FEASIBILITY OF SELECTIVE TARGETED SAMPLING FOR LUNG CANCER

ROUTINE SYSTEMATIC SAMPLING VS. SELECTIVE TARGETED SAMPLING OF LYMPH NODES DURING MEDIASTINAL STAGING: A FEASIBILITY RANDOMIZED CONTROLLED TRIAL

By KERRIE ANN SULLIVAN, BMSc. (Hon)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree: Master of Science

McMaster University © Copyright by Kerrie Ann Sullivan, June 2020

McMaster University MASTER OF SCIENCE (2020) Hamilton, Ontario

(Health Research Methods, Evidence & Impact)

TITLE: Routine Systematic Sampling vs. Selective Targeted Sampling of Lymph Nodes during Mediastinal Staging: A Feasibility Randomized Controlled Trial

AUTHOR: Kerrie Ann Sullivan, BMSc. (Hon)

SUPERVISOR: Dr. Waël C. Hanna, MDCM., MBA., FRCSC.

NUMBER OF PAGES: xiii, 83

LAY ABSTRACT

Before deciding on treatment for patients with lung cancer, a critical step in the investigations is finding out whether the lymph nodes in the chest contain cancer. This is best done with a needle that biopsies those lymph nodes through the walls of the airway, known as endobronchial ultrasound transbronchial needle aspiration. Guidelines require that every lymph node in the chest be biopsied through a process called Systematic Sampling. However, new research has suggested that some lymph nodes may not need a biopsy. These lymph nodes are ones with a very low chance of cancer, based on their imaging tests. In this study, Selective Targeted Sampling was introduced whereby lymph nodes that appeared normal were not initially biopsied. The study followed a feasibility design, which proved sufficient patient interest, adequate safety and possible benefits in pursuing a larger trial comparing Selective Targeted Sampling to Systematic Sampling.

ABSTRACT

Background: The standard of care for mediastinal staging during endobronchial ultrasound (EBUS) is Systematic Sampling (SS) where a minimum of 3 lymph node (LN) stations are biopsied, even if they appear normal on imaging. When LNs appear normal on PET and CT, the Canada Lymph Node Score can also identify if they appear normal on EBUS. For these Triple Normal LNs, the pretest probability of malignancy is < 6%, and routine biopsy may not be required. This preliminary study introduced Selective Targeted Sampling (STS), which omits biopsy of Triple Normal LNs and compared it firsthand to SS.

Methods: A prospective, feasibility RCT was conducted to determine whether the progression of a definitive trial was warranted. Primary outcomes and their progression criterium were recruitment rate (70% acceptable minimum); procedure length (no overlap between sampling methods' 95%CIs); and missed nodal metastasis (overlap between sampling methods' diagnostic accuracy 95%CIs and crossing of the null for the percent difference in diagnosis). cN0-N1 NSCLC patients undergoing EBUS were randomized to the STS or SS arm. Patients in the STS arm were then crossed over to the SS arm to receive standard of care. Wilson's CI method and McNemar's test of paired proportions were used for statistical comparison. Surgical pathology was the reference standard.

Results: Thirty-eight patients met the eligibility criteria, and all were recruited (100%; 95%CI: 90.82 to 100.00%). The median procedure lengths, in minutes, for STS and SS were 3.07 (95%CI: 2.33 to 5.52) and 19.07 (95%CI: 15.34 to 20.05) respectively. STS had a diagnostic accuracy of 100% (95%CI: 74.65% to 100.00%), whereas SS was 93.75% (95%CI: 67.71% to 99.67%) with the inclusion of inconclusive results. Percent difference in diagnosis between sampling method was 5.35% (95%CI: -0.54% to 11.25%).

Conclusion: With the progression criteria successfully met, a subsequent multicentered, non-inferiority crossover trial comparing STS to SS is warranted.

ACKNOWLEDGEMENTS

To my thesis supervisor, Dr. Waël C. Hanna, thank you for the countless hours spent supporting, guiding and mentoring me throughout both my undergraduate and graduate studies. You have played a pivotal role in my professional, academic and personal development.

To Dr. Forough Farrokhyar thank you for your expertise, dedication and continuous patience in teaching me sound statistics and research methodology throughout my graduate studies.

To Dr. Grigorios Leontiadis thank you for your extensive knowledge in research methodology and valuable insight in effectively communicating this thesis' findings.

To Danielle Hylton, thank you for both your insightful advice and dedication to research, which allowed this thesis project to become a reality.

To Yogita Patel, thank you for your thorough clinical research experience, knowledgeable advice and encouragement throughout this process.

To Isabella Churchill, as a fellow graduate student, thank you for your support.

I would also like to thank the Canadian Institutes of Health Research (CIHR) and the Canadian Imperial Bank of Commerce (CIBC) for funding this research endeavor through their student grant programs (CGS-M Program and CIBC Medical Oncology Bursary, respectively).

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

TABLES:

FIGURES:

LIST OF ABBREVIATIONS AND SYMBOLS

AE	Adverse event
BMI	Body mass index
CHS	Central hilar structure
CI	Confidence interval
CLNS	Canada lymph node score
cN	Clinical nodal stage
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus Disease 2019
CT	Computed tomography
EBUS	Endobronchial ultrasound
EBUS-TBNA	Endobronchial ultrasound-guided transbronchial needle aspiration
FN	False negative
FP	False positive
HiREB	Hamilton Integrated Research Ethics Board
IQR	Interquartile range
LLL	Left lower lobe
LN	Lymph node
LUL	Left upper lobe
NPV	Negative predictive value
NSCLC	Non-small cell lung cancer

Ottawa TM&M	Ottawa Thoracic Morbidity & Mortality System
PET	Positron emission tomography
PPV	Positive predictive value
RCT	Randomized controlled trial
REDCap	Research electronic data capture
RLL	Right lower lobe
RML	Right middle lobe
ROSE	Rapid onsite evaluation
RUL	Right upper lobe
SBRT	Stereotactic body radiation therapy
SD	Standard deviation
SJHH	St. Joseph's Healthcare Hamilton
STS	Selective targeted sampling
SS	Systematic sampling
SUV	Standardized uptake value
TN	True negative
TP	True positive

DECLARATION OF ACADEMIC ACHIEVEMENT

Kerrie A. Sullivan was involved in the conception and design of the study, development of all study documents, recruitment of patients, collection of data and performance of statistical analyses. She was the primary author of the subsequent manuscripts published based on this thesis work.

Dr. Waël C. Hanna was the primary supervisor of this thesis. He was involved in the conception and design of the study, screening of eligible patients, planning of relevant statistical analyses, and review of the manuscripts/conference proceedings.

Drs. Forough Farrokhyar and Grigorios Leontiadis provided insightful guidance on the research methodology of this study. They also reviewed its statistical analyses and manuscripts/conference proceedings.

CHAPTER 1: Introduction

Non-Small Cell Lung Cancer (NSCLC) is the deadliest cancer worldwide and accounts for approximately 21,000 deaths yearly in Canada (Bray et al., 2018; Canadian Cancer Statistics Advisory Committee, 2019). These sinister findings are attributed to the asymptomatic nature associated with early stages of the disease. To improve the early detection of NSCLC, Canada and other countries are developing initiatives for the systematic screening of high-risk individuals. With the help of several prognostic trials (Gohagan et al., 2005; Blanchon et al., 2007; Aberle et al., 2011; Pastorino et al., 2012; Horeweg et al., 2013 Cressman et al., 2014; Infante et al., 2015; Wille et al., 2016; Paci et al., 2017), specific factors have been identified in high-risk individuals, regardless of whether they are asymptomatic, and guidelines have been devised to improve their access to early screening. Current guidelines classify high-risk individuals as aged 55-74 years old whom are current smokers or have quit within the last 15 years and have amounted to a smoking history of 30-pack years or more (Moyer et al., 2014; Canadian Task Force on Preventive Health Care, 2016). The anticipated increase in screening has urged the research community to improve the various imaging modalities that assist in the detection, diagnosis and staging of NSCLC. The subsequent section delves into the three most commonly used imaging modalities for dictating treatment decisions and how they specifically contribute to the mediastinal staging of NSCLC patients.

1.1 Mediastinal Staging

In cases of early detection, patients often present to healthcare facilities for other health concerns, and incidentally discover they have pulmonary nodules (Gould et al., 2013). If there is suspicion of malignancy, most patients are further evaluated with a chest computed tomography (CT) scan. Several CT imaging characteristics are capable of discerning malignant and benign nodules. It has been shown that nodules with spiculated or ragged margins; increased size; irregular shape; non-calcification; and presence of a vessel sign or pleural retractions (Xu et al., 2008; Harders et al., 2011) have a greater probability of malignancy and thus, require additional investigation through positron emission tomography (PET) scans. The conjunction of CT and PET scans has the added benefit of providing data on the nodule's metabolic activity through the use of ^{18}F-fluorodeoxyglucose (FDG), a radiolabelled glucose analogue, and its corresponding maximum standardized uptake value (SUVmax). Malignant nodules generally have higher metabolic uptake than their surroundings (O et al., 2016), therefore, combined with the characteristics from CT scans, a more accurate assessment of malignancy can be conducted. If results from these two imaging modalities are indicative of NSCLC, the next step is to determine the most appropriate treatment pathway by mediastinal staging. In recent years, the staging pathway for suspected early-stage NSCLC has involved the diagnostic results of endobronchial ultrasound (EBUS), in addition to those of PET and CT scans. Collectively, these three imaging modalities are used to distinguish patients with resectable NSCLC, by determining if there is the presence of mediastinal nodal disease. The following sections outline the mediastinal assessment for each imaging modality.

1.1.1 Mediastinal Staging: CT

Besides detecting pulmonary nodules, CT scans have the ability to assess nodal disease by measuring the small axis-diameter of lymph nodes (LNs). Specifically, lymphadenopathy is correlated with a greater risk of cancer metastasis (Glazer et al., 1984). Over the years, thresholds for normal LNs have been explored with small-axis diameters ranging from 1.0 to 2.0 cm in size (Glazer et al., 1985; Kiyono et al., 1988; Ziyade et al., 2013). The consensus has been to use a threshold of 1.0 cm for benign LNs, despite the risks associated with false-positive results (Ganeshalingam et al., 2009; Xia et al., 2015). To mitigate these risks, clinicians will often complement the mediastinal results from CT scans with those found on PET.

1.1.2 Mediastinal Staging: PET

The PET portion of a PET/CT scan can complement the LN size assessment of CT by measuring the SUVmax of each LN under investigation. Similar to CT, SUV thresholds for benign LNs have been sought with studies recommending cut-offs between 2.5 and 5.3 (Vansteenkiste et al., 1998; Bryant et al., 2006; Hellwig et al., 2007). Nonetheless, an SUVmax of 2.5 was deemed to be the most appropriate cut-off, because it had the greatest balance between false negative and false positive results (Hellwig et al., 2007). Following PET/CT scan, patients are usually referred to EBUS for histological confirmation of mediastinal staging.

1.1.3 Mediastinal Staging: EBUS

Imaging through EBUS can be combined with biopsy in a procedure known as endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA), to acquire tissue samples of LNs. When pathology is successfully concluded, EBUS-TBNA is much more accurate at mediastinal staging than the other imaging modalities (Yasufuku et al., 2006). The use of EBUS-TBNA in the staging of NSCLC has become widespread since it proved non-inferiority in diagnosis to the more invasive procedure of mediastinoscopy. EBUS-TBNA demonstrated access to the hilar, interlobar and lobar LNs, whereas mediastinoscopy could not (Yasufuku et al., 2011). Since then, EBUS-TBNA has become standard of care in staging NSCLC. Current guidelines recommend that regardless of whether LNs appear benign on PET/CT scan, the lower paratracheal (4R, 4L) and subcarinal (7) nodal stations should be biopsied (De Leyn et al., 2014) through a process known as Systematic Sampling (SS).

1.2 A Need for Further Investigation

SS is a sampling method during EBUS that does not take into consideration the mediastinal findings of PET/CT scans. Consequently, benign mediastinal LNs are often biopsied and lead to inconclusive pathological diagnoses. Such inconclusive results have been found in up to 42.14% of cases (Ortakoylu et al., 2015) and often cause delays in timely treatment, given that they mandate repeat EBUS procedures or a subsequent mediastinoscopy (Jalil et al., 2015). Furthermore, in comparison to PET/CT scans, current EBUS mediastinal staging guidelines do not report any established imaging thresholds that could be used to identify benign LNs. As a result of these limitations to SS, the Canada

Lymph Node Score (CLNS) Project was created by a group of researchers motivated to improve mediastinal staging. The project's first step was to identify ultrasonographic features capable of predicting LN malignancy during EBUS.

1.3 Ultrasonographic Features of Malignancy

Researchers from the CLNS Project identified 13 studies (Fujiwara et al., 2010; Wang Memoli et al., 2011; Schmid-Bindert et al., 2012; Izumo et al., 2014; Jhun et al., 2014; Shafiek et al., 2014; Evison et al., 2015; Gogia et al., 2015; Nakajima et al., 2015; Rozman et al., 2015; Wang et al., 2015; Alici et al., 2016; Ayub et al., 2018) that had examined ultrasonographic features on the basis of whether they were able to discern benign and malignant mediastinal LNs. As a result, the CLNS Project published a systematic review summarizing the current evidence from these studies. They concluded that six ultrasonographic features consisting of: shape, echogenicity, margins, central hilar structure (CHS), central necrosis and small-axis diameter, had the potential to predict nodal malignancy during EBUS. Furthermore, the systematic review identified the next stage of research, which recommended the development and external validation of a diagnostic tool comprising these features (Hylton et al., 2018).

1.4 Validation of a Score

A prospective validation trial was the second step in the CLNS Project. Using both univariate and multivariate logistic regression models, the 6 ultrasonographic features described in the systematic review were analyzed as predictors of LN metastasis. Within the univariate analysis, echogenicity and shape were non-significant in predicting nodal disease ($p \geq 0.05$), therefore, they were excluded from the multivariate model. All remaining

variables were statistically significant ($p < 0.05$) predictors of malignancy in the multivariate model, except for central necrosis ($p = 0.096$). Nonetheless, it was still included within the model because it was viewed as a clinically relevant predictor and the model still demonstrated good discriminatory capability (Hylton & Turner et al., 2020). To that end, the Canada Lymph Node Score was created to assess the probability of malignancy for each LN. Each ultrasonographic LN feature could contribute one point to the score if its malignant form was present (**Table 1)**. Altogether, the study showed that the probability of malignancy increased proportionally with the CLNS. Furthermore, using logistic regression, a threshold was established to discern malignant and benign LNs. It was found that LNs with a CLNS ≥ 2 were deemed malignant, whereas a CLNS < 2 represented benignity (Hylton & Turner et al., 2020). Researchers from the CLNS Project then took the additional measure to externally validate the score across Canada. Endoscopists were assessed on their ability to appropriately differentiate the ultrasonographic features. A learning curve of 300 LNs was found in the study (Hylton & Turner et al., 2020), therefore, an Education Module was designed by the CLNS researchers to adequately train endoscopists in scoring the 4 ultrasonographic features (Hylton & Shargall et al., 2020). With the CLNS developed and validated, the subsequent goal of the CLNS Project was to incorporate the mediastinal findings from PET/CT scans.

