SF	85.4 (24.6)	87.1 (23.2)	0.29
CF	85.0 (21.4)	86.2 (18.0)	0.94
DY	29.4 (30.0)	22.5 (23.9)	0.075

Table 5-5 QOL Comparison by PS

Baseline QoL scales	PS>1 (n=24)	PS 0-1 (n=302)	p values
SumSc	72.7 (19.1)	83.9 (14.4)	0.004*
GH	56.3 (20.6)	69.7 (19.3)	0.001*
PF	64.8 (22.2)	85.4 (16.6)	<0.0001*
RF	62.5 (29.2)	84.3 (24.9)	0.0001*
EF	71.5 (26.7)	73.8 (24.8)	0.71
SF	68.1 (35.4)	87.7 (22.1)	0.0009*
CF	82.6 (18.7)	85.9 (19.7)	0.30
DY	36.1 (31.0)	25.0 (26.7)	0.064

Table 5-6 QOL Comparison by FEV1

Baseline QoL scales	FEV1<70 (n=64)	FEV1≥70 (n=262)	p values
SumSc	81.8 (14.7)	83.4 (14.7)	0.62
GH	63.2 (21.4)	70.1 (19.0)	0.015*
PF	82.0 (19.9)	84.3 (17.4)	0.38
RF	80.7 (26.9)	83.2 (25.6)	0.48
EF	74.1 (27.7)	73.5 (24.2)	0.47
SF	89.1 (21.9)	85.6 (24.3)	0.18
CF	85.2 (21.6)	85.8 (19.1)	0.76
DY	36.5 (27.0)	23.2 (26.6)	0.0001*

Table 5-7 QOL Comparison by DLCO

Baseline QoL scales	DLCO<70 (154)	DLCO≥70 (172)	p values
SumSc	81.4 (15.6)	84.6 (14.5)	0.02*
GH	63.8 (21.2)	73.2 (17.0)	<0.0001*
PF	80.7 (18.8)	86.6 (16.6)	0.001*
RF	79.5 (27.8)	85.5 (23.6)	0.04*
EF	73.5 (25.1)	73.7 (24.8)	0.99
SF	86.1 (23.0)	86.4 (24.6)	0.60
CF	85.0 (20.9)	86.2 (18.5)	0.86
DY	30.1 (28.7)	21.9 (25.1)	0.009*

5.5.1.2 Perioperative Changes Analysis

Sixty-six patients completed the EORTC QLQ C30 questionnaire three months after the operation. Table 5-8 shows the perioperative changes of QoL scales and SumSc in this population expressed as medians and Inter Quartile Ranges. PF showed a large and significant decline at three months. RF and SF showed a significant and medium decline (effect sizes 0.62 and 0.41, respectively).

Table 5-8 Perioperative changes in quality of life (n=66 patients).

EORTC QLQ C30 scale	Preoperative	Postoperative (3 months)	p-value	Standardized difference
GH	66.7 (58.3-83.3)	66.7 (50-83.3)	0.062	0.26
PF	86.7 (73.3-100)	73.3 (60-86.7)	<0.001	0.91
RF	100 (66.7-100)	83.3 (50-100)	<0.001	0.62
EF	75 (58.3-91.7)	83.3 (66.7-100)	0.066	-0.18
CF	83.3 (83.3-100)	83.3 (66.7-100)	0.41	0.07
SF	100 (83.3-100)	83.3 (66.7-100)	0.002	0.41
FA	22.2 (11.1-33.3)	33.3 (11.1-55.6)	0.002	-0.49
NV	0 (0-0)	0 (0-16.7)	0.073	-0.31
PA	0 (0.33.3)	16.7 (0-50)	0.004	-0.45
DY	33.3 (0-33.3)	33.3 (33.3-66.7)	<0.001	-0.88
SL	33.3 (0-66.7)	33.3 (0-66.7)	0.24	0.12
AP	0 (0-33.3)	0 (0-33.3)	0.027	-0.33
Co	0 (0-33.3)	0 (0-33.3)	0.18	-0.19
Di	0 (0-0)	0 (0-0)	0.42	-0.20
Fi	0 (0-0)	0 (0-33.3)	0.027	-0.32
SumSc	85.9 (76.5-92.3)	76.6 (67.3-87.7)	<0.001	0.48

Results are expressed as medians and inter-quartile ranges. GH: global health scale; PF: physical functioning, RF: role functioning; CF: cognitive functioning; EF: emotional functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnea; SL: insomnia; AP: appetite loss; Co; constipation; Di: diarrhea; Fi: financial impact; SS: summary score.

Effect sizes of 0.2, 0.5 and 0.8 indicate a small, medium and large difference, respectively. The sign before the effect size indicates the direction of the difference (a positive sign means that the preoperative value is greater than the postoperative one).

Some of the symptom scales (FA, PA, DY, FI) showed a postoperative significant increase in their values (worse symptoms). In particular, DY had the largest increase (effect size -0.88).

The change in GH was not significant after surgery (effect size 0.26, p=0.062). In contrast, SumSc decreased significantly at three months (effect size 0.48, p<0.001). From a clinical point perspective, we are expecting a decrease in HRQOL especially during the first three months after the operation (183).

5.5.1.3 Subgroups Analysis

Table 5-9 shows the perioperative changes of SumSc in different groups of patients with or without risk factors for surgery.

In particular, medium or large postoperative declines of SumSc were observed in both males and females, in patients with lower FEV1, lower performance score, and in those aged 70 years and above. Finally, a similar postoperative medium decline of SumSc was observed irrespective of the DLCO level.

Table 5-9 Subgroup changes in SumSc

Characteristics	Preoperative SumSc	Postoperative SumSc	P value	Effect size
FEV1<70% (12 patients)	85.3 (81.5-88.8)	71.8 (58.0-81.8)	0.015	1.84
FEV1≥70% (54 patients)	86.0 (76.3-92.9)	78.0 (69.0-88.7)	0.003	0.35
DLCO<70% (35 patients)	84.6 (74.9-92.7)	76.2 (67.9-85.6)	0.006	0.55
DLCO≥70% (31 patients)	87.2 (79.3-92.3)	80.1 (63.8-91.9)	0.007	0.55
Age>70 years (29 patients)	83.0 (76.3-92.7)	76.5 (69.4-85.6)	0.001	0.99
Age≤ 70 years (37 patients)	86.6 (77.4-92.1)	77.4 (66.9-88.7)	0.021	0.28
Males (20 patients)	91.2 (81.6-93.4)	90.9 (74.6-94.8)	0.056	0.50
Females (46 patients)	84.1 (74.9-90.9)	74.0 (66.5-83.5)	0.001	0.51
PS >1 (6 patients)	78.3 (44.3-93.2)	67.2 (51.1-69.4)	0.46	0.26
PS = 0-1 (60 patients)	86.4 (77.5-92.2)	78.0 (69.0-87.9)	<0.001	0.55

Results are expressed as medians and inter-quartiles range (IQR). FEV1: forced expiratory volume in one second; DLCO carbon monoxide lung diffusion capacity; PS: Eastern Cooperative Oncology Group performance score. Effect sizes of 0.2, 0.5 and 0.8 indicate small, medium and large differences,

respectively. All effect sizes were positive indicating that preoperative values were always greater than postoperative ones

5.6 Discussion

The SumSc was able to detect pre-operative difference between groups known to have difference in their clinical characteristics. The SumSc was also more sensitive than GH in identifying postoperative changes in the overall population. It was sensitive also to change also when exploring well-known groups before-after analysis.

SumSc was particularly sensitive in scoring preoperative QOL of older patients (>70 years) compared to the GH. This may be explained by the fact that the SumSc is calculated through multiple scales. In this case, as already demonstrated in literature (66), older patients have a statistically higher EF which may not have been easily expressed by the patients answering to the GH questions.

The fact that there is no difference within the groups with different FEV1% is not completely surprising. In fact several studies (184) have shown that FEV1 did not predict complications in those with FEV1< 70% where lung resection may cause only minimal loss or even the improvement in pulmonary function due to obstructed segments affected by Chronic Obstructive Pulmonary Disease (COPD). The limitation of our study is in the fact that we did not have information about predicted postoperative values of FEV1. The ppoFEV1 has been in fact more sensitive to predict postoperative events in similar populations compared to absolute values. SumSc also distinguished also patients with DLCO> or <70%, SumSc, confirming DLCO to be probably an objective parameter closely associated with QOL(63).

The Cardio-Thoracic speciality has been one of the first ones to introduce and champion the risk-adjusted outcome analysis for monitoring and improving quality of care(185).

However, there is an increasing interest in investigating the possible inclusion of the patient's own perspective in these scores (186). The first step in this field is to identify the most appropriate indicator of the patients' voice: the SumSc represents a valuable candidate for this, reflecting the patients' self-assessment of their daily life after a surgical operation for lung cancer. Overall these early findings provide encouraging evidence to validate the use of the SumSc in clinical practice in the mid-term follow up of patients treated with surgical resection. Confirming previously published data(86), in our cohort of surgical lung cancer patients, HRQOL decreased three months after the operation. This decline varied through the scales with meaningful effects confirmed in the Social, Physical and Role Functioning and in the Dyspnoea symptom score. The General Health score, which reflects the patient's consideration of their quality of life, did not change significantly between pre-and post-treatment.

On the other hand, our results show that the EORTC Summary Score was significantly reduced after surgery.

SumSc also detected important differences between subgroups of patients confirming the existing evidence. The analysis of subgroups of patients considered clinically at high risk for surgery, showed interesting results confirming that objective variables cannot be considered as surrogates of patient-reported quality of life. For example, Gender and FEV1 were factors not associated with different preoperative QoL scores. Patients with lower DLCO, an age older than 70 years or a PS>1 reported lower baseline value of SumSc. However, as demonstrated in the past (62, 158) DLCO was not associated with a greater QoL decline according to the SumSc analysis.

Giesinger and colleagues have already shown that the validity and responsiveness of the EORTC QLQ-C30 SumSc is equal or even superior to the original underlying QLQ-C30 scales scores(125). Most recently, the SumSc has also demonstrated good ability to

detect changes in subjects' quality of life among patients with unresectable hepatocellular carcinoma (187). Our results confirm their findings, showing that the SumSc performed better than the GH scale of QLQ-C30 especially in the changes over time.

Several investigations have described significantly reduction in HRQOL after surgical resection for lung cancer. This decrease is particularly evident in the first postoperative period, and improves, although not completely to preoperative levels, in the following months(183). The SumSc in our series confirmed this trend along with most of the single functioning and symptom scales, while the General Health score didn't.

Regarding the age-related sub analysis, several studies have demonstrated that age is not a major determinant of QOL after lung resection for cancer. Ferguson and colleagues found no difference in the postoperative QOL scores between patients younger or older than 70(158, 179). Burfeind and colleagues used the EORTC QLQ-C30 instrument in a prospective, longitudinal study to assess their lobectomy patients and found no significant difference between older (>70 years) and younger cohorts(65).

The SumSc in our population, showed a large decline (expressed as large ES) after three months from the operations only in patients older than 70 years, probably taking into consideration all the detailed aspects of the older patient's quality of life, which are not easily detected by answering to the generic question of the EORTC QLQ C-30 questionnaire.

Our results show that the gender-specific analysis does not follow the preoperative trend previously described in thoracic oncology surgery, for example Chang et al. reported gender was a significant determinant of the QOL aspects of physical, emotional and cognitive functioning(188).

The SumSc in our series was largely reduced in patient with forced expiratory volume in the first second (FEV1) lower than 70% of predicted, although at the baseline assessment there were no difference between these two groups. In our previous study of 220 patients who completed the SF-36 questionnaire before and after surgery, we selected patients with chronic obstructive pulmonary disease (COPD) to compare their HRQOL with a case-matched population of patients with normal respiratory function. This analysis did not find any differences between the groups in any of the preoperative and postoperative physical and mental QoL scales(14). Ferguson and co-authors, using the EORTC QLQ-C30 and LC-13 module, found that FEV1 was a consistent predictor of physical function, role function, fatigue, pain, and dyspnoea. However, this cross-sectional study has a much longer follow-up of (2.7 year), whilst our analysis was focused only on the very short-term(179).

In the US the first attempt to incorporate PROs in the Society of Thoracic Surgeons database with the National Institutes of Health (NIH) Patient Reported Outcome Measurement Information System (PROMIS) has demonstrated the feasibility of the future integration in lung cancer patients records(150). The American College of Chest Physicians (ACCP) has made the QOL data collection a class 2B recommendation for all lung cancer surgery patients(189). These data provided a reliable background to use the SumSc in our prospective study.

5.6.1 Limitations

This study has some potential limitations.

- First of all, only a small number of postoperative assessments were collected. This was due to logistical reasons with only patients who lived in the immediate surrounding area being followed up at the hospital, where the preoperative assessment took place. We

cannot rule out that the inclusion of all the patients may have affected our results. However, the characteristics of the patients who did not participate in the 3 months post-operative assessment were similar to those included in the full analysis except for a lower proportion of VATS procedures and a predominance of female patients.

- Most of the operations were performed using VATS. The minor incidence of the post-thoracotomy pain especially in the early postoperative period has been recently reported in a large randomized trial comparing the two surgical approaches (148).

We acknowledge that it would be interesting to verify whether similar results would be found analysing a population with a larger proportion of thoracotomies.

- We included only pulmonary lobectomy in our analysis. Future investigations are needed to explore the use of SumSc with different surgical operations for lung cancer such as pneumonectomy or wedge resections.

-Finally, our study is limited to the first three months follow-up after the operation. We have chosen this timeframe as this is when the major changes in HRQOL have been described previously and also to reduce the attrition rate. However, future investigations are warranted to evaluate the SumScore performance at later time points from surgery.

5.6.2 Conclusion

This study showed that the Summary Score of the EORTC QLQ C-30 questionnaire was more sensitive to early-changes in subjects' QOL, than the General Health score and distinguished better know-groups preoperatively. As suggested by the EORTC Quality of life group, the SumSc can avoid problems with potential type I errors that can arise because of multiple testing when making comparisons based on the 15 outcomes generated by this questionnaire. In addition, the use of the QLQ-C30 summary score can

reduce sample size requirements in clinical trials were HRQOL is a primary endpoint. In conclusion, our results confirmed potential use of SumSc in our prospective study described in the next three chapters. These initial findings in our surgical population, suggest that the SumSc can be used as a parsimonious and easy to interpreted patient-reported-outcome measure in multi-institutional databases and future clinical trials.

Chapter 6 Life After Lung Cancer (LILAC) prospective study: protocol and baseline analysis

6.1 Background

6.1.1 Overview

Chapter 3 described in a systematic review the lack of standardized and definitive evidence on PROMS in early stage NSCLC patients treated with radical intent. There are no randomized trials comparing the two available treatments. While there was some evidence of potential in terms of patient outcomes, robust evidence was limited to only one RCT and very few studies assessed the impact of these treatments on patient QOL. However, the use of patient reported outcomes (PROs) in cancer clinical trials is now widespread. The United States (US) Food and Drug Administration (FDA) have set out rigorous guidelines for the development and use of PROs as both primary and secondary trial outcomes(72).

Chapter 4 and 5 described collection of PROMS in the clinical setting and also confirmed the association between patient-reported health status and postoperative clinical outcomes in a cohort of surgical-only patients. This has also highlighted the possibility of inserting these outcomes in the delicate preoperative selection process. It also validated the use of a unique QOL score in a population of lung cancer patients treated with surgical resection. The latter has the potential to overcome the most criticized issue when collecting PROMS: the difficulty of finding a summary score to be used in clinical practice.

To address these gaps, I planned a prospective observational study assessing QOL and other PROMS over the first year after these two treatments.

6.2 Aims

The LILAC study is an observational prospective longitudinal study with repeated measures (PROMS), employing a convenience sample of consecutive patients planned to have VATS resections or SABR for early stage NSCLC. Outcome measures were collected prior to treatment and 1, 3, 6 & 12 months post-treatment, administering the questionnaires by the remote web-based system. Paper administration was also offered to patients without Internet access.

LILAC main aims were:

→ To compare changes before and after treatment of patient reported outcomes (QOL and Patient Satisfaction) after VATS lung resections or SABR in early stage NSCLC patients.

→ To correlate clinical data (baseline and 12 months demographic and clinical characteristics, complications and survival) with PROMS to determine objective predictors of clinical and patient-reported outcomes.

→ To identify specific factors, which have influenced the personal choice between the treatments (Decision Self-Efficacy Scale)

→ To establish recruitment and attrition rates and adherence to PROMS reporting during the study

→ To explore patient choice of electronic vs paper questionnaires

Using a mixed-methodology, I was able to correlate this information to the direct participants voice: staff and patients' feedback interviews were conducted in order to

give in-depth understanding of strengths and limitations of the project and define improved future study design. End-of study feedback questionnaires were also administered to help identifying patient opinions and implementation issues.

This chapter provides a detailed description of the methodology used for the LILAC study and a descriptive analysis of the baseline findings. This descriptive analysis aims to characterise the patient populations and baseline efficacy in decision making. It also describes the psychometric properties of the Decision-self efficacy questionnaire through a factor analysis in the lung cancer population.

The self-reported evolution of QOL during and in the first year following treatment for early-stage lung cancers will be described in Chapter 8. In addition, this chapter aims to evaluate the initial findings regarding the feasibility and acceptability of longitudinal PROMS data collection alongside routine clinical care over a period of one year using both electronic and paper methods. This is evaluated through consideration of recruitment, attrition rates and compliance. We have also investigated the patient opinions regarding taking part in this study. Results will help to overcome possible obstacles in the data collection and support the design of future studies. We have in fact run a set of interim and end-of-study interviews which will be analysed at the end of this chapter.

6.3 The Lilac Study protocol

I was involved in the Grant application for this prospective study awarded by Yorkshire Cancer Research (grant number: L399), in December 2015 to Professor Velikova. I have been responsible for involving our Patient representatives (PPI) in this study. Moreover, I have attended the Leeds Lung Cancer Supportive Group to discuss the participation of

a lung cancer patient who could give important insights not only during the grant application but during the entire study. The patient leaflet and documents have been also sent to patients involved in our Research Advisory Group (RAG).

The LiLAC protocol has been published in the Journal of Thoracic Disease(190).

6.3.1 Study design

Lilac is a prospective observational longitudinal study with repeated measures (PROMS) at baseline, 6weeks, 3,6 and 12 months after radical treatments for early stage NSCLC.

The study cohort comprised male and female consecutive patients undergoing treatments for early stage NSCLC (VATS anatomical lung resection or SABR) in Leeds Teaching Hospital Trust treated between February 2016 and March 2017. A prospective study involves a group of similar individuals (in this case, patients treated with surgery and radiotherapy) and following them up over time. Patients have been invited to self-report PROs measures using online secure access via QTool software at home or in clinic before the treatment and at 6 weeks, 3,6, and 12 months after. Paper administration has been offered to patients without Internet access.

Clinicians with access to the electronic patient medical record (PPM) can consult the Patients Reported Information in real-time during consultations for all the patients who completed the questionnaires. Appropriate training for the use of the QTool has been offered to both clinicians and patients as per previous experience in the group(103).

6.3.2 Treatment Groups

The study population comprised two different groups: patients treated with VATS anatomical lung resections and SABR.

Leeds is a tertiary referral centre in West Yorkshire, with expertise in delivering both minimally invasive thoracic surgery and SABR. A detailed description of the two treatments has been done in Chapter 1.

6.3.2.1 Surgical group

Leeds VATS Programme started in 2006. In 2014 200 curative lung resections were performed for stage I-II NSCLC (70% using VATS approach). 10% of the surgical patients were octogenarians, 50% were older than 70 years, 30% had moderate-to-severe COPD and more than 10% had a history of ischemic heart disease. As an internationally agreed definition of high-risk patient does not exist yet, there is equipoise among doctors treating these patients. In the same period a total of 225 patients have been submitted to anatomical lung resections for primary and secondary lung cancers.

6.3.2.2 SABR Group

The UK SABR Consortium was formed in 2007 to facilitate the safe introduction of the technique using national guidelines. Leeds Cancer Centre has treated over 500 patients, local and national referrals. Leeds was also the leading centre for a feasibility study of SABR versus surgery in patients considered at higher-risk of surgery (SABRTooth). In the same period a total of 189 patients have been treated by SABR in Leeds for primary or recurred NSCLC.

6.3.3 Ethical consideration

I have been responsible for writing the protocol, creating and submitting the HRA application and I have attended the Ethical Committee Meeting for the approval of the study.

This study has received ethical approval from The National Research Ethics Service Yorkshire and the Humber-Leeds East Committee - (REC Ref: 16/YH/0407).

Patients were eligible for the study if they were 18 years or older, able to read and understand English and were not exhibiting overt psychopathology or serious cognitive dysfunction and not participating another QOL study. All participants provided written informed consent.

I submitted two Substantial Amendments during this study which have been granted with no request of modifications. One Amendment (Substantial Amendment 1, 18/04/17) to send a letter with PIS and consent form to the patient before their clinical appointment and the second one (Substantial Amendment 3, 09/05/18) to modify the end-of-study interviews schedules. Details of these amendments will be described in the next chapter.

6.3.4 Patient sample

6.3.4.1 Patient eligibility

The following inclusion criteria were adopted:

- Age 18 years and over.
- Diagnosis of NSCLC either from histology or multi-disciplinary team meeting (MDT/Tumour Board) agreement on >95% likelihood of diagnosis based on radiological evidence or both.
- Decision for either surgery or SABR
- Able to give informed consent.
- Able to understand the language of the questionnaire.

These were the exclusion criteria:

- Advanced disease (III-IV stages)

- Patient included in other QOL study, which may increase patient burden and bias the answers of the questionnaires.

6.3.5 Patient approach and recruitment

Participants were recruited from Leeds Teaching Hospital Trust Lung MDT. Expert NSCLC teams identified potential patients for this study during MDTs with the help of lung cancer clinical nurse specialists (CNS).

6.3.5.1.1 First approach and introduction to the study and consent

Initially, we felt that the best time to approach patients for the study was after the first outpatient clinic with the surgeon or oncologist. At this stage there is a definite medical decision to proceed with active treatment and we can avoid the risk of approaching patients who subsequently do not receive treatment.

Following information provision, patients were given as long as they needed to consider participation to the study. If a patient has preferred more time to consider whether they wished to participate we gave them the opportunity to take the information home and discuss the study again with the researcher at their next visit (usually the booked pre-assessment for surgical patients and planning CT appointment for SABR patients). Patients also had the opportunity to decline participation at this point, consent to participation at this point or to consider being involved and speaking to the researcher at their next hospital visit.

Patients who preferred to fill in the questionnaires electronically received training in using the QTool system.

6.3.5.1.2 Second option to approach and consent the patients:

Patient candidates for surgical resection for NSCLC in Leeds Teaching Hospital Trust benefit from the recently implemented ERAS (Enhanced Recovery after Surgery) programme. However, that meant once they have had the consultation with the Thoracic Surgeon and the Lung Cancer Specialist Nurse, they will have, during the same morning an additional 20-minute conversation with the ERAS staff nurse. Following this, the patient will be sent immediately to the pre-assessment unit, where they will be seen by two or more Health Care Providers (Staff Nurses, Occupational Therapist and Pain Management Nurse). After that the patient is ready for surgery and he/she will return the day of the scheduled procedure. We noticed a slow recruitment rate after six months especially in the surgical group.

Many of the patients eligible to participate in LILAC study, were exhausted on that day of the pre-operative consultations and they declined to speak with an additional member of the research team.

In order to access this cohort of patients in a way that reduces patient burden, we sought approval to implement a new recruitment procedure. This allowed us to invite the patient to participate by letter after a decision has been made for surgery once they have an admission date. A member of the direct clinical team prepared the letter and sent it to the patient. The stamped addressed envelope was returned to the research team in order to check that a signed consent form had been sent back by the patient.

All patients were advised that in addition to the written information sent out by post they could contact the research team via the telephone or email to respond to any

questions or clarify any issues that they may have. The research team was available to assist the patient in signing the consent form or filling the questionnaires in clinic during the entire study period.

However, this second approach was prematurely stopped due to a formal complaint that we received from a patient. She was a candidate for surgical resection, however she did not have biopsy proven cancer. Nonetheless, the MDT considered her radiological and morphological features highly likely to be a cancer. As I received the complaint I promptly addressed the issue. We had a couple of phone calls and she explained to me that she felt distressed by receiving a leaflet about a research study on lung cancer before being operated on. I reassured her that this concern was given serious consideration and consequently we stopped sending the letter to patients without pathological confirmation of cancer.

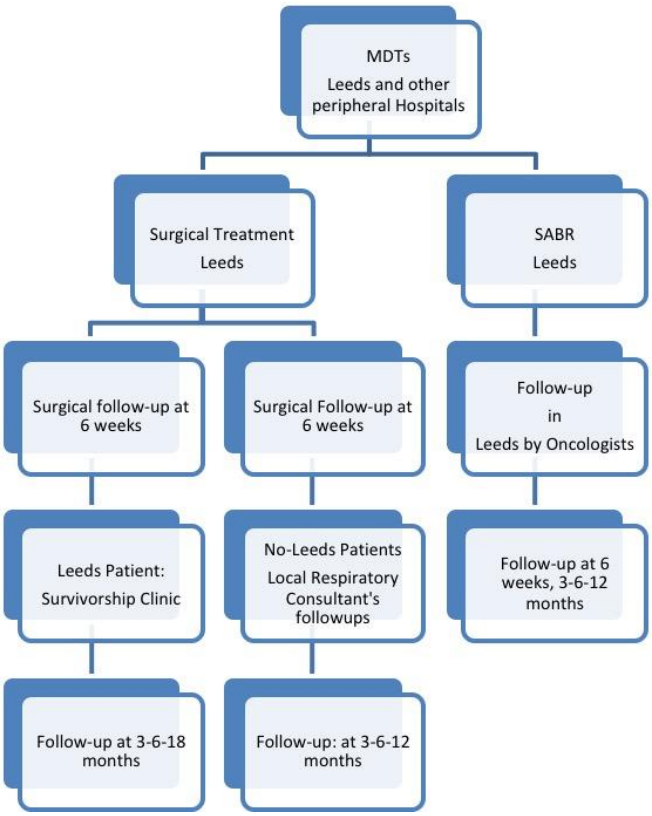
She appreciated my explanation and I invited her to join our research advisory group as I thought that these types of comments were extremely useful for our future research projects. She is now in fact attending all our Research Advisory Group (RAG) meetings and is very helpful in all our projects.

6.3.5.1.3 Follow-ups

Surgical Leeds patients were also asked to have an additional visit to St James Hospital to perform a Pulmonary Function Test at 12 months after their surgical procedures (the usual follow-up visit is at 18 month) as part of the Lilac study. This is not usually considered as a standard NHS practice, although recent guidelines identified this test as useful in following patient up functionally. The results of this exam were reviewed by a member of the Research Team with clinical competencies and if the results exceeded certain threshold the Respiratory Team was immediately contacted, and the patient

referred back to them without waiting for the next scheduled survivorship appointment. Follow up frequency are in line with current NHS practice. Current practice for early stage NSCLC is summarized in Figure 6-1.

Figure 6-1 Current NHS practice for NSCLC in Leeds



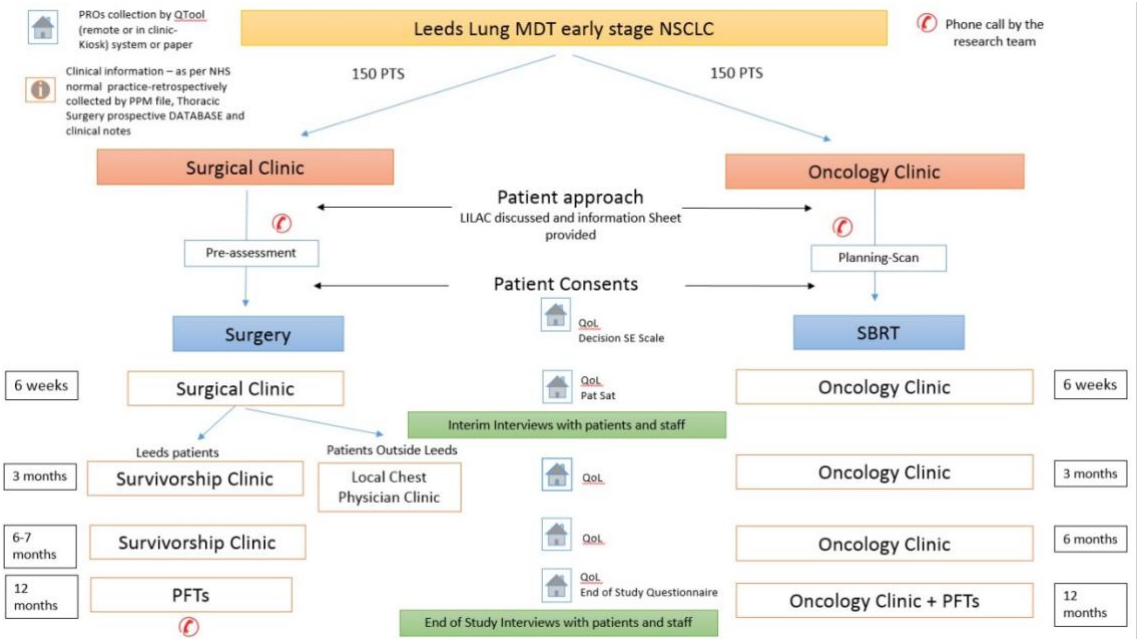
6.3.5.2 Assessment

Baseline QOL questionnaires were completed at time of consent prior to or within 1-15 days from the operation or 0-4 days from starting radiotherapy treatment. At this time patients opted to either complete the questionnaires online, receiving either an email or letter reminders, or to complete paper questionnaires which were posted to them. Patients were invited to complete the questionnaires at different time points. These time points broadly coincided with usual follow up schedules for patients. Specifically,

outcomes measures were collected prior to treatment and 6 weeks,3, 6 & 12 months afterwards. Quality of life questionnaires were collected each time (Baseline, 6W, 3,6,12months), the Decision-self Efficacy scale was administered only at baseline and the Patient Satisfaction questionnaire was collected only at the 6 weeks' time point.

The QOL assessment was scheduled according to the literature and to coincide with potentially the timing of greatest effect of these two treatments (see Figure 6-2 for LiLAC detailed assessments).

Figure 6-2 Overview of participant pathway through Lilac



For the remainder of the study two reminders at 7 and 14 days after the initial invitation were sent out by post or by text message/email according to the patients' preferences. Patients choosing the electronic completion were expected to receive the reminders by email and text. Email and letter reminders were managed using the Data Management System (DMS) set up previously in our group.

Electronic results completed online using QTool, were immediately available for viewing by clinical staff in patient's electronic health records (EHR). Following receipt of the

paper questionnaires, all paper results were also inputted into QTool and the results made available as soon as possible.

Clinal data were inputted immediately after patient recruitment in the DMS by the researcher and updated on a monthly basis.

6.3.5.2.1 Baseline demographic and clinical information

The followings clinical details were collected:

- Personal details and demographics including height, weight, and gender
- Date of diagnosis
- Pre-operative investigations results
- Confirmation of eligibility
- Confirmation of written informed consent
- MDT decision: to be extracted by clinical letter on PPM.
- Internet access
- Comorbidity, ECOG performance status, MRC dyspnoea score, Charlson Co-morbidity Index, Pulmonary Function Tests, NSCLC clinical Stage, Postcodes to calculate the Index of Multiple Deprivation (IMD).

6.3.5.2.2 Treatment and Post-Treatment Clinical Information

- Surgical group: Surgeon, details of previous thoracic operations, whether the outcome was curative, palliative or unresectable in the opinion of the surgeon at the time of operation, ASA status, type of operation, duration of hospital stay, operative and postoperative complications, readmissions, Peak-Flow, Diffusion Capacity of lung for carbon monoxide (DLCO) in Leeds patients only, Details of any local or distant recurrence, Any details of adjuvant treatment and its adverse events.

Complications have been defined according to the standardized definitions proposed by The Society of Thoracic Surgeons and the European Society of Thoracic Surgeons general thoracic surgery databases task forces(166). Complications are graded according to CTCAE version 4.0.

