Pulmonary Function Tests as Surrogate Markers for Interstitial Lung Disease Onset in Systemic Sclerosis

Melissa Caron

Department of Epidemiology, Biostatistics and Occupational Health

McGill University, Montreal

March 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy (Ph.D.)

Abstract

Systemic sclerosis (SSc) is a chronic and progressive autoimmune disease characterized by vasculopathy and widespread fibrosis. In addition to disfiguring skin involvement, SSc patients can suffer from extensive internal organ damage including interstitial lung disease (ILD) which is the leading cause of morbidity and mortality in this patient group. Pulmonary function tests (PFTs) are routinely used to monitor SSc-associated ILD (SSc-ILD) in clinical practice, epidemiologic studies and clinical trials for purposes of treatment initiation and follow-up. Yet, few validation studies have assessed which, if any, PFT measures are good surrogate markers for SSc-ILD onset.

The first contribution of this thesis is a systematic review of the literature that determined which PFT measures have been most commonly used as outcomes for SSc-ILD in experimental and observational studies. The systematic review also summarized the results of studies that validated PFT measures against either high-resolution computed tomography or lung biopsy results in SSc patients. Results showed that despite the current preference for the use of forced vital capacity (FVC) % predicted, available evidence does not overwhelmingly support its preferred status as a PFT surrogate marker for SSc-ILD.

The second contribution of this thesis is a methodologic study which used both simulated data and real data from a large Canadian observational cohort of SSc patients to evaluate the potential of hidden Markov models (HMMs) to validate and use PFT measures for SSc-ILD ascertainment. The HMM was of interest because it can use the full PFT measurement history to model the probability of SSc-ILD occurrence while simultaneously correcting for PFT variability. Its statistical performance when using FVC was compared to that of two commonly used definitions for possible SSc-ILD onset: <80% predicted FVC and ≥10% decline in FVC. The HMM had the highest specificity and lowest error rate compared to the cut-off and change in FVC algorithms.

The third contribution of this thesis is the first attempt to use HMMs to validate FVC, diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC) as surrogate markers for SSc-ILD onset in a large SSc patient cohort. Separate HMMs were evaluated using FVC, DLCO and TLC as well as for different bivariate and multivariate combinations of these PFT measures. The HMM using all three measures (FVC/DLCO/TLC) had the highest sensitivity likely making it the best PFT screening tool for SSc-ILD onset. However, all models had generally poor sensitivity for SSc-ILD. On the other hand, the joint TLC % predicted and FEV_1/FVC (ratio of forced expiratory volume in one second to FVC) model had a high specificity and low error rate, followed closely by DLCO % predicted and TLC absolute. These results suggest that TLC and DLCO may be better PFT surrogate markers for SSc-ILD onset than FVC and should also be considered as main PFT outcome measures in epidemiologic studies.

In summary, this thesis demonstrated that TLC and DLCO may be better PFT surrogate markers for SSc-ILD onset than FVC, thereby helping to improve the quality of evidence of future epidemiologic studies. Furthermore, these findings will inform clinical decision-making by demonstrating that serial measurements of FVC, DLCO and TLC should be used jointly to screen for SSc patients who should undergo further investigation for SSc-ILD. Nevertheless, the low sensitivities of PFT measures suggest that other avenues should be explored in the search for a suitable marker for SSc-ILD onset. Finally, from the methodologic standpoint, this thesis demonstrated that HMMs can be better suited to detect disease onset than fixed cut-offs or pre-specified changes in surrogate marker values.

Abrégé

La sclérodermie systémique (ScS) est une maladie auto-immune chronique à pathologie progressive caractérisée par une vasculopathie et une fibrose diffuse. En plus d'une atteinte cutanée déformante, les patients sclérodermiques souffrent de dommages importants aux organes internes. En particulier, la pneumopathie interstitielle (PI) est la cause principale de morbidité et de mortalité au cours de la ScS. Les explorations fonctionelles respiratoires (EFR) sont couramment utilisées en pratique clinique, ainsi que dans les études épidémiologiques, pour surveiller l'apparition de la PI liée à la ScS (PI-ScS) à des fins d'initiation du traitement et de suivi. À date, peu d'études ont validé les mesures d'EFR en tant que marqueurs de substitution pour le déclenchement de la PI-ScS.

La première contribution de cette thèse est une revue systématique de la littérature qui a déterminé quelles mesures d'EFR ont été utilisées les plus fréquemment comme marqueurs de substitution dans les études expérimentales et observationnelles de PI-ScS. Cette revue systématique a également résumé les résultats d'études qui ont validé les mesures d'EFR vis-à-vis la tomodensitométrie à haute résolution ou la biopsie pulmonaire chez les patients atteints de ScS. Les résultats ont démontré que, malgré la préférence actuelle pour l'utilisation du % prédit de la capacité vitale forcée (CVF), les preuves disponibles n'appuient pas son statut comme marqueur de substitution préféré pour la PI-ScS.

La deuxième contribution de cette thèse est une étude méthodologique qui a exploité à la fois des données simulées et des données réelles provenant d'une grande cohorte observationnelle de patients canadiens atteints de ScS. L'étude a évalué l'application d'un modèle de Markov caché (MMC) comme moyen d'établir l'apparition de la PI-ScS en utilisant des mesures d'EFR. En particulier, le MMC peut utiliser l'historique complet des mesures d'EFR pour prédire la probabilité d'occurrence de la PI-ScS tout en corrigeant pour les erreurs de mesure liées aux EFR. La performance d'un MMC utilisant des

mesures de CVF a été comparé à deux définitions communes de PI-ScS : un CVF <80% prédit et un déclin en CVF ≥10%. Le MMC a démontré la spécificité la plus élevée et le taux d'erreur le plus faible comparativement aux deux autres définitions.

La troisième contribution de cette thèse est l'utilisation d'un MMC pour valider la CVF, la capacité de diffusion du monoxyde de carbone (DLCO) et la capacité pulmonaire totale (CPT) comme marqueurs de substitution pour le déclenchement de la PI dans une grande cohorte de patients atteints de ScS. Des MMC distincts utilisant soit la CVF, la DLCO ou la CPT ainsi que différentes combinaisons de deux ou plusieurs mesures d'EFR ont été évalués. Le MMC utilisant conjointement la CVF, la DLCO et la CPT a démontré la sensibilité la plus élevée, ce qui en fait le meilleur outil de dépistage utilisant des mesures d'EFR pour la PI-ScS. Par contre, tous les modèles avaient de manière générale une basse sensibilité pour la PI-ScS. En revanche, le MMC utilisant le % prédit de la CPT conjointement avec le VEMS/CVF (volume expiratoire maximal seconde divisé par la CVF) avait une spécificité élevée et un faible taux d'erreur, suivi de près par le % prédit de la DLCO et le CPT en valeur absolue. Ces résultats suggèrent que la CPT et la DLCO pourraient être de meilleurs marqueurs de substitution que la CVF pour le déclenchement de la PI-ScS. Conséquemment, ces mesures devraient être considérées comme paramètres principaux dans les études épidémiologiques sur ce sujet.

En résumé, cette thèse a démontré que la CPT et la DLCO sont potentiellement de meilleurs marqueurs de substitution que la CVF pour le déclenchement de la PI-ScS et contribue ainsi à l'amélioration de la qualité de futures études épidémiologiques. De plus, les résultats de cette thèse démontrent que l'observation conjointe des mesures en série de la CVF, de la DLCO et de la CPT dans le cadre d'un MMC identifie le plus grand nombre de patients pouvant bénéficier d'une enquête plus approfondie. Ceci facilitera le dépistage de la PI-ScS. Étant donné la faible sensibilité des mesures d'EFR, d'autres marqueurs de substitution devraient être explorées pour le déclenchement de la PI-

ScS. Finalement, d'un point de vue méthodologique, cette thèse a démontré que les MMC peuvent être mieux adaptés pour détecter l'avènement d'une maladie que l'utilisation de seuils fixes ou de changements pré-spécifiés des valeurs de marqueurs de substitution.

Acknowledgements

The completion of my doctoral degree would not have been possible without the help, guidance and encouragement of my thesis co-supervisors, Dr. Kevin Schwartzman and Dr. Russell Steele. I would like to sincerely thank Kevin for being a remarkable mentor and an exemplary clinician-epidemiologist. I have learnt so much and gained invaluable insight, both methodologic and substantive, from his professional expertise. I would also like to express my deepest gratitude to Russ whose statistical knowledge is unparalleled. Since my time working under his supervision on an undergraduate research project, he has helped me grow as both a quantitative epidemiologist and an individual. His unwavering belief in my abilities encouraged me to discover my full potential.

I was very fortunate to have Dr. Marie Hudson as my thesis committee member and would like to profoundly thank her. She has been nothing short of inspirational and is a true role-model. I will forever be grateful for her encouraging me to write my first scientific article as an undergraduate student and subsequently mailing me the published manuscript. It still sits proudly in my library.

I would be remiss if I did not use this opportunity to acknowledge the Canadian Scleroderma Research Group (CSRG). Their summer studentship award in 2010 allowed me to discover and fall in love with epidemiology. I can truly say that this opportunity was life changing and I am so thankful that programs, such as this one, exist for young university students trying to find their way.

I would like to thank the professors and staff in the Department of Epidemiology, Biostatistics and Occupational Health as well as everyone in the Respiratory Epidemiology and Clinical Research Unit (RECRU). I was so lucky to have shared this journey with my phenomenal PhD friends. I cherish walking home from Purvis Hall with Daniala, decompressing in Zumba with Tina, sending motivational and accountability e-mails with Helen and Tanja, chatting with Hannah at the office and

celebrating some of our cohort's milestones with Brooke, Deepa, Farzin, Nichole, Pauley, Sonia and Tanya.

Lastly, I would like to give the most heartfelt thanks to my family for their endless love and support. To my mom Monica, dad Michel, and their respective partners Raymond and Lyne, who have encouraged me to always reach for the stars and give it my all. To my brother Martin and stepsister Steph who are downright the coolest and best siblings I could ever ask for. To my in-laws, Michael and Micheline, and my sister-in-law Caroline, for always having treated me as family. Finally, to my husband Mathew. You are my rock. I simply could not have done this without you. Thank you for believing in me and reminding me just how much I am capable of. We've experienced many adventures since we were 17 years old but completing our PhDs and moving to California has certainly been the most rewarding.

Contribution of Authors

The need for identifying a good pulmonary function test (PFT) surrogate marker for systemic sclerosis-associated interstitial lung disease (SSc-ILD) was initially discussed with my committee member, Dr. Marie Hudson. The thesis objectives were developed by myself and were agreed upon with my supervisors, Dr. Kevin Schwartzman and Dr. Russell Steele, and by Dr. Hudson. I proposed the use of hidden Markov models (HMMs) to achieve the overarching research aim and this was approved by my entire thesis committee. Below is a breakdown of the contribution of authors by study.

Manuscript 1

Caron M, Hoa S, Hudson M, Schwartzman K, Steele R. Pulmonary Function Tests as Outcomes for Systemic Sclerosis Interstitial Lung Disease. European Respiratory Review. 2018;27(148).

I developed the study protocol for this systematic review which was revised and approved by Drs. Schwartzman, Steele and Hudson. I was responsible for systematically identifying all potentially relevant studies. The selection of eligible studies through a primary and secondary screening strategy was performed by Dr. Hoa and me. I performed the data extraction and summarized all results. I wrote the draft for this article which was subsequently reviewed by Drs. Schwartzman, Steele, Hudson and Hoa. All authors approved the final version.

Manuscript 2

Caron M, Schwartzman K, Hudson M, Baron M, Steele RJ, and the Canadian Scleroderma Research Group. Ascertaining Disease Onset Using Surrogate Markers and Hidden Markov Models: An Application to Systemic Sclerosis-Associated Interstitial Lung Disease. Currently under review at the Journal of Clinical Epidemiology.

This study was conceptualized by me and Drs. Schwartzman, Steele and Hudson. Drs. Baron and Hudson provided data from the Canadian Scleroderma Research Group (CSRG) registry. I was responsible for preparing the data for analysis. I performed all statistical analyses and was aided by Dr. Steele. I interpreted the results and drafted the article which was critically reviewed by Drs. Schwartzman, Steele and Hudson for intellectual content. The final version was approved by all authors.

Manuscript 3

Caron M, Schwartzman K, Hudson M, Baron M, Steele RJ, and the Canadian Scleroderma Research Group. Validation of Pulmonary Function Test Measures as Surrogate Markers for Onset of Interstitial Lung Disease in Systemic Sclerosis. Currently under review at Rheumatology (Oxford).

I was primarily responsible for the conception and study design along with Drs. Schwartzman, Steele and Hudson. Drs. Baron and Hudson again provided data from the CSRG registry. I was responsible for preparing the data for analysis. I performed all statistical analyses and was aided by Dr. Steele. I interpreted the results and drafted the article which was reviewed by Drs. Schwartzman, Steele and Hudson who all provided feedback. In particular, Drs. Schwartzman and Hudson helped me formalize the clinical impact of my findings. The final version was approved by all authors.

Statement of Originality

The research encompassed in this thesis represents an original contribution to the fields of respiratory and rheumatologic epidemiology. More specifically, the results contribute to the advancement of knowledge in the arena of systemic sclerosis-associated interstitial lung disease (SSc-ILD). Furthermore, this work introduced hidden Markov models (HMMs) as a means of using surrogate markers to detect disease onset to the general field of clinical epidemiology.

The first study in this thesis, a systematic review of the literature including over 200 studies, was the first to juxtapose the systemic sclerosis (SSc) community's embracement of forced vital capacity (FVC) as the de facto surrogate marker for SSc-ILD with the lack of published evidence from validation studies to support this claim. The second study proposes HMMs as a sophisticated method of using surrogate markers, such as pulmonary function test measures, to detect disease presence. Though HMMs are often discussed in the statistical literature and occasionally employed in disease surveillance, they are rarely used in epidemiologic studies and have, to the best of my knowledge, never been validated using clinical data. Finally, the third study is the first to use HMMs to validate FVC, diffusing capacity for carbon monoxide (DLCO), and total lung capacity (TLC) as potential surrogate markers for SSc-ILD onset.

While I received unparalleled guidance from my thesis supervisors and committee member on the clinical and methodologic aspects of this research, I affirm that the conception, execution, and writing of this doctoral work are entirely my own.

Statement of Financial Support

I am extremely grateful for the financial support that I received throughout my doctoral research and studies. Entry into the epidemiology program was accompanied by the McGill University Graduate Excellence Award. The first two years of my degree were financed by the Fonds de recherche – Santé Québec (FRQS) Doctoral Training Award. I subsequently received the Frederick Banting and Charles Best Canada Graduate Scholarship (GSD – 146268), a Canadian Institutes of Health Research (CIHR) doctoral award which financed the remaining three years of my training. This award also covered all travel and conference-related expenses allowing me to attend and present my research at the 2017 Society for Epidemiologic Research (SER) Annual Meeting in Seattle, Washington; the 2018 American Thoracic Society (ATS) International Conference in San Diego, California; and the 2018 National Scleroderma Conference in Calgary, Alberta.

List of Abbreviations

ATS	American Thoracic Society
CI	Confidence Interval
CSRG	Canadian Scleroderma Research Group
CTD	Connective Tissue Disease
DLCO	Diffusing Capacity for Carbon Monoxide
DTA	Diagnostic Test Accuracy
ER	Error Rate
ERS	European Respiratory Society
FEV_1	Forced Expiratory Volume in the 1st Second of Forced Exhalation
FEV_1/FVC	Ratio of FEV_1 to FVC
FVC	Forced Vital Capacity
HMM	Hidden Markov Model
HRCT	High-Resolution Computed Tomography
ILD	Interstitial Lung Disease
IPF	Idiopathic Pulmonary Fibrosis
KCO	Carbon Monoxide Transfer
NSIP	Non-Specific Interstitial Pneumonia
NYHA	New York Hear Association
PAH	Pulmonary Arterial Hypertension
PFT	Pulmonary Function Test
PPV	Positive Predictive Value
S.D.	Standard Deviation
SLS I	Scleroderma Lung Study 1
SSc	Systemic Sclerosis
SSc-ILD	Systemic Sclerosis-Associated Interstitial Lung Disease
TLC	Total Lung Capacity
UIP	Usual Interstitial Pneumonia
+LR	Positive Likelihood Ratio
-LR	Negative Likelihood Ratio

List of Tables

List of Figures

Table of Contents

Chapter 1: Introduction

Systemic sclerosis (SSc) is a chronic and progressive autoimmune disease characterized by vasculopathy and excessive connective tissue production.[1] In addition to disfiguring skin involvement, SSc patients can suffer from extensive internal organ damage, including interstitial lung disease (ILD).[2] SSc-associated ILD (SSc-ILD) is the leading cause of morbidity and mortality in SSc patients and, in some cases, can progress very rapidly.[2-5] Unfortunately, existing treatment options cannot reverse the course of pulmonary fibrosis. Immunosuppressants can in fact only stabilize or slow disease progression and can be highly toxic.[3, 5, 6] Therefore, identifying SSc-ILD onset in a timely manner is essential to properly balance the risks and benefits of treatment.

High-resolution computed tomography (HRCT) is the clinical standard for diagnosing SSc-ILD.[7, 8] While new technology and advances aim to diminish radiation exposure,[9] clinicians are still reluctant to perform repeated HRCT scans to monitor SSc-ILD onset. Pulmonary function tests (PFTs), which can detect and follow physiologic parameters of lung restriction consistent with ILD,[10] are thus routinely used as screening tools for SSc-ILD in clinical practice following a baseline HRCT scan. PFT measures also act as surrogate markers for SSc-ILD in epidemiologic studies and as primary endpoints in clinical trials evaluating SSc-ILD treatment efficacy. Yet, there are no guidelines that definitively propose which PFT measure(s) to use.

Forced vital capacity (FVC) is the most commonly used PFT measure by clinicians and researchers, but no scientific rationale for its use has been provided. Indeed, very few validation studies have assessed which, if any, PFT measures best reflect the pathophysiology of SSc-ILD. Yet, FVC is often deemed superior to other PFT measures and is frequently recommended as a main outcome for SSc-ILD studies.[11] An understanding of the current status quo and a well-designed validation study are required to better contextualize and tackle this topic.

1.1 Research Objectives

The overarching goal of this thesis was to determine which PFT measure(s), if any, is (are) a suitable surrogate marker for SSc-ILD onset. The specific objectives were:

i. To describe the use and past validation of PFT measures as surrogate markers for SSc-ILD onset and progression in the literature.
ii. To evaluate the performance of hidden Markov models (HMMs) as a method for validating and using surrogate markers to detect disease onset.
iii. To validate PFT measures, specifically FVC, diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC), as surrogate markers for SSc-ILD onset using HMMs.

1.2 Organization of the Thesis

This is a manuscript-based thesis comprised of three main scientific articles, one per objective. It begins with some background information in Chapter 2 on the etiology, pathophysiology and epidemiology of SSc and SSc-ILD, as well as an overview of PFT measures. Chapter 3 describes the data source that was used and provides some theory on HMMs. Chapter 4 presents the results of a systematic review of the literature which aimed to determine which PFT measures have been most commonly used as markers for SSc-ILD onset and progression in observational and experimental studies. It also summarizes the results of studies that have validated PFT measures against HRCT and lung biopsy results in SSc patients. Chapter 5 is a methodologic study assessing the performance metrics of HMMs, which use complete longitudinal surrogate marker data to determine disease onset. The study uses both simulated and real data to evaluate how HMMs perform compared to the commonly used hard cut-offs and change in surrogate marker level methods of detecting disease presence. Chapter 6 validates FVC, DLCO and TLC as surrogate markers for SSc-ILD onset using HMMs and data from a large, multi-center Canadian cohort of SSc patients. Finally, Chapter 7

summarizes the findings of this research, describes the study limitations and challenges, and discusses the implications of this work for clinical practice and future research.

Chapter 2: Background

2.1 Systemic Sclerosis

Systemic sclerosis (SSc) is an autoimmune disease characterized by early vascular abnormalities followed by excessive collagen production (or fibrosis) ultimately disrupting the underlying integrity of affected tissues.[1] SSc is synonymous with scleroderma, a word derived from the Greek words *skleros* and *derma*, meaning "hard skin".[12] This term accurately describes the physical appearance of individuals who, in addition to disfiguring skin involvement, can have a wide variety of less visible features, including Raynaud's phenomenon, musculoskeletal complications, chronic gastrointestinal symptoms, interstitial lung disease (ILD), cardiac disease, and scleroderma renal crisis.[2]

The prominent physician Sir William Osler best described the gravity of SSc when he wrote: "In its more aggravated forms, [SSc] is one of the most terrible of all human ills. Like Tithonus, to "wither slowly", and like him to be "beaten down and marred and wasted" until one is literally a mummy, encased in an ever-shrinking, slowly contracting skin of steel, is a fate not pictured in any tragedy, ancient or modern."[2]

2.1.1 Etiology and Pathophysiology

The etiology of SSc is poorly understood, but environmental (including occupational and industrial) exposures and genetic factors (specifically polymorphisms in genes associated with the immune system) are posited to contribute to the disease's presentation, albeit weakly.[1, 2, 13] The pathogenesis is complex and involves an interplay of immune abnormalities, microvasculopathy, and disturbances in fibroblastic function.[13, 14]

It is believed that an autoimmune or external attack on endothelial cells is the trigger in the cascade of events leading to SSc. These attacks result in the production of reactive oxygen species leading to

further structural damage of the vasculature and vasoconstriction resulting in tissue ischemia. The overproduction of reactive oxygen species induces the proliferation of extracellular proteins, as well as of anomalous cytokines, growth factors and autoantibodies, thereby increasing inflammation and initiating the appearance of the first clinical symptoms of SSc. Subsequent stages of the disease are characterized entirely by fibrosis leading to derangement of visceral organs.[1, 13, 14] Eventually, a reduction in the factors responsible for the overproduction of extracellular matrix-producing cells causes these cells to undergo apoptosis and results in further internal organ damage.[1]

The intensity and timing of these events can differ from one patient to another, resulting in a disease that is highly heterogeneous with varying clinical manifestations.[14] This often hampers early diagnosis of SSc.[13] Consequently, the American College of Rheumatology (ACR) and the European League against Rheumatism (EULAR) jointly revised previous criteria for the diagnosis of SSc in 2013. With a sensitivity of 0.92 (early cases of SSc being those most often missed) and a specificity of 0.91, the new classification system is now widely accepted as the standard for identifying cases of SSc and has replaced the 1980 classification criteria which lacked senstivity.[15]

The sole sufficient criterion for the diagnosis of SSc is the presence of characteristic skin thickening on both hands proximal to the metacarpophalangeal joints. In the absence of this finding, a combination of manifestations is required to classify a patient as having SSc; these are described in Table 2-1.[2, 15] One such manifestation is the presence of SSc-specific anti-nuclear antibodies, the three most frequent being the anti-centromere, anti-topoisomerase I (or anti-Scl70), and anti-RNA polymerase III antibodies.[1, 2, 13]

Table 2-1: Systemic Sclerosis (SSc) Diagnostic Criteria. *This table was modified from Wigley et al.[2] Patients with a total score ≥9 are classified as having systemic sclerosis.*

Criterion	Sub-Criterion	Score
Skin thickening on both hands proximal to the metacarpophalangeal joints	-	9
Skin thickening of the fingers (Count highest score only.)	Puffy fingers	2
	Sclerodactyly of the fingers (distal to the metacarpophalangeal joints but proximal to the proximal interphalangeal joints)	4
Fingertip lesions (Count highest score only.)	Digital tip ulcers	2
	Fingertip pitting scars	3
Telangiectasia	-	2
Abnormal nailfold capillaries	-	2
Lung Involvement (pulmonary hypertension or interstitial lung disease)	-	2
Raynaud's phenomenon	-	3
SSc-specific autoantibodies (anti-centromere, anti-topoisomerase, anti-RNA polymerase III)	-	3

2.1.2 Subtypes: Limited Cutaneous vs. Diffuse Cutaneous

Since manifestations of SSc can be heterogeneous from patient to patient, experts often think of SSc as a syndrome with a wide spectrum of clinical phenotypes.[1] Classification of SSc into subtypes is valuable both clinically and scientifically, as these subtypes can be indicative of the risk of internal organ involvement and disease progression.[2]

The most common method of classification is based on the extent of skin involvement and thickening and encompasses two main groups: limited cutaneous SSc and diffuse cutaneous SSc.[16] In limited cutaneous SSc, fibrosis of the skin is primarily restricted to the face and limbs distal to the elbows and knees. Conversely, in diffuse cutaneous SSc, fibrosis is more widespread, affecting the trunk and/or proximal limbs, but sometimes sparing the face. Diffuse cutaneous SSc is also more frequently accompanied by multi-organ involvement (the esophagus, heart, kidneys and lungs being hit hardest)

and with mortality.[1, 2] A third but much less common group is sine SSc. Patients with sine SSc have systemic disease but do not experience skin involvement. For simplicity, they are often grouped with limited cutaneous SSc patients.[2] Approximately 55% of patients experience SSc in its limited form, while 35% have diffuse cutaneous SSc and 10% have sine SSc.[17]

2.1.3 Epidemiology and Prognosis

SSc is a rare disease, with a 2008 systematic review of the literature finding incidence estimates ranging from 0.6 to 122 cases per million per year and prevalence estimates fluctuating between 7 and 489 cases per million.[18] The wide range in estimates can be attributed to the use of differing study populations, disease definitions, and methods of ascertainment.[2, 18] In Canada, one study found SSc prevalence to range from 71 to 280 per million throughout Southwestern Ontario, while another estimated it to be 443 per million in Quebec.[19, 20] In general, SSc is believed to affect over 16,000 Canadians.[21] Although all age groups can be affected by SSc, its onset is most frequent between the ages of 35 and 50 years old.[2] Furthermore, SSc is anywhere from three to seven times more prevalent in women than in men.[2] In fact, a Quebec-based population sample study estimated an SSc prevalence in women of 744 cases per million, compared to 133 per million in men.[20]

The occurrence of multi-organ dysfunction and failure in SSc patients results in substantial morbidity and mortality.[1] Indeed, survival is considerably decreased in SSc, with published standardized mortality rates ranging from 1.46 to 7.1.[2] In a large Canadian cohort, the age- and sex-adjusted standardized mortality ratio among incident SSc cases was estimated to be 4.7 (95% confidence interval (CI): 3.6, 5.7), while the mean years of life lost was 23.8 for women and 22.9 for men.[22] Similarly, an American study estimated a median survival of approximately 11 years from SSc diagnosis,[23] with another calculating it to be 7.1 years in patients with diffuse SSc and 15.0 years in patients with limited SSc.[2] Indeed, the prognosis of SSc can be variable, as it is influenced by disease subset, the extent of internal

organ involvement, and the presence of comorbidities. Nevertheless, these estimates are the worst compared to other connective tissue diseases,[13] and have not improved in over 30 years.[22]

Despite being rare, SSc is associated with a high economic burden. One study estimated the annual direct and indirect costs of SSc in the United States to be ~1.5 billion$.[24] Canadian estimates were consistent, with average direct and indirect costs of over 5,000$ and 18,000$ per patient per year, respectively.[25]

SSc is undeniably a serious disease with impacts at both the personal and societal levels. It affects mostly women, during their most productive years and is associated with significant morbidity, increased mortality and high costs. It is also associated with psychological distress, as SSc patients are suddenly faced with an ill-understood disease affecting their physical, emotional and social functions, as well as their overall quality of life.[2] Research in this field has been slow, in large part due to the disease's rarity and heterogeneity.[13, 14] There is currently no known cure for SSc and no available anti-fibrotic agent to stop the progression of the disease. Current long-term combination therapy efforts simply aim to control the autoimmune inflammatory process and manage internal organ complications.[2, 14]

2.2 Systemic Sclerosis-Associated Interstitial Lung Disease

Pulmonary complications in SSc patients generally manifest themselves as ILD, which is often referred to as SSc-associated ILD (SSc-ILD), and/or pulmonary arterial hypertension (PAH). Together, they represent the leading cause of morbidity and mortality in SSc patients with approximately 35% and 28% of SSc-related deaths resulting from SSc-ILD and PAH, respectively.[3, 5] It follows that the study of pulmonary involvement in SSc is of extreme importance. ILD and PAH are however two distinct complications of SSc and should not be studied interchangeably. As such, this dissertation will focus specifically on SSc-ILD, the more common of these two complications.[2]

2.2.1 Etiology and Pathophysiology

While ILD can develop at any time, the risk is greatest in the first few years following SSc diagnosis.[26] The pathogenesis of SSc-ILD is complex and is believed to begin with injuries to both the pulmonary endothelial cells and the alveolar epithelial cells of the lung parenchyma. Inflammation ensues whereby various mediators are released causing an excess production of extracellular matrix proteins by fibroblasts leading to pulmonary fibrosis.[27, 28]

Generally, pulmonary fibrosis can be divided into two main histologic forms depending on the disease's pattern and characteristics: non-specific interstitial pneumonia (NSIP) and usual interstitial pneumonia (UIP).[2, 28] NSIP is defined by homogeneous lesions characterized by inflammation and little tissue destruction, while UIP presents a heterogeneous pattern, along with cellular infiltrates and advanced fibrotic destruction.[5, 27] NSIP is more commonly observed in SSc-ILD patients, accounting for 76% of all ILD cases. On the other hand, UIP is only observed in 11% of cases.[27] In fact, UIP is most often associated with idiopathic pulmonary fibrosis, *i.e.,* pulmonary fibrosis unassociated with any systemic condition or identified exposure. The prognosis of patients with idiopathic pulmonary fibrosis and UIP has been found to be worse than that of patients with SSc-ILD and NSIP,[14, 27] highlighting the fact that these two diseases are pathologically different.

2.2.2 Clinical Manifestations

Clinical presentation of SSc-ILD is often delayed, thereby impeding its early diagnosis.[27] In its early stages, the disease may be completely asymptomatic and undetectable using chest radiography.[2] Furthermore, initial symptoms, such as dyspnea upon exertion and fatigue, are often non-specific and can be related to other features of SSc. Dry, non-productive cough only tends to develop in the later stages of the disease.[2, 27] Perhaps the most telling physical sign of SSc-ILD is the presence of bilateral inspiratory crackles (or "Velcro" rales) on physical examination.[2, 5]

2.2.3 Epidemiology and Prognosis

The prevalence of ILD among SSc patients has been difficult to accurately capture, and estimates have varied widely from 16% to 100%.[27] In particular, ILD prevalence in a Canadian cohort of patients was estimated to be 52%.[29] These discrepancies are likely a result of differing study populations, SSc-ILD definitions, and methods of ascertainment. For instance, autopsy results have revealed parenchymal involvement in up to 100% of patients, while 90% of patients showed interstitial abnormalities on high-resolution computed tomography (HRCT) and 40-75% experienced deteriorating pulmonary function tests (PFTs).[3]

The prognosis of patients with SSc-ILD is poor and survival ranges from a median of three to 15 years depending on the pathologic subset of fibrosis (UIP vs. NSIP).[28] While SSc-ILD is stable or slowly progressing in most,[30] approximately 15% of SSc patients experience rapidly progressive ILD generally in the first years following disease onset, resulting in considerable loss of lung function.[3, 4] The cumulative survival of SSc-ILD patients at ten years from diagnosis is estimated to be between 29% and 69%, lower than the estimated 75% ten-year survival rate of SSc patients without ILD.[3]

ILD is not believed to be reversible in most patients, and as such, stabilisation of fibrosis with slowing of further progression is the target of SSc-ILD treatment.[3, 11] Recent randomized trials have provided some evidence in this regard.[31-34] Stabilisation is most often achieved through immunosuppressive and anti-fibrotic agents, although these therapies are not appropriate for all patients nor are they always sustainable given their toxicity.[3, 5, 6, 34] Indeed, they are associated with serious side effects, including infections, bone marrow suppression and long-term risks of secondary malignancies.[6]

2.3 Monitoring Systemic Sclerosis-Associated Interstitial Lung Disease Onset

Given the disease's delayed clinical presentation and the lack of available treatments to reverse the course of the disease once it becomes more severe, constant screening for the disease is of the utmost

importance. Indeed, it is imperative to identify SSc patients at risk of developing ILD and to identify its onset in a timely manner. Despite the severity of SSc-ILD, there are currently no clinical practice guidelines that suggest how to properly screen for ILD in SSc patients beyond performing an HRCT scan at baseline.[35, 36]

2.3.1 High-Resolution Computed Tomography

HRCT is the clinical standard for diagnosing and assessing the structural extent of SSc-ILD.[7, 8] HRCT is a non-invasive, reliable and reproducible technique that is more sensitive and accurate than chest radiography.[2, 5, 27] The first abnormality to be observed in SSc-ILD patients is the increased presence of subpleural lung markings. This is generally followed by distortions in the architecture of the lungs due to ground-glass opacities and fine reticular patterns (both of which are evocative of NSIP), and occasionally honeycombing (which is more commonly associated with UIP).[2, 3, 5]

Nevertheless, much debate surrounds the routine use and prognostic value of HRCT in screening for SSc-ILD. In fact, most patients with a normal first HRCT scan continue to obtain normal results during follow-up.[37, 38] Though recent advances in technology can help in diminishing radiation exposure through HRCT, the standard radiation dose associated with HRCT of the lungs is of 1.6 millisievert (mSv).[9] The exposure of patients to harmful radiation and the high costs of HRCT thus preclude its repeated use. For these reasons, PFTs at regular intervals are the most commonly used and suggested method to screen for SSc-ILD.[2, 38]

2.3.2 Pulmonary Function Tests

PFTs can serve as a potential screening and monitoring tool for SSc-ILD as they can reveal the presence of restrictive physiologic lung impairment due to ILD, while also being safe, non-invasive, clinically feasible and relatively cheap. PFT measurements are easily standardized, have well-defined ranges of normality, and can be obtained in a uniform manner according to American Thoracic Society

(ATS) standards.[39] They can be expressed in absolute terms or as the percent of a predicted value. These predicted ("reference") values are based on data from healthy subjects and account for sex, age, and height.[40-42] Furthermore, PFT measurements can be obtained repeatedly over time at a relatively low cost. The acquisition of such longitudinal data improves detection of ILD by using patients as their own controls, rather than by simply comparing them with reference values from the general population.

Many PFT measures are impaired in SSc patients with ILD due to a reduction in lung volume, airflow and diffusive conductance.[10] These measures are described in Table 2-2.

Table 2-2: Effect of Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) on Different Pulmonary Function Test (PFT) Measures.

Pulmonary Function Test (PFT) Measure	**Description**[10]	**Observed Effect in Patients with Interstitial Lung Disease (ILD)**[10]
Diffusing Capacity for Carbon Monoxide (DLCO)	Total diffusing capacity of the lungs for carbon monoxide [mL/(min)(mm HG)]	Reduced
Forced Vital Capacity (FVC)	Total volume expelled by a forced exhalation from a maximal inspiration (L)	Reduced
Total Lung Capacity (TLC)	Volume of gas in the lungs at the end of a maximal inspiration (L)	Reduced
Ratio of Forced expiratory volume in 1 second/Forced Vital Capacity (FEV_1/FVC)	Proportion of vital capacity that can be expired in the first second of forced exhalation (L/s)	Normal or high (reduced airflow is proportionate to the reduction in lung volume)

There is currently no consensus on which PFT measure(s) can act as the best screening tool for SSc-ILD onset. A good screening tool should emphasize sensitivity and the reduction of false negative results since SSc-ILD is a serious disease with a short window for the successful administration of stabilizing treatment. Yet, a previous report found PFT measures to have poor sensitivity along with

a high rate of false negative results especially in early cases of ILD.[43] Furthermore, PFT abnormalities can also reflect other unrelated abnormalities in SSc, including anemia, pulmonary arterial hypertension (PAH), myopathy and heart failure.[10]

In addition to being considered as screening tools, PFT measures are also commonly used as outcomes and as surrogate markers for SSc-ILD onset and progression in epidemiologic studies and clinical trials. An ideal surrogate marker should exhibit good accuracy (validity) and reproducibility,[44] yet PFT validation studies are scarce. Thus, no optimal PFT surrogate marker for SSc-ILD has been proposed.

Within the epidemiologic literature, forced vital capacity (FVC) seems to be the preferred PFT marker for SSc-ILD.[45] However, there are some indications that it may not be the best option. For instance, SSc-ILD clinical trials in which FVC was used as the primary endpoint only identified at best a modest treatment effect,[31-33] while observational studies using FVC as an outcome of interest have not been able to identify consistent predictors of SSc-ILD.[46] Furthermore, FVC's minimal clinically important difference in SSc-ILD is approximately 3-5%,[47] yet yearly random variations in FVC of 10-15% are well-documented.[48] Finally, FVC remains generally stable over time in SSc,[49] despite evidence of radiological progression in over 65% of patients.[37, 50]

Diffusing capacity for carbon monoxide (DLCO) is sometimes also used as a surrogate marker as it is sensitive for SSc-ILD, but is generally not favoured since it can be confounded by the presence of pulmonary vascular disease, such as PAH and anemia.[45, 51] Yet, DLCO is known to be one of the first PFT measures to be impaired by ILD onset.[2] On the other hand, ATS clearly characterizes restrictive lung diseases such as ILD by a reduction in total lung capacity (TLC).[42]

It is clear that uncertainty abounds and yet many decisions relating to SSc-ILD rely heavily on PFT measures. For instance, they help rheumatologists identify which SSc patients should be investigated

further for ILD. They are also used by researchers as markers of SSc-ILD progression, to help identify predictors of rapid decline.[46] Perhaps most importantly, they serve as primary endpoints in randomized clinical trials on which millions of dollars are spent to study the efficacy of novel drugs for SSc-ILD.[31-34] With much at stake, it is imperative to identify which PFT measure(s), if any, are good markers of SSc-ILD. Proposing a suitable PFT surrogate marker for SSc-ILD would improve both screening practices and the validity of epidemiologic research in this area.

Chapter 3: Overview of Data Source and Analytic Methodology

A multi-center cohort design using retrospective data was implemented to address the second and third objectives of this thesis. The data for these two studies was provided by the Canadian Scleroderma Research Group (CSRG) registry, the largest longitudinal cohort of Canadian systemic sclerosis (SSc) subjects.

3.1 Canadian Scleroderma Research Group

The CSRG is a Canadian Institutes of Health Research (CIHR)-funded and internationally recognized multidisciplinary team of medical doctors and scientists who aim to improve the lives of SSc patients by maximizing the potential for high-impact SSc research in Canada.[52] Since its inception in 2003, the CSRG has involved 15 sites across Canada, one site in Mexico, and has recruited over 1,600 patients, accounting for approximately 10% of all SSc patients in Canada.[21] This makes the CSRG registry one of the largest SSc cohorts in the world.

Subjects enrolled in the CSRG registry must have a rheumatologist-confirmed diagnosis of SSc, be 18 years of age or older, be fluent in English or French, and be likely to comply with study procedures. Over 98% of the cohort meet the 2013 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) classification criteria for SSc.[53]

Enrolled patients are required to complete a baseline visit and are subsequently followed prospectively with yearly standardized visits. Socio-demographic characteristics, symptoms and other patient-reported outcomes are collected in a patient case report form. Physicians document the subjects' clinical histories and physical examinations using a physician case report form. Blood tests, pulmonary function tests (PFTs), chest X-rays and cardiac echocardiograms are performed yearly per protocol,

while high-resolution computed tomography (HRCT) scans are only performed at the discretion of the study physician, presumably when pulmonary complications are suspected.

3.1.1 Cohort and Baseline Characteristics

All analyses in the second and third objectives of this thesis were performed using CSRG data stemming from participants' recorded study visits having occurred during the period from September 2004 to July 2017. This dataset included 1,665 SSc subjects with as many as 13 annual visits and five visits on average. The total number of visits was 8,314 (Table 3-1).

Table 3-1: Frequency Table of Study Visits in Canadian Scleroderma Research Group (CSRG) Patients.

Number of Visits	Number of SSc Patients
1	306
2	177
3	167
4	182
5	162
6	141
7	140
8	116
9	93
10	68
11	46
12	33
13	34

Women comprised 86.4% of patients in the dataset while men accounted for 13.6%. Most study subjects (86.1%) identified as white. Approximately 30% of patients had interstitial lung disease (ILD) at baseline. ILD status was assessed using the combination gold-standard described in Section 3.1.3 below. The study subjects' baseline characteristics stratified by ILD status are summarized in Table 3-2.

Table 3-2: Baseline Characteristics of Patients Enrolled in the Canadian Scleroderma Research Group (CSRG) Registry.