1.5 Triple Normal Lymph Nodes

Inconclusive diagnoses may be reduced by omitting biopsies of normal-appearing mediastinal LNs. Given that the CLNS Project was able to identify a threshold of benignity for EBUS, the identification of normal-appearing LNs was simplified. As such, the CLNS Project developed the Triple Normal criteria to easily identify benign mediastinal LNs during EBUS. To uphold the criteria, mediastinal LNs are required to have a small axis diameter < 1 cm on CT, an SUV < 2.5 on PET and a CLNS < 2 during EBUS. A correlational analysis was conducted by the researchers of the CLNS Project and they demonstrated that Triple Normal LNs had a 94.4% (95% Confidence Interval (CI): 89.3% to 97.6%) chance of being truly benign (Hylton et al., Submitted March 2020 to CHEST). This suggests that the practiced standard of SS may be unnecessary and inefficient.

1.6 Rationale for a Feasibility Trial

The CLNS Project successfully developed a diagnostic tool, the CLNS, which is capable of predicting LN malignancy using ultrasonographic features during EBUS. In a separate study, the CLNS Project created the Triple Normal criteria, such that mediastinal findings from PET/CT scans could be incorporated in the EBUS staging process for patients with early stage NSCLC. The next step in continuing the CLNS Project forward was to combine the CLNS and Triple Normal criteria into a novel sampling method, known as Selective Targeted Sampling (STS). STS is a tailored approach of SS for early stage NSCLC, which only biopsies LNs with a high probability of malignancy (CLNS≥2). The goal of STS is to refine the mediastinal staging process by eliminating unnecessary biopsies, in hopes that this could reduce procedure length, inconclusive results, patient

discomfort and healthcare costs. In order for this sampling method to be brought to clinical utility, its diagnostic measures must first be compared directly to the practiced standard, SS. The purpose of this feasibility trial is to obtain preliminary results of this comparison as they relate to patient interest and diagnostic safety. Moreover, given the novelty of STS as a sampling method, the feasibility trial is a way to acquire valuable information regarding its benefits and limitations within a clinical setting.

CHAPTER 2: Methods

This study was a continuation of the CLNS Project and followed the CONSORT 2010 statement: extension to randomised pilot and feasibility trials (Eldridge et al., 2016). **Appendix 2** outlined the complete CONSORT reporting checklist with corresponding pages for each of its item.

2.1 Primary Objectives

Three primary outcomes and their objectives were chosen based on their clinical importance to assess the feasibility of progressing to a large-scale definitive trial. They included the following:

1) **Recruitment Rate:** To determine if there was sufficient patient interest in conducting a trial that compared two different sampling methods, STS and SS, for early stage NSCLC.
2) **Difference in Procedure Length:** To quantitatively investigate a potential clinically relevant benefit of STS when compared to SS.
3) **Missed Nodal Metastasis:** To assess the adequacy of diagnostic safety and potential risk of misdiagnosis between both sampling methods.

2.1.1 Progression Criteria

Each outcome had a progression criterium that, if attained, would favour the continuation of a definitive trial.

1) **Recruitment Rate:** A minimum acceptable recruitment rate of 70% was chosen as the progression criterium. This threshold was based on the current clinical trial

recruitment rates at St. Joseph's Healthcare Hamilton (SJHH), which range from 60%-80% (Hylton & Turner et al., 2020).

2) **Difference in Procedure Length:** STS demonstrating a decreased procedure length with no overlap between the two sampling methods' 95% CIs was chosen as the progression criterium. With this study having a small sample size, the CIs were expected to be wide. Therefore, if the CIs did not overlap despite the imprecision, it was anticipated that an even greater effect could be observed through a definitive trial.
3) **Missed Nodal Metastasis:** STS providing improved or similar diagnostic accuracy and no diagnostic percent difference to SS, as observed by their corresponding 95% CIs, was chosen as the minimum progression criterium. These diagnostic parameters were selected over others, because they were the least likely to be affected by the small sample size. To elaborate further, if any two cells in a crosstabulation were zero, at least one of the following parameters would have been unmeasurable: sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV), because there would be an error in the denominator. If this progression criterium was reached, it would suggest that STS did not have any extreme diagnostic safety risks and clinical equipoise would be established for a subsequent definitive trial.

All components of the progression criteria had to be reached, for a large-scale definitive trial to be justified.

2.2 Secondary Objectives

Secondary outcomes were exploratory in nature and comprised accrual period, prevalence of CLNS values, frequency of biopsies for each sampling method, number of inconclusive biopsies per arm and adverse events (AEs). The secondary outcomes as a whole served the following objectives:

1) To provide further information regarding the study population and STS as a sampling method.
2) To identify study limitations.

2.3 *Study Design & Patient Selection*

In continuation of the CLNS Project, a prospective, patient-blinded, pilot randomized controlled trial (RCT) with a predefined protocol was conducted to determine the feasibility of a large-scale definitive trial comparing STS to SS. NSCLC patients that met the eligibility criteria at the tertiary care center, SJHH, were consecutively enrolled prior to their EBUS procedure. The eligibility criteria involved the inclusion of those patients that were 18 years of age or older; had both pre-treatment PET and CT scans completed; were referred to have EBUS for mediastinal staging of confirmed or suspected NSCLC; and demonstrated clinical nodal stage 0 or 1 (cN0 or cN1) disease based on their PET and CT scans. Those patients who had a combination of peripheral tumours, cN0 disease and tumours less than 2 cm in diameter were excluded because these individuals did not require mediastinal staging. Moreover, evidence of cN2 or higher on PET and CT scans was also an exclusion criterium for patients, because the mediastinal LNs would not fulfill the Triple Normal standard that is at the essence of STS.

2.3.1. Amendment to Protocol

The original eligibility criteria of this study were only inclusive towards cN0 NSCLC patients, as per their pre-treatment PET/CT scans. However, shortly following trial commencement, the prevalence of cN0 patients was found to be considerably low and the accrual period prolonged. To mitigate this issue, the study opened the enrollment to cN1 patients, in addition to those with cN0 disease. The rationale was that cN1 patients would maintain the Triple Normal criteria because their mediastinal LNs would still appear normal on PET/CT scans.

2.4 Randomization

For this study, patients were randomized in a 1:1 ratio to receive either SS alone or STS followed by crossover to SS. An impartial research system analyst computer-generated the randomization sequence with the central web-based Research Electronic Data Capture (REDCap) randomization module (Patridge & Bardyn, 2018). The computer-generated randomization sequence used a random permutated block design, with blocks of varying sizes, to ensure that an approximate equal number of patients were allocated to each treatment arm. All members of the study team were blinded to the randomization sequence until the trial's end. Allocation concealment was maintained by obtaining consent and enrolling patients prior to randomization. Once patients signed consent, the REDCap mobile app was used to assign patients to their respective treatment arm. REDCap was configured to ensure that once patients were initially assigned to a treatment arm, allocation could not be modified.

2.5 Controlling for Bias

Several measures were set in place to minimize the opportunities for bias. With the study following an RCT methodology, one of the major biases that had to be considered was selection bias. Selection bias occurs in RCTs when recruiters selectively enroll patients into the study based on their prior knowledge of the patient's treatment allocation. This bias is often evident when there are no strategies for allocation concealment, or the recruiters are aware of the block sequence in the case of permuted block randomization (Higgins et al., 2019). To avoid the risk of selection bias, this study implemented two safeguards. Firstly, the computer-generated randomization sequence was developed by an impartial research analyst whom did not disclose such sequence to anyone until all patients were randomized. Secondly, the study team obtained written consent and enrolled eligible patients prior to their treatment allocation. Once randomized to a treatment arm, no changes could be made to the allocation.

During the actual EBUS procedure, it is important to consider performance bias. In this study, there were two opportunities in which this bias could arise. Firstly, STS was reliant on the endoscopist's assessment of the CLNS. Therefore, it was imperative that the endoscopist had a sufficient amount of training to correctly identify the four features that comprise the score. To ensure these standards were met, this study's endoscopist had to obtain a passing grade on the CLNS Online Education Module (Hylton & Shargall et al., 2020) prior to study commencement. Lastly, to prevent performance bias in the SS arm, the endoscopist was required to provide adequate nodal samples to pathology. To ascertain the adequacy of the samples, an impartial on-site cytopathologist was present to examine

the cellular material. If the sample was insufficient, the cytopathologist would advise the endoscopist to acquire a maximum of two additional aspirations, as per recommended guidelines (Lee et al., 2008; Hwangbo et al., 2010).

Detection bias is an overarching term for various biases that relate to systematic differences in how outcomes are measured (Higgins et al., 2019). Included within this category are diagnostic review/incorporation and observer biases. Diagnostic review/incorporation bias could have arisen in this study if the surgical pathology (reference test) took into account the diagnostic results from EBUS, when formulating the final diagnosis (Schmidt & Factor, 2013; Kea et al., 2019). To circumvent this opportunity of bias within the study, pathologists did not have access to the EBUS diagnosis when concluding the surgical pathology. With respect to observer bias, this could have occurred if the pathologists were not blinded to the treatment allocation of patients when concluding both the EBUS and surgical pathologies. Consequently, observer bias may have led to inflations in treatment effect estimates (Schulz et al., 1995; Poolman et al., 2007; Hrobjartsson et al., 2012). To address this bias, the pathologists in this study were impartial and blinded to the treatment allocation of patients.

Spectrum bias can develop when indeterminate or ambiguous results are excluded from the study, because the spectrum of disease is not a true reflection of what is found in clinical practice (Hall et al., 2019). To prevent this bias, the study recorded and analyzed all inconclusive biopsies for both treatment arms.

2.6 Sample Size & Recruitment Strategy

One of the primary outcomes of this feasibility study was to recruit at least 70% of the total number of eligible patients. The decision to use 70% as the minimum recruitment rate was based on the existing recruitment standards in place for clinical trials at SJHH. Therefore, assuming a recruitment rate of 70%, an absolute precision of 10% and previous hospital literature reporting approximately 6 EBUS procedures conducted per week at SJHH, a sample size of 53 patients would account for marginal error and produce a 95% CI of ± 10% at the 0.05 level of significance. The sample size was increased to 54 patients to allow for equal allocation of patients to the treatment arms (i.e. 27 patients per arm). This sample size was calculated using 95% CI of proportions with consideration of marginal error. A 10% absolute precision was deemed appropriate provided this was a feasibility RCT that intended to determine whether a larger scale trial was warranted. Patient recruitment began once research ethics approval was obtained. With the understanding that approximately 6 EBUS procedures were conducted each week, it was conservatively anticipated that enrolment of patients would occur over 4-6 months.

2.6.1 Early Stoppage

It was decided to introduce the opportunity for early stoppage ad hoc, due to the possibility of a prolonged accrual period. For early stoppage to be implemented, it was necessary that 70% of the target sample size (n = 54) be successfully recruited. Therefore, the earliest possible stoppage point would occur at 38 eligible patients (70% of the targeted sample size), if they all agreed to partake in the study (100% ongoing recruitment). The rationale for this early

stoppage was that even if the next 16 eligible patients were to decline trial participation, the primary outcome would still have been reached with 70% of patients recruited.

2.7 EBUS Procedure

All patients signed informed consent and were enrolled prior to their EBUS procedure. Patients were brought to the endoscopy suite where they all received deep sedation intravenously as per endoscopic guidelines (Wahidi et al., 2016). Once sedation was optimized, a conventional flexible bronchoscope was passed orally to examine the trancheobronchial tree. Following bronchoscopy, the EBUS procedure was initiated using a convex probe to assess the mediastinal LNs stations as outlined by the International Association for the Study of Lung Cancer LN map (El-Sherief et al., 2014). A minimum of three mediastinal LN stations was examined during procedure, typically stations 4L, 4R and 7. Tumour and additional nodal station assessments were at the discretion of the endoscopist. Static images were taken for each of the examined LNs, whereby the axis diameters (short and long) were clearly marked. It is at this point where the two treatment arms differed:

1) Selective Targeted Sampling: During the LN assessment, the CLNS was applied to each of the examined LNs. Once all four features of the CLNS were assessed, the endoscopist assigned a score from 0 to 4 per LN. Provided that all eligible patients demonstrated normal-appearing mediastinal LNs on PET (SUV < 2.5) and CT (small axis < 1 cm) scans, mediastinal LNs that had a CLNS < 2 fulfilled the Triple Normal criteria. As such they were marked as “Not for Biopsy”,

whereas all other LNs were biopsied using a needle that was passed through the hollow center of the bronchoscope. To ensure standard of care, patients were subsequently crossed over to the SS arm and all Triple Normal LNs were biopsied in a similar fashion.

2) Systematic Sampling: Patients in this arm did not have their LNs assessed by the CLNS. Instead all examined LNs were biopsied regardless of whether they fulfilled the Triple Normal criteria.

All aspirated cellular materials from the biopsies were underwent rapid onsite evaluation (ROSE) by a cytopathologist to determine the adequacy of the sample. The acquisition of static images, assessment of LNs and collection of samples were all conducted by the same endoscopist at SJHH. In the weeks following, the EBUS pathology was obtained in order to compare its findings to those of the CLNS in the STS arm.

2.8 Surgical Pathology

For patients referred to have surgical resection as part of their NSCLC treatment, the gold reference test for their nodal staging was the surgical pathology. The nodal stage was reported as directed by the 8th Edition TNM Lung Cancer Stage Classification (Detterbeck et al., 2017). The surgical pathology was used to determine both the number of patients and their respective LNs that were upstaged, in addition to the diagnostic accuracy of both STS and SS.

2.9 Statistical Analyses

The study followed an intention-to-treat analysis for several reasons. Firstly, it was important that the randomization in this trial be maintained so as to ensure that unknown

confounders and prognostic factors be considered balanced between the two treatment arms. Secondly, with the sample size already considerably small, the exclusion of patients who deviated from protocol would have led to greater uncertainty and imprecision surrounding study results. Lastly, this trial was exploratory in nature given its feasibility design, therefore, such exclusion of patients could have caused exaggerated estimates of the primary outcomes like that of patient interest and safety of both sampling methods (Ranganathan et al., 2016).

Descriptive statistics of both the patients and assessed LNs were reported with categorical data represented as counts and percentages. Continuous variables were recorded as means with a corresponding standard deviation (SD) when normally distributed, whereas skewed variables were recorded as medians with an interquartile range (IQR). Comorbidities known to impact EBUS procedure length or increase the risk of complication, both intra- and post-operatively, were collected for each patient. Such comorbidities included anxiety, asthma, cardiovascular history, chronic pain, chronic obstructive pulmonary disease (COPD), diabetes mellitus, hypertension, obesity (characterized by a body mass index [BMI]≥30) and sleep apnea (American Society of Anesthesiologists, 2002; Pino, 2007; Bader & Pothier, 2009; Long et al., 2012).

2.9.1 Primary Outcomes

Recruitment rate was reported as the percentage of eligible patients that enrolled in the study over the total number of eligible patients. Wilson's 95% CI was used to visually depict the level of imprecision (Wilson, 1927; Newcombe, 1998).

The distribution of procedure length was assessed for each treatment arm. If the variable was normally distributed, an independent two sample t-test was employed with a corresponding 95% CI. Conversely, if normality was violated but the shape distribution between arms was similar, the Mann-Whitney U test (non-parametric equivalent) was used instead with a median and corresponding 95% CI reported as described in Fritz et al. (2012). If the treatment arms were skewed and of different shape distributions, then approximate 95% CIs for the median were calculated using the Hettmansperger-Sheather interpolated order statistics method (Hettmansperger & Sheather, 1986; Nyblom, 1992). Such method relied on ordering the procedure lengths in each arm from shortest to longest, then identifying the upper and lower quantile procedure lengths' whose difference in cumulative distribution function would approximate a probability of 95%. Outlier from both sampling methods were assessed individually to determine their inclusion or exclusion within the analysis. A backwards multiple linear regression model was developed to explore other procedure variables that might impact the duration. Total number of predictor variables included in the model followed the 10 patient per predictor rule devised by Harrell (2015). These predictor variables were removed from the model if they had a $p > 0.1$. Multicollinearity was considered present if the variables had a variance inflation factor (VIF) ≥ 5 and a tolerance ≤ 0.2 (Kutner et al., 2004; O'Brien, 2007; Sheather, 2009). If present, multicollinearity was addressed by either combining the variables or by exclusion of one based on clinical judgement with reasons clearly stated.