- SABR group: dose planned, dose delivered, dates of delivery, post-SABR complications, unexpected admissions, DLCO, details of any local or distant recurrence toxicity will be evaluated using the CTCAE version 4.0. Data will be extracted from MOSAIQ (radiotherapy delivery system) or the electronic patient record (EPR) as necessary.

6.3.6 Outcome measures

As detailed in the previous chapters, patients completed the validated cancer specific European Organization for Research and Treatment of Cancer quality of life questionnaire (EORTC QLQ-C30)(88) and the disease specific module for the relevant cancer site (Lung Cancer Module LC13)(89) as quality of life tools. For the majority of items, a four-point Likert-type scale is used for the questionnaire and the item responses are converted through a linear transformation for both individual and scaled items onto a 0-100 scale. Higher scores for symptom items reflect a higher level of symptoms and higher scores for the function items reflect a better level of functioning(88). For single symptom items, a score of 0 relates to a patient response of ‘not at all’, a score of 33.3 scores is a linear transformation from the patient response category ‘a little’, as compared to 66.6 corresponding to a response ‘quite a bit’ and 100 ‘very much’. For functional items a score of 100 represents a patient with no impact on their functioning, 66.6 relates to ‘a little’ impact on function, 33.3 ‘quite a bit’ and 0 ‘very much’. Minimal important differences were classified as a change in score of up to 10 points, moderate

differences as a change between 10-20 points and large differences to be change in score were greater than 20 points(138).

Patient's satisfaction was assessed through the administration of the Patient Satisfaction Questionnaire Short Form (PSQ-18)(94), which is a cross-cultural validated survey for use in different settings. The questionnaire is an 18-item self-administered survey including different scales reflecting the perceived level of satisfaction in relation to the care provided by doctors. The team behind this Likert scale questionnaire proposed seven dimensions of patient satisfaction directed toward their doctors. These are general satisfaction, technical quality, interpersonal manner, communication, financial aspects, time spent with doctor, and accessibility and convenience (94). All items have been scored so that higher score represent high degree of satisfaction with care. Scores ranged from 1 to 5. Missing items have been managed as described in the scoring manual (94).

Decision Self-efficacy Scale measures self-confidence or belief in one's abilities in decision-making, including shared decision-making(98). Patients are asked to reflect on how confident they feel in making an informed choice on a scale ranging from "Not at all confident" to "Very confident". Scores range from 0 [not at all confident] to 100 [very confident]. A score of 0 means 'extremely low self-efficacy' and a score of 100 means 'extremely high self-efficacy'.

6.3.6.1 Sample Size

In 2014, we reported as reference value for the sample size calculation, 100 pre-surgical quality of life questionnaires collected in 5 months from outpatients' clinic at Leeds Cancer Centre (LCC). For longitudinal studies involving regular PROMs, we typically see 70% consent rate and 30-35% attrition over 3 months (75). Therefore, our expectations

were to be able to recruit 150 VATS and 150 SABR patients over 12 months with 12 months of follow-up. We determined the sample size of the surgical arm by using our historical cohort of 115 anatomic lung resections (operated in 2014). Their average baseline EORTC QLQ-C30 Global Health Scale value was 65 with a standard deviation of 21.5. Therefore, in order to detect a minimum peri-operative (baseline to 3 months) difference of 6.5 points (10% from baseline), with a two-sided alfa level of 0.05 and a statistical power of 90%, a sample size of 115 patients in the surgical arm was estimated. A similar assumption was used for SABR patients where we do not have any available data.

6.3.7 Feedback Interviews

Patients were interviewed at the end of the 12 months study period to gain more in-depth feedback on their experience of using Lilac. The former interview schedule explored patients' views of the accessibility and acceptability of Lilac, in addition to their general views of using the electronic system and how participating in this study have facilitated the discussion of QOL in clinic.

The initial interview questions and schedule was generated based on the experience of the group with electronic PROMS collection. We were expecting at least half of on-line completers as per the previous audit, however, after the interim interviews and acknowledging a reduced electronic participation, the schedule was modified to better capture the views of Lilac patients. In particular, we removed detailed questions on the use of the QTool website and added others to explore barriers or facilitators to internet access. We also added more specific questions about the discussion of QOL in clinic regardless of the use of Lilac after having noticed less participation of the clinical staff

then predicted. This may have given more information to improve both electronic completion and staff participation in future studies.

Although initially, it was planned to interview 5-10 patients from each completion's type group to ensure data saturation, recruitment of electronic completers was substantially lower than expected. As a result, a higher proportion of patients were recruited from the patients who decided to fill the survey on paper to reflect the overall patient sample. Patients were recruited consecutively until data saturation was reached.

6.4 Analysis

6.4.1 Quantitative analysis

Data was analysed using Stata/SE 15 (StataCorp. 2013. Stata Statistical Software: Release 13.1. College Station, TX: StataCorp LP). PROMS scores over time and clinical data were summarised using descriptive statistics. Analysis of the EORTC QLQ-C30 and the disease specific module and the handling of missing responses within the questionnaire were performed according to EORTC guidelines, using a process of imputing for missing values in scaled responses(88). The reasons for patients leaving the trial will be presented along with an evaluation of the rate of recruitment to the study in the CONSORT flow-chart. Initial analysis explored recruitment sample including comparison between electronic and paper completion.

6.4.1.1 Psychometric properties analysis of DSE questionnaire

During the baseline assessment patients were asked to also complete the Decision self-efficacy questionnaire (DSE). There is paucity of evidence assessing the views of patients with stage I NSCLC on aspects of Shared decision-making (SDM) considered to be of

greatest importance in the decision making process between surgery and SABR(191). However, the majority of patients reported not being offered both treatment options (surgery and SABR), indicating that SDM was not taking place in many consultations. This makes the decision-efficacy the first shared decision element to explore, compare to other like decision conflict (which assumed a discussion and choice offered). Other studies have showed self-efficacy has direct and indirect effects on quality of life in patients with resected lung cancer(192). For these reasons I decided to assess the self-efficacy in this population.

A number of studies have explored the use of a principal component analysis (PCA) as a means to explain and compress the correlated variability of a questionnaire in a specific group of patients.

PCA is a statistical tool for establishing patterns between correlated items within a data set. The process effectively groups together correlated data points, reducing the data to a few parameters which describe the whole data set. This questionnaire is composed by eleven questions which are very different from each other and investigates significantly different aspects of the decision-making. However, it generates one score, so we explored the construct validity of the DSE and specifically to see whether the single questions were grouped in more than one single factor as described in the user manual. PCA quantifies the variability in a clinical datasets and separates out items with similar morphology.

The Decision Self-Efficacy Scale (DSE) measures self-confidence or belief in one's ability to make informed decisions and participate in shared decision making with health professionals(98). There are two versions of the Decision Self-Efficacy Scale, one with 5 response categories and one with 3 response categories. It is a 11-item instrument with a five-point response scale ranging from 0 (not at all confident) to 4 (very confident). An

example question is: *'I understand the information enough to be able to make a choice'.* Internal consistency has been evaluated in women considering hormone replacement therapy (alpha coefficient 0.89). Scores are linearly transformed: score of 0 means 'extremely low self-efficacy' and a score of 100 means 'extremely high self-efficacy'.

A principal component analysis (PCA) was used to examine the DSE structural validity. The minimum recommended sample size to conduct a PCA is 100(193).

An exploratory (principal axis) factor analysis was conducted on the data set of 11 (N= 158 cases) items using IBM SPSS version 24. In addition to the total variance explained, the scree plot, eigenvalues, and component loadings were assessed to verify the factor structure of DSE in this cohort.

As a measure of construct validity, Pearson correlation coefficients were generated to determine relationships between the items. For highly related constructs, moderate to strong associations (r~ +/- 0.40 to 0.80) between these determinants and the factors of the DSE were expected(194).

To explore the clinical applicability of the measure, we hypothesized a difference between the two groups' (surgery and SABR) DSE results. As all the DSE variables were not normally distributed, they were compared across groups by the Mann–Whitney U test.

6.4.2 Qualitative analysis

After 6 months from the first questionnaire, a subset of participants per group were purposively sampled by gender and age and invited for interim-interview. At the end of the study a subset of patients (excluding the one who participated in the interim interviews) were invited for interview.

We thought that with the interim interviews we would have identified any issues raised and allowed us the chance to correct them during the rest of the study. We have also invited staff to take part in the end-of-study interviews.

End of study questionnaires were sent electronically or by postal mail to all the participants along with the last quality of life survey (at 12 months). In this thesis we will not analyse the staff interviews and the feedback questionnaires. These will be part of the next step of the qualitative analysis to submit as a manuscript to peer-reviewed journal.

Twenty end-of-study patient interviews were conducted and analysed using the framework approach to thematic analysis (195, 196). An analysis team was assembled (CP and FB), and all the interviews were coded by the 2 researchers. An initial basic framework of themes was drafted based on the modifications made on the interviews schedule and a codebook was drafted and subsequently amended at every interview. Following coding of the first three interviews, the framework was reviewed by the analysis team and amendments were made in accordance to the content of the interviews (i.e., to allow further themes and subthemes to emerge from the data). Thus, an iterative approach was adopted so that changes could be made to the coding framework as new themes and relationships between themes emerged. Regular meetings were scheduled to discuss any queries or discrepancies, and these were resolved by discussion and consensus. As learnt during the Oxford Course, I also explored the technique of creating a “one sheet of paper” (OSOP)(197), which I found very useful in visually organizing my codes.

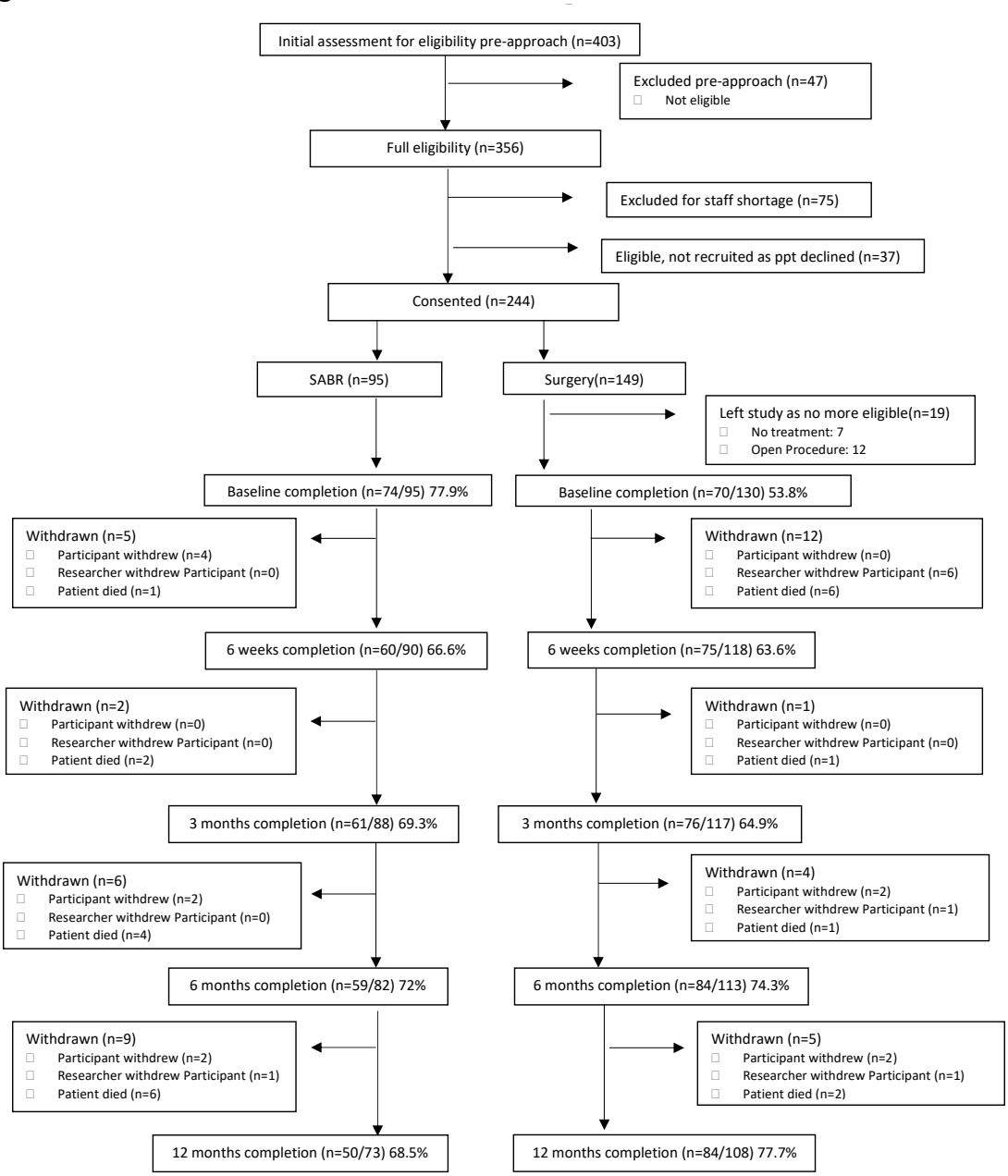
6.5 Recruitment Results

In total, 356 patients were eligible to participate in the study between 25th of February 2017 and 9th April 2018, meeting the estimated rate of recruitment and 244 consented to take part (60%) (Figure 6-3). Of these, nineteen patients were excluded from the analysis: 7 patients did not receive any treatments for oncological reasons and 12 patients became not eligible before the treatment as they had a further consultation with the surgeon and they were considered to be suitable for an open operation. Of the remaining 225 patients, 44 left the trial over the course of the follow up period: 23 patients died (13 in the SABR group and 10 in the Surgical one); nine patients became not eligible during the course of the follow up and 12 patients (8 patients in the SABR group and 4 in the surgical group) actively withdrew. In 5 cases the reason was that too much else going on at the time to consider taking part. The remaining 7 did not provide a reason.

At the time of writing, all patients have completed the twelve months (Week 52) follow up time point (last patient's questionnaire due for the questionnaire the 9th of April 2019), but the analysis will be focusing on the six-months results. However, I have completed the 12 months complications/censoring collection and the completion data in order to provide a clear and definitive picture of the study.

Data checking has been carried out in 5% of returned questionnaires.

Figure 6-3 LiLAC CONSORT Flow-chart



Two hundred and twenty-five patients were included in the baseline analysis (95 received SABR and 130 were submitted to VATS resection).

Recruitment in the SABR subgroup was slower than expected. More SABR centres have been opened during the last few years reducing the numbers of patients treated in Leeds.

Of the entire population consented, ninety-nine participants were male (44%) and 126 were female (56%). The mean age of patients in the study was 71.8 years (SD 9.22; range

40 to 95 years) but this differed according to treatment: SABR patients had the mean age of 74 years (SD 9.2; range 45 – 95); surgical patients had a mean age of 70 years (SD 8.8; 40-90). A summary of patient demographics, treatment and tumour characteristics are provided in Table 6-1.

Table 6-1 Baseline Characteristics of patients included in the study

	SABR (95)			Surgery (130)			
	Mean or Count	Median	IQR	Mean or Count	Median	IQR	p value
Age	74.3 (9.2)	75	70-80	70.0 (8.8)	71	65-76	0.0001
Sex males	37 (39%)			62 (48%)			0.21
BMI	27.3 (5.4)	26.5	24.5-30.0	26.3 (5.3)	26.5	22.9-29.0	
FEV1	76.6 (26.3)	77	60-90	88.0 (22.4)	91	73-101	0.0001
DLCO	71.0 (22.1)	71	57-82	83.4 (21.1)	82	68-97	<0.0001
CCI	2.1 (1.3)	2	1-3	1.2 (1)	1	0-2	<0.0001
PS>1	54 (57%)			21 (16%)			<0.0001
CAD	32 (34%)			9 (6.9%)			
CVD	12 (13%)			4 (3.1%)			
CKD	7 (7.4%)			1 (0.8%)			
Diabetes	22 (23%)			10 (7.6%)			
COPD	44 (46%)			38 (29%)			
current smokers	25			30			
ex smokers	67			85			
never smokers	3			15			
Tumour size (cm)	2.1 (1)	2	1.4-2.5	2.5 (1.4)	2.2	1.4-3.1	
Pre-treatment path diagnosis	38 (40%)			74 (57%)			
IMD quintiles (n,%)							
1 (least deprived)	31 (32%)			44 (34%)			
2	21 (22%)			21 (16%)			
3	14 (15%)			20 (15%)			
4	15 (16%)			27 (21%)			
5 (most deprived)	14 (15%)			18 (14%)			
All Complications	85 (89.4%)			95 (73%)			0.002
Minor complications (Grades 1-2)	64 (67%)			68 (53%)			0.028
Major Complications (grades 3-5)	21 (22.1%)			27 (20.7%)			0.809
Treatment related deaths within 90 days	3 (3.2%)			7 (5.3%)			0.52
Deaths at 1 year	14 (15%)			11 (8.4%)			0.14
Completed MITS				115 (88%)			
Converted to open surgery				15 (12%)			

1

At the time of recruitment only 27 patients (11%) stated they would complete the questionnaires online; however, through the course of the study 4 of them converted to receive the reminders of the questionnaires on paper. The reasons given at the time were varied: not checking emails regularly, using someone else's email/computer and

[1] BMI: body mass index; FEV1; forced expiratory volume in one second expressed as percentage of normal for age sex and height; DLCO: carbon monoxide lung diffusion capacity expressed as percentage of normal for age, sex and height; PS: Performance Status; CCI: Charlson Comorbidity Index; CAD: history of coronary artery disease, CVD: history of cerebrovascular disease; CKD: history of chronic kidney disease; COPD: Chronic Obstructive Pulmonary Disease; IMD: Index of Multiple Deprivation; MITS: minimally-invasive Surgery(i.e. VATS).

perceived ease of having the reminder on paper. 12.6% of female patients opted for the on-line compared to 11.1% of male ones (Table 6-2).

Table 6-2 Percentage of electronic completers among the known groups

Variable		Elecronic completion %	p value
Gender	female	12.6	0.837
	male	11.1	
PS	0-1	14	0.276
	>1	8	
Age	<70	14.5	0.401
	>70	10.6	
IMD	>5	12.6	0.837
	<5	11.3	
Education Skill	>5	11.4	1
	<5	12.3	
Treatment	SABR	7.4	0.095
	Surgery	15.4	

Electronic completion rates were not statistically different in known groups: income deprivation decile>5 (11.3% Vs 12.6% p=0.837), education skills decile>5 (12.3% Vs 11.4% p=1), Age>70 (14.5% Vs 10.6 p=0.401), PS>1 (14% Vs 8% p=0.276). There was a marginally significant difference between electronic completion rate in the surgical group (15.4%) compared to SABR one (7.4%), p=0.095.

Comparing the available demographic data of the total patients treated in the same period, we did not find major differences. This confirmed that the Lilac cohort is representative of the total population treated in the same period (Table 6-3).

Table 6-3 Characteristics of total patients treated in Leeds from Feb 2017 to March 2018

VATS lung resections (225)	Mean or Count	Median	IQR	SABR (189)	Mean or Count	Median	IQR
Age	69.3 (9.6)	70	64-75	Age	74 (9.1)	74	69-81
Sex males	104 (46.2%)			Sex males	80 (42.3%)		

6.5.1.1 Overall completion

Figure 6-3 shows an overview of recruitment and questionnaire completion rates by study group. At the time of the interim analysis, all patients had been invited to complete the final questionnaire at twelve months. Figure 6-4and Figure 6-5 show the overall completion rates of the quality of life questionnaires over the 12months period. There is a completely different trend between the two groups: surgical patients were less likely to fill the questionnaires at the beginning of the study. During the period between the preoperative visit and the first two months after surgery, lung resection patients are experiencing the highest rate of morbidity(198-200). These, along with the already described issue with the recent adoption of the ERAS programme, may explain the lower response rate in the initial part of the study for the surgical group. In the SABR group instead, we have a good return rate at the beginning and a slight constant reduction over the following time points. This may be explained by the baseline characteristics of those patients who are frailer and with poorer PS.

Figure 6-4 SABR completion rates over 12months

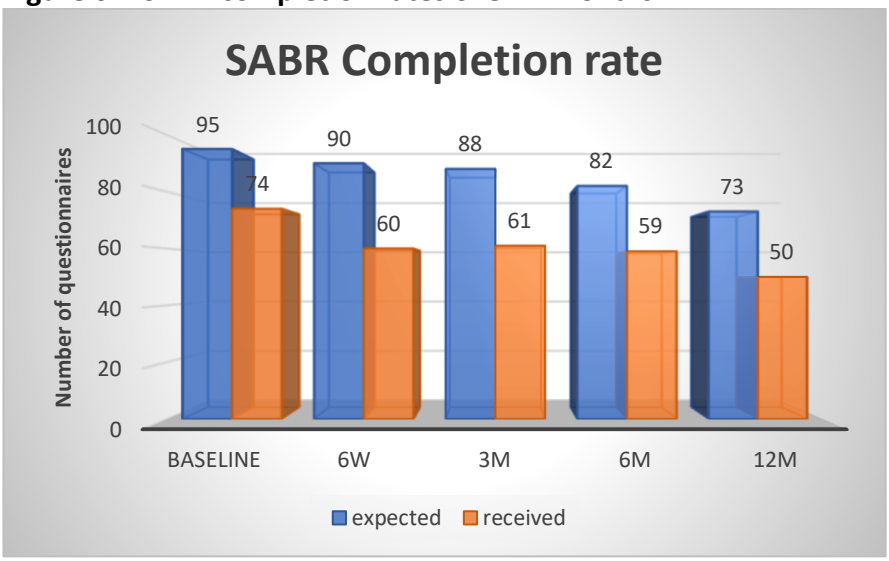
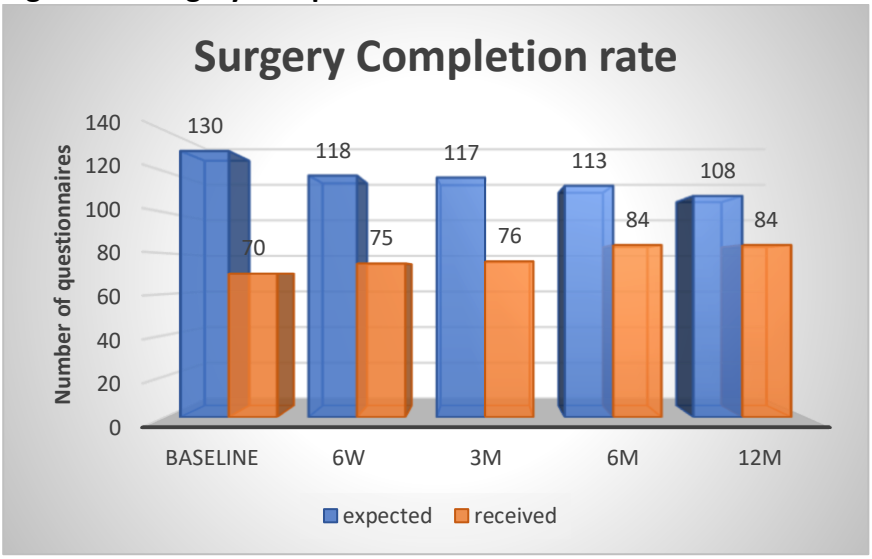


Figure 6-5 Surgery Completion rate over 12months



6.6 Decision-self Efficacy scale Psychometric Results

A total of 244 patients consented to the study of which 158 (64.7%) returned the Decision Self-Efficacy Scale. Of these, 73 (46.2%) patients were treated using SABR, and 85 (53.8%) had surgical resection. We have kept the 6 patients who were excluded by the analysis as they became not eligible before the treatment, had a further consultation with the surgeon and were considered to be suitable for an open operation. We consider that this psychometric analysis will not have been affected.

We did not find any baseline difference between patients who completed the DSE and those who did not complete (Table 6-4).

Table 6-4 Difference in baseline characteristics between DSE completers and patients who did not complete it.

	DSE compl (n=158)	no DSE compl (n=86)	p value
age (mean,SD)	72.4 (8.6)	70.6 (9.5)	0.17
gender (m,%)	69 (43.6%)	43 (50%)	0.34
highps	54 (34.1%)	23 (26.7%)	0.23

The baseline objective characteristics of the participants completing the DSE are in listed in Table 6-5.

Table 6-5 Characteristics of patients completing the DSE

Variable	Pat with DSE (n=158)	SABR (73)	Surgery (85)
Treatment (surgery, %)	85 (53.8)		
Gender (m, %)	69 (43.6)	26 (35.6)	43 (50.5)
Age (years, SD)	72.4 (8.6)	74.5 (9.3)	70.5 (7.5)
Comorbidity (yes,%)	135 (85.4)	67 (91.7)	68 (80)
FEV1% (SD)	83.5 (25.1)	75.5 (27.4)	89.2 (21.8)
DLCO% (SD)	77.6 (22.2)	69.6 (22.8)	83.5 (19.9)
Current Smokers (n, %)	34 (22.6%)	19 (27.1)	15 (18.7)
PS>1 (n, %)	104 (62.4)	39 (53.4)	15 (17.6)

The mean value of the DSE was 81.7 (SD 23). In the Surgical group the mean score was 83.6 (SD 22.9) and in the SABR group it was 79.5 (SD 23). DSE is the main score representing the overall efficacy in making the decision. 100 corresponds to the high level of efficacy.

A principal axis factor analysis was conducted on the 11 items with oblique rotation (direct oblimin) as it was expected that the factors would not be independent. The Kaiser-Meyer-Olkin (KMO) measure verified sampling adequacy for the analysis (KMO =.91) well above the minimum criterion of .50. In addition, all KMO values for individual items were ≥ .88 and Bartlett's test of sphericity was also significant at $p < .001$. An initial analysis was run to obtain eigen values for each factor in the data. Two factors had Eigen values over Kaisers criterion of 1 and explained 81.2% of the variance. The scree plot depicted two inflections confirming Eigen values over 1. Table 6-6 shows a factor loading, proportion of variance explained, regression coefficient and Eigen values after rotation, suggesting that factor 1 represents overcoming barriers and factor 2 represents information seeking.

Table 6-6 Factor Analysis results of DSE

Item on the DSE scale	Overcoming barriers	Information seeking
	Factor 1	Factor 2
5 Ask questions without feeling dumb	1.01	
6. Express my concerns about each choice	0.954	
7. Ask for advice	0.83	
8. Figure out the choice that best suits me	0.726	
9. Handle unwanted pressure from others in making my choice	0.911	
10. Let the clinic team know what's best for me	0.754	
11. Delay my decision if I feel I need more time	0.68	
1. Get the facts about medication choices available to me		0.999
2. Get the facts about the benefits of each choice		0.98
3. Get the facts about the benefits and risks of each choice		0.848
4. Understand the information enough to be able to make a choice		0.631
Eigen values	7.91	1.02
% of variance	71.96	9.13

The items correlated significantly at p=0.001 (range 8.88-5.55) only one was excessively large >.9 (between items 1 and 2). However, a determinant value of 2.09E-006 above the necessary value of 0.00001 revealed that the level of collinearity would not be detrimental to the analysis. Therefore, no items were removed. The correlation matrix is provided in the Appendices.

There is no difference between the two groups in terms of self-efficacy: SABR 79.5, Surgery 83.6 (p=0.09). When looking at all the eleven items linearly transformed, as described in the scoring manual, no statistical difference has been found between the two groups (see Table 6-7).

Table 6-7 DSE questions results in SABR and Surgical groups[2]

	Surgery (n 85)	SABR (n 73)	p value
DSE	83.6 (22.9)	79.5 (23.1)	0.09
*F1	84.3 (24.1)	80.6 (23.8)	0.13
**F2	83.2 (23)	78.8(26.2)	0.19
Q1	85.2 (23.5)	80.4(26.1)	0.13
Q2	83.5 (26)	79.1 (26.0)	0.1
Q3	84.1 (25.8)	80.1(27.3)	0.21
Q4	84.4 (26.1)	82.8 (24.6)	0.33
Q5	84.4 (30.8)	79.7 (30.8)	0.34
Q6	83.5 (26.8)	78 (30)	0.16
Q7	87 (26.8)	82.8(26.3)	0.11
Q8	84.1 (23.7)	79.1 (28.8)	0.35
Q9	80.8 (29.2)	77.3 (30.6)	0.51
Q10	83.8 (22)	77.7 (31)	0.5
Q11	79.1 (31.5)	76.7 (31.5)	0.88

DSE was not statistically different between female and male (p=0.37). Male patients have a mean DSE value of 84 (SD 21.3) and female of 79.9 (SD 24.2).

When comparing DSE among younger and older people we could not find any statistical difference (p=0.4). I have utilized as cut-off for the definition of older the median value of the population which is 72 years. In particular, older people have a DSE mean of 82.5 (SD 23.7) and younger people of 81 (SD 22.6).

Interestingly we found the patients with a PS>1 have a less efficacy in making their decision during the preoperative period. In fact, patients with PS>1 have a DSE mean of 73.8 (SD 26) compare to patients with a PS 0-1 who have a DSE mean of 85.8 (SD 20.3 p=0.0024). Furthermore, these patients maintained the statistical significance in both the subscales identified in our factor analysis: F1(PS>1:76.1 Vs PS 0-1:85.9 p=0.0052) and F2 (PS>1:71.6 Vs PS 0-1:85.4 p=0.0004).

2 *F2=Q1-Q4. **F1=Q5-Q11

6.7 Feedback Interviews results

6.7.1 Reflective account

Reflexivity expresses the ways in which the researcher and the research process may shape the data collection, including the role of prior backgrounds and experience(201). Within the context of the current study, I needed to consider the ways in which the interactions with participants might be influenced by my own professional background, experiences and prior assumptions. In particular, it is important to take into consideration possible impact on participants' willingness to talk openly about experiences, or how this might have shaped what they said during the interviews. This may also have affected the iterative process and conclusions drawn through the thematic analysis.

I am a qualified thoracic surgeon and I have had clinical responsibility for lung cancer patients since 2007. I also have had the opportunity to work in different countries and cultures, to learn how to adapt the clinical conversation to different social and cultural backgrounds.

As clinician, however, I am used to leading consultations and interactions with patients and obtaining information using traditional clinical interview methods. In qualitative research different skills and communication techniques are needed to elicit information from interviewees and make people feel at ease to disclose information and express opinions. My prior knowledge and clinical pre-assumptions on how colleagues and patients will think or behave, can also have influenced how interviews are conducted and what people were elicited to openly express their opinions.

I have been involved in QOL research for 10 years now, and this has completely changed my attitude towards patients and understanding of the impact cancer and treatment has on patients' lives. This has in turn influenced my approach to conduct increasingly patient-centred clinical consultations. My knowledge of QOL components, their correlation with the types of questions and issues by commonly expressed by patients in clinic, will have positively influenced how I conducted and analysed the interviews in this study. My approach to delivering the interviews was also shaped by working in a multidisciplinary team within PCOR (many of my colleagues in this group have social science or nursing backgrounds rather than medical). I had the opportunity to assist some interviews done by other experienced colleagues, but also had the chance over three years to regularly interact and learn from these researchers and their experiences of investigating and understanding the experiences of cancer patients.

It is important to recognise the potential influence of gender stereotypes in the surgical setting which lead the patient to not easily consider a female doctor the responsible surgeon(202). Stereotypes have a powerful impact when forming an impression. This can be confirmed by the fact that none of the patients that I approached about the study had recalled me as the surgeon. The possible bias in answering to our questions may have been limited by this, especially regarding the interaction with the doctors. Nevertheless, it has been noted that female physicians are perceived by patients to engage in communication that more broadly relates to the larger life context by addressing psychosocial issues and counselling(203). This may have influenced the interviews and the way patients engaged with me.

Including myself three researchers performed the interviews with staff and patients. All three had different backgrounds: medical, nursing and social science. The support of Dr Florien Boele, a social scientist, in analysing the end-of study interviews also assisted in

giving a more independent view to these results as she has not been previously involved in the Lilac study. These combined perspectives helped to reduce the potential bias coming from the interviewer's background.