	All Patients (N = 1,665)		**Patients without Interstitial Lung Disease at Baseline (N = 1,121)**		**Patients with Interstitial Lung Disease at Baseline (N = 497)**	
	Mean (% or S.D)	**Missing Values (%)**	**Mean (% or S.D)**	**Missing Values (%)**	**Mean (% or S.D)**	**Missing Values (%)**
Sociodemographic Characteristics						
Sex (%) *						
Women	1,438 (86.4)	0 (0)	994 (88.7)	0 (0)	410 (82.5)	0 (0)
Men	227 (13.6)		127 (11.3)		87 (17.5)	
Age (S.D.)	55.2 (12.3)	2 (0.1)	54.6 (12.5)	0 (0)	56.9 (11.8)	2 (0.4)
Ethnicity (%) *						
White	1,329 (86.1)	122 (7.3)	897 (86.5)	84 (7.5)	399 (85.4)	30 (6.0)
Black	20 (1.3)		10 (1.0)		9 (1.9)	
Smoking Status (%) *						
Never	626 (40.7)	128 (7.7)	412 (39.8)	85 (7.6)	199 (42.8)	32 (6.4)
Past Smoker	704 (45.8)		463 (44.7)		230 (49.5)	
Current Smoker	207 (13.5)		161 (15.5)		36 (7.7)	
Clinical Manifestations of Disease						
Disease Duration, Years (S.D.)	9.8 (9.3)	26 (1.6)	9.6 (9.2)	7 (0.6)	10.1 (9.5)	2 (0.4)
Disease Extent (%) *						
Limited/Sine	1,040 (63.4)	24 (1.4)	769 (69.0)	6 (0.5)	256 (51.6)	1 (0.2)
Diffuse	601 (36.6)		346 (31.0)		240 (48.4)	
Pulmonary Hypertension (%) *	145 (10.3)	260 (15.6)	76 (8.0)	168 (15.0)	69 (15.9)	63 (12.7)
Anti-Nuclear Antibodies (%) *	1,106 (95.3)		748 (95.2)		339 (95.5)	
Anti-Centromere	399 (34.4)	504 (30.3)	330 (42.0)	335 (29.9)	61 (17.2)	142 (28.6)
Anti-Topoisomerase	177 (15.9)	549 (33.0)	84 (11.3)	375 (33.5)	91 (25.9)	145 (29.2)
Anti-RNA Polymerase III	147 (18.3)	860 (51.7)	94 (17.7)	591 (52.7)	53 (20.3)	236 (47.5)
Shortness of Breath						
Patient-Reported (S.D.) (Numerical Rating Scale 0 to 10)	2.0 (2.6)	138 (8.3)	1.6 (2.3)	93 (8.3)	2.9 (2.8)	34 (6.8)

NYHA Function Class (%) *						
Class I	850 (52.1)	35 (2.1)	681 (61.4)	11 (1.0)	149 (30.3)	6 (1.2)
Class II	641 (39.3)		365 (32.9)		268 (54.6)	
Class III	120 (7.4)		56 (5.0)		63 (12.8)	
Class IV	19 (1.2)		8 (0.7)		11 (2.2)	
Pulmonary Function Tests						
DLCO (S.D.)						
Absolute, mL/(min)(mm HG)	15.9 (5.3)	452 (27.1)	17.0 (5.1)	292 (26.0)	13.3 (5.0)	126 (25.4)
% Predicted	70.3 (21.4)	426 (25.6)	75.3 (19.7)	272 (24.3)	59.0 (20.8)	120 (24.1)
FVC (S.D.)						
Absolute, L	3.0 (0.8)	369 (22.2)	3.1 (0.8)	228 (20.3)	2.6 (0.8)	108 (21.7)
% Predicted	91.8 (19.4)	262 (15.7)	96.9 (17.3)	166 (14.8)	80.4 (19.4)	68 (13.7)
TLC (S.D.)						
Absolute, L	4.7 (1.1)	475 (28.5)	4.9 (1.0)	303 (27.0)	4.1 (1.0)	138 (27.8)
% Predicted	93.6 (18.3)	368 (22.1)	98.7 (15.7)	240 (21.4)	82.1 (18.6)	100 (20.1)
FEV_1/FVC (S.D.)	78.7 (9.7)	336 (20.2)	77.8 (9.3)	212 (18.9)	80.6 (10.3)	91 (18.3)

**Patients with missing data were not included in the denominator of the % calculations.*

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; NYHA = New York Heart Association; S.D. = Standard Deviation; TLC = Total Lung Capacity

Patients were on average 55.2 years of age and ranged from 18 to 88 years old at baseline. The mean duration of SSc disease presence since the first non-Raynaud's symptom was 9.8 years. Among patients without ILD, 69.0% had limited cutaneous or sine SSc while the remaining 31.0% had the diffuse cutaneous form of disease. On the other hand, almost half of patients with ILD (48.4%) had diffuse cutaneous SSc. Pulmonary hypertension was rare but was more common in patients with ILD than in those without (15.9% vs. 8.0%, respectively).

Unsurprisingly, CSRG patients with ILD at baseline were more likely to experience signs and symptoms consistent with restrictive lung disease, such as dyspnea. Dyspnea was evaluated using a patient-reported numerical rating scale from 0 to 10 indicating no to severe limitations to daily activities respectively. It was also assessed by treating physicians using the New York Heart Association (NYHA) functional class.[54] Both methods of evaluation found that patients with ILD were more limited by shortness of breath in their physical activity. Additionally, their baseline measurements of absolute and % predicted forced vital capacity (FVC), diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC) were all reduced, while the ratio of forced expiratory volume in one second (FEV_1) to FVC (FEV_1/FVC) was slightly elevated compared to patients without ILD.

3.1.2 Availability of High-Resolution Computed Tomography Results

Ideally, HRCT results would be the reference test used to validate the different PFT measures in search of the best surrogate marker for SSc-associated ILD (SSc-ILD) onset. However, a review of the availability of HRCT scans in the CSRG dataset revealed that relying solely on these results would be problematic. Throughout the course of the almost thirteen-year follow-up period, only 797 HRCT scans on 574 subjects were recorded in the database. This is in stark contrast to the large number of available PFT measurements recorded. For illustrative purposes, the numbers of patients, PFT

measurements and HRCT scans available during patients' first five study visits are depicted in Figure 3-1.

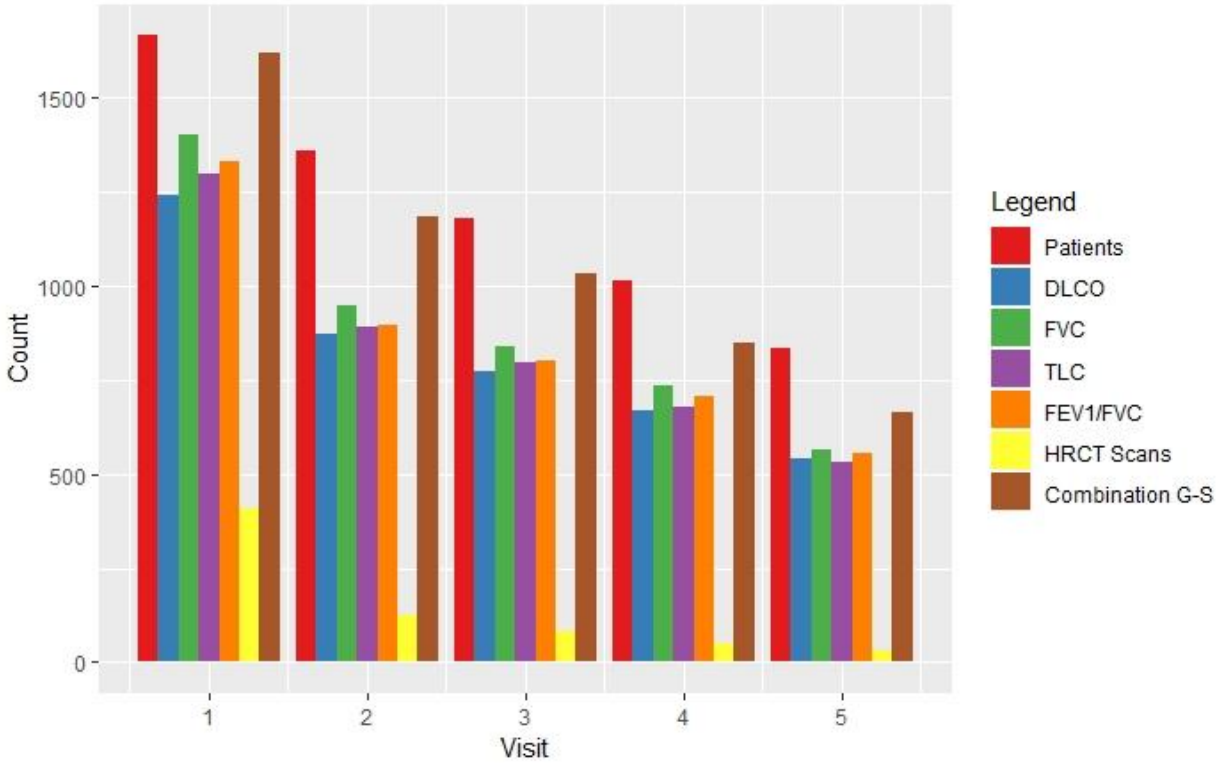


Figure 3-1: Number of Available Patients, Pulmonary Function Test (PFT) Measurements, High-Resolution Computed Tomography (HRCT) and Combination Gold-Standard Results at Canadian Scleroderma Group (CSRG) Visits One through Five. *The combination gold-standard is defined and discussed in Section 3.1.3.*
Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; G-S = Gold-Standard; TLC = Total Lung Capacity

As discussed in section 2.3.1, HRCT is not routinely performed due to repeated radiation exposure and high costs.[9] Hence the group of CSRG patients who undergo an HRCT evaluation is primarily comprised of patients in whom lung disease is suspected due to worsening PFT results and/or symptoms. Indeed, a 2018 global survey of rheumatologists and SSc experts found that abnormal FVC and DLCO values were an important indication for chest HRCT.[35] Ordering an HRCT based on abnormal or lowered PFT measurements is also what is recommended in leading textbooks.[2]

Consequently, the use of HRCT scans as the reference test in our validation study would lead to partial verification bias.

Partial verification bias occurs in studies of diagnostic test accuracy when only a subset of subjects receives or undergoes the reference test. Although the effect of verification bias remains to be completely quantified, partial verification bias will generally overestimate an index test's sensitivity by underestimating the number of false negatives.[55] Furthermore, the strict use of HRCT results as the reference test would result in a greater prevalence of ILD cases in the validation sample. Accordingly, we would expect this to artificially increase positive predictive value and decrease negative predictive value.

3.1.3 The Combination Gold-Standard

To circumvent the potential for verification bias, cases of SSc-ILD were diagnosed in our studies using a combination gold-standard. The combination gold-standard was inspired by a 2012 clinical decision rule by Steele et al. used to predict the presence of SSc-ILD. This clinical decision rule surmises that ILD is present if at least one of the following two conditions is met: (1) chest X-ray reports the presence of increased interstitial markings or fibrosis; or (2) basilar Velcro-like crackles are reported by a study physician on lung auscultation.[29] The latter have been shown to predict the presence of ILD on HRCT.[56, 57]

The diagnostic properties of this clinical decision rule were calculated using patient data from the CSRG registry. HRCT was used as the reference test and multiple imputation was performed to prevent verification bias. The algorithm was found to have a fair sensitivity of 51.3% (95% confidence interval (CI): 44.3, 58.2), but otherwise good performance characteristics with a specificity of 86.5% (95% CI: 82.7, 90.3), a positive predictive value of 80.7% (95% CI: 75.2, 86.2), and a negative predictive value of 61.7% (95% CI: 52.6, 70.8).[29]

The clinical decision rule was also validated externally in 2017 by Hax et al. in a Brazilian cohort of 177 SSc patients. In that study, varying degrees of lung involvement on HRCT (any ILD, extent of ILD ≥ 10%, extent of ILD ≥ 20%) were used as the reference tests. The algorithm's sensitivity varied from 58.6 to 94.7 depending on the reference test used. Its specificity ranged from 58.9 to 65.6, while its positive predictive value and negative predictive value varied from 37.5 to 70.8 and 46.7 to 97.7, respectively.[58]

The combination gold-standard was created by modifying the clinical decision rule slightly to include HRCT results when available. In other words, the combination gold-standard would use HRCT results when available and otherwise would diagnose a case of ILD if: (1) interstitial markings or fibrosis were present on chest X-ray; or (2) basilar Velcro-like crackles were present on lung auscultation**.** Since chest X-ray and lung auscultations were routinely administered, most CSRG patients had yearly results available for the combination gold-standard thereby precluding the risk of any verification bias (Figure 3-1).

3.2 Hidden Markov Models

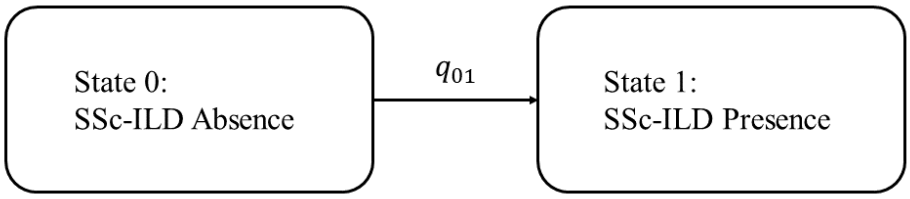
The use of hidden Markov models (HMMs) figures prominently in this thesis as they were the analytic tool chosen to ascertain SSc-ILD onset using PFT measures. They are briefly described in this section and are discussed in more detail in Appendix H.

3.2.1 Markov Model Theory

HMMs are an application of Markov models, a type of stochastic model that describes the process of transitioning from one discrete state to another. These processes can be depicted graphically using separate boxes for each state and arrows representing possible or allowable transitions between states. In the context of SSc-ILD onset, the Markov chain would include two disease states with states 0 and 1 representing SSc-ILD absence and presence, respectively. Since fibrosis of the lungs is an irreversible

process,[11] only transitions from state 0 to state 1 are permitted (Figure 3-2a). An additional third state could also be included to represent death (Figure 3-2b). Such states are considered all-absorbing as it is impossible to transition out of them.[59-62]

a)



b)

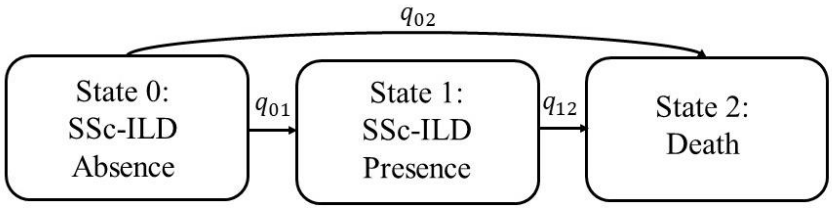


Figure 3-2: Multi-State Markov Models Depicting Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Onset. *a) Two-state model for SSc-ILD onset. b) Three-state model for SSc-ILD onset with a death absorbing state.* q_{01} *represents the instantaneous risk of progressing from SSc-ILD absence to SSc-ILD presence;* q_{02} *represents the instantaneous risk of death in SSc subjects without ILD;* q_{12} *represents the instantaneous risk of death in SSc subjects with ILD.*
Abbreviations: ILD = Interstitial Lung Disease; SSc = Systemic Sclerosis; SSc-ILD = Systemic Sclerosis-Associated Interstitial Lung Disease

For each pair of states *r* and *s*, a transition intensity q_{rs} denotes the instantaneous risk of transitioning from state *r* to state *s*. Mathematically, this can be written as:

$$q_{rs} = \lim_{\delta t \to 0} P(S(t + \delta t) = s | S(t) = r)/\delta t$$

where $S(t)$ represents the state occupied at time *t*.

Thus, for an *n*-state model, an *n* x *n* transition intensity matrix *Q* can be formalized, where the *(r, s)* entry corresponds to the instantaneous risk of transitioning from state *r* to state *s*. An entry of zero

indicates an undefined transition. The entries along the diagonal are chosen such that the rows of the matrix sum to zero. Hence, in the case of SSc-ILD onset with death, the 3 x 3 transition intensity matrix would take the form $Q = \begin{bmatrix} -q_{01} - q_{02} & q_{01} & q_{02} \\ 0 & -q_{12} & q_{12} \\ 0 & 0 & 0 \end{bmatrix}$ where q_{01} denotes the instantaneous risk of ILD onset, q_{02} represents the instantaneous risk of death in SSc subjects without ILD onset, and q_{12} represents the instantaneous risk of death in SSc subjects with ILD.[61-63]

The likelihood of observed data is calculated from the transition probability matrix $P(\Delta t)$ in which entry *(r, s)* corresponds to the probability of transitioning from states *r* to *s* during a time interval of length Δt. This probability matrix is equal to the matrix exponential of Q:

$$P(\Delta t) = \exp(\Delta t Q)$$

The Markov property (or memoryless assumption) is required to build the full likelihood of Q from the transition probability matrix $P(\Delta t)$ and thus estimate the transition intensities. It assumes that future progression is conditional only upon the current state occupied by a subject and not on the historical trajectory taken to reach the current state.[64, 65]

Markov models are very useful tools when state occupancy is known throughout the study period. However, states are sometimes indirectly observed or inferred using surrogate markers. For instance, SSc-ILD tends to be monitored using pulmonary function testing instead of directly observed through HRCT scans. Markov models are not suitable in these circumstances since surrogate markers and biomarkers are generally not distributed bimodally impeding the possibility of directly observing separate states in the study population. Furthermore, disease markers can be prone to random variation and measurement error. HMMs offer a potential solution.

3.2.2 Hidden Markov Model Theory

HMMs are so aptly named as they can be used when state occupancy is unknown (or hidden) and is instead inferred using marker measurements. HMMs will use a marker's entire measurement history to estimate the instantaneous risk of progressing from one state to another while simultaneously correcting for potential misclassification of the observed state due to measurement error in the marker. This is possible due to the HMM's hierarchical nature: it is comprised of both a measurement error model and a Markov model (Figure 3-3).[63, 66]

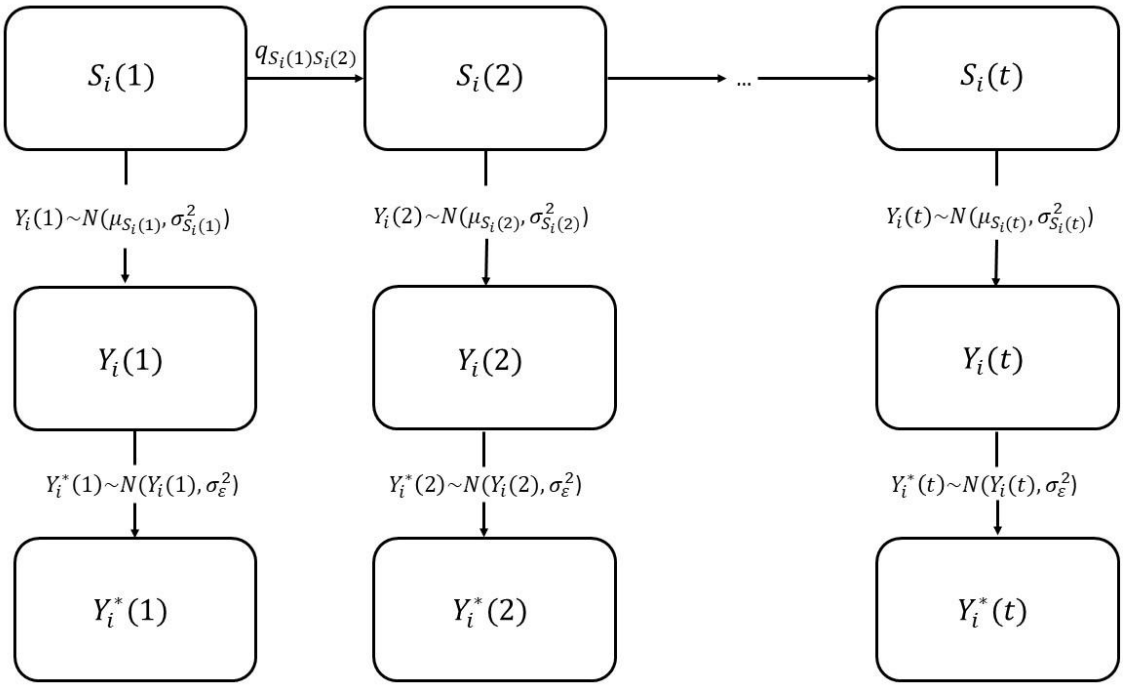


Figure 3-3: Schematic Representation of a Hidden Markov Model (HMM). *The top portion of the figure represents a typical Markov process for a subject i who finds themselves in state $S_i(1)$ at time point 1, $S_i(2)$ at time point 2, and so on. The probability of transitioning from one state to the next is governed by a specific transition intensity q. The middle component of the figure assumes an underlying distribution for the marker given the disease state, The bottom component corresponds to the measurement error model which relates observed levels (Y*) of the marker to their true levels (Y) using a classical measurement error model. In this figure, the true marker levels Y are assumed to be normally distributed with means and variances dependent on the state occupied.*

As transitions from one state to another occur, the occupancy of each state will in turn give rise to a *true* marker level denoted by random variable *Y*. It follows that *Y* can only take on certain values

conditional on the state occupied. Thus, for each state an underlying probability distribution for *Y* can be specified. For example, PFT values can be normally distributed in both the absence and presence of SSc-ILD but centered around different means and characterized by different spreads. This relationship between states and *true* marker levels is represented in the top half of Figure 3-3.

When markers are subject to measurement error, random variable *Y* is not directly observed. Instead, what is observed is the imperfectly measured random variable *Y**. Thus, an *observed* marker measurement for patient *i* at time *t*, Y^*_{it}, can differ from their underlying *true* marker value, Y_{it}. The measurement error component of the HMM thus relates these two variables using a classical measurement error model:

$$Y_{it}^* = Y_{it} + \varepsilon_{it}$$

where ε_{it} are assumed to be independent and identically normally distributed, with a mean of zero and common variance σ^2. This is depicted in the bottom half of Figure 3-3.[63, 66] It is worth noting that when using an underlying normal distribution for *Y*, the resulting normal-normal implies that the common variance parameter σ^2 cannot be estimated as it is confounded with and cannot be disentangled from the variability of *Y*.

Nevertheless, HMMs can be very useful tools in the realm of SSc-ILD onset. They can use all PFT values recorded in an SSc cohort to estimate the instantaneous risk of ILD onset while correcting for measurement error associated with these PFT measures. However, HMMs are made even more powerful by their ability to predict ILD presence for each patient at any point in time given their history of PFT measurements. This is achieved by the forward-backward algorithm.

3.2.3 Forward-Backward Algorithm

The forward-backward algorithm is an inference algorithm specific to HMMs which allows us to compute the marginal probability of disease state occupancy at time *t* given the available sequence of marker measurements. While its functioning is beyond the scope of this thesis, it can be thought of, in the simplest of terms, as an application of Bayes' rule.[67]

According to Bayes' theorem:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

where:

$P(A|B) = P(S_i(t)|Y_i^*(t), \dots, Y_i^*(2), Y_i^*(1))$. This is the probability of ILD presence in subject *i* given all their PFT measurements recorded up until time point *t*.

$P(B|A) = P(Y_i^*(t), \dots, Y_i^*(2), Y_i^*(1)|S_i(t_k))$. This probability can easily be calculated using the Markov property.

$P(A) = P(S_i(t_k))$. This probability can be obtained through the estimated transition intensities.

$P(B) = P(Y_i^*(t), \dots, Y_i^*(2), Y_i^*(1))$. This corresponds to the joint distribution of the observed PFT measurements.

Chapter 4: Historical Use and Performance of Pulmonary Function Test Measures as Outcomes for Systemic Sclerosis-Associated Interstitial Lung Disease

4.1 Preface to Manuscript 1

The use of forced vital capacity (FVC) as a pulmonary function test (PFT) surrogate marker for systemic sclerosis-associated interstitial lung disease (SSc-ILD) is ubiquitous. Yet, there is reason to believe that it may not be the best option. A limited search of the literature failed to reveal how and why FVC became so popular. Indeed, there appears to exist few studies having evaluated its validity, as well as that of other PFT measures. We aimed to better understand whether FVC's apparent universal use is warranted and whether there was probable cause to explore the performance of other PFT measures as surrogate markers for SSc-ILD.

We performed a systematic review of the literature to outline the historical use and validation of PFT measures as surrogate markers for SSc-ILD. Five electronic databases were searched for all eligible studies which either used at least one PFT measure as a longitudinal outcome for SSc-ILD and/or reported at least one classical measure of validity. While the remainder of this thesis focuses specifically on SSc-ILD onset, this systematic review took a broader approach to the topic by dealing with both SSc-ILD onset and progression.

The resulting manuscript, entitled "Pulmonary Function Tests as Outcomes for Systemic Sclerosis Interstitial Lung Disease", was published in European Respiratory Review: an official journal of the European Respiratory Society (ERS) (2018; 27:148). ERS journals allow authors to retain the rights to use of their whole article in a thesis and as such no written permission from the publisher was required.

4.2 Title Page

Title: Pulmonary Function Tests as Outcomes for Systemic Sclerosis Interstitial Lung Disease

Authors: Melissa Caron[1, 3]; Sabrina Hoa[1, 2]; Marie Hudson[2]; Kevin Schwartzman[1, 3]; Russell Steele[4]

Affiliations:

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

[2] Division of Rheumatology, Jewish General Hospital, Montreal, QC, Canada

[3] Respiratory Epidemiology and Clinical Research Unit (RECRU), Montreal Chest Institute, McGill University Health Centre, Montreal, QC, Canada

[4] Department of Mathematics and Statistics, McGill University, Montreal, QC, Canada

Correspondence: Melissa Caron, Respiratory Epidemiology and Clinical Research Unit (RECRU), 5252 de Maisonneuve West, Montreal, QC, H4A 3S5, Canada. E-mail: melissa.caron@mail.mcgill.ca

Take-Home Message: Despite FVC's popularity, available evidence does not favour it as the best PFT surrogate marker for SSc-ILD.

Conflict of Interest: M. Caron reports grants from the Fonds de Recherche du Québec (Santé PhD Studentship) and from the Canadian Institutes of Health Research (GSD – 146268) during the conduct of the study. S. Hoa reports grants from the Université de Montréal Rheumatology Clinical Fellowship Program (Abbvie educational grant) and from the Arthritis Society's Postdoctoral Fellowship Award, during the conduct of the study.

4.3 Abstract

Interstitial lung disease (ILD) is the leading cause of morbidity and mortality in systemic sclerosis (SSc). We performed a systematic review to characterize the use and validation of pulmonary function tests (PFTs) as surrogate markers for SSc-ILD progression.

Five electronic databases were searched to identify all relevant studies. Included studies either used at least one PFT measure as a longitudinal outcome for SSc-ILD progression (*i.e.,* outcome studies) and/or reported at least one classical measure of validity for the PFTs in SSc-ILD (*i.e.,* validation studies).

This systematic review included 169 outcome studies and 50 validation studies. Diffusing capacity for carbon monoxide (DLCO) was cumulatively the most commonly used outcome until 2010 when it was surpassed by forced vital capacity (FVC). FVC% was the primary endpoint in 70.4% of studies, compared to 11.3% for DLCO%. Only five studies specifically aimed to validate the PFTs: two concluded that DLCO was the best measure of SSc-ILD extent, while the others did not favour any PFT. These studies also showed respectable validity measures for total lung capacity (TLC).

Despite the current preference for FVC, available evidence suggests that DLCO and TLC should not yet be discounted as potential surrogate markers for SSc-ILD progression.

4.4 Introduction

Systemic sclerosis (SSc) is a chronic and progressive autoimmune disease involving a complex interplay of microvasculopathy, disturbances in fibroblastic function and abnormalities of the immune system.[1, 13] In addition to disfiguring skin involvement, SSc patients can suffer from extensive internal organ damage, including interstitial lung disease (ILD).[68]

SSc-ILD is the leading cause of morbidity and mortality in SSc,[28] and is estimated to occur in over 50% of patients.[29] The prognosis of patients with SSc-ILD can be poor, as it is estimated that approximately 15% of patients will experience rapidly progressive ILD.[28, 69] For these reasons, ILD presence and progression are routinely monitored using pulmonary function tests (PFTs).

Since SSc-ILD was first described in 1949,[70] various PFT measures have been used to characterize its progression. Yet, forced vital capacity (FVC) has become the preferred surrogate marker for SSc-ILD, despite the paucity of any validation studies. To better understand the rationale behind FVC as the preferred outcome measure for SSc-ILD, we performed a systematic review of the literature aiming to outline the historical use and validation of PFT measures as surrogate markers for SSc-ILD progression.

The first objective was to determine the frequency at which the different PFT measures were used as outcomes in the study of SSc-ILD onset and progression. We then aimed to summarize the results of studies that validated PFT measures against either high-resolution computed tomography (HRCT) or lung biopsy results in SSc patients. The results of this systematic review thus not only describe changing practices in the use of PFTs over time in SSc-ILD, but also assess the extent to which these trends were supported by available evidence.

4.5 Material and Methods

The protocol for this review was registered with the Prospero international prospective register of systematic reviews (CRD42016039565). IRB approval was not required.

Eligibility Criteria

Only studies published in English or in French were considered. To be as inclusive and comprehensive as possible, all original research, including published conference abstracts and clinical trial registrations, were eligible for inclusion. No restrictions were placed on study design. Reviews, summary articles, case reports, commentaries, letters and editorials were excluded.

Eligible studies had to include patients with a diagnosis of SSc. Furthermore, SSc-ILD had to be the focus of the study. All included studies either: (1) used at least one PFT measure as a longitudinal outcome for SSc-ILD; and/or (2) reported at least one classical measure of validity for the association between PFTs and either HRCT findings or lung biopsy results in SSc. The former studies are referred to hereafter as outcome studies and the latter as validation studies. Finally, a minimum of 20 SSc patients was required for inclusion in the review.

Information Sources and Search Strategy

The MEDLINE (PubMed), Embase (Ovid), and Web of Science databases were searched on July 5th, 2016 to identify all potentially relevant studies since January 1949. ClinicalTrials.gov and the World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) were also searched for clinical trial registrations. Additionally, the reference lists of all studies identified via the electronic searches and adhering to the inclusion criteria were manually curated for further potentially eligible studies.

The search strategy, designed in consultation with a professional librarian, consisted of three main search terms: (1) systemic sclerosis, (2) interstitial lung disease, and (3) pulmonary function test. The detailed search strategy for the MEDLINE database is available in Appendix A.

Study Selection

The title and abstract of each article were assessed for potential eligibility independently by two reviewers. Articles were only excluded if both reviewers came to this decision unanimously. Next, the reviewers separately performed a full-text assessment of the remaining articles to confirm their inclusion. Any disagreements were resolved by consensus. The kappa statistic was calculated at both stages of the screening process to assess inter-reviewer agreement.

Data Collection

Data from the included studies was extracted using a standardised, pre-piloted form. Data extracted from the outcome studies included the PFT measures used as outcomes for SSc-ILD progression, as well as the reasoning provided for these choices. Data collected from the validation studies included the chosen clinical reference standard, the HRCT scoring system utilized (if applicable) and the reported measures of validity.

Synthesis of Results

For the outcome studies, the cumulative use over time of the different PFT measures was graphically depicted. The reasoning provided for the choice in primary PFT outcome measure was also summarized. For all analyses, all variations of diffusing capacity for carbon monoxide (DLCO), including DLCO adjusted for alveolar volume and DLCO corrected for haemoglobin, were grouped together into one DLCO measure.

Among the validation studies, we identified those whose primary goal was specifically to validate the different PFT measures in SSc-ILD. We focused explicitly on these validation studies and summarized their reported measures of validity in forest plots. Given the variability in study populations and ILD scoring methods, no attempts were made to meta-analyze the data. The results of the remaining validation studies, whose aim was not specifically to validate PFT measures, are presented as supporting information.

Risk of Bias

The quality of the outcome studies was not assessed, as the purpose of this review was not to summarize measures of effect. The quality of the studies whose specific aim was to validate PFTs in SSc-ILD was evaluated using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[71] Publication bias was not assessed as there are, to our knowledge, no available methods to do so in the context of screening/diagnostic test accuracy.

4.6 Results

An adaptation of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram, outlining the study selection process, is available in Figure 4-1.[72] The detailed search strategy retrieved 3,258 records. Following de-duplication, the titles and abstracts of 2,056 records were screened to assess eligibility. Both co-reviewers independently agreed on the exclusion of 1,415 records, with an overall percent agreement of 0.91 and a kappa statistic of 0.78. To confirm inclusion, a full-text assessment was performed on the remaining 641 records of which 384 were excluded. The overall percent agreement and kappa statistic for the secondary screening were 0.85 and 0.69, respectively. The reasons for exclusion at both the primary and secondary screening stages are presented in Appendix B.

Following the screening process, an additional 55 abstracts and clinical trial registrations were excluded since they matched subsequently published full-text articles. Finally, the reference lists of the 202 included studies were manually searched for further potentially relevant studies. This resulted in the addition of 10 records, for a total of 212 studies included in this systematic review. Of the 212 included records, 169 qualified as outcome studies,[30-32, 43, 49, 73-236] while 50 were listed as validation studies.[43, 107, 133, 169, 206, 212, 225, 237-279] Seven records satisfied the inclusion criteria for both outcome and validation studies.[43, 107, 133, 169, 206, 212, 225]

Outcome Study Results

The relevant study characteristics of the 169 outcome studies are summarized in Appendix C. The earliest identified longitudinal study using PFTs as outcomes for SSc-ILD progression was published in 1982.[30] For most of the study period, DLCO was cumulatively the most commonly used PFT outcome followed closely by FVC. In 2010, FVC surpassed DLCO as the PFT most often used (Figure 4-2a). Similarly, DLCO's percent cumulative use (*i.e.,* the yearly ratio of the cumulative count of a given PFT to the cumulative number of published articles) gradually decreased over time in favour of an increase in the percent cumulative use of FVC. These trends were especially prominent after 2005 (Figure 4-2b).

Interesting trends in the use of absolute and percent predicted (denoted by %) PFT values were also observed and are further described in Appendix D.

Among the 169 outcome studies, 71 clearly specified a single primary PFT endpoint for SSc-ILD progression. While DLCO% comprised only a small proportion of primary PFT endpoints overall (11.3%), Figure 4-3 reveals that it was a relatively important main outcome in the 1980's and 1990's, only to be surpassed by FVC% in the mid-2000's (which accounted for 70.4% of studies overall).

Of these 71 studies, 42 (59.2%) did not report a reason to support their choice in outcome, 20 (28.2%) alluded to the main PFT measure's use in previous SSc-ILD studies, five (7.0%) stated that the chosen PFT had been previously validated, while four (5.6%) reported using it either due to its high sensitivity or specificity.

Interestingly, of the five studies that cited prior validation to justify their choice in PFT (FVC% in four cases,[49, 129, 171, 174] and DLCO% in the last),[95] only the study using DLCO% referenced an article with reported measures of validity. The four studies which chose their main PFT outcome measure based on its high sensitivity (in the case of DLCO),[73, 76] or high specificity (in the case of FVC),[188, 227] either did not provide a citation to support this claim,[227] referred to studies performed outside the field of SSc-ILD,[73, 76, 188] or cited SSc studies with fewer than twenty patients.[73, 76]

Validation Study Results

The relevant characteristics of the 50 validation studies are summarized in Appendix E, while their measures of validity are reported in Appendix F. Among these studies, only five had the clear objective of validating the different PFT measures in SSc-ILD.[43, 238, 268, 271, 279] Ideally, identifying the best surrogate marker for SSc-ILD progression should involve the comparison of longitudinal variations in PFT measures with concurrent variations in either HRCT scores or lung biopsy results. This would ensure that the surrogate marker changes along with the reference standard. Yet, all five of these validation studies performed cross-sectional assessments of PFT results against HRCT-assessed SSc-ILD severity.

Three of these five studies correlated PFT values with HRCT scores.[238, 268, 279] Forest plots of the correlation coefficients are available in Figure 4-4. The first of these studies, published in 1997, concluded that DLCO was the best measure of SSc-ILD extent (Figure 4-4a).[238] The authors of the

second study, published in 2013, reported that PFTs alone may not be sufficient to identify cases of SSc-ILD (Figure 4-4b).[268] Finally, the third study, published in 2016, concluded in favour of DLCO% being the best surrogate marker for SSc-ILD extent at any one point in time (Figure 4-4c).[279]

The remaining two studies to have validated PFT measures reported measures of sensitivity and specificity (Figure 4-5).[43, 271] Neither study concluded in favour of any specific PFT. Rather, they both warned against using only PFTs to identify ILD in SSc patients and suggested evaluating more inclusive screening tools.

The quality of these five validation studies is described in Appendix G. All studies exhibited some risk of bias, generally due to the potential for verification bias. Further issues included a lack of information about the timing between the PFTs and HRCT scans and about whether proper blinding was performed.

Of the remaining 45 studies, more than half aimed to validate specific HRCT scoring systems using PFTs as reference standards. Thus, despite HRCT being the gold-standard for assessing SSc-ILD presence, no consensus exists on the best way in which to score SSc-ILD severity. This explains the large variety of different HRCT scoring systems used in SSc-ILD and renders it challenging to summarize results. It is also worth noting that one study correlated PFTs with lung biopsy results,[237] while only six reported longitudinal correlations between PFTs and HRCT scores.[133, 225, 241, 250, 255, 274] While no formal synthesizing analysis was performed on these studies, their results do not overwhelmingly appear to support the superiority of any particular PFT (Appendix F).

4.7 Discussion

We performed a comprehensive systematic review of the literature to explore the use of PFTs as surrogate markers for SSc-ILD onset and progression. This review confirmed the predominant

selection of FVC as a primary endpoint in recent longitudinal SSc-ILD studies, despite a paucity of evidence from validation studies that FVC has better performance characteristics than other PFT measures.

Our results show that both FVC and DLCO were historically used to monitor SSc-ILD progression. However, given that many authors did not report why they chose a given PFT measure as an outcome, it is difficult to fully understand the recent shift in preference to FVC. A possible explanation is the widespread use of FVC as a therapeutic endpoint in idiopathic pulmonary fibrosis (IPF),[280, 281] a comparable disease especially given the lack of connective tissue disease (CTD)-ILD trials. However, a lack of consensus still surrounds FVC as the best outcome measure for this disease.[282, 283] Furthermore, IPF is characterized by usual interstitial pneumonia, a pathologically different subtype of ILD than the non-specific interstitial pneumonia pattern commonly observed in SSc-ILD.[284] We suggest two additional reasons, supported by the results of this systematic review, which may explain FVC's current popularity in SSc.

First, given that DLCO can be confounded by the presence of pulmonary arterial hypertension, FVC is believed to be more specific than DLCO to ILD.[45, 285] However, evidence suggests that DLCO may be more sensitive than FVC.[73, 238, 279] While the appropriate trade-off between sensitivity and specificity is debatable, it is worth noting that a lower specificity due to confounding by pulmonary arterial hypertension can be corrected for in analyses by using simple epidemiological techniques such as restriction and adjustment.

In addition to being less specific to ILD, DLCO is also regarded as being inherently more variable than FVC given its dependence on the measurement of both carbon monoxide transfer (KCO) and alveolar volume. Indeed, DLCO is often not chosen as a measure of SSc-ILD progression because it is considered to not be as reproducible as FVC. However, it is worth noting that variability in FVC

can be equally concerning. In fact, a recent study calculated FVC's minimal clinically important difference in SSc-ILD to be well within the range of measurement error at the individual level.[47]

Next, we suggest that the pivotal Scleroderma Lung Study 1 (SLS I) also played an important role in the sudden decrease in use of DLCO in favour of FVC. SLS I was the first double-blind, randomized, placebo-controlled trial to study the effect of oral cyclophosphamide on SSc-ILD. The baseline-adjusted mean absolute 12-month difference in FVC, the study's primary endpoint, was 2.53% (0.28% – 4.79%), favouring cyclophosphamide. This improvement persisted at 24 months, at which point the mean absolute difference in FVC was 1.95% (1.2% – 2.6%). Secondary outcomes included DLCO and TLC with adjusted mean 12-month differences of -1.04% (p-value = 0.43) and 4.09% (0.49% – 7.65%), respectively.[31] Despite the slight variability in these results and the modest treatment effects observed, FVC was subsequently described as having been validated as an endpoint in randomized controlled trials.[286] In fact, the quality of FVC as a measure for SSc-ILD is often erroneously assessed by its observed treatment effect, rather than by proper validation techniques against a gold-standard. Interestingly, after the 2006 publication of SLS I, our analyses revealed a steep increase in the use of FVC, as well as a steep decrease in the use of DLCO following years of relative stability (Figure 4-2b).

It is likely that FVC's specificity for ILD and perceived validity led to its current acceptance as the best surrogate marker for ILD within the SSc community. However, we suggest that this preference is not yet fully warranted, given the results of studies which validated PFT measures against HRCT-assessed SSc-ILD severity. In fact, these five studies revealed that FVC does not have conclusively better performance characteristics than other PFT measures. Moreover, two of these studies suggested that DLCO best captured SSc-ILD severity.[238, 279]

Also noteworthy was total lung capacity (TLC)'s performance which overlapped considerably with those of FVC and DLCO (Figures 4-4 and 4-5). Despite these results and the fact that the American Thoracic Society defines restrictive lung diseases, such as ILD, by a reduction in TLC in the presence of a preserved or elevated FEV_1/FVC (forced expiratory volume in the 1st second of forced exhalation to FVC ratio),[42] TLC has rarely been used as an outcome in SSc-ILD (Figure 4-2) and was never used as a main outcome (Figure 4-3).

Ultimately, the best surrogate marker for SSc-ILD onset and progression may be a composite outcome consisting of a combination of two or more PFT measures. In fact, composite outcomes were used in several studies included in this systematic review. The most common ones included decreases in FVC or DLCO ≥10%,[107, 115, 127, 155, 203] a decline ≥10% in FVC or 15% in DLCO,[108, 114, 143, 144, 154, 161, 169, 170, 200, 201], and decreases in FVC or DLCO ≥15%.[82, 83, 136, 177, 185, 193] Few studies used a FVC-TLC composite endpoint,[74, 153, 165] and only one study jointly considered FVC, TLC and DLCO.[118] Nevertheless, the validity of such endpoints remains to be fully evaluated, as only cross-sectional validation studies have been performed so far. These studies found composite outcomes consisting of FVC and DLCO, and of FVC, DLCO and TLC to have high sensitivity, but low specificity (Figure 4-5).

This systematic review highlights the need to continue to focus efforts on identifying the best sole or composite PFT surrogate marker for SSc-ILD. Presently, the widespread use of FVC has translated into the identification of few predictors of SSc-ILD progression and into attenuated treatment effects in SSc-ILD randomized controlled trials,[31-33] suggesting that FVC may weakly reflect the extent of SSc-ILD. In fact, it is possible that FVC may also be affected by the cutaneous involvement of the chest wall in SSc.[287] Following a recent consensus exercise, the OMERACT (Outcome Measures in Rheumatology) CTD-ILD working group agreed to and proposed a ≥10% decline in FVC or a ≥5%

to <10% decline in FVC in the presence of a ≥15% decline in DLCO as a clinically meaningful outcome in clinical trials of CTD-ILD. However, they also recognized that these measures remain to be validated.[288] Indeed, thorough PFT validation studies are needed to better assess the validity of different PFT measures. In particular, longitudinal studies should explore the association between concurrent changes in PFTs and clinical reference standards. Future studies should build up recent research aiming to develop composite measures of SSc-ILD severity including, for instance, measures of lung physiology, disease manifestations and patient-reported outcomes.[220] This will ultimately strengthen the quality and results of clinical trials and epidemiological studies and assist clinical decision-making.