Missed nodal metastasis applied to both patients and their respective LNs. Given this study included an RCT and crossover design, both the independent Pearson's chi square test and McNemar's test on paired proportions were attempted, respectively, to determine diagnostic statistics between the two sampling methods. When comparing both arms to surgical pathology independently, the calculations for sensitivity, specificity, NPV, PPV and diagnostic accuracy were attempted with their corresponding 95% CIs. Diagnostic accuracy was defined as the percentage of LNs with the correct diagnosis during EBUS (i.e. true positives and true negatives) when compared to surgical pathology over the total number of LNs assessed by both procedures. In the case where at least one of the sampling methods were unable to calculate sensitivity or specificity, the corresponding 95% CIs for diagnostic accuracy were derived using Wilson's method (Wilson, 1927; Newcombe, 1998). In contrast, for the crossover analysis between STS and SS, the percent difference in diagnosis was calculated using McNemar's test on paired proportions.

2.9.2 Secondary Outcomes

Accrual period was calculated in months by first determining the monthly eligibility rate of patients then dividing the study's sample size by this rate. The accrual period was reported for both the targeted and early stoppage sample sizes, if applicable. Prevalence of cN0-N1 NSCLC was also calculated as a percentage with this outcome. Another prevalence of interest was the percentage of LNs with each CLNS value in this study.

Percentages of inconclusive diagnoses and frequency of biopsies for each sampling method were reported. Both the two-proportions Z test and McNemar's test on paired proportions were attempted based on the data being independent or paired, respectively. The 95% CIs were reported for the percent difference between the two sampling methods.

Patients were assessed both intra- and post-EBUS procedure for AEs using the Ottawa Thoracic Morbidity & Mortality (Ottawa TM&M) System (Seely et al., 2010; Ivanovic et al., 2011). Specifically, all chart notes transcribed from the starting of the patient's EBUS procedure to the end of the 48-hour time period following procedure were reviewed for potential AEs. This timeline was justified based on previous literature demonstrating that most EBUS-related AEs occurred within 24 hours of procedure (Eapen et al., 2013). If identified, AEs were graded as outlined by the Ottawa TM&M System and followed until resolved.

Data that violated the assumptions of the intended statistical analyses were described narratively, using a qualitative approach. Other than the inclusion criteria of the multiple regression model and the normality assumption tests, p-values were not reported because this study was not hypotheses-driven. All statistical analyses were conducted using SPSS Statistics v25.0 (IBM Corp 2017).

2.10 Ethical Considerations

This study was approved by the Hamilton Integrated Research Ethics Board (HiREB) prior to study recruitment (Project Number: 5829) and was registered to ClinicalTrials.gov (NCT03859349). During the course of HiREB submission, ethical

considerations were addressed to ensure that patient risks were minimal throughout the entirety of the trial. Specifically, this study was originally planned as an RCT whereby those patients allocated to the STS arm would not subsequently crossover to the SS arm. Therefore, Triple Normal LNs would not be biopsied. This proposition in study design was based on the preliminary data of the Triple Normal correlational analysis by Hylton et al., (Submitted March 2020 to CHEST), which demonstrated an NPV of 94.4%. This meant that the probability of malignancy for Triple Normal LNs was ~6%. Despite this low chance of malignancy, the ethics review board felt there was too high of a risk for missed nodal metastasis in those patients allocated to the STS arm. Their reasoning was although Triple Normal LNs and the CLNS were assessed separately in previous studies, there was an absence of literature that compared both through STS. As such, they felt that STS was not tested sufficiently enough in clinical setting to understand the full implications of omitting Triple Normal LNs from biopsy. To mitigate this ethical issue and acquire further understanding of the clinical utility of STS, the study design was altered so that those patients in the STS arm would eventually crossover to the SS arm. In doing so, this ensured the following: all patients received standard of care, the risks of participating in this study were minimal, and STS could still be compared directly to SS. From a logistics standpoint, the minimal risks granted the ability to recruit and enroll patients on the same day as their EBUS procedure, which was the only timeline compatible with the endoscopy patient booking process at SJHH. Herein, the findings of the ethically improved feasibility RCT are reported.

CHAPTER 3: Results

A complete CONSORT flow diagram (**Figure 1**) was included to visually demonstrate flow of participants and assessment of LNs for each treatment arm. Between May 2019 and March 2020, 136 consecutive patients referred to the endoscopy clinic under the same endoscopist were screened for possible eligibility. Of these patients, 38 patients met the eligibility criteria and were recruited. Nineteen patients were randomized to the STS arm, whereas the remaining 19 patients were allocated to the SS arm.

3.1 Baseline Demographics

3.1.1 Patient Characteristics

Patient baseline characteristics were categorized by treatment arm; **Table 2** summarizes these findings. In both the STS and SS arms, 19 cN0-N1 patients with suspected or confirmed NSCLC were included (n_{total} = 38). Specifically, both arms had 73.68% (n = 14/arm) of patients appearing with cN0 disease based on PET/CT scans, whereas the remaining 26.32% (n = 5/arm) seemed to demonstrate cN1 disease. The average age of patients in the STS arm was 72.74 ± 9.02 years, with 52.63% (n = 10) being female. Comparatively, patients allocated to SS had an average age of 70.53 ± 12.44 years and 63.16% (n = 12) were female. For STS and SS, most patients were ex-smokers (n = 8, 42.11%; n = 9, 47.37%, respectively) or smokers (n = 9, 47.37%; n = 7, 36.84%, respectively). Patient comorbidities that could impact the level of sedation and risk of complication during EBUS were identified for each treatment arm. They included anxiety, asthma, cardiovascular history, chronic pain, COPD, diabetes mellitus, hypertension and obesity.

3.1.2 Lymph Node Characteristics

From the 38 patients recruited in this study, 118 LNs in total were assessed during EBUS (**Table 3**). Both treatment arms examined an equal number of LNs (n = 59). The most commonly scored nodal stations in the STS arm were 4L (n = 19, 32.20%), 4R (n = 19, 32.20%) and 7 (n = 19, 32.20%). Meanwhile, these nodal stations were also the most biopsied in the SS arm. Specifically, seventeen LNs (28.81%) from 4L, 19 LNs (32.20%) from 4R and another 19 LNs (32.20%) from 7 were sampled in the SS arm. In terms of sampling method diagnosis, STS identified 98.30% (n = 58) of the LNs as benign, whereas 1 LN (1.70%) was recognized as malignant. In contrast, SS identified 81.35% (n = 48) of LNs as benign and 1.70% (n = 1) of LNs as malignant. The remaining 16.95% (n = 10) of LNs sampled in this arm were non-diagnostic as they had inconclusive pathology results. From the 118 LNs assessed during EBUS, 31 (26.27%) were further examined during surgery with STS and SS contributing 15 and 16 LNs, respectively. Surgical pathology confirmed all LNs from the STS arm to be benign, whereas SS had 15 LNs (93.75%) diagnosed as benign and 1 LN (6.25%) proven to be malignant.

3.1.3 Patient Diagnoses

Once mediastinal staging was complete, most patients were referred for surgical resection (n = 19, 50.00%). However, in the STS arm, both surgery (n = 8, 42.11%) and stereotactic body radiation therapy (SBRT; n = 8, 42.11%) were the most commonly referred treatments (**Table 4**). The SS arm had 11 patients (57.90%) referred to surgery and 6 patients (31.58%) referred to SBRT. Only 2 (10.53%) patients

were diagnosed as having benign disease and they were both allocated to the SS arm. The remaining 36 patients were diagnosed with NSCLC. The most common histology for patients randomized to STS was squamous cell carcinoma (n = 10, 52.63%), while most patients in the SS arm had adenocarcinoma (n = 8, 47.06%). Patients commonly presented with a lesion in the right upper lobe (RUL) for both arms (n = 7 [36.84%], n = 8 [42.11%] for STS and SS arms, respectively). Of those patients with malignancy, T2a (n = 5, 26.32%) and T3 (n = 5, 26.32%) were the most prevalent T-stages for STS, while T2a (n = 8, 47.06%) was the most common T-stage for SS (**Table 4**).

3.2 Recruitment Rate

As described in the CONSORT flow diagram (**Figure 1**), 136 patients were screened and 38 of these patients met the eligibility criteria. In terms of recruitment rate, all 38 patients agreed to participate in the trial, and none withdrew at any point in time. Therefore, the recruitment rate was 100% (95% CI: 90.82% to 100.00%). Given that the criteria for early stoppage was fulfilled in that at least 70% of the targeted sample size (n = 54) was recruited, the study terminated recruitment following the 38$_{th}$ patient.

3.3 Procedure Length

Based on visual examination of histograms and the Shapiro-Wilk test of normality, procedure length was found to be skewed for both arms ($p < 0.05$). Median procedure lengths for STS and SS were 3.07 (IQR = 3.19) and 19.07 (IQR = 4.67) minutes, respectively. From the box plots in **Figure 2**, 4 outliers were observed for STS, while SS had one. Given that it is not uncommon for EBUS procedures to last around 30 minutes (Medford et al., 2010; Burkett et al., 2017) and that the STS method biopsied LNs with a

CLNS $\geq$ 2, these cases were justifiably kept in the analysis. Use of the non-parametric Mann-Whitney U test was originally intended to explore the median differences in procedure length between both arms. However, as **Figure 2** demonstrated, STS and SS did not have similar shape distributions, therefore, the medians could not be interpreted using the Mann-Whitney U test. Instead, the proposed Hettmansperger-Sheather interpolated order statistics method was used to derive approximate 95% CIs for median procedure length. With STS demonstrating a median procedure length of 3.07 minutes, the approximation was 95% CI: 2.33 minutes to 5.52 minutes. Meanwhile, SS had a larger median of 19.07 minutes with an approximate 95% CI: 15.34 minutes to 20.05 minutes. The 95% CIs for both sampling method did not overlap, therefore, this trend suggested that the STS arm had a reduced procedure length compared to SS.

3.3.1 Factors Affecting Procedure Length

During EBUS, certain factors other than treatment arm could have affected the procedure length. With the option to biopsy the lung and the number of additional LNs (other than the minimum 3) scored/biopsied being at the discretion of the endoscopist, these factors were explored for possible influence on procedure length. In a backwards multiple linear regression model, three potential predictor variables were entered: 1) treatment arm (i.e. STS or SS), 2) number of LNs scored/biopsied (i.e. 3 or 4 LNs), and 3) lung biopsied (i.e. yes or no). Given the small sample size (n = 38), only these 3 predictor variables were considered, so as to uphold the appropriate number of predictors and minimize the risk of overfitting the model (Harrell et al., 2015). Using the inclusion criterion of a $p \leq 0.1$, lung biopsied was removed ($p = 0.115$), thus

leaving treatment arm and number of LNs scored/biopsied in the model (**Table 5**). Between these two predictor variables, multicollinearity was not observed as both VIF and tolerance were 1.000. Therefore, in addition to treatment arm, the number of LNs scored/biopsied would be important to consider when analyzing the procedure length for a subsequent definitive trial.

3.4 Missed Nodal Metastasis

The number of missed nodal metastases observed in both treatment arms were assessed in two different ways: 1) upstaging per person, 2) upstaging per LN. These categories could be further divided into upstaging comparing each sampling method to surgical pathology (independent analysis) and between STS and SS (crossover analysis).

3.4.1 Upstaging per Person

For both STS and SS, most patients had an N-stage of N0 (n = 10, 52.63%; n = 12, 63.16%, respectively). When both treatment arms were compared to surgical pathology, eight and 11 patients had surgery from the STS and SS arms, respectively (**Table 4**). Of these patients, 3 from the STS arm (37.50%) and 1 from the SS arm (9.09%) were upstaged from N0 to N1 (**Table 5**). However, in all STS cases, nodal involvement was found in N1 LNs that were not assessed during EBUS. Comparatively, the involved nodal station (i.e. 11R) was biopsied during EBUS for the SS case, but the results of the biopsy were inconclusive.

3.4.1.1 Upstaging per Person: Crossover Analysis

The other form of upstaging occurred when patients in the STS arm crossed over to SS. Specifically, two patients from the STS arm (10.52%) were upstaged from N0 to N2 once they crossed over (**Table 6**). This upstaging was further explored in the following section by using LNs as the unit of analysis.

3.4.2 Upstaging per LN

A total of 31 LNs from 19 patients were assessed both during surgery and EBUS (**Table 4**). To compare the surgical pathology (reference standard) diagnoses to those of STS and SS, three crosstabulations were created (**Table 7**). From these tables, the diagnostic accuracy for both sampling methods were determined. Two of these tables (**Table 7A-B**) were dedicated to SS to account for its inconclusive diagnoses in alternative ways. Pearson's chi-squared test was not possible with this data, as multiple cells had values of 0. Nonetheless, the diagnostic accuracy of STS was 100.00% (95% CI: 74.65% to 100.00%) in that all LNs were correctly diagnosed as benign by STS when compared to the ground truth of surgical pathology (**Table 7C**). With there being only true negative diagnoses, the only other measurable diagnostic parameters were specificity and NPV with values of 100.00% (95% CI: 78.20% to 100.00%) and 100.00% (95% CI: unmeasurable), respectively. In contrast, SS had a diagnostic accuracy of 93.75% (95% CI: 67.71% to 99.67%) and 100.00% (95% CI: 73.24% to 100.00%) depending on the inclusion or exclusion of inconclusive diagnoses, respectively. As for other diagnostic parameters, SS with the inclusion of inconclusive results had a sensitivity of 0.00% (95% CI: 0.00% to 97.50%), a specificity of 100.00%

(95% CI: 78.20% to 100.00%) and an NPV of 93.75% (95% CI: 93.75% to 93.75%; **Table 7A**). The exclusion of inconclusive results only allowed specificity and NPV to be quantifiable, with values of 100% (95% CI: 76.84% to 100.00%) and 100.00% (95% CI: unmeasurable), respectively (**Table 7B**). Regardless of inconclusive diagnoses however, both sampling methods demonstrated similar diagnostic accuracy based on the overlap of their 95% CIs.

3.4.2.1 Upstaging per LN: Crossover Analysis

To further elaborate on those patients upstaged from N0 to N2, when crossing over from STS to SS each LN scored and biopsied during EBUS was assessed. Two crosstabulations tables were generated (**Tables 8A-B**) to depict the diagnoses of each sampling method when the inconclusive results of SS were accounted for or not. When accounted for, inconclusive results were combined with the benign diagnoses because staging followed a "benign until proven otherwise" rationale. With the inclusion of inconclusive results STS and SS agreed that 92.86% of LNs (n = 52) were benign and 1.79% of LNs (n = 1) were malignant. Disagreement in diagnoses occurred for 5.35% of LNs (n = 3) because STS originally identified these LNs as benign, whereas SS determined they were malignant (**Table 8A**). These numbers shifted to 91.30% (n = 42), 2.17% (n = 1) and 6.52% (n = 3), respectively, when inconclusive results of SS were removed (**Table 8B**). McNemar's test on paired proportions reported a difference in diagnosis between the sampling methods of 5.35% (95% CI: -0.54% to 11.25%) and 6.52% (95% CI: -0.61% to 13.66%) depending on the

inclusion or exclusion of inconclusive results, respectively. Provided both 95% CIs crossed the null of 0, these estimates would trend towards there being no difference in diagnosis between STS and SS.