6.7.2 Pilot Interviews

A subset of patients (6 patients, 3 declined) were approached at 6 months and asked to take part in a semi-structured interview about their experiences of using LILAC. Patients were approached consecutively as they completed the six months questionnaire, with an aim to interview 3 patients overall from each group (Surgery Vs SABR). Interviews took place in a private room in the oncology outpatient clinic at St James' University Hospital, Leeds or over the phone. We have conducted 6 interviews in total also because not many patients have decided to fill in the questionnaire on line. The main results of those interviews were that the interview schedule was not able to capture the main issue with Lilac as was too directed in exploring the experiences with electronic completion of PROMS. As a result, I have involved in the following advisory group meeting Dr Simon Pini, Dr Kate Absolom, Dr Florien Boele and Dr Trish Holch, who are experts in qualitative research. We have analysed the interview schedule and made the necessary changes in order to better understand the experience of patients with the Lilac study. In fact, the former interview schedule was based on that used for the end of study interviews in the eRAPID usability in the breast clinic where the patient demographics are quite different.

I produced a new schedule for both staff and patients and I submitted these to IRAS form to get HRA approval. It was labelled as substantial amendment and granted the approval on the 09/05/18.

Specific questions were also added on the clinician's use of the symptom graphs, which allowed them to view responses over time and also type of symptom/questions that were not captured by the questionnaire but important to them. These themes had emerged as an important factor in the interim phase. Some general questions about patients' experiences of hospital admissions and treatment changes during the 12 months were also added. Detailed questions about the electronic procedures were removed, as only a modest part of the patients completed questionnaires on line. Rather more information about the reason why they have not chosen the electronic platform was explored. The full amended semi-structured interview schedule is presented in Figure 6-6.

Figure 6-6 Patients interviews schedule

Theme	Questions
Consent	1. If you can remember we approached you on (DATE), what was happening at that time? (Prompts – consultant, CNS, pre-assessment etc.) 2. Were you approached by a doctor/nurse/radiographer about the study before the researcher came to speak to you? 3. Can you think of a different way/time you would have preferred to have been approached to take part in the study?
Questionnaire Completion Method	1. Can you tell me about why you chose to complete the questionnaires online/on paper? Follow up questions (if not already mentioned): a. Are there any reasons why were this completion method preferable to completing on paper/online? (If not covered in first answer) b. How often do you use a computer or smartphone? c. How often would you say you use the internet? (if never, ask "do you have access to the internet?" direct/indirect") d. (if yes) Do you use the internet to access health information? e. (If yes) How confident would you say you are in using the internet? f. (If no) Do you have a family member/friend who could help you use the internet? g. Do you have an email address or smart phone that you feel confident using?
Patient Engagement	1. I can see that you completed X number of questionnaires. Can you think back to what was happening at X time point? (approximate date, e.g. 6 weeks post-op). Were there any potential barriers to you completing questionnaires at that time? (prompts re treatment) 2. What would be your preference for receiving reminders to complete the questionnaires? (i.e. phone call, email, text, letter). 3. How would you feel about receiving a phone call to remind you at each time point (i.e. before treatment, 6 weeks, 3 months, 6 months and 12 months after treatment)?
Content and Frequency of questionnaires	1. What areas of your life were affected most by the illness and treatment? 2. How useful do you think the questions were in enabling you to express how you were feeling at the time? 3. Did you think the style of questions were useful in enabling you to tell us how you were feeling? 4. Was there anything about the questionnaires that you felt weren't relevant to you? 5. Were there any questions that you didn't like/want to answer? (If so which?) (prompts re questions/areas of QoL asked about) 6. Do you have any suggestions for questions which weren't there but you would have liked to have seen? 7. You were asked to complete the questionnaires at 6 weeks, 3 months, 6 months and 12 months following treatment. What do you think about the time points at which you needed to complete the questionnaires? (i.e. would it be easier closer together or was it better that there was time in between)
Clinician Involvement	1. Typically, in your appointments what sort of issues would you discuss with your clinician or them with you? 2. In your appointments with your clinician have you discussed any aspect quality of life with them? (i.e. ADLs, social activities and impact of symptoms/pain) 3. (If yes) Did you or your clinician refer to your answers to the questionnaires? (Follow up) if yes/no how did that make you feel when they did/didn't refer to your answers? 4. (If no) How important do you feel it is to discuss your quality of life with your clinician?

6.7.3 End of Study interviews

We approached 37 patients, 6 of them declined for various reasons, most of them being too overwhelmed by the events (2) and two as having other commitments so not able to answer the telephone and have time to speak. The remaining two did not give explanations. 11 did not answer phone calls, although they were informed and

consented at the beginning of the Lilac study for these interviews. A final set of 20 patients' interviews was reached. Data saturation started after 18 interviews, after which additional two interviews were conducted to confirm we had reached data saturation.

I personally conducted 9 patient end-of-study interviews. The rest of the interviews were conducted by other two member of the research staff (ES, BC). The amended semi-structured interview schedule was used and if open-ended questions elicited brief response, prompts were offered. Seven interviews were conducted in person, the remaining over the phone as preferred by the patients. When face-to-face, participants were presented with the questionnaire to help them recall the questions. The interviews were transcribed verbatim. 8 SABR patients and 12 Surgical patients were interviewed reflecting the actual proportion of patients recruited per group. Less electronic completers were interviewed as we had tried to invite most of them but unfortunately, they have not answered or declined. Details of the patients involved is presented in Table 6-8.

Table 6-8 Interviewed Patients characteristics

Participant	Gender	Age	Completion type	Group
Patient 1	F	70	P	Surgery
Patient 2	M	80	P	SABR
Patient 3	M	79	P	SABR
Patient 4	F	75	EL	Surgery
Patient 5	F	77	EL	SABR
Patient 6	F	72	EL	Surgery
Patient 7	M	72	P	Surgery
Patient 8	M	84	P	Surgery
Patient 9	M	70	P	SABR
Patient 10	F	76	P	Surgery
Patient 11	F	80	P	SABR
Patient 12	F	82	P	SABR
Patient 13	M	65	P	Surgery
Patient 14	F	72	P	Surgery
Patient 15	F	72	P	SABR
Patient 16	F	42	P	Surgery
Patient 17	M	66	P	Surgery
Patient 18	M	83	P	Surgery
Patient 19	F	76	P	Surgery
Patient 20	F	82	P	SABR

6.7.3.1 Thematic analysis of interviews

Four main themes were identified in relation to patient engagement with the study and the benefits of the quality of life data collection for patients. Themes and subthemes are outlined in Table 6-9.

Table 6-9 Themes and sub-themes of end-of-study interviews

Consent process	Introduction to the study
	Timing of invitation
	Psychological situation
	Member of the staff
Completion methods barriers and facilitators	Perceived facilitators for paper completion
	Perceived facilitators for online completion
	Perceived barriers for paper completion
	Perceived barriers for on-line completion
	Computer/Internet use
	Family support for Lilac participation
Patient's engagement with Lilac	Reminders preferences
	Frequency of assessment
	Barriers to complete questionnaires
	Questionnaire role in expressing symptoms/issues
	Comments on questions and suggestions
Perceived influence on clinical care	Involvement of clinicians in quality of life discussion
	Clinician and staff perceived engagement with Lilac

The themes were as follows:

1) General comments on the approach and consent process.

2) Completion methods barriers and facilitators. This theme has highlighted the patients perceived difficulties in choosing filling the study's questionnaire on paper or on-line.
3) Engagement with Lilac. This theme explores the experience of the patients with the Lilac questionnaires, reminders and frequency of assessment.
4) Perceived influence on clinical care. This theme describes patients' experiences of the impact of Lilac on their consultations with clinical staff.

6.7.3.2 General comments on the approach and consent process.

Patients were generally very satisfied by the consent process, preferring the face-to-face approach to a phone call or a posted letter.

I preferred it that it was personal rather than through a letter or something like that.
(pat 20, female 82Y SABR)

Most of the patients were happy to have the research study introduced to them by a doctor, feeling reassured that they were aware of what was going on at that moment. However, patients could not recall too much on detail on what was discussed. They did feel that it was always very nicely asked and explained.

"You know the day you see the surgeon and discuss all about what they're going to do, and then asked if you'd do the survey, I think it's the only way you can do it."
(PAT 14, female 72Y surgery)

"I'm glad he did ask because he knew what he was talking about. (PAT 6 female 72Y Surgery)"

With regard to the psychological moment the patients were going through during the period of consent to LILAC, they indicated that they had received a lot of traumatic news and it was difficult to concentrate on additional details.

"*when* somebody tells you that you, 'we're treating you for cancer of the lung' you don't think about research!... You're thinking about yourself a little bit and no, I mean I just went along with whatever they approached me with" (PAT 8, male, 84Y surgery)

This is particularly repeated by surgical patients. One patient however, indicated that the study was perceived as a welcome distraction from this difficult moment.

6.7.3.3 Completion methods barriers and facilitators

Patients were asked to comment on their choice of either paper-based, or online questionnaires and also on their IT literacy. This point was particularly important as we had less on-line completers than expected.

Patients completing on paper mentioned that their choice was driven by the fact that they found it easier to access and allowed them to have time to reflect, sit down and have this piece of paper in front of them.

“At my age, I think I find it easier to do things on paper where I can actually see and then if you go wrong when you’re doing it online, it kicks you back to the beginning or it won’t let you carry on and that’s annoying to me” (PAT 15 female 72Y SABR)

Patients choosing the on-line completion methods found it easier as one patient had handwriting that is difficult to read, or reported choosing the online method for ecological reasons. However, some who tried the electronic website for Lilac found very irritating not to have the possibility of skipping questions and come back to them later. The majority of patients we interviewed did not consider themselves to have a good level of IT literacy. Either they did not have the access to internet or they did have devices but did not use them regularly. One patient reported to be scared to be involved in on-line medical advice or websites. Another patient reported that being an older lady she did not feel skilled enough to complete the questionnaire on the computer.

“I’ve got a mobile phone but I don’t, I just use it for calls and I don’t ever use the Internet really” (PAT 1 female 70Y Surgery)

“Because the only line I've got, love, is a washing line” (PAT 3 Male 79Y SABR)

“I’m still a bit illiterate like but it was all before my time, I’m 80, this was all before my time it was” (PAT 11 female 80Y SABR)

I've got the only mobile telephone of me own, that never goes out of the house, I’m that way with computers and telephones and stuff, I just don't get on with them, no..” (Pat 18 Male 83Y Surgery)

One patient described how his granddaughter supported him to complete the symptom reports, and they went through the questionnaire together. The majority of patients however, said they are not keen to ask family for help in completing the study although they are aware they do have access to internet. Despite this limited level of IT literacy overall, five patients mentioned they are using internet to access the GP practice on-line services.

6.7.3.4 Engagement with Lilac

Most of the patients agreed that reminders should be sent to help people to not forget about the questionnaires. Preferences on reminder mode varied from person to person.

"I'm happy online but I think you have to do both because some older people are not very happy using the internet. So for me, you can send me an email, I'm quite happy to go that way…. I'm sort of technologically minded... …but a lot of people are not" (PAT 5 female 77Y SABR)

Interestingly, many of them preferred a posted letter or text message. One patient highlighted the fear of fishing calls, having lost trust in people ringing for commercial reasons.

In general, patients were happy with the frequency of the questionnaires. Two patients mentioned the difficulty in completing the earliest one after treatment (six weeks). One patient suggested having another one at 18 months after treatment and another additional assessment to better discriminate symptoms over time.

"I was getting over, I was just getting over the operation because the my heart played up a little bit, so I had a few complications and the district nurse was coming out, and then when that first questionnaire come through, it took me, I think it were three or four days before I could actually figure out, that I felt like I could do it" (PAT 14 Female 72Y Surgery)

Patients identified some barriers in completing the questionnaires: a few patients reported not receiving some surveys, two acknowledged losing it and two patients identified their health issues as a reason for not being able to complete it.

Furthermore, patients appreciated that answering the questions allowed them to express problems and health issues. One patient even stressed that writing down is a way to share the sorrow and psychological issues. Another patient pointed out the importance of completing these questionnaires to follow the quality of life changes over time.

Patients especially indicated that it allowed them not to feel alone and be reassured that someone is looking after them.

"Because like in-between these questionnaires I've spoke to other people as well, you know, and like see how far I'd come and that. Yeah, when I'd filled one questionnaire in and then I filled another one three months, I felt "oh I'm getting better here" (PAT 1 female 70Y Surgery)

"..you know, I think it was a good way, sort of like you are sharing your, you know, sorrows and your pain with, you know, through paper" (PAT 16 Female 42Y Surgery)

"But just I think it would reassure people that somebody's actually interested" (PAT 4 Female 75 Surgery)

Patients were content with the questions in general, but two of the responders indicated that some of the questions did not apply to them. Another patient reported the questions to be too generic. Main suggestions were in fact to add space for explanations next to the questions, or a footnote were to put comments as sometimes the pain for example was not coming from the lung cancer treatment, so they would have preferred to specified this. A few suggested to add space for medications, other health conditions and surgical-specific notes.

The younger patient also said that it would help to include specific question about family with kids arrangements.

6.7.3.5 Perceived influence on clinical care

Most of the patients indicated that they discussed the everyday life with the clinical team. Five patients reported that they have not discussed QOL, however sometimes, acknowledging doctors are very busy. One patient mentioned that often the question is a generic one that made him to answer only for specific health issues, not if everything is alright. Another patient said that clinical visits may be avoided if more QOL aspects would be discussed.

"..it won't take 10 minutes or something like that because I appreciate people are very, very busy. But I think a lot of, after telephone, visits and things, could be avoided if a few caring questions could be asked" (PAT4 Female 75Y Surgery)

"No, I didn't, no, love, no, I just discussed about walking and that, yeah" (PAT 19 76Y Female Surgery)

All the patients but one said that the LILAC questionnaire result was not mentioned during their doctor's consultations. Only one patient indicated that the clinician has referred to his answers.

6.8 Discussion

This Chapter showed the feasibility of collecting PROMS in a cohort of early stage NSCLC patients treated with VATS lung resection and SABR. We have demonstrated that the collection of the Decision Self-efficacy Scale questionnaires is feasible in a cohort of lung cancer patients. We demonstrated furthermore, that the only characteristic of the patient impacting on the decision self-efficacy was the PS. This data showed that only 11% of patients chose to fill in the questionnaires on-line. The baseline clinical characteristics confirmed the marked difference between these two populations, with the SABR patients less fit than the surgical ones.

Our results showed a good response rate considering the fact that few patients consented to complete the questionnaires on line. For longitudinal studies involving regular PROMS, we typically see 70% consent rate and 30-35% attrition over 3 months(75). However, these studies have not involved lung cancer patients, with older age and multiple comorbidities. In lung cancer studies the response rates at 12months are very different as pointed out in Chapter 3. Our attrition rates are in line with the recent literature. A 12 months longitudinal surgical RCT reported a lower QOL questionnaire return rate at baseline (148). In the SABR group, our results showed a

better rate of questionnaire return at 12 months in the SABR group compared to published longitudinal data(139, 143).

We believe that the low surgical completion rate at baseline may be attributed to the difficult logistic situation during the preoperative appointment. Major efforts should be done in the future to remind the patients to return the questionnaire before the treatment (with phone-calls or mail) considering the relative importance of the baseline information in QOL studies.

We also provided evidence for the validity of the DSE as a11-item measure of two main subscales: the main factor (1) explaining almost 72% related to overcoming barriers to decision making (items 5-11) and the second factor explaining the additional 9.1% of the variance (items 1-4) was related to information acquisition. This questionnaire is summed to create one global item that measures the patient's self-confidence or belief in their ability to obtain relevant decision-making information including shared decision making taking into account several aspect of the decision process(98). However, this questionnaire was not tested in a cancer population, but only in menopausal woman and psychiatry patients(100, 204), limiting the comparison of our results.

We have not identified a difference in efficacy between our two treatment groups, however, the sample size was too limited to draw conclusions. We showed indeed that patients with poor performance status were most likely to be less confident in making their decision for treatment. We acknowledge that patients were filling the DSE questionnaire after the decision for their treatment has been made. However, the more compromised ones were still feeling that they were not involved completely in the decision-making process. Patients who have a worse PS and limited functional capacity tend to have more difficulty tolerating rigorous cancer treatments. These patients have less favourable outcomes than more fit patients with better PS, regardless of the

treatments given (205, 206). These patients in fact reported less confidence in both seeking information from doctors and overcoming barriers aspects when making their best decision on treatment. One explanation can be that regardless the type of treatment offered when the patient is less independent physically (as in those with poorer PS score), they show more difficulty in making their own medical decisions.

On the other hand, it has been demonstrated that amongst English clinicians, performance status, cancer stage, comorbidities, ability to tolerate treatment, biological age and toxicity of treatment were all significantly more influential than chronological age. In addition, performance status and cancer stage were significantly more influential than biological age. In this sense, clinicians may involve less those patients with poorer PS, where they are more concerned about an expected higher morbidity and mortality. In those cases, patients may perceive similarly less confidence in making that decision which is more "physician-driven".

This data confirms the importance to identify high-risk patient subgroups which will benefit of programmes aimed to improve their participation in treatment decision-making contexts.

The conflict, more than the social and emotional component of the difficult-decision making may be considered when evaluating the routine data collection in complex clinical area like this one (207). This may help in identify people with a greater need of help and support in making decision and will help in tailoring specific decision aids.

One of the most important aspects of this PhD has been to link quantitative data to the qualitative one: being one of the first times I conducted semi-structured interviews, I required adaptation to this new method with formal and informal training throughout the course of my PhD. For this reason, since the beginning of this project I have been shadowing people in the group that are experts in qualitative research. Since the

beginning of the Lilac protocol drafting, I have organized meetings with other research staff like Dr Simon Pini and Dr Trish Holch to better understand the importance of a good schedule for the interviews. I also shadowed them in some interviews and listened to the previous recorded interviews for training purposes. This has been particularly useful after the first interim interviews with patients where we have noticed that the interviews scheduled needed modifying to better meet the aims of Lilac. Ultimately, I also attended the University of Oxford two day course on analysing and interpreting qualitative interviews which allowed me to better understand the value of the patient's voice in a research project. I strongly believe that the clinical experience has given me the possibility to deeply understand the patient's experience during clinician/HCP consultations and to focus this aspect during the interviews. What I learnt in clinical consultation with patients has facilitated the understanding of patient's comments on clinical care and participation in research during the difficult period they were going through. This may be considered a bias on one hand, on the other hand the qualitative part of this study was an important learning experience as a researcher and as clinician because what the patient was reporting during the interviews is absolutely unique and valuable.

The aim of this work was to use qualitative methods to explore barriers and motivators for patient engagement with Lilac and the impact of Lilac on patients' experiences of care. Findings supported the demographic characteristics already found in the quantitative assessment. Lilac was well accepted by patients, but not extensively used in clinical setting by the staff. The level of IT literacy was the main barrier to participate on-line on the study and even patients who described themselves as being able to receive help from highly IT-literate family members, reported not wanting to bother them to access and complete the questionnaires. This may justify the fact that compared

to other studies including oncological patients in the same centre, Lilac has not included more than 11% of online responders. Patients have suggested maintaining the paper modality as an option.

Although the questionnaire was carefully selected as cancer specific and internationally validated, it was not able to capture specific issues of surgical patients. Also, SABR patients were concerned about questions being too generic which calls for additional notes to explain their answers. Regarding this issue I had already received some feedback from patients who rang me during the study to explain for example that the pain they were indicating was not related to their lung treatment but to their back problems. Future publication of the updated version of the LC-13 questionnaire is expected to help in overcoming these barriers.

The patients were happy to receive the reminders through their post or via text. Most mentioned that the first approach should be done face-to-face and not by letter. My experience in dealing with a patient complaint after receiving the PIS by post, was another example of this important point raised by patients. What she discussed with me in two meetings was that the stress that they are experiencing in those weeks before surgery is enormous. This stress may be exacerbated by receiving an unexpected leaflet clearly naming lung cancer. The crucial role of the face-to-face approach was also indicated by all the positive comments on the modalities and explanations during the consent process.

While the majority of patients perceived the completion of QOL questionnaires helping with the self-awareness of their QOL and providing valuable additional information to share with their clinical team, most of patients did not discuss the results of the LILAC in clinic.

Lung cancer patients are a subgroup of patients with a low level of IT literacy which has limited the use of the LILAC study through the on-line platform. To increase the uptake of LILAC by clinicians, more effort should be done to train the clinicians and nurses to use patients' responses. Most importantly these qualitative interviews have highlighted the importance for patients to discuss QOL aspects with their clinical team. The lack of discussion of QoL with the clinical team when there are no issues, may also unveil the necessity of better understanding of the patient-clinicians communication and relationship.

This study is an observational study on a relatively large sample size, collecting multiple clinical information and repeated PROMS over 12 months after treatment. It is also the first study collecting more than 158 questionnaires exploring the self-efficacy around the early stage lung cancer treatment decision process. The information gathered during the patient's interviews was also the first to our knowledge to explore the IT literacy and the QOL topic in this elderly population.

6.9 Limitations

This study is an observational longitudinal study and comparing the two groups outcomes was not one of the aims. I decided to adopt this methodology instead of an RCT as previous trials directly comparing the effect of these two treatments on QOL failed due to low recruitment (41, 42). The populations described in this study are completely different from their clinical characteristics, so the evolution of their post-treatment QOL may have been affected by other factors, not only by the surgery or SABR. However, the main outcome was to follow these two populations in a consistent and standardized way.

Results from the qualitative part of the SABRTooth will give more understanding of reasons of non-randomization in this setting and may help in overcoming these barriers in future trials design.

This work has demonstrated the acceptability of PROMs collection in clinical setting. However, the recruitment has been lower than expected and lower than the audit surgical study. Our study is limited by its single-centre setting and small numbers of patients.

The low recruitment in the SABR group has been in part explained by the implementation of SABR treatment in other hospital of the closest regions. Furthermore, the sample size for this group was estimated based on surgical numbers as this is the first experience of PROMs collection in SABR patients in our setting.

Conversely, in the surgical group the logistic issues have played an important role in the recruitment. The research assistant job was delayed over six months. I was the only responsible person of recruitment for the first 5 months. The additional help and support of the PCOR staff and of the research radiographer have ensured that the recruitment carried on with no major issues. However, I cannot exclude that a dedicated research assistant from the initial part of recruitment may have increased the number of consented patients.

The qualitative work has been characterized by limited participation of the electronic completers, possibly limiting the generalizability of the results. Most of the patients displayed evident issues with memory retrieval. Further interviews at 6 months may have given more understanding and feedback of the approach and consent process.

Although not in the aim of this thesis, more efforts should be made in training the clinical team in the ePROMS consultation. This may not exclude an improvement in patient engagement with the study if the clinical team is providing more feedback during

consultations. The future analysis of the staff interviews and patient feedback questionnaires may give more insights on how to support health care providers and patients in QOL results discussion.

6.10 Conclusions

In summary this chapter shows the feasibility and acceptability of PROMs collection in early-stage NSCLC patients treated with radical intent. Barriers to the electronic completion and consultation may be considered in designing future feasibility studies.

We confirmed that the two groups cannot be directly compared as SABR patients were likely sicker given their lower FEV1%, DLCO%, high CCI, higher presence of ischemic heart disease and more people with diagnosed COPD. This information needs to be considered when analysing the PROMs results in the next chapter.

Chapter 7 Life after Lung Cancer (LILAC) prospective study: Quality of Life Evolution and Patient Satisfaction results

7.1 Background

The effect of surgical treatments on QOL has been investigated in a growing number of trials in recent years; however, the instruments and metrics used for analysis have varied, which complicates the interpretation and generalization of such studies. Regarding SABR, a recently widely implemented technique, the studies investigating the effect on QOL are limited by small sample size or assessments as demonstrated in the chapter 3.

It has been shown that, independently by the instrument used, patients submitted to surgical treatment for lung cancer, experienced the most consistent decline in their QOL during the first three months after surgery. The aspect of QOL mostly affected is physical function (PF). This decline partly recovers in the next 3-12 months, but the standardized mean difference remains at medium relevance(86). Several factors have been found associated to the prediction of postoperative QOL: the extent of surgery has been consistently reported as the most important factor influencing QOL. Patients submitted to pneumonectomy report the most relevant decline in QOL, particularly in the physical domain. Consequently, we decided to exclude pneumonectomy from this study (64, 208, 209).

Although we acknowledge the difficulty in designing a full RCT comparing these two treatments from a patient-reported outcomes perspective, the LILAC study is the first to follow the two treatment groups systematically for twelve months after the start of treatments.

7.2 Overview

The previous Chapter described the methodology of the Lilac prospective study and has given all the baseline overview of the patients involved in this study. This Chapter focuses on describing the QOL trajectory of the two patient groups (surgical and SABR) over the first six months after their treatments. It also describes the results of the patients' satisfaction with care assessed after 6weeks from the initial date of treatment. The results presented answer the main research question of this prospective study builds on previous work to further explore the potential of Lilac to investigate in a more rigorous way the role of PROMs during treatment for NSCLC.

7.3 Methods and Aims

As described in the previous Chapter, two hundred and twenty-five patients were included in the analysis (95 received SABR and 130 were submitted to VATS resection). Baseline questionnaires were completed at time of consent prior to or within 0-14 days of starting treatment. At this time patients opted to either complete the questionnaires online, receiving either email or letter reminders, or to complete paper questionnaires which were posted to them. Patients were invited to complete the questionnaires at the same time points. These time points broadly coincided with usual follow up schedules for patients to allow them to ask advice if needed. In particular, they were asked to

complete the questionnaires at week six, then at 3, 6 and 12 months after the initial date of treatment.

Electronic results were immediately available for viewing by clinical staff in patient's electronic health records (EHR). From the start of the study all paper results were also inputted into QTool and the results made available with a slight delay due to the need to wait for the paper survey to be posted back.

As already described, quality of life was investigated through the validated cancer specific EORTC core questionnaire (EORTC QLQ-C30) and the disease specific module for the relevant cancer site (Lung Cancer Module LC13) (88, 89). In this thesis only results from the QLQ C-30 will be presented.

Patient's satisfaction was assessed through the administration of the Patient Satisfaction Questionnaire Short Form (PSQ-18), which is a cross-cultural validated survey for use in different settings. The questionnaire is a 18-item self-administered survey including different scales reflecting the perceived level of satisfaction in relation to the care provided by doctors (94).

The research aims of this chapter are to:

- Describe the longitudinal QOL evolution of early stage NSCLC patient treated with SABR and VATS resection over the first six months after treatment (through descriptive statistics)
- Explore the clinical significance of the QOL changes over the first six months after treatment (through responder analysis).
- Evaluate the association of PROMS and clinical factors (through statistical modelling).

7.3.1 Statistical Analysis

This chapter will use descriptive statistics to describe the evolution of quality of life over the six months period. All quality of life scores were calculated according to the scoring manual (88). The summary score was calculated according to Giesinger et al.(125) which led to a single score ranging from 0 (worst) to 100 (best). Descriptive statistics included frequencies and percentages, medians and inter-quartile ranges. Normality of distribution of numeric variables (including the QoL scores) was assessed by the Shapiro Wilk test. Between groups calculations made use of the independent t-test for numeric variables with normal distribution or of the Wilcoxon rank-sum test for those without normal distribution. Categorical variables were compared by using the Chi-square test or the Fisher's exact test (in case of 10 or fewer variables in at least one of the cells).

All analyses were exploratory in nature, thus significant p-values ($p < 0.05$) can not be interpreted as confirming a priori hypothesis. The analysis was performed on Stata 15.0 statistical software (Stata Co., College Station, TX).

For a better understanding of the clinical impact of the quality of life information we also performed responder analysis (210, 211). This approach can help identify important differences at the individual level compared to traditional approaches using summaries of group means and standard deviations. Responder rates can be understood more intuitively than a difference in means of rating scales and has been proposed or recommended by regulatory guidance or clinical communities to be used in clinical trials. Draft guidance from the FDA(72, 212) on patient-reported outcomes specifically endorsed the responder analysis as an alternative approach to assessing clinical relevance. One of the major problems with the responder analysis is, however, the well-known issue that dichotomization tends to result in a loss of statistical power compared

to an analysis of the original continuous variable(213, 214). The other one is the choice of cut-off to define response. Yet in many disease areas across different clinical trials, various definitions of response have been used, and there is no consensus as to which is the most appropriate one(215). We performed individual level analysis of 'change scores', using the FDA recommendation for individual level analysis of change score and assuming that a clinically meaningful change = ½ SD.

The reason why we have decided to run this analysis is clearly stated in the FDA document defining a minimum important difference: "*Many PRO instruments are able to detect mean changes that are very small; accordingly, it is important to consider whether such changes are meaningful. Therefore, it is appropriate for a critical distinction to be made between the mean effect seen (and what effect might be considered important) and a change in an individual that would be considered important, perhaps leading to a definition of a responder.*". Responder analysis can be used to interpret the results in a meaningful way for clinician and patients (216).

I have also explored possible associations of baseline characteristics with the three main outcomes of the treatments: clinical outcomes (postoperative major complications), patient-reported outcomes (QoL), and quality outcomes (patients' satisfaction).

Three different sets of statistical modelling have been run to explore possible variables association with these outcomes in each treatment group (SABR and surgery).

Zoe Rogers, the research assistant working on Lilac, conducted the Responder Analysis and the Patient Satisfaction modelling. I performed all the rest of the statistical analysis.

7.3.1.1 Clinical model

For the clinical outcome (major complications) logistic regression analyses were used through a stepwise approach with backward elimination (p for variable retention <0.1).

Major complications were defined as those with Common Terminology Criteria for Adverse Events (CTCAE) grade>=3 (217).

In particular, the following variables were initially screened as factors to be entered in the regression models by testing their univariate association with post-treatment major complications: age, sex, body mass index (BMI), forced expiratory volume in 1 second (FEV1), carbon monoxide lung diffusion capacity (DLCO), history of coronary artery disease (CAD), cerebrovascular disease (CVD), chronic kidney disease (CKD), diabetes, performance status >1, IMDR, Charlson's Comorbidity Index (CCI), baseline quality of life scales (GH or SumSc).

Variables with p value<0.2 at univariate analysis were all entered in the logistic regression analysis (dependent variable: major complications). The final model included only those variables with a $p < 0.1$.

7.3.1.2 QoL model

Variables associated with variation of quality of life scales over time were tested using a random effect panel regression applying a backward stepwise approach with p<0.1 to retain variables in the model. The following variables were initially entered in the model (age, FEV1, DLCO, sex, CAD, CVD, CKD, Diabetes, PS>1, IMD, BMI, CCI). For these analyses, the dependent variables (quality of life scales) were analysed as panel longitudinal data.

There are two kinds of information in cross-sectional time-series data: the cross-sectional information reflected in the differences between subjects, and the time-series or within-subject information reflected in the changes within subjects over time. Panel data regression techniques allow you to take advantage of these different types of information. Although it is possible to use ordinary multiple regression techniques on

panel data, they may not be optimal (17). The estimates of coefficients derived from regression may be subject to omitted variable bias—a problem that arises when there is some unknown variable or variables that cannot be controlled for that affect the dependent variable. With random effect panel data model, it is possible to control for some types of omitted variables even without observing them, by observing changes in the dependent variable over time. This controls either for omitted variables that differ between cases but are constant over time, or for omitted variables that vary over time but are constant between cases.

One of the main aims of this thesis was to explore the role of QoL in the pre-treatment risk prediction model for early stage NSCLC. I was particularly interested in learning the statistical methods of modelling and logistic regression. For this purpose, I attended the two-days course on statistical modelling at the Imperial College of London in December 2018. We have run two separate cumulative multivariable analysis of the association between demographic, psychosocial and clinical factors and Global QoL and SumSc at 6 months. We aimed to investigate the relationship that demographic factors (age, gender and Index of Multiple Deprivation), clinical variables (forced expiratory volume in one second (FEV1%), Charlson Comorbidity Index (CCI), Performance Score (PS)) and patient-reported factors (e.g. Decision Self-Efficacy (DSE)), have with six months Quality of Life (QoL) in patients undergoing treatment for early stage non-small cell lung cancer (NSCLC) adjusting for the treatment received and their baseline QoL. Outcome measure was QoL at six months evaluated through the two generic scores (GH and SumSc). For the multivariable analysis, a Bonferroni correction was made to the significance level for multiple comparisons in order to limit the possibility of a type I error. The adjusted significance level used was 0.008 (i.e. $\alpha = .05/6$). The results of these two last models are provided in the Appendix.