Our systematic review is not without certain limitations. First and foremost, we did not include studies which validated PFT measures against mortality given the already large scope of this review. Such studies are also integral in identifying the best PFT surrogate marker for SSc-ILD progression. Despite this exclusion, our results are complemented by those of a 2010 systematic review which found that DLCO was more consistently associated with mortality in SSc-ILD than FVC.[46] Since the publication of this review, further evidence has mounted favouring the utility of DLCO. Indeed, one recent study found a decline in DLCO and KCO to be most predictive of adverse outcomes, including death.[289] Likewise, a second recent study suggested that one-year trends in an FVC and DLCO composite endpoint and two-year trends in measures of gas transfer were most predictive of mortality.[290]

Additional limitations include the decision to restrict validation studies to studies with classical measures of validity and to not account for studies which reported regression coefficients and p-values. While this translated into a loss of information, it permitted a better focus on higher quality validation data and allowed for standardization across included studies. Finally, while there is potential

for publication bias, we believe that the nature of our research question, in addition to the large number of included studies and the extensive search of multiple databases, alleviates this concern.

In summary, available evidence does not overwhelmingly favour one PFT measure as the best surrogate marker for SSc-ILD, rendering it challenging to support the current preference for FVC. Indeed, the perceived superiority of FVC is not reflected in rigorous PFT validation studies. While FVC has the potential to be a viable surrogate marker for SSc-ILD, it would be ill-advised at this stage to discount other potentially interesting PFT measures, such as DLCO and TLC.

4.8 Acknowledgements

The authors would like to thank the McGill University epidemiology liaison librarian Genevieve Gore for her assistance in designing the detailed search strategy.

4.9 Figures

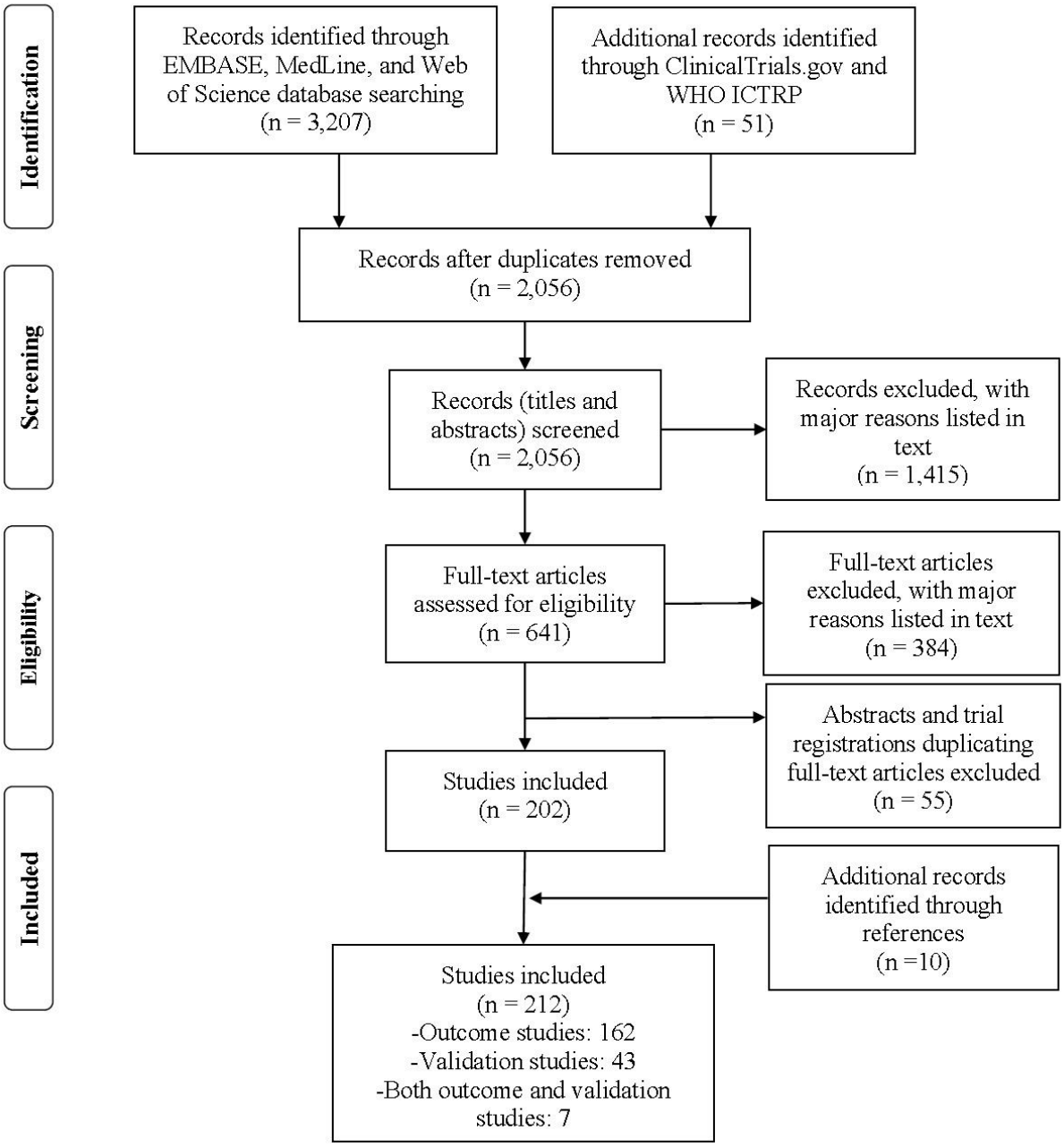
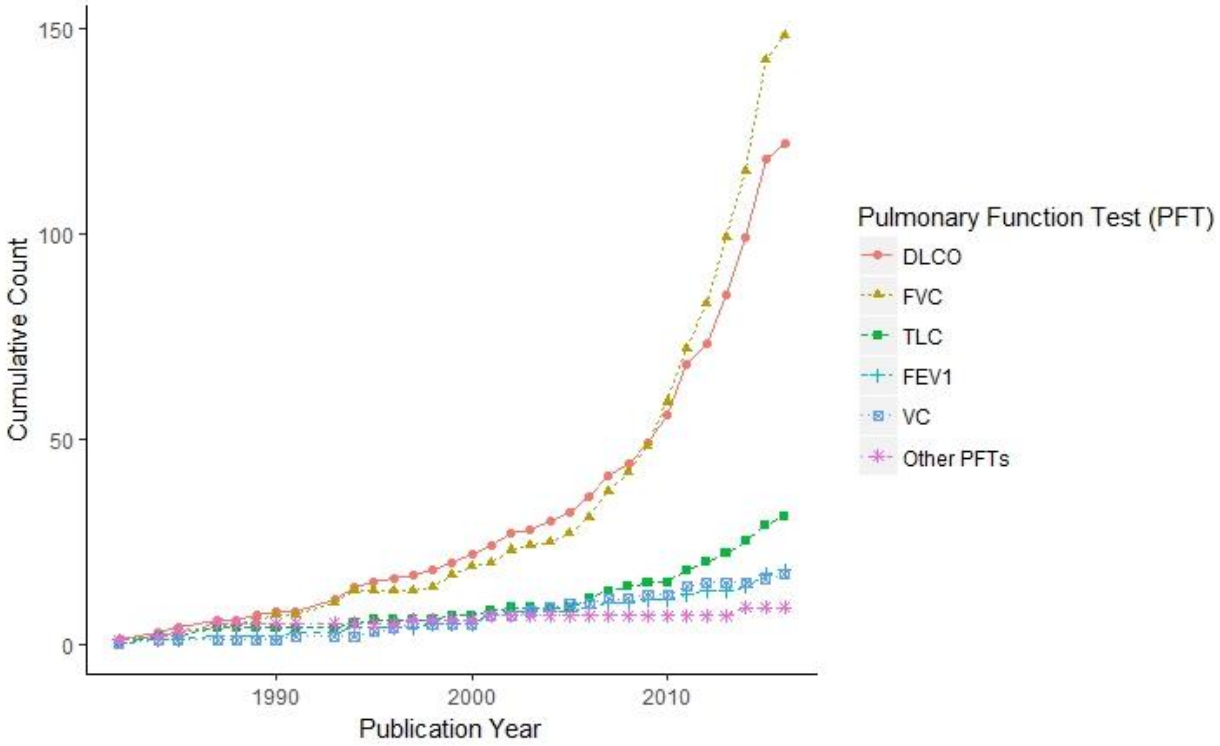


Figure 4-1: Adapted PRISMA Flow Diagram.[72]
Abbreviations: ICTRP = International Clinical Trials Registry Platform; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses; WHO = World Health Organization

a)


150
100
50
0
Cumulative Count
1990
2000
2010
Publication Year
Pulmonary Function Test (PFT)
DLCO
FVC
TLC
FEV1
VC
Other PFTs

b)

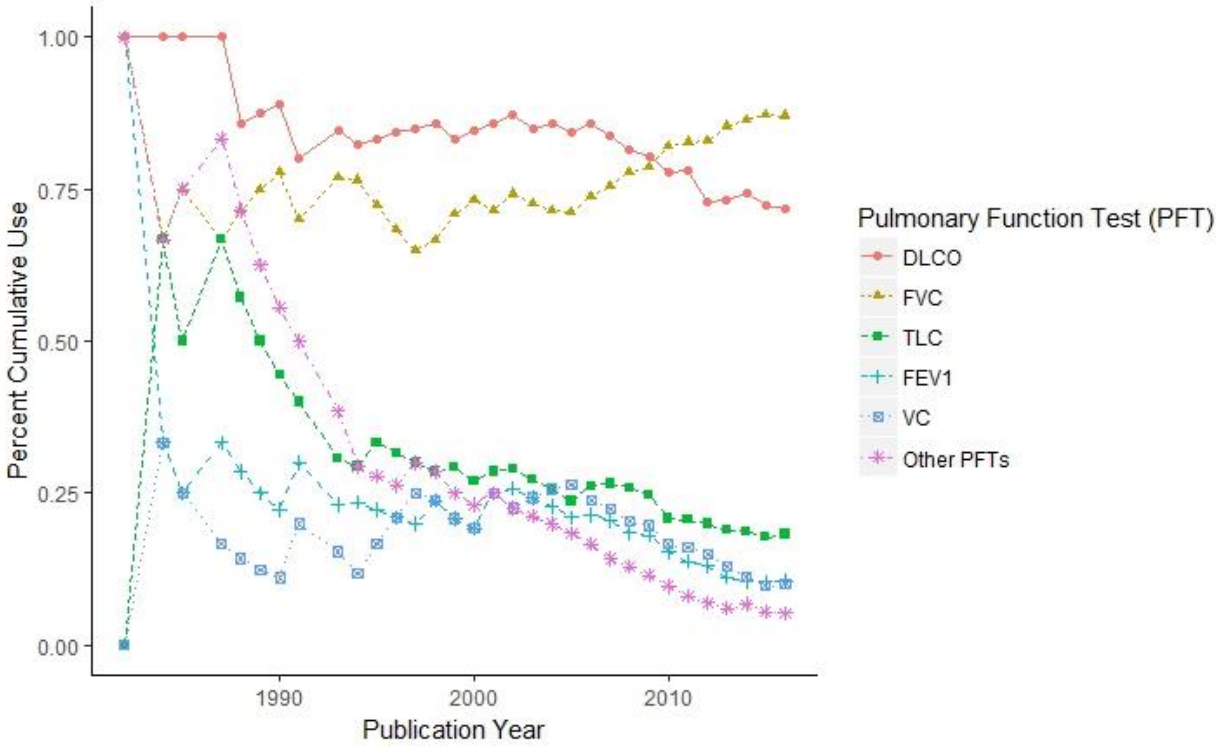

1.00
0.75
0.50
0.25
0.00
Percent Cumulative Use
1990
2000
2010
Publication Year
Pulmonary Function Test (PFT)
DLCO
FVC
TLC
FEV1
VC
Other PFTs

Figure 4-2: a) Cumulative Use and b) Percent Cumulative Use of the Different Pulmonary Function Test (PFT) Measures as Longitudinal Outcomes for Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Progression. *For each year, percent cumulative use was calculated by dividing the cumulative use of each PFT measure by the cumulative number of published articles. No distinctions were made between absolute and % predicted PFT values. All variations of DLCO were grouped together into one all-encompassing DLCO measure. Other PFTs included measures of forced expiratory flow over mid-half of FVC, FEV_1/FVC, FEV_1/VC, functional residual capacity, and residual volume.*
Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FVC = Forced Vital Capacity; FEV_1 = Forced Expiratory Volume in the 1st Second of Forced Exhalation; TLC = Total Lung Capacity; VC = Vital Capacity

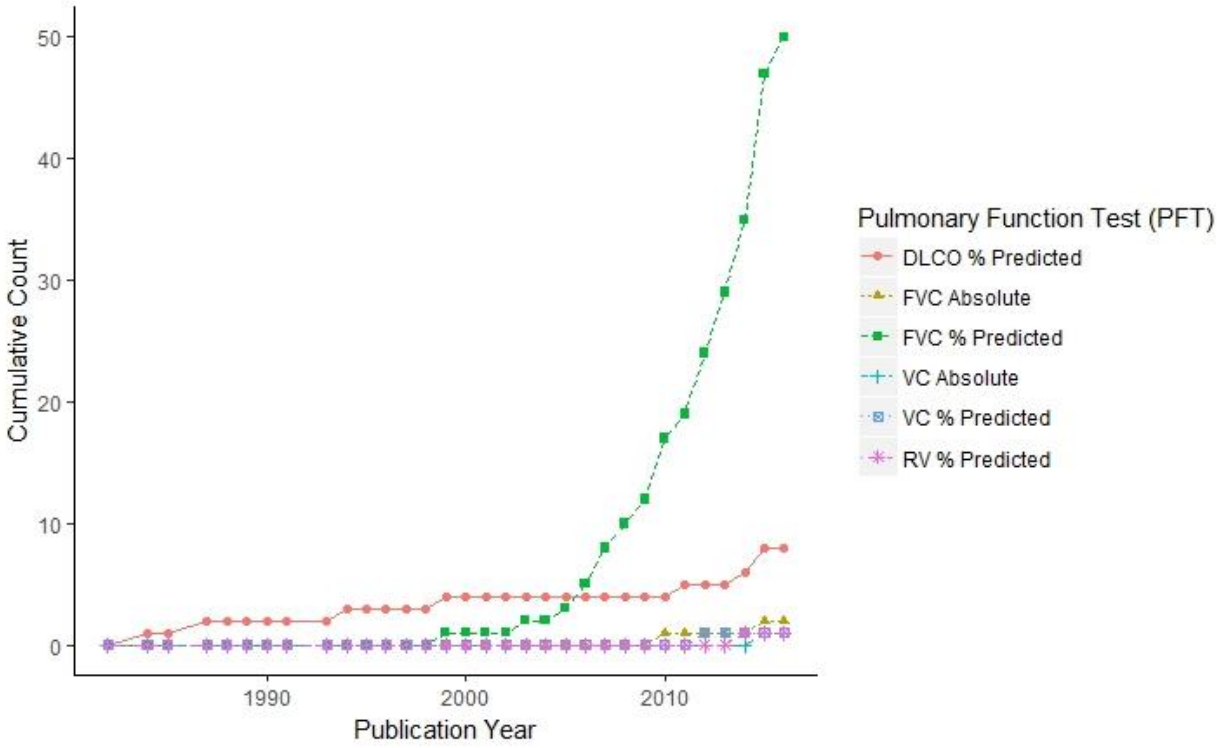
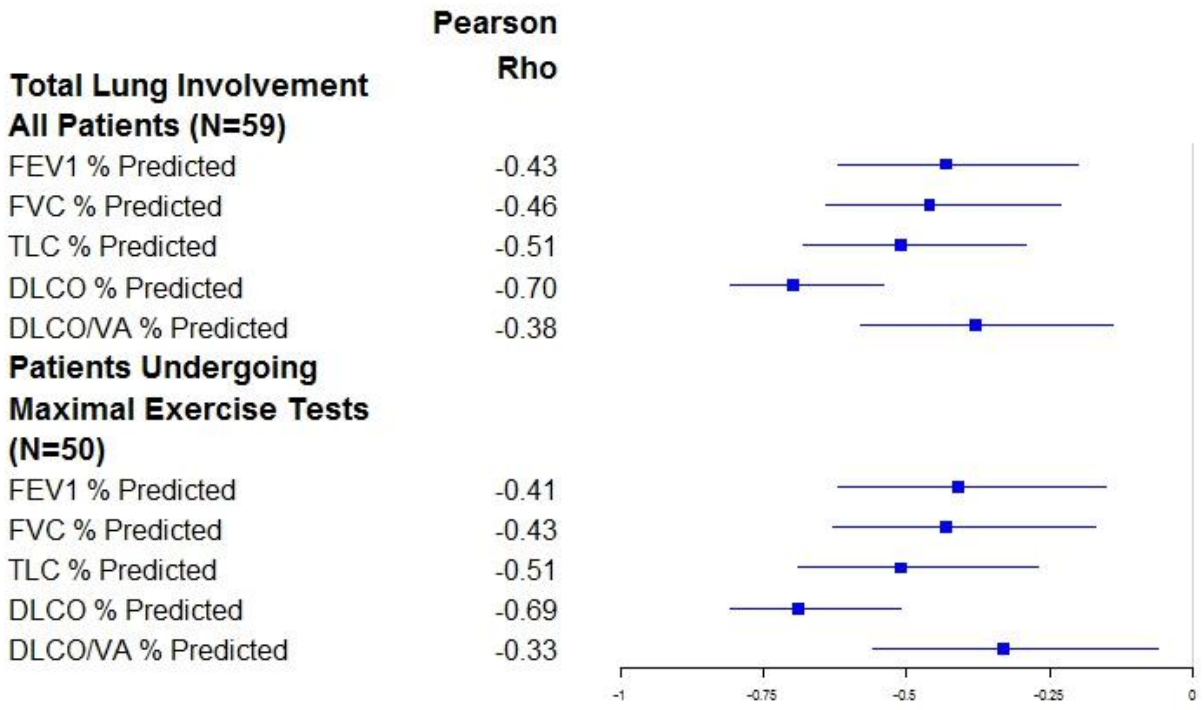
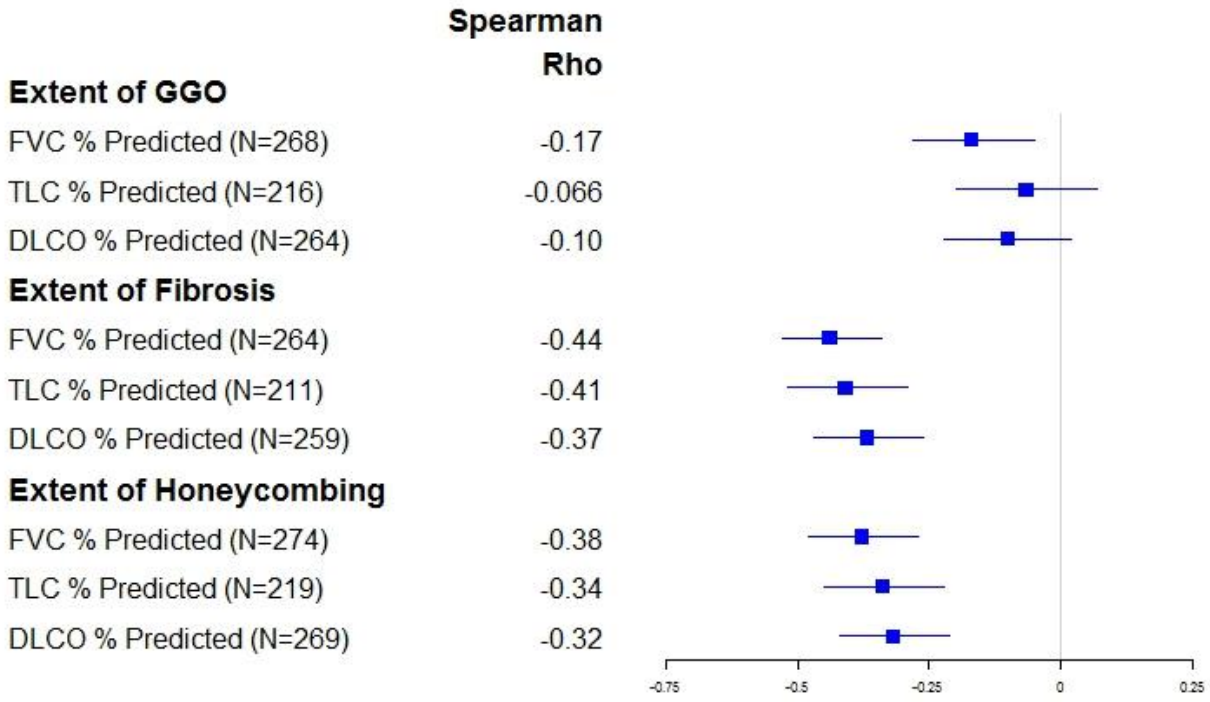


Figure 4-3: Cumulative Use of the Different Primary Pulmonary Function Test (PFT) Outcomes in Longitudinal Studies of Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Progression. *Four studies which used FVC but which did not make a distinction between absolute and % predicted value were excluded from this figure.*
Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FVC = Forced Vital Capacity; RV = Residual Volume; VC = Vital Capacity

a)


Pearson Rho
Total Lung Involvement All Patients (N=59)
FEV1 % Predicted -0.43
FVC % Predicted -0.46
TLC % Predicted -0.51
DLCO % Predicted -0.70
DLCO/VA % Predicted -0.38
Patients Undergoing Maximal Exercise Tests (N=50)
FEV1 % Predicted -0.41
FVC % Predicted -0.43
TLC % Predicted -0.51
DLCO % Predicted -0.69
DLCO/VA % Predicted -0.33
-1
-0.75
-0.5
-0.25
0

b)


Spearman Rho
Extent of GGO
FVC % Predicted (N=268) -0.17
TLC % Predicted (N=216) -0.066
DLCO % Predicted (N=264) -0.10
Extent of Fibrosis
FVC % Predicted (N=264) -0.44
TLC % Predicted (N=211) -0.41
DLCO % Predicted (N=259) -0.37
Extent of Honeycombing
FVC % Predicted (N=274) -0.38
TLC % Predicted (N=219) -0.34
DLCO % Predicted (N=269) -0.32
-0.75
-0.5
-0.25
0
0.25

c)

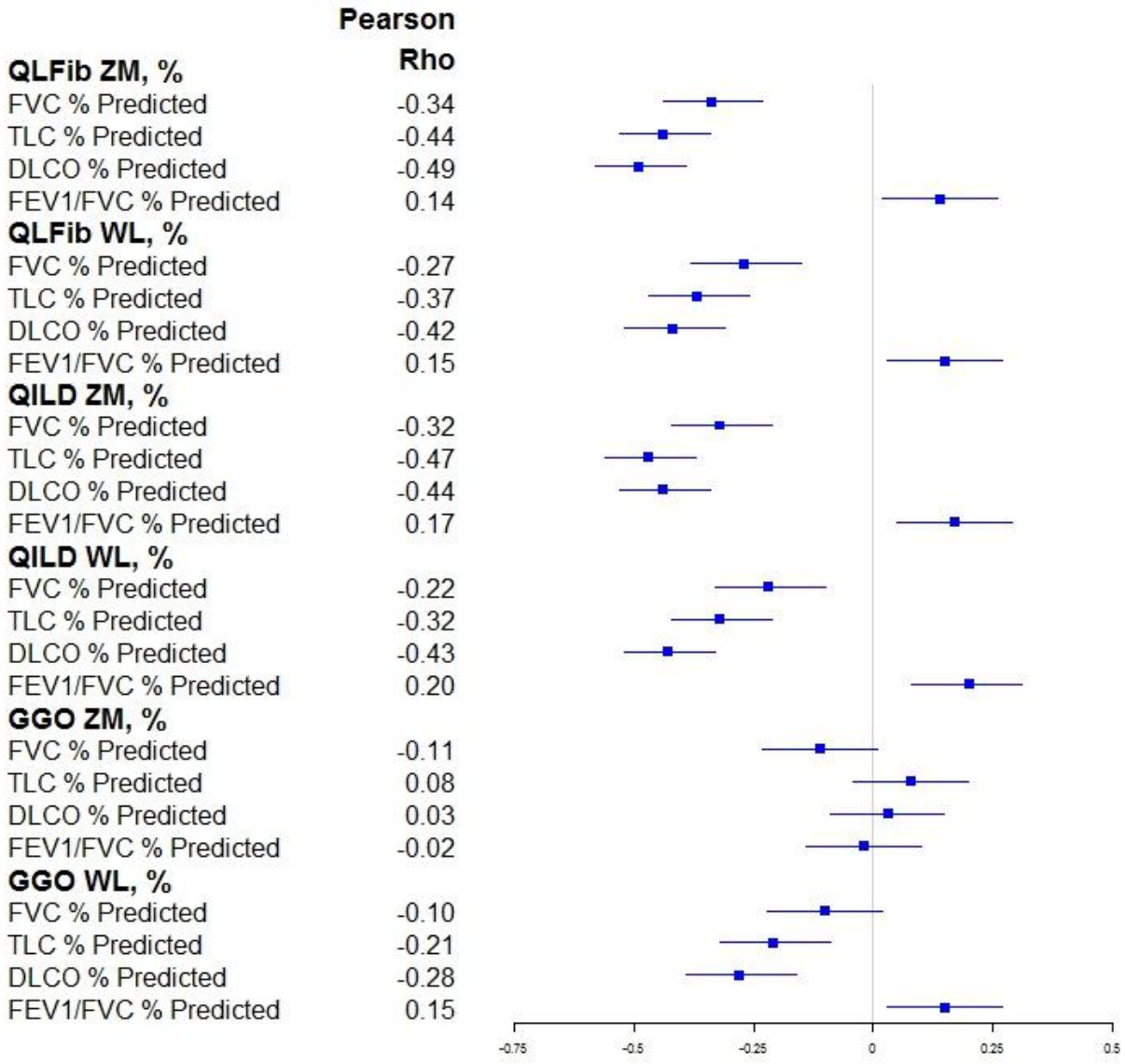
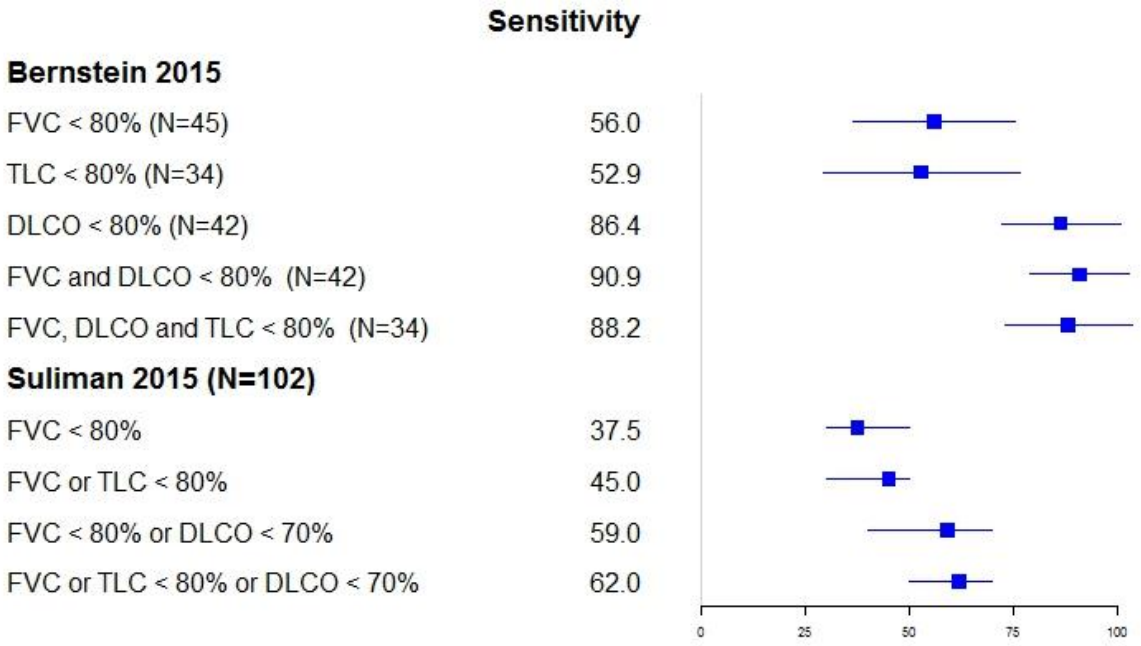


Figure 4-4: Forest Plots of the Correlation Between High-Resolution Computed Tomography (HRCT) Scores and Pulmonary Function Test (PFT) Values. *The 95% confidence intervals were calculated using the Fisher transformation. a) Wells 1997 Study: Total abnormal lung involvement on HRCT was measured to the nearest 5%.*[238] *b) Zamora 2013 Study: The scoring method for this study was not reported.*[268] *c) Tashkin 2016 Study: HRCT scans were scored using quantitative imaging analyses and DLCO was corrected for haemoglobin. While the number of subjects included in the analyses was not reported, we assumed that all 261 SSc subjects having undergone HRCT testing were included n the correlation analyses.*[279]
Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; DLCO/VA = Diffusing Capacity for Carbon Monoxide Corrected for Alveolar Volume; FEV_1 = Forced Expiratory Volume in the 1st Second of Forced Exhalation; FVC = Forced Vital Capacity; GGO = Ground-Glass Opacity; QILD = Quantitative Interstitial

Lung Disease; QLFib = Quantitative Extent of Lung Fibrosis; TLC = Total Lung Capacity; WL = Whole Lung; ZM = Zone of Maximal Involvement

a)



b)

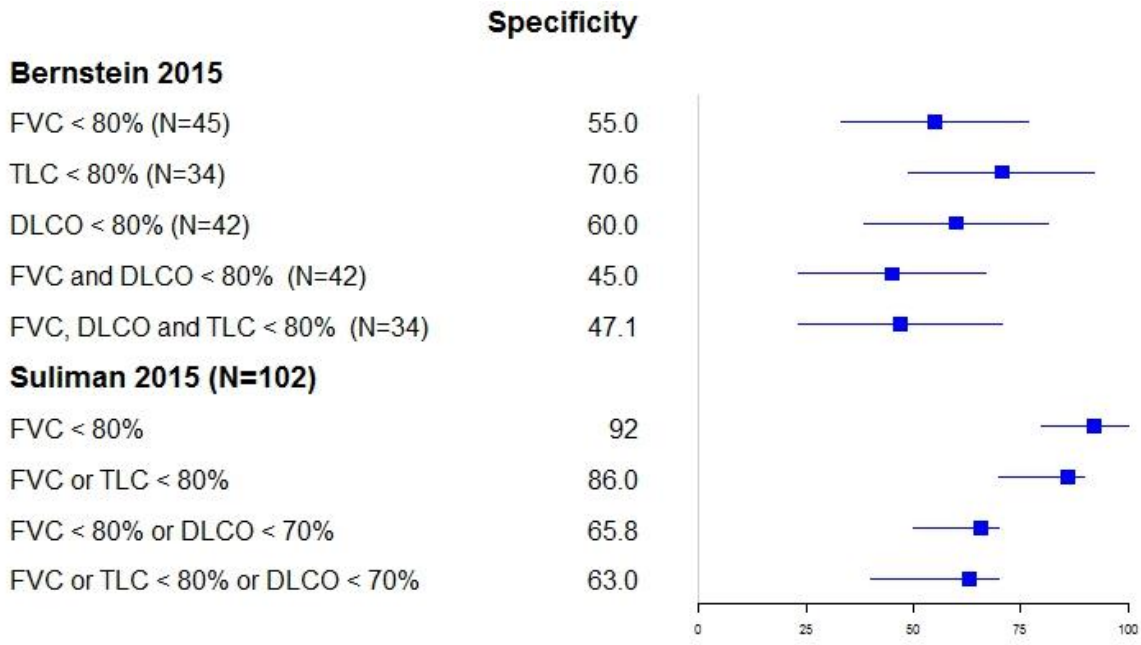


Figure 4-5: Forest Plot of the a) Sensitivities and b) Specificities of Different Pulmonary Function Test (PFT) Screening Algorithms for Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Presence.[43, 271] *95% confidence intervals for the Bernstein 2015 study were calculated using reported measures of sensitivity, specificity, positive predictive value and negative predictive value.*

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FVC = Forced Vital Capacity; TLC = Total Lung Capacity

Chapter 5: Evaluation of Hidden Markov Models as a Method of Ascertaining Systemic Sclerosis-Associated Interstitial Lung Disease Onset using Pulmonary Function Test Measures

5.1 Preface to Manuscript 2

Chapter 4 confirmed via a systematic review of the literature that forced vital capacity (FVC) is currently the pulmonary function test (PFT) surrogate marker most often used for systemic sclerosis-associated interstitial lung disease (SSc-ILD), despite a lack of evidence showing superior diagnostic test properties to other PFT measures. In fact, diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC) were often shown either to be comparable or to outperform FVC in the few existing validation studies on this topic.

However, when assessing SSc-ILD presence, these validation studies used crude definitions of disease presence in the form of hard cut-offs in PFT measurements (*e.g.* FVC <80% predicted, etc.) Equally common in the SSc-ILD literature is the use of fixed changes in PFT values, such as a ≥10% decrease between consecutive PFT measurements to indicate SSc-ILD presence. These definitions ignore patients' entire history of PFT measurements and simplify the uncertainty surrounding SSc-ILD status to a binary outcome. Furthermore, they do not account for the inherent variability associated with PFT measurements. We aimed to propose a better alternative for how PFT measures could be used to ascertain SSc-ILD onset and presence.

In this next study, we used both simulated and real FVC data to assess the ability of hidden Markov models (HMMs) to identify cases of SSc-ILD compared to the traditional <80% cut-off and 10% decline in FVC definitions. HMMs are a potentially more sophisticated alternative, as they can use the full history of PFT measurements to model the probability of SSc-ILD occurrence while

simultaneously correcting for measurement error. Measures of diagnostic test accuracy were calculated for all three algorithms using a combination of high-resolution computed tomography, chest X-ray and lung auscultation results as a reference standard, as described in Section 3.1.3.

The resulting manuscript, entitled "Ascertaining Disease Onset Using Surrogate Markers and Hidden Markov Models: An Application to Systemic Sclerosis-Associated Interstitial Lung Disease", is prepared for submission to the Journal of Clinical Epidemiology where it is currently under review. To reach a broader audience of clinical epidemiologists, the article is written to reflect how HMMs can use surrogate markers in general to detect disease onset using SSc-ILD as a motivating example.

5.2 Title Page

Title: Ascertaining Disease Onset Using Surrogate Markers and Hidden Markov Models: An Application to Systemic Sclerosis-Associated Interstitial Lung Disease

Authors: Melissa Caron[a, b]; Kevin Schwartzman[a, b]; Marie Hudson[c]; Murray Baron[c]; Russell J Steele[d]; and the Canadian Scleroderma Research Group

Affiliations:

[a] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, QC, H3A 1A2, Canada

[b] Respiratory Epidemiology and Clinical Research Unit (RECRU), Montreal Chest Institute, McGill University Health Centre, 5252 de Maisonneuve West, Montreal, QC, H4A 3S5, Canada

[c] Division of Rheumatology, Sir Mortimer B Davis Jewish General Hospital, 3755 chemin de la Côte Sainte-Catherine, Montreal, QC, H3T 1E2, Canada

[d] Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, QC, H3A 2K6, Canada

Corresponding Author: Melissa Caron, Respiratory Epidemiology and Clinical Research Unit (RECRU), 5252 de Maisonneuve West, Montreal, QC, Canada, H4A 3S5. E-mail: melissa.caron@mail.mcgill.ca

Declarations of Interest: None.

Funding: Melissa Caron would like to acknowledge grants from the Fonds de Recherche du Québec (Santé PhD Studentship) and from the Canadian Institutes of Health Research (GSD – 146268). These funding sources were not involved in the conduct of the research or preparation of the article.

Highlights/What is new?

- A hidden Markov model can be used to detect disease onset with surrogate markers.
- It uses full marker measurement history while also correcting for measurement error.
- It has a lower error rate than the use of cut-offs or fixed changes in marker level.
- It can be used as both a screening tool or statistical model for disease presence.

5.3 Abstract

Objective. Disease onset is often ascertained using surrogate markers and basic decision rules, such as the attainment of a specific marker level. Hidden Markov models (HMMs) are a potentially more sophisticated alternative, as they can use the full history of surrogate marker measurements to model the probability of disease while correcting for measurement error. We aimed to assess the statistical performance of HMMs (using predicted probability ≥0.50 to indicate disease presence) in the context of systemic sclerosis-associated interstitial lung disease onset and compare it to the traditional <80% cut-off and 10% decline in forced vital capacity algorithms.

Study Design and Setting. Measures of diagnostic test accuracy were calculated using both simulated data and real data. A combination of high-resolution computed tomography, chest X-ray and lung auscultation results was used as a gold-standard for the ascertainment of interstitial lung disease.

Results. The results showed that the HMM had the highest specificity and lowest error rate suggesting it can be an alternative to determining disease status. The HMM also achieved higher sensitivity than the cut-off and change in marker level algorithms when the predicted probability threshold was lowered.

Conclusion. HMMs are useful screening tools for disease onset when using surrogate markers.

Keywords: biomarkers; interstitial lung disease; hidden Markov models; surrogate markers; systemic sclerosis

Running Title: Ascertaining Disease Onset Using Surrogate Markers and Hidden Markov Models

5.4 Introduction

Surrogate markers and biomarkers are commonly used to ascertain the onset of chronic diseases both clinically and in epidemiologic studies.[291-294] Often this involves rudimentary decision rules, such as the use of hard cut-offs or of pre-specified changes in the marker level. This framework ignores patients' entire history of marker measurements and simplifies the uncertainty surrounding disease status to a binary outcome. These problems are further exacerbated by the fact that disease markers are often subject to measurement error due to imperfect testing instruments and/or random variation. A hidden Markov model (HMM) addresses these issues by using the full marker history to model the probability of disease occurrence while simultaneously correcting for measurement error.

Though HMMs are often discussed in the statistical literature and occasionally employed in disease surveillance,[63, 295, 296] they are rarely used in epidemiologic studies and have never been validated using clinical data. Thus, the overarching aim of this study was to assess the statistical performance and applicability of HMMs in the context of systemic sclerosis-associated interstitial lung disease (SSc-ILD) onset.

Systemic sclerosis (SSc) is a rare autoimmune disease involving an excess production of connective tissue.[1, 13] SSc patients have disfiguring skin involvement in addition to pulmonary, gastrointestinal, renal and cardiac complications.[2] Interstitial lung disease (ILD) is the leading cause of morbidity and mortality in SSc.[28] SSc-ILD onset is routinely monitored using pulmonary function test (PFT) measures, in particular forced vital capacity (FVC) which corresponds to the total volume of air released during forced exhalation following a maximal inspiration.[10] This measure is often expressed as the % of a predicted value, which is based on data from healthy subjects of similar age, sex and height.[42]

SSc-ILD onset is often suspected based on either an FVC cut-off value of <80% predicted or a 10% decline between two consecutive FVC measurements,[45, 200, 201] despite yearly random variations in FVC of 10-15% being well-documented.[48, 297] We hypothesized that HMMs may be better suited to monitor SSc-ILD onset given the natural fluctuations observed in FVC measurements.

We first present results from a simulation study comparing the performance metrics of the HMM to those of the 80% cut-off and 10% absolute decline in FVC algorithms. We then provide a practical example using data from the Canadian Scleroderma Research Group (CSRG), one of the largest prospective cohorts of SSc patients in the world.

5.5 Materials and Methods

Hidden Markov Models

HMMs use all available disease marker data to estimate the instantaneous risk (or transition intensity) of progressing from one disease state to another, while correcting for measurement error in the marker. This correction is possible as HMMs are hierarchical in nature: they are comprised of both a classical measurement error model and a traditional Markov model.[64, 66]

The classical measurement error model relates the imperfectly measured observed marker level to the true marker level: $Y_{it}^{*} = Y_{it} + \varepsilon_{it}$, where Y^{*}_{it} is the observed marker measurement for patient i at time t and Y_{it} is the true marker value. The random variable Y is associated with the disease states through an underlying probability distribution. The transitions between states and corresponding transition intensities are subsequently modeled by the Markov model. HMMs are especially powerful as they can use the estimated transition intensities to then calculate the probability of state occupancy at any time for any study subject using their personal history of marker measurements.[64, 66] A more detailed explanation of HMMs is available in Appendix H.

A simple two-state HMM with separate states for disease absence and presence can be used to study disease onset. It can be viewed as a sophisticated version of a running average whereby a subject's newly measured marker value is compared to their average marker level to determine whether the trend is indicative of disease occurrence or not.

In our study, two-state HMMs were used to calculate patients' probability of ILD occurrence given their prior FVC measurements. The models were programmed via the `msm` package in `R`.[65, 298] Since fibrosis of the lungs is an irreversible process,[11] only transitions from ILD absence to presence were permitted. To account for between-subject heterogeneity in FVC measurements, the HMMs adjusted for baseline FVC. Consequently, every subject's first visit was excluded. Appendix I describes how the HMMs were initialized.

Simulation Study

The simulation study was modeled after SSc patients in the CSRG registry at risk of developing ILD during follow-up. We simulated 1,150 SSc patients with five annual visits. All subjects were ILD-free at baseline. ILD onset could occur during visits two through five and was simulated using a Markov process in the `msm` package. ILD status was thus known for each patient-visit. A complete description of how FVC values were subsequently simulated is available in Appendix J. A comparison of the simulated FVC values with those recorded in CSRG subjects is depicted in Figure 5-1.

We fit a discrete-time HMM to the simulated data using visit number as the timescale of interest. Sample code is provided in Appendix K. The simulations were performed 50 times and measures of validity were averaged.