3.5 Secondary Outcomes

3.5.1 Accrual Period

Prior to the inclusion of cN1 patients, a total of 55 patients were screened and only 14 met the eligibility criteria (i.e. cN0 disease on pre-treatment PET/CT scans) over a 4.5-month period. This amounted to an eligibility rate of approximately 3 patients/month. Furthermore, the prevalence of eligible patients in the study population was just 25.45%. With these numbers, it would have required 18 months to reach the targeted sample size (n = 54) and 12.67 months to reach the earliest stoppage sample size (n = 38). These projections described accrual periods that were much longer than the anticipated study period of 4-6 months.

Once the eligibility criteria included cN1 patients and the trial implemented early stoppage, the study accrual period was 10 months with 38 eligible patients identified from a total of 136 screened (**Figure 1**). Although the accrual period of this study exceeded the projected timeframe, the eligibility rate was improved to 4 eligible patients/month, the prevalence of eligible patients in the study population increased to 27.94% and the accrual period was decreased by 2.67 months. Therefore, both inclusion of cN1 patients and early stoppage proved to be beneficial protocol amendments that did not hinder the integrity of the study.

3.5.2 Prevalence of Each CLNS

There were five potential CLNS values that each LN could possess. However, in this study only three of these values were identified with 41 LNs (69.49%) possessing a CLNS = 0 and another 13 LNs (22.03%) having a CLNS = 1. The remaining 5 LNs (8.48%) were found to have a CLNS = 2 (**Table 9**).

3.5.3 Frequency of LNs Biopsied

A two-proportions Z test was not possible to compare the frequency of LNs biopsied between both independent sampling methods, because SS had a proportion of 1.

3.5.3.1 Frequency of LNs Biopsied: Crossover Analysis

Prior to patients crossing over from the STS arm to SS, the median number of LN biopsies for STS was 0 per patient (range: 0-2), because most LNs had a CLNS = 0 or 1. By comparison, the median number of LN biopsies in the SS arm was 3 per patient (range: 2-4). **Table 10** illustrated the increase in LN biopsies for patients in the STS arm (n = 5 biopsies) as they crossed over to the SS arm (n = 51 biopsies). With McNemar's test of paired proportions, the difference was 86.44% (95% CI: 77.70% to 95.18%), which indicated a decreased number of LN biopsies in the STS arm.

3.5.4 Inconclusive Biopsies

A two-proportions Z test was not possible to compare the number of inconclusive diagnoses between both independent sampling methods, because STS had a proportion of 0.

3.5.4.1 Inconclusive Biopsies: Crossover Analysis

A total of 10 LNs (16.95%) in the SS had inconclusive EBUS pathology results, whereas 0 LNs were inconclusive based on STS (**Table 3**). By examining those LNs that crossed over from STS to SS, the study demonstrated that STS provided a diagnosis of benignity for 10 LNs (17.86%) that were diagnosed as inconclusive by SS (**Table 3**). Crosstabulation (**Table 11**) with McNemar's test on paired proportions expressed a difference in inconclusive diagnoses of 17.68% (95% CI: 7.83% to 27.89%), thus suggesting that STS had a decreased number of inconclusive results when compared to SS.

3.5.5 Adverse Events

Despite patients exhibiting comorbidities that increased the risk of complications, none of the patients experienced an AE during their EBUS procedure or within the 48-hours following procedure. The absence of AEs further demonstrated the safety of STS in that it did not put patients at a heightened risk of danger compared to SS.

CHAPTER 4: Discussion

4.1 Interpretation of Findings

All primary objectives met their progression criterium, therefore, a definitive large-scale trial is warranted. The minimum acceptable recruitment rate of 70% was surpassed in that all eligible patients (n = 38) were enrolled and none of these individuals withdrew from the study afterwards. With a recruitment rate of 100%, the study convincingly demonstrated patient interest for this field of research. All patients had the study's risks, benefits and responsibilities explained in detail prior to their enrollment, therefore, the high recruitment rate suggests that patients felt comfortable and safe in participating. Safety was further solidified with there being no AEs during and after the patients' EBUS procedures.

To further assess the safety of the trial, the number of missed nodal metastases for each arm was explored. Both STS and SS had similar diagnostic accuracy when compared to surgical pathology, which can be seen from the major overlap in their CIs. In fact, none of the LNs assessed by STS were found to be malignant by surgical pathology. Comparatively, SS had one LN upstaged to malignant because the SS biopsy was inconclusive. For STS, upstaging from N0 to N1 occurred for nodal stations in the hilar/interlobar and peripheral zones (Rusch et al., 2009) that were not assessed during EBUS. This trend is not surprising, as clinical practice guidelines for EBUS mediastinal staging prioritize N2 biopsies (specifically stations 4R, 4L and 7), given their results differentiate between chemotherapeutic and surgical treatment (De Leyn et al., 2014). Nonetheless, it would be advisable for a subsequent trial to record this form of upstaging,

as it could provide insight on whether certain N1 nodal stations should be examined more frequently during EBUS to improve staging.

The other form of upstaging was from N0 to N2, which occurred when STS patients crossed over to SS. Only 3 LNs were upstaged by SS through histologically confirmed pathology. Whether inconclusive results were included or not, the percent difference in diagnosis between the two sampling methods trended towards no difference, because both CIs crossed the null.

Therefore, the outcome as a whole for number of missed nodal metastases between STS and SS met its progression criterium. There were no occasions in which the CIs showed a marked decrease in diagnostic capability for STS. Instead, clinical equipoise was established between the two sampling methods, which justifies the progression of a definitive trial with a non-inferiority design.

The final primary objective was to compare the procedure lengths between the two sampling methods in order to explore a possible benefit of STS and account for any confounding variables during the procedure. Median procedure length for STS was shorter than SS and their corresponding CIs did not overlap. These findings favour that STS may be beneficial in reducing procedure times, which inherently could decrease both the healthcare costs and patient discomfort associated with EBUS mediastinal staging. Since there was no overlap in the CIs, this outcome also met its progression criterium because it provided sufficient incentive to conclusively investigate the trend further through a definitive trial that allows for causal conclusions. In order to fully assess the causal relationship between sampling method and procedure length, other variables impacting

procedure length needed to be accounted for. The multiple linear regression model in this study was intended for this purpose. Results of the model demonstrated that the number of LNs scored/biopsied also affected procedure length, while lung biopsies did not. Therefore, when moving forward with a definitive trial, the number of LNs assessed for each staging method should be reported and considered for procedure length analysis.

The primary objectives were essential in deciding the progression of a definitive large-scale trial, however, the secondary objectives were important in providing novel information regarding the diagnostic characteristics of STS. Moreover, these outcomes proactively identified study design limitations, such that solutions could be devised in advance of a definitive trial. For instance, accrual period was a feasibility outcome that answered a paucity of literature surrounding the prevalence of cN0 patients within the SJHH lung cancer population. It became apparent soon after trial commencement, that the prevalence of cN0 patients within this population was low at roughly 25%. As a result, the accrual period was prolonged and this study necessitated the development of strategies to resolve this concern, which included the introduction of early stoppage and extending the eligibility criteria to cN1 patients.

The prevalence of each CLNS value within the study was a secondary outcome, which provided confidence in the patient eligibility criteria. Specifically speaking, the absence of any LN with a CLNS value of 3 or 4 demonstrated that the study population of cN0-N1 patients had a large number (n = 54, 91.52%) of Triple Normal LNs (CLNS < 2). Given that STS and SS differ in their assessment of Triple Normal LNs, the high percentage reinforced that the patient population was ideal in comparing the diagnostic properties of

both sampling methods. Another point to consider was the likelihood ratios associated with the CLNS values. Hylton & Turner et al., (2020) described a level of uncertainty in determining whether LNs with a CLNS = 1 or 2 were malignant, because both their positive and negative likelihood ratios were close to 1. Comparatively, LNs with a CLNS = 0 had a greater probability of benignity with a positive likelihood ratio of 1 and a sensitivity of 100%. As such, with the majority of scored LNs possessing a CLNS = 0 in this study, it strengthened the diagnostic certainty of STS.

The frequency of biopsies and the number of inconclusive diagnoses between the sampling methods were outcomes that identified additional benefits associated with STS. Biopsies incur additional healthcare costs, because they increase the usage of EBUS equipment, scopes, needles and balloons, while warranting more scope repairs (Wahidi & Yasufuku, 2013; Wahidi et al., 2016). Similarly, previous studies have demonstrated inconclusive biopsy rates as high as 42.14% (Ortakoylu et al., 2015) for SS, which transcends to delays in lung cancer treatment. In this study, STS trended towards a decrease in both these outcomes compared to SS. Therefore, if proven through a definitive trial that STS is non-inferior to SS, assessing these outcomes could provide additional reasons for bringing STS to clinical utility.

One last key observation was the number of lymph nodes that were upstaged from SS to surgery. One lymph node out of 16 was upstaged (**Table 3**), which meant a missed nodal metastasis rate of 6.25%. This rate is very similar to the one found for STS in the Triple Normal study, which was 5.60% (Hylton et al., Submitted March 2020 to CHEST).

These numbers reinforce the concept that although SS is the current practiced standard, it too is not 100% accurate in its diagnosis.

4.2 Limitations

Rather than hypotheses-driven outcomes, the purpose of feasibility trials is to assess outcomes on the basis of whether they justify the progression of a definitive trial and attempt to identify potential limitations ahead of time. The sample sizes of feasibility trials are often not statistically powered to provide conclusive statements regarding outcomes; therefore, interpretation of p-values is considered inappropriate (Eldridge et al., 2016). Instead, the "CONSORT Guidelines for Pilot and Feasibility Trials" recommends the use of descriptive statistics, estimates and CIs when exploring the effect size of outcomes (Eldridge et al., 2016). The intent of reporting CIs is to maintain transparency about the imprecision surrounding results. Therefore, this study's exploratory outcomes (i.e. procedure length, missed nodal metastasis, frequency of biopsies and inconclusive biopsies) must be assessed with a level of uncertainty in that their conclusions are not definitive in nature. Nonetheless, they provide an idea of direction for the outcomes moving forward and establish clinical equipoise. Additionally, for outcomes whereby the CIs of both treatment arms did not overlap (e.g. procedure length), the size of effect may increase with the implementation of a statistically powered trial, because the CIs are expected to be more precise.

Along these lines, sample size is an area of limitation requiring elaboration. By implementing early stoppage after the recruitment of 38 patients, accrual period was successfully shortened, however, it inadvertently increased the difficulty of statistical

analyses. This is due in part to the increased risk of the data violating the normality assumption, which is a requirement for parametric tests (Altman et al., 1995). As such, non-parametric tests were considered, which increased the difficulty of providing valid effect estimates with easily interpretable CIs (Perme et al., 2019). The decreased sample size also prevented the opportunity of independent sample tests like Pearson's chi-square or the two-proportion Z-test, because there was an increased likelihood of cell counts equal to 0 and invalid proportions (i.e. 1 and 0), respectively. Consequently, this study was limited in the diagnostic statistics it could calculate between STS and SS, because there was an absence of true positive, false positive and false negative diagnoses. Moreover, the sample size limited the number of predictor variables allowed in the procedure length multiple linear regression model. This limitation omitted the possibility to explore interaction terms, which could have provided more insight on whether treatment arm impacted the other predictor variables' effects differently. Despite these limitations in reporting relevant information, it was still possible to convey the outcomes' findings in meaningful ways such that informed decisions could be made with respect to the progression of a definitive trial.

This study was the first of its kind in the CLNS Project to prospectively restrict the eligibility criteria to only cN0 NSCLC patients. Therefore, there was minimal literature to fully anticipate the study accrual period of cN0 NSCLC patients at SJHH. After 4.5 months of screening with a limited number of patients meeting the eligibility criteria, it became apparent that accrual period was an unexpected limitation of the trial. To shorten this period, ad hoc amendments to protocol saw the inclusion of cN1 patients and the opportunity for early stoppage if recruitment achieved its progression criterium. Both

amendments did reduce the accrual period to 10 months, however it was still a considerably long timeframe for just 38 eligible patients. Therefore, a subsequent trial should follow the amended eligibility criteria, but it should also consider other strategies to reduce the accrual period.

From the descriptive statistics of this study, an interesting observation was found in the number of patients (n = 14, 36.84%) referred to SBRT following benign EBUS pathology. This came as a surprise because it was anticipated that most patient with benign results would receive surgery. However, in the STS arm both surgery and SBRT were tied as the most commonly referred treatment (**Table 4**). Reasons for referral to SBRT over surgery included patient preference or multiple comorbidities resulting in failed pulmonary function tests, which contraindicated the patient's surgical candidacy. With many patients undergoing SBRT instead of surgery, this limited the number of LNs that were assessed between surgical pathology (reference standard) and EBUS diagnosis. Specifically, out of the 118 LNs examined during EBUS, only 31 LNs (26.27%) were also assessed during surgery (**Table 3**). The substantial decrease in LNs diagnosed by both EBUS and surgical pathology highlighted an inevitable bias within this study. Partial verification bias (also denoted verification bias) was clearly introduced because not all patients received the reference standard. Only patients with benign mediastinal (N2) LNs during EBUS underwent surgery, whereas patients with malignant N2 LNs were referred to chemoradiation. The absence of true and false positives further demonstrates the impact of this bias, as only the specificity and NPV were consistently quantifiable. Sensitivity was only possible when the SS arm included inconclusive results within the analysis, however,

it was very imprecise with its 95% CI covering values from 0% to 97.50%. Unfortunately, given the nature of mediastinal staging, there is no full-proof way to prevent partial verification bias. Usually patients with malignant mediastinal LNs identified during EBUS benefit most from chemoradiation rather than surgery (De Leyn et al., 2014). Therefore, subjecting these patients to surgery as an adjunct could raise ethical concerns, because it could cause detrimental delays in receiving more appropriate treatment. Despite this inability to fully resolve partial verification bias, there are a couple ways that the definitive trial could mitigate its effects. Firstly, as noted by the patients that were upstaged from N0 to N1 through surgery in this study, the malignant LNs were from N1 stations not assessed during EBUS. It was previously mentioned that it would be important for a definitive trial to report these cases, but it may also be beneficial for the trial endoscopists to sample more N1 stations during EBUS. The rationale being that patients with N1 nodal disease are still eligible for surgical resection (Asamura et al., 2015), therefore, N1 nodal station assessments during EBUS could contribute to true and false positive diagnoses when analyzing the data with surgical pathology. From a statistical standpoint, estimates of diagnostic statistics can account for partial verification bias and be corrected by using multiple imputation (Harel & Zhou, 2006) or the Begg and Greenes method (Begg & Greenes, 1983; de Groot et al., 2011). In general, the definitive trial should consider deriving their sample size calculations on the number of LNs that are assessed by both EBUS and surgical pathology, instead of EBUS alone. Such consideration would ensure that the trial has sufficient statistical power to discern clinically important diagnostic

differences between STS and SS when the partial verification bias correction methods are implemented.

Notwithstanding, this feasibility trial did attempt to take extra precautionary measures against methodological biases that could be avoided. Patients were blinded to whether they received STS; the study did not assess any patient-reported outcome; and the pathology from EBUS and surgery were ascertained by impartial pathologists that were blinded to treatment allocation. Furthermore, all research personnel were blinded to the randomization sequence and allocation concealment was maintained using an electronic randomization module. Nevertheless, the statistical analyses were conducted by an assessor that was not blinded to the treatment allocations. Most outcomes were determined from objective clinical judgement and the study did not make any definitive conclusions regarding the exploratory outcomes, however, there still risks the possibility of detection bias (Higgins et al., 2019). With the subsequent trial being hypothesis-driven in nature, detection bias should be avoided. Therefore, it is suggested that all statistical analyses be undertaken by an impartial biostatistician.