7.3.1.3 Patient satisfaction model

I also wanted to investigate if any specific preoperative factors, either clinical or patient-reported were associated to a different level of patient satisfaction with care at 6 weeks after treatment. Variables initially used in the analyses were the following: age, preoperative FEV1%, gender, DSE (decision self-efficacy scale score), IMD (index of multiple deprivation) decile, baseline GH score of the EORTC QLQ C-30, PS (>1) and CCI (Charlson Comorbidity Index). Two different models were run as the group were too heterogeneous. Variables with a p-level <0.1 were used as independent predictors in logistic regression analyses.

7.4 Clinical Outcomes

Two hundred and twenty-five patients were included in the analysis (95 received SABR and 130 were submitted to VATS resection).

44 left the trial over the course of the 12 months follow up period: 23 patients died (13 in the SABR group and 10 in the Surgical one); nine patients became not eligible during the course of the follow up and 12 patients (8 patients in the SABR group and 4 in the surgical group) actively withdrew. In 5 cases the reason was that too much else going on at the time to consider taking part. The remaining 7 did not provide a reason.

At the time of writing, all patients have completed the twelve months (Week 52) follow up time point (last patients due for the questionnaire the 9th of April 2019), but the quality of life analysis of this thesis will focus on the trajectory up to the six-month's time-point.

Clinical outcomes analysis demonstrated that patients treated with SABR experienced a higher degree of Grade 1-2 complications, as defined by the Common Terminology

Criteria for Adverse Events-CTCAE (217)(Table 7-1). No statistical difference has been found in terms of 12 months mortality rates.

Table 7-1 Complications rates

	SABR	Surgery	
All complications	85 (89%)	95 (73%)	0.002
Minor complications	64 (67%)	69 (53%)	0.031
Major complications	21 (22%)	26 (20%)	0.7
90 day mortality	3 (3.2%)	7 (5.3%)	0.52
1 year mortality	14 (15%)	11 (8.4%)	0.14

[3]

7.4.1 Modelling for the prediction of postoperative complications

Variables were screened using univariate analysis. Variables with p value<0.2 at univariate analysis (Table 7-2) were entered in a stepwise logistic regression analysis (dependent variable: major complications).

[3] Complications were defined by CTCAE grades to be consistent in the two groups. Major Complications are defined those with a grade ≥ 3.

Table 7-2 SABR univariate analysis results for major complications

	With complications (21)	Without complications (74)	p-value
Age	75.5 (7.9)	74.0 (9.6)	0.62
Sex male	7 (33.3)	30 (40.5)	0.61
FEV1	76.0 (28.9)	76.8 (25.8)	0.95
DLCO	68.4 (22.5)	71.7 (22.1)	0.37
BMI	25.7 (5.2)	27.7 (5.4)	0.15
CAD	8 (38)	24 (32.4)	0.61
CVD	3(14)	9 (12.2)	0.72
Diabetes	6(28.6)	16 (21.6)	0.56
COPD	12(57.1)	32 (43.2)	0.32
CKD	1 (4.7)	6 (8.1)	1
CCI	2.3 (1.7)	2.0 (1.2)	0.63
PS>1	15 (71.4)	39 (52.7)	0.14
DSE	74.3 (22.6)	80.9 (23.3)	0.19
GH*	46.7 (16.9)	55.6 (24.8)	0.14
SumSc*	63.9 (20.4)	72.0 (18.2)	0.17
PF*	51.6 (19.1)	58.4 (24.2)	0.32
RF*	60.0 (27.3)	63.8 (32.8)	0.51
EF*	61.7 (34.8)	67.7 (28.6)	0.71
DY*	53.3 (24.6)	48.6 (32.9)	0.6

[4]

BMI, DSE, GH at baseline and high PS were entered in the logistic regression as their p value were <0.2. The morbidity rate in patients with PS>1 was 27% vs. 15% in those with PS = 0 or 1. No risk factor resulted associated with major complications after SABR at the logistic regression.

[4] *: for the baseline quality of life scales the number of available measurements in the SABR group was as follows: 15 in the complicated group and 59 in the non-complicated group. SumSc: Summary Score; GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; DY: dyspnoea; Functional scales: higher score represents good function. Symptoms scales: higher score represents worse symptom. DSE: decision-self efficacy score.

**: Continuous variable values are expressed as mean (SD); categorical variable as number (%)

Table 7-3 Surgical Univariate Analysis results for major complications

	With complications (26)	Without complications (104)	p-value
Age	69.4 (10.4)	70.2 (8.5)	0.71
Sex male	20 (76.9)	42 (40.4)	0.001
FEV1	77.9 (21.9)	90.5 (21.9)	0.013
DLCO	85.9 (28.8)	82.8 (18.8)	0.92
BMI	25.7 (6.2)	26.4 (5.0)	0.38
CAD	0	9 (8.6)	0.2
CVD	2 (7.7)	2 (1.9)	0.18
Diabetes	2 (7.7)	8 (7.7)	1
COPD	11 (42.3)	27 (25.9)	0.1
CKD	0	1 (0.9)	1
CCI	1.4 (1.1)	1.2 (1.0)	0.19
PS>1	5 (19.2)	16 (15.4)	0.77
DSE	83.6 (25.3)	84.8 (22.9)	0.96
GH*	66.7 (21.4)	72.3 (15.3)	0.56
SumSc*	86.1 (11.3)	82.3 (14.4)	0.45
PF*	83.8 (14.3)	81.8 (18.8)	0.99
RF*	82.1 (24.0)	83.9 (23.6)	0.83
EF*	72.6 (31.8)	73.9 (24.0)	0.81
DY*	23.8 (20.4)	26.8 (27.3)	0.89

[5]

In the Surgical group as shown in the Table 7-3, the following factors were screened for stepwise logistic regression: sex, FEV1, CAD, CVD, CCI, COPD.

Male sex and FEV1 remained significantly associated with major complications after surgery at the logistic regression.

[5] *: for the baseline quality of life scales the number of available measurements in the surgical group is as follows: 14 in the complicated group and 56 in the non-complicated group.

7.5 EORTC QLQ C-30 results

Due to the small sample size in each study arm the following results are presented as a descriptive analysis. The baseline clinical differences in these two populations make a direct comparison of QOL difficult as the effect of treatment on PROMS may have been masked by the underlying comorbidity or functional status. I tried to simultaneously follow the two trajectories with different methodologies in order to provide more clinical meaning to the results.

Before the surgical procedure the mean GH score was 71.2 vs 53.8 before SABR. This reflects the fact that the Lilac study is not RCT comparing the two treatments for the same clinical group, but is following two different populations, with markedly different clinical characteristics at baseline and through the first year after treatment. This is because the indication for SABR is for those patients who are not fit for surgery or less frequently decline the operation. Thus, SABR patients are typically those with the most compromise cardio-respiratory values. This is reflected in the lower baseline QOL for the SABR patients, especially in the global score.

The recently published general population normative data for the EORTC QLQ C-30 (218), show that the GH mean of 62.3 (SD 23.7) in UK, confirming the lower trend of the SABR population, also when compared to general population. Patients treated with VATS resection for example have a baseline GH higher then general UK population.

At baseline, SABR patients have consistently low scores in all the scales except for the Emotional functioning (Table 7-4).

Table 7-4 Baseline comparison of EORTC QLQ C-30 subscales between SABR Vs Surgery

BS	SumSc	GH	PF	RF	EF	FA	DY	CF	SF	PA
Surgery	83.1	71.2	82.2	83.6	73.6	23.7	26.2	83	83	11.4
SABR	70.3	53.8	57	63.1	66.4	43.5	49.5	73	63.5	25.5
p value	<0.0001	<0.0001	<0.0001	<0.0001	0.17	<0.0001	<0.0001	0.0277	0.0014	0.001

[6]

At six months, the differences in most of the scales are maintained, except for the Dyspnoea and Emotional Function where no statistical difference between the groups is observed (Table 7-5). This may highlight the major impact of the surgical treatment at six month time point.

Table 7-5 Six months comparison of EORTC QLQ C-30 subscales between SABR Vs Surgery

6m	SumSC	GH	PF	RF	EF	FA	DY	CF	SF	PA
Surgery	82.7	71.4	77.5	75.6	83.4	29	39.9	77.1	70.1	29.1
SABR	70.9	52.2	60	59.8	74.4	46.5	44.8	86.5	82.9	16.5
p vaue	0.0001	<0.0001	<0.0001	0.0015	0.063	<0.0001	0.31	0.0376	0.0079	0.0068

[7]

All the other QOL scores by group are shown in Table 7-6 and Table 7-7.

I have also performed a comparison between baselines and six months data within each group. Surgical patients experienced a statistically and clinically difference in the Emotional functioning (improvement) when comparing baseline to 6 months follow-up (Table 7-6). However, they showed a worsening in the dyspnea and insomnia which is statistically and clinically significant. The trend of dyspnea is in fact the expected one, with the worse value immediately after the resection, and a slight recovery which usually is stable or recover up to one year post resection agrees with other published data(86). However, from 6 weeks to 3 months the changes are not clinically, defined as

[6] *Results are expressed as means. SumSc: Summary Score; GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; FA: fatigue; DY: dyspnoea; SF: social functioning; PA: pain. Functional scales: higher score represent good function. Symptoms scales: higher score represents worse symptom.*

[7] *Results are expressed as means. SumSc: Summary Score; GH: global health status; PF: physical functioning; RF: role functioning; EF: emotional functioning; FA: fatigue; DY: dyspnoea; SF: social functioning; PA: pain. Functional scales: higher score represent good function. Symptoms scales: higher score represents worse symptom.*

a difference greater than less than 10 points, different. Also, the Fatigue scale worsened at six weeks of 17 points which recovered in the following months, showing the same trend of dyspnea but not reaching statistical significance.

Table 7-6 Surgery cumulative results of EORTC QLQ C-30 with comparison Baseline-Six months values.

Surgery	Baseline (70)	6 weeks (75)	3 months (76)	6 months (84)
SumSc	83.1 (13.9)	72.8 (19.3)	76.4 (18.6)	82.7 (14.5)^
GH	71.2 (16.6)	63.5 (19.3)	62.7 (22.1)	71.4 (20.3)^
PF	82.2 (17.9)	73.0 (21.3)	73.4 (23.7)	77.5 (21.6)^
RF	83.6 (23.5)	60.4 (32.4)	69.5 (28.1)	75.6 (27.2)a
EF	73.6 (25.5)	75.9 (25.1)	75.3 (26.8)	83.4 (20.8)b
CF	83.1 (18.5)	83.3 (21.6)	83.6 (20.8)	86.5 (19.1)c
SF	82.9 (21.4)	68.7 (28.1)	73.6 (28.6)	82.9 (24.3)^
FA	23.7 (19.5)	39.3 (25.8)	34.4 (27.1)	28.0 (23.2)^
NV	3.8 (10.7)	11.9 (20.0)	8.6 (17.3)	4.0 (8.8)^
PA	11.4 (19)	28.9 (28.5)	24.1 (28.9)	16.5 (24.1)^
DY	26.2 (25.9)	42.9 (32.6)	37.3 (28.8)	39.9 (30.5)d
SL	32.9 (33.8)	36.9 (32.9)	27.6 (29.0)	20.8 (26.9)e
AP	11.9 (23.9)	27.5 (34.2)	17.5 (25.8)	10.2 (19.4)^
CO	11.0 (20.0)	23.0 (28.6)	21.5 (28.7)	13.1 (21.3)^
DI	4.8 (14.2)	8.0 (19.6)	9.8 (18.8)	5.6 (13.5)^
FI	8.6 (21.7)	12.4 (26.1)	15.1 (29.1)	11.1 (23.9)^

[8]

When exploring the total means of all the data collected (Table 7-7), SABR patients had a clinically significant improvement in the Emotional Functioning too (but not statistical

[8] Wilcoxon matched-pairs signed-rank test used for comparison of 6 months vs. baseline values ^: p>0.2; a: p=0.12; b: p=0.004; c: p=0.07; d: p=0.007; e: p=0.03

Results are expressed as means and standard deviations. SumSc: Summary Score; GH: global health status; PF2: physical functioning; RF2: role functioning; EF: emotional functioning; CF: cognitive functioning; SF: social functioning; FA: fatigue; NV: nausea and vomiting; PA: pain; DY: dyspnoea; SL: insomnia; AL: appetite loss; CO: constipation; DI: diarrhoea; FI: financial difficulties. Scores are from 0 to 100. Functional scales: higher score represent good function. Symptoms scales: higher score represents worse symptom.

different p=0.09), and a statistically significant worsening in the Role Functioning (no>10 points). The Dyspnoea scores did not reach clinical nor statistical significance confirming that in this population the radiotherapy treatment has not affected the quality of life of patients reported outcomes in the first six months.

Table 7-7 SABR Cumulative results of EORTC QLQ C-30 with comparison Baseline-Six months values.

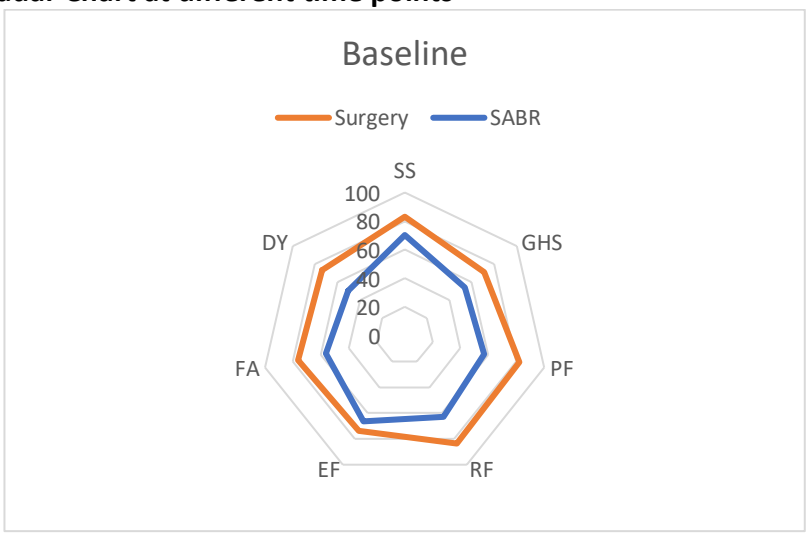
SABR	Baseline (74)	6 weeks (60)	3 months (61)	6 months (59)
SumSc	70.3 (18.8)	69.2 (19.6)	73.0 (16.6)	70.9 (18.8)^
GH	53.8 (23.6)	51.5 (22.7)	52.1 (24.2)	52.2 (22.0)^
PF	57.0 (23.3)	58.9 (24.4)	58.6 (24.7)	60.0 (27.0)^
RF	63.1 (31.6)	55.3 (33.7)	59.3 (32.6)	59.8 (31.4)a
EF	66.4 (29.8)	69.4 (27.5)	76.5 (25.3)	74.4 (27.6)b
CF	73.0 (26.7)	73.3 (25.3)	77.3 (24.0)	77.1 (23.7)c
SF	63.5 (35.7)	63.1 (35.3)	68.3 (29.0)	70.1 (31.9)^
FA	43.5 (26.0)	50.8 (29.0)	47.4 (26.0)	46.5 (23.5)^
NV	8.8 (15.9)	11.1 (24.3)	6.7 (12.7)	8.2 (16.9)^
PA	25.5 (29.2)	28.6 (32.3)	23.0 (29.2)	29.1 (33.3)^
DY	49.5 (31.3)	54.8 (32.0)	52.0(32.3)	44.8 (32.2)^
SL	33.8 (34.5)	36.1 (34.9)	32.8 (33.0)	37.0 (33.1)^
AP	23.4 (31.1)	29.9 (34.3)	23.3 (28.3)	28.6 (32.1)^
CO	20.3 (29.1)	21.7 (30.6)	17.2 (25.9)	21.4 (30.1)^
DI	8.6 (19.9)	7.8 (18.8)	6.6 (13.4)	11.1 (23.3)^
FI	13.2 (24.0)	10.6 (26.4)	10.0 (24.8)	7.5 (18.1)^

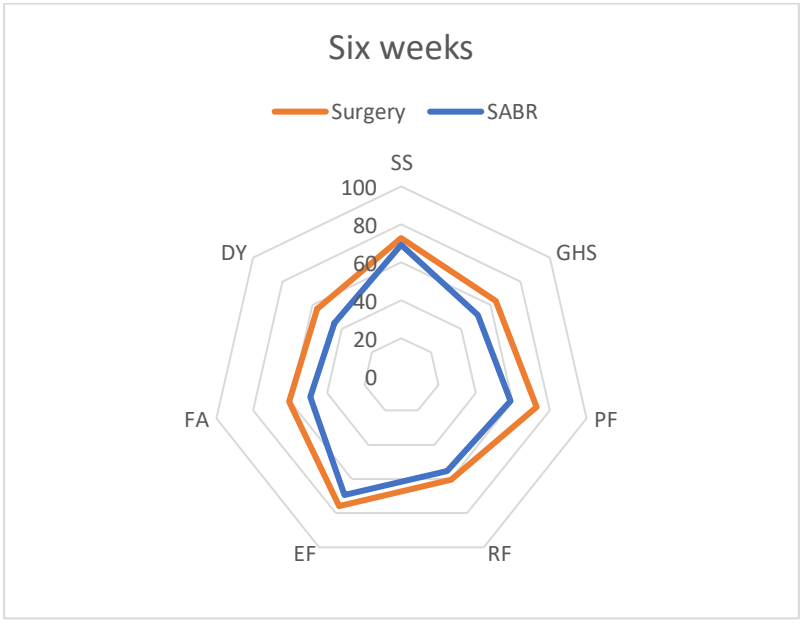
[9]

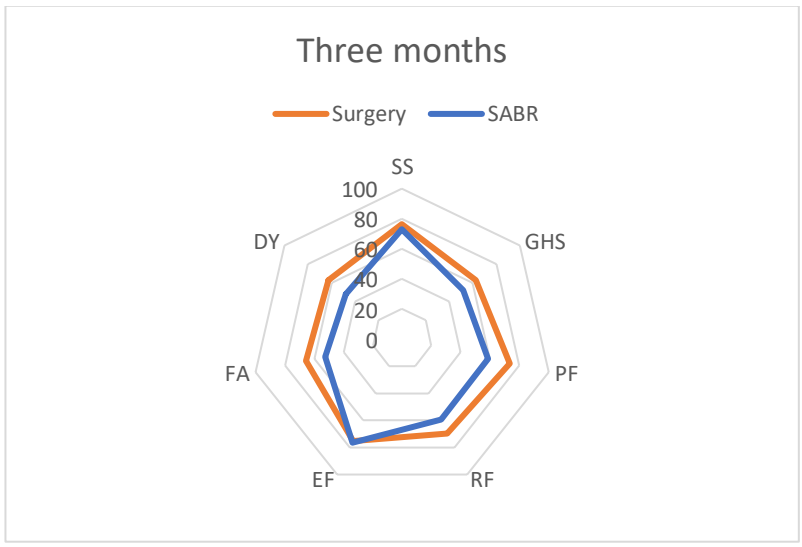
In the following radar charts (Figure 7-1), we can see the difference of subscales scores at all time points of the most relevant scales (symptoms scores are reversed for this purpose).

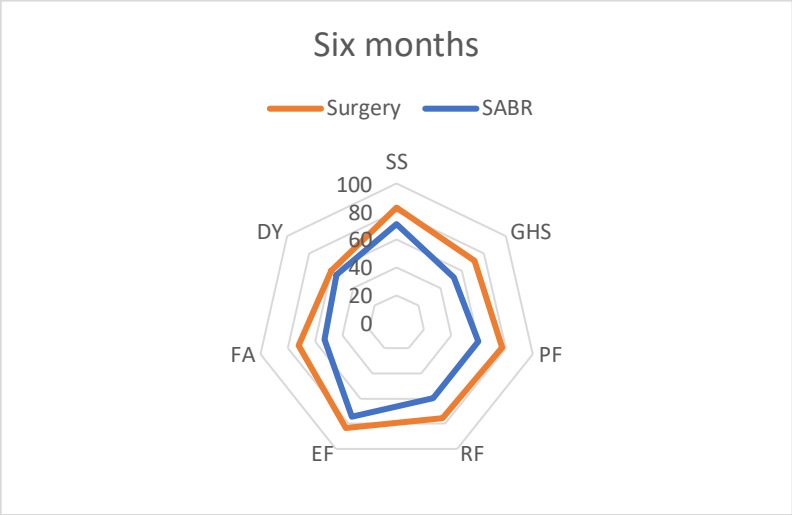
[9] Wilcoxon matched-pairs signed-rank test used for comparison of 6 months vs. baseline values. ^: p>0.2; a: p=0.016; b: p=0.09; c: p=0.14

Figure 7-1 Radar Chart at different time points



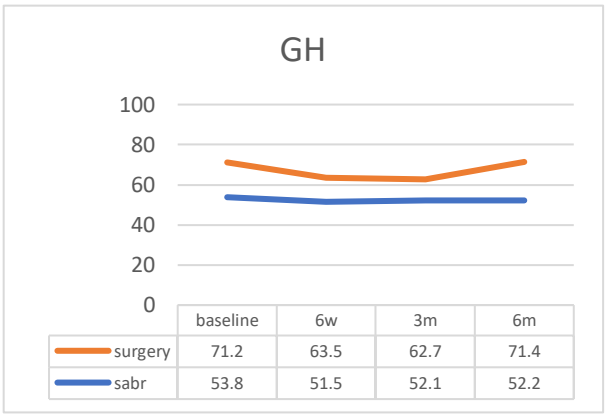


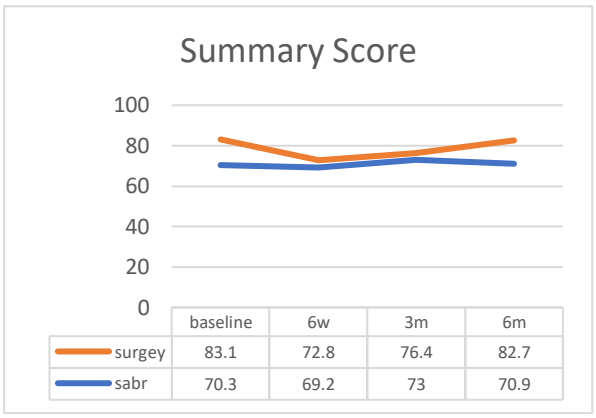


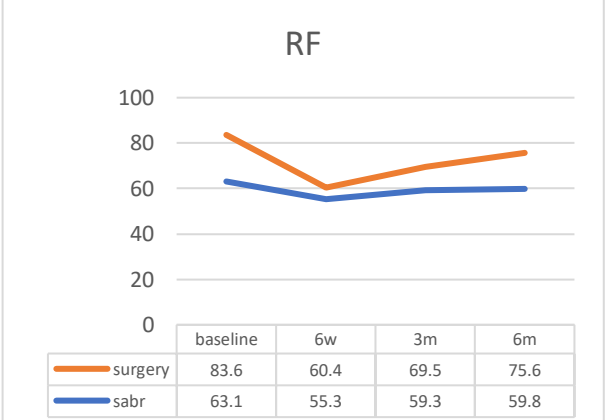


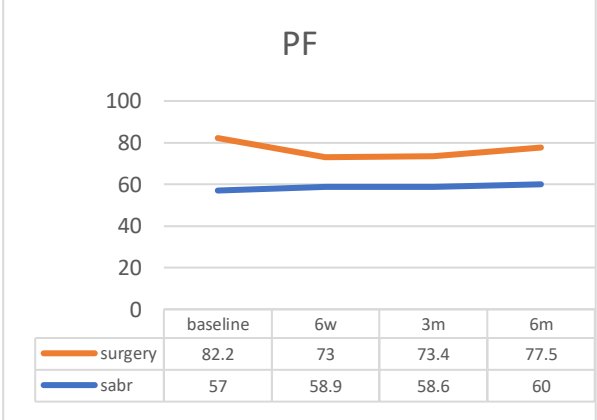
At six months most of the symptoms and functioning scores narrowed between the two groups, showing that surgical patients experienced in these first months a worsening in their QoL which made them closer to their less-fit counterparts. This is particularly evident in the single scales line charts (Figure 7-2) where the symptoms and functional scales being similar to each other at six months, although their baseline values were always different (especially Dyspnoea).

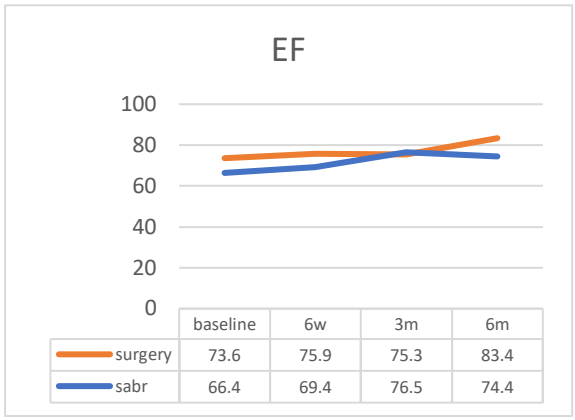
Figure 7-2 Line Graphs of the overall EORTC scales scores means

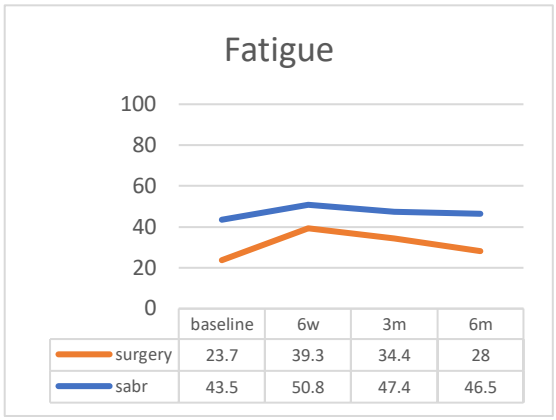


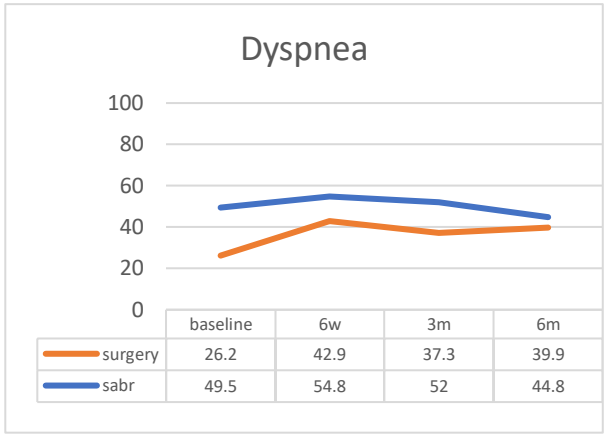






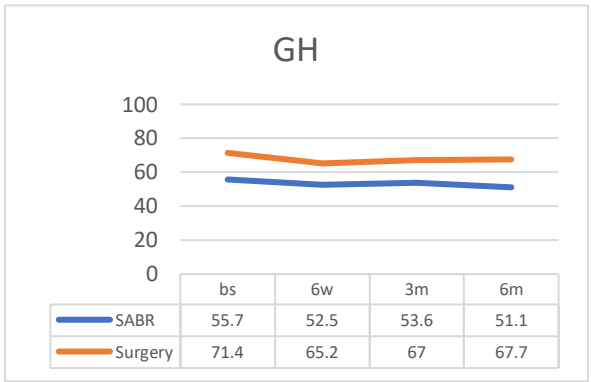


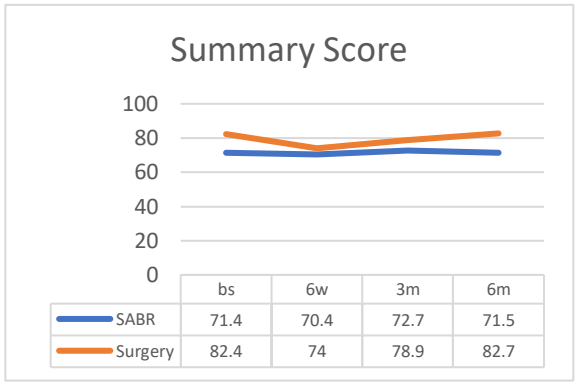


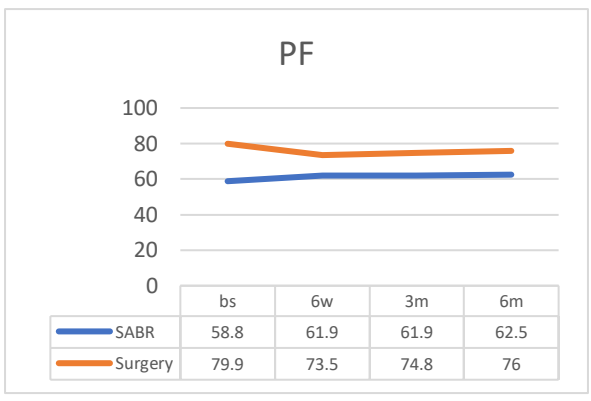


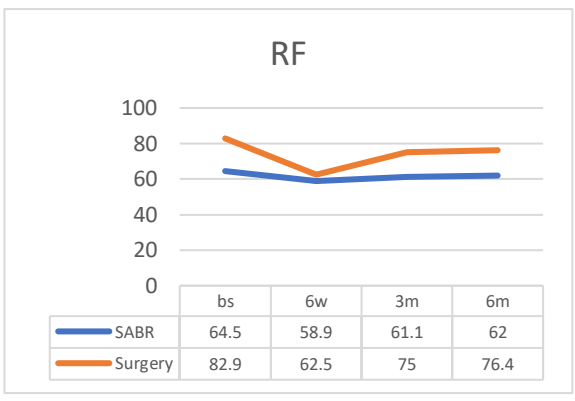
Same results were also showed when patients with all the four assesment has been selected (Figure 7-3).

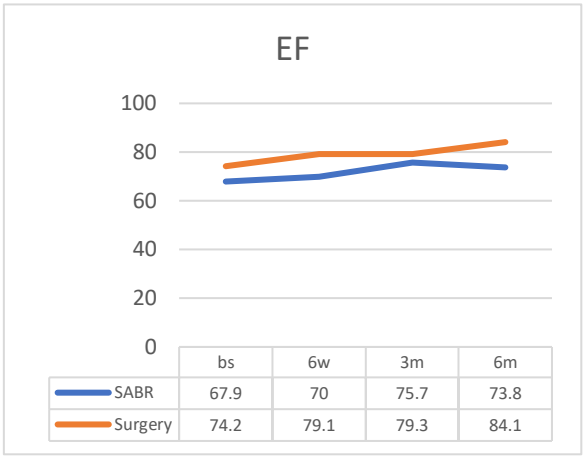
Figure 7-3 Line Graphs of EORTC scores of the 79 patients with all completion

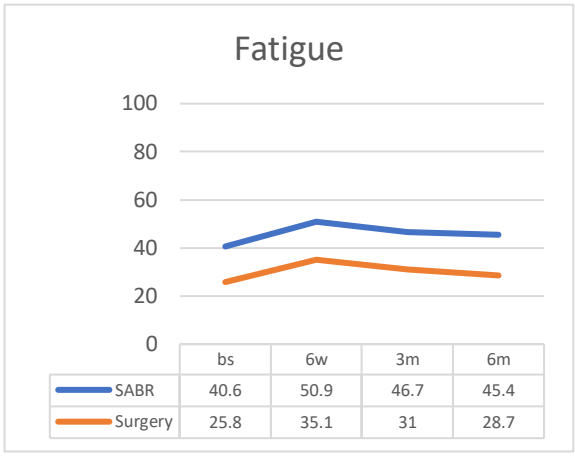


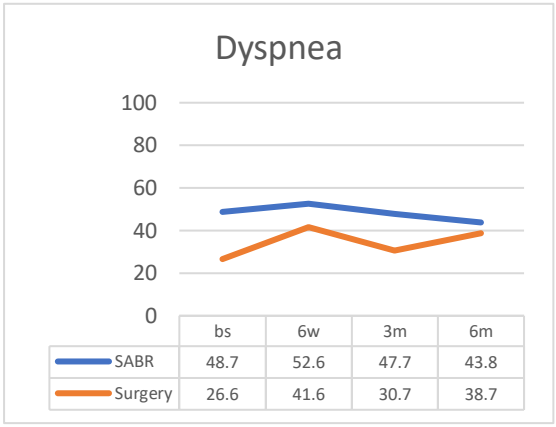












Surgical patients experienced a deterioration at 6 weeks in their general QOL particulary demonstrated by the SumSc (10.3 points) in the overall analysis. At six weeks the patients experienced a reduced value of Role Functioning after surgery which is accompained by a more marked increased in Dyspnea.