Application to the Canadian Scleroderma Research Group Data

Only CSRG patients with at least three visits and without ILD at baseline were included in these analyses since a minimum of three FVC measurements was required to run an HMM adjusting for baseline FVC. ILD status was assessed using a combination gold-standard which has previously been validated in both the CSRG registry and externally.[29, 58] While high-resolution computed tomography (HRCT) is the standard for determining ILD presence, HRCT scans are not routinely ordered for CSRG patients. In the absence of an HRCT scan, the combination gold-standard suggests that ILD is present if there is evidence of fibrosis on chest X-rays or if crackles are heard during lung auscultation.[29] Both chest X-rays and lung auscultations are regularly recorded. If no such results were available, the visit was assumed to be ILD-negative. Once a patient was diagnosed by the gold-standard as having developed ILD, they could not subsequently revert to being ILD-free.

Missing FVC values (28.6% missingness) were imputed using multiple imputation by chained equations implemented via the `mice` package in `R`.[299] The missing FVC data were imputed in long format using a two-level normal model with homogeneous within-patient variances. Predictors included HRCT results, type of SSc (limited or diffuse), current smoking status, presence of pulmonary hypertension, and presence of auto-antibodies predictive of ILD (anti-topoisomerase, anti-centromere, and anti-RNA polymerase III).[4, 27, 193, 223, 300-302] Twenty-five imputed datasets were generated.

The calendar dates on which FVC was measured were used to fit a continuous-time HMM. Since PFTs are performed annually, missing dates were imputed by adding or removing a year from the previous or next available test date, respectively. Since data on patient mortality were available, a death state was also included in the HMM.

A running average approach to ascertaining SSc-ILD onset was also analyzed as a less complicated version of an HMM. This approach calculated the difference between a subject's current FVC value

and their running average which was then compared to a pre-specified threshold to assign ILD status. Various thresholds and their performances were assessed.

Ethics approval for the collection of CSRG data was obtained at each study site and patients provided informed written consent to participate in the registry. Since this study involved the secondary use of deidentified data, no consent was required.

Measures of Validity

For both studies, classical measures of diagnostic test accuracy (DTA) and their corresponding standard errors were computed for the HMM, as well as for the <80% FVC cut-off and 10% absolute decline in FVC algorithms. These values were averaged across all 25 imputed datasets. Given that the HMM generates a probability of ILD rather than a definitive binary decision, a predicted probability ≥0.50 was chosen to define ILD presence. It is important to note in the case of the HMM that at no time were the same data used to both fit and evaluate the model. The HMM was fit using only FVC marker measurements, while ILD outcomes were only used to evaluate it.

A detailed description of how the measures of validity were calculated is available in Appendix L. In addition to classical measures of validity, we also computed incident measures by only using subjects' first ILD-positive result. For instance, incident sensitivity refers to the ability to correctly identify incident ILD. Alternatively, prevalent sensitivity is the proportion of patients with ILD whose disease was missed at onset but was subsequently identified. Cases correctly detected at onset were therefore excluded from the numerator of this proportion. Finally, the overall error rate (ER) corresponded to the sum of false negatives and false positives divided by the total number of comparisons.

5.6 Results

Simulation Study

The percentage of simulated patients having developed ILD by the end of the five-year study period was 22.7% (standard deviation (S.D.) = 1.3), while the mean prevalence of ILD among all patient-visits was 11.7% (S.D.=0.7). This closely resembled the CSRG registry where 21.4% of patients developed ILD by their fifth visit and 12.6% of these visits were ILD-positive.

The performance metrics of the three algorithms are outlined in Table 5-1. The HMM had the lowest ER (overall ER: 15.9%; incident ER: 14.2%); while the 10% decline algorithm had the highest (overall ER: 28.4%; incident ER: 28.1%). The 10% decline algorithm had both the highest overall and incident sensitivities (65.2% and 61.3%, respectively) and identified the greatest percentage of ILD cases (72.0% vs. 35.3% and 62.1% for the 80% cut-off algorithm and HMM, respectively). The HMM had the highest specificity (88.7%), positive predictive value (PPV) (overall PPV: 46.5%; incident PPV: 37.7%) and positive likelihood ratio (+LR). Its negative likelihood ratio (-LR) was comparable to that of the 10% decline algorithm.

The receiver operating characteristic (ROC) curve in Figure 5-2 depicts the effect of varying the HMM's predicted probability threshold for SSc-ILD. It illustrates that the HMM can surpass the 10% decline's high sensitivity as lower probability thresholds are chosen.

Table 5-2 depicts the cumulative percentage of SSc-ILD cases caught in the year of and following ILD onset. As expected, the 10% decline algorithm and HMM performed better than the 80% cut-off algorithm given their higher sensitivities. However, the incident sensitivity of the HMM improved markedly from 13.0% to 61.9% when ILD occurred at visits 2 and 3, respectively. Indeed, as FVC measurements accrue, more data are available for the HMM to detect ILD, unlike the 80% cut-off and 10% decline algorithms which can only account for the one or two most recent FVC measurements, respectively.

Though the 10% decline algorithm and HMM are comparable in their capacity to identify cases of SSc-ILD, the HMM is accompanied by a much lower false positive rate. Under the scenario where ILD onset occurred at visit 2, Figure 5-3 demonstrates that the HMM had few instances of missed ILD cases and a low cumulative false positive rate. By visit 5, the HMM had incorrectly classified on average 18.3% of non-ILD visits, compared to 39.7% for the 10% decline algorithm.

Kappa statistics were calculated to determine the agreement between the three methods of ILD ascertainment (Appendix M). The results found the agreement between all methods to be poor even when restricting to ILD-negative and positive visits, indicating very little overlap in their behavior.

Application to the Canadian Scleroderma Research Group Data

This study included 3,640 paired FVC measures and combination gold-standard results collected from 779 CSRG subjects. The performance metrics of the HMM and both algorithms are presented in Table 5-3. Overall, the results were not as favorable as in the simulation study.

The 10% decline algorithm again demonstrated the highest overall sensitivity, but only at 47.8%, while the 80% cut-off algorithm had the greatest incident sensitivity (39.9%). The HMM exhibited the greatest specificity (80.2%), as well as the highest incident PPV (21.0%). Its overall PPV was equivalent to that of the 80% cut-off algorithm (27.6% vs. 28.6%). The 80% cut-off algorithm also exhibited the highest +LR and lowest -LR. However, the HMM had the lowest overall (29.1%) and incident ERs (22.9%). The results obtained when only HRCT scans were used as the gold-standard are available in Appendix N. Kappa statistics are available in Appendix M; again, there was very little agreement between the algorithms.

The running average approach performed similarly to the HMM when a threshold of 12 was chosen (Table 5-3). This was unsurprising as the HMM estimated the means of the FVC probability distributions in the absence and presence of ILD to be 12% apart. As with the HMM, increasing the running average threshold for SSc-ILD onset resulted in a lower incident sensitivity and greater specificity (Appendix O).

5.7 Discussion

Our analyses have demonstrated that HMMs have potential as an analytic tool to assess disease onset using surrogate markers or biomarkers. They can act as a more refined alternative to hard cut-offs and pre-determined changes in markers by using complete longitudinal data and correcting for potential measurement error.

The <80% predicted and 10% decline in FVC definitions have long been used in both epidemiologic studies of SSc-ILD and in clinical practice.[45, 200, 201] Yet, the 80% cut-off algorithm tends to be too restrictive as many SSc patients with ILD have FVC measurements that lie above this value.[43] On the other hand, the 10% decline algorithm is too inclusive given the inherent variability of FVC.[43, 48, 297] In contrast, we found that the HMM is accompanied by a low ER and excellent specificity.

The importance of these DTA measures with respect to one another depends on the reason for which the HMM or algorithms are being used. If they are intended as screening tools, emphasis should be placed on sensitivity. Alternatively, if they are intended to ascertain disease status in epidemiologic studies, both sensitivity and specificity should be valued equally, and this is well reflected by the error rate. Our study has shown that the HMM can be an appropriate choice for both uses. It can maximize the identification of disease presence while minimizing the potential for false positives. Furthermore, its sensitivity can be increased by lowering its predicted probability threshold for ILD (Figure 5-2) or

by increasing the number of marker measurements taken prior to disease onset. From a substantive standpoint, this would require that PFTs be performed more frequently.

A drawback of the HMM is that it can be complex and difficult to use in clinical practice. A running average may be a more intuitive approach capable of rendering similar results, albeit with a loss in specificity (Table 5-3). However, unlike the HMM, this method does not provide a probability of ILD onset.

To the best of our knowledge, this is the first study to compare the performance of the HMM to that of crude definitions of disease onset using surrogate markers. A key strength of this study was the use of a simulated dataset in which disease status was known for every visit and could be used to calculate measures of DTA unaffected by selection bias. However, results from simulation studies are often criticized for being too idealistic and not applicable to real clinical settings.[303, 304] For this reason, we extended our study to also include an analysis on a real cohort of SSc patients.

A further strength of our study was the use of the combination gold-standard to deal with the generalizability issue that arises when using only HRCT results as the reference standard, as patients on whom HRCT scans are available tend to have symptoms of lung involvement and/or worsened PFT results. This enabled us to tackle the issue of verification bias and prevented us from discarding useful data. Furthermore, at no point were the combination gold-standard results used to train the HMM precluding the possibility of overfitting.

However, the combination gold-standard also contributed to the study's limitations as chest X-rays and lung auscultation are less sensitive measures of SSc-ILD than HRCT.[29, 58] Furthermore, despite the CSRG being one of the largest SSc cohorts, our sample size was limited due to the rarity of SSc. While the number of patients and visits could have been increased in the simulation study, we chose

to reproduce the size of the CSRG to get a better idea of how the HMM would perform in the context of a limited clinical dataset. In addition, this allowed us to obtain a lower limit for the actual power of the HMM, since its ability to ascertain disease onset improves as data accrues.

A further drawback is the noticeable discrepancy in the sensitivities obtained from the simulation (Table 5-1) and CSRG studies (Table 5-3). These inconsistent results are likely due to the data generating mechanism employed for ILD status in the simulation study. FVC values were simulated to drop by approximately 10% upon ILD onset, whereas this is more likely to occur in a gradual manner. This could artificially increase the ability of the HMM and 10% decline to detect cases of SSc-ILD in the simulation study.

While only a relatively small number of datasets were generated in the simulation study, the low variability across simulations (S.D. of the simulated FVC means = 0.2%) and the small 95% confidence intervals (Table 5-1) eliminate the need for increasing the number of simulated datasets.

Our study assumes that the underlying parametric models for the Markov process and measurement error component are correct. Furthermore, we assume that FVC is a good surrogate marker for SSc-ILD onset and provides good discrimination between ILD absence and presence. Indeed, the performance of the HMM can only be as good as the surrogate marker it uses. The better results observed in the simulation study suggest that FVC may in fact not be the best PFT surrogate marker for ILD onset.

Given the mixed results when using real data, future studies should more thoroughly investigate the performance of HMMs in such settings, perhaps by mimicking slower declines in the marker value upon disease onset. Furthermore, our analyses could be extended to validate multi-state HMMs in the context of disease progression. From a substantive perspective, future work should also determine

whether the performance of the HMM improves as PFT measures other than FVC are considered as potential surrogate markers for SSc-ILD onset. Unfortunately, these were all beyond the scope of this study.

In conclusion, we believe that this study illustrates that the HMM can be a novel and sophisticated tool for epidemiologists to study disease onset when using a disease marker affected by measurement error. The model's use of complete longitudinal data and correction of measurement error make it a more sophisticated alternative to hard cut-offs and pre-determined changes in disease markers.

5.8 Tables and Figures

Table 5-1: Performance of the 80% Cut-Off, 10% Decline and Hidden Markov Model Algorithms in Identifying Interstitial Lung Disease (ILD) Onset in Simulated Systemic Sclerosis (SSc) Patients.

	80% Cut-Off in FVC *		10% Decline in FVC *		Hidden Markov Model *	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	32.6	32.0, 33.3	65.2	64.3, 66.2	57.3	56.5, 58.1
Incident Sensitivity	27.3	26.5, 28.0	61.3	60.3, 62.2	42.0	41.0, 43.0
Prevalent Sensitivity	8.0	7.4, 8.5	10.7	10.2, 11.2	20.1	18.8, 21.4
Specificity	84.7	84.4, 85.0	72.6	72.3, 73.0	88.7	88.4, 88.9
Overall PPV	26.8	26.1, 27.4	29.0	28.5, 29.5	46.5	45.7, 47.3
Incident PPV	25.1	24.4, 25.8	25.3	24.8, 25.7	37.7	36.7, 38.7
NPV	88.0	87.6, 88.2	92.4	92.2, 92.6	92.4	92.1, 92.6
Overall +LR †	2.1	2.1, 2.2	2.4	2.3, 2.4	5.1	5.0, 5.2
Incident +LR †	1.8	1.7, 1.8	2.2	2.2, 2.3	3.7	3.6, 3.9
Overall -LR †	0.80	0.79, 0.80	0.48	0.47, 0.49	0.48	0.47, 0.49
Incident -LR †	0.86	0.85, 0.87	0.53	0.52, 0.55	0.65	0.64, 0.67
Overall ER	23.0	22.7, 23.2	28.4	28.1, 28.8	15.9	15.7, 16.2
Incident ER	18.9	18.6, 19.2	28.1	27.7, 28.4	14.2	14.0, 14.5

** Values are expressed as the mean and 95% confidence interval of the 50 simulated datasets.*

† All measures are expressed as percentages, except for the positive and negative likelihood ratios.

Abbreviations: CI = Confidence Interval; ER = Error Rate; FVC = Forced Vital Capacity; NPV = Negative Predictive Value; PPV = Positive Predictive Value; +LR = Positive Likelihood Ratio; -LR = Negative Likelihood Ratio

Table 5-2: Cumulative Percentage of Simulated Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Cases Identified by the 80% Cut-Off, 10% Decline and Hidden Markov Model (HMM) Algorithms.

ILD Onset at: *	ILD Identified at Visit 2: †		ILD Identified at Visit 3: †		ILD Identified at Visit 4: †		ILD Identified at Visit 5: †		ILD Never Identified: †	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
80% Cut-Off in FVC										
Visit 2 (N = 71.8 ± 1.2)	25.8%	24.2%, 27.4%	32.9%	31.4%, 34.4%	36.7%	35.1%, 38.2%	38.7%	37.2%, 40.3%	61.3%	59.7%, 62.8%
Visit 3 (N = 67.2 ± 1.1)			27.5%	25.9%, 29.2%	34.5%	32.7%, 36.2%	38.4%	36.8%, 40.0%	61.6%	60.0%, 63.2%
Visit 4 (N = 63.2 ± 1.0)					27.7%	26.3%, 29.1%	34.4%	32.9%, 35.8%	65.6%	64.2%, 67.1%
10% Decline in FVC										
Visit 2 (N = 71.8 ± 1.2)	49.1%	47.5%, 50.7%	58.1%	56.6%, 59.7%	63.9%	62.3%, 65.4%	68.7%	67.2%, 70.3%	31.3%	29.7%, 32.8%
Visit 3 (N = 67.2 ± 1.1)			59.9%	58.1%, 61.7%	68.3%	66.7%, 69.9%	73.1%	71.4%, 74.8%	26.9%	25.2%, 28.6%
Visit 4 (N = 63.2 ± 1.0)					67.3%	65.6%, 68.9%	75.3%	73.7%, 76.8%	24.7%	23.2%, 26.3%
Hidden Markov Model										
Visit 2 (N = 71.8 ± 1.2)	13.0%	9.0%, 17.1%	59.7%	58.1%, 61.3%	67.5%	66.0%, 69.1%	71.0%	69.4%, 72.5%	29.0%	27.5%, 30.6%
Visit 3 (N = 67.2 ± 1.1)			61.9%	60.4% 63.4%	69.5%	67.9%, 71.1%	72.6%	71.1%, 74.2%	27.4%	25.8%, 28.9%
Visit 4 (N = 63.2 ± 1.0)					54.6%	52.4%, 56.8%	61.3%	59.2%, 63.4%	38.7%	36.6%, 40.8%

**This column indicates the average number of simulated patients having developed interstitial lung disease at visits 2-4. Values are expressed as the mean and standard error of the 50 simulated datasets.*

†Values are expressed as the mean and 95% confidence interval of the 50 simulated dataset.

Abbreviations: CI = Confidence Interval; FVC = Forced Vital Capacity; ILD = Interstitial Lung Disease

Table 5-3: Performance of the 80% Cut-Off, 10% Decline and Hidden Markov Model (HMM) Algorithms in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	80% Cut-Off in FVC *		10% Decline in FVC *		Hidden Markov Model *		Running Average: 12% Threshold *	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	39.9	35.7, 44.1	47.8	41.0, 54.6	31.4	22.7, 40.2	31.8	26.9, 36.8
Incident Sensitivity	39.9	32.4, 47.4	36.0	27.7, 44.4	22.5	14.1, 30.8	21.9	15.0, 28.8
Prevalent Sensitivity	7.1	2.7, 11.5	18.2	11.2, 25.1	16.0	5.9, 26.0	13.3	7.5, 19.0
Specificity	76.4	74.3, 78.5	56.0	53.2, 58.7	80.2	73.3, 87.2	73.8	71.1, 76.5
Overall PPV	28.6	25.3, 32.0	20.5	17.8, 23.2	27.6	22.4, 32.8	22.4	18.8, 26.0
Incident PPV	18.3	13.1, 23.6	14.5	10.7, 18.3	21.0	14.7, 27.3	16.4	11.0, 21.8
NPV	84.2	82.7, 85.7	81.9	79.6, 84.2	83.1	81.5, 84.7	82.0	80.3, 83.7
Overall +LR †	1.7	1.5, 1.9	1.1	0.9, 1.2	1.6	0, Inf	1.2	1.0, 1.4
Incident +LR †	1.7	1.5, 1.9	0.8	0.6, 1.0	1.2	0.6, 1.7	0.8	0.5, 1.2
Overall -LR †	0.79	0.72, 0.86	0.93	0.81, 1.03	0.85	0, Inf	0.92	0.85, 1.00
Incident -LR †	0.79	0.66, 0.91	1.14	1.00, 1.28	0.97	0.86, 1.07	1.06	0.96, 1.15
Overall ER	30.6	28.7, 32.6	45.6	43.2, 48.0	29.1	24.8, 33.5	34.3	31.8, 36.7
Incident ER	25.6	23.6, 27.7	45.1	42.5, 47.7	22.9	16.6, 29.3	29.1	26.5, 31.7

** Values are expressed as the mean and 95% confidence interval of the 25 imputed datasets.*

† All measures are expressed as percentages, except for the positive and negative likelihood ratios.

Abbreviations: CI = Confidence Interval; FVC = Forced Vital Capacity; Inf = Infinity; NPV = Negative Predictive Value; PPV = Positive Predictive Value; +LR = Positive Likelihood Ratio; -LR = Negative Likelihood Ratio

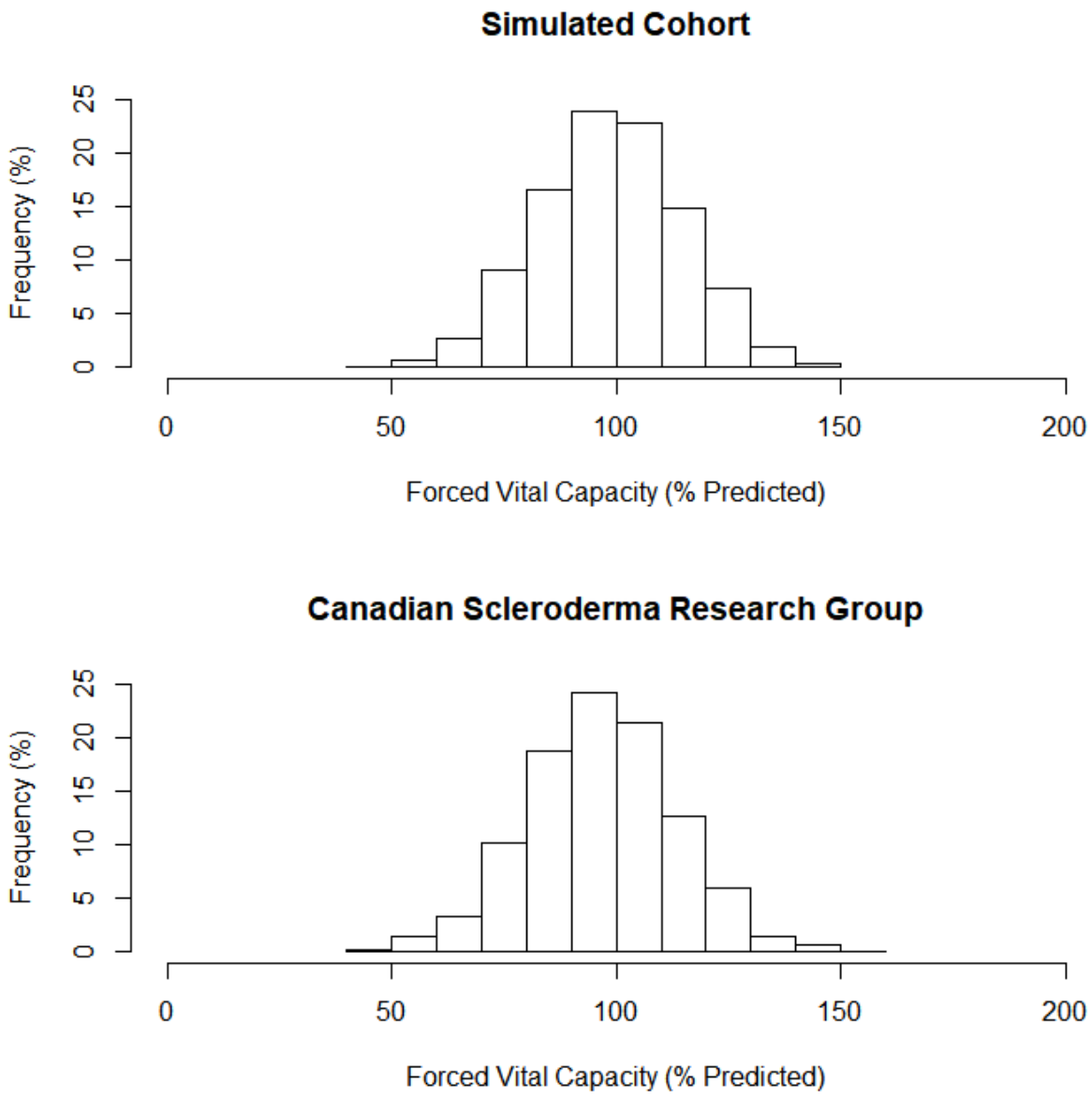


Figure 5-1: Distribution of Forced Vital Capacity (FVC) Values in the Simulated Systemic Sclerosis (SSc) Cohort and in the Canadian Scleroderma Research Group Registry (CSRG).

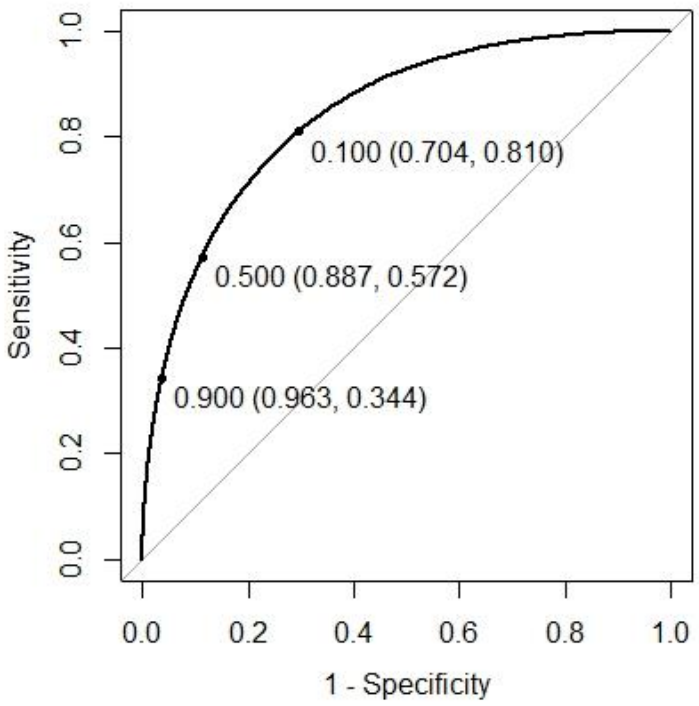


Figure 5-2: Receiver Operating Characteristic (ROC) Curve Illustrating the Effect of Varying the Predicted Probability Cut-Off of the Hidden Markov Model (HMM) on Overall Sensitivity and Specificity for Interstitial Lung Disease (ILD) Presence in Simulated Systemic Sclerosis (SSc) Patients. *The 0.10%, 0.50% and 0.90% cut-off coordinate points are depicted with their corresponding specificity and sensitivity.*

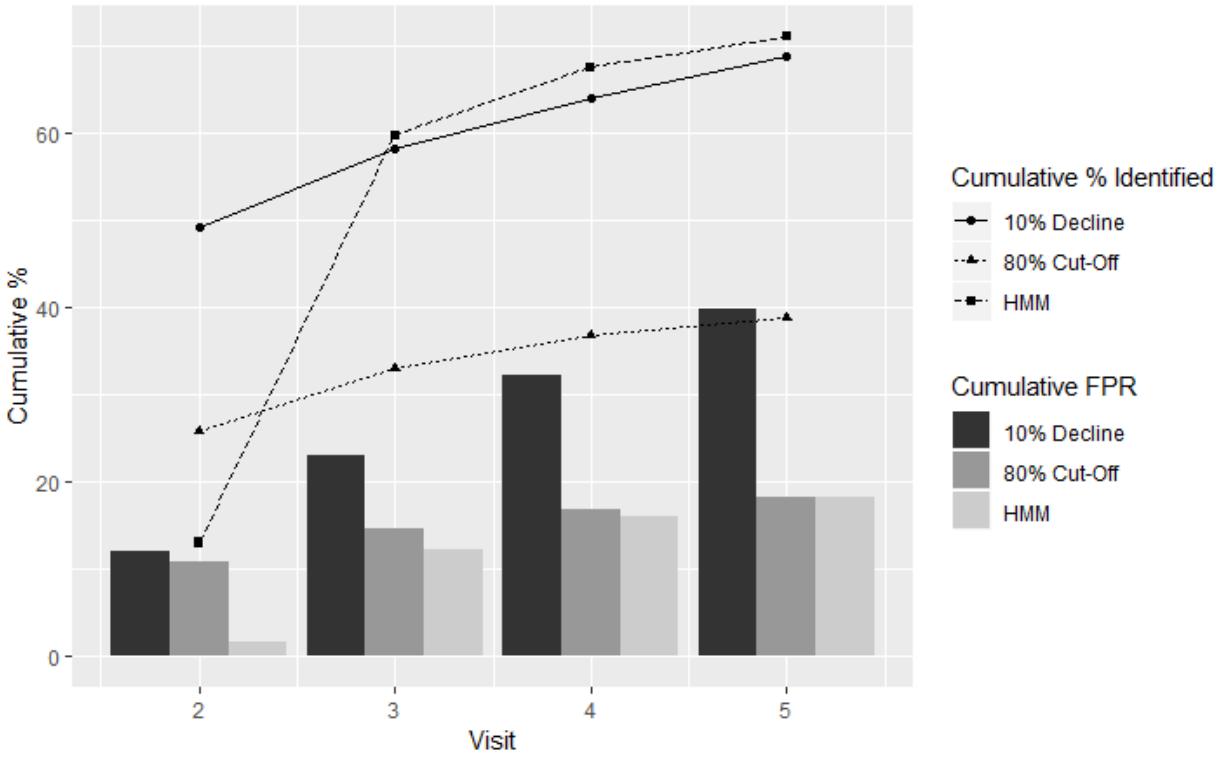


Figure 5-3: Cumulative Percentage of Simulated Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Cases Identified by the 80% Cut-Off, 10% Decline and Hidden Markov Model (HMM) Algorithms and Associated Cumulative False Positive Rates when Interstitial Lung Disease (ILD) Onset Occurred at Visit 2.
Abbreviations: FPR = False Positive Rate; HMM = Hidden Markov Model

Chapter 6: Validation of Pulmonary Function Test Measures as Surrogate Markers for Systemic Sclerosis-Associated Interstitial Lung Disease Onset Using Hidden Markov Models

6.1 Preface to Manuscript 3

An overview of previous pulmonary function test (PFT) validation studies in Chapter 4 suggested the need to further investigate diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC) in addition to forced vital capacity (FVC) as potential surrogate markers for systemic sclerosis-associated interstitial lung disease (SSc-ILD) onset. In Chapter 5, we introduced hidden Markov models (HMMs) as a promising new method of using PFT measures to detect SSc-ILD. In particular, we showed how they may be more suitable than the use of fixed cut-offs and changes in PFT measurements. The final objective of this thesis aimed to validate different univariate and multivariate HMMs using longitudinal measurements of FVC, DLCO and/or TLC to detect SSc-ILD.

Data from the Canadian Scleroderma Research Group (CSRG) were used for this validation study of both absolute and % predicted PFT measures. For each single and composite PFT measure, we fit a three-state HMM to represent states of SSc-ILD absence, SSc-ILD presence and death. All models adjusted for baseline PFT measurements. Measures of diagnostic test accuracy were calculated using a combination of high-resolution computed tomography, chest X-ray and lung auscultation results as a gold-standard for the presence of ILD.

The resulting manuscript, entitled "Validation of Pulmonary Function Test Measures as Surrogate Markers for Onset of Interstitial Lung Disease in Systemic Sclerosis", is prepared for submission to Rheumatology (Oxford) where it is currently under review.

6.2 Title Page

Title: Validation of Pulmonary Function Test Measures as Surrogate Markers for Onset of Interstitial Lung Disease in Systemic Sclerosis

Authors: Melissa Caron[1, 2], Kevin Schwartzman[1, 2], Marie Hudson[3], Murray Baron[3], Russell J. Steele[4] for the Canadian Scleroderma Research Group

Affiliations:

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

[2] Respiratory Epidemiology and Clinical Research Unit (RECRU), Montreal Chest Institute, McGill University Health Centre, Montreal, QC, Canada

[3] Division of Rheumatology, Sir Mortimer B Davis Jewish General Hospital, Montreal, QC, Canada

[4] Department of Mathematics and Statistics, McGill University, Montreal, QC, Canada

Corresponding Author: Melissa Caron, 1020 Pine Ave W, Montreal, QC, Canada, H3A 1A2. E-mail: melissa.caron@mail.mcgill.ca

Funding: No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Keywords: systemic sclerosis; interstitial lung disease; pulmonary function test; validation study

Key messages:

- DLCO and TLC may be more appropriate surrogate markers for SSc-ILD onset than FVC.
- A combined FVC/DLCO/TLC screening test achieved the greatest sensitivity for SSc-ILD onset.
- Overall, PFT measures had low sensitivity for SSc-ILD onset suggesting the need for further research.

6.3 Abstract

Objectives. Interstitial lung disease (ILD) is the leading cause of morbidity and mortality in systemic sclerosis (SSc) patients. Pulmonary function test (PFTs) are routinely used as screening tools and surrogate markers for SSc-ILD, yet few studies have evaluated their validity. We aimed to assess the performance of forced vital capacity (FVC), diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC) as surrogates for SSc-ILD onset.

Methods. A longitudinal validation study was performed using data from a large SSc cohort. Hidden Markov models (HMMs) were chosen to validate univariate and composite PFT measures as they use the full history of PFT measurements to model the probability of ILD onset while simultaneously correcting for measurement error. Measures of diagnostic test accuracy were calculated using a combination of high-resolution computed tomography, chest X-ray and lung auscultation results as a gold-standard for the presence of ILD.

Results. All PFT measures had poor sensitivity for SSc-ILD. The highest sensitivity was estimated for the HMM using a composite absolute FVC/DLCO/TLC measure [34.4% (29.2, 39.7)]. The joint TLC % predicted and forced expiratory volume in one second (FEV_1)/FVC model had the greatest specificity [88.8% (85.5, 92.1)] and lowest error rate [24.9% (22.5, 27.4)], followed closely by DLCO % predicted and TLC absolute.

Conclusion. These results demonstrate that, despite FVC's popularity, TLC and DLCO should also be considered when screening for SSc-ILD onset. However, PFT measures' low sensitivities suggest that other avenues should be explored in the search for a suitable marker for SSc-ILD onset.

6.4 Introduction

Systemic sclerosis (SSc) is a chronic autoimmune disease characterized by the presence of vasculopathy, immune dysregulation and fibrosis of the skin and internal organs.[1] Of the many clinical complications associated with SSc, interstitial lung disease (ILD) is the leading cause of morbidity and mortality.[28] Prevalence estimates for SSc-associated ILD (SSc-ILD) vary considerably,[27] but there is growing evidence that at least 50% of patients are affected.[29, 86] The course of SSc-ILD is highly heterogeneous, but approximately 15% of patients experience rapidly progressive lung function decline in the first few years following disease onset.[3, 4, 30] Since ILD is irreversible and since immunosuppressant and anti-fibrotic medications can only slow the course of the disease while being associated with serious toxicities,[3, 5, 6, 11, 34] a good screening tool is imperative to identify new cases of SSc-ILD in a timely manner and to balance the risks and benefits of treatment.

High-resolution computed tomography (HRCT) is generally considered the clinical standard for assessing the presence and severity of SSc-ILD.[7, 8] However, its repeated use is costly and could unnecessarily expose asymptomatic SSc patients to high doses of radiation.[9] Consequently, pulmonary function tests (PFTs), which include physiologic parameters of lung restriction consistent with ILD, are routinely used as screening tools and surrogate markers for SSc-ILD. However, few studies have evaluated their validity and there is currently no consensus on the best PFT measure to screen for onset of SSc-ILD.

Forced vital capacity (FVC) is the PFT measure most commonly used to detect SSc-ILD onset, despite a lack of published evidence that it has good discriminatory capacity.[305] In fact, among the few existing studies aiming to validate PFT measures, most found diffusing capacity for carbon monoxide (DLCO) and total lung capacity (TLC) to have the best diagnostic accuracy.[238, 279, 305] However, these validation studies evaluated SSc-ILD extent and not onset. Furthermore, they did not account for any of the

variability inherent in PFT measurement and were cross-sectional thereby ignoring patients' entire PFT history.

In addition to these methodologic limitations, PFT measures themselves have been criticized for being poor screening tools for SSc-ILD. For instance, normal PFT values are often considered to range from 80% to 120% of the predicted mean for age- and sex-matched reference populations,[306] while many SSc patients with ILD exhibit PFT values that lie well within this range of normality.[43, 307] Furthermore, PFT measures, particularly DLCO, can be affected by the presence of pulmonary hypertension, the second most common form of lung involvement in SSc patients,[45, 285] and anemia, which is not uncommon in SSc.

In this study, we aimed to validate FVC, DLCO and TLC as surrogate markers for SSc-ILD onset while using an advanced statistical model to address all these concerns.

6.5 Methods

Study Design and Population

We performed a longitudinal validation study following the STARD recommendations for diagnostic accuracy studies.[308] The study utilized data from the Canadian Scleroderma Research Group (CSRG), a large, multi-center Canadian cohort of SSc patients. The CSRG registry has recruited over 1,600 patients from 15 sites across Canada and one site in Mexico. Patients have a rheumatologist-confirmed diagnosis of SSc, are 18 years of age or older, fluent in English or French, and likely to be compliant with study procedures and yearly visits. Over 98% of the cohort meet the 2013 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) classification criteria for SSc.[53] Annual visits include standardized physical evaluations, chest X-rays and pulmonary function testing.

Study Inclusion and Exclusion Criteria

CSRG subjects with ILD at baseline were excluded in order to study ILD occurrence. A minimum of three PFT measurements was required to run the statistical model of interest. Consequently, only patients with a baseline and two follow-up annual visits were included in the study.

Combination Gold-Standard (Reference Test)

Since HRCT scans were not performed annually, ILD status was assessed using a combination gold-standard which used HRCT results when available, but otherwise used a previously validated clinical decision rule for ILD presence. This rule suggests that ILD is present if there is evidence of fibrosis on chest X-ray or if crackles are heard during lung auscultation,[29] information which is collected yearly on each CSRG patient.

Pulmonary Function Tests (Index Test)

PFTs were performed at participating centers in accordance with American Thoracic Society (ATS) standards. Both absolute and percent (%) predicted values for FVC, DLCO (single-breath, corrected for hemoglobin) and TLC were considered in this study.

Shortness of Breath

Shortness of breath was used as a comparator to better contextualize the performance of PFT measures. It was measured in two ways. The first was patient-reported and used a numerical rating scale from 0 to 10 indicating no to severe limitations to daily activities. The second shortness of breath measure was physician-assessed as per the New York Heart Association (NYHA) functional class.[54]

Statistical Analysis

FVC, DLCO and TLC were evaluated as surrogate markers for SSc-ILD onset using a hidden Markov model (HMM). HMMs have already been used to study the onset and progression of other chronic diseases, such as bronchiolitis obliterans post-lung transplant, chronic obstructive pulmonary disease and HIV infection.[64, 309, 310]

Our study used HMMs with two distinct states for ILD absence and ILD presence. The HMM uses patients' entire PFT history to estimate the instantaneous risk (also referred to as a transition intensity) of transitioning from the former state to the latter.[63, 66] HMMs are particularly powerful tools as they can estimate this transition intensity while simultaneously correcting for the random variability that can accompany the collection of PFT measures. The HMM can subsequently use the estimated transition intensity along with a patient's full PFT measurement history to predict the probability of ILD presence on the patient's most recent pulmonary function testing date.

Censoring due to death was accounted for in our study by including a death state in the HMM in addition to the ILD-absence and presence states. We thus fit three-state HMMs using the `msm` package in `R`.[65, 298] Separate models were fit for FVC, DLCO and TLC, as well as for their bivariate and multivariate combinations. An additional DLCO HMM was considered in which we adjusted for the presence of pulmonary hypertension. We also considered a bivariate model for TLC and the ratio of forced expiratory volume in one second (FEV_1) to FVC, since ATS defines restrictive lung diseases such as ILD by a reduction of TLC in the presence of a preserved or elevated FEV_1/FVC.[42] Each model was fit twice using either absolute or % predicted values for each PFT variable.

Two additional HMMs using patient-reported and physician-assessed shortness of breath served as points of comparison in order to assess the additional validity and precision gained by using PFT measures rather than symptoms of dyspnea as markers for SSc-ILD.

Running HMMs requires specification of the underlying probability distribution taken by each PFT measure in both the absence and presence of SSc-ILD. Given the reasonably symmetric distribution of PFT values, we assumed these measures to be normally distributed with initial means and standard deviations (S.D.s) reflecting those observed in ILD-absent and present visits of CSRG patients with at least three study visits and who were ILD-free at baseline. The dyspnea variables were assumed to follow a Poisson distribution with rates equal to the mean numerical rating observed in these patients' ILD-absent and present visits.

All models adjusted for baseline PFT measurements to account for the heterogeneity between subjects. Since the current `msm` package does not yet support the addition of covariates in bivariate and multivariate HMMs, this adjustment was made by subtracting patients' baseline PFT value from all their subsequent annual measurements. For this reason, baseline visits were not included in the analyses.

The probability of ILD presence on the dates for which patients had an available combination gold-standard result were obtained from the HMM. A cut-off ≥ 0.50 was chosen to indicate HMM-detected ILD-presence. This result was then compared to the true ILD status (known from the combination gold-standard) to calculate measures of validity. The combination gold-standard results were strictly used to evaluate the HMMs; only PFT measurements were used to train them.

An error rate corresponding to the sum of false negatives and false positives divided by the total number of comparisons was calculated in addition to standard measures of validity. The measures of diagnostic test accuracy were also assessed using only the subjects' first ILD-positive result to determine the PFT measures' ability to detect *incident* cases of ILD. Finally, a Brier score was calculated for each model as an indicator of the accuracy of their probabilistic predictions. Brier scores were

calculated as the mean of the difference between the HMM predicted probabilities and the actual ILD status (0 if ILD-absent, 1 if ILD-present). Models with lower Brier scores are thus better calibrated.[311]

Missing Data

Missing absolute and % predicted values for the PFT measures were imputed using multiple imputation by chained equations.[299] Predictors included results from HRCT scans (normal or abnormal), SSc type (limited or diffuse), current smoking status, presence of pulmonary hypertension, shortness of breath, and presence of auto-antibodies predictive of ILD (anti-topoisomerase, anti-centromere, and anti-RNA polymerase III).[4, 27, 193, 223, 300-302] Twenty-five imputed datasets were generated and the measures of diagnostic test accuracy were averaged across all datasets.

In the case of missing HRCT, chest X-ray and physical examination data for the combination gold-standard, the patient was assumed to be free of ILD. Since fibrosis of the lungs is an irreversible process, once a patient was diagnosed by the gold-standard as having ILD they could not subsequently revert to being ILD-free.[11]

Finally, since PFTs were performed annually, missing PFT dates were imputed by adding or removing a year from the previous or next available test date, respectively. If no PFT dates were available, the date on which the study visit occurred was used.

Ethical Considerations

Ethics approval for the CSRG registry and its data collection protocol was obtained at McGill University (Montreal, Canada) and at all participating study centers. All subjects provided written informed consent to participate in the registry. Since this study involved the secondary use of deidentified data, no further consent was required.

6.6 Results

This study included 831 CSRG patients with an average of 6.6 (S.D. = 2.7), minimum of three and maximum of 13 study visits. Table 6-1 outlines the subjects' baseline characteristics. Throughout the follow-up period, 165 patients (19.9%) developed ILD and 80 deaths due to all-causes were recorded.

A complete overview of the performance metrics of all evaluated HMMs is available in Appendix P (Appendix Tables P-1 – P-3). The sensitivities of the various single and composite PFT HMMs are illustrated in Figure 6-1. As a general trend, absolute PFT measures appeared to have a greater sensitivity than their % predicted counterparts. Unsurprisingly, bivariate and multivariate PFT combinations (depicted in green and red, respectively) were more sensitive than using single PFT measures (depicted in blue).