4.3 Implication for Future Trial

The subsequent definitive trial will adopt specific aspects of this study, but it will also incorporate some reasonable changes to address identified study limitations.

4.3.1 Crossover Design

Originally, this feasibility trial was intended to follow a strict RCT design, whereby STS and SS were only assessed as independent samples. However, as stated previously, REB was not accepting of this design because the clinical implications of

STS had not been explored thoroughly. Consequently, the study adapted a crossover design, which ensured that all patients in the STS arm would eventually receive standard of care (i.e. SS). This methodological amendment happened to improve recruitment rate, as it minimized the perceived risks of partaking in the study. Moreover, the crossover of patients from STS to SS allowed the study to conduct statistical analyses on both paired and independent samples. As was seen with the results, the paired data were more feasible to compare than the independent data, because they did not violate statistical assumptions. Therefore, it is recommended that the subsequent trial adapt a complete crossover design. All patients would start off in the STS arm then they would cross over to the SS arm, whereby patients would serve as their own control. Fittingly, this study design was used for the reputable trial which demonstrated non-inferiority between EBUS-TBNA (i.e. SS) and mediastinoscopy (Yasufuku et al., 2011). Through this trial, SS became the new practiced standard for mediastinal staging. Therefore, it would be sensible to adopt a similar study design for a definitive trial that aims to further improve the mediastinal staging process.

4.3.2 Multicentered

One of the more prominent limitation of this study was the prolonged accrual period. Although strategies were set in place to reduce this period throughout the study, it is still a limitation that requires consideration for the subsequent trial. In addition to adopting the amended eligibility criteria, the accrual period could be reduced by making the definitive trial multicentered. Not only will this provide access to a greater number of eligible patients, it will also improve the generalizability of results by

recruiting patients from various lung cancer populations. Fortunately, the CLNS Project has already conducted a pan-Canadian trial, which had multiple endoscopists from different provinces assess the CLNS and successfully pass its Online Education Module (Hylton & Shargall et al., 2020; Hylton & Turner et al., 2020). As such, there would be a sufficient number of endoscopists proficient in assessing LNs with STS. Possible subsites could include Toronto General Hospital, Health Sciences Centre (Winnipeg), Royal Alexandra Hospital (Edmonton), Centre Hospitalier de L'Université de Montréal and McGill University Health Centre (Montréal).

4.3.3 Non-Inferiority Margin

Many diagnostic statistics, including the margin between both sampling methods' NPVs, were unable to be calculated due to partial verification bias and the small number of LNs assessed by both surgery and EBUS. Consequently, estimations regarding a possible non-inferiority margin for the subsequent trial were not possible. To that end, alternative means to determine an appropriate non-inferiority margin should be explored. Such margin should be established a priori to the subsequent trial with its rationale clearly stated (Piaggio et al., 2012). A possible avenue to explore is the minimal clinically important difference (MCID) of diagnostic tests in the field of mediastinal staging. This MCID could be identified using the Delphi method (Dalkey & Helmer, 1963; Dalkey et al., 1969) whereby focus groups of mediastinal staging experts and patients would congregate to deduce a meaningful non-inferiority margin with the help of relevant literature such as the National Cancer Care Network guidelines (National Comprehensive Cancer Network, 2018).

4.3.4 Per-Protocol Analysis

Although there were no patients that deviated from protocol, this feasibility study decided a priori to follow an intention-to-treat analysis to account for randomization, small sample size and to avoid overestimates of the primary outcomes. However, for a definitive multicentered non-inferiority crossover trial, it is highly advised that a per-protocol analysis be followed instead of intention-to-treat. The rationale for this change in analysis is two-tiered. First, the definitive trial is to adopt a complete crossover design with patients serving as their own control, therefore, randomization is no longer necessary to ensure balanced baseline characteristics. Furthermore, for non-inferiority trials, a per-protocol analysis is more conservative with its treatment effect estimates. Specifically, such analysis is better at identifying inferiority because intention-to-treat analyses do not exclude deviated patients who can bias the two treatments in appearing similar (D'Agostino et al., 2003). Therefore, if non-inferiority between the two sampling methods could be established with the more conservative per-protocol analysis, it would solidify the diagnostic relationship between STS and SS.

4.3.5 Coronavirus Pandemic

The world has seen dramatic changes due to the Coronavirus Disease 2019 (COVID-19) pandemic, especially the healthcare field. Policies and regulations for hospitals will likely be altered for a very long time, if not permanently. Short-term wise, these new trends could cause delays in recruitment for the definitive trial as many endoscopic services have been disrupted and most non-COVID related research has

been put at a standstill. However, with Canada slowly introducing the reopening stages of its economy, research is expected to start back up with precautionary measures set in place. Long-term wise and amidst the development of a vaccine, COVID-19 is expected to undergo a second and possibly a third wave of infection where countries will likely see a spike in cases (Moore et al., 2020; Vaid et al., 2020). To reduce the incidence rate from these waves, hospitals are recommended to maintain stringent policies whereby patients, if possible, limit the time spent in hospitals (St. Joseph's Healthcare Hamilton, 2020; University Health Network: Patient Relations and Patient Education & Engagement, 2020). For these reasons, the definitive trial could be beneficial for early NSCLC patients and stakeholders alike as its goal is in line with COVID policies, in that it aims is to introduce a novel sampling method with suspected reductions in procedure length, patient discomfort and healthcare costs. The potential for the definitive trial to improve COVID-policy for mediastinal staging is an important and relevant component of its rationale to be considered for ethic review boards, patients and stakeholders.

4.3.6 Other Considerations

Diagnostic statistics, procedure length, prevalence of CLNS values, frequency of biopsies and inconclusive diagnoses would all be valuable study outcomes to include in the definitive multicentered non-inferiority crossover trial, because causal relationships could be concluded regarding the potential benefits of STS. Nonetheless with the trial adopting a complete crossover design, all the data will be paired, therefore, special considerations should be taken when assessing the various predictors

of procedure length, specifically with respect to any regression analysis (e.g. paired differences). Another concept to consider for the definitive trial is a cost analysis between STS and SS. This study's findings showed potential in STS reducing healthcare costs, as the sampling method indicated slight decreases in procedure length, frequency of biopsies and inconclusive diagnoses. For that reason, reporting the costs per sampling method as they relate to EBUS equipment, scope repairs and services would be advantageous in assessing the economic implications.

Lastly, **Table 12** outlines the diagnostic parameters intended for the definitive trial and the possible clinical consequences associated with false positive, false negative and inconclusive LN diagnoses from both sampling methods.

4.4 Generalizability

The population of interest for this study was specific towards those patients with early stage lung cancer. Although the study's eligibility criteria were strict in only enrolling cN0-N1 patients, this stringency was purposefully implemented for the following reason. Hylton & Turner et al., (2020) had already validated the CLNS across all stages of NSCLC. Therefore, the next step was to target the lung cancer population that would benefit the most from the CLNS through STS, which was evidently those with early stages of the disease. Specifically, SS has demonstrated drawbacks (i.e. inconclusive results, false negatives) in successfully staging patients with normal-appearing mediastinum on pre-treatment PET/CT scans (Shingyoji et al., 2014; Ong et al., 2015; Dooms et al., 2015). Moreover, with the advancements in lung cancer screening, it is expected that more patients

will be identified with cN0-N1 disease on pre-treatment imaging modalities. As such, improvements in staging this specific, yet important, population are needed.

This study was only conducted at one tertiary cancer facility, so there is always the possibility of variation in results and methodology for other healthcare sites. Moreover, they may very well differ in their prevalence of cN0-N1 NSCLC patients. Nonetheless, this study could serve as a template of general strategies to reduce study limitations for other feasibility trials with different designs and settings.

With respect to moving forward with a multicentered crossover trial, subsites may be confronted by limitations that were not identified in this study. For instance, SJHH was fortunate to have ROSE during each EBUS procedure, such that performance bias could be minimized. Unfortunately, not all healthcare facilities have the resources to access this service, therefore, alternative means to prevent performance bias should be explored. If a subsite does not have access to ROSE, past literature has shown that two to three aspirations per LN yields the best possibility of conclusive diagnosis (Lee et al., 2008; Hwangbo et al., 2010). This could be a standard implemented within the definitive trial to minimize performance bias.

Although this study was considered preliminary in nature, the study design recommendations for the definitive trial included much consideration for the generalizability of its results. Hence, it is believed that the multicentered component of the definitive trial will provide an accurate assessment of mediastinal staging for the Canadian cN0-cN1 NSCLC population.

CHAPTER 5: Conclusion

This study was a feasibility RCT comparing two sampling methods, which successfully met the progression criteria for a definitive large-scale trial. Through the exploration of various outcomes, study limitations were identified and addressed with possible solutions. With the advancement of a subsequent trial, this study recommends that it follow a per-protocol multicentered, non-inferiority crossover design, whereby the eligibility criteria include both cN0 and cN1 NSCLC patients as determined by their pre-treatment PET/CT scans.

WORKS CITED:

Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., … Sicks, J. R. D. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, *365*(5), 395–409. https://doi.org/10.1056/NEJMoa1102873

Alici, I. O., Yılmaz Demirci, N., Yılmaz, A., Karakaya, J., & Özaydın, E. (2016). The sonographic features of malignant mediastinal lymph nodes and a proposal for an algorithmic approach for sampling during endobronchial ultrasound. *Clinical Respiratory Journal*, *10*(5), 606–613. https://doi.org/10.1111/crj.12267

Altman, D. G., & Bland, J. M. (1995). Statistics notes: The Normal distribution. *BMJ*, *310*(6975), 298. https://doi.org/10.1136/bmj.310.6975.298

American Society of Anesthesiologists. (2002). Practice Guidelines for Sedation and Analgesia by Non-Anesthesiologists. *Anesthesiology*, *96*(2), 1004–1017. https://doi.org/10.1097/00000542-200204000-00031

Asamura, H., Chansky, K., Crowley, J., Goldstraw, P., Rusch, V. W., Vansteenkiste, J. F., … Rami-Porta, R. (2015). The International Association for the Study of Lung Cancer Lung Cancer Staging Project: Proposals for the Revision of the N Descriptors in the Forthcoming 8th Edition of the TNM Classification for Lung Cancer. *Journal of Thoracic Oncology*, *10*(12), 1675–1684. https://doi.org/10.1097/JTO.0000000000000678

Ayub, I. I., Mohan, A., Madan, K., Hadda, V., Jain, D., Khilnani, G. C., & Guleria, R. (2018). Identification of specific EBUS sonographic characteristics for predicting benign mediastinal lymph nodes. *Clinical Respiratory Journal*, *12*(2), 681–690. https://doi.org/10.1111/crj.12579

Bader, A. M., & Pothier, M. M. (2009). Out-of-Operating Room Procedures: Preprocedure Assessment. *Anesthesiology Clinics*, *27*(1), 121–126. https://doi.org/10.1016/j.anclin.2008.10.001

Begg, C. B., & Greenes, R. A. (1983). Assessment of Diagnostic Tests When Disease Verification is Subject to Selection Bias. *Biometrics*, *39*(1), 207–215. https://doi.org/10.2307/2530820

Blanchon, T., Bréchot, J.-M., Grenier, P. A., Ferretti, G. R., Lemarié, E., Milleron, B., … Flahault, A. (2007). Baseline results of the Depiscan study: A French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). *Lung Cancer*, *58*(1), 50–58. https://doi.org/10.1016/j.lungcan.2007.05.009

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *0*(0), 1–31. https://doi.org/10.3322/caac.21492

Bryant, A. S., Cerfolio, R. J., Klemm, K. M., & Ojha, B. (2006). Maximum Standard Uptake Value of Mediastinal Lymph Nodes on Integrated FDG-PET-CT Predicts Pathology in Patients with Non-Small Cell Lung Cancer. *Annals of Thoracic Surgery*, *82*(2), 417–423. https://doi.org/10.1016/j.athoracsur.2005.12.047

Burkett, A., Sekhon, H. S., Burkett, C., Hakim, S. W., & Amjadi, K. (2017). An Algorithmic Approach for Assessment of Mediastinal Lesions Using Conventional Transbronchial Needle Aspiration and Endoscopic Ultrasonography in a Single Procedure. *Canadian Respiratory Journal*, *2017*, 1–7. https://doi.org/10.1155/2017/1971629

Canadian Cancer Statistics Advisory Committee. (2019). Canadian Cancer Statistics 2019. In *Canadian Cancer Statistics 2019.*

Canadian Task Force on Preventive Health Care. (2016). Recommendations on screening for lung cancer. *CMAJ*, *188*(6), 425–432. https://doi.org/10.1503/cmaj.151421

Cochrane Handbook for Systematic Reviews of Interventions. (2019). In J. P. T. Higgins, J. Thomas, J. Chandler, T. Li, M. Page, & V. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (2nd ed.). https://doi.org/10.1002/9781119536604

Cressman, S., Lam, S., Tammemagi, M. C., Evans, W. K., Leighl, N. B., Regier, D. A., … Peacock, S. J. (2014). Resource utilization and costs during the initial years of lung cancer screening with computed tomography in Canada. *Journal of Thoracic Oncology*, *9*(10), 1449–1458. https://doi.org/10.1097/JTO.0000000000000283

D'Agostino, R. B., Massaro, J. M., & Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues - The encounters of academic consultants in statistics. *Statistics in Medicine*, *22*(2), 169–186. https://doi.org/10.1002/sim.1425

Dalkey, N. (1969). An experimental study of group opinion: The Delphi method. *Futures*, *1*(5), 408–426. https://doi.org/10.1016/S0016-3287(69)80025-X

Dalkey, N., & Helmer, O. (1963). An Experimental Application of the DELPHI Method to the Use of Experts. *Management Science*, *9*(3), 458–467. https://doi.org/10.1287/mnsc.9.3.458

de Groot, J. A. H., Janssen, K. J. M., Zwinderman, A. H., Bossuyt, P. M. M., Reitsma, J. B., & Moons, K. G. M. (2011). Correcting for Partial Verification Bias: A Comparison of Methods. *Annals of Epidemiology*, *21*(2), 139–148. https://doi.org/10.1016/j.annepidem.2010.10.004

De Leyn, P., Dooms, C., Kuzdzal, J., Lardinois, D., Passlick, B., Rami-Porta, R., … Zielinski, M. (2014). Revised ests guidelines for preoperative mediastinal lymph node staging for non-small-cell lung cancer. *European Journal of Cardio-Thoracic Surgery*, *45*(5), 787–798. https://doi.org/10.1093/ejcts/ezu028

Detterbeck, F. C., Boffa, D. J., Kim, A. W., & Tanoue, L. T. (2017). The Eighth Edition Lung Cancer Stage Classification. *Chest*, *151*(1), 193–203. https://doi.org/10.1016/j.chest.2016.10.010

Dooms, C., Tournoy, K. G., Schuurbiers, O., Decaluwe, H., De Ryck, F., Verhagen, A., … De Leyn, P. (2015). Endosonography for mediastinal nodal staging of clinical N1 non-small cell lung cancer: A prospective multicenter study. *Chest*, *147*(1), 209–215. https://doi.org/10.1378/chest.14-0534