SABR patients in general were more stable over the first six months. However, they experienced a clinically significant increase in Fatigue at six months (from 40.6 to 50.9) which is similar to the surgical group (from 25.8 to 35.1). In both groups emotional functioning improved and at six months was reported higher than before the treatment. However, in surgical patients it reached both clinical and statistical significance.

As this analysis was not intended to directly compare directly the two treatments the following radar charts show the main results of each group highlighting the relatively stability of the SABR patients and the more marked changes after the surgery (Figure 7-4 and Figure 7-5).

In general, SABR patients maintained their QOL over the first 6months. The surgical operation impacted more on patient-reported outcomes: at six weeks they reported the worse scores but they returned the baseline in most of the scales except in dyspnea and Role Functioning.

For these radar charts symptoms scores are reversed (higher score means less symptom).

Figure 7-4 Radar Charts Showing evolution of EORTC subscales in SABR patients

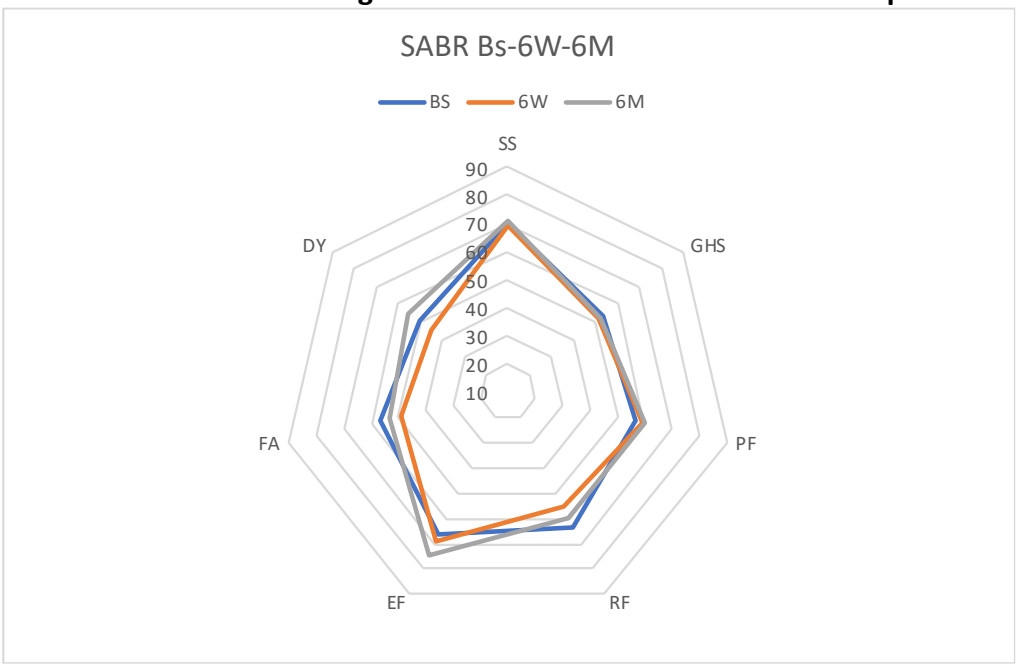
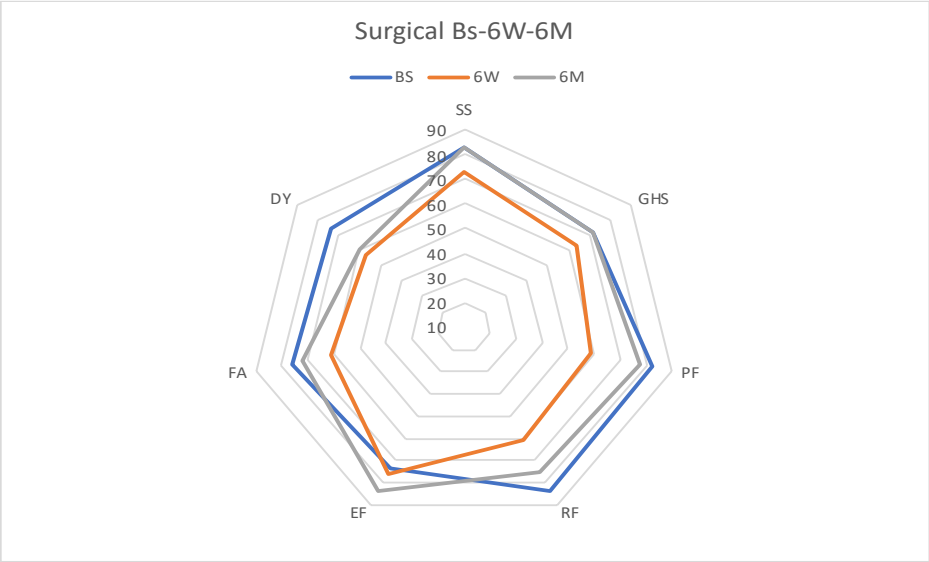


Figure 7-5 Radar Charts Showing evolution of EORTC subscales in Surgical patients



7.5.1 Responder Analysis on EORTC QLQ C-30 GH and SumSc

Guidelines from the FDA recommend that responder analysis can be applied to patient reported outcomes (PRO) measures at an individual level (defining a responder) rather than at treatment group level (212). We used FDA recommended methodology for defining a responder as an individual who reaches a particular threshold for clinically meaningful change over a pre-determined time period. This method will be particularly useful for a non-RCT to follow and compare, although in a non-randomized way, the evolution of QoL in two groups of patients with different clinical characteristics.

We used a method of analysis as described in the paper by Farrar and colleagues(210), to calculate the cumulative proportion of patients who reached a pre-defined response rate (responders) in change in EORTC Global QoL, calculated and displayed for all possible response rate cut-offs. Patient data were sorted in order of highest to lowest level of response and separated by treatment group. For each level of response, the proportion of patients that equal or exceed that level of response was calculated by

dividing the number of patients reaching that response by the total number of patients in the group.

We applied a distribution-based approach to define our responder thresholds for clinically meaningful deterioration, improvement or no change (stability) in Global QoL. A clinically meaningful response was defined as ½ of a standard deviation (SD) as is recommended for changes in health-related quality of life (219). The SD of the Global QoL change score from Baseline to Six months in both treatment groups combined was 20.70 and therefore a score of ½ SD=10.35 was deemed to be clinically meaningful. The same process has been applied to the SumSc analysis resulting with a 6.5 as clinical meaningfulness.

7.5.1.1 Results

The table below (Table 7-8) shows the distribution of the change in Global QoL scores (GH) and the proportion of patients showing Improvement (>10.35 change in Global QoL), no change/stability (10.35 to 10.35 change in Global QoL) or worsening/deterioration (<-10.35 change in Global QoL). This analysis includes 50 SABR patients and 51 Surgical patients as those are the patients with both baseline and 6 months completions.

Table 7-8 Proportion of SABR and Surgery patients reporting an Improvement (green), No change (amber) or Worsening/Deterioration (red) in EORTC Global QoL at 6 months after Baseline.

SABR				**Surgery**			
Change in EORTC Global score	Number of patients	Cumulative %	Proportion improved / No change / Worsened	Change in EORTC Global score	Number of patients	Cumulative %	Proportion improved / No change / Worsened
50	1	2.0		50	0	0.0	
33.33	2	6.0		33.33	4	7.8	
25	1	8.0		25	2	11.8	
16.67	8	24.0	24.0	16.67	7	25.5	25.5
8.33	5	34.0		8.33	6	37.3	
0	12	58.0		0	15	66.7	
-8.33	4	66.0	42.0	-8.33	5	76.5	51.0
-16.67	7	80.0		-16.67	8	92.2	
-25	1	82.0		-25	0	92.2	
-33.33	3	88.0		-33.33	3	98.0	
-41.67	3	94.0		-41.67	0	98.0	
-50	3	100.0	34.0	-50	1	100.0	23.5
SD of change Global QoL for **Surgery** patients from BL to SixMonth = 18.14 so 9.07 should be considered clinically significant change for the **Surgery** group							
SD of change Global QoL for **SABR** patients from BL to SixMonth = 22.91 so 11.46 should be considered clinically significant for the **SABR** group							
SD of change Global QoL for BOTH Surgery+SABR from BL to SixMonths = 20.70 so 10.35 should be considered clinically significant for both groups.							

The same analysis was performed on 41 SABR patients and 47 Surgical patients with complete data to calculate the Summary Score.

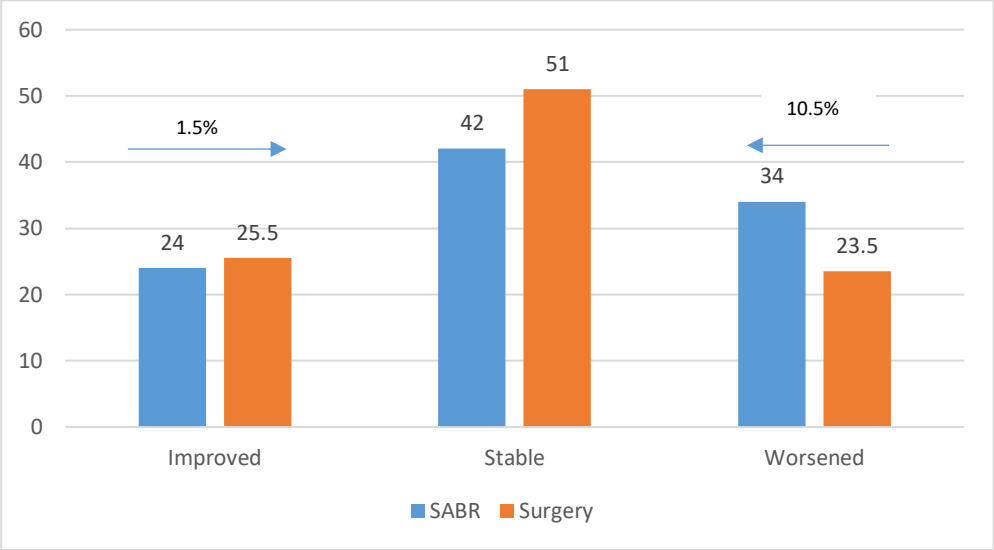
The difference between the two groups represents the absolute risk reduction (ARR) which can be calculated for any chosen cut-off point (210).

To calculate the ARR of the non-responders (i.e. clinically meaningful deterioration), we subtracted the proportion of SABR patients who reported a deterioration in Global QoL (34%) from the proportion of Surgery patients reporting a deterioration (23.5%). Therefore, the reduction in risk that a surgical patient (as compared to a SABR patient) will report a clinically meaningful worsening/deterioration of Global QoL at 6 months is 10.5%.

Similarly, the ARR can be calculated for the responders (clinically meaningful improvement). We subtracted the proportion of Surgery patients who reported an improvement in Global QoL (25.5%) from the proportion of SABR patients reporting an improvement (24%). Therefore, the reduction in risk that a SABR patient (as compared to a surgical patient) will report a clinically meaningful improvement of Global QoL at 6 months is 1.5%. This may be alternatively stated as 1.5% *increase in likelihood* that a

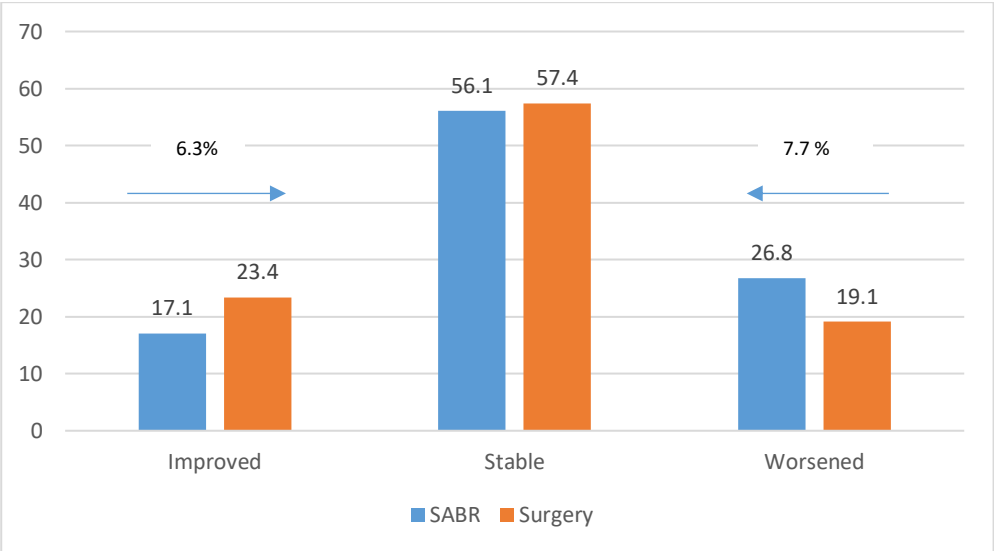
surgical patient (as compared to a SABR patient) will report a clinically meaningful improvement in Global QoL at 6 months (Figure 7-6). To summarise, in the SABR group, ¼ of patients will improve, 42% remain stable and 34% deteriorate. In the Surgical group, ¼ will improve, ½ remain stable and ¼ worse (Figure 7-6).

Figure 7-6 Proportion of patients reporting a clinically meaningful improvement in Global QoL at 6 months (n=101).



We have also repeated the same analysis for the Summary Score (Figure 7-7).

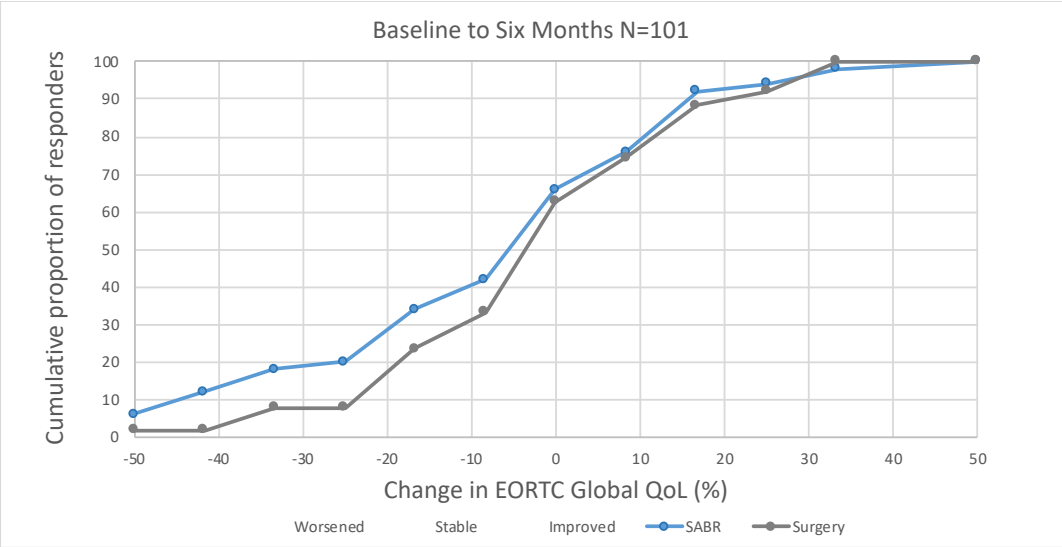
Figure 7-7 Proportion of patients reporting a clinically meaningful improvement in Summary Score at 6 months (n=88).

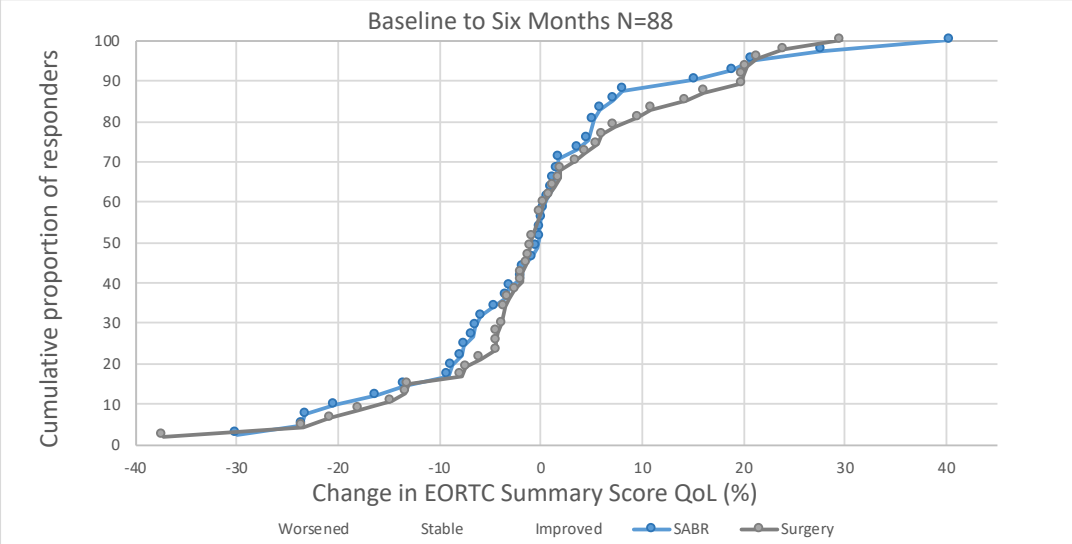


With the SumSc the changes at 6 months were similar, however a larger proportion of surgical patient improved their score (6.3% SumSc Vs 1.5% Global Score). Both scores reported a similar percentage of stable patients in their first 6 months after treatment.

The Cumulative Proportion of Responders Analysis (CPRA) graph was produced by placing every possible responder level on the χ-axis and plotting the associated proportion of patients who reached that response level on the y-axis (Figure 7-8). This type of graph is helpful for graphically following the single patient and plots the groups.

Figure 7-8 CPRA graphs showing clinically meaningful Worsening/Deterioration (red), No change (amber), and Improvement (green) in Global QoL at 6 months and in Summary Score.





The difference between the groups on a CPRA graph represents the absolute risk reduction (ARR)(210). It should be noted that the ARR is not stable across all the cut-off points on the CPRA graph above.

The Global QOL graph confirms a major proportion of SABR patients clinically deteriorating at 6M with no skewed patients in both sides.

Instead, there is crossover of the two groups in the SumScore plot meaning the SABR group are reporting the most improvement in Global QOL at 6 months, despite being

less likely of improvement overall. In other words, those SABR patients who improved were more likely to have a greater perceived gain in QoL captured by this scale.

We have also explored the use of a waterfall plot to present each individual patient's response to QoL. The horizontal (x) axis across the plot may represent the no change value; vertical bars are drawn for each patient, either above or below the baseline. The vertical (y) axis may be used to measure maximum percent change from baseline. Those vertical bars that are above the line represent "responders", those who had an improvement in their quality of life scores. Vertical bars below the baseline (x) axis are drawn for each patient that has achieved some degree of quality of life reduction, depicted as negative percent.

The Waterfall plots below (Figure 7-9) are showing changes from baseline to 6M (A) in Global QoL in SABR patients, (B) Global QoL in Surgical patients, (C) in SumSc in SABR patients, and (D) in SumSc in Surgical patients. Only patients with a baseline and 6M assessment are included.

Figure 7-9 Waterfall plots of Baseline to six months changes in Global QoL and SumSc (green=improvement, yellow=stable, red=worse).

In conclusion the responder analysis demonstrated that at 6months, surgical patients report clinically improved overall QOL evaluated through GH and SumSc. Also, SABR patients reported a higher degree of improvement. Similar percentages of patients remained stable (not reaching the clinically meaningful difference in change) in both scores.

Interestingly, more SABR patient deteriorated at 6months however, the surgical patients reported the highest degree of deterioration captured by the SumScore and not by the Global QoL.

As Dyspnoea has been found in the analysis of cumulative changes to be an influential symptom in the surgical group in the first six months, we have repeated the responder analysis for this scale (data in the Appendix). The cumulative difference may be

explained clinically, with more SABR patients improving at six weeks (10% more than surgical patients) and relatively similar percentage of patients deteriorating (difference of 2.7%). The difference between baseline and six months instead, is more marked: 20.1% more SABR patients improved, and 11.8% more Surgical patients clinically worsened in their reported dyspnea.

7.6 EORTC QLQ C-30 Modelling

I have focused the analysis the Global QoL score and the SumSc for the purpose of this logistic regression analysis. I wanted to explore possible factors associated with the evolution of quality of life at six months. The Results of the Panel regression Analysis for the SumSc and GH score are shown in the following tables (Table 7-9 and Table 7-10).

Table 7-9 Results of random effects time-series cross-sectional regression analysis (dependent variable: Summary Score)

Variable	coefficients	SE	p-value
Surgical			
age	0.38	0.16	0.016
CKD	35.1	14.8	0.018
Diabetes	-21.1	5.5	<0.0001
SABR			
Diabetes	-9.5	4.1	0.02
High PS	-13.4	3.4	<0.0001

[10]

The following variables were initially entered in the model (age, FEV1, DLCO, sex, CAD, CVD, CKD, Diabetes, PS>1, IMD, BMI, CCI).

[10] CKD: Chronic Kidney Disease. PS: Performance Score. High PS is defined>1.

In the SABR group factors which remained associated with SumSc were high PS and the presence of diabetes. In the Surgical group instead, the PS lost significance. When the SumScore is the dependent variable and only associated factors were the presence of diabetes, chronic kidney disease and age. The diabetes variable may be not considered as based on limited number of observations.

Table 7-10 Results of random effects time-series cross-sectional regression analysis (dependent variable: GH)

Variable	coefficients	SE	p-value
Surgical			
FEV1	0.16	0.56	0.003
DLCO	0.14	0.61	0.021
High PS	-7.78	4.01	0.008
Diabetes	-8.2	4	0.04
SABR			
Male	-6.5	3.9	0.1
High PS	-11.4	3.7	0.003
CVD	13.5	6.1	0.028
BMI	0.062	0.33	0.057

However, when looking at the Global Heath Score, FEV1, DLCO, diabetes and high PS showed a statistically significant association (Table 7-10). In the SABR population, factors influencing the changes in Global Health score were high PS and the presence of cerebro-vascular disease (CVD). The main difference between the two groups is the FEV1 (expression of pulmonary function) was never associated to the evolution of the other QoL scales in the SABR group but in most of the scales in the surgical patients (data not shown). This data confirms the relative difference in the respiratory function of the two groups. SABR patients were overall below certain values where probably difference in FEV1 scores doesn't impact on their QOL after the treatment. Surgical patients instead are experiencing a significant negative change in QOL after the surgical intervention if their preoperative FEV1 was lower.

PS was in both groups sensitive to the QOL change, stressing the importance of patient's relative factors like independence not easily captured by more objective parameters.
The two models aiming to identify predictors of six months QOL showed in both cases that only the baseline QOL value was significantly associated both in the univariate and multivariate analyses.

7.7 Patients satisfaction results

Patients' satisfaction was only assessed at the first post-treatment time-point (6 weeks). The amount of time post-operatively and post-admission is useful in giving people pause to heal and reflect on their admission more objectively. This was done also to maximize response rate, as suggested also by Bredart (92). Moreover, assessing satisfaction close to hospital recovery could allow for a better distinction among elements of satisfaction and higher response variability (220).
Patient satisfaction was completed by 60 SABR (66.6%) and 74 surgical patients (62.7%). As already discussed, this low response rate at this time-point is explained by the fact that in this period patients in both groups are experiencing the highest level of post-treatment symptoms. The QoL results are in fact lower at 6 weeks in both groups. We did not find any difference between the two groups in all the subscales of the PS-18 questionnaires. In all the scales patients reported moderate level of satisfaction with the care provided (Table 7-11).

Table 7-11 Patient satisfaction results according to the treatment.

PS18 scales	SABR (n:60)	Surgery (n:74)	p value
General satisfaction (mean, SD)	3.82 (1)	3.89 (1)	0.6
Technical quality	3.88 (0.8)	3.95 (0.6)	0.81
Interpersonal manner	4.05 (0.99)	4.37 (0.6)	0.11
Communication	3.95 (0.93)	3.96 (0.81)	0.7
Financial aspects	4 (0.98)	4.08 (0.92)	0.7
Time Spent with Doctor	3.78 (0.9)	3.85 (0.9)	0.55
Accessibility and Convenience	3.44 (0.7)	3.47 (0.8)	0.62

[11]

The comparison between known groups general satisfaction showed that patients living in more deprived areas (higher IMD scores across) were more satisfied. Patients experiencing minor and major complications were also more satisfied after 6 weeks from treatment (Table 7-12). I have categorized the continuous variables splitting the group with the mean value.

Table 7-12 Patient Satisfaction known groups comparison

IMD>4.6	IMD<4.6	p value
4.02 (0.9)	3.66 (1)	0.02
Employment domain >4.5	Employment domain <4.5	
4 (1)	3.71 (1)	0.06
Education domain >4.6	Education domain <4.6	
4.03 (0.9)	3.66 (1)	0.02
Health and Disability domain>4	Health and Disability domain <4	
4.06 (1)	3.66 (1)	0.01
Living Enviroment domain >4.4	Living Enviroment domain <4.4	
3.72 (1.1)	4.03 (0.86)	0.16
Crime domain >4.6	Crime domain <4.6	
3.96 (1)	3.76 (1)	0.16
Age>72	Age<72	
3.8 (1)	3.89 (0.9)	0.85
PS>1	PS<1	
3.7 (1)	3.9 (0.9)	0.24
major complications Y	N	
4.06 (0.97)	3.72 (1)	0.04
minor compl Y	N	
4.06 (0.99)	3.72 (1)	0.03
female	male	
3.83 (1)	3.88 (1)	0.64

[12]

[11] Patient satisfaction score are ranging 1-5. Higher score corresponds to high grade of satisfaction.

[12] Groups have been categorized with mean. **IMD**: Index of multiple deprivation decile: decile 1 fall within the least deprived are, whereas the decile 10 is the most deprived area and this applies to all the sub-domains: **Employment Deprivation** Domain measures the proportion of the working-age population in an area involuntarily excluded from the labour market. **Education Skills and Training Deprivation Domain** - The domain measures the lack of attainment and skills in the local population. **Health Deprivation and Disability** Domain measures the risk of

7.7.1 Patient Satisfaction model

Of the 225 participants, 91 patients failed to complete a Patient Satisfaction questionnaire at the Six-week time point and thus were excluded from the analysis. Of the 134 remaining patients, 26 had failed to complete the EORTC QLQ-C30 at Baseline and so these patients' data were also excluded from the analysis. The univariate analysis of Patient Satisfaction was therefore carried out on the remaining 108 patients. The multivariable analysis excluded a further 6 patients who had not completed a DSE. Results of the regression are shown in Table 7-13.

Table 7-13 Regression output of the two pat sat models.

Variable	Total SABR participants (*N* = 48) (%)		Coef	*p*-value
Baseline QoL	48	(100)	0.005	0.47
Age (years)	48	(100)	0.01	0.40
Gender				
Female	31	(65)		
Male	17	(35)	-0.03	0.92
DSE	48	(100)	0.02	0.003*
FEV1P	48	(100)	0.007	0.26

* Statistical significance at <0.01

Variable	Total Surgery participants (N = 54) (%)		Coef	p-value
Baseline QoL	54	(100)	0.01	0.29
Age (years)	54	(100)	0.01	0.54
Gender				
Female	29	(54)		
Male	25	(46)	-0.09	0.76
DSE	54	(100)	0.008	0.20
FEV1P	54	(100)	-0.001	0.89

* Statistical significance at <0.01

premature death and the impairment of quality of life through poor physical or mental health **The Living Environment Deprivation Domain** measures the quality of the local environment. **Crime** Domain measures the risk of personal and material victimisation at local level.

Multivariable regression analysis showed that the only preoperative factor that remained independently associated with patient satisfaction was the efficacy in making decision in the SABR group. Patient reporting more efficacy during the decision-making period before radiotherapy, were also the more satisfied with the care provided. No predictive variables were found in the surgical group.

7.8 Discussion

This study longitudinally and consistently collected QOL information with validated cancer specific questionnaires in an acceptable number of SABR and Surgical treated early stage NSCLC patients. In addition, by collecting a range of clinical variables, we have been able to explore associations with patient reported outcomes.

QOL data confirmed the expectation that patients selected for SABR have more comorbidities, worse PS and also worse baseline QOL. However, the treatment doesn't impact markedly on their QOL over the first six months after treatment.

Surgical patients , although reporting higher levels of QOL preoperatively, have a major impact, especially on respiratory symptoms immediately after the operation.

In both groups, Emotional functioning confirmed findings from the wider the literature and supersedes the pre-treatment values.

A more detailed discussion on PROMS analysis and the statistical models follow.

7.8.1 PROMS

Quality of life at baseline differs substantially in these two groups: SABR patients reported statistically significant lower scores in all the functioning scales and higher scores in the symptoms items. Only Emotional Functioning did not reach statistical significance difference between the two groups. When looking at 6months, the

emotional functioning still remains no different but Dyspnoea also lost its significance, meaning that surgical patients showed a major increase in this symptom.

Our data showed a substantial stable trend in the quality of life of SABR patient. The only statistical, but not clinical, difference was related to Role Functioning. So, although their baseline quality of life was lower, as expected by their physical characteristics, the treatment did not clearly affect their quality of life in the first 6 months (25% of patients will improve, 42% remain stable and 34% deteriorate). Early post-treatment Qol scores indicate that SABR is a well-tolerated treatment for patients with early stage NSCLC.

Only two studies reported clinically significant deteriorations (141, 144), with one reporting a deterioration in fatigue and the other in dyspnoea which are the two symptoms where we have also found the major change, albeit not statistically different after 6 months. Because many of these NSCLC patients have coexisting chronic obstructive pulmonary disease, changes in dyspnoea scores after SABR appear to be in keeping with the natural history of this disease, suggesting that SABR might only have had a minor contribution.

Surgical patients instead had reported a more marked effect on their quality of life (¼improved, ½ remained stable and ¼ worsened). This effect was particularly evident at the first time point (6weeks) which is where patients can be struggling with the postoperative pain or other complications.

The reduction in Role functioning at six weeks may be explained by the clear difficulties of surgical patients in returning to normal social life and activities before their first outpatient appointment with the surgeon (usually at 6-7 weeks from discharge). We have reported a statistical difference in Role Functioning also in the SABR patients, but without reaching clinical significance. Of note the surgical and SABR patients scores became more similar at 6weeks: this can be explained as most of the SABR patients are

already less fit with restricted social commitments due to this disease or coexisting comorbidities.

Dyspnoea and fatigue are also the two symptoms that at 6 weeks are increasing more in surgical patients compare to SABR. Interestingly, the patient reported dyspnoea improved at 6months in both groups, but in SABR patients returned to lower than baseline, and remained higher than baseline in surgical patients. This supports the acute effect of a surgical resection on patients' respiratory function.

Emotional functioning, as described in other studies (64, 86, 221) and shown in the CH 3, improved in both groups, reaching the statistical and clinical significance in the patients who underwent resection.

Although we have not yet analysed the results of the Lung Cancer Specific questionnaire, but we are not expecting other major differences in the Qol evolution as no other scale would be able to capture symptoms specific to these two populations.

Interestingly, the SumSc showed a different trend in both groups, which was more evident when only patients with all the questionnaire data were analysed. In fact, in patients submitted to surgical resection, the data indicated a sudden decrease which recovered during the other time points and returning to baseline after six months. This was not detected by the GH, highlighting once again, that the SumSc appears more sensitive to all the changes across domains and more comprehensively captures quality of life as a whole. This confirms the results illustrated in Ch 6 and in our previously published paper (178). One possible explanation of these differences may be in the construct of these two scales: if the SumSc covers a range of symptoms and functions, the GH is reflecting more a patient judgment of their overall QOL.