All PFT measures exhibited low sensitivity and none outperformed NYHA functional class with its sensitivity of 63.2% (56.7-69.6). The PFT HMM with the greatest sensitivity used a composite measure of absolute FVC, DLCO and TLC [34.4% (95% CI: 29.2, 39.7)]. However, its sensitivity increased to 60.5% (95% CI: 48.3, 72.7) when a predicted probability cut-off of ≥0.10 was used. The corresponding receiver operating characteristic (ROC) curve for this model is illustrated in Figure 6-2. This model's incident sensitivity, or ability to detect an incident case of ILD, was surpassed only by the composite % predicted FVC, DLCO and TLC measure [21.1% (95% CI: 14.5, 27.8) vs. 22.2% (95% CI: 15.8, 28.6), respectively] (Appendix Tables P-2 – P-3).

Figure 6-3 summarizes the specificities of the different PFT HMMs, all of which were greater than the specificity associated with the NYHA functional class shortness of breath variable. Specificities were generally higher for the single PFT measures than for the bivariate and multivariate PFT measures except for the TLC % predicted and FEV_1/FVC composite measure which had the highest specificity [88.8% (95% CI: 85.5, 92.1)]. However, its specificity was largely equivalent to the HMMs

using DLCO % predicted [87.0% (95% CI: 80.1, 93.8)], TLC absolute and FEV1/FVC [86.7% (95% CI: 83.1, 90.3)], and TLC absolute [86.0% (95% CI: 82.3, 89.8)]. In contrast, the FVC HMM's specificity was lower, but was accompanied by a wide 95% CI [FVC % predicted: 82.1% (95% CI: 75.3, 88.9); FVC absolute: 83.1% (95% CI: 75.6, 90.5)].

The overall error rates are shown in Figure 6-4. All were lower than for the HMMs using shortness of breath, specifically that using NYHA function class. The TLC % predicted and FEV_1/FVC composite measure had the lowest error rate [24.9% (95% CI: 22.5, 27.4)], followed by the HMMs using TLC absolute [26.1% (95% CI: 23.3, 29.0)], TLC absolute and FEV1/FVC [26.2% (95% CI: 23.6, 28.7)], and DLCO % predicted [26.3% (95% CI: 21.6, 31.0)]. Again, FVC had a greater error rate but wide 95% CI [FVC % predicted: 29.0% (95% CI: 24.7, 33.4); FVC absolute: 28.0% (95% CI: 22.5, 33.5)].

Similarly to the specificities and error rates, the HMMs with the lowest Brier scores were those for TLC % predicted and FEV1/FVC [0.217 (95% CI: 0.215, 0.220)], TLC absolute [0.218 (95% CI: 0.214, 0.222)], DLCO % predicted [0.222 (95% CI: 0.215, 0.230)] and TLC absolute and FEV1/FVC [0.226 (95% CI: 0.223, 0.229)] (Appendix Tables P-2 – P-3). These were lower than the Brier score for FVC absolute HMM [0.230 (95% CI: 0.223, 0.238)] and significantly lower than that for FVC % predicted HMM [0.243 (95% CI: 0.237, 0.249)].

6.7 Discussion

This study demonstrated that DLCO and TLC should not be disregarded when using PFT measures as either screening tools or surrogate markers for SSc-ILD onset. Despite FVC's popularity in the SSc-ILD arena, our results do not suggest that its exclusive use is warranted.

SSc-ILD is a serious disease with a short window for the successful administration of stabilizing treatment. A screening tool should thus emphasize sensitivity and the reduction of false negative

results. Our findings suggest that jointly observing a patient's history of FVC, DLCO and TLC measurements is more sensitive than focusing solely on any one of these PFT measures. Consequently, treating physicians should consider the combined change in these PFT measures when assessing whether to investigate further. This could be especially useful in clinical practice to monitor patients with a normal baseline HRCT.

Alternatively, if PFT measures are used as surrogate markers in epidemiologic studies to ascertain SSc-ILD status then the accuracy of these studies will depend not only on sensitivity, but on other components of validity including specificity. It follows that the error rate and Brier score are good measures of PFT measures' overall calibration and discrimination properties, as they incorporate the effects of both sensitivity and specificity. In this case, the HMM with the lowest error rate and Brier score was that using a composite TLC % predicted and FEV_1/FVC measure, suggesting that this might be the best PFT surrogate marker for SSc-ILD presence. This result supports ATS's stance that restrictive lung diseases are best defined by TLC and FEV_1/FVC.[42] Its performance did however overlap considerably with that of TLC absolute, a joint TLC absolute and FEV_1/FVC measure, and DLCO % predicted.

The latter result supports the existing body of evidence that has found DLCO to be a good measure of SSc-ILD extent.[238, 279] These past validation studies however did not validate a joint TLC and FEV_1/FVC outcome. To the best of our knowledge, our study is the first to do so. It is also the first to evaluate absolute PFT measures (albeit while adjusting for baseline values) which have been overlooked in prior validation studies, despite the fact that they have been widely used historically and are still occasionally considered as primary endpoints in SSc-ILD studies.[34, 305]

Our study used a sophisticated statistical model which allowed us to address limitations of previous validation studies. Indeed, these studies often used firm cut-offs or pre-specified changes in PFT

values to determine SSc-ILD onset which ignore patients' entire history of PFT measurements. The use of HMMs allowed for the validation of PFT measures using complete longitudinal data. Another advantage of the HMM is that it accounts for the random variability associated with PFT measurements that can arise from subject, instrument, and/or methodologic sources.

The idea of using a statistical model which studies transitions through SSc-ILD states has previously been considered. Indeed, a similar statistical strategy has been used to evaluate the effect of drug therapy on the probability of transitioning from one SSc-ILD severity pattern to another.[50, 312] These studies used HRCT results, but in the absence of such data, HMMs can use surrogate markers, such as PFT measures, to study state transitions while correcting for PFT measurement error.

HMMs could be an excellent method to screen for SSc-ILD in clinical practice. Indeed, if a HMM using a composite FVC/DLCO/TLC measure were used as a screening tool for SSc-ILD onset, it would act as a sophisticated version of running three screening tests in parallel. Essentially, it would look for separate declining trends in FVC, DLCO and TLC while underscoring any joint deterioration in two or more of these PFT measures. Furthermore, HMMs would be advantageous as they would reduce subjectivity when evaluating a patient's full history of PFT measurements by providing a probability of SSc-ILD presence. This enables clinicians to choose a cut-off value that ensures the achievement of a desired level of sensitivity and specificity. While the HMM is admittedly a complex tool to use in clinical practice, it could be developed into a user-friendly web-application where a treating physician need only enter a patient's PFT measurements and the dates on which these were collected to obtain a probability of ILD presence.

Similarly, if researchers used HMMs together with a joint TLC % predicted and FEV_1/FVC outcome to determine SSc-ILD status in the absence of HRCT, they would increase the reliability of their studies by mitigating measurement error in these PFT parameters. It follows that more accurate

ascertainment of SSc-ILD status would better enable the identification of predictors of SSc-ILD onset and the evaluation of treatment efficacy. A description of how possible predictor and treatment variables can be included as covariates in HMMs and how the results can be interpreted is available elsewhere.[63, 64]

Our validation study has several strengths in addition to the use of HMMs. We adjusted for baseline PFT measurements, thereby allowing patients to be compared to their own "normal" values. Also, in the case of DLCO, we explored adjusting for the presence of pulmonary hypertension, but this did not significantly affect the results (Appendix Tables P-2 – P-3). Finally, the use of the combination gold-standard as the reference test allowed us to prevent verification bias that would have occurred had we only used HRCT results,[55] as CSRG patients for whom HRCT scans are available tend to have symptoms of lung involvement and/or worsened PFT results.

Nevertheless, the combination gold-standard also contributed to the study's limitations as plain chest radiographs and lung auscultation are less sensitive measures of SSc-ILD than HRCT.[29, 58] An additional drawback of our results is the considerable overlap in 95% CIs precluding the possibility of definitively suggesting the best PFT surrogate marker for SSc-ILD. This is potentially due to the limited sample size of the CSRG registry, an unfortunate consequence of studying rare diseases such as SSc.

It is also possible that PFT measures themselves may simply not be good surrogate markers for SSc-ILD. This is highlighted by their overall low sensitivities. In fact, the sensitivities of PFT measures were lower than those of patient and physician-reported shortness of breath. These results support a previous report which also found PFT measures to have poor sensitivity along with a high rate of false negative results especially in early stages of ILD.[43] Future work should explore other avenues for SSc-ILD detection and surrogate markers, including the use of composite outcomes comprised of PFT

results and patient-reported outcomes.[220] Additionally, recent studies have shown promising results when investigating the use of serum biomarkers as potential surrogate markers for SSc-ILD.[44, 313-315]

Further research should also evaluate the validity of PFT measures as surrogate markers for SSc-ILD progression since our study dealt specifically with SSc-ILD onset. This could be achieved by increasing the number of ILD states in HMMs (e.g. different states to represent mild, moderate and severe SSc-ILD).

In summary, we used an advanced statistical model to examine the validity of different PFT measures for SSc-ILD onset. Despite its frequent use, we found that FVC is not superior to other PFT measures. Our results suggest that clinicians and researchers should at the very least consider using DLCO and TLC in addition to FVC. In fact, a combined FVC/DLCO/TLC outcome could serve as a more sensitive PFT screening test to detect possible ILD, while a composite TLC % predicted and FEV_1/FVC measure may be the best PFT surrogate marker to ascertain ILD status in SSc-ILD studies. Nevertheless, PFTs have low overall sensitivity for SSc-ILD and future research should focus on the performance of other potential surrogate markers for SSc-ILD onset in the absence of HRCT.

6.8 Acknowledgements

M.C. would like to acknowledge doctoral grants from the Fonds de Recherche du Québec (Santé PhD Studentship) and from the Canadian Institutes of Health Research (GSD – 146268).

6.9 Tables and Figures

Table 6-1: Baseline Characteristics of Canadian Scleroderma Research Group (CSRG) Patients Included in this Study, N = 831. *Patients with missing data were not included in the denominator of the % calculations.*

	Mean or %	S.D. or n
Sociodemographic Characteristics		
Sex		
Women	88.6%	736
Men	11.4%	95
Age (Years)	54.0	12.0
Smoking Status		
Never/Past Smoker	83.7%	671
Current Smoker	16.3%	131
Clinical Manifestations of Disease		
Disease Extent		
Limited/Sine	67.4%	558
Diffuse	32.6%	270
Pulmonary Hypertension	5.3%	39
Anti-Nuclear Antibodies	95.1%	343
Anti-Centromere	42.6%	278
Anti-Topoisomerase	10.4%	64
Anti-RNA Polymerase III	18.0%	81
Shortness of Breath		
Patient-Reported (Numerical Rating Scale 0 to 10)	1.4	2.1
NYHA Functional Class		
Class I	62.0%	510
Class II	34.9%	287
Class III	2.8%	23
Class IV	0.4%	3
Pulmonary Function Tests		
FVC		
Absolute (L)	3.1	0.8
% Predicted	96.9	16.7
DLCO		
Absolute [mL/(min)(mm HG)]	17.1	5.0
% Predicted	74.7	19.0
TLC		
Absolute (L)	5.0	1.0
% Predicted	98.8	15.3
FEV_1/FVC	77.6	9.3

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; NYHA = New York Heart Association; TLC = Total Lung Capacity

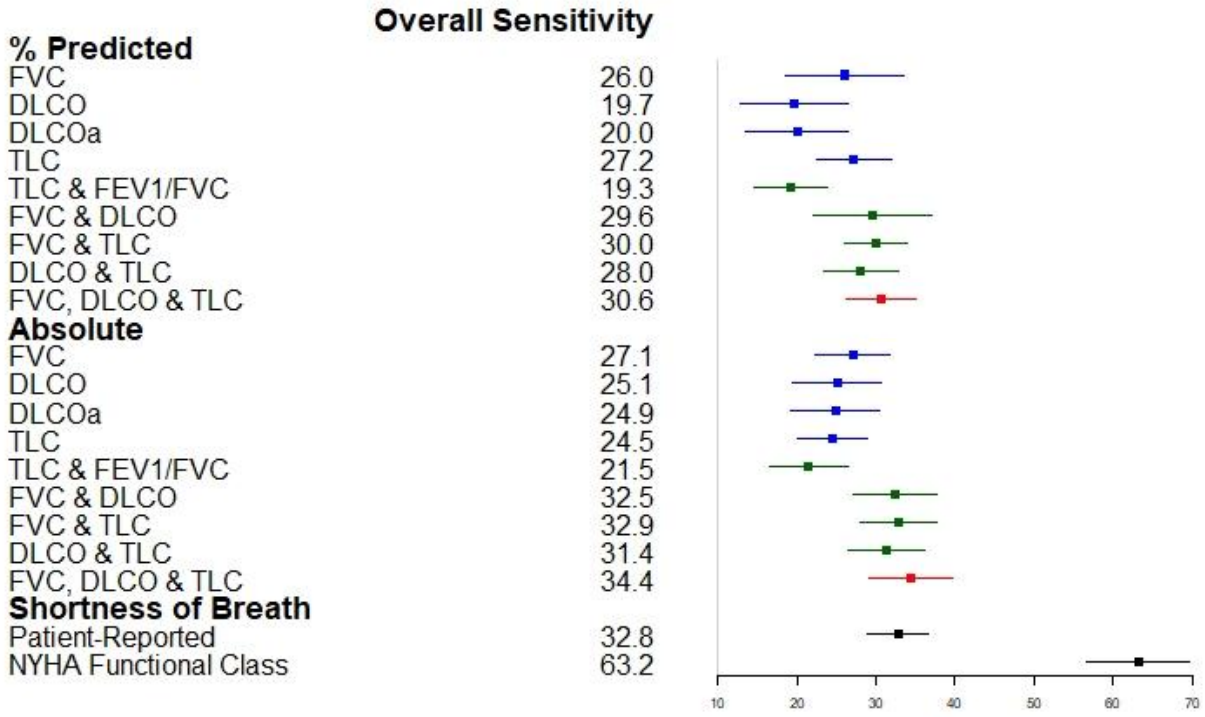


Figure 6-1: Overall Sensitivity of Pulmonary Function Test (PFT) Measures when Used as Surrogate Markers for Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) in Hidden Markov Models (HMMs). *Values are expressed as the mean and 95% confidence interval of the 25 imputed datasets. Univariate, bivariate and trivariate PFT measures are depicted in blue, green and red, respectively. Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; DLCOa = Diffusing Capacity for Carbon Monoxide Adjusted for Pulmonary Hypertension; FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; TLC = Total Lung Capacity*

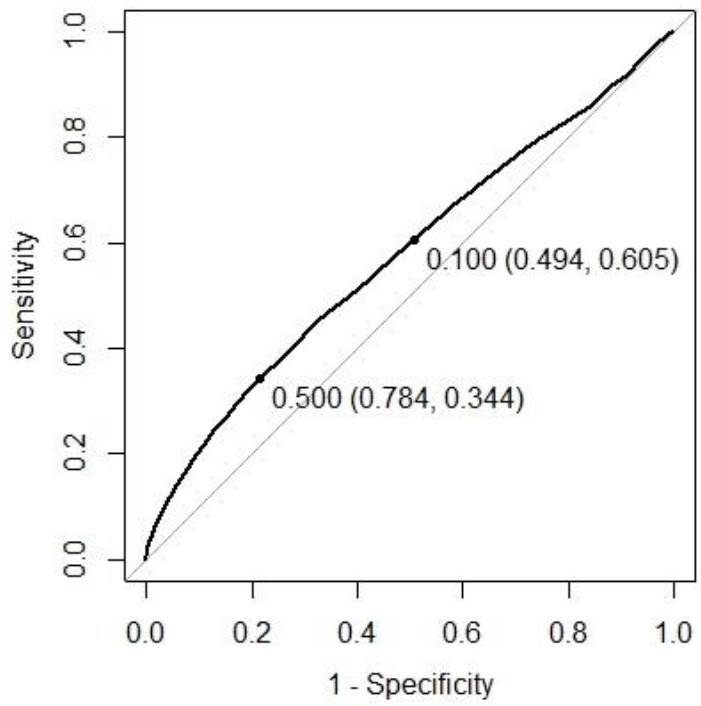


Figure 6-2: Effect of Varying the Probability Cut-Off of the Joint Absolute Forced Vital Capacity (FVC), Diffusing Capacity for Carbon Monoxide (DLCO), and Total Lung Capacity (TLC) Hidden Markov Model (HMM). *The ≥0.10 and ≥0.50 cut-off coordinate points are depicted with their corresponding specificity and sensitivity. The Area Under the Curve (AUC) was 0.577 (95% CI: 0.573, 0.582). Values are expressed as the mean of the 25 imputed datasets.*

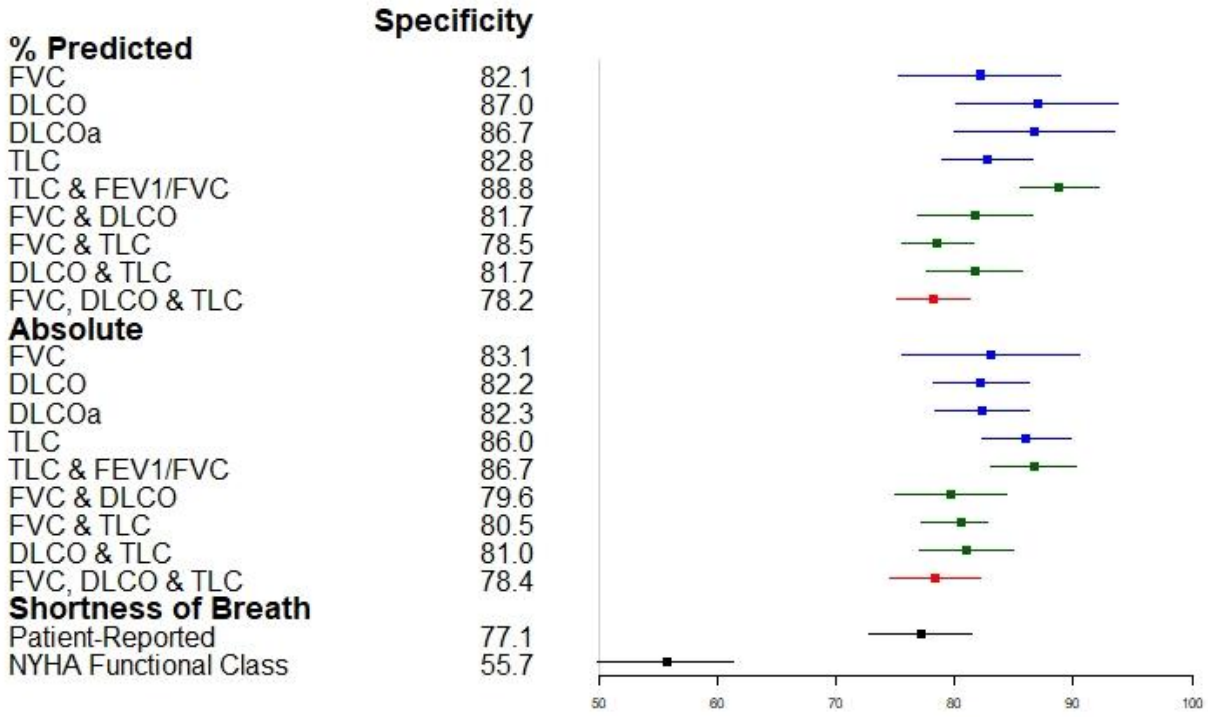


Figure 6-3: Specificity of Pulmonary Function Test (PFT) Measures when Used as Surrogate Markers for Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) in Hidden Markov Models (HMMs). *Values are expressed as the mean and 95% confidence interval of the 25 imputed datasets. Univariate, bivariate and trivariate PFT measures are depicted in blue, green and red, respectively. Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; DLCOa = Diffusing Capacity for Carbon Monoxide Adjusted for Pulmonary Hypertension; FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; TLC = Total Lung Capacity*

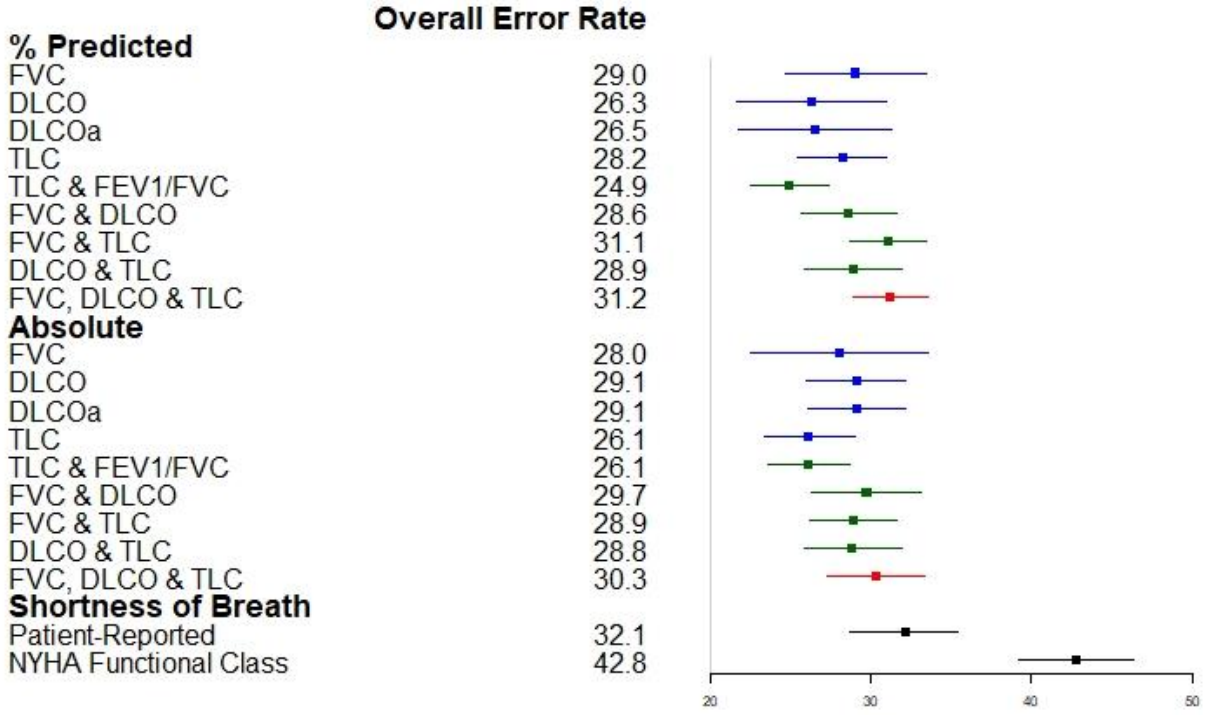


Figure 6-4: Overall Error Rate of Pulmonary Function Test (PFT) Measures when Used as Surrogate Markers for Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) in Hidden Markov Models (HMMs). *Values are expressed as the mean and 95% confidence interval of the 25 imputed datasets. Univariate, bivariate and trivariate PFT measures are depicted in blue, green and red, respectively. Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; DLCOa = Diffusing Capacity for Carbon Monoxide Adjusted for Pulmonary Hypertension; FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; TLC = Total Lung Capacity*

Chapter 7: Discussion

7.1 Summary of Findings

The first study in this thesis was a comprehensive and systematic review of the literature describing the historic use and validation of pulmonary function test (PFT) measures as surrogate markers for systemic sclerosis-associated interstitial lung disease (SSc-ILD). A search of five electronic databases identified 169 studies having used one or multiple PFT measures as outcomes for SSc-ILD onset or progression. Upon review, we found that diffusing capacity for carbon monoxide (DLCO) was cumulatively the most commonly used SSc-ILD outcome until 2010 when it was manifestly surpassed by forced vital capacity (FVC). Among the 71 studies which had a clear primary PFT endpoint, 70.4% chose FVC % predicted while 11.3% chose DLCO % predicted and none opted for total lung capacity (TLC). We identified only five studies with the precise aim of validating PFT measures for SSc-ILD. Two of these validation studies concluded that DLCO % predicted was the best measure of SSc-ILD extent while the remaining three did not favour any PFT measure. The results of this study confirmed our hypothesis that FVC is the preferred PFT surrogate marker for SSc-ILD despite a lack of evidence that its validity surpasses that of other PFT measures.

The second study used both simulated data and real data from a large Canadian cohort of systemic sclerosis (SSc) patients to evaluate the potential of hidden Markov models (HMMs) to use longitudinal PFT measurements to detect SSc-ILD onset. We considered HMMs as an alternative to the use of hard cut-offs and fixed changes in PFT values since they can use a patient's entire history of PFT measurements to predict their probability of having interstitial lung disease (ILD) while simultaneously correcting for random variability inherent in pulmonary function testing. In this study, we fit a HMM to simulated and real FVC measurements and used a predicted probability threshold ≥ 0.50 to indicate ILD presence. The HMM's statistical performance was compared to that of two common definitions

of SSc-ILD onset: <80% cut-off and ≥10% decline in FVC. Using a combination of high-resolution computed tomography (HRCT), chest X-ray and lung auscultation results as a gold-standard for SSc-ILD, we found that the HMM had the highest specificity and lowest error rate compared to the hard cut-off and change in FVC algorithms. The HMM also achieved a higher sensitivity than these two algorithms when the predicted probability threshold was lowered. These results suggest that HMMs can act as screening tools and can be useful to ascertain ILD status in epidemiologic studies.

In the third study, HMMs were used to validate FVC, DLCO and TLC as surrogate markers for SSc-ILD onset. In so doing, we addressed a major limitation of the five previously identified PFT validation studies: their use of cross-sectional data only and the failure to consider patients' entire PFT history. Both absolute and % predicted values were considered, as well as composite PFT measures. We found all PFT measures to have low sensitivity for SSc-ILD. The HMM with the highest sensitivity used a composite absolute FVC/DLCO/TLC outcome and achieved a sensitivity of only 34.4% (29.2, 39.7). However, when using a predicted probability cut-off ≥0.10, this model's sensitivity increased to 60.5% (48.3, 72.7) making it the best PFT screening tool for SSc-ILD. On the other hand, the joint TLC % predicted and ratio of forced expiratory volume in one second (FEV_1) to FVC (FEV_1/FVC) model had the greatest specificity [88.8 %(85.5, 92.1)] and lowest error rate [24.9% (22.5, 27.4)], followed closely by DLCO % predicted and TLC absolute suggesting that these measures may be good overall surrogate markers for the determination of ILD status in epidemiologic studies.

Overall, this thesis offers two major contributions. From a substantive standpoint, it demonstrates that, despite FVC's popularity, FVC should not be considered as a stand-alone marker for SSc-ILD onset. It should at the very least be considered in conjunction with DLCO and TLC. In particular, we have shown that the latter two PFT measures can determine ILD status with greater accuracy, but if one is interested in maximizing the number of ILD cases captured by PFT screening then all three of

these PFT measures should be used. Although this may be more intuitive for respirologists, other physicians caring for SSc patients, including rheumatologists and internists, should integrate this important information into their clinical practice. From a methodologic standpoint, this thesis has shown that, when using surrogate markers or biomarkers, HMMs can be better suited to detect disease status than the use of fixed cut-offs or of pre-specified changes in marker values. These models can easily be implemented in epidemiologic studies using existing software packages and, if developed into a user-friendly web-application, could also assist clinicians in objectively screening for the presence of disease.

7.2 Limitations and Challenges

7.2.1 Systemic Sclerosis-Associated Interstitial Lung Disease Onset versus Progression

With the exception of the first study, this thesis focused specifically on SSc-ILD onset. We had initially set out to identify the best PFT surrogate markers for both SSc-ILD onset and progression, that is the transitioning from no to mild, moderate and finally severe ILD. Such a study would require knowledge of not only ILD presence, but also of ILD severity. This is most often assessed using one of a variety of different HRCT scoring methods which quantify the extent of lung involvement by each proposing different ways of sectioning the lungs into important anatomical locations and assessing the severity of parenchymal abnormalities in these regions.[120, 238, 316] We sought on multiple occasions to secure funds to retrieve and score existing HRCT scans of Canadian Scleroderma Research Group (CSRG) patients, but unfortunately we were ultimately unsuccessful. Consequently, this thesis focused solely on SSc-ILD onset. The development of ILD could however be viewed as a form of progression in and of itself meaning that the results of this thesis do provide some insight on which PFT measures to use to monitor decline in lung function.

7.2.2 Measurement of Systemic Sclerosis-Associated Interstitial Lung Disease Presence

Given that HRCT scans were not routinely performed on CSRG patients, yearly ascertainment of their ILD status was not straightforward. In the absence of HRCT results, we chose to define ILD presence based on evidence of fibrosis on chest X-ray or of crackles heard during physical examination, both of which are recorded annually. This method of ILD ascertainment is not perfect and will inevitably result in some form of information bias. In particular, the lower sensitivities of chest radiography and physical examinations compared to HRCT scans may translate into the missed identification of SSc patients with less extensive ILD, including patients with early disease.[29] This subgroup of patients has the greatest risk of SSc-ILD progression.[46] For this reason, this algorithm is not (nor should it be) a standard for diagnosing SSc-ILD in the clinical setting. In fact, this often involves multi-disciplinary and expert discussion. However, in the context of clinical research, specificity is often preferred over sensitivity when studying low-prevalence diseases in order to minimize false-positives and increase the accuracy of patient classification by ILD status. The algorithm's larger specificity allows for just that. Furthermore, it circumvents the potentially more problematic verification bias which would occur if only using available HRCT results. While it is difficult to quantify the effect of either of these two biases, the use of chest X-ray and lung auscultation results has been found to have good specificity and fair sensitivity upon both internal and external validation.[29, 58]

We also chose to assume that the absence of a combination gold-standard result meant that no ILD was present. We felt this was a reasonable assumption to make given that patients suspected to have pulmonary complications are more likely to undergo testing to confirm ILD presence. Furthermore, this decision enabled us to increase the power of our model given that we were not discarding approximately 10% of visits due to a missing ILD status.

7.2.3 Selection of Included Patients

Both the second and third studies of this thesis required eligible CSRG patients to have at least three annual study visits. This inclusion criterion was necessary to run the HMMs while adjusting for baseline PFT measurements. Indeed, for the HMM to formulate the likelihood of the data given an underlying Markov process, at least two marker measurements per subject are required. Since the first study visit was removed from the analyses such that baseline PFT measurements could be accounted for, it was fundamental that each patient have at least three visits. It follows that this may have introduced selection bias into our sample by potentially removing patients with rapidly progressive disease. However, these patients may already have been excluded from the study by virtue of already having ILD at their baseline visit. If patients who rapidly progress from no to severe ILD were included in the studies, they would have better represented this notion of distinct states for ILD absence and presence, which is favoured by the HMM. Therefore, the exclusion of these patients may imply that our second and third studies likely provide a lower limit to the performance metrics of the HMM. While other hierarchical modeling approaches, such as linear mixed models and generalized estimating equations, could be performed without requiring a minimum of three study visits, these models would not account for measurement error in the PFTs and would not provide a probability of ILD presence.

An additional limitation related to the population under study is that it is a cohort comprised of well-established cases of SSc, with patients having a mean disease duration of approximately 10 years at baseline (Table 3-2). Yet ILD often occurs early in the course of SSc, generally within the first year or two following disease onset.[26, 86] This precludes us from generalizing our results in the last study to early cases of SSc. To establish how the PFT measures perform in early SSc, a preliminary sensitivity analysis was conducted using the same statistical approach on a subgroup of CSRG patients with a disease duration at baseline of fewer than three years. Unfortunately, the results were inconclusive as

the smaller sample size resulted in larger and overlapping confidence intervals. However, based on the point estimates, the joint FVC/DLCO/TLC model had the greatest overall sensitivity, while FVC % predicted had the highest specificity and lowest overall error rate with TLC absolute not far behind.

On a similar and more general note, the second and third studies of this thesis were limited by their small sample sizes. The small number of study subjects potentially precluded a definitive conclusion about the best PFT surrogate marker for SSc-ILD. Indeed, the 95% confidence intervals surrounding the diagnostic test accuracy point estimates were moderately-sized and resulted in substantial overlap between the different PFT measures investigated. However, circumventing small sample size is difficult when studying rare diseases such as SSc. Furthermore, the CSRG has a world-class registry accounting for approximately 10% of all SSc patients in Canada.

7.2.4 Hidden Markov Model Assumptions

Important assumptions are made when using an HMM and are worth emphasizing. First, it assumes the correct specification of the underlying probability distributions for the true surrogate marker levels. In other words, that PFT measures are indeed normally distributed in both the absence and presence of SSc-ILD. Since all PFT measures considered in this thesis were reasonably symmetrical, we do not believe this assumption to be violated.

An additional limitation of the HMM is that it works best if the disease states under consideration are clearly distinct from one another. In the case of SSc-ILD, this assumes that a clear distinction can be made between absence and presence of disease. It is however possible that ILD classification is not clear nor straightforward for SSc patients as ILD may present as a continuous spectrum of disease. If this clear distinction between SSc-ILD absence and presence were to be violated, the HMM would likely not perform at its full potential. Nevertheless, its ability to use longitudinal data enabled it to outperform common definitions of SSc-ILD presence.

Finally, the stationarity assumption, or assumption that transition intensities are constant over time may not always hold in populations of SSc since ILD is most likely to occur early in the course of the disease. This may not be problematic in our particular patient population given their longer disease duration, but future studies focusing on early SSc should consider increasing the number of states in the Markov process to account for the potential of a varying transition intensity.

7.3 Generalizability of Findings

The CSRG data used in the last two studies of this thesis was collected at 15 study sites spanning much of the Canadian territory. Given the spread and size of this registry, the information it contains is likely to be representative of most Canadian SSc patients. However, it is perhaps limited in its representation of populations living in Northern Canada which are known to have higher rates of SSc than non-North American native populations.[317] On a global scale, we surmise that our results are generalizable to most of the North American and European population of SSc patients.

As previously mentioned, the results of this thesis apply to the onset of SSc-ILD and cannot be generalized to SSc-ILD progression. It is nevertheless possible that TLC and DLCO may also be the best PFT surrogate markers for SSc-ILD extent, yet this would require confirmation through further testing.

Finally, the eligibility criteria were such that SSc patients progressing rapidly from no to severe ILD may have been omitted from the last two studies. The exclusion of this patient population impedes the ability to definitively state whether the PFT measures investigated can capture a quick transition to severe SSc-ILD. Similarly, the use of the combination gold-standard may have hampered the evaluation of the PFT measure's capabilities of detecting early forms of SSc-ILD.

7.4 Implications of Findings

The results of this thesis can have a positive impact on the screening of SSc patients for ILD in clinical practice as well as on the way in which ILD status is ascertained in future research on this topic.

7.4.1 Screening and Monitoring for SSc-ILD

There is currently no officially recognized method to screen for SSc-ILD onset. While some experts recommend performing a baseline HRCT scan on all SSc patients, there is no accepted protocol for the subsequent monitoring of patients without ILD initially.[318] In many instances, treating rheumatologists look for declining trends in PFT measures, particularly in FVC. Our results have corroborated this practice and shown the importance of also following DLCO and TLC. We now suggest that the use of a composite FVC/DLCO/TLC outcome in an HMM provides an objective manner of achieving what clinicians attempt to evaluate by gestalt. Indeed, HMMs will statistically assess whether worsening trends in these PFT measures are indicative of ILD presence while simultaneously correcting for random variability associated with pulmonary function testing. Given the importance of beginning treatment for SSc-ILD as soon as possible, screening measures for lung disease should aim to have a high sensitivity. As such, decreasing the HMM-predicted probability threshold will help in achieving this.

We support the creation of a web-based application that could easily be used by treating clinicians to monitor the probability of SSc-ILD onset given their patient's entire history of FVC, TLC and DLCO measurements. Such an application would only require the input of all past PFT values and the dates on which they were measured. A resulting current probability of having ILD would subsequently be generated empowering physicians through evidence-based decision-making. For example, this probability could then be used to decided about reordering HRCTs in patients previously deemed not to have ILD.

Should SSc clinics choose to proceed with the use of HMMs as a screening tool for ILD in their practice, we also suggest increasing the frequency at which patients perform PFTs. Presently, CSRG patients undergo pulmonary function testing once a year. More frequent testing would translate into a more powerful HMM and the ability to detect potential SSc-ILD incidence earlier. An optimal delay of three to six months between pulmonary function testing sessions has been suggested in SSc patients but remains to be determined.[287]

7.4.2 Determination of SSc-ILD Status in Epidemiologic Studies

The first study of this thesis demonstrated that FVC % predicted is demonstrably the most commonly used outcome for SSc-ILD extent in epidemiologic studies. Often, FVC is used along with a <80% cut-off or 10% decline between consecutive measurements to indicate ILD presence. This designation is in turn used as an outcome in studies which seek to identify predictors of SSc-ILD onset. However, we found in our third study that FVC is likely not the best PFT surrogate marker for SSc-ILD onset. Additionally, we showed in our second study that HMMs may be more suitable to assess disease onset than the use of clean cut-offs or pre-specified changes in marker levels.

Considering these results, we suggest that future studies should screen patients without ILD using HMMs. Furthermore, these models should consider using PFT measures such as TLC and DLCO % predicted rather than FVC. While our third study stopped short of definitively proposing the best PFT surrogate marker for SSc-ILD onset between TLC and FEV_1/FVC, TLC absolute, and DLCO % predicted, it did show that solely using FVC % predicted is suboptimal. This is important as the use of a more accurate PFT surrogate marker for SSc-ILD onset will enhance the validity of studies on this topic.

7.5 Suggestions for Future Research

Given that PFT measures may not be sensitive to the detection of SSc-ILD, future work should investigate other possible avenues for SSc-ILD surrogate markers. This could include creating composite endpoints using PFT measures and patient-reported outcomes.[220] Or focus could be shifted to new potential surrogate markers for SSc-ILD, including serum biomarkers. Research in this field has already begun with some studies showing promising results for lung epithelial-derived surfactant protein D (SP-D) and others.[44, 313-315]

Future research in this area should also evaluate whether results similar to those presented in this thesis are obtained when studying SSc-ILD progression. This could easily be done by increasing the number of states in the HMM. It would however require the availability or procurement of repeat HRCT scans on SSc patients, as well as quantitative scoring of lung disease involvement visible on these scans. Identifying a good surrogate marker for SSc-ILD progression is of the utmost importance as it would allow for the identification of predictors of rapid decline in lung function. This would allow clinicians to readily flag at-risk patients who would in turn benefit most treatment.

Finally, it may be interesting to consider the use of certain statistical methods to deal with the imperfect nature of the combination gold-standard. Indeed, estimates of accuracy resulting from incorrect misclassification of disease status could be corrected based on external evidence of the extent of imperfection of the reference standard or with statistical techniques such as, for example, latent class analysis.[324]

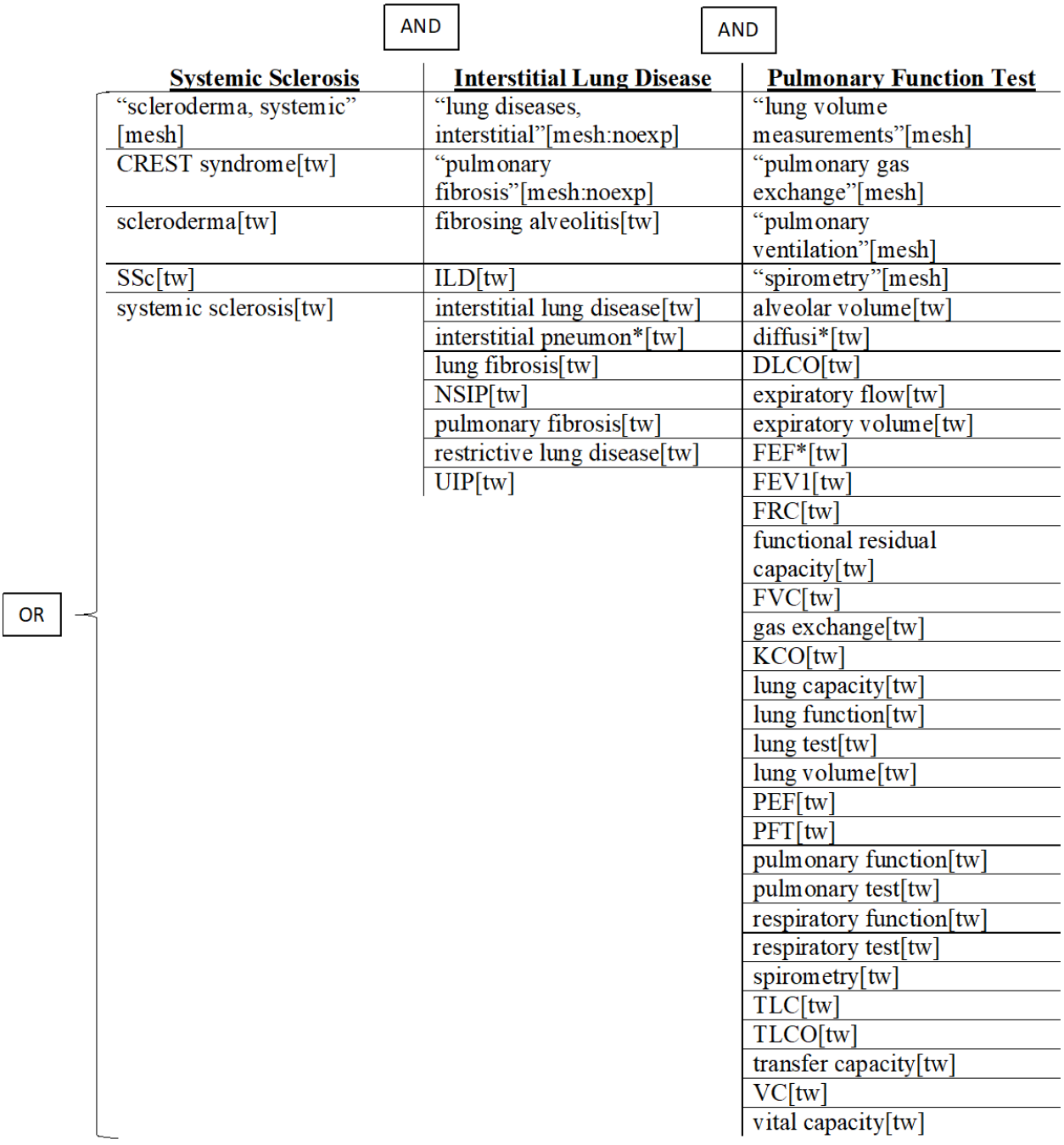
7.6 Conclusion

This thesis showed, through a comprehensive review of the literature, that the current status quo of preferring FVC as a surrogate marker for SSc-ILD onset and progression is unjustified given the available evidence. Furthermore, the systematic review identified methodologic limitations in existing

PFT validation studies which often only made use of cross-sectional data and basic definitions of SSc-ILD presence. This led to the evaluation and proposal of HMMs as a valuable mean of using patients' entire history of PFT measurements to detect SSc-ILD onset. Finally, validation of different univariate and multivariate HMMs using longitudinal measurements of FVC, DLCO and/or TLC indicated that FVC alone was not the best PFT surrogate marker for SSc-ILD. Indeed, DLCO and TLC were found to more accurately determine SSc-ILD status overall, while a joint FVC/DLCO/TLC HMM detected the most cases of SSc-ILD.