Eapen, G. A., Shah, A. M., Lei, X., Jimenez, C. A., Morice, R. C., Yarmus, L., … Ost, D. E. (2013). Complications, consequences, and practice patterns of endobronchial ultrasound-guided transbronchial needle aspiration: Results of the AQuIRE Registry. *Chest*, *143*(4), 1044–1053. https://doi.org/10.1378/chest.12-0350

El-Sherief, A. H., Lau, C. T., Wu, C. C., Drake, R. L., Abbott, G. F., & Rice, T. W. (2014). International Association for the Study of Lung Cancer (IASLC) lymph node map: Radiologic review with CT illustration. *Radiographics*, *34*(6), 1680–1691. https://doi.org/10.1148/rg.346130097

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., … Tugwell, P. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *Pilot and Feasibility Studies*, *355*(i5239), 1–29. https://doi.org/10.1186/s40814-016-0105-8

Evison, M., Morris, J., Martin, J., Shah, R., Barber, P. V., Booton, R., & Crosbie, P. A. J. (2015). Nodal staging in lung cancer: A risk stratification model for lymph nodes classified as negative by EBUS-TBNA. *Journal of Thoracic Oncology*, *10*(1), 126–133. https://doi.org/10.1097/JTO.0000000000000348

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Fujiwara, T., Yasufuku, K., Nakajima, T., Chiyo, M., Yoshida, S., Suzuki, M., … Yoshino, I. (2010). The utility of sonographic features during endobronchial ultrasound-guided transbronchial needle aspiration for lymph node staging in patients with lung cancer: A standard endobronchial ultrasound image classification system. *Chest*, *138*(3), 641–647. https://doi.org/10.1378/chest.09-2006

Ganeshalingam, S., & Koh, D.-M. (2009). Nodal staging. *Cancer Imaging*, *9*(1), 104–111. https://doi.org/10.1102/1470-7330.2009.0017

Glazer, G. M., Gross, B. H., Quint, L. E., Francis, I. R., Bookstein, F. L., & Orringer, M. B. (1985). Normal mediastinal lymph nodes: Number and size according to American Thoracic Society mapping. *American Journal of Roentgenology*, *144*(2), 261–265. https://doi.org/10.2214/ajr.144.2.261

Glazer, G. M., Orringer, M. B., Gross, B. H., & Quint, L. E. (1984). The mediastinum in non-small cell lung cancer: CT-surgical correlation. *American Journal of Roentgenology*, *142*(6), 1101–1105. https://doi.org/10.2214/ajr.142.6.1101

Gogia, P., Insaf, T. Z., McNulty, W., Boutou, A., Nicholson, A. G., Zoumot, Z., & Shah, P. L. (2015). Endobronchial ultrasound: Morphological predictors of benign disease. *ERJ Open Research*, *2*(1), 1–8. https://doi.org/10.1183/23120541.00053-2015

Gohagan, J. K., Marcus, P. M., Fagerstrom, R. M., Pinsky, P. F., Kramer, B. S., Prorok, P. C., … Zylak, C. (2005). Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. *Lung Cancer*, *47*(1), 9–15. https://doi.org/10.1016/j.lungcan.2004.06.007

Gould, M. K., Donington, J., Lynch, W. R., Mazzone, P. J., Midthun, D. E., Naidich, D. P., & Wiener, R. S. (2013). Evaluation of individuals with pulmonary nodules: When is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians' evidence-based clinical practice guidelines. *Chest*, *143*(5 Suppl), e93S-e120S. https://doi.org/10.1378/chest.12-2351

Hall, M. K., Kea, B., & Wang, R. (2019). Recognising Bias in Studies of Diagnostic Tests Part 1: Patient Selection. *Emergency Medicine Journal*, *36*(7), 431–434. https://doi.org/10.1136/emermed-2019-208446

Harders, S. W., Madsen, H. H., Rasmussen, T. R., Hager, H., & Rasmussen, F. (2011). High resolution spiral CT for determining the malignant potential of solitary pulmonary nodules: Refining and testing the test. *Acta Radiologica*, *52*(4), 401–409. https://doi.org/10.1258/ar.2011.100377

Harel, O., & Zhou, X. H. (2006). Multiple imputation for correcting verification bias. *Statistics in Medicine*, *25*(22), 3769–3786. https://doi.org/10.1002/sim.2494

Harrell, F. E. J. (2015). Multivariable Modeling Strategies. In *Regression Modeling Strategies* (Springers, pp. 63–102). https://doi.org/10.1007/978-3-319-19425-7_4

Hellwig, D., Graeter, T. P., Ukena, D., Groeschel, A., Sybrecht, G. W., Schaefers, H.-J., & Kirsch, C.-M. (2007). 18F-FDG PET for mediastinal staging of lung cancer: Which SUV threshold makes sense? *Journal of Nuclear Medicine*, *48*(11), 1761–1766.

Horeweg, N., Van Der Aalst, C. M., Vliegenthart, R., Zhao, Y., Xie, X., Scholten, E. T., … De Koning, H. J. (2013). Volumetric computed tomography screening for lung cancer: Three rounds of the NELSON trial. *European Respiratory Journal*, *42*(6), 1659–1667. https://doi.org/10.1183/09031936.00197712

Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., … Brorson, S. (2012). Observer bias in randomised clinical trials with binary outcomes: Systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*, *344*(3), e1119. https://doi.org/10.1136/bmj.e1119

Hwangbo, B., Lee, G. K., Lee, H. S., Lim, K. Y., Lee, S. H., Kim, H. Y., … Zo, J. I. (2010). Transbronchial and transesophageal fine-needle aspiration using an ultrasound bronchoscope in mediastinal staging of potentially operable lung cancer. *Chest*, *138*(4), 795–802. https://doi.org/10.1378/chest.09-2100

Hylton, D. A., Kidane, B., Spicer, J., Turner, S., Sullivan, K., Churchill, I., … Hanna, W. C. (n.d.). Endobronchial Ultrasound Staging of Operable NSCLC: Do Triple Normal Lymph Nodes Require Routine Biopsy? *CHEST*, *Submitted*, 1–12.

Hylton, D. A., Shargall, Y., Finley, C., Agzarian, J., Fahim, C., & Hanna, W. C. (2020). A novel online education module to teach clinicians how to correctly identify ultrasonographic features of mediastinal lymph nodes during endobronchial ultrasound. *Canadian Journal of Surgery. Journal Canadien de Chirurgie*, *63*(1), E62–E68. https://doi.org/10.1503/cjs.000119

Hylton, D. A., Turner, S., Kidane, B., Spicer, J., Xie, F., Farrokhyar, F., … Hanna, W. C. (2020). The Canada Lymph Node Score for prediction of malignancy in mediastinal lymph nodes during endobronchial ultrasound. *Journal of Thoracic and Cardiovascular Surgery*, *159*(6), 2499–2507. https://doi.org/10.1016/j.jtcvs.2019.10.205

Hylton, D. A., Turner, S., Shargall, Y., Finley, C., Agzarian, J., Yasufuku, K., … Hanna, W. C. (2018). Ultrasonographic characteristics of lymph nodes as predictors of malignancy during endobronchial ultrasound (EBUS): A systematic review. *Lung Cancer*, *126*(2018), 97–105. https://doi.org/10.1016/j.lungcan.2018.10.020

IBM Corp. Released. (2017). IBM SPSS Statistics version 25.0. *2017*.

Infante, M., Cavuto, S., Lutman, F. R., Passera, E., Chiarenza, M., Chiesa, G., … Alloisio, M. (2015). Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography. *American Journal of Respiratory & Critical Care Medicine*, *191*(10), 1166–1175. https://doi.org/10.1164/rccm.201408-1475OC

Ivanovic, J., Al-Hussaini, A., Al-Shehab, D., Threader, J., Villeneuve, P. J., Ramsay, T., … Seely, A. J. E. (2011). Evaluating the reliability and reproducibility of the Ottawa thoracic morbidity and mortality classification system. *Annals of Thoracic Surgery*, *91*(2), 387–393. https://doi.org/10.1016/j.athoracsur.2010.10.035

Izumo, T., Sasada, S., Chavez, C., Matsumoto, Y., & Tsuchida, T. (2014). Endobronchial ultrasound elastography in the diagnosis of mediastinal and hilar lymph nodes. *Japanese Journal of Clinical Oncology*, *44*(10), 956–962. https://doi.org/10.1093/jjco/hyu105

Jalil, B. A., Yasufuku, K., & Khan, A. M. (2015). Uses, Limitations, and Complications of Endobronchial Ultrasound. *Baylor University Medical Center Proceedings*, *28*(3), 325–330. https://doi.org/10.1080/08998280.2015.11929263

Jhun, B. W., Um, S.-W., Suh, G. Y., Chung, M. P., Kim, H., Kwon, O. J., … Lee, K.-J. (2014). Clinical value of endobronchial ultrasound findings for predicting nodal metastasis in patients with suspected lymphadenopathy: A prospective study. *Journal of Korean Medical Science*, *29*(12), 1632–1638. https://doi.org/10.3346/jkms.2014.29.12.1632

Kea, B., Hall, M. K., & Wang, R. (2019). Recognising bias in studies of diagnostic tests part 2: Interpreting and verifying the index test. *Emergency Medicine Journal*, *36*(8), 501–505. https://doi.org/10.1136/emermed-2019-208447

Kiyono, K., Sone, S., Sakai, F., Imai, Y., Watanabe, T., Izuno, I., … Watanabe, M. (1988). The number and size of normal mediastinal lymph nodes: A postmortem study. *American Journal of Roentgenology*, *150*(4), 771–776. https://doi.org/10.2214/ajr.150.4.771

Kutner, M., Nachtsheim, C., & Neter, J. (2004). Multiple Regression. In *Applied Linear Regression Models* (4th ed.). McGraw Hill/Irwin Series: Operations and Decision Sciences.

Lee, H. S., Lee, G. K., Lee, H. S., Kim, M. S., Lee, J. M., Kim, H. Y., … Hwangbo, B. (2008). Real-time endobronchial ultrasound-guided transbronchial needle aspiration in mediastinal staging of non-small cell lung cancer: How many aspirations per target lymph node station? *Chest*, *134*(2), 368–374. https://doi.org/10.1378/chest.07-2105

Long, Y., Liu, H. H., Yu, C., Tian, X., Yang, Y. R., Wang, C., & Pan, Y. (2012). Pre-existing diseases of patients increase susceptibility to hypoxemia during gastrointestinal endoscopy. *PLoS ONE*, *7*(5), e37614. https://doi.org/10.1371/journal.pone.0037614

Medford, A. R., Agrawal, S., Free, C. M., & Bennett, J. A. (2010). A prospective study of conventional transbronchial needle aspiration: performance and cost utility. *Respiration*, *79*(6), 482–489. https://doi.org/10.1159/000277931

Moore, K. A., Lipsitch, M., Barry, J. M., & Osterholm, M. T. (2020). The Future of the COVID-19 Pandemic: Lessons Learned from Pandemic Influenza. In *COVID-19: The CIDRAP Viewpoint* (pp. 1–8).

Moyer, V. A. (2014). Screening for lung cancer: U.S. preventive services task force recommendation statement. *Annals of Internal Medicine*, *160*(5), 330–338. https://doi.org/10.7326/m13-2771

Nakajima, T., Inage, T., Sata, Y., Morimoto, J., Tagawa, T., Suzuki, H., … Yoshino, I. (2015). Elastography for predicting and localizing nodal metastases during endobronchial ultrasound. *Respiration*, *90*(6), 499–506. https://doi.org/10.1159/000441798

National Comprehensive Cancer Network. (2018). Non-Small Cell Lung Cancer.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, *17*(8), 857–872. https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

Nyblom, J. (1992). Note on interpolated order statistics. *Statistics and Probability Letters*, *14*(2), 129–131. https://doi.org/10.1016/0167-7152(92)90076-H

O, J. H., Lodge, M. A., & Wahl, R. L. (2016). Practical PERCIST: A simplified guide to PET response criteria in solid tumors 1.0. *Radiology*, *280*(2), 576–584. https://doi.org/10.1148/radiol.2016142043

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Ong, P., Grosu, H., Eapen, G. A., Rodriguez, M., Lazarus, D., Ost, D., … Casal, R. F. (2015). Endobronchial ultrasound-guided transbronchial needle aspiration for systematic nodal staging of lung cancer in patients with N0 disease by computed tomography and integrated positron emission tomography-computed tomography. *Annals of the American Thoracic Society*, *12*(3), 415–419. https://doi.org/10.1513/AnnalsATS.201409-429OC

Ortakoylu, M. G., Iliaz, S., Bahadir, A., Aslan, A., Iliaz, R., Ozgul, M. A., & Urer, H. N. (2015). Diagnostic value of endobronchial ultrasound-guided transbronchial needle aspiration in various lung diseases. *Jornal Brasileiro de Pneumologia*, *41*(5), 410–414. https://doi.org/10.1590/S1806-37132015000004493

Paci, E., Puliti, D., Lopes Pegna, A., Carrozzi, L., Picozzi, G., Falaschi, F., … Goldoni, C. A. (2017). Mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial. *Thorax*, *72*(9), 825–831. https://doi.org/10.1136/thoraxjnl-2016-209825

Pastorino, U., Rossi, M., Rosato, V., Marchianò, A., Sverzellati, N., Morosi, C., … La Vecchia, C. (2012). Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. *European Journal of Cancer Prevention*, *21*(3), 308–315. https://doi.org/10.1097/CEJ.0b013e328351e1b6

Patridge, E. F., & Bardyn, T. P. (2018). Research Electronic Data Capture (REDCap). *Journal of the Medical Library Association*, Vol. 106, pp. 142–144. https://doi.org/10.5195/JMLA.2018.319

Perme, M. P., & Manevski, D. (2019). Confidence intervals for the Mann–Whitney test. *Statistical Methods in Medical Research*, *28*(12), 3755–3768. https://doi.org/10.1177/0962280218814556

Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J. W., & Altman, D. G. (2012). Reporting of noninferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. *JAMA - Journal of the American Medical Association*, *308*(24), 2594–2604. https://doi.org/10.1001/jama.2012.87802

Pino, R. M. (2007). The nature of anesthesia and procedural sedation outside of the operating room. *Current Opinion in Anaesthesiology*, *20*(4), 347–351. https://doi.org/10.1097/ACO.0b013e32827035c7

Poolman, R. W., Struijs, P. A. A., Krips, R., Sierevelt, I. N., Marti, R. K., Farrokhyar, F., & Bhandari, M. (2007). Reporting of outcomes in orthopaedic randomized trials: Does blinding of outcome assessors matter? *Journal of Bone and Joint Surgery*, *89*(3), 550–558. https://doi.org/10.2106/JBJS.F.00683

Ranganathan, P., Pramesh, C., & Aggarwal, R. (2016). Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. *Perspectives in Clinical Research*, *7*(3), 144–146. https://doi.org/10.4103/2229-3485.184823

Rozman, A., Marc Malovrh, M., Adamic, K., Subic, T., Kovac, V., & Flezar, M. (2015). Endobronchial ultrasound elastography strain ratio for mediastinal lymph node diagnosis. *Radiology and Oncology*, *49*(4), 334–340. https://doi.org/10.1515/raon-2015-0020

Rusch, V. W., Asamura, H., Watanabe, H., Giroux, D. J., Rami-Porta, R., & Goldstraw, P. (2009). The IASLC lung cancer staging project: A proposal for a new international lymph node map in the forthcoming seventh edition of the TNM classification for lung cancer. *Journal of Thoracic Oncology*, *4*(5), 568–577. https://doi.org/10.1097/JTO.0b013e3181a0d82e

Schmid-Bindert, G., Jiang, H., Kähler, G., Saur, J., Henzler, T., Wang, H., … Pilz, L. R. (2012). Predicting malignancy in mediastinal lymph nodes by endobronchial ultrasound: A new ultrasound scoring system. *Respirology*, *17*(8), 1190–1198. https://doi.org/10.1111/j.1440-1843.2012.02223.x

Schmidt, R. L., & Factor, R. E. (2013). Understanding sources of bias in diagnostic accuracy studies. *Archives of Pathology and Laboratory Medicine*, *137*(4), 558–565. https://doi.org/10.5858/arpa.2012-0198-RA

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials. *JAMA: The Journal of the American Medical Association*, *273*(5), 408–412. https://doi.org/10.1001/jama.273.5.408

Seely, A. J. E., Ivanovic, J., Threader, J., Al-Hussaini, A., Al-Shehab, D., Ramsay, T., … Sundaresan, R. S. (2010). Systematic classification of morbidity and mortality after thoracic surgery. *Annals of Thoracic Surgery*, *90*(3), 936–942. https://doi.org/10.1016/j.athoracsur.2010.05.014

Shafiek, H., Fiorentino, F., Peralta, A. D., Serra, E., Esteban, B., Martinez, R., … Cosío, B. G. (2014). Real-Time Prediction of Mediastinal Lymph Node Malignancy by Endobronchial Ultrasound. *Archivos de Bronconeumologia*, *50*(6), 228–234. https://doi.org/10.1016/j.arbr.2014.05.003

Sheather, S. (2009). Multiple Linear Regression. In *A Modern Approach to Regression with R* (pp. 125–149). New York: Springer Texts in Statistics.