Patients satisfaction results did not show marked difference between the groups. However, patients who experienced minor or major complication reported higher

scores of patient satisfaction. This is in agreement with Barlesi and colleagues who in similar thoracic surgical patients(222) found that the absence of postoperative complications expressed as a post-operative quality score was the only quality index that showed weak but significant correlation (r = 0.3) with an index, rating the satisfaction with the structure (i.e. room arrangement, food, waiting time).

However, the Lilac results regarding patient satisfaction and post-treatment morbidity are discordant from what we have previously published in lung resection patients (223) and in other studies using the EORTC InPATSAT32 (91). It could certainly be argued that some of the dissatisfaction caused by adverse events after surgery may not be alleviated by clear and informative communication or the intensive care that complicated patients are supposed to receive. However, it is not easy to put these results in the context of studies exploring possible correlation between socio-economic level and patient's satisfaction. Comparability across studies is hampered by heterogeneous reporting and differences in patient's population demographics. A review of previous studies on lung cancer patients show that socio-economic factors influence the use of health care services and subsequent cancer survival rates. In previous reports from our centre, there does not appear to be a significant relationship between socio-economic status and stage, performance status or outcome from lung cancer(224). Results from the UK National Cancer Patient Experience Survey (2012) showed that the significant differences that exist among IMD and satisfaction of care are not uni-dimensional(225): however, there was a certain degree of consistency about the kinds of questions which were less highly ranked by patients in the most deprived areas as identified by the Index of Multiple Deprivation (IMD) quintile, with most of the items relating to information giving and understanding. Probably there will be some other factor that in our population which has helped most deprived area to have patients more generically

satisfied with care. Future analysis comparing our results with more updated data from the 2017 National Cancer Patient Experience Survey may help to depict these differences and help commissioners to reduce inequalities.

7.8.2 Prediction of outcomes

Our results show that in this group of patients' preoperative quality of life was not associated with the occurrence of post-treatment major complications. In the surgical group alone, male gender and lower FEV1 were identified as predictors of major complications. This may again be related to the fact that surgical patients have better respiratory function in general but lower values are sensibly affecting the postoperative clinical outcomes. In a recent analysis on the ESTS database on 62774 lung resection patients showed that factors influencing morbidity were: type of resection, age, presence of cardiac comorbidity, ppoFEV1 and surgical approach(198). Many studies have shown the association between FEV1 or its derivate, predicted postoperative FEV1 (ppoFEV1), and surgical risk (226-228). In particular the risk of pulmonary morbidity and mortality has been shown to increase when FEV1 is below 50-60% or ppoFEV1 <30% as reported in the most updated preoperative evaluation guidelines (45). PpoFEV1 and male gender are two of the factors included in the Eurolung 1 risk score(48, 53).

Nevertheless, in our analysis, the definition of complications in the surgical group was through the CTCAE grading, in order to be consistent with the radiotherapy published literature. This may explain the difference of results compared to our previously published paper described in Ch 4(155).

Male gender has also been defined as risk factor for postoperative morbidity by the Thoracoscore, one of the most important preoperative risk scores used so far in our speciality(46). Furthermore, an analysis from the American STS multi-institutional

database(229) has recently demonstrated that women have lower postoperative morbidity and mortality after lung cancer surgery.

The other important finding is that performance score was the preoperative factor mostly associated to the evolution of quality of life at six months across groups. This was captured for SABR patients in both the overall QoL scores (GH and SumSc)but only in the GH analysis for the Surgery group.. Patients with less independence at the baseline assessments were those at high risk of worse quality of life. Performance status, which is a clinician-assigned score, is extensively used in oncology research and practice. Compared with HRQOL, there is more literature on the relationship between performance status and outcomes. Performance status has been shown to be associated with short andlong-term survival outcomes after cancer treatment. However, the majority of studies have been in advanced, metastatic cancer patients in palliative care(230, 231). These results suggested that PS may be also used to stratify patients (in particular SABR) when discussing the prediction of their post-treatment QoL. The fact that the overall models with the six-month QoL as outcome measures, have not identified other predictors except the baseline QoL, is not surprising. It has instead highlighted the importance of the correct use of longitudinal data regression analysis and the limits of the cross-sectional studies.

We hypothesized that there is an association between efficacy in making decision for their lung cancer treatment and the post-treatment satisfaction. The Lilac data supporting this and highlighted in the SABR patients an association with more positive preoperative DSE scores and post-treatment satisfaction. This was not replicated in the surgical group. These results may be also be affected by the specific time point where we have decided to assess satisfaction (6weeks). The attrition was higher at this point, especially in the surgical group, but we decided to assess satisfaction closer to hospital

recovery to allow for a better distinction among elements of satisfaction and response variability (220). The decision self-efficacy is not a measure of the quality of the decision-making process, however it has given for the first-time provided information of patient attitude surrounding treatment choices. Understanding the role of post-treatment PROMS and their association with decision-making factors and patient's satisfaction may give important information not only for clinical practice but also to future research.

7.9 Limitations

The major limitation of this prospective study is the fact that this is not a RCT. The two populations are completely different from a clinical point of view. However, the aim of our study was not to compare them, but rather to follow these patients longitudinally and get information of the QoL evolution to be discussed during their clinical consultations. Sensitivity analysis on the QoL data will be warranted to explore possible bias due to attrition during the different time-points.

Another important limitation of this part of my thesis is that the use of the EORTC QLQ C-30 and the LC-13 has still the barrier of being designed to explore QoL after systemic therapy. Several questions are too generic or not relevant for the population included in this study as confirmed during the qualitative work. Furthermore, the LC-13 version does not include some surgery-specific items (i.e. postoperative pain; postoperative functional recovery; wound cosmetic) that are instead inserted in the updated version(154). More treatment-specific questionnaires would increase reliability of QoL results in these patients.

It is important to note that a certain percentage of patients die during surgical treatment in hospital unlike SABR in which death is highly unlikely during the first months as

confirmed in our study; this may have biased the sample, since data from the sickest, most at-risk surgical patients who die during surgery would not have been included in the study.

Our methodology focussed only on short-term follow-up of patients' service evaluation. The amount of time post-treatment and post-admission is useful in giving people pause to heal and reflect on their admission more objectively. A comparative evaluation at six-months to one-year post-treatment may reduce social desirability bias, a confounding variable which the participants of this study could have experienced during their episode of care. Furthermore, the PS-18 questionnaire is designed to assess generic satisfaction without specifying if it was referring to inpatient or outpatient care. I have chosen this questionnaire instead of the most common EORTC PATSAT-32 as SABR is an outpatient treatment. We cannot rule out that this questionnaire may not be capturing the same domains in these two different groups.

7.10 Conclusions

In conclusion, we were able to depict the trajectory of QOL after 6 months from radical treatment for early stage NSCLC. This information, together with the longer follow-up data, will support the clinicians in discussing treatment alternatives.

These preliminary results will inform future research to possibly include PROMS in patients' selection algorithms or risk scores.

Chapter 8 Conclusions, discussions and future directions

8.1 Conclusions

This PhD had two overall main objectives:

- To explore the QOL evolution after early-stage NSCLC treatments and assess the relationship between patient reported outcome measures (PROMS) and clinical factors.
- To evaluate the feasibility and acceptability of routine PROMS collection within clinical practice in early stage NSCLC radically treated.

In order to test these objectives, important methodological and practice issues were addressed:

- Selection of the optimal PROM to effectively measure QOL through a systematic review of the available literature and preferred time-points to better capture the changes in QOL in early-stage NSCLC patients treated with radical intent.
- Finalise, collect, collate and integrate the quality of life data from an audit of a prospectively maintained surgical clinical database in order to explore associations with these data and patient perioperative details.
- Set-up a prospective study comparing two group of patients receiving different treatments (surgery or stereotactic ablative radiotherapy) to obtain QOL and PROMS data and perform analysis.

- Set up of electronic integration of PROMS data into individual patients' electronic health records (EHR) for use in clinical practice and a reminder system to monitor patient questionnaire collection on the prospective study.

The studies reported in this thesis were carried out over a three-year period. This chapter discusses and synthesises the key findings.

8.1.1 Chapter 3: Systematic and Literature reviews

The results of the systematic review highlighted the lack of current QOL reporting before and after radical treatment for early-stage NSCLC. Only one small RCT was identified that provided evidence exploring a direct comparison of QOL outcomes following SABR and VATS lung resection for lung cancer. The remaining 16 studies evaluated QOL in a single speciality and were characterized by small sample sizes and limited follow-up. These studies assessed a total 832 SABR patients and 686 receiving anatomical VATS resections, and confirmed that in general the physical components of QOL worsen immediately after treatment for up to 3-months, returning to baseline after 1 year. This was a consistent feature in the surgical papers and SABR studies demonstrated more stability across QOL subscales during the first year. Emotional functioning supersedes the pre-operative values across treatments. The most commonly used QoL instrument was the cancer specific EORTC QOL C-30, which in a few studies had been administered with the Lung Cancer specific module. These two instruments were taken forward to the prospective observational study where their content was analysed quantitatively and also through interviews with patients.

8.1.2 Chapter 4: Quality of life and risk prediction

This Chapter explored the role of patient-reported quality of life in the context of preoperative selection of patients for NSCLC surgical resection. In a cohort of 330 patients submitted to anatomical lung resections in a single centre, we demonstrated that preoperative QOL predicts the occurrence of a prolonged length of inpatient stay. In particular, poor Role Functioning (RF) remained significantly associated with prolonged hospital stay after adjusting the analysis for other potential confounders. This concept is an important addition to the previous discussion on complications as it provides evidence that length of postoperative stay is not influenced only by postoperative events.

100 patients had RF lower than 67. Thirty-one of them experienced a prolonged hospital stay 31/100=30% (vs. 18% of those with a higher value, p=0.008). We demonstrated that when the analysis is limited to VATS, preoperative RF remained significantly worse in patients with the prolonged length of stay. Nevertheless, the Physical Functioning was also an independent factor (p=0.018) significantly associated with prolonged stay along with low DLCO (p=0.008), history of stroke (p=0.005) and low FEV1 (p=0.07). For the VATS procedure, which is considered less invasive and indicated for less complex and extended tumours, the physical component of QoL and other clinical factors are more useful predictors of postoperative length of stay.

This chapter explored QOL in different classes of mortality risk identified by the Eurolung Risk Score. Patients considered by their clinical characteristics to be at higher risk of death after the operation by the Eurolung risk score, have a Role and Emotional function lower than those with greater chances of survival. The indications for radical treatment of lung cancer are rapidly changing with new technologies (robotic surgery, more

advanced radiotherapy treatments, intraoperative chemotherapy); therefore, the use of FEV1 and DLCO in isolation may no longer be sufficient to select patients for surgery or alternative treatments. Our data provides preliminary evidence that PROs should be considered to improve stratification of these patients.

8.1.3 Chapter 5: Summary Score evaluation

Chapter 5 reports the assessment of the recently developed EORTC SumSc in our population of surgical lung cancer patients. The SumSc was able to detect pre-operative difference between groups known to have difference in their baseline clinical characteristics. The SumSc was also more sensitive than generic QOL score (i.e.GH) to identify postoperative changes in the overall population. It was sensitive to change also when looked at known groups before-after analysis. We demonstrated that a score that includes most of the QOL domains is more sensitive than one generated by two generic questions on QOL (GH). This may also be explained in our cohort of patients by the relative superiority of scales like RF and PF in detecting changes in the perioperative period described in the previous chapters. Based on these results we would recommend QOL analysis on lung cancer patients to include SumSc.

The results of the systematic review highlighted the lack of QOL reporting before and after radical treatment for early-stage NSCLC. Only one small RCT was identified that provided evidence addressing the specific question to directly compare the QOL results after SABR and VATS lung resection for lung cancer. The remaining 16 studies evaluating QOL in the single speciality, were characterized by small sample sizes and limited follow-up. These studies assessed a total 832 SABR patients and 686 receiving anatomical VATS resections, and confirmed that in general the physical components of QOL decrease immediately after treatment up to 3-months, returning to baseline after 1 year. This was

a consistent feature in the surgical papers and SABR studies demonstrated a more stability across scales during the first year. Emotional functioning supersedes the pre-operative values across treatments. The most commonly used QOL instruments was a cancer-specific one, the EORTC QOL C-30, which in few studies has been administered with the Lung Cancer specific module. These two instruments were taken forward to the prospective observational study where their content was analysed quantitively and also through interviews with patients. Chapter 6: LILAC Prospective study methods and baseline results including Interviews

We successfully recruited 244 patients in a prospective longitudinal study. It provided encouraging findings in the use of longitudinal PROMS collection in routine practice over the first 12-month period after two radical treatments (surgery and SABR) for early-stage NSCLC. Recruitment and attrition rates were acceptable and response rates were similar to other studies collecting PROMS with the exception for a lower baseline response rate, especially in the surgical group. This was because surgical patients were less likely to fill the questionnaires at the beginning of the study where they are experiencing more psychological burden, being approached during the same day of the first discussion with the surgeon. SABR patients instead have a longer period to digest and prepare to face the cancer treatment, as the planning scan comes after at least 5-10 days from the initial oncology outpatient appointment. Indeed, SABR patients instead showed a slight progressive reduction in returning the questionnaires over the following time points.

Only 11% of the consented patients decided to fill the questionnaire on-line. Although not statistically different, 15.4% of Surgical patients completed on-line, versus 7.4% of the SABR ones. The SABR population is more physiologically and functionally compromised, reflecting the fact that this treatment is indicated for patients not fit for

surgical intervention. Patients major limitation to the on-line completion was the lack of IT literacy and computer availability.

When looking at the psychometric analysis of the DSE questionnaire collected at baseline we have provided evidence for the validity in our group of patients. We showed that DSE is a 11-item measure factoring in two main subscales: the main factor (1) explaining almost 72% related to overcoming barriers to decision making (items 5-11) and the second factor explaining the additional 9.1% of the variance (items 1-4) was related to information acquisition. We have not identified a difference in efficacy between our two treatment groups.

Overall the findings from the patients end-of study interviews showed that Lilac study was well perceived by the patients. We did not identify any major barriers or downside in the method of our approach or consent process. Patients appreciated sensibly the value of filling a questionnaire about their QOL not only to help future patients with research but also to self-monitoring their symptoms. In general, there was a lack of clinical team feedback and engagement on Lilac study which deserves further focus in the next projects.

The fact that the clinicians did not mention the questionnaire results which were available to them in clinic, is accompanied by the overall impression that doctors are asking generic questions about QOL and focusing on respiratory symptoms only when issues are identified. These early findings suggest longitudinal electronic PRO collection with integration into EHR is feasible and acceptable to patients but deserve more attention to improve patient's electronic completion and clinicians' consultations of results.

8.1.4 Chapter 6: LILAC Prospective study methods and baseline results including Interviews

We successfully recruited 244 patients in a prospective longitudinal study with repeated QOL assessments. It provided encouraging findings supporting the collection of longitudinal PROMS in routine practice over the first 12-month period after two radical treatments (surgery and SABR) for early-stage NSCLC. Recruitment and attrition rates were acceptable and response rates were similar to other studies collecting PROMS with the exception of a lower baseline response rate, especially in the surgical group. This was the result of the practical and psychological burden for this group prior to surgery where they are typically approached during the same day for their first discussion with the surgeon. SABR patients instead have a longer period to process information and prepare for treatment, as the planning scan comes after at least 5-10 days from the initial oncology outpatient appointment. Indeed, SABR patients showed a slight progressive reduction in returning the questionnaires over the subsequent time points. Only 11% of participants chose to complete the questionnaire online. Although not statistically different, 15.4% of surgical patients completed online, versus 7.4% of the SABR group. The SABR population is more physiologically and functionally compromised, reflecting the fact that this treatment is indicated for patients not fit for surgical intervention. Patients' major limitation to the online completion was the lack of IT literacy and computer availability.

When looking at the psychometric analysis of the DSE questionnaire collected at baseline we have provided evidence for validity in our group of patients. We showed that DSE is a 11-item measure with two main subscales: the main factor (1) explaining almost 72% of the variance related to overcoming barriers to decision making (items 5-

11) and the second factor explaining the additional 9.1% of variance (items 1-4) related to information acquisition. We did not identify a difference in efficacy between our two treatment groups.

Overall the findings from the patients' end-of study interviews showed that the Lilac study was well received. We did not identify any major barriers or downside in the method of our approach or consent processes. Patients appreciated the value of completing a questionnaire about their QOL not only to help future patients but also for self-monitoring their symptoms. In general, there was a lack of clinical team feedback and engagement in the LILAC study and using the patient reported data in clinical encounters. This is something which deserves more attention in future projects.

The fact that the clinicians did not refer to the questionnaire results which were available to them in clinic, is accompanied by the overall impression that doctors ask generic questions about QOL, focusing on respiratory symptoms only when issues are identified. These early findings suggest longitudinal electronic PROMS collection with integration into EHR is feasible and acceptable to patients but deserve more attention to improve patient's electronic completion and clinicians' utilisation of results.

8.1.5 Chapter 7: PROMS trajectory in the LILAC Study

We reported 6 months results describing the trajectory of QOL after two different NSCLC treatments. SABR patients had significantly lower baseline scores in all the QOL domains measured with the EORTC QoL C-30 except for EF. This could be due to the fact that the SABR patients are generally less well given their significantly higher likelihood of having lung-related comorbidities and heart disease. SABR patients experience an

improvement in EF, in line with reported literature, and non-clinically significant worsening of RF. However, the other scales remained stable over time.

Surgical patients deteriorated significantly in terms of dyspnoea, especially during the first 6 weeks after the operation. They improved in the EF domain as already reported. The SumSc appeared to be more sensitive in detecting these changes.

When evaluating the clinical significance of QOL changes during these first six months using responder analysis we found interesting results. In the SABR group, ¼ of patients improved, 42% remained stable and 34% deteriorated. In the Surgical group, ¼ improved, ½ remained stable and ¼ worsened (Figure 7-6).

Interestingly, slightly more SABR patients experienced a clinically significant QOL deterioration at 6months. However, surgical patients reported the highest degree of deterioration captured by the SumScore and not by the Global QOL. Future analysis will be performed to check the clinical difference in QOL at 12 months.

The only consistent predictor of the evolution of QOL up to six months detected by SumSc was PS in SABR patients. In Surgical patients it was age and chronic kidney disease (CKD). When looking at the GH after surgery, low FEV1 and high PS showed a statistically significant association with worsening in QOL. In the SABR population, factors influencing the changes in Global Health score were high PS and the presence of cerebro-vascular disease (CVD).

We did not find any difference between the two groups in all the subscales of the PS-18 questionnaires. However, we showed that the only factors that remained independently associated with patient satisfaction were efficacy in treatment decision making in the SABR group.

8.2 Methodological aspects

8.2.1 Study design

The use of a mixed methods approach for this project allowed a multi-faceted exploration of the role of QOL collection within clinical practice and to investigate its role in the preoperative risk-assessment process. This mixed methods approach was used to ensure all important aspects of QOL data collection and their possible challenges are considered and analysed.

The retrospective studies have allowed maximal data collection and important preliminary data to shape the following prospective study. This design has, however, limitations due to the absence of a formal design. These are lack of resources, lack of expertise or advice in project design and analysis, no collection of detailed information of attrition rate and organisational impediments. The data collection was paper based and I have inserted all the QOL data into the database. It has important strengths in the fact that being performed by the clinical team involved in the patients care, it has more access to daily clinics and wards for the recruitment. It is also important to note that the initial third of those patients were recruited before the ERAS programme had started. Patients were admitted the day before, and often the questionnaire was chased or collected in the ward. This may have increased the return rate.

The selection of a prospective design in the LILAC study allowed for data collection from a relatively large participant sample to assess consistently among these two groups the QOL trajectory and the relationship to PROMS, clinical and hospital factors. The prospective study design provides a good model to address the feasibility and

acceptability of longitudinal PROMS data collection and to test the clinical utility of electronic PROMS integration in patients' records.

This study was not designed as RCT due to concerns about recruitment particularly acknowledging the difficulties of recent feasibility trials (41, 110) in the same setting. We wanted to gather information first to inform patients about their quality of life evolution, for each treatment modality. However, observational studies are potentially subject to bias. Whilst attempts are made to account for all potential confounding factors within the analysis some factors may remain hidden and unrecorded.

Whilst this analysis provides only descriptive data describing the trajectory of QOL, the data generated from both treatment groups provides useful information to inform clinical practice and the design of future clinical trials.

The qualitative research was limited by the relatively small sample size in the electronic PROMs completion group and therefore may lack generalisability. However, we have adopted an in-depth approach. The qualitative analyses illuminated potential barriers for the electronic data collection, their value in QOL discussions in clinic and possible suggestions for the type of questions selected and timing of assessment.

All studies were conducted in Leeds Cancer Centre, a specialised tertiary referral centre for thoracic surgery and radiotherapy. However, whilst all patients received SABR and surgical treatment in Leeds, some of the patients are followed up at referring satellite hospitals. This may lead to differences in the availability of the online results for the clinicians as the QTool platform is not available outside Leeds. Health professionals who follow patients up within Leeds were shown how to access the results but the use of the results and impact on care was not measured within this study.

The Enhanced Recovery After Surgery programme has been recently adopted by the surgical unit in Leeds (2015). The ERAS nurse meets the patient the same day as they

are seen by the surgeon, the staff nurse and the lung cancer specialist nurse to discuss the patient educational programme. This journey is not as well-established as in other centres(232). This means that sometimes the patient is required to wait for the ERAS nurse to find a clinic room, and/or need to wait for hours to be seen in the pre-assessment unit which is in another wing of the Hospital. This may have influenced the willingness of the patients to speak with an additional person regarding the LILAC study. This was confirmed as a theme in the qualitative interviews.

The PCOR research team has been active for over twenty years and clinicians working within the hospital have been regularly involved in studies involving PROMS research. This experience has supported me enormously to engage clinical staff members who treat and support patients treated with SABR which was clearly not my area of clinical and research expertise. The same important support also helped with the set-up of the electronic platform to access patient QOL data and patient' electronic questionnaire completion reminders and tracker.

8.2.2 Patient Reported Outcome Measures

The EORTC-QLQ C-30 questionnaire was selected for use within these clinical studies. I have also adopted the Lung Cancer specific module (LC-13), which has not been analysed for the purpose of this thesis but will be integrated to the final results with the 12 months data collection. The QLQ C-30 questionnaire has been consistently validated and it is the most commonly used in cancer settings. However, we acknowledge that the questions may be too generic for lung cancer patients especially those who have not received systemic treatment. The results of the qualitative work have championed the importance of a surgical-specific or more generic treatment-specific submodules for lung cancer patient which will be implemented soon from the EORTC QOL group. The

DSE questionnaires, although demonstrating consistency in this groups of patients, did not identify any differences between the two groups. Future analysis may be useful to investigate more on the decision-conflict rather than efficacy to get important information in the delicate decision-making process.

8.3 Strengths and limitations

The mixed method design ensured that the study appropriately tackled the research objectives, enabling a detailed exploration of what actually happens in in the first year after radical treatment for early stage NSCLC. The mixed methods approach used in this thesis worked well to address the research aim to explore feasibility and acceptability of QOL collection.

We successfully recruited 244 patients with early stage NSCLC with satisfactory completion rates over the first year that are similar to those reported in the QOL literature. We collected robust data describing the trajectory of QOL after VATS resection and SABR.

Strengths also include that the research was carried out in a real-world setting, using an observational design which increased the validity of the findings because it provides objective information about this context. In addition, gaining patients' perspectives was used to get a more subjective view of the effect of these treatments of the patients' everyday life.

This work is contributing to an emerging field of clinical research that is growing in importance. With new UK lung cancer screening studies, e.g. Yorkshire Lung Cancer Screening Trial (YLST) and Manchester Early Detection of Lung Disease (MEDLD), and increased awareness of lung cancer, the expectation is that more early stage NSCLC will

be diagnosed. This will translate to an increase incidence of early-stages NSCLC and an ageing population with the increased presence of comorbidities when compared to the younger population. For these higher risk patients, it is not known whether surgery or non-surgical treatments is the best approach for an individual patient. The failure of previous feasibility trials in this setting has limited the evidence available. Ongoing studies like VALOR and STABLEMATES are already showed major barriers in recruiting patients(152, 153). The results of this study will provide additional information about the PROMs evolution after the two accepted radical treatments for early-stage NSCLC. This can be used not only in clinical practice, but also to help in re-design feasibility RCT for these treatments.

Healthcare is becoming increasingly linked to IT and health informatics developments. "Big Data" is changing approaches to data collection and linkage of care systems. ePROMS have potential to support patients and clinical team in dealing with remote self-management(101, 104). However, patient engagement with systems is complex especially in frail populations like lung cancer patients, and evidence is needed to inform future development, evaluation and implementation. Staff interviews will offer more insight to improve lung cancer clinical team involvement in QoL electronic record consultation during patient appointment.

The methodological limitations of the individual studies are discussed in the relevant chapters. However, there were some more general limitations. The sample size of the prospective study has been limited by the implementation of new SABR centres in the region and by the important logistical issues for the surgical group. The differences of data available at baseline may have limited the results generated in this thesis. The completed dataset with 12 months data collection will improve the robustness of our findings. The collection of pulmonary function test data at 12 months also will help in

comparing the objective characteristics with PROMS. Sensitivity analysis may also help in validating our findings.

A further limitation of this study is its low electronic completion. As outlined in the Lilac study, part of the study design, like sample size, frequency of completion of outcome measures etc., were based on the experience of previous studies and on the integration of ePROMS in other cancer sites, such as breast cancer. It was clear that approaching lung cancer patients in the context of a service evaluation, through the established clinical resources, has streamlined the data collection of the work described in the Chapters 4-6. The patient's different demographics and IT literacy levels have in part also explained the marked difference in electronic completion rate between the Lilac and studies like eRAPID(104). The possibility of a future feasibility RCT on the acceptance of electronic monitoring of QOL in these patients may be considered, although patient feedback still strongly recommends the option of paper completion.

There is also a need to be conscious of the additional patient burden to complete outcome measures. For example, additional completion time-points, or additional patient interviews mid-way may have provided additional insight. It should be considered in any case the number and type of questions to be answered as already described in our previous work (233) and also considering the patients feedback on the lack of tailored questions and possibility of providing additional comments.

8.4 Implications for practice and future research

8.4.1 Using PROMS for risk assessment of patients

This study has demonstrated that routine collection and integration of PROMS results into patient EHR using a combination of electronic and paper methods is feasible and

acceptable to patients treated with SABR and surgery for early-stage NSCLC in clinical practice.

Longitudinal PROMS data capture in clinical practice using electronic methods may offer a number of benefits to improve and potentially risk stratify long term follow up of patients. However, more efforts should be made in the lung cancer setting to help older patients access electronic facilities: through their family, or at primary care level, with kiosks that allow them to report their QOL every time they are going to the GP. The impact of the electronic PROMs collection on cancer care was not an endpoint of this study, but it is evident that more strategies are need to be put thus in place and increase engagement of clinical teams.

As the surveillance practice after early stage NSCLC surgical resection is not yet completely standardized internationally (234-236), PROMS might also be considered as a method of screening patients presenting with specific symptoms patterns after resection or SABR, providing additional data to evaluate and implement new survivorship programs.

8.4.2 Developing predictive models of patient reported risk assessment and informing future trial design

We have demonstrated with retrospective data that baseline QOL predicts complications and prolonged hospital stay in surgical NSCLC patients. In the same group of patients, we also showed that the SumSc is more sensitive to perioperative difference. However, the prospective data did not confirm the role of QOL in predicting complications. In fact, one of the challenges with the models developed in both treatment groups was how to standardize the complications coding after treatments with such different morbidity as demonstrated by recent papers (25, 237). We decided

to adopt the Thoracic Morbidity and Mortality system based on Clavien-Dindo schema(238) for the surgical cohort to be consistent with the SABR patients. However, the retrospective studies coded complications according to ESTS database complication classification(166) which has already demonstrated differed in our surgical patients from the TM&M grading(239). Future analysis may attempt to model the QOL with more specific complications coding. Furthermore, we have also started collecting data about length of stay and the 12months pulmonary function tests in surgical patients, which will be analysed at the end of the 12 months collection.

By incorporating these new variables in the QOL model and incorporating the longer follow-up (12 months) we will attempt to confirm that our previous results are robust to justify consideration of integration of QoL in preoperative risk assessment.

Although complex models are less appealing to clinicians as they lack transparency in the decision-making process, we need to give more information to the patients with early-stage NSCLC at higher risk but not deemed completely unfit for surgery. These are the patients who should receive additional tailored information to make an informed choice.

The QOL preliminary data presented in this PhD work may help this process in two ways: in giving more information about the expected changes after the two treatments and in informing the design of future clinical trials in terms of when and how to collect larger scale PROMS data to investigate the inclusion of QOL in preoperative risk scores.

Once the full LILAC follow-up data is analysed, we will be able to confirm the factor/s predicting QoL at 12months. This information will be used to better stratify the "acceptance of risk" of perceived dyspnoea that the BTS guidelines refer to in the surgical preoperative algorithm(45). Survival and recurrence analysis may also help the stratification of these patients.

In addition to developing predictive models for patient reported outcomes, the data collected in this study may warrant the collection of SumSc data in larger scale datasets, for example a prospective multi-institutional database. If regular PROMs data collection is established in clinical practice and the role of QOL data in predicting postoperative morbidity is confirmed, the large observational datasets will allow for the investigation of QOL in risk scores like Eurolung. As these are already launching an App where inserting the factors of the patients we have an estimate of Mortality and Morbidity risk, it would be important to see if the inclusion of SumSc may influence the percentages.

National PROMS data collection is under consideration, and results from this study may contribute to inform future large-scale projects involving lung cancer patients at a wider population level.

In summary, the results from this project have found PROMS data collection and integration with EHR to be acceptable and feasible in patients treated with SABR and surgical resection for NSCLC over a twelve-month follow up period. The trajectory of PROMS during the first 6months is quite different between these two treatments and requires the 12months follow-up data to draw more comprehensive conclusions and inform clinical discussion and shared decision-making.

Use of PROMS as an adjunct to traditional pre-treatment clinical factors may be a more specific and able to support more tailored risk assessment models which will requires evaluation in larger multi-centre trials.

Future work will evaluate these models further using the prospective study data to validate and develop more accurate methods to optimise individual patient selection for early-stage NSCLC treatments.

References

1. UK CR. Cancer Statistics Key Facts 2017 [Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/Allcancerscombined/.

2. Office of National Statistics. General Lifestyle Survey 2008. Smoking and drinking among adults, . 2008-2010.

3. National Collaborating Centre for Cancer (UK). The Diagnosis and Treatment of Lung Cancer (Update). Cardiff (UK): National Collaborating Centre for Cancer (UK); 2011 Apr. (NICE Clinical Guidelines, No. 121.)

4. Snell N, Strachan D, Hubbard R, Gibson J, Limb E, Gupta R, et al. Burden of lung disease in the UK; findings from the British Lung Foundation's 'respiratory health of the nation' project. European Respiratory Journal. 2016;48(suppl 60):PA4913.

5. Jackman DM, Johnson BE. Small-cell lung cancer. The Lancet. 2005;366(9494):1385-96.

6. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2015;10(9):1243-60.

7. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA: a cancer journal for clinicians. 2017;67(1):7-30.

8. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. Chest. 2017;151(1):193-203.

9. Franks KN, Jain P, Snee MP. Stereotactic ablative body radiotherapy for lung cancer. Clinical oncology (Royal College of Radiologists (Great Britain)). 2015;27(5):280-9.

10. Broggio J BN. Cancer survival by stage at diagnosis for England (experimental statistics): Adults diagnosed 2012, 2013 and 2014 and followed up to 2015. 2016.

11. Walters S, Maringe C, Coleman MP, Peake MD, Butler J, Young N, et al. Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-2007. Thorax. 2013;68(6):551-64.

12. Ramnath N, Dilling TJ, Harris LJ, Kim AW, Michaud GC, Balekian AA, et al. Treatment of stage III non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e314S-e40S.