The results of this thesis contribute to the limited body of evidence on the validation of PFT measures in this context. Furthermore, they can act as a gateway to the use of HMMs in the realm of SSc-ILD and beyond as a tool to more accurately study disease onset using surrogate markers or biomarkers. Finally, these findings can be used to empower treating rheumatologists and researchers working towards the common goal of better managing ILD in SSc patients.

Appendix A: Detailed MEDLINE (PubMed) Search Strategy



Screens: 1949 – Present

Appendix B: Reasons for Study Exclusion by Screening Stage

Primary Screen: Titles and Abstracts (1,415 Excluded Records)		Secondary Screen: Full-Text Assessment (384 Excluded Records)	
-Duplicate	25	-Duplicate	4
-Not in English or in French	87	-Not in English or in French	6
-Not Original Research	467	-Not Original Research	18
-Unrelated to SSc-ILD	654	-Unrelated to/Unclear if SSc-ILD	25
-Did Not Validate or Use PFTs as Outcomes in SSc-ILD	51	-Did Not Validate or Use PFTs as Outcomes in SSc-ILD	86
		-Unclear which PFTs Used	33
-Did Not Have a Minimum of 20 SSc Patients at Baseline	131	-Did Not Have a Minimum of 20 SSc Patients at Baseline	11
		-Did Not Focus Primarily on SSc-ILD	56
		-Other Reason	11
		-Cross-Sectional Outcome Study	134

Abbreviations: ILD = Interstitial Lung Disease; PFT = Pulmonary Function Test; SSc = Systemic Sclerosis

Appendix C: Characteristics of the Outcome Studies (N = 169)

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Schneider et al. 1982[30]	United States	Cohort Study (Hospital Discharge Records)	38 (74% Female)	44 (17-65)	DLCO Absolute, FEV_1 Absolute, FEV_1/FVC, FVC Absolute	N/A	N/A
Konig et al. 1984[73]	Germany	Cohort Study	101	N/R	DLCO%, TLC%, VC%	DLCO%	Most Sensitive PFT
Peters-Golden et al. 1984[74]	United States	Cohort Study (Rheumatology Unit)	24 (92% Female)	46.1 ± 2.61 (19-67)	DLCO Absolute, FEV_1/FVC, FVC Absolute, FVC%, TLC Absolute, TLC%	N/A	N/A
Steen et al. 1985[75]	United States	Cohort Study (Hospital Records)	92 (73% Female)	N/R	DLCO%, FEV_1/FVC, FVC%	N/A	N/A
De Clerck et al. 1987[76]	Belgium	Cohort Study	23	47.6 ± 10.3 (28-63)	DLCOcorr%, DLCOcorr/LV%, FEV_1/FVC, TLC%	DLCO%	Most Sensitive PFT

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Greenwald et al. 1987[77]	United States	Cohort Study (Chlorambucil Clinical Trial)	61 (87% Female)	47 ± 12	DLCO Absolute, DLCO%, $FEF_{25-75\%}$ Absolute, $FEF_{25-75\%}$%, FEV_1 Absolute, FEV_1%, FEV_1/FVC, FRC Absolute, FRC%, FVC Absolute, FVC%, TLC Absolute, TLC%	N/A	N/A
McCarthy et al. 1988[78]	Canada	Cohort Study (Rheumatic Disease Unit)	36 (75% Female)	48.5 ± 11 (23-66)	FVC Absolute	FVC Absolute	N/R
Zarafonetis et al. 1989[79]	United States	Cohort Study (Hospital Records)	390	N/R	DLCO%, FVC%	N/A	N/A
Silver et al. 1990[80]	United States	Cohort Study (General Clinical Research Center)	43 (60% Female)	43.9 ± 11.6 (21-63.4)	DLCO Absolute, FVC Absolute	N/A	N/A
Abramson et al. 1991[81]	Australia	Cohort Study (Clinical Notes and Lung Function Records)	113 (81% Female)	50.6 ± 14 (16-81)	FEV_1 Absolute, VC Absolute	N/A	N/A
Wells et al. 1993[82]	United Kingdom	Cohort Study (ILD Unit)	66 (76% Female)	50.1 ± 12.0	DLCO%, FVC%	N/A	N/A
Wells et al. 1993[83]	United Kingdom	Cohort Study (Hospital Records)	53 (74% Female)	49 ± 12	DLCO Absolute, FVC Absolute	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Wells et al. 1993[84]	United Kingdom	Cohort Study (Hospital Records)	35	N/R	DLCO%, FVC%	N/A	N/A
Dujic et al. 1994[85]	Croatia	Cohort Study (Department of Dermatology)	29 (86% Female)	51.5 ± 12.7 (27-75)	DLCO%	DLCO%	Previous Use
Steen et al. 1994[86]	United States	Cohort Study (Division of Rheumatology and Clinical Immunology)	890 (92% Female)	42	FVC Absolute, FVC%	FVC	Previous Use
Steen et al. 1994[87]	United States	Cohort Study (Hospital Records)	122	45	DLCO%, FVC Absolute, FVC%	FVC	N/R
Tashkin et al. 1994[88]	United States	Cohort Study (Chlorambucil Clinical Trial)	90	47 ± 11	DLCO Absolute, DLCO%, FEV_1 Absolute, FVC Absolute, TLC Absolute	N/A	N/A
Behr et al. 1995[89]	Germany	Case-Control Study	43 (65% Female)	54.3 ± 3.0 (15-71)	DLCO%, TLC%, VC%	N/A	N/A
Behr et al. 1996[90]	Germany	Cohort Study (Department of Internal Medicine)	79 (67% Female)	50.4 ± 1.2	DLCO%, VC%	N/A	N/A
Jacobsen et al. 1997[91]	Denmark	Cohort Study (Participating Clinical Centres Chart Records)	176 (85% Female)	41 (4-74)	DLCO%, DLCO/VA%, FEV_1/VC%, VC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Greidinger et al. 1998[92]	United States	Cohort Study (Scleroderma Center)	101 (75% Female)	49.5 ± 13.5	DLCO%, FEV_1%, FVC%	N/A	N/A
Atamas et al. 1999[93]	United States	Case-Control Study (Scleroderma Center)	37 (68% Female)	44.4 ± 13.0 (18-69)	DLCO Absolute, DLCO%, FVC Absolute, FVC%	N/A	N/A
Kon et al. 1999[94]	United Kingdom	Case-Control Study	37 (89% Female)	49.6 ± 11.6 (24-79)	FVC%	FVC%	N/R
Witt et al. 1999[95]	Germany	Cohort Study (Pneumological Outpatient Clinic)	73 (78% Female)	54.4 ± 9.6 (20–80)	DLCO%, FVC%, TLC%	DLCO%	Validation
White et al. 2000[96]	United States	Cohort Study (Scleroderma Center)	103 (69% Female)	48 (30-59)	DLCO Absolute, DLCO%, FVC Absolute, FVC%	N/A	N/A
Yuhara et al. 2000[97]	Japan	Cohort Study (Hospital Records)	24 (96% Female)	36.7 ± 10.6	DLCO%, DLCO/VA%, FVC%	N/A	N/A
Marie et al. 2001[98]	France	Cohort Study (Hospital Records)	43 (86% Female)	59 (33-79)	DLCO%, FEV_1%, FEV_1/VC%, FRC%, FVC%, RV%, TLC% VC%	N/A	N/A
Scorza et al. 2001[99]	Italy	Experimental Study (Outpatient Clinic)	46 (85% Female)	53 (25-75)	DLCO%, FEV_1%, VC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Bouros et al. 2002[100]	United Kingdom	Cohort Study (Hospital Records)	80	N/R	DLCO%, FVC%	N/A	N/A
Giacomelli et al. 2002[101]	Italy	Cohort Study (Outpatient Clinic Centers)	23 (83% Female)	57.3 (39-67)	DLCO%, FEV_1%, FVC%	N/A	N/A
Pakas et al. 2002[102]	Greece	Experimental Study (Rheumatology Outpatient Clinic)	28 (82% Female)	48.3	DLCO%, FVC%, TLC%	N/A	N/A
Kowal-Bielecka et al. 2003[103]	Poland	Case-Control Study	30 (100% Female)	46 (24-62)	FVC%	FVC%	N/R
Yanaba et al. 2003[104]	Japan	Case-Control Study	39 (85% Female)	49 (2-72)	DLCO%, VC%	N/A	N/A
Airo et al. 2004[105]	Italy, United Kingdom	Individual Patient Data Meta-Analysis	53	N/R	DLCO%, FVC%	N/A	N/A
Yanaba et al. 2004[106]	Japan	Case-Control Study	42 (86% Female)	49 ± 18	DLCO%, VC%	N/A	N/A
De Santis et al. 2005[107]	Italy	Cohort Study (Outpatient Clinic of the Division of Rheumatology)	100 (92% Female)	55.4 ± 11.9	DLCO%, FVC%	N/A	N/A
Kodera et al. 2005[108]	Japan	Case-Control Study	123 (86% Female)	51 ± 14	DLCO%, VC%	N/A	N/A
Kowal-Bielecka et al. 2005[109]	Poland	Cohort Study	21 (100% Female)	52 (25–66)	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Hoyles et al. 2006[32]	United Kingdom	Experimental Study (Fibrosing Alveolitis in Scleroderma Trial (FAST))	45 (71% Female)	55 (18-75)	DLCO% DLCO/VA%, DLCOcorr%, FEV_1%, FVC%, TLC%	N/A	N/A
Plastiras et al. 2006[110]	Greece	Cohort Study (Outpatient University Rheumatology Clinic)	78 (85% Female)	45.9 ± 13.5	DLCO%, FVC%	FVC%	Previous Use
Tashkin et al. 2006[31]	United States	Experimental Study (Scleroderma Lung Study I (SLS I))	158 (70% Female)	47.9 ± 1.0 (19.6-83.1)	DLCO%, DLCO/VA%, FVC%, TLC%	FVC%	Previous Use
Beretta et al. 2007[111]	Italy	Case-Control Study (Outpatient Clinical Immunology and Allergology Clinic)	204 (91% Female)	48.6 ± 13.2 (16-75)	FVC%	FVC%	N/R
Beretta et al. 2007[112]	Italy	Cohort Study (Outpatient Allergology, Clinical Immunology and Rheumatology Clinic)	33 (79% Female)	49.7 ± 10.4	DLCO Absolute, DLCO%, VC Absolute, VC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Clements et al. 2007[113]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	158 (70% Female)	48 ± 13	DLCO%, FEV_1%, FVC%, TLC%	N/A	N/A
Goh et al. 2007[114]	United Kingdom	Cohort Study (ILD Unit)	141 (81% Female)	47.3 ± 12.2	DLCO%, FVC%	N/A	N/A
Mittoo et al. 2007[115]	United States	Cohort Study (Scleroderma Center)	25 (64% Women)	43.5 ± 12.5 (16-67)	DLCO Absolute, DLCO%, FVC Absolute, FVC%	N/A	N/A
Tashkin et al. 2007[116]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	145 (70% Female)	47.9 ± 1.0	DLCO%, DLCO/VA%, FVC%, TLC%	FVC%	Previous Use
Tzelepis et al. 2007[117]	Greece	Cohort Study (University Rheumatology Clinic)	59 (81% Female)	47.5 ± 13.9	FVC%	FVC%	Previous Use
Berezne et al. 2008[118]	France	Cohort Study (National Reference Centers for Systemic Sclerosis)	27 (74% Female)	49.4 ± 15	DLCO%, FVC%, TLC%	N/A	N/A
Boin et al. 2008[119]	United States	Case-Control Study (Scleroderma Center)	62 (84% Female)	51.1	FVC%	FVC%	N/R
Goh et al. 2008[120]	United Kingdom	Cohort Study (Hospital Records)	215 (81% Female)	49.1 ± 13.0	DLCO%, FVC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Strange et al. 2008[121]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	141 (72% Female)	48.6 ± 12.0	FVC%	FVC%	Previous Use
Assassi et al. 2009[122] (Abstract)	United States	Cohort Study	36	N/R	FVC%	FVC%	N/R
De Souza et al. 2009[123]	Brazil	Cohort Study (Hospital Records)	28 (100% Female)	44.89 ± 8.74	DLCO%, FEV_1%, FVC%	N/A	N/A
Gordon et al. 2009[124] (Abstract)	United States	Experimental Study	30	N/R	DLCO%, FVC%	N/A	N/A
Khanna et al. 2009[125]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	158 (71% Female)	48.5 ± 12.3	FVC%	FVC%	Previous Use
Ottewell et al. 2009[126] (Abstract)	United Kingdom	Cohort Study	22 (91% Female)	56 (31-79)	DLCO%, VC%	N/A	N/A
Schmidt et al. 2009[127]	Germany	Case-Control Study	32 (72% Female)	58.5 (30-72)	DLCO%, FVC%, TLC%	N/A	N/A
Wanchu et al. 2009[128]	India	Cohort Study (Rheumatology Clinic)	36 (94% Female)	37.5 ± 10.5	DLCO%, FVC Absolute, FVC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Assassi et al. 2010[129]	United States	Cohort Study (Genetics versus Environment in Scleroderma Outcome Study (GENISOS))	266 (80% Female)	48.63 ± 13.5	FVC%	FVC%	Validation
Boin et al. 2010[130] (Abstract)	United States	Cohort Study	22	N/R	FVC%	FVC%	N/R
Colaci et al. 2010[131]	Italy	Cohort Study (Rheumatology Unit)	26 (77% Female)	47.8 ± 10.5	DLCOcorr%, FVC%	N/A	N/A
Cuomo et al. 2010[132] (Abstract)	Italy	Cohort Study	20 (90% Female)	46 (18-57)	DLCO%, FVC%	N/A	N/A
Gilson et al. 2010[133]	France	Cohort Study (Department of Rheumatology)	105 (86% Female)	52.7 ± 11.8	DLCO%, FVC%	FVC%	Previous Use
Mittoo et al. 2010[134] (Abstract)	Canada	Cohort Study (Canadian Scleroderma Research Group (CSRG))	67 (88% Female)	54.5 ± 12.1	DLCO%, FVC%	FVC%	N/R
Schorr et al. 2010[135] (Abstract)	United States	Cohort Study (Scleroderma Specialty Center Database)	91	N/R	FVC%	FVC%	N/R
Seibold et al. 2010[136]	United States	Experimental Study	152 (74% Female)	52.5 (15-80)	DLCO%, FVC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Shahane et al. 2010[137] (Abstract)	United States	Cohort Study (Clinic Scleroderma Database)	133	N/R	DLCO, FVC	N/A	N/A
Steen et al. 2010[138] (Abstract)	United States	Cohort Study	1,029	N/R	DLCO%, FVC%	N/A	N/A
Theodore et al. 2010[139]	United States	Cohort Study	24	N/R	FVC	FVC	Previous Use
Abhishek et al. 2011[140]	United Kingdom	Cohort Study (Rheumatology Day-Case Unit Databases)	36 (75% Female)	54.26 ± 14.03	DLCO Absolute, FVC Absolute	N/A	N/A
De Santis et al. 2011[141]	Italy	Case-Control Study (Outpatient Clinic of Rheumatology Division)	46 (78% Female)	55.1 ± 14	DLCO%, FVC%	N/A	N/A
Espinosa et al. 2011[142]	Spain	Cohort Study (Autoimmune Disease and Internal Medicine Departments)	37 (81% Female)	43.0 ± 12.4	DLCO%, FVC%	N/A	N/A
Goh et al. 2011[143]	United Kingdom	Cohort Study	168 (82% Female)	49.5 ± 13.2	DLCO Absolute, DLCO%, FVC Absolute, FVC%	N/A	N/A
Hasegawa et al. 2011[144]	Japan	Case-Control Study	92 (87% Female)	52.3 ± 13.5	DLCO%, VC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Hoshino et al. 2011[145]	Japan	Case-Control Study (Hospital Records)	314 (88% Female)	44.9	FVC%	FVC%	N/R
Jayaweera et al. 2011[146] (Abstract)	Australia	Cohort Study (Australia Scleroderma Interest Group (ASIG))	43	N/R	DLCO%, FVC%	N/A	N/A
Khanna et al. 2011[147]	United States	Experimental Study	20 (65% Female)	46.1 ± 14.2	DLCO%, FVC%, TLC%	N/A	N/A
Khanna et al. 2011[148]	United States	Cohort Study (Scleroderma Lung Study I (SLS I) Placebo Group)	77 (62% Female)	48.3 ± 12.5	DLCO%, FVC%	N/A	N/A
Mittoo et al. 2011[149]	United States	Cohort Study (Scleroderma Center Database)	38 (68% Female)	44.3 ± 11.4 (21-74)	DLCO%, FVC%	N/A	N/A
Poormoghim et al. 2011[150]	Iran	Cohort Study (Rheumatology Clinic)	91 (93% Female)	44.10 ± 14.88	DLCO%, FVC%	N/A	N/A
Rosato et al. 2011[151]	Italy	Cohort Study (Clinical Immunology Unit-Scleroderma Center)	41 (90% Female)	47.5 (23-70)	DLCOcorr%, FEV_1%, TLC%, VC%	DLCOcorr%	N/R
Roth et al. 2011[152]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	112 (71% Female)	46.9 ± 0.9	FVC%	FVC%	Previous Use

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Tiev et al. 2011[153]	France	Cohort Study	83 (88% Female)	53.5 ± 12.2	FVC%, TLC%	N/A	N/A
Volpinari et al. 2011[154]	Italy	Cohort Study	79 (90% Female)	55 ± 13	DLCO%, FVC%	N/A	N/A
Abignano et al. 2012[155] (Abstract)	United Kingdom	Cohort Study (Medical Records)	45	N/R	DLCO%, FVC%	N/A	N/A
De Santis et al. 2012[156]	Italy	Cohort Study (Outpatient Clinic of Rheumatology Division)	110 (87% Female)	54.9 ± 12.6	DLCO%, FVC%	N/A	N/A
Hesselstrand et al. 2012[157]	Sweden	Cohort Study (SSc Cohort)	244	N/R	VC%	VC%	N/R
Kishore Babu et al. 2012[158] (Abstract)	India	Cohort Study (Discharge Summaries)	23 (78% Female)	35.9 ± 4.5	FVC%	FVC%	N/R
Kuwana et al. 2012[159] (Abstract)	Japan	Cohort Study (Institutional SSc Database)	50	N/R	FVC%	FVC%	N/R
Kuwana et al. 2012[160] (Abstract)	Japan	Cohort Study (Institutional SSc Database)	50	N/R	FVC%	FVC%	N/R
Le Gouellec et al. 2012[161] (Abstract)	France	Cohort Study	75	N/R	DLCO%, FVC%	N/A	N/A
Schupp et al. 2012[162] (Abstract)	Germany	Cohort Study	126	N/R	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Sfriso et al. 2012[163]	Italy	Case-Control Study (Rheumatology Unit)	32 (100% Female)	55.1 ± 9.2 (45.6-67.4)	DLCO%, FVC%	N/A	N/A
Soriano et al. 2012[164] (Abstract)	Italy	Cohort Study	31	N/R	FEV_1%, TLC%	N/A	N/A
Tiev et al. 2012[165]	France	Cohort Study (Department of Internal Medicine)	105 (88% Female)	54.8 ± 12.9	FVC%, TLC%	N/A	N/A
Ananyeva et al. 2013[166] (Abstract)	Russia	Cohort Study	27 (96% Female)	45	DLCO%, FVC%	N/A	N/A
Ando et al. 2013[167]	Japan	Cohort Study (Department of Respiratory Medicine at Tertiary Care Center)	71 (82% Female)	58.2 ± 13.9	FVC%	FVC%	Previous Use
Burt et al. 2013[168]	Brazil, United States	Cohort Study (Previous Study or Compassionate Basis)	90 (81% Female)	42 (16-71)	DLCOcorr%, FVC%, TLC%	N/A	N/A
Celeste et al. 2013[169]	Italy	Cohort Study (Outpatient Clinic Referral Center for Systemic Autoimmune Diseases)	221 (90% Female)	45.5	DLCO%, FVC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
De Lauretis et al. 2013[170]	United Kingdom	Case-Control Study (Interstitial Lung Disease Unit)	286 (78% Female)	51.0	DLCO%, FVC%	N/A	N/A
Elhaj et al. 2013[171]	United States	Case-Control Study (Genetics versus Environment in Scleroderma Outcome Study (GENISOS))	266 (83% Female)	48.6 ± 13.5	DLCOcorr%, FVC%	FVC%	Validation
Enghelmayer et al. 2013[172] (Abstract)	Argentina	Cohort Study	24	N/R	DLCO Absolute, FVC Absolute, FVC%	N/A	N/A
Koneva et al. 2013[173] (Abstract)	Russia	Cohort Study (Institute of Rheumatology)	44 (93% Female)	49 ± 13	DLCO%, FVC%	N/A	N/A
Liu et al. 2013[174]	United States	Case-Control Study (Genetics versus Environment in Scleroderma Outcome Study (GENISOS))	266 (83% Female)	48.6 ± 13.5	FVC%	FVC%	Validation
Panopoulos et al. 2013[175]	Greece	Case-Control Study (Department of Therapeutics)	26 (92% Female)	47.1	DLCO%, FVC%, TLC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Radic et al. 2013[176] (Abstract)	Germany	Cohort Study	153	N/R	DLCO%, FVC%	N/A	N/A
Stock et al. 2013[177]	United Kingdom	Case-Control Study (Tertiary Referral Centre Clinics)	440 (81% Female)	52.8 (15-83)	DLCO%, FVC%	N/A	N/A
Vacca et al. 2013[178] (Abstract)	Italy	Cohort Study (Rheumatology Unit)	22	N/R	DLCO%, FVC%	N/A	N/A
Wu et al. 2013[179] (Abstract)	United States	Cohort Study (Genetics versus Environment in Scleroderma Outcome Study (GENISOS))	266	N/R	FVC%	FVC%	N/R
Zhang et al. 2013[180]	Canada	Cohort Study (Canadian Scleroderma Research Group (CSRG))	1,043 (86% Female)	55.74 ± 11.88	FVC%	FVC%	Previous Use
Ananyeva et al. 2014[181] (Abstract)	Russia	Cohort Study (Rheumatology Clinic Lung Study Program)	77 (94% Female)	38	DLCO%, FVC%, FVC%/DLCO %	N/A	N/A
Chakr et al. 2014[182] (Abstract)	Brazil	Cohort Study (SSc Clinic)	28 (86% Female)	49.7 ± 14.2	DLCO%, FEV_1%, FVC%	N/A	N/A
Christmann et al. 2014[183]	Brazil, United States	Cohort Study	28	N/R	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Cottrell et al. 2014[184]	United States	Cohort Study (Scleroderma Center)	2,205 (83% Female)	46.2 ± 13.6	FVC%	FVC%	N/R
Fraticelli et al. 2014[185]	Italy	Experimental Study	30 (70% Female)	51 (41.75 - 62)	DLCO Absolute, FVC Absolute	N/A	N/A
Guillen-Del Castillo et al. 2014[186]	Spain	Cohort Study (Hospital Records)	63 (86% Female)	43.0 (33.0-54.0)	DLCO/VA%, FVC%	FVC%	N/A
Hoffmann-Vold et al. 2014[187] (Abstract)	Norway	Cohort Study (Norwegian Systemic Connective Tissue Disease and Vasculitis Registry (NOSVAR))	305 (79% Female)	48.0	DLCO%, FVC%	N/A	N/A
Kumanovics et al. 2014[188]	Hungary	Cohort Study (Tertiary Care Centre)	173 (89% Female)	57.6 ± 11.3	DLCO%, FVC%	FVC%	Most Specific PFT
Kwon et al. 2014[189] (Abstract)	South Korea	Cohort Study (Rheumatology Clinic)	32 (84% Female)	47.5 ± 9.4	FVC%	FVC%	N/R
Lambrecht et al. 2014[190]	Belgium	Case-Control Study (Scleroderma Clinic)	119	N/R	DLCO%	DLCO%	N/R
Le Gouellec et al. 2014[191] (Abstract)	France	Cohort Study	75 (76% Female)	N/R	DLCO%, FVC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Narvaez et al. 2014[192] (Abstract)	Spain	Cohort Study (Hospital Recruitment)	30 (87% Female)	54	DLCO%, FVC%, TLC%	N/A	N/A
Nihtyanova et al. 2014[193]	United Kingdom	Cohort Study (Tertiary Referral Center)	398 (86% Female)	41	DLCO%, FVC%	N/A	N/A
Parida et al. 2014[194] (Abstract)	India	Experimental Study	30	N/R	DLCO%, FVC%, TLC%	FVC%	N/R
Pham et al. 2014[195] (Abstract)	United States	Cohort Study	20	N/R	DLCO%, FVC%	N/A	N/A
Poormoghim et al. 2014[196]	Iran	Cohort Study (Hospital SSc Database)	36 (83% Female)	N/R Azathioprine Group: 35.0 (30.1–45.0); Cyclophospamide Group: 33.0 (29.0–40.5)	DLCOcorr%, FVC%	N/A	N/A
Rotondo et al. 2014[197] (Abstract)	Italy	Cohort Study	70 (90% Female)	59.7 ± 4.5	DLCO%, FVC%, RV%, TLC%	RV%	N/R
Ariani et al. 2015[198] (Abstract)	Italy	Cohort Study (Multi-Centre Study)	149	N/R	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Balbir-Gurman et al. 2015[199]	Israel	Cohort Study (Tertiary Care Rheumatology Unit)	26 (77% Female)	50.7 ± 12.7	DLCO%, FVC%	N/A	N/A
Bosello et al. 2015[200]	Italy	Experimental Study	20 (85% Female)	41.4 ± 13.1	DLCO%, FEV_1%, FVC%, TLC%	N/A	N/A
De Luca et al. 2015[201]	Italy	Case-Control Study (Rheumatology Inpatient Clinic)	120	N/R	DLCO%, FEV_1% FVC%	N/A	N/A
Hoffmann-Vold et al. 2015[202]	Norway	Cohort Study (Norwegian Systemic Connective Tissue Disease and Vasculitis Registry (NOSVAR))	305 (79% Female)	48 ± 15.0	DLCO%, FVC%	FVC%	Previous Use
Iudici et al. 2015[203]	Italy	Cohort Study (Rheumatology Unit)	45 (91% Female)	49.86 ± 13.33	DLCOcorr%, FVC%	N/A	N/A
Jordan et al. 2015[204]	Switzerland	Case-Control Study (European Scleroderma Trial and Research (EUSTAR) Centres)	63 (71% Female)	50.9 ± 1.6	DLCO%, FVC%	FVC%	Previous Use

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Khanna et al. 2015[205] (Abstract)	United States	Experimental Study (LOTUSS Study)	63 (83% Female)	50.6 ± 12.3	DLCO%, FVC%	N/A	N/A
Khanna et al. 2015[206]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	93 (73% Female)	47.19 ± 11.72	DLCO%, FVC%	N/A	N/A
Koneva et al. 2015[207] (Abstract)	Russia	Cohort Study	54 (81% Female)	48.5 ± 12.9	DLCO%, FVC%	N/A	N/A
Lepri et al. 2015[208] (Abstract)	Australia, France, Italy, Spain, Switzerland	Cohort Study (Multi-Centre Study)	23	N/R	DLCO%, FVC%	FVC%	N/R
Man et al. 2015[49]	United States	Cohort Study (SSc Referral Centre)	254 (80% Female)	49 ± 13	FVC%	FVC%	Validation
Mani et al. 2015[209] (Abstract)	India	Experimental Study (Tertiary Care Hospital)	62	N/R	FVC%	FVC%	N/R
Mateos-Toledo et al. 2015[210] (Abstract)	Mexico	Cohort Study	46	N/R	DLCO%, FVC Absolute, FVC%	N/A	N/A
Narvaez et al. 2015[211] (Abstract)	Spain	Cohort Study	31	59	DLCO%, FVC%, TLC%	N/A	N/A

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Ninaber et al. 2015[212]	The Netherlands	Cohort Study (Referrals to Tertiary Outpatient Targeted Multidisciplinary Healthcare Program)	41 (76% Female)	50.9	DLCO%, FVC%	DLCO%	N/R
Radic et al. 2015[213] (Abstract)	Croatia, Germany, Switzerland	Cohort Study (European Scleroderma Trial and Research (EUSTAR) Database)	124	N/R	DLCO%, FVC%	N/A	N/A
Sakamoto et al. 2015[214]	Japan	Case-Control Study (Hospital Records)	33 (70% Female)	63 (54-70)	VC Absolute	VC Absolute	Previous Use
Saketkoo et al. 2015[215] (Abstract)	United States	Cohort Study (Pulmonary Hypertension Recognition and Outcomes in Scleroderma (PHAROS) Registry)	256	N/R	FVC%	FVC%	N/R
Schulam et al. 2015[216] (Abstract)	United States	Cohort Study	672	N/R	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Shirai et al. 2015[217] (Abstract)	Japan	Cohort Study (SSc Database)	58	N/R	FVC%	FVC%	N/R
Suliman et al. 2015[43]	Switzerland	Cohort Study (Division of Rheumatology)	102 (77% Female)	58.5 (28-90)	FVC%	FVC%	Previous Use
Tanaseanu et al. 2015[218]	Romania	Cohort Study	40 (95% Female)	34 ± 12	DLCO%, FEV_1%	N/A	N/A
Tashkin et al. 2015[219] (Abstract)	United States	Experimental Study (Scleroderma Lung Study II (SLS II))	142	N/R	DLCO%, FVC%	FVC%	N/R
Volkmann et al. 2015[220]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	82 (73% Female)	47.2	FVC%, TLC%	N/A	N/A
Volkmann et al. 2015[221] (Abstract)	United States	Cohort Study (Scleroderma Lung Study II (SLS II))	136	N/R	DLCO%	DLCO%	Previous Use
Wallace et al. 2015[222]	United States	Cohort Study (Combined Response Index in Systemic Sclerosis (CRISS) Database)	177 (75% Female)	50.5 ± 11.7	DLCO%, FVC%, TLC%	N/A	N/A
Fava et al. 2016[223]	United States	Cohort Study (Scleroderma Center)	27 (78% Female)	51.3 ± 9.6	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
Hoffmann-Vold et al. 2016[224]	Norway	Case-Control Study (Norwegian Systemic Connective Tissue Disease and Vasculitis Registry (NOSVAR))	298 (82% Female)	48 ± 15.4	DLCO%, FVC%	FVC%	N/R
Kloth et al. 2016[225]	Germany	Cohort Study (Radiology Department Database)	26 (54% Female)	37.45 ± 9.83 (11-51)	DLCO Absolute, DLCO%, FEV_1 Absolute, FEV_1%, FVC Absolute, FVC%, TLC Absolute, TLC% VC Absolute, VC%	N/A	N/A
Owen et al. 2016[226]	Australia	Cohort Study (Australia Scleroderma Cohort Study (ASCS))	47 (79% Female)	54.6	DLCO Absolute, FVC Absolute	FVC Absolute	Previous Use
Shenoy et al. 2016[227]	India	Cohort Study (Rheumatology Outpatient Department)	57 (86% Female)	45.55	FVC%	FVC%	Most Specific PFT

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
ClinicalTrials.gov ID: NCT00319033[228] (Registration)	Canada, France, Germany, Israel, Italy, Republic of Korea, The Netherlands, Sweden, Switzerland, United Kingdom, United States	Experimental Study (BUILD 2 OL)	132	N/A	DLCO, FVC	N/A	N/A
EudraCT #: 2008-000224-27[229] (Registration)	United Kingdom	Cohort Study	20 (Anticipated)	N/A	DLCO%, FVC%	N/A	N/A
ClinicalTrials.gov ID: NCT01570764[230] (Registration)	France	Experimental Study (SCLEROCYC)	50 (Anticipated)	N/A	DLCO%, FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
ClinicalTrials.gov ID: NCT01559129[231] (Registration)	Australia, France, Germany, Italy, Poland, Russia. Spain, Switzerland, United Kingdom, United States	Experimental Study	23	N/A	FVC Absolute	FVC Absolute	N/R
ClinicalTrials.gov ID: NCT01858259[232] (Registration)	France, Germany, Hungary, Italy, Switzerland, United Kingdom	Cohort Study (DeSScipher)	1,372	N/A	DLCO%, FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
ClinicalTrials.gov ID: NCT02597933[233] (Registration)	Australia, Belgium, Canada, China, Denmark, France, Germany, Greece, India, Ireland, Israel, Italy, Japan, The Netherlands, Poland, Portugal, Spain, Switzerland, United Kingdom, United States	Experimental Study	520 (Anticipated)	N/A	DLCO%, FVC Absolute, FVC%	FVC Absolute	Previous Use
ClinicalTrials.gov ID: NCT02588625[234] (Registration)	Canada, Poland, United Kingdom, United States	Experimental Study	N/R	N/A	FVC%	FVC%	N/R

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	PFT Measure(s) Used	Main PFT Measure	Reason for Using Main PFT
ClinicalTrials.gov ID: NCT02370693[235] (Registration)	United States	Experimental Study	30 (Anticipated)	N/A	FVC	FVC	N/R
ClinicalTrials.gov ID: NCT02745145[236] (Registration)	Argentina, Australia, Canada, Israel, Italy, Poland, Spain, United Kingdom, United States	Experimental Study	175 (Anticipated)	N/A	DLCO%, DLCO/VA, FVC Absolute, FVC%, TLC%	FVC Absolute	N/R

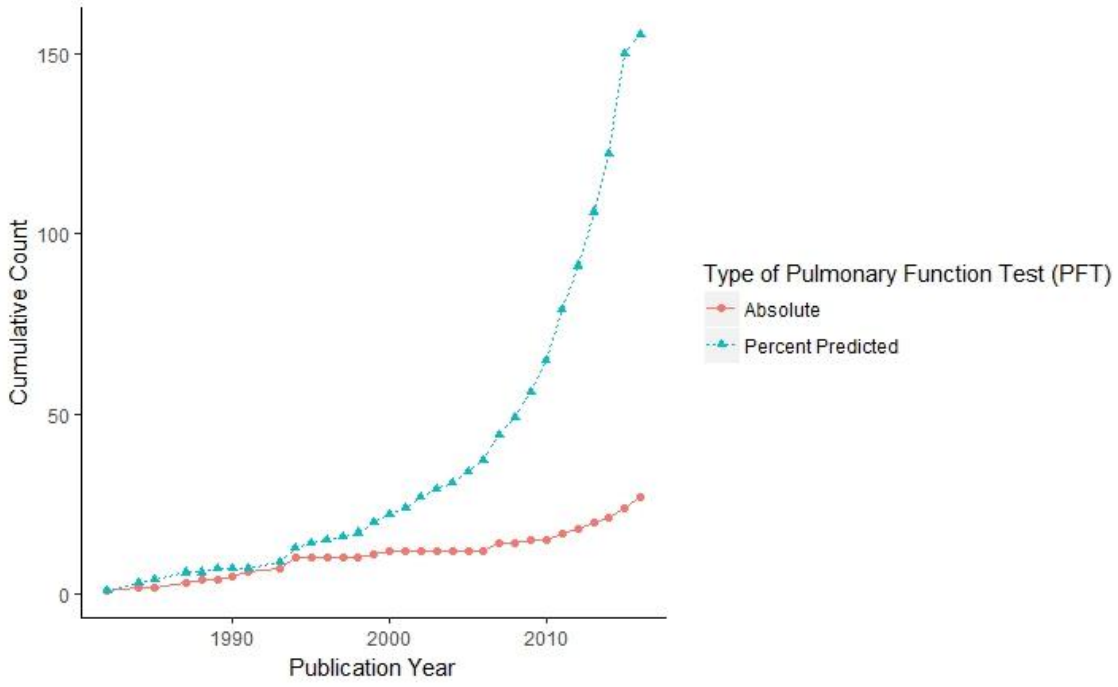
Abbreviations: % = Percent Predicted; DLCO = Diffusing Capacity for Carbon Monoxide; DLCO/VA = Diffusing Capacity for Carbon Monoxide Corrected for Alveolar Volume; DLCOcorr = Diffusing Capacity for Carbon Monoxide Corrected for Haemoglobin; DLCOcorr/LV% = Diffusing Capacity for Carbon Monoxide Corrected for Haemoglobin and Lung Volume; $FEF_{25-75\%}$ = Forced Expiratory Flow over Mid-Half of FVC; FEV_1 = Forced Expiratory Volume in the 1st Second of Forced Exhalation; FVC = Forced Vital Capacity; ILD = Interstitial Lung Disease; N/A = Not Applicable; N/R = Not Reported; PFT = Pulmonary Function Test; RV = Residual Volume; SSc = Systemic Sclerosis; TLC = Total Lung Capacity; VC = Vital Capacity

Appendix D: Cumulative Use and Percent Cumulative Use of Absolute and % Predicted Values

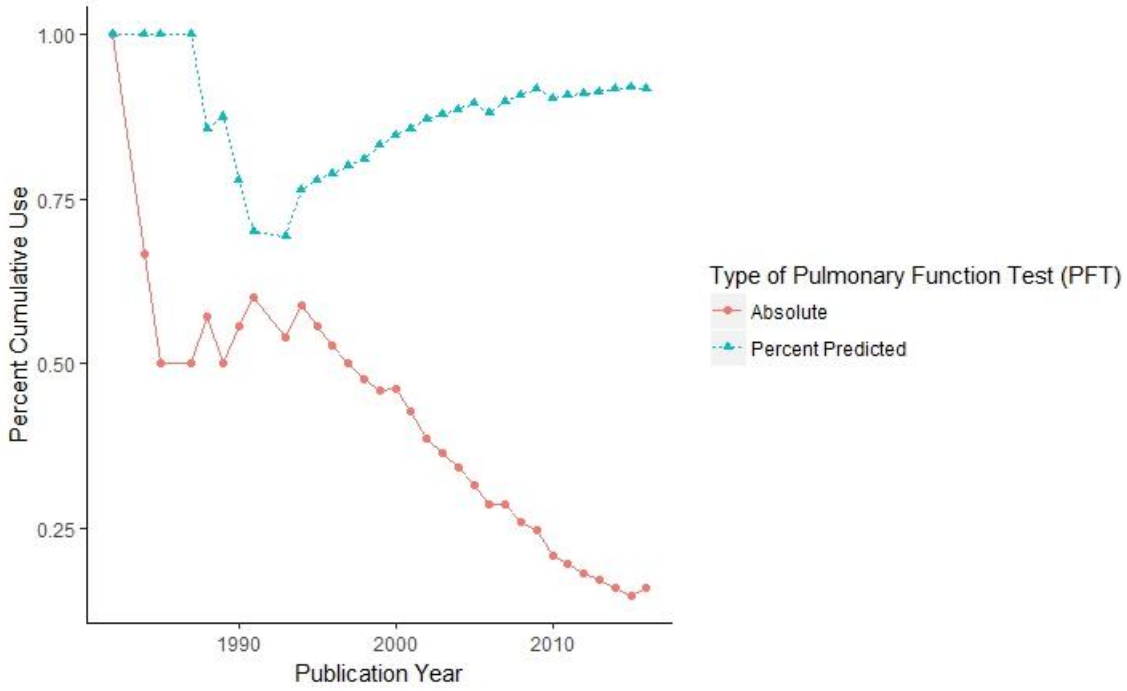
Percent predicted pulmonary function test (PFT) values were cumulatively more commonly used than absolute measures as outcomes for systemic sclerosis-associated interstitial lung disease (SSc-ILD) progression throughout the study period. However, their use was similar to that of absolute PFT values until the mid- 1990's, at which point the cumulative use of % predicted values skyrocketed (Appendix Figure D-1a). By the end of the study period, PFTs were expressed as % predicted values in 155 (93.9%) of the 169 outcome studies, while they were reported as absolute values in 27 (16.4%) of studies. This is further depicted in Appendix Figure D-1b which illustrates the percent cumulative use of absolute and % predicted PFT values. Indeed, while the cumulative percent use of percent predicted PFT values increased steadily as of the mid-1990's, the cumulative percent use of absolute PFT values plummeted in a consistent manner.