Shingyoji, M., Nakajima, T., Yoshino, M., Yoshida, Y., Ashinuma, H., Itakura, M., … Iizasa, T. (2015). Endobronchial ultrasonography for positron emission tomography and computed tomography-negative lymph node staging in non-small cell lung cancer. *Annals of Thoracic Surgery*, *98*(5), 1762–1768. https://doi.org/10.1016/j.athoracsur.2014.05.078

St. Joseph's Healthcare Hamilton. (2020). Information for Current Patients or Patients with an Upcoming Appointment. Retrieved June 23, 2020, from St. Joseph's Healthcare Hamilton Official Website website: https://www.stjoes.ca/coronavirus/information-for-patients-and-visitors/have-an-appointment-during-covid-19

University Health Network: Patient Relations and Patient Education & Engagement. (2020). *Limits on Who Can Enter UHN During the COVID-19 Pandemic*. Retrieved from https://www.uhn.ca/PatientsFamilies/Health_Information/Health_Topics/Documents/Limits_On_Who_Can_Enter_UHN_During_COVID19.pdf

Vaid, S., McAdie, A., Kremer, R., Khanduja, V., & Bhandari, M. (2020). Risk of a second wave of Covid-19 infections: using artificial intelligence to investigate stringency of physical distancing policies in North America. *International Orthopaedics*, e1-9. https://doi.org/10.1007/s00264-020-04653-3

Vansteenkiste, J. F., Stroobants, S. G., De Leyn, P. R., Dupont, P. J., Bogaert, J., Maes, A., … Demedts, M. G. (1998). Lymph node staging in non-small-cell lung cancer with FDG-PET scan: A prospective study on 690 lymph node stations from 68 patients. *Journal of Clinical Oncology*, *16*(6), 2142–2149. https://doi.org/10.1200/JCO.1998.16.6.2142

Wahidi, M. M., Herth, F., Yasufuku, K., Shepherd, R. W., Yarmus, L., Chawla, M., … Feller-Kopman, D. J. (2016). Technical aspects of endobronchial ultrasound-guided transbronchial needle aspiration CHEST guideline and expert panel report. *Chest*, *149*(3), 816–835. https://doi.org/10.1378/chest.15-1216

Wahidi, M. M., & Yasufuku, K. (2013). Point: Should endobronchial ultrasound guide every transbronchial needle aspiration of lymph nodes? Yes. *Chest*, *144*(3), 732–734. https://doi.org/10.1378/chest.13-0702

Wang Memoli, J. S., El-Bayoumi, E., Pastis, N. J., Tanner, N. T., Gomez, M., Huggins, J. T., … Silvestri, G. A. (2011). Using endobronchial ultrasound features to predict lymph node metastasis in patients with lung cancer. *Chest*, *140*(6), 1550–1556. https://doi.org/10.1378/chest.11-0252

Wang, L., Wu, W., Hu, Y., Teng, J., Zhong, R., Han, B., & Sun, J. (2015). Sonographic features of endobronchial ultrasonography predict intrathoracic lymph node metastasis in lung cancer patients. *Annals of Thoracic Surgery*, *100*(4), 1203–1209. https://doi.org/10.1016/j.athoracsur.2015.04.143

Wille, M. M. W., Dirksen, A., Ashraf, H., Saghir, Z., Bach, K. S., Brodersen, J., … Pedersen, J. H. (2016). Results of the randomized danish lung cancer screening trial with focus on high-risk profiling. *American Journal of Respiratory and Critical Care Medicine*, *193*(5), 542–551. https://doi.org/10.1164/rccm.201505-1040OC

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, *22*(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

Xia, Y., Zhang, B., Zhang, H., Li, W., Wang, K. P., & Shen, H. (2015). Evaluation of lymph node metastasis in lung cancer: Who is the chief justice? *Journal of Thoracic Disease*, *7*(Suppl 4), S231–S237. https://doi.org/10.3978/j.issn.2072-1439.2015.11.63

Xu, D. M., van Klaveren, R. J., de Bock, G. H., Leusveld, A., Zhao, Y., Wang, Y., … Oudkerk, M. (2008). Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. *European Journal of Radiology*, *68*(2), 347–352. https://doi.org/10.1016/j.ejrad.2007.08.027

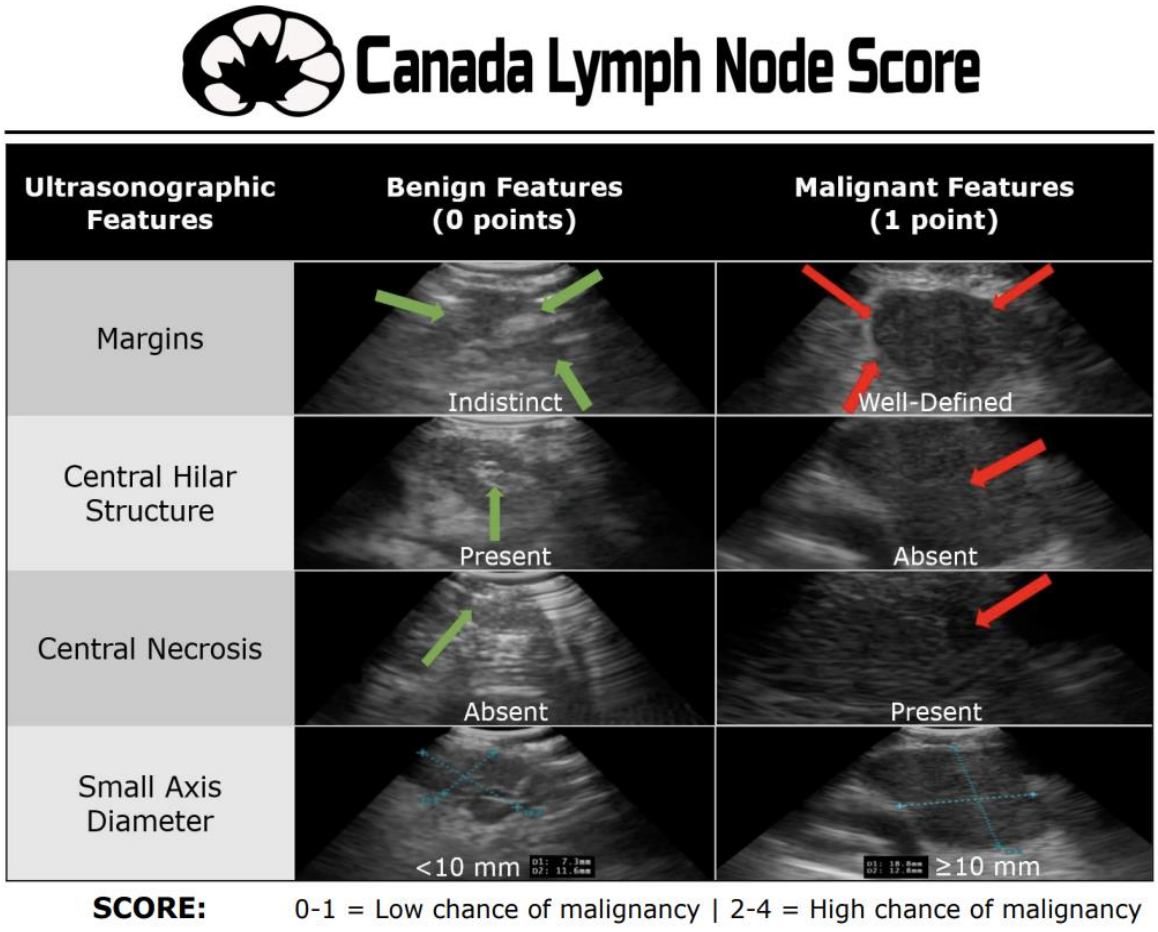
Yasufuku, K., Nakajima, T., Motoori, K., Sekine, Y., Shibuya, K., Hiroshima, K., & Fujisawa, T. (2006). Comparison of endobronchial ultrasound, positron emission tomography, and CT for lymph node staging of lung cancer. *Chest*, *130*(3), 710–718. https://doi.org/10.1378/chest.130.3.710

Yasufuku, K., Pierre, A., Darling, G., De Perrot, M., Waddell, T., Johnston, M., … Keshavjee, S. (2011). A prospective controlled trial of endobronchial ultrasound-guided transbronchial needle aspiration compared with mediastinoscopy for mediastinal lymph node staging of lung cancer. *Journal of Thoracic and Cardiovascular Surgery*, *142*(6), 1393–1400. https://doi.org/10.1016/j.jtcvs.2011.08.037

Ziyade, S., Pinarbasili, N. B., Ziyade, N., Akdemir, O. C., Sahin, F., Soysal, Ö., & Toker, A. (2013). Determination of standard number, size and weight of mediastinal lymph nodes in postmortem examinations: Reflection on lung cancer surgery. *Journal of Cardiothoracic Surgery*, *8*(94), 1–8. https://doi.org/10.1186/1749-8090-8-94

APPENDIX 1: Tables & Figures (in order as they appear within texts)

Table 1. CLNS visual assessment criteria



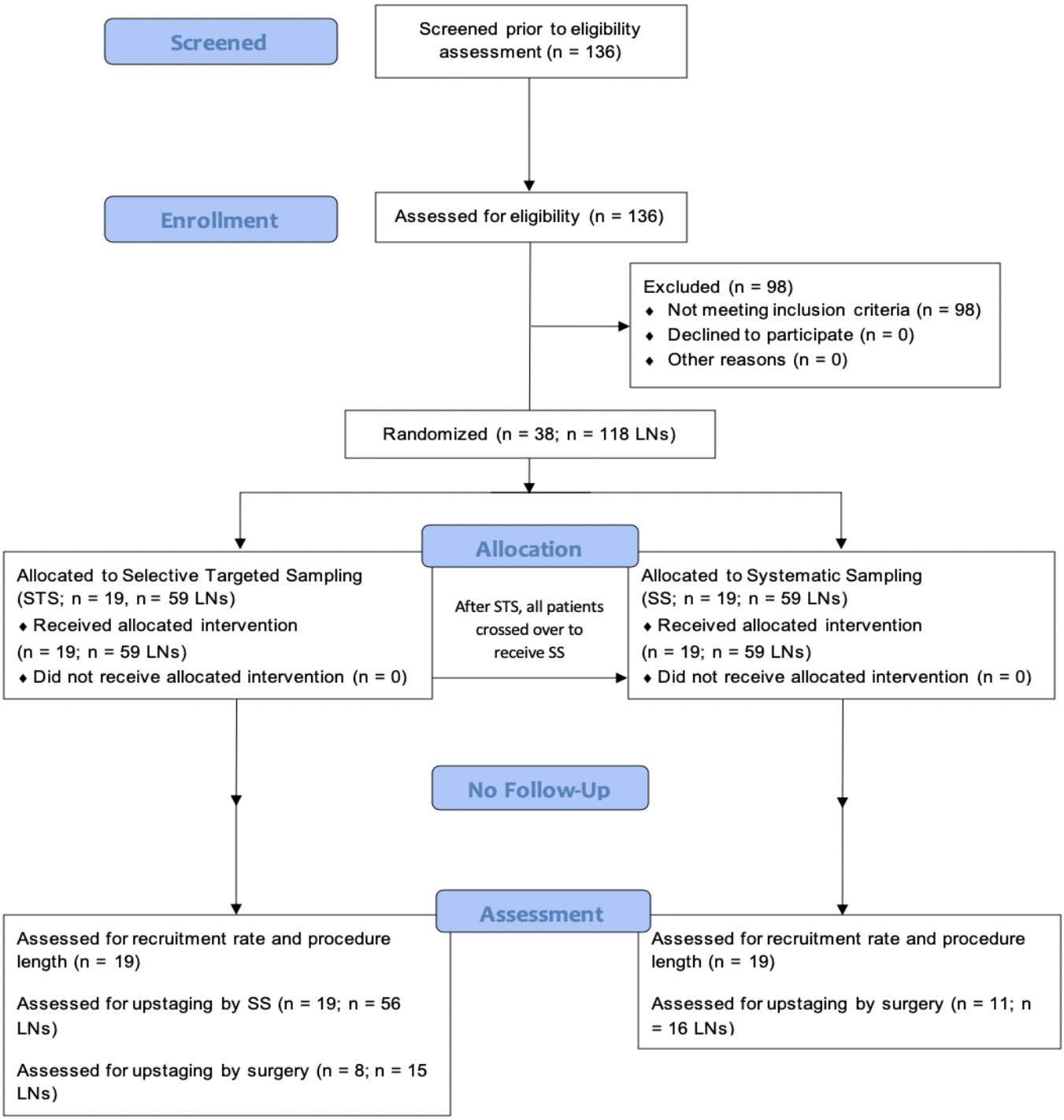


Figure 1. CONSORT flow diagram. Patients in the STS arm crossed over to receive SS, but primary objectives data for the STS arm were collected prior to this crossover. LNs = lymph nodes.

Table 2. Patient baseline characteristics

n_{total} = 38	Selective Targeted Sampling (n = 19)	Systematic Sampling (n = 19)
Age in years, mean (SD)	72.74 (9.02)	70.53 (12.44)
Sex, n (%)		
Female	10 (52.63)	12 (63.16)
Male	9 (47.37)	7 (36.84)
Smoking status, n (%)		
Non-smoker	2 (10.53)	3 (15.79)
Ex-smoker†	8 (42.11)	9 (47.37)
Smoker	9 (47.37)	7 (36.84)
Comorbidities, n (%)		
Anxiety	3 (15.79)	3 (15.79)
Asthma	4 (21.05)	1 (5.27)
Cardiovascular history¥	4 (21.05)	6 (31.58)
Chronic pain	3 (15.79)	1 (5.27)
COPD	8 (42.11)	8 (42.11)
Diabetes mellitus	6 (31.58)	3 (15.79)
Hypertension	12 (63.16)	8 (42.11)
Obesity (BMI≥30)	7 (36.84)	5 (26.32)
Imaging N-stage, n (%)§		
N0	14 (73.68)	14 (73.68)
N1	5 (26.32)	5 (26.32)

†Patients needed to have successfully quit smoking for a year, in order to be considered an ex-smoker.