13. National Lung Cancer Audit Annual Report 2017 - for the audit period 2016. .

14. Pompili C, Brunelli A, Refai M, Xiume F, Sabbatini A. Does chronic obstructive pulmonary disease affect postoperative quality of life in patients undergoing lobectomy for lung cancer? A case-matched study. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2010;37(3):525-30.

15. Scott WJ, Howington J, Feigenberg S, Movsas B, Pisters K. Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines (2nd edition). Chest. 2007;132(3 Suppl):234s-42s.

16. Brunelli A, Charloux A, Bolliger CT, Rocco G, Sculier JP, Varela G, et al. ERS/ESTS clinical guidelines on fitness for radical therapy in lung cancer patients (surgery and chemo-radiotherapy). The European respiratory journal. 2009;34(1):17-41.

17. Islam KM, Jiang X, Anggondowati T, Lin G, Ganti AK. Comorbidity and Survival in Lung Cancer Patients. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2015;24(7):1079-85.

18. De Leyn P, Dooms C, Kuzdzal J, Lardinois D, Passlick B, Rami-Porta R, et al. Revised ESTS guidelines for preoperative mediastinal lymph node staging for non-small-cell lung cancer. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2014;45(5):787-98.

19. Lardinois D, De Leyn P, Van Schil P, Porta RR, Waller D, Passlick B, et al. ESTS guidelines for intraoperative lymph node staging in non-small cell lung cancer. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2006;30(5):787-92.

20. Harada H, Okada M, Sakamoto T, Matsuoka H, Tsubota N. Functional advantage after radical segmentectomy versus lobectomy for lung cancer. The Annals of thoracic surgery. 2005;80(6):2041-5.

21. Yang CF, D'Amico TA. Thoracoscopic segmentectomy for lung cancer. The Annals of thoracic surgery. 2012;94(2):668-81.

22. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. The Annals of thoracic surgery. 1995;60(3):615-22; discussion 22-3.

23. Ettinger DS, Wood DE, Akerley W, Bazhenova LA, Borghaei H, Camidge DR, et al. Non-Small Cell Lung Cancer, Version 6.2015. Journal of the National Comprehensive Cancer Network : JNCCN. 2015;13(5):515-24.

24. Whitson BA, Groth SS, Duval SJ, Swanson SJ, Maddaus MA. Surgery for early-stage non-small cell lung cancer: a systematic review of the video-assisted thoracoscopic

surgery versus thoracotomy approaches to lobectomy. The Annals of thoracic surgery. 2008;86(6):2008-16; discussion 16-8.

25. Cao C, Manganas C, Ang SC, Peeceeyen S, Yan TD. Video-assisted thoracic surgery versus open thoracotomy for non-small cell lung cancer: a meta-analysis of propensity score-matched patients. Interactive cardiovascular and thoracic surgery. 2013;16(3):244-9.

26. 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(5 Suppl):e166S-90S.

27. Paul S, Altorki NK, Sheng S, Lee PC, Harpole DH, Onaitis MW, et al. Thoracoscopic lobectomy is associated with lower morbidity than open lobectomy: a propensity-matched analysis from the STS database. The Journal of thoracic and cardiovascular surgery. 2010;139(2):366-78.

28. Fernando HC, Landreneau RJ, Mandrekar SJ, Nichols FC, DiPetrillo TA, Meyers BF, et al. Analysis of longitudinal quality-of-life data in high-risk operable patients with lung cancer: Results from the ACOSOG Z4032 (Alliance) multicenter randomized trial. The Journal of thoracic and cardiovascular surgery. 2015;149(3):718-26.

29. Rizk NP, Ghanie A, Hsu M, Bains MS, Downey RJ, Sarkaria IS, et al. A prospective trial comparing pain and quality of life measures after anatomic lung resection using thoracoscopy or thoracotomy. The Annals of thoracic surgery. 2014;98(4):1160-6.

30. Demmy TL, Nwogu C. Is video-assisted thoracic surgery lobectomy better? Quality of life considerations. The Annals of thoracic surgery. 2008;85(2):S719-28.

31. Li WW, Lee TW, Lam SS, Ng CS, Sihoe AD, Wan IY, et al. Quality of life following lung cancer resection: video-assisted thoracic surgery vs thoracotomy. Chest. 2002;122(2):584-9.

32. Timmerman R, Paulus R, Galvin J, Michalski J, Straube W, Bradley J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. Jama. 2010;303(11):1070-6.

33. Timmerman R, McGarry R, Yiannoutsos C, Papiez L, Tudor K, DeLuca J, et al. Excessive toxicity when treating central tumors in a phase II study of stereotactic body radiation therapy for medically inoperable early-stage lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2006;24(30):4833-9.

34. Louie AV, van Werkhoven E, Chen H, Smit EF, Paul MA, Widder J, et al. Patient reported outcomes following stereotactic ablative radiotherapy or surgery for stage IA non-small-cell lung cancer: Results from the ROSEL multicenter randomized trial. Radiotherapy and Oncology. 2015;117(1):44-8.

35. Jain P, Baker A, Distefano G, Scott AJ, Webster GJ, Hatton MQ. Stereotactic ablative radiotherapy in the UK: current status and developments. The British journal of radiology. 2013;86(1029):20130331.

36. Murray P, Franks K, Hanna GG. A systematic review of outcomes following stereotactic ablative radiotherapy in the treatment of early-stage primary lung cancer. The British journal of radiology. 2017:20160732.

37. Baumann P, Nyman J, Hoyer M, Wennberg B, Gagliardi G, Lax I, et al. Outcome in a Prospective Phase II Trial of Medically Inoperable Stage I Non–Small-Cell Lung Cancer Patients Treated With Stereotactic Body Radiotherapy. Journal of Clinical Oncology. 2009;27(20):3290-6.

38. Nagata Y, Hiraoka M, Shibata T, Onishi H, Kokubo M, Karasawa K, et al. Prospective Trial of Stereotactic Body Radiation Therapy for Both Operable and Inoperable T1N0M0 Non-Small Cell Lung Cancer: Japan Clinical Oncology Group Study JCOG0403. International Journal of Radiation Oncology*Biology*Physics. 2015;93(5):989-96.

39. Ricardi U, Filippi AR, Guarneri A, Giglioli FR, Ciammella P, Franco P, et al. Stereotactic body radiation therapy for early stage non-small cell lung cancer: Results of a prospective trial. Lung Cancer. 2010;68(1):72-7.

40. Shirvani SM, Jiang J, Chang JY, Welsh JW, Gomez DR, Swisher S, et al. Comparative Effectiveness of 5 Treatment Strategies for Early-Stage Non-Small Cell Lung Cancer in the Elderly. International Journal of Radiation Oncology*Biology*Physics. 2012;84(5):1060-70.

41. Chang JY, Senan S, Paul MA, Mehran RJ, Louie AV, Balter P, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials. The Lancet Oncology. 2015;16(6):630-7.

42. Snee MP, McParland L, Collinson F, Lowe CM, Striha A, Baldwin DR, et al. The SABRTooth feasibility trial protocol: a study to determine the feasibility and acceptability of conducting a phase III randomised controlled trial comparing stereotactic ablative radiotherapy (SABR) with surgery in patients with peripheral stage I non-small cell lung cancer (NSCLC) considered to be at higher risk of complications from surgical resection. Pilot and feasibility studies. 2016;2:5.

43. Brunelli A, Cassivi SD, Fibla J, Halgren LA, Wigle DA, Allen MS, et al. External validation of the recalibrated thoracic revised cardiac risk index for predicting the risk of major cardiac complications after lung resection. The Annals of thoracic surgery. 2011;92(2):445-8.

44. Brunelli A, Varela G, Salati M, Jimenez MF, Pompili C, Novoa N, et al. Recalibration of the revised cardiac risk index in lung resection candidates. The Annals of thoracic surgery. 2010;90(1):199-203.

45. Lim E, Baldwin D, Beckles M, Duffy J, Entwisle J, Faivre-Finn C, et al. Guidelines on the radical management of patients with lung cancer. Thorax. 2010;65 Suppl 3:iii1-27.

46. Falcoz PE, Conti M, Brouchet L, Chocron S, Puyraveau M, Mercier M, et al. The Thoracic Surgery Scoring System (Thoracoscore): risk model for in-hospital death in 15,183 patients requiring thoracic surgery. The Journal of thoracic and cardiovascular surgery. 2007;133(2):325-32.

47. 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 : official journal of the European Association for Cardio-thoracic Surgery. 2005;28(2):306-11.

48. Brunelli A, Salati M, Rocco G, Varela G, Van Raemdonck D, Decaluwe H, et al. European risk models for morbidity (EuroLung1) and mortality (EuroLung2) to predict outcome following anatomic lung resections: an analysis from the European Society of Thoracic Surgeons database. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2017;51(3):490-7.

49. Brunelli A, Berrisford RG, Rocco G, Varela G. The European Thoracic Database project: composite performance score to measure quality of care after major lung resection. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2009;35(5):769-74.

50. Brunelli A, Rocco G, Van Raemdonck D, Varela G, Dahan M. Lessons learned from the European thoracic surgery database: the Composite Performance Score. European

journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2010;36 Suppl 1:S93-9.

51. O'Dowd EL, Luchtenborg M, Baldwin DR, McKeever TM, Powell HA, Moller H, et al. Predicting death from surgery for lung cancer: A comparison of two scoring systems in two European countries. Lung Cancer. 2016;95:88-93.

52. 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(9):826-34.

53. Pompili C, Shargall Y, Decaluwe H, Moons J, Chari M, Brunelli A. Risk-adjusted performance evaluation in three academic thoracic surgery units using the Eurolung risk models. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2018.

54. Brunelli A. European Society of Thoracic Surgeons Risk Scores. Thoracic surgery clinics. 2017;27(3):297-302.

55. Ferguson MK, Stromberg JD, Celauro AD. Estimating lung resection risk: a pilot study of trainee and practicing surgeons. The Annals of thoracic surgery. 2010;89(4):1037-42; discussion 42-3.

56. Cook C. Is clinical gestalt good enough? The Journal of manual & manipulative therapy. 2009;17(1):6-7.

57. MacCormick AD, Parry BR. Judgment Analysis of Surgeons' Prioritization of Patients for Elective General Surgery. Medical Decision Making. 2006;26(3):255-64.

58. Ferguson MK, Farnan J, Hemmerich JA, Slawinski K, Acevedo J, Small S. The impact of perceived frailty on surgeons' estimates of surgical risk. The Annals of thoracic surgery. 2014;98(1):210-6.

59. Timmermans D, Kievit J, van Bockel H. How do surgeons' probability estimates of operative mortality compare with a decision analytic model? Acta psychologica. 1996;93(1-3):107-20.

60. Dijksterhuis A, Bos MW, Nordgren LF, van Baaren RB. On making the right choice: the deliberation-without-attention effect. Science (New York, NY). 2006;311(5763):1005-7.

61. Oh S, Miyamoto H, Yamazaki A, Fukai R, Shiomi K, Sonobe S, et al. Prospective analysis of depression and psychological distress before and after surgical resection of lung cancer. General Thoracic and Cardiovascular Surgery. 2007;55(3):119-24.

62. Brunelli A, Socci L, Refai M, Salati M, Xiumé F, Sabbatini A. Quality of Life Before and After Major Lung Resection for Lung Cancer: A Prospective Follow-Up Analysis. The Annals of thoracic surgery. 2007;84(2):410-6.

63. Handy JR, Jr., Asaph JW, Skokan L, Reed CE, Koh S, Brooks G, et al. What happens to patients undergoing lung cancer surgery? Outcomes and quality of life before and after surgery. Chest. 2002;122(1):21-30.

64. Brunelli A, Pompili C, Koller M. Changes in quality of life after pulmonary resection. Thoracic surgery clinics. 2012;22(4):471-85.

65. Burfeind WR, Tong BC, O'Branski E, Herndon JE, Toloza EM, D'Amico TA, et al. Quality of life outcomes are equivalent after lobectomy in the elderly. The Journal of thoracic and cardiovascular surgery. 2008;136(3):597-604.

66. Salati M, Brunelli A, Xiumè F, Refai M, Sabbatini A. Quality of life in the elderly after major lung resection for lung cancer☆. Interactive cardiovascular and thoracic surgery. 2009;8(1):79-83.

67. Pompili C, Brunelli A, Refai M, Xiumè F, Sabbatini A. Does chronic obstructive pulmonary disease affect postoperative quality of life in patients undergoing lobectomy for lung cancer? A case-matched study☆. European Journal of Cardio-Thoracic Surgery. 2010;37(3):525-30.

68. Cykert S, Kissling G, Hansen CJ. Patient preferences regarding possible outcomes of lung resection: what outcomes should preoperative evaluations target? Chest. 2000;117(6):1551-9.

69. Bottomley A. The Cancer Patient and Quality of Life. 2002;7(2):120-5.

70. Velikova G, Stark D, Selby P. Quality of life instruments in oncology. European journal of cancer (Oxford, England : 1990). 1999;35(11):1571-80.

71. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. Controlled clinical trials. 2004;25(6):535-52.

72. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and Quality of Life Outcomes. 2006;4:79.

73. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. Journal of evaluation in clinical practice. 1999;5(4):401-16.

74. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2012;21(8):1305-14.

75. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2004;22(4):714-24.

76. Velikova G, Brown JM, Smith AB, Selby PJ. Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. British journal of cancer. 2002;86(1):51-9.

77. Ahmed S, Berzon RA, Revicki DA, Lenderking WR, Moinpour CM, Basch E, et al. The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Medical care. 2012;50(12):1060-70.

78. Black N. Patient reported outcome measures could help transform healthcare. BMJ (Clinical research ed). 2013;346:f167.

79. Pompili C, Salati M, Refai M, Berardi R, Onofri A, Mazzanti P, et al. Preoperative quality of life predicts survival following pulmonary resection in stage I non-small-cell lung cancer. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2013;43(5):905-10.

80. Bouazza YB, Chiairi I, El Kharbouchi O, De Backer L, Vanhoutte G, Janssens A, et al. Patient-reported outcome measures (PROMs) in the management of lung cancer: A systematic review. Lung Cancer. 2017;113:140-51.

81. Pompili C, Novoa N, Balduyck B. Clinical evaluation of quality of life: a survey among members of European Society of Thoracic Surgeons (ESTS). Interactive cardiovascular and thoracic surgery. 2015;21(4):415-9.

82. Chen H, Louie AV, Boldt RG, Rodrigues GB, Palma DA, Senan S. Quality of Life After Stereotactic Ablative Radiotherapy for Early-Stage Lung Cancer: A Systematic Review. Clinical lung cancer. 2016;17(5):e141-e9.

83. M. M. Relationship-Based Health Care in the United States, United Kingdom, Canada, Germany, South Africa, and Japan. A Comparative Study of Patient and Physician Perceptions Worldwide.; 2003.

84. Tamblyn R, Abrahamowicz M, Dauphinee D, Wenghofer E, Jacques A, Klass D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. Jama. 2007;298(9):993-1001.

85. Colombo R, Doherty DJ, Wilson CM, Krzys K, Lange S, Maynes H. Implementation and Preliminary Analysis of FACT-G Quality of Life Questionnaire within an Oncology Survivorship Clinic. Cureus. 2018;10(3):e2272.

86. Pompili C. Quality of life after lung resection for lung cancer. Journal of thoracic disease. 2015;7(Suppl 2):S138-44.

87. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical care. 1992;30(6):473-83.

88. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-76.

89. Koller M, Warncke S, Hjermstad MJ, Arraras J, Pompili C, Harle A, et al. Use of the lung cancer-specific Quality of Life Questionnaire EORTC QLQ-LC13 in clinical trials: A systematic review of the literature 20 years after its development. Cancer. 2015;121(24):4300-23.

90. Brunelli A, Rocco G. Clinical and nonclinical indicators of performance in thoracic surgery. Thoracic surgery clinics. 2007;17(3):369-77.

91. Avery KN, Metcalfe C, Nicklin J, Barham CP, Alderson D, Donovan JL, et al. Satisfaction with care: an independent outcome measure in surgical oncology. Ann Surg Oncol. 2006;13(6):817-22.

92. Bredart A, Bottomley A, Blazeby JM, Conroy T, Coens C, D'Haese S, et al. An international prospective study of the EORTC cancer in-patient satisfaction with care measure (EORTC IN-PATSAT32). European journal of cancer (Oxford, England : 1990). 2005;41(14):2120-31.

93. Pompili C, Brunelli A, Rocco G, Salvi R, Xiume F, La Rocca A, et al. Patient satisfaction after pulmonary resection for lung cancer: a multicenter comparative analysis. Respiration; international review of thoracic diseases. 2013;85(2):106-11.

94. Marshall GNaRDH. The Patient Satisfaction Questionnaire Short Form (PSQ-18). 1994.

95. Hotta K, Kiura K, Takigawa N, Yoshioka H, Hayashi H, Fukuyama H, 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. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2010;5(10):1668-72.

96. Gartner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: A systematic review. PloS one. 2018;13(2):e0191747.

97. Scholl I, Koelewijn-van Loon M, Sepucha K, Elwyn G, Legare F, Harter M, et al. Measurement of shared decision making - a review of instruments. Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen. 2011;105(4):313-24.

98. AM OC. User Manual – Decision Self-Efficacy Scale. Ottawa: Ottawa Hospital Research Institute. 1995.

99. O'Connor AM. Validation of a decisional conflict scale. Medical decision making : an international journal of the Society for Medical Decision Making. 1995;15(1):25-30.

100. Bunn H, O'Connor A. Validation of client decision-making instruments in the context of psychiatry. The Canadian journal of nursing research = Revue canadienne de recherche en sciences infirmieres. 1996;28(3):13-27.

101. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. Journal of Clinical Oncology. 2016;34(6):557-65.

102. Wright EP, Selby PJ, Crawford M, Gillibrand A, Johnston C, Perren TJ, et al. Feasibility and compliance of automated measurement of quality of life in oncology practice. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2003;21(2):374-82.

103. Holch P, Warrington L, Bamforth LCA, Keding A, Ziegler LE, Absolom K, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. Annals of oncology : official journal of the European Society for Medical Oncology. 2017;28(9):2305-11.

104. Absolom K, Holch P, Warrington L, Samy F, Hulme C, Hewison J, et al. Electronic patient self-Reporting of Adverse-events: Patient Information and aDvice (eRAPID): a randomised controlled trial in systemic cancer treatment. BMC Cancer. 2017;17(1):318.

105. Murray L, Ramasamy S, Lilley J, Snee M, Clarke K, Musunuru HB, et al. Stereotactic Ablative Radiotherapy (SABR) in Patients with Medically Inoperable Peripheral Early Stage Lung Cancer: Outcomes for the First UK SABR Cohort. Clinical Oncology. 2016;28(1):4-12.

106. Cheyne L, Esterbrook G, Viadyanathan S, Milton R, Smith G, Blaxill P, et al. S109 The introduction of stereotactic ablative radiotherapy increases overall radical treatment rates for stage I lung cancer but does not reduce surgical resection rates–a two centre study. 2013;68(Suppl 3):A57-A.

107. Cancer in Yorkshire. Statistics. [Available from: http://yorkshirecancerresearch.org.uk/cancerinyorkshire/.

108. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. The New England journal of medicine. 2017;376(2):105-8.

109. Bush N, Donaldson G, Moinpour C, Haberman M, Milliken D, Markle V, et al. Development, feasibility and compliance of a web-based system for very frequent QOL and symptom home self-assessment after hematopoietic stem cell transplantation. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2005;14(1):77-93.

110. Franks K, McParland L, Webster J, Baldwin D, Sebag-Montefiore D, Evison M, et al. P2.16-16 SABRTOOTH: A Fasibility Study of SABR Versus Surgery in Patients with Peripheral Stage I NSCLC Considered to be at Higher Risk for Surgery. Journal of Thoracic Oncology. 2018;13(10).

111. Abstracts Presented at the 22nd Annual Conference of the International Society for Quality of Life Research. Quality of Life Research. 2015;24:A5-209.

112. Tariq S, Woodman J. Using mixed methods in health research. JRSM short reports. 2013;4(6):2042533313479197.

113. Curry LA, Nembhard IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. Circulation. 2009;119(10):1442-52.

114. Ivankova NV, Creswell JW, Stick SL. Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. Field Methods. 2006;18(1):3-20.

115. Tashakkori A, Creswell JW. Editorial: Exploring the Nature of Research Questions in Mixed Methods Research. Journal of Mixed Methods Research. 2007;1(3):207-11.

116. Guetterman TC, Fetters MD, Creswell JW. Integrating Quantitative and Qualitative Results in Health Science Mixed Methods Research Through Joint Displays. Annals of family medicine. 2015;13(6):554-61.

117. Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). Journal of evaluation in clinical practice. 2018;24(3):459-67.

118. Mackway-Jones K, Carley SD, Morton RJ, Donnan S. The best evidence topic report: a modified CAT for summarising the available evidence in emergency medicine. Journal of accident & emergency medicine. 1998;15(4):222-6.

119. Carley SD, Mackway-Jones K, Jones A, Morton RJ, Dollery W, Maurice S, et al. Moving towards evidence based emergency medicine: use of a structured critical appraisal journal club. Journal of accident & emergency medicine. 1998;15(4):220-2.

120. Khan OA, Dunning J, Parvaiz AC, Agha R, Rosin D, Mackway-Jones K. Towards evidence-based medicine in surgical practice: best BETs. Int J Surg. 2011;9(8):585-8.

121. Mabvuure NT, Klimach S, Eisner M, Rodrigues JN. An audit of best evidence topic reviews in the International Journal of Surgery. Int J Surg. 2015;17:54-9.

122. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP journal club. 1995;123(3):A12-3.

123. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC medical informatics and decision making. 2007;7:16-.

124. Sedgwick P. Bias in observational study designs: cross sectional studies. BMJ (Clinical research ed). 2015;350:h1286.

125. Giesinger JM, Kieffer JM, Fayers PM, Groenvold M, Petersen MA, Scott NW, et al. Replication and validation of higher order models demonstrated that a summary score for the EORTC QLQ-C30 is robust. J Clin Epidemiol. 2016;69:79-88.

126. Calvert M, Thwaites R, Kyte D, Devlin N. Putting patient-reported outcomes on the 'Big Data Road Map'. Journal of the Royal Society of Medicine. 2015;108(8):299-303.

127. Batchelor TJP, Naidu B, Rasburn NJ, Abdelnour-Berchtold E, Gonzalez M, Brunelli A, et al. Guidelines for enhanced recovery after lung surgery: recommendations of the Enhanced Recovery After Surgery (ERAS®) Society and the European Society of Thoracic Surgeons (ESTS). European Journal of Cardio-Thoracic Surgery. 2018;55(1):91-115.

128. Paul S, Lee PC, Mao J, Isaacs AJ, Sedrakyan A. Long term survival with stereotactic ablative radiotherapy (SABR) versus thoracoscopic sublobar lung resection in elderly people: national population based study with propensity matched comparative analysis. 2016;354:i3570.

129. Verstegen NE, Oosterhuis JW, Palma DA, Rodrigues G, Lagerwaard FJ, van der Elst A, et al. Stage I-II non-small-cell lung cancer treated using either stereotactic ablative radiotherapy (SABR) or lobectomy by video-assisted thoracoscopic surgery (VATS): outcomes of a propensity score-matched analysis. Annals of oncology : official journal of the European Society for Medical Oncology. 2013;24(6):1543-8.

130. Crabtree TD, Denlinger CE, Meyers BF, El Naqa I, Zoole J, Krupnick AS, et al. Stereotactic body radiation therapy versus surgical resection for stage I non-small cell lung cancer. The Journal of thoracic and cardiovascular surgery. 2010;140(2):377-86.

131. Ashley L, Keding A, Brown J, Velikova G, Wright P. Score equivalence of electronic and paper versions of the Social Difficulties Inventory (SDI-21): a randomised crossover

trial in cancer patients. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2013;22(6):1435-40.

132. Lagerwaard FJ, Aaronson NK, Gundy CM, Haasbeek CJ, Slotman BJ, Senan S. Patient-reported quality of life after stereotactic ablative radiotherapy for early-stage lung cancer. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2012;7(7):1148-54.

133. Pompili C, Absolom K, Franks K, Velikova GJJoTD. Are quality of life outcomes comparable following stereotactic radiotherapy and minimally invasive surgery for stage I lung cancer patients? 2018. 2018;10(12):7055-63.

134. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine. 2009;6(7):e1000097.

135. Fenton L, Lauckner H, Gilbert R. The QATSDD critical appraisal tool: comments and critiques. Journal of evaluation in clinical practice. 2015;21(6):1125-8.

136. Brundage M, Blazeby J, Revicki D, Bass B, de Vet H, Duffy H, et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2013;22(6):1161-75.

137. Campbell M, Katikireddi SV, Sowden A, McKenzie JE, Thomson H. Improving Conduct and Reporting of Narrative Synthesis of Quantitative Data (ICONS-Quant): protocol for a mixed methods study to develop a reporting guideline. 2018;8(2):e020064.

138. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. Journal of Clinical Oncology. 1998;16(1):139-44.

139. Lagerwaard FJ, Aaronson NK, Gundy CM, Haasbeek CJA, Slotman BJ, Senan S. Patient-Reported Quality of Life After Stereotactic Ablative Radiotherapy for Early-Stage Lung Cancer. Journal of Thoracic Oncology. 2012;7(7):1148-54.

140. Mathieu D, Campeau MP, Bahig H, Larrivee S, Vu T, Lambert L, et al. Long-term quality of life in early-stage non-small cell lung cancer patients treated with robotic stereotactic ablative radiation therapy. Pract Radiat Oncol. 2015;5(4):e365-73.

141. Ubels RJ, Mokhles S, Andrinopoulou ER, Braat C, van der Voort van Zyp NC, Aluwini S, et al. Quality of life during 5 years after stereotactic radiotherapy in stage I non-small cell lung cancer. Radiat Oncol. 2015;10:98.

142. van der Voort van Zyp NC, Prevost JB, van der Holt B, Braat C, van Klaveren RJ, Pattynama PM, et al. Quality of life after stereotactic radiotherapy for stage I non-small-cell lung cancer. International journal of radiation oncology, biology, physics. 2010;77(1):31-7.

143. Widder J, Postmus D, Ubbels JF, Wiegman EM, Langendijk JA. Survival and Quality of Life After Stereotactic or 3D-Conformal Radiotherapy for Inoperable Early-Stage Lung Cancer. International Journal of Radiation Oncology*Biology*Physics. 2011;81(4):e291-e7.

144. Ferrero C, Badellino S, Filippi AR, Focaraccio L, Giaj Levra M, Levis M, et al. Pulmonary function and quality of life after VMAT-based stereotactic ablative radiotherapy for early stage inoperable NSCLC: a prospective study. Lung Cancer. 2015;89(3):350-6.

145. Jain S, Poon I, Soliman H, Keller B, Kim A, Lochray F, et al. Lung stereotactic body radiation therapy (SBRT) delivered over 4 or 11days: A comparison of acute toxicity and quality of life. Radiotherapy and Oncology. 2013;108(2):320-5.

146. Videtic GMM, Reddy CA, Sorenson L. A prospective study of quality of life including fatigue and pulmonary function after stereotactic body radiotherapy for medically inoperable early-stage lung cancer. Supportive Care in Cancer. 2013;21(1):211-8.

147. Sun V, Kim JY, Williams AC, Raz DJ, Sampath S, Ferrell B. Quality of life and symptoms following stereotactic body radiotherapy in early-stage lung cancer patients. The Journal of community and supportive oncology. 2014;12(11):407-14.

148. Bendixen M, Jørgensen OD, Kronborg C, Andersen C, Licht PB. Postoperative pain and quality of life after lobectomy via video-assisted thoracoscopic surgery or anterolateral thoracotomy for early stage lung cancer: a randomised controlled trial. The Lancet Oncology. 2016;17(6):836-44.

149. Handy JJR, Asaph JW, Douville EC, Ott GY, Grunkemeier GL, Wu Y. Does video-assisted thoracoscopic lobectomy for lung cancer provide improved functional outcomes compared with open lobectomy? European Journal of Cardio-Thoracic Surgery. 2010;37(2):451-5.

150. Khullar OV, Rajaei MH, Force SD, Binongo JN, Lasanajak Y, Robertson S, et al. Pilot Study to Integrate Patient Reported Outcomes After Lung Cancer Operations Into The Society of Thoracic Surgeons Database. The Annals of thoracic surgery. 2017;104(1):245-53.

151. Fagundes CP, Shi Q, Vaporciyan AA, Rice DC, Popat KU, Cleeland CS, et al. Symptom recovery after thoracic surgery: Measuring patient-reported outcomes with the MD Anderson Symptom Inventory. The Journal of thoracic and cardiovascular surgery. 2015;150(3):613-9.e2.

152. Health UNIo. Stereotactic Body Radiotherapy (SBRT) Versus Sublobar Resection for High-Risk Patients Non-Small Lung Cancer 2015.

153. Affairs UDoV. CSP #2005 - Veterans Affairs Lung Cancer Surgery Or Stereotactic Radiotherapy trial (VALOR 2015).

154. Koller M, Hjermstad MJ, Tomaszewski KA, Tomaszewska IM, Hornslien K, Harle A, et al. An international study to revise the EORTC questionnaire for assessing quality of life in lung cancer patients. Annals of oncology : official journal of the European Society for Medical Oncology. 2017;28(11):2874-81.

155. Pompili C, Velikova G, White J, Callister M, Robson J, Dixon S, et al. Poor preoperative patient-reported quality of life is associated with complications following pulmonary lobectomy for lung cancer. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2017;51(3):526-31.

156. Martinez-Alario J, Tuesta ID, Plasencia E, Santana M, Mora ML. Mortality prediction in cardiac surgery patients: comparative performance of Parsonnet and general severity systems. Circulation. 1999;99(18):2378-82.

157. Cykert S. Risk acceptance and risk aversion: patients' perspectives on lung surgery. Thoracic surgery clinics. 2004;14(3):287-93.

158. Ganai S, Ferguson MK. Quality of life in the high-risk candidate for lung resection. Thoracic surgery clinics. 2012;22(4):497-508.

159. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases. 1987;40(5):373-83.

160. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg. 1991;78(3):355-60.

161. Bradley A, Marshall A, Abdelaziz M, Hussain K, Agostini P, Bishay E, et al. Thoracoscore fails to predict complications following elective lung resection. European Respiratory Journal. 2012;40(6):1496-501.

162. Qadri SS, Jarvis M, Ariyaratnam P, Chaudhry MA, Cale AR, Griffin S, et al. Could Thoracoscore predict postoperative mortality in patients undergoing pneumonectomy? European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2014;45(5):864-9.

163. Pompili C, Falcoz PE, Salati M, Szanto Z, Brunelli A. A risk score to predict the incidence of prolonged air leak after video-assisted thoracoscopic lobectomy: An analysis from the European Society of Thoracic Surgeons database. The Journal of thoracic and cardiovascular surgery. 2017;153(4):957-65.

164. Boyd P. The requirements of the Data Protection Act 1998 for the processing of medical data. Journal of medical ethics. 2003;29(1):34-5.

165. Switula DJS, Ethics E. Principles of good clinical practice (GCP) in clinical research. 2000;6(1):71-7.

166. 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. The Annals of thoracic surgery. 2015;99(1):368-76.

167. Blackstone EH. Breaking down barriers: helpful breakthrough statistical methods you need to understand better. The Journal of thoracic and cardiovascular surgery. 2001;122(3):430-9.

168. Brunelli A. A synopsis of resampling techniques. Journal of thoracic disease. 2014;6(12):1879-82.

169. Erus S, Tanju S, Kapdagli M, Ozkan B, Dilege S, Toker A. The comparison of complication, pain, quality of life and performance after lung resections with thoracoscopy and axillary thoracotomy. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2014;46(4):614-9.

170. Goodman SH, Sewell DR, Cooley EL, Leavitt N. Assessing levels of adaptive functioning: the Role Functioning Scale. Community mental health journal. 1993;29(2):119-31.

171. Sherbourne CD, Sturm R, Wells KB. What outcomes matter to patients? Journal of general internal medicine. 1999;14(6):357-63.

172. Anatchkova MD, Bjorner JB. Health and role functioning: the use of focus groups in the development of an item bank. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2010;19(1):111-23.