A probable explanation for the increasing popularity of % predicted values throughout the study period is the proposal of new PFT standardization and interpretative guidelines. In fact, in a 1987 report, the American Thoracic Society (ATS) focused mainly on absolute values, while also discussing the new development of standardizing PFTs using reference values.[319] Following a 1991 ATS statement on the selection and interpretation of PFT reference values,[320] the ATS's subsequent guideline on the standardization of PFTs in 1995 placed emphasis solely on % predicted values.[321] It is perhaps this report that can explain the decrease in percent cumulative use of absolute values in favour of % predicted values in the mid-1990s.

a)



b)



Appendix Figure D-1: a) Cumulative Use and b) Percent Cumulative Use of Absolute and % Predicted Pulmonary Function Test (PFT) Measures as Longitudinal Outcomes for Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Progression. *For each year, percent cumulative use was calculated by dividing the cumulative use of both absolute and % predicted PFT measures by the cumulative number of published articles.*

Appendix E: Characteristics of the Validation Studies (N = 50)

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Harrison et al. 1991[237]	United Kingdom	Case-Control Study (Hospital Records)	34 (76% Female)	45 (22-67)	Lung Biopsy (Four-Point Scoring System for Interstitial Fibrosis, Four-Point Scoring System for Loss of Lung Architecture)	Cross-Sectional	DLCO%
Wells et al. 1997[238]	United Kingdom	Cross-Sectional Study (Interstitial Lung Disease Unit)	64 (80% Female)	48.8 ± 11.6	HRCT (Nearest 5% - Overall Lung Involvement)	Cross-Sectional	DLCO%, DLCO/VA%, FEV_1%, FVC%, TLC%
Wells et al. 1997[239]	United Kingdom	Cross-Sectional Study (Interstitial Lung Disease Unit)	57 (82% Female)	48 ± 12	HRCT (Nearest 5% - Overall Lung Involvement)	Cross-Sectional	DLCO%
Diot et al. 1998[240]	France	Cross-Sectional Study (Hospital Referrals)	52 (98% Female)	53.71 ± 14.12 (23-79)	HRCT (Warrick Total Score (0-30))	Cross-Sectional	DLCO%, TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Kim et al. 2001[241]	South Korea	Cohort Study (Tertiary Hospital Clinical Records)	40 (85% Female)	54 (27-76)	HRCT (Nearest 5% - Ground-Glass Opacity, Nearest 5% - Honeycombing, Nearest 5% - Irregular Linear Opacity, Nearest 5% - Overall Lung Involvement)	Longitudinal	DLCO%, FEV_1 Absolute, FEV_1/FVC%, FVC Absolute
Shahin et al. 2001[242]	Egypt	Case-Control Study (Department of Rheumatology and Rehabilitation)	22 (95% Female)	37.6 ± 14.3	HRCT (Total Score (0-At Least 21))	Cross-Sectional	DLCO%
Han et al. 2003[243] (Abstract)	South Korea	Cross-Sectional Study	43	46.8 ± 11.9	HRCT (Absence/Presence of Pulmonary Fibrosis)	Cross-Sectional	DLCO%
Ooi et al. 2003[244]	China	Cross-Sectional Study (Division of Rheumatology)	45 (89% Female)	48.5 ± 13.4	HRCT (Fibrosis Index (0-48), Inflammatory Index (0-48), Total Score (0-96))	Cross-Sectional	DLCO%, FEV_1%, FVC%, TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
De Santis et al. 2005[107]	Italy	Cohort Study (Outpatient Clinic of the Division of Rheumatology)	100 (92% Female)	55.4 ± 11.9	HRCT (Kazerooni Alveolar Score (0-5), Kazerooni Interstitial Score (0-5))	Cross-Sectional	DLCO%, FVC%
Orlandi et al. 2006[245]	Italy	Cross-Sectional Study (SSc Outpatients)	39 (87% Female)	58 ± 13 (18-80)	HRCT (Inspiratory Volume/Body Surface Area, Low-Dose Volumetric Kurtosis, Low-Dose Volumetric Mean Lung Attenuation, Low-Dose Volumetric Total Lung Skewness, Total Mean Lung Attenuation, Total Lung Kurtosis, Total Lung Skewness, Warrick Total Score (0-30))	Cross-Sectional	DLCO%, TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Beretta et al. 2007[246]	Italy	Cross-Sectional Study (Centre of Expertise for Systemic Autoimmune Diseases Outpatient Clinic)	28 (82% Female)	52.2 ± 10.6	HRCT (Warrick Total Score (0-30))	Cross-Sectional	DLCOcorr%, FVC%, TLC%
Camiciottoli et al. 2007[247]	Italy	Cross-Sectional Study	48 (88% Female)	57 ± 13 (18-80)	HRCT (Total Lung Kurtosis, Total Lung Skewness, Total Mean Lung Attenuation, Warrick Total Score (0-30))	Cross-Sectional	DLCO%, FEV_1%, FRC%, FVC%
Goldin et al. 2008[248]	United States	Cross-Sectional Study (Scleroderma Lung Study I (SLS I))	162 (70% Female)	51 ± 12.3	HRCT (Global Fibrosis Score (0-4), Global Ground-Glass Opacity Score (0-4), Global Honeycombing Score (0-4))	Cross-Sectional	DLCO%, FEV_1%, FEV_1/FVC%, FVC%, TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Bellia et al. 2009[249]	Italy	Cross-Sectional Study (Department of Rheumatology)	31 (97% Female)	54 ± 10.4	HRCT (Warrick Alveolitis Index (0-4), Warrick Extent Score (1-3), Warrick Fibrosis Index (0-26), Warrick Severity Score (1-5), Warrick Total Score (0-30))	Cross-Sectional	DLCO%, FEV_1%, TLC%
Goldin et al. 2009[250]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	98 (74% Female)	46.6	HRCT (Global Fibrosis Score (0-4))	Longitudinal	DLCO%, FVC%, TLC%
Vonk et al. 2009[251]	The Netherlands	Cross-Sectional Study (Pulmonary Hypertension Screening, a Multidisciplinary Approach in Scleroderma (POEMAS) and Nationwide Survey)	1,000	N/R	HRCT (Scoring Not Reported)	Cross-Sectional	TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Ananyeva et al. 2010[252] (Abstract)	Russia	Cross-Sectional Study	138 (90% Female)	47 ± 13	HRCT (Extent of Lung Involvement – Scoring Not Reported)	Cross-Sectional	DLCO%, FVC%
Gilson et al. 2010[133]	France	Cohort Study (Department of Rheumatology)	105 (86% Female)	52.7 ± 11.8	HRCT (Wells Total Score (0-3))	Longitudinal	FVC%
Peng et al. 2010[253] (Abstract)	China	Cross-Sectional Study (Scleroderma Study of Peking Union Medical College Hospital (PUMCH))	68	N/R	HRCT (Extent of Lung Involvement – Scoring Not Reported)	Cross-Sectional	DLCO%, FVC%, TLC%
Kim et al. 2011[254] (Abstract)	United States	Cross-Sectional Study (Anonymized Research Database)	119	48 ± 10.6	HRCT (Quantitative Percentage with Fibrosis in Whole Lung)	Cross-Sectional	DLCO%, FEV_1%, FVC%, TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Kim et al. 2011[255]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	83	N/R	HRCT (Quantitative Percentage with Fibrosis in Highest Score Zone at Baseline, Quantitative Percentage with Fibrosis in Whole Lung)	Longitudinal	FVC%, TLC%
Moghadam et al. 2011[256]	Iran	Cross-Sectional Study (Rheumatology Research Center)	55 (91% Female)	38.4 ± 1.3 (17–63)	HRCT (Wells Total Score (0-4))	Cross-Sectional	DLCO Absolute, DLCO%, FVC Absolute, FVC%, TLC Absolute, TLC%
Parra et al. 2011[257] (Abstract)	Brazil	Cross-Sectional Study	30 (77% Female)	N/R	HRCT (Extent of Lung Involvement – Scoring Not Reported)	Cross-Sectional	DLCO%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Assayag et al. 2012[258] (Abstract)	Canada	Cross-Sectional Study (Canadian Scleroderma Research Group (CSRG))	54 (89% Female)	58.5	HRCT (Global Fibrosis Score (0-4), Global Ground-Glass Opacity Score (0-4), Global Honeycombing Score (0-4), Global Severity Score (0-12))	Cross-Sectional	DLCO%, FVC%
Mantero et al. 2012[259] (Abstract)	Italy	Case-Control Study	32 (84% Female)	62.5 (59-73)	HRCT (Extent of Lung Involvement – Scoring Not Reported)	Cross-Sectional	FVC%
Mittal et al. 2012[260] (Abstract)	India	Cross-Sectional Study	23 (91% Female)	35.3 ± 9.9	HRCT (Total Score (0-24))	Cross-Sectional	FVC%
Pernot et al. 2012[261]	France	Case-Control Study (Department of Dermatology, Department of Internal Medicine)	35 (83% Female)	60.1	HRCT (Absence/Presence of ILD)	Cross-Sectional	DLCO%
Perrin et al. 2012[262] (Abstract)	France	Cross-Sectional Study	72	N/R	HRCT (Absence/Presence of ILD)	Cross-Sectional	DLCO%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Wilsher et al. 2012[263]	New Zealand	Cross-Sectional Study (Rheumatology Clinics)	30 (80% Female)	47 ± 12 (18-70)	HRCT (Total Extent of Ground-Glass Opacity (0-24), Total Extent of Reticular Pattern (0-24))	Cross-Sectional	DLCO%, FEV_1%, VC%
Zimmermann et al. 2012[264] (Abstract)	Brazil	Cross-Sectional Study	45	N/R	HRCT (Tomographic Index)	Cross-Sectional	DLCO, Final Expiratory Volume, FVC, RV, TLC
Celeste et al. 2013[169]	Italy	Cohort Study (Outpatient Clinic Referral Center for Systemic Autoimmune Diseases)	221 (90% Female)	45.5	HRCT (Nearest 5% - Overall Lung Involvement)	Cross-Sectional	DLCO%, FVC%
Gatta et al. 2013[265]	Italy	Cross-Sectional Study (Hospital Information System)	42 (14% Female)	48 (27-66)	HRCT (Modified Warrick Total Score (0-115))	Cross-Sectional	DLCO%, DLCO/VA%, FVC%, TLC%
Nguyen-Kim et al. 2013[266] (Abstract)	Switzerland	Cross-Sectional Study	37 (95% Female)	57 ± 12.5	HRCT (Total Lung Kurtosis, Total Lung Skewness)	Cross-Sectional	DLCO, FEV_1, FVC, TLC

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Piorunek et al. 2013[267]	Poland	Cross-Sectional Study	37 (84% Female)	43.2 ± 13.9	HRCT (Warrick Total Score (0-30))	Cross-Sectional	DLCO%
Zamora et al. 2013[268] (Abstract)	United States	Cross-Sectional Study (Pulmonary Hypertension Recognition and Outcomes in Scleroderma (PHAROS) Registry)	336	N/R	HRCT (Total Extent of Fibrosis – Scoring Not Reported, Total Extent of Ground-Glass Opacity – Scoring Not Reported, Total Extent of Honeycombing – Scoring Not Reported)	Cross-Sectional	DLCO%, FVC%, TLC%
Colaci et al. 2014[269]	Italy	Cross-Sectional Study (Rheumatology Centre)	107 (81% Female)	52.1 ± 12.3	HRCT (Modified Schurawitzki Total Score (0-18))	Cross-Sectional	DLCOcorr/VA%, FEV_1%, FVC%, TLC%, VC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Ariani et al. 2015[270]	Italy	Cross-Sectional Study (Units of Rheumatology)	257 (91% Female)	60.0 ± 13.4	HRCT (Fibrosis Ratio, Parenchymal Kurtosis, Parenchymal Mean Lung Attenuation, Parenchymal Skewness, Parenchymal Standard Deviation, Total Lung Kurtosis, Total Lung Skewness, Total Mean Lung Attenuation, Total Lung Standard Deviation)	Cross-Sectional	DLCO%, DLCO/VA%, FVC%, TLC%
Bernstein et al. 2015[271] (Abstract)	United States	Cross-Sectional Study (Prospective Registry of Early Systemic Sclerosis (PRESS))	91 (68% Female)	52.0 ± 15.3	HRCT (Absence/Presence of ILD)	Cross-Sectional	DLCO%, FVC%, TLC%
Ghandour et al. 2015[272] (Abstract)	Egypt	Cross-Sectional Study (Outpatient Clinic of Rheumatology & Rehabilitation Department)	40 (100% Female)	(17-57)	HRCT (Extent of Lung Involvement – Scoring Not Reported)	Cross-Sectional	FEV_1/FVC%, FVC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Guarnieri et al. 2015[273]	Italy	Case-Control Study (Outpatient Clinic of Rheumatology Unit)	37 (81% Female)	54	HRCT (Global Severity Score (0-12))	Cross-Sectional	DLCO%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Khanna et al. 2015[206]	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	93 (73% Female)	47.19 ± 11.72	HRCT (Maximum Fibrosis Score in Zone of Maximal Involvement (0-4), Nearest 5% - Overall Lung Involvement, Quantitative Percentage with Fibrosis in Whole Lung, Quantitative Percentage with Fibrosis in Zone of Maximal Involvement, Quantitative Total Extent of Interstitial Lung Disease in Whole Lung, Quantitative Total Extent of Interstitial Lung Disease in Zone of Maximal Involvement)	Cross-Sectional	DLCO%, FVC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Kim et al. 2015[274] (Abstract)	United States	Cohort Study (Scleroderma Lung Study I (SLS I))	76	N/R	HRCT (Quantitative Percentage with Fibrosis in Whole Lung, Total Lung Kurtosis)	Cross-Sectional, Longitudinal	DLCO%, FVC%
Ninaber et al. 2015[212]	The Netherlands	Cohort Study (Referrals to Tertiary Outpatient Targeted Multidisciplinary Healthcare Program)	41 (76% Female)	50.9	HRCT (% High Attenuation Areas, 85th Percentile Density Score)	Cross-Sectional	DLCO%, FVC%
Salaffi et al. 2015[275]	Italy	Cross-Sectional Study	79 (85% Female)	59 ± 9.7	HRCT (Computer-Aided Method Pulmonary Fibrosis Fraction (%))	Cross-Sectional	DLCO%, FEV_1%, FVC%
Suliman et al. 2015[43]	Switzerland	Cohort Study (Division of Rheumatology)	102 (77% Female)	58.5 (28-90)	HRCT (Absence/Presence of ILD)	Cross-Sectional	DLCOcorr%, FVC%, TLC%
Antoniou et al. 2016[276]	United Kingdom	Cross-Sectional Study (Hospital Records, Centre for Rheumatology and Pulmonary Hypertension)	333 (78% Female)	54.4 ± 13.1	HRCT (Nearest 5% - Overall Lung Involvement)	Cross-Sectional	DLCO%, FEV_1%, FVC%, FVC/DLCO Absolute

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
Cetincakmak et al. 2016[277]	Turkey	Cross-Sectional Study (Referrals to Department of Radiology Clinic)	38 (95% Female)	41	HRCT (Left Percentage of Lower Lobe Volume, Right Percentage of Lower Lobe Volume, Total Percentage of Lower Lobe Volume)	Cross-Sectional	DLCOcorr%, FEV_1/FVC%, FVC%
Kloth et al. 2016[225]	Germany	Cohort Study (Radiology Department Database)	26 (54% Female)	37.45 ± 9.83 (11-51)	HRCT (Mean Lung Density)	Longitudinal	FEV_1
Salaffi et al. 2016[278]	Italy	Cross-Sectional Study (Department of Rheumatology)	126 (84% Female)	60.68 ± 10.74 (22-78)	HRCT (Computer-Aided Method Pulmonary Fibrosis Fraction (%))	Cross-Sectional	DLCO%, FVC%
Tashkin et al. 2016[279]	United States	Cross-Sectional Study (Scleroderma Lung Study I (SLS I), Scleroderma Lung Study II (SLS II))	300 (72% Female)	50.3	HRCT (Quantitative Ground-Glass Opacity in Whole Lung, Quantitative Ground-Glass Opacity in Zone of Maximal Involvement, Quantitative Percentage with	Cross-Sectional	DLCOcorr%, FEV_1/FVC%, FVC%, TLC%

Record	Country	Study Design (SSc Recruitment Method/Site)	SSc Subjects in Study (% Female)	Age, Years (Range)	Gold Standard (Scoring Method)	Type of Validation	PFT Measures Considered
					Fibrosis in Whole Lung, Quantitative Percentage with Fibrosis in Zone of Maximal Involvement, Quantitative Total Extent of Interstitial Lung Disease in Whole Lung, Quantitative Total Extent of Interstitial Lung Disease in Zone of Maximal Involvement)		

Abbreviations: % = Percent Predicted; DLCO = Diffusing Capacity for Carbon Monoxide; DLCO/VA = Diffusing Capacity for Carbon Monoxide Corrected for Alveolar Volume; DLCOcorr = Diffusing Capacity for Carbon Monoxide Corrected for Haemoglobin; DLCOcorr/VA = Diffusing Capacity for Carbon Monoxide Corrected for Alveolar Volume and Haemoglobin; FEV_1 = Forced Expiratory Volume in the 1st Second of Forced Exhalation; FRC = Functional Residual Capacity; FVC = Forced Vital Capacity; HRCT = High-Resolution Computed Tomography; ILD = Interstitial Lung Disease; N/R = Not Reported; PFT = Pulmonary Function Test; RV = Residual Volume; TLC = Total Lung Capacity; VC = Vital Capacity

Appendix F: Results of the Validation Studies whose Aim was Other than to Validate Pulmonary Function Tests (N = 45)

The measures of validity are grouped alphabetically by scoring system used.

HRCT: % High Attenuation Areas	
Ninaber et al. 2015[212]	DLCO%: r = -0.48 (p = 0.002) FVC%: r = -0.62 (p < 0.001)
HRCT: 85th Percentile Density Score	
Ninaber et al. 2015[212]	DLCO%: r = -0.49 (p = 0.001) FVC%: r = -0.64 (p < 0.001)
HRCT: Absence/Presence of Pulmonary Fibrosis/ILD	
Han et al. 2003[243] (Abstract)	DLCO% = 68%: AUC = 0.814
Pernot et al. 2012[261]	DLCO% = 67%: AUC = 0.75 (p = 0.005) Se = 54% Sp = 91%
Perrin et al. 2012[262] (Abstract)	DLCO% = Unclear: AUC = 0.67
Bernstein et al. 2015[271] (Abstract)	DLCO% < 80%: Se = 86.4% Sp = 60.0% PPV = 70.4% NPV = 80.0% FVC% < 80%: Se = 56.0% Sp = 55.0% PPV = 60.9% NPV = 50.0% TLC% < 80%: Se = 52.9% Sp = 70.6% PPV = 64.3% NPV = 60.0% DLCO% & FVC% < 80%: Se = 90.9% Sp = 45.0% PPV = 64.5% NPV = 81.8% DLCO% & FVC & TLC% < 80%: Se = 88.2% Sp = 47.1% PPV = 62.5% NPV = 80.0%

Suliman et al. 2015[43]	FVC% < 80%: FNR = 62.5% FPR = 7.9% Se = 37.5% (0.3-0.5) Sp = 92% (0.8-1.0) LR+ = 4.7 (1.5-4.7) LR- = 0.7 (0.5-0.8) DLCOcorr <70% or FVC% < 80%: FNR = 41.0% FPR = 34.3% Se = 59.0% (0.4-0.7) Sp = 65.8% (0.5-0.7) LR+ = 1.7 (1.0-2.8) LR- = 0.6 (0.4-0.9) FVC% or TLC% < 80%: FNR = 55.0% FPR = 13.2% Se = 45.0% (0.3-0.5) Sp = 86.0% (0.7-0.9) LR+ = 3.4 (1.4-8.1) LR- = 0.6 (0.4-0.8) DLCOcorr% < 70% or FVC% or TLC% < 80%: FNR = 37.0% FPR = 37.0% Se = 62.0% (0.5-0.7) Sp = 63.0% (0.4-0.7) LR+ = 1.7 (1.0-2.6) LR- = 0.6 (0.4-0.8)
HRCT: Computer-Aided Method Pulmonary Fibrosis Fraction (%)	
Salaffi et al. 2015[275]	DLCO%: r = -0.490 (p < 0.0001) FEV_1%: r = -0.675 (p < 0.0001) FVC%: r = -0.653 (p < 0.0001)
Salaffi et al. 2016[278]	DLCO%: r = -0.556 (p < 0.0001) FVC%: r = -0.670 (p < 0.0001)
HRCT: Extent of Lung Involvement – Scoring Not Reported	
Ananyeva et al. 2010[252] (Abstract)	DLCO%: r = -0.42 (p = 0.00) FVC%: r = -0.31 (p = 0.0002)
Peng et al. 2010[253] (Abstract)	DLCO%: r = -0.496 (p = 0.000) FVC%: r = -0.324 (p = 0.009) TLC%: r = -0.465 (p = 0.000)
Parra et al. 2011[257] (Abstract)	DLCO%: r = -0.601 (p = 0.01)
Mantero et al. 2012[259] (Abstract)	FVC%: r = -0.77 (p < 0.0001)
Ghandour et al. 2015[272] (Abstract)	FEV_1/FVC%: r = 0.593 (p = 0.000) FVC%: r = 0.373 (p = 0.018)
HRCT: Fibrosis Index (0-48)	
Ooi et al. 2003[244]	FVC%: r = -0.31 (p = 0.05) TLC%: r = -0.38 (p = 0.02)
HRCT: Fibrosis Ratio	

Ariani et al. 2015[270]	DLCO%: r = -0.10 (NS) DLCO/VA%: r = 0.06 (NS) FVC%: r = -0.18 (p = 0.0038) TLC%: r = -0.04 (NS)
HRCT: Global Fibrosis Score (0-4)	
Goldin et al. 2008[248]	DLCO%: r = -0.44 (p = 0.0001) FEV_1%: r = -0.05 (p = 0.54) FEV_1/FVC%: r = 0.31 (p = 0.0002) FVC%: r = -0.22 (p = 0.007) TLC%: r = -0.36 (p = 0.0001)
Goldin et al. 2009[250]	**Longitudinal Validation:** DLCO%: Kendall τ = 0.199 (p = 0.053) FVC%: Kendall τ = 0.21 (p = 0.041) TLC%: Kendall τ = 0.22 (p = 0.035)
Assayag et al. 2012[258] (Abstract)	DLCO%: r = -0.587 (p < 0.005) FVC%: r = -0.535 (p < 0.005)
HRCT: Global Ground-Glass Opacity Score (0-4)	
Goldin et al. 2008[248]	DLCO%: r = 0.05 (p = 0.52) FEV_1%: r = 0.19 (p = 0.02) FEV_1/FVC%: r = 0.02 (p = 0.76) FVC%: r = 0.14 (p = 0.08) TLC%: r = -0.03 (p = 0.7)
Assayag et al. 2012[258] (Abstract)	DLCO%: r = -0.521 (p < 0.005) FVC%: r = -0.450 (p < 0.005)
HRCT: Global Honeycombing Score (0-4)	
Goldin et al. 2008[248]	DLCO%: r = -0.25 (p = 0.002) FEV_1%: r = -0.07 (p = 0.41) FEV_1/FVC%: r = -0.005 (p = 0.59) FVC%: r = -0.04 (p = 0.61) TLC%: r = -0.19 (p = 0.02)
Assayag et al. 2012[258] (Abstract)	DLCO%: r = -0.398 (p < 0.005) FVC%: r = -0.458 (p < 0.005)
HRCT: Global Severity Score (0-12)	
Assayag et al. 2012[258] (Abstract)	DLCO%: r = -0.617 (p < 0.0001) FVC%: r = -0.580 (p < 0.0001)
Guarnieri et al. 2015[273]	DLCO%: r = 0.45 (p = 0.01)
HRCT: Inflammatory Index (0-48)	
Ooi et al. 2003[244]	DLCO%: r = -0.43 (p = 0.008)
HRCT: Inspiratory Volume/Body Surface Area	
Orlandi et al. 2006[245]	DLCO%: r = 0.56 (p < 0.01) TLC%: r = 0.69 (p < 0.01)
HRCT: Kazerooni Alveolar Score (0-5)	
De Santis et al. 2005[107]	DLCO%: r = -0.53 (p < 0.0001) FVC%: r = -0.51 (p < 0.0001)
HRCT: Kazerooni Interstitial Score (0-5)	

De Santis et al. 2005[107]	DLCO%: r = -0.35 (p = 0.0006) FVC%: r = -0.32 (p = 0.0016)
HRCT: Left Percentage of Lower Lobe Volume	
Cetincakm et al. 2016[277]	DLCOcorr%: r = 0.076 (p = 0.750) FEV_1/FVC%: r = -0.037 (p = 0.873) FVC%: r = 0.579 (p = 0.006)
HRCT: Low-Dose Volumetric Kurtosis	
Orlandi et al. 2006[245]	DLCO%: r = 0.72 (p < 0.01) TLC%: r = 0.75 (p < 0.01)
HRCT: Low-Dose Volumetric Mean Lung Attenuation	
Orlandi et al. 2006[245]	DLCO%: r = -0.68 (p < 0.01) TLC%: r = -0.74 (p < 0.01)
HRCT: Low-Dose Volumetric Total Lung Skewness	
Orlandi et al. 2006[245]	DLCO%: r = 0.72 (p < 0.01) TLC%: r = 0.71 (p < 0.01)
HRCT: Maximum Fibrosis Score in Zone of Maximal Involvement (0-4)	
Khanna et al. 2015[206]	Placebo Group: DLCO%: r = -0.46 (p = 0.001) FVC%: r = -0.21 (p = 0.15) Cyclophosphamide Group: DLCO%: r = -0.44 (p = 0.003) FVC%: r = -0.16 (p = 0.29)
HRCT: Mean Lung Density	
Kloth et al. 2016[225]	**Longitudinal Validation:** FEV_1: r = 0.733 (p = 0.016)
HRCT: Modified Schurawitzki Total Score (0-18)	
Colaci et al. 2014[269]	DLCOcorr/VA%: r = -0.124 (NS) TLC%: r = -0.206 (p = 0.033) VC%: r = -0.310 (p < 0.001)
HRCT: Modified Warrick Total Score (0-115)	
Gatta et al. 2013[265]	DLCO%: r = -0.741 (p = 2.02E-08) DLCO/VA%: r = -0.687 (p = 0.0000005) FVC%: r = -0.509 (p = 0.000575) TLC%: r = -0.654 (p = 0.00000264)
HRCT: Nearest 5% - Ground-Glass Opacity	
Kim et al. 2001[241]	**Longitudinal Validation:** DLCO%: Kendall τ = 0.57 (NS) FEV_1 Absolute: Kendall τ = -0.221 (NS) FEV_1/FVC%: Kendall τ = -0.134 (NS) FVC Absolute: Kendall τ = -0.125 (NS)
HRCT: Nearest 5% - Honeycombing	
Kim et al. 2001[241]	**Longitudinal Validation:** DLCO%: Kendall τ = -0.411 (p = 0.049) FEV_1 Absolute: Kendall τ = -0.295 (p > 0.05) FEV_1/FVC%: Kendall τ = -0.276 or -0.020 (p > 0.05) FVC Absolute: Kendall τ = -0.272 (p > 0.05)

HRCT: Nearest 5% - Irregular Linear Opacity	
Kim et al. 2001[241]	**Longitudinal Validation:** DLCO%: Kendall τ = -0.172 (NS) FEV_1 Absolute: Kendall τ = 0.0 (NS) FEV_1/FVC%: Kendall τ = 0.256 (NS) FVC Absolute: Kendall τ = 0.20 (NS)
HRCT: Nearest 5% – Overall Lung Involvement	
Wells et al. 1997[238]	All Patients: DLCO%: r = -0.70 ($p < 0.0005$) DLCO/VA%: r = -0.38 ($p < 0.003$) FEV_1%: r = -0.43 ($p < 0.001$) FVC%: r = -0.46 ($p < 0.0005$) TLC%: r = -0.51 ($p < 0.0005$) Patients Undergoing Maximal Exercise Tests: DLCO%: R^2 = 0.52 DLCO%: r = -0.69 ($p < 0.0005$) DLCO/VA%: r = -0.33 ($p < 0.02$) FEV_1%: r = -0.41 ($p = 0.003$) FVC%: r = -0.43 ($p < 0.002$) TLC%: r = -0.51 ($p < 0.0005$)
Wells et al. 1997[239]	Patients with Predominant Ground-Glass Attenuation (HRCT Grade 1): DLCO%: r = -0.68 Patients with Mixed Appearances (HRCT Grade 2): DLCO%: r = -0.78 Patients with Predominance of a Reticular Pattern (HRCT Grade 3): DLCO%: r = -0.76 Patients with Reversible Disease on Serial HRCT: DLCO%: r = -0.79 Patients with No Regression of Disease at Follow-Up HRCT: DLCO%: r = -0.72
Kim et al. 2001[241]	**Longitudinal Validation:** DLCO%: Kendall τ = -0.124 (NS) FEV_1 Absolute: Kendall τ = -0.249 (NS) FEV_1/FVC%: Kendall τ = -0.168 (NS) FVC Absolute: Kendall τ = -0.172 (NS)
Celeste et al. 2013[169]	DLCO%: r = -0.52 ($p < 0.0001$) (95% CI: -0.64, -0.38) FVC%: r = -0.456 ($p < 0.0001$) (95% CI: -0.599, -0.29)
Khanna et al. 2015[206]	Placebo Group: DLCO%: r = -0.48 ($p = 0.001$) FVC%: r = -0.05 ($p = 0.75$) Cyclophosphamide Group: DLCO%: r = -0.51 ($p = 0.001$) FVC%: r = -0.25 ($p = 0.09$)

Antoniou et al. 2016[276]	DLCO%: r = -0.56 FEV_1%: r = -0.28 FVC%: r = -0.35 FVC/DLCO Absolute: r = 0.36 ($p < 0.0005$)
HRCT: Parenchymal Kurtosis	
Ariani et al. 2015[270]	DLCO%: r = 0.42 ($p < 0.0001$) DLCO/VA%: r = 0.13 (NS) FVC%: r = 0.51 ($p < 0.0001$) TLC%: r = 0.50 ($p < 0.0001$)
HRCT: Parenchymal Mean Lung Attenuation	
Ariani et al. 2015[270]	DLCO%: r = -0.41 ($p < 0.0001$) DLCO/VA%: r = -0.09 (NS) FVC%: r = -0.52 ($p < 0.0001$) TLC%: r = -0.52 ($p < 0.0001$)
HRCT: Parenchymal Skewness	
Ariani et al. 2015[270]	DLCO%: r = 0.41 ($p < 0.0001$) DLCO/VA%: r = 0.13 (NS) FVC%: r = 0.49 ($p < 0.0001$) TLC%: r = 0.46 ($p < 0.0001$)
HRCT: Parenchymal Standard Deviation	
Ariani et al. 2015[270]	DLCO%: r = -0.33 ($p < 0.0001$) DLCO/VA%: r = -0.12 (NS) FVC%: r = -0.4 ($p < 0.0001$) TLC%: r = -0.43 ($p < 0.0001$)
HRCT: Quantitative Ground-Glass Opacity in Whole Lung	
Tashkin et al. 2016[279]	DLCOcorr%: r = -0.28 ($p < 0.0001$) FEV_1/FVC%: r = 0.15 ($p < 0.01$) FVC%: r = -0.10 ($p \geq 0.01$) TLC%: r = -0.21 ($p < 0.0001$)
HRCT: Quantitative Ground-Glass Opacity in Zone of Maximal Involvement	
Tashkin et al. 2016[279]	DLCOcorr%: r = 0.03 ($p \geq 0.01$) FEV_1/FVC%: r = -0.02 ($p \geq 0.01$) FVC%: r = -0.11 ($p \geq 0.01$) TLC%: r = 0.08 ($p \geq 0.01$)
HRCT: Quantitative Percentage with Fibrosis in Highest Zone at Baseline	
Kim et al. 2011[255]	**<u>Longitudinal Validation:</u>** FVC%: r = -0.40 ($p = 0.0003$) TLC%: r = -0.18 ($p = 0.12$)
HRCT: Quantitative Percentage with Fibrosis in Whole Lung	

Kim et al. 2011[254] (Abstract)	Evaluation Set: DLCO%: r = -0.35 ($p < 0.0001$) FEV_1%: r = -0.23 ($p < 0.0001$) FVC%: r = -0.31 ($p < 0.0001$) New Cohort: DLCO%: r = -0.35 ($p < 0.0001$) FEV_1%: r = -0.45 ($p < 0.0001$) FVC%: r = -0.53 ($p < 0.0001$)
Kim et al. 2011[255]	**Longitudinal Validation:** FVC%: r = -0.33 ($p = 0.003$) TLC%: r = -0.16 ($p = 0.17$)
Khanna et al. 2015[206]	Placebo Group: DLCO%: r = -0.22 ($p = 0.13$) FVC%: r = -0.17 ($p = 0.26$) Cyclophosphamide Group: DLCO%: r = -0.20 ($p = 0.20$) FVC%: r = -0.25 ($p = 0.11$)
Kim et al. 2015[274] (Abstract)	DLCO%: r = -0.50 FVC%: r = -0.49 **Longitudinal Validation:** FVC%: r = -0.39 ($p = 0.0007$)
Tashkin et al. 2016[279]	DLCOcorr%: r = -0.42 ($p < 0.0001$) FEV_1/FVC%: r = 0.15 ($p \geq 0.01$) FVC%: r = -0.27 ($p < 0.0001$) TLC%: r = -0.37 ($p < 0.0001$) Scleroderma Lung Study I (SLS I): DLCOcorr%: $R^2 = 0.46$ ($p < 0.0001$) Scleroderma Lung Study II (SLS II): DLCOcorr%: $R^2 = 0.39$ ($p < 0.0001$)
HRCT: Quantitative Percentage with Fibrosis in Zone of Maximal Involvement	
Khanna et al. 2015[206]	Placebo Group: DLCO%: r = -0.43 ($p = 0.002$) FVC%: r = -0.45 ($p = 0.002$) Cyclophosphamide Group: DLCO%: r = -0.41 ($p = 0.005$) FVC%: r = -0.39 ($p = 0.02$)
Tashkin et al. 2016[279]	DLCOcorr%: r = -0.49 ($p < 0.0001$) FEV_1/FVC%: r = 0.14 ($p \geq 0.01$) FVC%: r = -0.34 ($p < 0.0001$) TLC%: r = -0.44 ($p < 0.0001$) Scleroderma Lung Study I (SLS I): DLCOcorr%: $R^2 = 0.48$ ($p < 0.0001$) Scleroderma Lung Study II (SLS II): DLCOcorr%: $R^2 = 0.44$ ($p < 0.0001$)
HRCT: Quantitative Total Extent of ILD in Whole Lung	

Khanna et al. 2015[206]	Placebo Group: DLCO%: r = -0.35 (p = 0.01) FVC%: r = -0.38 (p = 0.008) Cyclophosphamide Group: DLCO%: r = -0.07 (p = 0.63) FVC%: r = -0.08 (p = 0.61)
Tashkin et al. 2016[279]	DLCOcorr%: r = -0.43 (p < 0.0001) FEV_1/FVC%: r = 0.20 (p < 0.01) FVC%: r = -0.22 (p < 0.0001) TLC%: r = -0.32 (p < 0.0001) Scleroderma Lung Study I (SLS I): DLCOcorr%: R^2 = 0.44 (p < 0.0001) Scleroderma Lung Study II (SLS II): DLCOcorr%: R^2 = 0.37 (p < 0.0001)
HRCT: Quantitative Total Extent of ILD in Zone of Maximal Involvement	
Khanna et al. 2015[206]	Placebo Group: DLCO%: r = -0.41 (p = 0.005) FVC%: r = -0.27 (p = 0.07) Cyclophosphamide Group: DLCO%: r = -0.24 (p = 0.12) FVC%: r = -0.19 (p = 0.23)
Tashkin et al. 2016[279]	DLCOcorr%: r = -0.44 (p < 0.0001) FEV_1/FVC%: r = 0.17 (p ≥ 0.01) FVC%: r = -0.32 (p < 0.0001) TLC%: r = -0.47 (p < 0.0001) Scleroderma Lung Study I (SLS I): DLCOcorr%: R^2 = 0.55 (p < 0.0001) Scleroderma Lung Study II (SLS II): DLCOcorr%: R^2 = 0.44 (p < 0.0001)
HRCT: Right Percentage of Lower Lobe Volume	
Cetincakm et al. 2016[277]	DLCOcorr%: r = 0.115 (p = 0.628) FEV_1/FVC%: r = -0.041 (p = 0.860) FVC%: r = 0.536 (p = 0.012)
HRCT: Scoring Not Reported	
Vonk et al. 2009[251]	TLC%: r = 0.527 (p < 0.01)
HRCT: Tomographic Index	
Zimmermann et al. 2012[264] (Abstract)	DLCO: r = 0.31 (p = 0.04) Final Expiratory Volume: r = 0.31 (p = 0.03) FVC: r = 0.40 (p = 0.005) RV: r = 0.33 (p = 0.02) TLC: r = 0.55 (p < 0.001)
HRCT: Total Extent of Fibrosis – Scoring Not Reported	
Zamora et al. 2013[268] (Abstract)	DLCO%: r = -0.37 (p < 0.0001) FVC%: r = -0.44 (p < 0.0001) TLC%: r = -0.41 (p < 0.0001)
HRCT: Total Extent of Ground-Glass Opacity – Scoring Not Reported	

Zamora et al. 2013[268] (Abstract)	DLCO%: r = -0.10 (p = 0.11) FVC%: r = -0.17 (p = 0.005) TLC%: r = -0.066 (p = 0.34)
HRCT: Total Extent of Ground-Glass Opacity (0-24)	
Wilsher et al. 2012[263]	DLCO%: r = -0.57 (p = 0.01) FEV_1%: r = -0.38 (p = 0.05) VC%: r = -0.36 (p = 0.07)
HRCT: Total Extent of Honeycombing – Scoring Not Reported	
Zamora et al. 2013[268] (Abstract)	DLCO%: r = -0.32 (p < 0.0001) FVC%: r = -0.38 (p < 0.0001) TLC%: r = -0.34 (p < 0.0001)
HRCT: Total Extent of Reticular Pattern (0-24)	
Wilsher et al. 2012[263]	DLCO%: r = -0.53 (p = 0.01) FEV_1%: r = -0.19 (p = 0.33) VC%: r = -0.13 (p = 0.51)
HRCT: Total Lung Kurtosis	
Orlandi et al. 2006[245]	DLCO%: r = 0.75 (p < 0.01) TLC%: r = 0.78 (p < 0.01)
Camiciottoli et al. 2007[247]	DLCO%: r = 0.58 (p < 0.0001) FEV_1%: r = 0.56 (p < 0.0001) FRC%: r = 0.57 (p < 0.0001) FVC%: r = 0.71 (p < 0.0001)
Nguyen-Kim et al. 2013[266] (Abstract)	DLCO: r = 0.29 (p = 0.1) FEV_1: r = 0.45 (p = 0.01) FVC: r = 0.39 (p = 0.03) TLC: r = 0.29 (p = 0.1)
Ariani et al. 2015[270]	DLCO%: r = 0.38 (p < 0.0001) DLCO/VA%: r = 0.07 (NS) FVC%: r = 0.51 (p < 0.0001) TLC%: r = 0.49 (p < 0.0001)
Kim et al. 2015[274] (Abstract)	DLCO%: r = 0.45 FVC%: r = 0.42 **<u>Longitudinal Validation:</u>** FVC%: r = 0.14 (p = 0.24)
HRCT: Total Lung Skewness	
Orlandi et al. 2006[245]	DLCO%: r = 0.73 (p < 0.01) TLC%: r = 0.77 (p < 0.01)
Camiciottoli et al. 2007[247]	DLCO%: r = 0.62 (p < 0.0001) FEV_1%: r = 0.52 (p < 0.0005) FRC%: r = 0.58 (p < 0.0001) FVC%: r = 0.67 (p < 0.0001)
Nguyen-Kim et al. 2013[266] (Abstract)	DLCO: r = 0.34 (p = 0.056) FEV_1: r = 0.38 (p = 0.03) FVC: r = 0.47 (p = 0.006) TLC: r = 0.34 (p = 0.056)

Ariani et al. 2015[270]	DLCO%: r = 0.41 (p < 0.0001) DLCO/VA%: r = 0.11 (NS) FVC%: r = 0.52 (p < 0.0001) TLC%: r = 0.51 (p < 0.0001)
HRCT: Total Lung Standard Deviation	
Ariani et al. 2015[270]	DLCO%: r = -0.15 (p = 0.0226) DLCO/VA%: r = 0.05 (NS) FVC%: r = -0.23 (p = 0.0005) TLC%: r = -0.11 (NS)
HRCT: Total Mean Lung Attenuation	
Orlandi et al. 2006[245]	DLCO%: r = -0.66 (p < 0.01) TLC%: r = -0.77 (p < 0.01)
Camiciottoli et al. 2007[247]	DLCO%: r = -0.55 (p < 0.0001) FEV_1%: r = -0.58 (p < 0.001) FRC%: r = -0.59 (p < 0.0001) FVC%: r = -0.66 (p < 0.0001)
Ariani et al. 2015[270]	DLCO%: r = -0.41 (p < 0.0001) DLCO/VA%: r = -0.07 (NS) FVC%: r = -0.54 (p < 0.0001) TLC%: r = -0.52 (p < 0.0001)
HRCT: Total Percentage of Lower Lobe Volume	
Cetincakm et al. 2016[277]	DLCOcorr%: r = 0.121 (p = 0.61) FEV_1/FVC%: r = -0.062 (p = 0.792) FVC%: r = 0.539 (p = 0.012)
HRCT: Total Score (0-At Least 21)	
Shahin et al. 2001[242]	DLCO%: r = 0.64 (p < 0.01)
HRCT: Total Score (0-24)	
Mittal et al. 2012[260] (Abstract)	FVC%: r = -0.48
HRCT: Total Score (0-96)	
Ooi et al. 2003[244]	DLCO%: r = -0.43 (p = 0.008) FEV_1%: r = -0.37 (p = 0.03) FVC%: r = -0.43 (p = 0.008) TLC%: r = -0.47 (p = 0.003 or 0.008)
HRCT: Warrick Alveolitis Index (0-4)	
Bellia et al. 2009[249]	DLCO%: r = -0.46 (p = 0.01) TLC%: r = -0.28 (p = 0.13)
HRCT: Warrick Extent Score (1-3)	
Bellia et al. 2009[249]	DLCO%: r = -0.41 (p = 0.02) FEV_1%: r = -0.33 (p = 0.06) TLC%: r = -0.37 (p = 0.04)
HRCT: Warrick Fibrosis Index (0-26)	
Bellia et al. 2009[249]	DLCO%: r = -0.38 (p = 0.04) TLC%: r = -0.35 (p = 0.05)
HRCT: Warrick Severity Score (1-5)	