¥History included myocardial infarct, stroke, atrial fibrillation, coronary artery disease, and peripheral artery disease.

§N-stage based on pre-treatment PET/CT scans.

SD = standard deviation.

Table 3. Characteristics of the lymph nodes examined during EBUS†

n_{total} = 118	**Selective Targeted Sampling (n = 59)**	**Systematic Sampling (n = 59)**
Nodal station, n (%)		
2R	0	1 (1.70)
4L	19 (32. 20)	17 (28.81)
4R	19 (32. 20)	19 (32.20)
7	19 (32. 20)	19 (32. 20)
10R	1 (1.70)	2 (3.39)
11L	1 (1.70)	0
11R	0	1 (1.70)
Diagnosis based on sampling method, n (%)		
Benign	58 (98.30)	48 (81.35)
Malignant	1 (1.70)	1 (1.70)
Inconclusive	0	10 (16.95)
Biopsied		
n_{total} = 115	**(n = 56)¥**	**(n = 59)**
Diagnosis based on EBUS pathology, n (%)		
Benign	42 (75.00)	48 (81.35)
Malignant	4 (7.14)	1 (1.70)
Inconclusive	10 (17.86)	10 (16.95)
Assessed during Surgery		
n_{total} = 31	**(n = 15)**	**(n = 16)**
Nodal stations assessed during EBUS and surgery, n (%)		
4L	1 (6.67)	0
4R	6 (40.00)	6 (37.50)
7	8 (53.33)	9 (56.25)
11R	0	1 (6.25)
Surgical pathology diagnosis, n (%)		
Benign	15 (100.00)	15 (93.75)
Malignant	0	1 (6.25)

†Lymph nodes are the unit of analysis. SD = standard deviation
¥For STS, this includes all LNs biopsied regardless of their CLNS.

Table 4. Histologically confirmed diagnosis of patients†

n_{total} = 38	Selective Targeted Sampling (n = 19)	Systematic Sampling (n = 19)
Referred treatment, n (%)		
Surgery	8 (42.11)	11 (57.90)
SBRT	8 (42.10)	6 (31.58)
Chemotherapy	3 (15.79)	1 (5.26)
Surveillance	0	1 (5.26)
Disease diagnosis, n (%)		
Benign	0	2 (10.53)
Malignant	19 (100.00)	17 (89.47)
Location of lesion, n (%)		
LUL	2 (10.53)	3 (15.79)
LLL	4 (21.05)	4 (21.05)
RUL	7 (36.84)	8 (42.11)
RML	1 (5.26)	0
RLL	5 (26.32)	4 (21.05)
Malignant Cases n_{total} = 36	**(n = 19)**	**(n = 17)**
Histology, n (%)		
Adenocarcinoma	6 (31.58)	8 (47.06)
Squamous cell carcinoma	10 (52.63)	6 (35.29)
Other	3 (15.78)	3 (17.65)
T-stage, n (%)		
T1b	0	2 (11.77)
T1c	2 (10.53)	1 (5.88)
T2a	5 (26.32)	8 (47.06)
T2b	3 (15.79)	1 (5.88)
T3	5 (26.32)	4 (23.53)
T4	4 (21.05)	1 (5.88)

†Staging based on the 8th Edition of TNM in Lung Cancer.

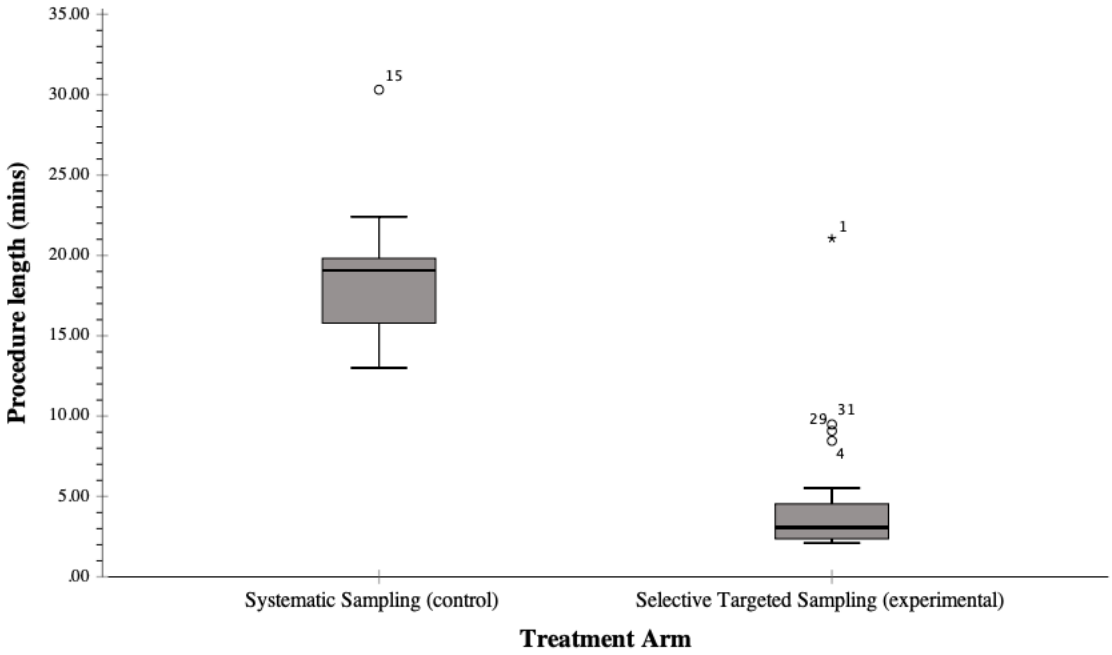


Figure 2. Boxplots of procedure lengths for SS and STS. Asterisks represent extreme outliers, whereas circles represent mild outliers. Coinciding numbers correspond to the study PID. The procedure length, in minutes, for each outlier were 1 = 21.05, 4 = 8.46, 15 = 30.31, 29 = 9.08, and 31 = 9.48. All outliers had justifiable reasons for inclusion within the analysis.

Table 5. Multiple linear regression model for procedure length†

Predictors of procedure length¥	Unstandardized β	Std. Error	95% CI	p-value
Treatment Arm	-13.77	1.31	-16.43 to -11.12	<0.0001
Number of LNs Biopsied	3.82	1.71	0.36 to 7.28	0.032
Constant	6.72	5.38	-4.20 to 17.64	0.220

†Backwards elimination at $p > 0.1$.
¥ Coding: {SS = 0, STS = 1}, {3 LNs = 0, 4 LNs = 1}.

Table 6. Final N stage of patients and number of missed nodal metastases

n_{total} = 38	Selective Targeted Sampling (n = 19)	Systematic Sampling (n = 19)
Final N-stage, n (%)		
N0	10 (52.63)	13 (68.42)
N1	6 (31.58)	5 (26.32)
N2	3 (15.79)	1 (5.26)
Upstaged, proportion (%)		
N0 → N1†	3/8 (37.50)	1/11 (9.09)
N0 → N2¥	2/19 (10.53)	0

†Only patients that underwent surgery had N0→N1 upstaging.
¥N0→N2 upstaging only occurred when patients in the STS arm crossed over to the SS arm.

Table 7. Nodal disease diagnoses, with LNs as the unit of analysis, for STS and SS when compared to the surgical pathology (reference standard)

Table 7A. SS versus surgical pathology (inconclusive results considered benign diagnosis)

Systematic Sampling Diagnosis	Surgical Pathology Diagnosis		Total
	Malignant	*Benign*	
Malignant	0	0	0
Benign	1	15	16
Total	1	15	**16**
Diagnostic Parameters			
Parameter	*Value (%)*	*95% CI (%)*	
Sensitivity	0.00	0.00 to 97.50	
Specificity	100.00	78.20 to 100.00	
PPV	-	-	
NPV	93.75	93.75 to 93.75	
Accuracy†	93.75	67.71 to 99.67	

Table 7B. SS versus surgical pathology (inconclusive results excluded)

Systematic Sampling Diagnosis	**Surgical Pathology Diagnosis**		**Total**
	Malignant	*Benign*	
Malignant	0	0	0
Benign	0	14	14
Total	0	14	**14**
Diagnostic Parameters			
Parameter	*Value (%)*	*95% CI (%)*	
Sensitivity	-	-	
Specificity	100.00	76.84 to 100.00	
PPV	-	-	
NPV	100.00	-	
Accuracy†	100.00	73.24 to 100.00	

Table 7C. STS versus surgical pathology

Selective Targeted Sampling Diagnosis	**Surgical Pathology Diagnosis**		**Total**
	Malignant	*Benign*	
Malignant	0	0	0
Benign	0	15	15
Total	0	15	**15**
Diagnostic Parameters			
Parameter	*Value (%)*	*95% CI (%)*	
Sensitivity	-	-	
Specificity	100.00	78.20 to 100.00	
PPV	-	-	
NPV	100.00	-	
Accuracy†	100.00	74.65 to 100.00	

"-" symbolizes that it was not possible to calculate the diagnostic parameter.

† The 95% CI were calculated using Wilson's method.

Table 8. Crossover analysis of nodal disease diagnoses between STS and SS with LNs serving as the unit of analysis

Table 8A. Inconclusive results considered benign SS diagnosis

		Systematic Sampling		Total
		Malignant	*Benign*	
Selective Targeted Sampling	*Malignant*	1	0	1
	Benign	3	52	55
	Total	4	52	**56**
Diagnostic Parameter				
Parameter		*Value (%)*	*95% CI (%)*	
Percent Difference		5.36	-0.54 to 11.25	

Table 8B. Inconclusive results excluded

		Systematic Sampling		Total
		Malignant	*Benign*	
Selective Targeted Sampling	*Malignant*	1	0	1
	Benign	3	42	45
	Total	4	42	**46**
Diagnostic Parameter				
Parameter		*Value (%)*	*95% CI (%)*	
Percent Difference		6.52	-0.61 to 13.66	

Table 9. Prevalence of the CLNS values for each LN assessed by STS

	Selective Targeted Sampling (n = 59)
CLNS values, n (%)	
0	41 (69.49)
1	13 (22.03)
2	5 (8.48)

Table 10. Crossover analysis for number of LNs biopsied between STS and SS

		Systematic Sampling		Total
		Biopsied	*Not Biopsied*	
Selective Targeted Sampling	*Biopsied*	5	0	5
	Not Biopsied	51	3†	54
	Total	56	3	**59**

†These LNs could not be found once crossed over to SS.

Table 11. Crossover analysis of inconclusive LN diagnoses between STS and SS

		Systematic Sampling		Total
		Inconclusive	*Conclusive*	
Selective Targeted Sampling	*Inconclusive*	0	0	0
	Conclusive	10	46	56
	Total	10	46	**56**

Table 12. Diagnostic parameters intended for definitive trial and consequences of false positive and false negative diagnoses from both sampling methods

		Surgical Diagnosis	
		Malignant	*Benign*
Sampling Method Diagnosis	*Malignant*	TP	FP
	Benign	FN	TN

Diagnostic Parameters

Parameters	*Formula (×100)*	*Description*
Accuracy	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$	○ Ability of the sampling method to correctly detect **LN malignancy when present** and **LN benignity when malignancy is absent**
Sensitivity	$\frac{TP}{(TP + FN)}$	○ How often the sampling method correctly identifies a **malignant LN**
Specificity	$\frac{TN}{(TN + FP)}$	○ How often the sampling method correctly identifies a **benign LN**

PPV	$\frac{TP}{(TP + FP)}$	○ Probability that a LN with a **malignant sampling diagnosis** truly has nodal disease
NPV	$\frac{TN}{(TN + FN)}$	○ Probability that a LN with a **benign sampling diagnosis** truly does not have nodal disease

Consequence of Incorrect Sampling Diagnosis

Incorrect Diagnosis	*Consequence*
FP	○ Likely would NOT undergo surgery and receive the reference test ○ Likely referred to chemotherapy when actually a candidate for surgery or SBRT ○ Exposed to a more intense treatment regimen
FN	○ Likely would undergo either surgery or SBRT ○ If surgical candidate, would receive reference test, which would confirm false negative diagnosis and result in subsequent referral to chemotherapy ○ Delays to most appropriate treatment (chemotherapy)
Inconclusive	○ Delays to appropriate treatment. Further testing often needed (e.g. mediastinoscopy, repeat EBUS, bronchoscopy)

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

APPENDIX 2: CONSORT checklist for reporting information in feasibility and pilot trials

Section/ Topic	Item #	Checklist item	Reported on page #
Title and abstract			
	1a	Identification as a pilot or feasibility randomised trial in the title	i-ii
	1b	Structured summary of pilot trial design, methods, results, and conclusions (for specific guidance see CONSORT abstract extension for pilot trials)	iv-v
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale for future definitive trial, and reasons for randomised pilot trial	1-8
	2b	Specific objectives or research questions for pilot trial	9, 11
Methods			
Trial design	3a	Description of pilot trial design (such as parallel, factorial) including allocation ratio	11
	3b	Important changes to methods after pilot trial commencement (such as eligibility criteria), with reasons	12, 15
Participants	4a	Eligibility criteria for participants	11-12
	4b	Settings and locations where the data were collected	11
	4c	How participants were identified and consented	16
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	16-17
Outcomes	6a	Completely defined prespecified assessments or measurements to address each pilot trial objective specified in 2b, including how and when they were assessed	9-11
	6b	Any changes to pilot trial assessments or measurements after the pilot trial commenced, with reasons	N/A
	6c	If applicable, prespecified criteria used to judge whether, or how, to proceed with future definitive trial	9-10
Sample size	7a	Rationale for numbers in the pilot trial	15-16
	7b	When applicable, explanation of any interim analyses and stopping guidelines	15-16
Randomisation:			

Sequence generation	8a	Method used to generate the random allocation sequence	12
	8b	Type of randomisation(s); details of any restriction (such as blocking and block size)	12
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	12
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	12-13
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	12-14
	11b	If relevant, description of the similarity of interventions	N/A
Statistical methods	12	Methods used to address each pilot trial objective whether qualitative or quantitative	17-21
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were approached and/or assessed for eligibility, randomly assigned, received intended treatment, and were assessed for each objective	23, 71
	13b	For each group, losses and exclusions after randomisation, together with reasons	N/A
Recruitment	14a	Dates defining the periods of recruitment and follow-up	23
	14b	Why the pilot trial ended or was stopped	25
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	72-74
Numbers analysed	16	For each objective, number of participants (denominator) included in each analysis. If relevant, these numbers should be by randomised group	25-32, 76-79
Outcomes and estimation	17	For each objective, results including expressions of uncertainty (such as 95% confidence interval) for any estimates. If relevant, these results should be by randomised group	25-32, 75-79
Ancillary analyses	18	Results of any other analyses performed that could be used to inform the future definitive trial	26-27
Harms	19	All-important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	32
	19a	If relevant, other important unintended consequences	N/A
Discussion			
Limitations	20	Pilot trial limitations, addressing sources of potential bias and remaining uncertainty about feasibility	37-41

Generalisability	21	Generalisability (applicability) of pilot trial methods and findings to future definitive trial and other studies	46-47
Interpretation	22	Interpretation consistent with pilot trial objectives and findings, balancing potential benefits and harms, and considering other relevant evidence	33-37
	22a	Implications for progression from pilot to future definitive trial, including any proposed amendments	41-46
Other information			
Registration	23	Registration number for pilot trial and name of trial registry	21
Protocol	24	Where the pilot trial protocol can be accessed, if available	N/A
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	vi
	26	Ethical approval or approval by research review committee, confirmed with reference number	21