173. HINZ A, EINENKEL J, BRIEST S, STOLZENBURG J-U, PAPSDORF K, SINGER S. Is it useful to calculate sum scores of the quality of life questionnaire EORTC QLQ-C30? 2012;21(5):677-83.

174. Nordin K, Steel J, Hoffman K, Glimelius B. Alternative methods of interpreting quality of life data in advanced gastrointestinal cancer patients. British journal of cancer. 2001;85(9):1265-72.

175. Gundy CM, Fayers PM, Groenvold M, Petersen MA, Scott NW, Sprangers MA, et al. Comparing higher order models for the EORTC QLQ-C30. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2012;21(9):1607-17.

176. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella DJQoLR. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. 2009;18(7):873-80.

177. L. Backhus P. Bhandari CP, A. Brunelli, N. Novoa, K. Naunheim, M. Edwards. Surgeon Practices for Post Resection Lung Cancer Surveillance: Comparisons of STS and ESTS Members. IASLC 18th World Conference on Lung Cancer; 15-18 October 2017; Yokohama, Japan2017.

178. Pompili C, Koller M, Velikova G, Franks K, Absolom K, Callister M, 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.

179. Ferguson MK, Parma CM, Celauro AD, Vigneswaran WT. Quality of life and mood in older patients after major lung resection. The Annals of thoracic surgery. 2009;87(4):1007-12; discussion 12-3.

180. Valentine J.C CH. Effect size substantive interpretation guidelines: issues in the interpretation of effect sizes. In: Clearinghouse DW, editor. Washington2003.

181. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. American journal of clinical oncology. 1982;5(6):649-55.

182. Salati M, Brunelli A. Risk Stratification in Lung Resection. Current surgery reports. 2016;4(11):37-.

183. Pompili C, Brunelli A. Quality of life after lung resection is not associated with functional objective measures. The European respiratory journal. 2013;42(1):283-5.

184. 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. The Annals of thoracic surgery. 2002;74(4):999-1003.

185. Hudson J, Semenkovich T, Puri V. Oncologic Quality Indicators in Thoracic Surgery. Thoracic surgery clinics. 2017;27(3):227-44.

186. Kozower BD, O'Brien SM, Kosinski AS, Magee MJ, Dokholyan R, Jacobs JP, et al. The Society of Thoracic Surgeons Composite Score for Rating Program Performance for Lobectomy for Lung Cancer. The Annals of thoracic surgery. 2016;101(4):1379-86; discussion 86-7.

187. Phillips R, Gandhi M, Cheung YB, Findlay MP, Win KM, Hai HH, et al. Summary scores captured changes in subjects' QoL as measured by the multiple scales of the EORTC QLQ-C30. J Clin Epidemiol. 2015;68(8):895-902.

188. Chang N-W, Lin K-C, Hsu W-H, Lee S-C, Chan JY-H, Wang K-Y. The effect of gender on health-related quality of life and related factors in post-lobectomy lung-cancer patients. European Journal of Oncology Nursing. 2015;19(3):292-300.

189. Colt HG, Murgu SD, Korst RJ, Slatore CG, Unger M, Quadrelli S. Follow-up and surveillance of the patient with lung cancer after curative-intent therapy: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143(5 Suppl):e437S-e54S.

190. Pompili C, Franks KN, Brunelli A, Hussain YS, Holch P, Callister ME, et al. Patient reported outcomes following video assisted thoracoscopic (VATS) resection or stereotactic ablative body radiotherapy (SABR) for treatment of non-small cell lung cancer: protocol for an observational pilot study (LiLAC). Journal of thoracic disease. 2017;9(8):2703-13.

191. Hopmans W, Damman OC, Senan S, Hartemink KJ, Smit EF, Timmermans DRM. A patient perspective on shared decision making in stage I non-small cell lung cancer: a mixed methods study. BMC Cancer. 2015;15(1):959.

192. Chen HL, Liu K, You QS. Self-efficacy, cancer-related fatigue, and quality of life in patients with resected lung cancer. European journal of cancer care. 2018;27(6):e12934.

193. MacCallum RC, Widaman, Keith F.,Zhang, Shaobo,Hong, Sehee. Sample size in factor analysis. Psychological Methods, Vol 4(1), Mar 1999, 84-99.

194. Thorndike RM. Book Review : Psychometric Theory (3rd ed.) by Jum Nunnally and Ira Bernstein New York: McGraw-Hill, 1994, xxiv + 752 pp. Applied Psychological Measurement. 1995;19(3):303-5.

195. Castleberry A, Nolen A. Thematic analysis of qualitative research data: Is it as easy as it sounds? Currents in pharmacy teaching & learning. 2018;10(6):807-15.

196. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nursing & health sciences. 2013;15(3):398-405.

197. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). Medical education. 2006;40(5):405-14.

198. Salati M, Brunelli A, Decaluwe H, Szanto Z, Dahan M, Varela G, 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(6):1041-8.

199. Falcoz PE, Puyraveau M, Thomas PA, Decaluwe H, Hurtgen M, Petersen RH, et al. Video-assisted thoracoscopic surgery versus open lobectomy for primary non-small-cell lung cancer: a propensity-matched analysis of outcome from the European Society of Thoracic Surgeon database. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2016;49(2):602-9.

200. Brunelli A, Dinesh P, Woodcock-Shaw J, Littlechild D, Pompili C. Ninety-Day Mortality After Video-Assisted Thoracoscopic Lobectomy: Incidence and Risk Factors. The Annals of thoracic surgery. 2017;104(3):1020-6.

201. Mortari L. Reflectivity in Research Practice:An Overview of Different Perspectives. 2015;14(5):1609406915618045.

202. Delgado A, López-Fernández L-A, Luna JdD, Saletti-Cuesta L, Gil N, Jiménez M. The role of expectations in preferences of patients for a female or male general practitioner. Patient education and counseling. 2011;82(1):49-57.

203. Roter DL, Hall JA, Aoki Y. Physician Gender Effects in Medical CommunicationA Meta-analytic Review. Jama. 2002;288(6):756-64.

204. Cranney A, O'Connor AM, Jacobsen MJ, Tugwell P, Adachi JD, Ooi DS, et al. Development and pilot testing of a decision aid for postmenopausal women with osteoporosis. Patient education and counseling. 2002;47(3):245-55.

205. Firat S, Byhardt RW, Gore E. Comorbidity and Karnofksy performance score are independent prognostic factors in stage III non-small-cell lung cancer: an institutional analysis of patients treated on four RTOG studies. International Journal of Radiation Oncology*Biology*Physics. 2002;54(2):357-64.

206. Cerny T, Anderson H, Bramwell V, Thatcher N, Blair V. Pretreatment prognostic factors and scoring system in 407 small-cell lung cancer patients. 1987;39(2):146-9.

207. Legare F, Kearing S, Clay K, Gagnon S, D'Amours D, Rousseau M, et al. Are you SURE?: Assessing patient decisional conflict with a 4-item screening test. Canadian family physician Medecin de famille canadien. 2010;56(8):e308-14.

208. Schulte T, Schniewind B, Walter J, Dohrmann P, Kuchler T, Kurdow R. Age-related impairment of quality of life after lung resection for non-small cell lung cancer. Lung Cancer. 2010;68(1):115-20.

209. 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(3):423-31.

210. Farrar JT, Dworkin RH, Max MB. Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: making clinical trial data more understandable. J Pain Symptom Manage. 2006;31(4):369-77.

211. Lin Y. Robust inference for responder analysis: Innovative clinical trial design using a minimum p-value approach. Contemp Clin Trials Commun. 2016;3:65-9.

212. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res. 2011;11(2):163-9.

213. Deyi BA, Kosinski AS, Snapinn SM. Power considerations when a continuous outcome variable is dichotomized. Journal of Biopharmaceutical Statistics. 1998;8(2):337-52.

214. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ (Clinical research ed). 2006;332(7549):1080.

215. Leucht S, Davis JM, Engel RR, Kane JM, Wagenpfeil S. Defining 'Response' in Antipsychotic Drug Trials: Recommendations for the Use of Scale-Derived Cutoffs. Neuropsychopharmacology. 2007;32:1903.

216. Velikova G, Williams LJ, Willis S, Dixon JM, Loncaster J, Hatton M, et al. Quality of life after postmastectomy radiotherapy in patients with intermediate-risk breast cancer (SUPREMO): 2-year follow-up results of a randomised controlled trial. The Lancet Oncology. 2018;19(11):1516-29.

217. Trotti A, Colevas AD, Setser A, Rusch V, Jaques D, Budach V, et al. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. Seminars in radiation oncology. 2003;13(3):176-81.

218. Nolte S, Liegl G, Petersen MA, Aaronson NK, Costantini A, Fayers PM, et al. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. European journal of cancer (Oxford, England : 1990). 2019;107:153-63.

219. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical care. 2003;41(5):582-92.

220. Kane RL, Maciejewski M, Finch M. The relationship of patient satisfaction with care and clinical outcomes. Medical care. 1997;35(7):714-30.

221. Poghosyan H, Sheldon LK, Leveille SG, Cooley ME. Health-related quality of life after surgical treatment in patients with non-small cell lung cancer: a systematic review. Lung Cancer. 2013;81(1):11-26.

222. Barlesi F, Boyer L, Doddoli C, Antoniotti S, Thomas P, Auquier P. The place of patient satisfaction in quality assessment of lung cancer thoracic surgery. Chest. 2005;128(5):3475-81.

223. Pompili C, Tiberi M, Salati M, Refai M, Xiume F, Brunelli A. Patient satisfaction with health-care professionals and structure is not affected by longer hospital stay and complications after lung resection: a case-matched analysis. Interactive cardiovascular and thoracic surgery. 2015;20(2):236-41.

224. Cheyne L, Taylor A, Milton R, Fear J, Callister MEJ. Social deprivation does not affect lung cancer stage at presentation or disease outcome. Lung Cancer. 2013;81(2):247-51.

225. Health Q. National Cancer Patient Experience Survey 2011/12 - National Report. 2012.

226. Nakahara K, Ohno K, Hashimoto J, Miyoshi S, Maeda H, Matsumura A, et al. Prediction of postoperative respiratory failure in patients undergoing lung resection for lung cancer. The Annals of thoracic surgery. 1988;46(5):549-52.

227. Magdeleinat P, Seguin A, Alifano M, Boubia S, Regnard JF. Early and long-term results of lung resection for non-small-cell lung cancer in patients with severe ventilatory impairment. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2005;27(6):1099-105.

228. Berry MF, Hanna J, Tong BC, Burfeind WR, Jr., Harpole DH, D'Amico TA, et al. Risk factors for morbidity after lobectomy for lung cancer in elderly patients. The Annals of thoracic surgery. 2009;88(4):1093-9.

229. Tong BC, Kosinski AS, Burfeind WR, Jr., Onaitis MW, Berry MF, Harpole DH, Jr., et al. Sex differences in early outcomes after lung cancer resection: analysis of the Society of Thoracic Surgeons General Thoracic Database. The Journal of thoracic and cardiovascular surgery. 2014;148(1):13-8.

230. Kawaguchi T, Takada M, Kubo A, Matsumura A, Fukai S, Tamura A, et al. Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2010;5(5):620-30.

231. Kikuchi N, Ohmori K, Kuriyama S, Shimada A, Nakaho T, Yamamuro M, et al. Survival Prediction of Patients with Advanced Cancer: The Predictive Accuracy of the Model Based on Biological Markers. Journal of Pain and Symptom Management. 2007;34(6):600-6.

232. Rogers LJ, Bleetman D, Messenger DE, Joshi NA, Wood L, Rasburn NJ, et al. The impact of enhanced recovery after surgery (ERAS) protocol compliance on morbidity

from resection for primary lung cancer. The Journal of thoracic and cardiovascular surgery. 2018;155(4):1843-52.

233. Pompili C, Brunelli A, Xiume F, Refai M, Salati M, Socci L, et al. Prospective external convergence evaluation of two different quality-of-life instruments in lung resection patients. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2011;40(1):99-105.

234. Pompili C, Edwards M, Hasegawa S, Yoshino I, Chida M, Novoa N, et al. P2.15-21 Post Resection Lung Cancer Surveillance: Comparisons of Practice Between STS, ESTS and JACS Members. Journal of Thoracic Oncology. 2018;13(10):S825-S6.

235. Fahim C, Hanna WC. Surveillance after surgical resection of stage I non-small cell lung cancer. AME Medical Journal. 2017;2:128-.

236. Hanna WC, Keshavjee S. How to follow up patients after curative resection of lung cancer. Semin Thorac Cardiovasc Surg. 2013;25(3):213-7.

237. Ackerson BG, Tong BC, Hong JC, Gu L, Chino J, Trotter JW, et al. Stereotactic body radiation therapy versus sublobar resection for stage I NSCLC. Lung Cancer. 2018;125:185-91.

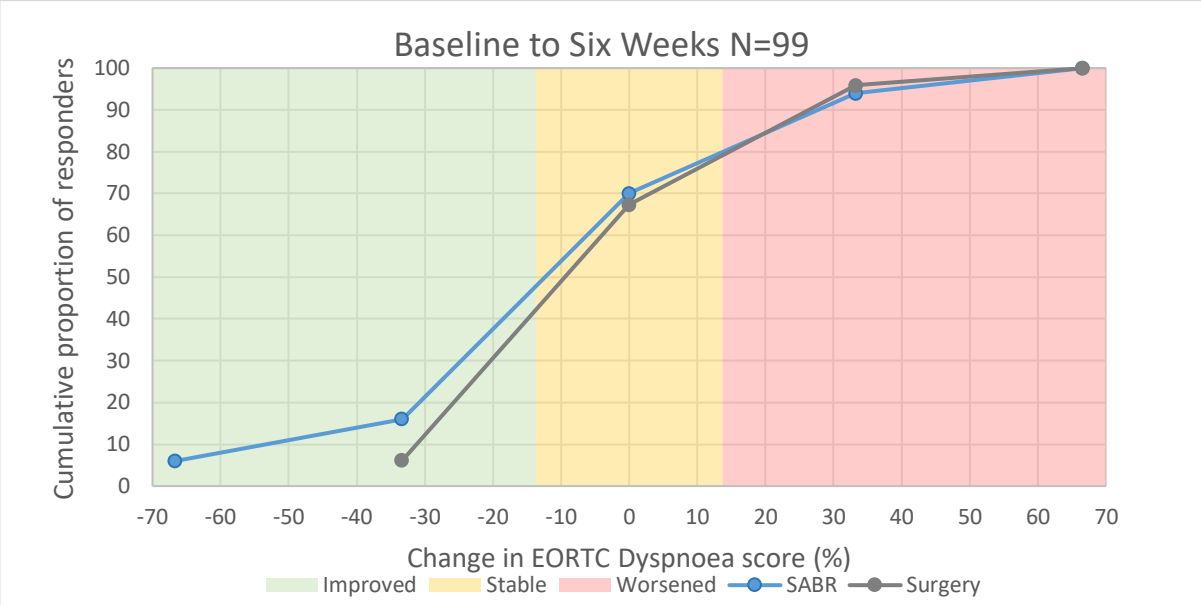
238. Seely AJ, Ivanovic J, Threader J, Al-Hussaini A, Al-Shehab D, Ramsay T, et al. Systematic classification of morbidity and mortality after thoracic surgery. The Annals of thoracic surgery. 2010;90(3):936-42; discussion 42.

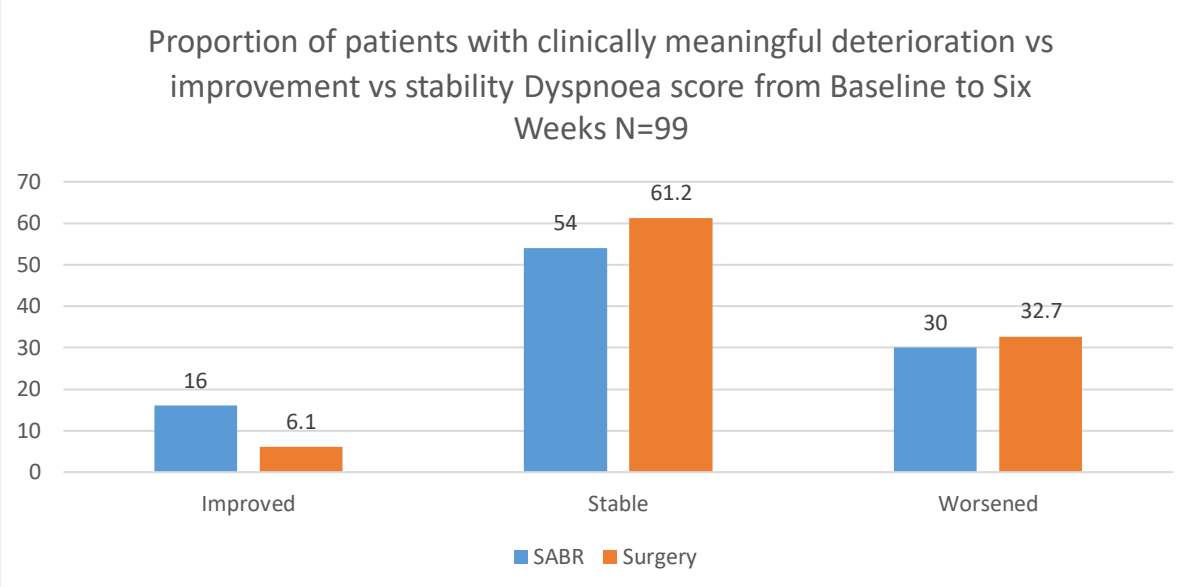
239. Sandri A, Papagiannopoulos K, Milton R, Kefaloyannis E, Chaudhuri N, Poyser E, et al. Major morbidity after video-assisted thoracic surgery lung resections: a comparison between the European Society of Thoracic Surgeons definition and the Thoracic Morbidity and Mortality system. Journal of thoracic disease. 2015;7(7):1174-80.

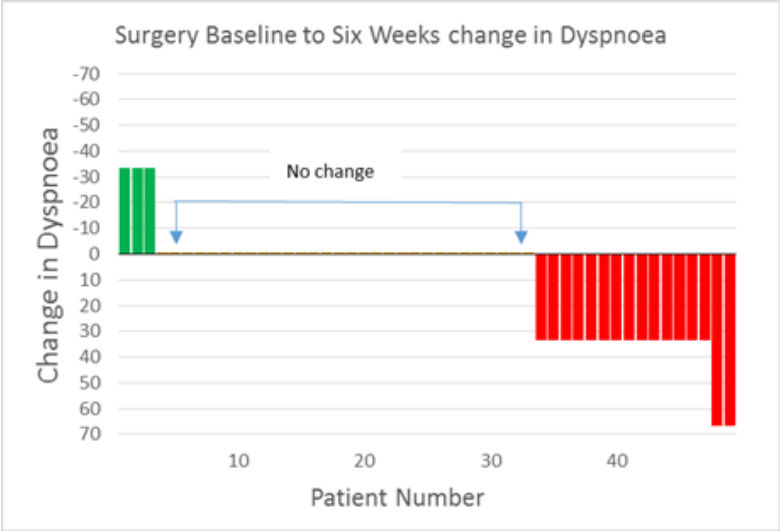
Appendices

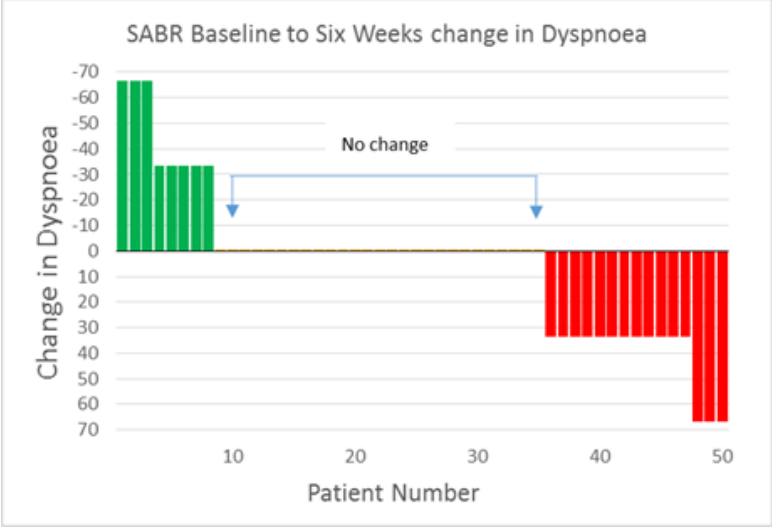
Appendix 1

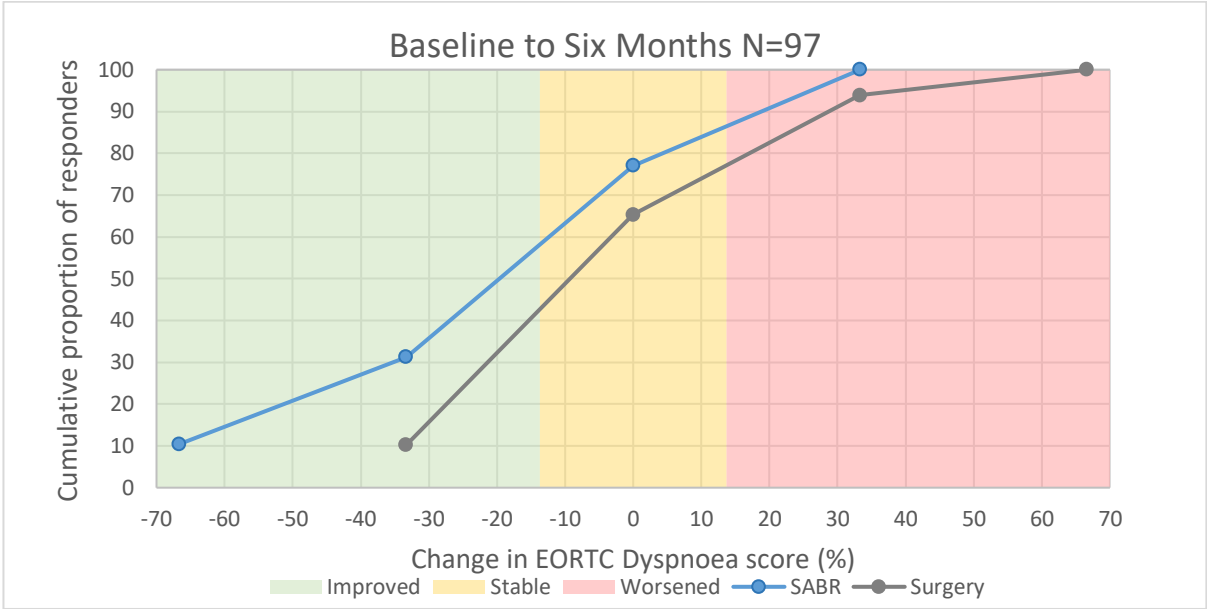
Dyspnoea Responder Analysis

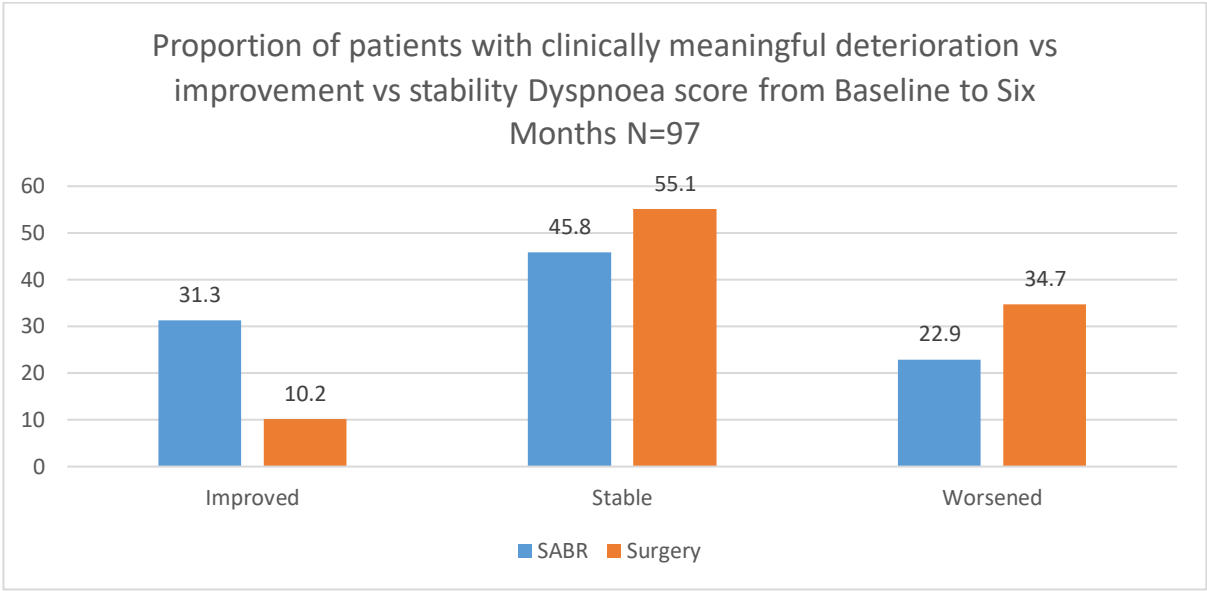





Surgery Baseline to Six Weeks change in Dyspnoea
Change in Dyspnoea
-70
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
70
No change
10
20
30
40
Patient Number


SABR Baseline to Six Weeks change in Dyspnoea
Change in Dyspnoea
-70
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
70
No change
10
20
30
40
50
Patient Number


Baseline to Six Months N=97
Cumulative proportion of responders
100
90
80
70
60
50
40
30
20
10
0
-70
-60
-50
-40
-30
-20
-10
0
10
20
30
40
50
60
70
Change in EORTC Dyspnoea score (%)
Improved
Stable
Worsened
SABR
Surgery


Proportion of patients with clinically meaningful deterioration vs improvement vs stability Dyspnoea score from Baseline to Six Months N=97
60
50
40
30
20
10
0
31.3
10.2
45.8
55.1
22.9
34.7
Improved
Stable
Worsened
SABR
Surgery

Appendix 2

Six Months GH Quality of Life Multivariate Model

Out of 225 patients who participated, 82 patients failed to complete the EORTC QLQ-C30 at the Six month time point and thus these patients' data were excluded from the analysis. Of the 143 remaining patients, a further two patients did not answer either of Q29 and Q30 of the EORTC QLQ-C30 so a Six month Global QoL score could not be calculated. Forty-two of the remaining patients had not completed the Baseline EORTC QLQ-30 and were excluded. A further 5 patients had not completed the Decision Self-efficacy scale, and so multivariable analysis was carried out on 96 patients.

Table 1: Multivariable analysis of the association between demographic, psychosocial and clinical factors and Global QoL at 6 months.

		Univariable			Multivariable		
Variable	**Total participants (*N* = 96) (%)**	**% change in QoL at Six months (95% CI)**		***p*-value**	**% change in QoL at Six months (95% CI)**		***p*-value**
Treatment							
SABR	46 (48)	0			0		
Surgery	50 (52)	21.43	(12.88 - 29.98)	<0.001	8.92	(0.24 - 17.59)	0.04
Baseline QoL	96 (100)	0.63	(0.47 - 0.80)	<0.001	0.52	(0.33 - 0.71)	<0.001*
Age (years)	96 (100)	- 0.18	(-0.77 - 0.41)	0.55	- 0.27	(-0.76 - 0.21)	0.28
Gender							
Female	57 (59)	0			0		
Male	39 (41)	5.23	(-4.46 - 14.92)	0.29	7.12	(-0.91 - 15.16)	0.08
FEV1P	96 (100)	0.27	(0.09 - 0.45)	0.004	0.10	(-0.07 - 0.27)	0.26
CCI	96 (100)	- 6.30	(-10.76 - -1.83)	0.006	- 0.85	(-4.84 - 3.15)	0.68
DSE	96 (100)	0.19	(-0.04 - 0.38)	<0.06	0.02	(-0.14 - 0.18)	0.81

* Statistical significance at <0.01
Performance not included as non-significant when adjusting for Baseline QoL

Baseline Global QoL was significantly associated with Six month Global QoL both in the univariable and multivariable analyses [unadjusted % change in QoL = 0.63, CI = 0.47 to 0.80, $p<0.001$; adjusted % change in QoL = 0.52, CI = 0.33 to 0.71, $p<0.001$].

Six Months SumSc Quality of Life Multivariate Model

Out of 225 patients who participated, 82 patients failed to complete the EORTC QLQ-C30 at the Six month time point and thus these patients' data were excluded from the analysis. Of the 143 remaining patients, a further 15 patients did not answer all thirteen of the EORTC QLQ-C30 function and symptom subscales and so a Six month QoL Summary score could not be calculated. Thirty nine of the remaining patients had not completed the Baseline EORTC QLQ-30 and were excluded and one patient did not answer all thirteen of the EORTC QLQ-C30 function and symptom subscales and so a Baseline QoL Summary score could not be calculated. A further 4 patients had not completed the Decision Self-efficacy scale, and so multivariable analysis was carried out on 84 patients.

QoL SumSc at Six months		**Univariable**			**Multivariable**		
Variable	**Total participants (*N* = 84) (%)**	**Coef (95% CI)**		***p*-value**	**Coef (95% CI)**		***p*-value**
Treatment							
SABR	38 (45)	0			0		
Surgery	46 (55)	14.09	(6.76 - 21.42)	<0.001	1.52	(-4.39 - 7.42)	0.62
Baseline QoL	84 (100)	0.79	(0.63 - 0.96)	<0.001	0.64	(0.46 - 0.83)	<0.001*
Age (years)	84 (100)	- 0.46	(-0.96 - 0.03)	<0.07	- 0.27	(-0.63 - 0.08)	0.13
Gender							
Female	50 (60)	0			0		
Male	34 (40)	4.07	(-3.93 - 12.08)	0.32	4.82	(-0.88 - 10.51)	<0.10
FEV1P	84 (100)	0.21	(0.06 - 0.36)	0.007	0.12	(0.01 - 0.24)	<0.04
CCI	84 (100)	- 7.42	(-10.96 - -3.87)	<0.001	- 2.82	(-5.72 - 0.08)	<0.06
DSE	84 (100)	0.21	(0.04 - 0.39)	<0.02	0.03	(-0.10 - 0.16)	0.69

* Statistical significance at <0.008
Univariate analysis of categorical variables and QoL using Wilcoxon rank-sum test (not shown) and Generalised linear regression (see above)
Univariate analysis of continuous variables and QoL using Generalised linear regression (see above)
Multivariable analysis using Generalised linear regression (see above)
Performance not included in multivariable model as non-significant when adjusting for Baseline QoL

Baseline QoL Summary score was significantly associated with Six month QoL Summary score both in the univariable and multivariable analyses [unadjusted % change in QoL =

0.79, CI = 0.63 to 0.96, p<0.001; adjusted % change in QoL = 0.64, CI = 0.46 to 0.83, p<0.001].

Appendix 3

Factor Analysis Correlation Matrix

CORRELATION	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Q1	1.00	.907	.797	.760	.527	.563	.641	.653	.514	.574	.650
Q2	.907	1.00	.828	.732	.552	.601	.645	.648	.576	.588	.658
Q3	.797	.828	1.00	.804	.583	.706	.610	.712	.637	.651	.594
Q4	.760	.732	.804	1.00	.659	.706	.662	.702	.614	.607	.627
Q5	.527	.552	.583	.649	1.00	.839	.773	.701	.734	.646	.645
Q6	.563	.601	.706	.706	.893	1.00	.759	.750	.835	.714	.681
Q7	.641	.645	.610	.662	.773	.759	1.00	.665	.720	.599	.698
Q8	.653	.684	.721	.702	.701	.750	.665	1.00	.710	.786	.691
Q9	.514	.576	.637	.614	.734	.835	.720	.710	1.00	.637	.651
Q10	.574	.588	.651	.607	.646	.714	.599	.786	.637	1.00	.735
Q11	.650	.658	.594	.627	.645	.681	.698	.691	.651	.735	1.00

The items were correlated with each other as shown however only one exceeded the threshold (none > .9) therefore no evidence of extreme multicollinearity was evident as the determinant of the R-matrix is (R-2.09E-006) which is > 0.00001