Bellia et al. 2009[249]	DLCO%: r = -0.39 (p = 0.03) FEV_1%: r = -0.33 (p = 0.07) TLC%: r = -0.34 (p = 0.06)
HRCT: Warrick Total Score (0-30)	
Diot et al. 1998[240]	DLCO%: r = -0.50 (p < 0.0002) TLC%: r = -0.39 (p < 0.005)
Orlandi et al. 2006[245]	DLCO%: r = -0.45 (p < 0.01) TLC%: r = -0.69 (p < 0.01)
Beretta et al. 2007[246]	DLCOcorr%: r = -0.18 (p ≥ 0.05) FVC%: r = -0.25 (p ≥ 0.05) TLC%: r = -0.42 (p < 0.05)
Camiciottoli et al. 2007[247]	DLCO%: r = -0.39 (p < 0.01) FEV_1%: r = -0.53 (p < 0.005) FRC%: r = -0.51 (p < 0.001) FVC%: r = -0.56 (p < 0.0001)
Bellia et al. 2009[249]	DLCO%: r = -0.43 (p = 0.02) FEV_1%: r = -0.36 (p = 0.05) TLC%: r = -0.38 (p = 0.04)
Piorunek et al. 2013[267]	DLCO%: r = -0.36 (p < 0.05)
HRCT: Wells Total Score (0-3)	
Gilson et al. 2010[133]	**<u>Longitudinal Validation:</u>** FVC%: Concordance $\varkappa$ = 0.6
HRCT: Wells Total Score (0-4)	
Moghadam et al. 2011[256]	DLCO Absolute: r = -0.513 (p < 0.001) DLCO%: r = -0.657 (p < 0.001) FVC Absolute: r = -0.429 (p = 0.001) FVC%: r = -0.523 (p < 0.001) TLC Absolute: r = -0.375 (p = 0.005) TLC%: r = -0.549 (p < 0.001)
Lung Biopsy: Four-Point Scoring System for Interstitial Fibrosis	
Harrison et al. 1991[237]	DLCO%: r = -0.46 (p < 0.01)
Lung Biopsy: Four-Point Scoring System for Loss of Lung Architecture	
Harrison et al. 1991[237]	DLCO%: r = -0.4 (p < 0.05)

Abbreviations: % = Percent Predicted; AUC = Area Under the Curve; CI = Confidence Interval; DLCO = Diffusing Capacity for Carbon Monoxide; DLCO/VA = Diffusing Capacity for Carbon Monoxide Corrected for Alveolar Volume; DLCOcorr = Diffusing Capacity for Carbon Monoxide Corrected for Haemoglobin; DLCOcorr/VA = Diffusing Capacity for Carbon Monoxide Corrected for Alveolar Volume and Haemoglobin; FRC = Functional Residual Capacity; FVC = Forced Vital Capacity; FEV_1 = Forced Expiratory Volume in the 1st Second of Forced Exhalation; FNR = False Negative Rate; FPR = False Positive Rate; HRCT = High-Resolution Computed Tomography; ILD = Interstitial Lung Disease; LR- = Negative Likelihood Ratio; LR+ = Positive Likelihood Ratio; NPV = Negative Predictive Value; NS = Not Significant; p = p-value; PPV = Positive Predictive Value; R^2 = Coefficient of Determination; RV = Residual Volume; Se = Sensitivity; Sp = Specificity; TLC = Total Lung Capacity; VC = Vital Capacity

Appendix G: Quality of the Five Validation Studies Evaluating the Performance of Pulmonary Function Tests Against High-Resolution Computed Tomography

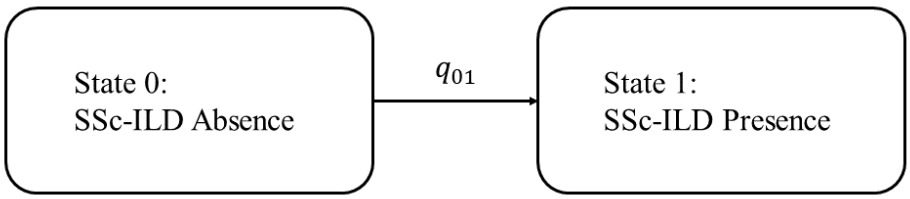
Study quality assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standards
Wells et al. 1997[238]	Low Risk	Unclear Risk	Unclear Risk	High Risk	Unclear Risk	Low Risk	Low Risk
Zamora et al. 2013[268] (Abstract)	Low Risk	Unclear Risk	Unclear Risk	High Risk	Unclear Risk	Low Risk	Low Risk
Bernstein et al. 2015[271] (Abstract)	Low Risk	Unclear Risk	Unclear Risk	High Risk	Low Risk	Low Risk	Low Risk
Suliman et al. 2015[43]	Low Risk	Unclear Risk	Low Risk	Unclear Risk	Unclear Risk	Low Risk	Low Risk
Tashkin et al. 2016[279]	Low Risk	Unclear Risk	Low Risk	High Risk	Unclear Risk	Low Risk	Low Risk

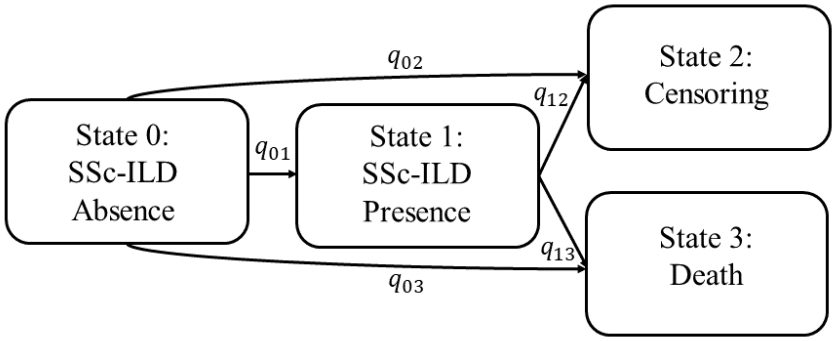
Appendix H: Hidden Markov Models

Hidden Markov models (HMMs) are an application of multi-state Markov models, a type of stochastic model that describes the process by which subjects transition from one discrete (disease) state to another.[59] These processes are often graphically depicted using separate boxes for each state and arrows representing possible transitions between states. In the simplest of terms, one can imagine a two-state model for systemic sclerosis-associated interstitial lung disease (SSc-ILD) onset with states 0 and 1 representing disease absence and presence, respectively, and where only transitions from state 0 to state 1 are permitted (Appendix Figure H-1a). Such models can easily be extended to include *n* states with arrows for all plausible transitions. They can also be further generalized to include absorbing states from which it is impossible to transition out of, such as censoring and death (Appendix Figure H-1b).[60, 63]

a)



b)



Appendix Figure H-1: Multi-State Models Depicting the Stochastic Process of Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Onset. *a) Two-state model.* q_{01} *represents the instantaneous risk of progressing from SSc-ILD absence to SSc-ILD presence. b) Four-state model with censoring and death absorbing states.* q_{01} *represents the instantaneous risk of progressing from SSc-ILD absence to SSc-ILD presence;* q_{02} *represents the instantaneous risk of censoring in SSc subjects without ILD;* q_{03} *represents the instantaneous risk of death in SSc subjects without ILD;* q_{12} *represents the instantaneous risk of censoring in SSc subjects with ILD;* q_{13} *represents the instantaneous risk of death in SSc subjects with ILD.*
Abbreviations: SSc-ILD = Systemic Sclerosis-Associated Interstitial Lung Disease

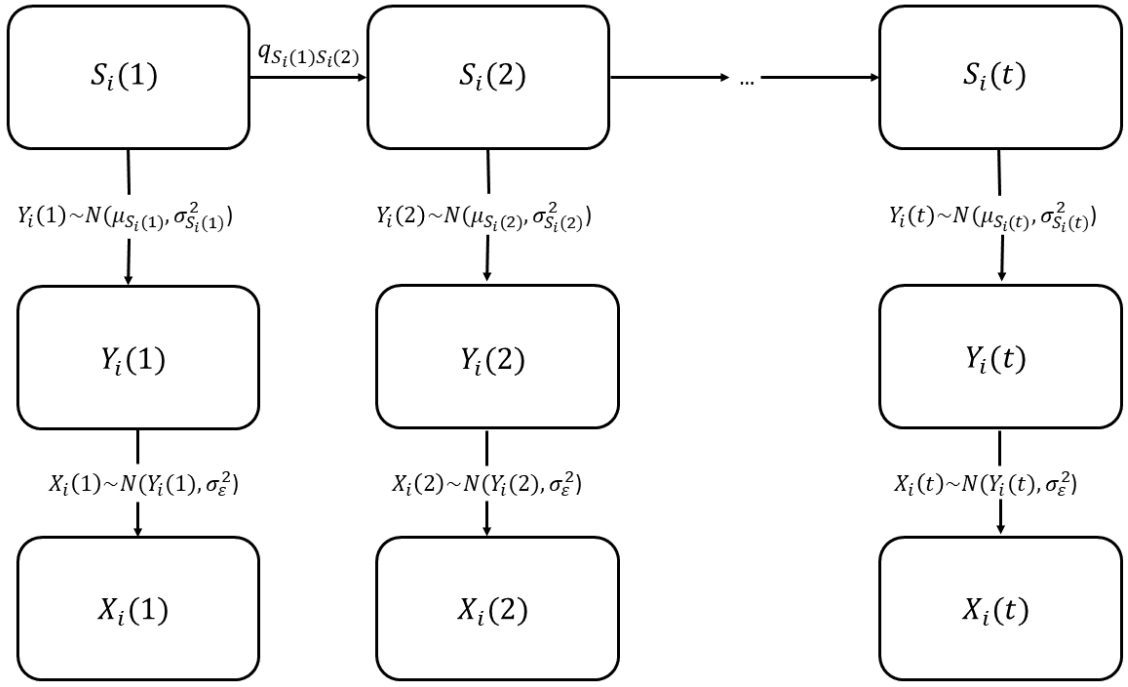
When a subject occupies state *r*, the probability that they next transition to state *s* is dependent on a certain transition intensity. For each pair of states *r* and *s*, a transition intensity q_{rs} denotes the instantaneous risk of transitioning from state *r* to state *s*. Mathematically, this can be written as $q_{rs} = \lim_{\delta t \to 0} P(S(t+\delta t) = s | S(t) = r)/\delta t$, where *S(t)* represents the state occupied at time *t*.[60, 63]

If the exact times at which subjects' transition from one state to another are observed, multi-state models can easily estimate transition intensities using maximum likelihood estimation. However, in most epidemiologic studies, data are often collected at arbitrary time points rendering the exact transition times unknown. In these instances, the Markov property (or memoryless assumption) is required to compute the likelihood of intermittently observed multi-state processes. Consequently, Markov models represent a distinct type of multi-state model as they assume that future progression is conditional only upon the current state occupied by a subject and not on the underlying trajectory taken to reach that state.[59, 63]

Markov models are useful tools when disease states can be diagnosed with accuracy and precision. However, disease states are sometimes not directly observed, but rather are inferred using surrogate markers or biomarkers. Often, such markers are prone to random variation and measurement error. In this case, HMMs may offer a solution.

As with Markov models, HMMs estimate the instantaneous risk of progressing from one disease state to another. However, this is done while simultaneously correcting for potential misclassification of

the observed disease state due to measurement error in the disease marker. This correction is possible as HMMs are hierarchical: they are comprised of a measurement error model accounting for the variability in the disease marker in addition to the above-described traditional Markov model (Appendix Figure H-2).[63, 66]



Appendix Figure H-2: Schematic Representation of a Hidden Markov Model. *The top portion of the figure represents a typical Markov process for a subject i who finds themselves in state $S_i(1)$ at time point 1, $S_i(2)$ at time point 2, and so on. The probability of transitioning from one state to the next is governed by a specific transition intensity q. The middle component of the figure assumes an underlying distribution for the marker given the disease state. The bottom component corresponds to the measurement error model which relates observed levels (X) of the marker to their true levels (Y) using a classical measurement error model. In this figure, the true marker levels Y are assumed to be normally distributed with means and variances dependent on the state occupied.*

When a patient occupies a specific disease state, this gives rise to a *true* biomarker level denoted by random variable *Y*. For example, the *true* FVC value, Y_{it}, for subject *i* at time point *t* will depend on their underlying SSc-ILD disease state (absence or presence). Thus, for each disease state an underlying probability distribution for *Y* can be specified (middle portion of Appendix Figure H-2).[63, 66] For

instance, FVC values can be normally distributed in both the absence and presence of SSc-ILD but centered around different means.

When markers are subject to measurement error, random variable Y is not observed but rather the imperfectly measured disease marker, denoted by random variable Y^*. Thus, an *observed* FVC measurement for patient i at time t, Y^*_{it}, can differ from their underlying *true* FVC value, Y_{it}. The measurement error component of the HMM thus relates these two variables using a classical measurement error model: $Y^*_{it} = Y_{it} + \varepsilon_{it}$, where ε_{it} are assumed to be independent and identically normally distributed, with a mean of zero and common variance σ^2. It is worth noting that when using an underlying normal distribuion for Y, the resulting normal-normal implies that the common variance parameter σ^2 cannot be estimated as it is confounded with and cannot be disentangled from the variability of Y.[63, 66]

In sum, the HMM will use a cohort of subjects' observed marker values Y^* to estimate the instantaneous risk of progressing from one disease state to another while correcting for measurement error in the marker. HMMs are especially powerful as they can calculate the probability of state occupancy at any time for a study subject using their history of marker measurements, the marker's underlying probability density function for each disease state, and the estimated transition intensities.[63, 66]

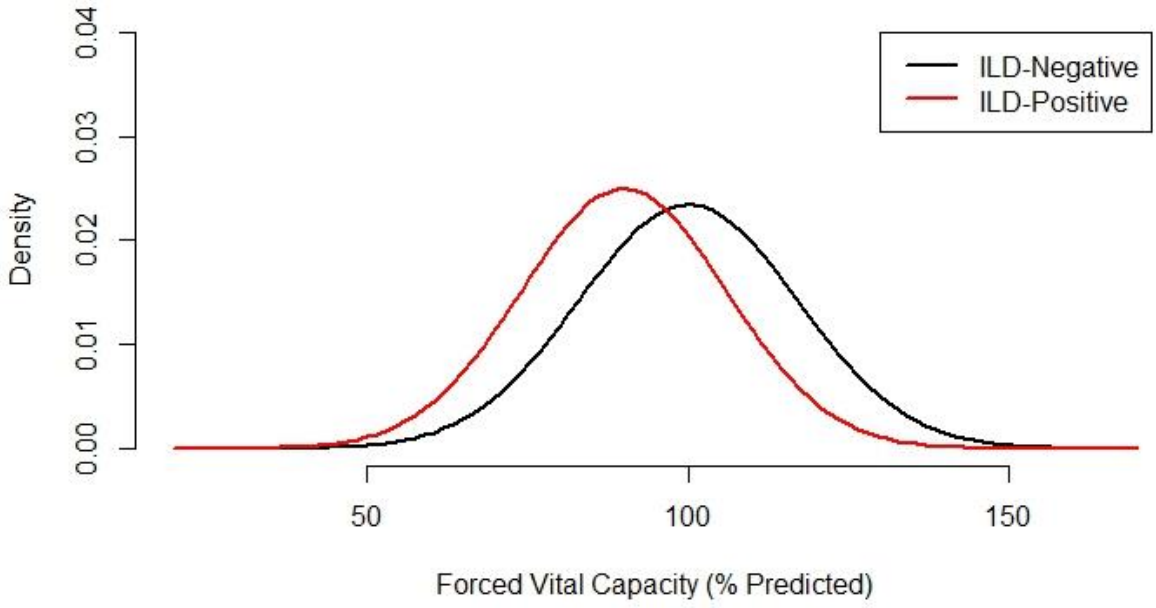
HMMs can support both discrete or continuous-time data.[322] They can also be extended to include multiple disease states to study disease progression and can account for censoring and death.[65] The identification of predictors of state transition can also be investigated by including covariates in the Markov model component of the HMM. These covariates, collectively denoted A, can be specified for each disease state using a proportional hazards model relating them to the transition intensities:

$q_{rs}(A) = q_{rs}^{(0)} \exp(\beta_{rs} A)$, where q_{rs} denotes the instantaneous risk (or transition intensity) of transitioning from state *r* to state *s* and β_{rs} corresponds to the log-hazard ratio. Similarly, covariates can be included in the measurement error component of the HMM. If a confounder *X* is known to affect measurement of the marker, it can be adjusted for in the measurement error model as follows: $Y_{it}^{*} = Y_{it} + \beta X + \varepsilon_{it}$.[64, 65]

The HMM has three main assumptions. As with traditional Markov models, the Markov property (or memoryless assumption) requires that the future of the Markov process be independent of the past state trajectory, *i.e.,* where one transitions to next only depends on the current state occupied. The stationarity assumption states that the transition intensities must be homogeneous and independent of time and individuals. Finally, the observation independence assumption assumes the conditional independence of the marker values given the underlying disease state.[66]

Appendix I: Initializing the Hidden Markov Model

When running a hidden Markov model (HMM), it is necessary to specify an initial underlying distribution of the disease marker given the different possible disease states. We assumed that the *true* forced vital capacity (FVC) values followed a normal distribution and assigned an initial mean of 100% for the interstitial lung disease (ILD)-absence distribution and of 90% for the ILD-presence distribution. The standard deviations were based on those observed in the first five visits of Canadian Scleroderma Research Group (CSRG) registry subjects who were ILD-free at baseline (assessed using the combination gold-standard) and who had at least five visits. Thus, the standard deviations were set to 16 in CSRG patients with ILD and to 17 in those without ILD. These distributions are illustrated in Appendix Figure I-1. While we assumed that these underlying distributions were normally distributed, other permissible options include categorical, uniform, exponential, Poisson, binomial and so on.[322]



Appendix Figure I-1: Forced Vital Capacity (FVC) Density Functions in the Absence (Black) and Presence (Red) of Systemic Sclerosis-Associated Interstitial Lung Disease (SSc-ILD) Used to Inform a Two-State Hidden Markov Model (HMM).

Appendix J: Simulation Study Details

The simulation study was modeled after systemic sclerosis (SSc) patients in the Canadian Scleroderma Research Group (CSRG) registry at risk of developing interstitial lung disease (ILD) during follow-up. We simulated 1,150 SSc patients with five annual visits. All subjects were ILD-free at baseline. ILD onset could occur during visits two through five and was simulated using a Markov process in the `msm` package.[65] The Markov process only allowed transitions from ILD absence to presence and a transition intensity of 0.065 was chosen in order to obtain an ILD incidence and prevalence comparable to those observed in the CSRG.

Since forced vital capacity (FVC) measurements can be quite variable amongst healthy subjects without ILD, we ensured that every simulated subject had their own unique FVC trajectory with highly correlated consecutive measurements. We began by assigning each subject their own *true* ILD-negative FVC value. This measurement was obtained by drawing from a normal distribution with a mean of 100% predicted and standard deviation (S.D.) of 15. This S.D. corresponds to the variation surrounding the random intercept of a longitudinal mixed model studying the natural progression of FVC during the first five visits of ILD-free CSRG subjects. ILD status was assessed using the combination gold-standard.

Next, *observed* FVC measurements for all ILD-negative visits were drawn from normal distributions with means equivalent to the subjects' own *true* FVC value and with a S.D. of six, corresponding to the residuals' variation in the above described mixed model. This introduced measurement error into the FVC measurements. The *observed* FVC measurements for ILD-positive visits were drawn from normal distributions with the same S.D. of six but with a mean *true* FVC value ten percentage points lower to introduce an effect of ILD. The inverse-sampling method was used to perform all draws.[323]

Appendix K: Sample R Code for Simulation Study and Hidden Markov Model

```
library(msm)

#Simulate 1150 patients with 5 visits each.
df<-data.frame(subject = rep(1:1150, rep(5,1150)), time = rep(1:5, 1150))

#Define the transition intensity matrix of the Markov process assuming 2
states.
qmatrix.2<- rbind(c(0,  0.065),

                  c(0,  0))

#Assign true ILD states to each patient-visit.
sim.df<-simmulti.msm(df, qmatrix.2, drop.absorb=FALSE)



##Simulate FVC Values

#Choose normal distribution parameters used to simulate FVC.
ll<--Inf #Lower limit of distribution
ul<-Inf #Upper limit of distribution
mu0<-100 #Mean of ILD absence distribution
mu1<-90 #Mean of ILD presence distribution
sd0<-15 #Standard deviation of ILD absence distribution
sd.me<-6 #Measurement error

#Draw ILD-absent FVC means for each subject
#using the inverse sampling method.
sim.df$true.fvc<-NA
for(i in 1:nrow(sim.df)){
      if(sim.df$time[i]==1)
            sim.df$true.fvc[i]=qnorm(runif(1, pnorm(ll,mu0,sd0),
            pnorm(ul,mu0,sd0)), mu0, sd0)
      else
            sim.df$true.fvc[i]=sim.df$true.fvc[i-1]

}

#For ILD-absent visits, draw the observed FVC.
sim.df$fvc.err<-NA
for(i in 1:nrow(sim.df)){
  if(sim.df$state[i]==1)
      sim.df$fvc.err[i]=qnorm(runif(1, pnorm(ll,sim.df$true.fvc[i], sd.me),
      pnorm(ul, sim.df$true.fvc[i], sd.me)), sim.df$true.fvc[i], sd.me)
}

#For ILD-present visits, draw the observed FVC.
for(i in 1:nrow(sim.df)){
  if(sim.df$state[i]==2)
      sim.df$fvc.err[i]=qnorm(runif(1, pnorm(ll, sim.df$true.fvc[i]-10,
      sd.me), pnorm(ul, sim.df$true.fvc[i]-10, sd.me)), sim.df$true.fvc[i]-
      10, sd.me)
}
```

```
#Create a variable for the baseline FVC.
sim.df$base.fvc<-NA
for(i in 1:nrow(sim.df)){
  sim.df$base.fvc[i]<-sim.df$fvc.err[sim.df$subject==sim.df$subject[i] &
      sim.df$time==1]
}


##Hidden Markov Model

#Specify initial values for the FVC standard deviations.
sd.noild<-17
sd.ild<-16

#Specify and initialize underlying distribution for FVC.
hmodel.2<-list(hmmNorm(mean=mu0, sd=sd.noild),
             hmmNorm(mean=mu1, sd=sd.ild))

#Run HMM adjusting for baseline FVC.
sim.df<-subset(sim.df, sim.df$time>1) #Remove baseline visit
sim.hmm<-msm(fvc.err~time, subject=subject, data=sim.df, qmatrix=qmatrix.2,
      hmodel=hmodel.2, hcovariates=list(~base.fvc, ~base.fvc),
      initprobs=c(0.93, 0.07), center=FALSE)

sim.hmm #Estimated parameters

#Calculate probability of ILD at each visit.
sim.vit<-viterbi.msm(sim.hmm)
sim.hmm.df<-data.frame(sim.df, sim.vit$pstate[,2])
colnames(sim.hmm.df)[8]<-"ILD.prob"
head(sim.hmm.df)
```

Appendix L: Measures of Validity

Calculating measures of validity in the simulation study was straightforward given that the visit number was used as the time scale of interest. Indeed, for every visit, a forced vital capacity (FVC) value was available as well as a known interstitial lung disease (ILD) status. This ensured that the results of the hidden Markov model (HMM), 80% cut-off and 10% decline in FVC algorithms could be compared to the true ILD status at the same time point (*i.e.*, visit).

In contrast, since calendar dates were used in the Canadian Scleroderma Research Group (CSRG) data analysis, a proper comparison would require that FVC values and gold-standard results be obtained on the same date which was very rare. Thus, for the 80% cut-off and 10% decline in FVC algorithms we compared each available combination gold-standard result to the chronologically next available algorithm result. This ensured that we were assessing the algorithms' diagnostic properties rather than their predictive properties.

For the HMM, estimated probabilities of ILD presence were available for the dates on which an FVC value was recorded, but not on the dates for which a combination gold-standard result was available. We therefore calculated the probability of ILD-presence on these dates using the following formula:

$$P(ILD_{it}) = (1 - e^{(-q\Delta t)}) \cdot (1 - P\big(ILD_{i(t-\Delta t)}\big) + P(ILD_{i(t-\Delta t)})$$

where:

$P(ILD_{it})$: HMM-estimated probability of ILD presence for patient *i* at time point *t*;

$P(ILD_{i(t-\Delta t)})$: HMM-estimated probability of ILD presence for patient *i* at previous pulmonary function testing date;

q: HMM-estimated transition intensity for the transition from the ILD-absence state to the ILD-presence state;

Δt: number of days separating the combination gold-standard date from the prior pulmonary function testing date.

Several assumptions and decisions were made to ensure that the HMM and both algorithms were fairly compared. First, since the probability of a patient having ILD as predicted by the HMM increases with each consecutive visit (as a general property), positive results for ILD detected by the 80% cut-off and 10% decline in FVC algorithms were carried forward throughout all subsequent visits. In other words, once these algorithms perceived a visit as being ILD-positive, all subsequent visits for that patient were automatically considered to have been identified as ILD-positive by the algorithms as well. Second, since ILD status at a patient's first visit can not be assessed by the 10% decline in FVC algorithm (by definition) and by the HMM (since the baseline visit was excluded from the analysis in order to adjust for baseline FVC), these first visits were excluded when assessing the performance of the 80% cut-off algorithm.

For both the simulation and CSRG studies, classical measures of diagnostic test accuracy were computed by comparing the HMM or algorithms' decision on ILD status to the true ILD status. Given that the HMM generates a probability of ILD rather than a definitive binary decision, a predicted probability ≥ 0.50 was chosen to define ILD presence.

Two methods were used to calculate certain measures of validity. The first method used all available comparisons, treating them as independent events. The second aimed to provide a better understanding of each algorithm's ability to specifically detect *incident* ILD. Therefore, only subjects' first ILD-positive result was included in these calculations. For example, overall sensitivity refers to the ability to correctly identify all ILD-positive results, while incident sensitivity refers to the ability to

correctly identify incident ILD. The latter corresponds to the proportion of patients with ILD whose disease was diagnosed at onset. We also computed a prevalent sensitivity – that is the proportion of patients with ILD whose disease was not detected at onset but was subsequently identified. Thus, adding the incident and prevalent sensitivities indicates the overall proportion of subjects with ILD who were positively identified during follow-up. Finally, the overall error rate corresponded to the sum of false negatives and false positives divided by the total number of comparisons.

Appendix M: Kappa Statistics

Appendix Table M-1 and Appendix Table M-2 outline the agreement between the 80% cut-off, 10% decline in FVC and HMM in the simulation and Canadian Scleroderma Research Group (CSRG) studies, respectively. The results found the agreement between all methods of ILD ascertainment to be poor, even when restricting to ILD-negative and positive visits, indicating very little overlap in their behavior and method of operation.

Appendix Table M-1: Kappa Coefficients Measuring the Agreement in Interstitial Lung Disease (ILD) Ascertainment of Simulated Systemic Sclerosis (SSc) Patients Between the 80% Cut-Off, 10% Decline and Hidden Markov Model (HMM) Algorithms.

Methods Compared	All Visits *		ILD-Negative Visits *		ILD-Positive Visits *	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
80% Cut-Off & 10% Decline	0.14	0.13, 0.15	0.11	0.10, 0.12	0.08	0.07, 0.10
Maximum $\varkappa$	0.61	0.60, 0.62	0.65	0.64, 0.66	0.41	0.40, 0.43
80% Cut-Off & HMM	0.24	0.23, 0.25	0.18	0.17, 0.19	0.20	0.19, 0.21
Maximum $\varkappa$	0.97	0.96, 0.97	0.83	0.81, 0.84	0.53	0.51, 0.55
10% Decline & HMM	0.33	0.33, 0.34	0.24	0.24, 0.25	0.28	0.27, 0.29
Maximum $\varkappa$	0.62	0.61, 0.63	0.51	0.49, 0.52	0.83	0.81, 0.85

**Values are expressed as the mean and 95% confidence interval of the 50 simulated datasets.*
Abbreviations: CI = Confidence Interval; HMM = Hidden Markov Model; ILD = Interstitial Lung Disease; $\varkappa$ = Kappa Coefficient

Appendix Table M-2: Kappa Coefficients Measuring the Agreement in Interstitial Lung Disease (ILD) Ascertainment of Systemic Sclerosis (SSc) Patients Between the 80% Cut-Off, 10% Decline and Hidden Markov Model (HMM) Algorithms.

Methods Compared	All Visits *		ILD-Negative Visits *		ILD-Positive Visits *	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
80% Cut-Off & 10% Decline	0.18	0.17, 0.19	0.18	0.18, 0.19	0.14	0.13, 0.16
Maximum $\varkappa$	0.62	0.61, 0.63	0.56	0.56, 0.57	0.84	0.82, 0.86
80% Cut-Off & HMM	0.25	0.21, 0.29	0.25	0.21, 0.28	0.21	0.17, 0.25
Maximum $\varkappa$	0.87	0.83, 0.91	0.88	0.84, 0.92	0.82	0.78, 0.85
10% Decline & HMM	0.24	0.22, 0.26	0.23	0.21, 0.25	0.27	0.23, 0.30
Maximum $\varkappa$	0.52	0.49, 0.55	0.48	0.45, 0.51	0.67	0.63, 0.71

**Values are expressed as the mean and 95% confidence interval of the 25 imputed datasets.*

Abbreviations: HMM = Hidden Markov Model; ILD = Interstitial Lung Disease; $\varkappa$ = Kappa Coefficient

Appendix N: Validating the Canadian Scleroderma Research Group Data Algorithms Using High-Resolution Computed Tomography as the Gold-Standard

When high-resolution computed tomography (HRCT) results alone were used as the gold-standard for determining interstitial lung disease (ILD) status, only 150 forced vital capacity (FVC) measures from 113 patients could be used to validate the hidden Markov model (HMM) and 80% cut-off and 10% decline in FVC algorithms (Appendix Table N-1).

The 10% decline algorithm had the highest overall sensitivity (48.2%), while the 80% cut-off algorithm had the greatest incident sensitivity (45.7%). The HMM had both the lowest overall (34.8%) and incident (29.4%) sensitivities. Specificity was highest for the HMM (84.3%). The positive predictive values (PPVs) were highest for the 80% cut-off algorithm (overall PPV: 74.5%; incident PPV: 69.5%). Finally, the 80% cut-off algorithm had the lowest error rates (ERs) (overall ER: 38.0%; incident ER: 34.8%).

It is worth noting that a generalizability issue arises when using only HRCT results as a gold-standard, as patients on whom HRCT scans are available tend to have symptoms of lung involvement and/or worsened pulmonary function test results. Indeed, when our analyses were restricted only to patients on which an HRCT scan was available, the prevalence of ILD (54.0%) surpassed what would be observed in a general SSc population. Predictably, this caused the positive predictive value to increase, while the negative predictive value decreased compared to the simulation study (Appendix Table N-1 vs. Table 5-1). However, when the combination gold-standard was used, the prevalence of ILD in the validation sample (19.2%) more closely resembled that in the simulated study (11.7%), resulting in

comparable positive predictive values and negative predictive values to those in the simulation study (Table 5-3 vs. Table 5-1).

Appendix Table N-1: Performance of the 80% Cut-Off, 10% Decline and Hidden Markov Model (HMM) Algorithms in Identifying Interstitial Lung Disease (ILD) Onset when Validating Using Only High-Resolution Computed Tomography (HRCT) Results in Patients Enrolled in the Canadian Scleroderma Research Group (CSRG) Registry.

	80% Cut-Off in FVC *		10% Decline in FVC *		Hidden Markov Model *	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	45.1	33.7, 56.5	48.2	35.2, 61.2	34.8	22.9, 46.7
Incident Sensitivity	45.7	32.7, 58.7	43.1	29.2, 56.9	29.4	16.1, 42.6
Prevalent Sensitivity	0.5	0, 2.8	5.2	0, 12.4	3.7	0, 9.2
Specificity	81.9	72.6, 91.1	73.9	61.7, 86.2	84.3	74.7, 94.0
Overall PPV	74.5	62.0, 87.0	68.5	55.0, 82.0	72.3	56.3, 88.3
Incident PPV	69.5	54.6, 84.4	60.4	44.1, 76.8	65.0	45.3, 84.6
NPV	56.0	46.1, 65.8	54.9	44.1, 65.7	52.4	42.7, 62.2
Overall +LR †	2.5	1.8, 3.2	1.9	1.1, 2.6	2.3	0, Inf
Incident +LR †	2.5	1.9, 3.2	1.7	1.0, 2.4	1.9	0.9, 3.0
Overall -LR †	0.67	0.44, 0.90	0.70	0.43, 0.98	0.77	0, Inf
Incident -LR †	0.66	0.40, 0.93	0.77	0.49, 1.06	0.84	0.62, 1.05
Overall ER	38.0	29.9, 46.1	40.0	30.9, 49.0	42.4	33.4, 51.4
Incident ER	34.8	26.4, 43.3	40.3	30.7, 49.9	41.0	31.5, 50.5

**Values are expressed as the mean and 95% confidence interval of the 25 imputed datasets.*

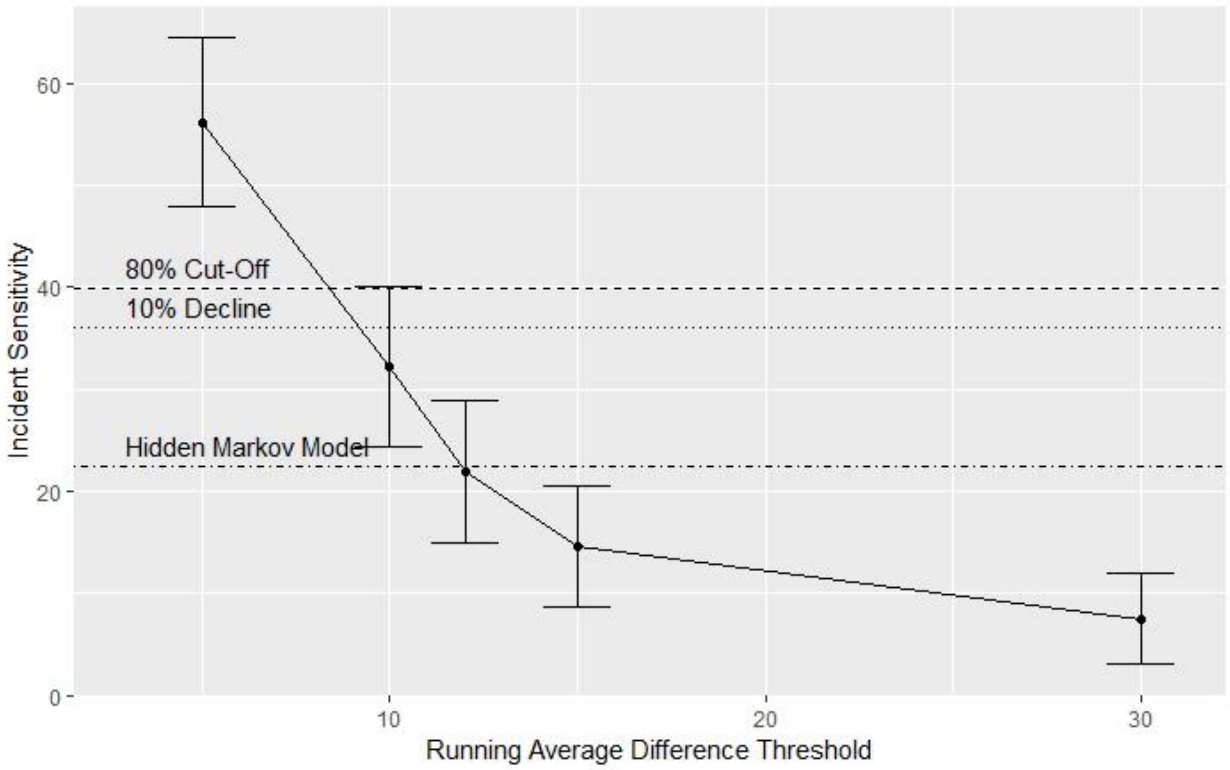
†All measures are expressed as percentages, except for the positive and negative likelihood ratios.

Abbreviations: FVC = Forced Vital Capacity; Inf = Infinity; NPV = Negative Predictive Value; PPV = Positive Predictive Value; +LR = Positive Likelihood Ratio; -LR = Negative Likelihood Ratio

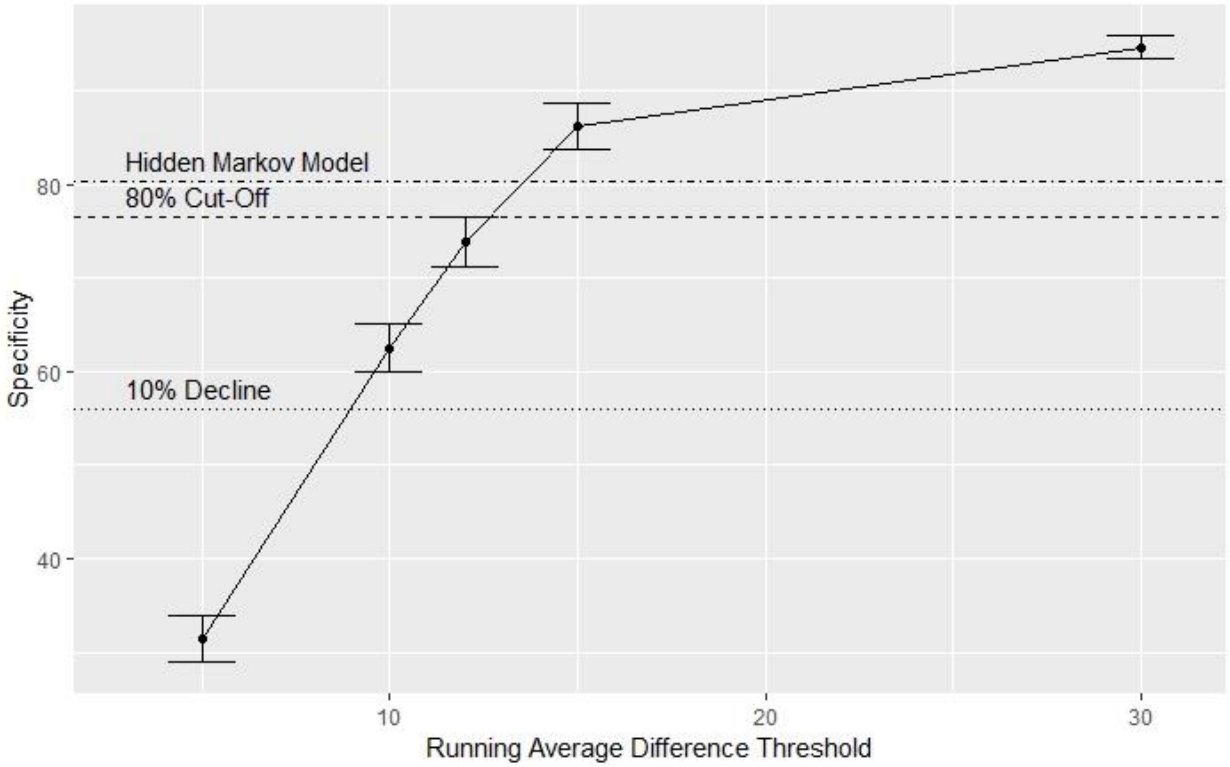
Appendix O: Effect of Varying the Running Average Threshold to Detect Interstitial Lung Disease Onset

Appendix Figure O-1 illustrates the effect of varying the threshold for the difference between a systemic sclerosis patient's current forced vital capacity value and their running average to detect interstitial lung disease onset on a) incident sensitivity and b) specificity.

a)



b)



Appendix Figure O-1: Result of Varying the Threshold for the Difference Between a Systemic Sclerosis (SSc) Patient's Current Forced Vital Capacity (FVC) Value and their Running Average to Detect Interstitial Lung Disease (ILD) Onset on a) Incident Sensitivity and b) Specificity.

Appendix P: Performance Metrics of the Different Hidden Markov Models

The tables in this appendix present the measures of diagnostic test accuracy of the different pulmonary function test (PFT) measures when used as univariate or multivariate outcomes in hidden Markov models (HMMs) for the detection of interstitial lung disease (ILD) in patients enrolled in the Canadian Scleroderma Research Group (CSRG) registry. Appendix Table P-2 outlines the results for % predicted PFT measures, while Appendix Table P-3 reports the results for absolute PFT measures. Appendix Table P-1 shows the measures of validity for patient-reported shortness of breath and for the New York Heart Association (NYHA) functional class as a comparator to better contextualize the performance of PFT measures.

Appendix Table P-1: Performance of Hidden Markov Models (HMMs) Using Shortness of Breath in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	Patient-Reported Shortness of Breath		NYHA Functional Class	
	Mean	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	32.8	28.9, 36.6	63.2	56.7, 69.6
Incident Sensitivity	27.6	21.0, 34.3	27.9	20.1, 35.6
Prevalent Sensitivity *	12.2	5.8, 18.6	56.1	48.8, 63.4
Specificity	77.1	72.7, 81.5	55.7	49.9, 61.4
Overall Positive Predictive Value	27.2	23.2, 31.3	27.1	25.0, 29.2
Incident Positive Predictive Value	19.9	14.4, 25.5	23.0	19.3, 26.7
Negative Predictive Value	81.5	79.9, 83.0	85.3	83.6, 87.0
Overall Positive Likelihood Ratio †	1.4	0, Inf	1.4	0, Inf
Incident Positive Likelihood Ratio †	1.2	0.9, 1.6	0.6	0.4, 0.9
Overall Negative Likelihood Ratio †	0.87	0, Inf	0.66	0, Inf
Incident Negative Likelihood Ratio †	0.94	0.83, 1.05	1.30	1.19, 1.40
Overall Error Rate	32.1	28.7, 35.4	42.8	39.2, 46.3
Incident Error Rate	25.8	21.7, 30.0	46.0	40.8, 51.2
Brier Score †	0.278	0.273, 0.283	0.259	0.256, 0.262

**Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*
†All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.
Abbreviations: Inf = Infinity; NYHA = New York Heart Association

Appendix Table P-2: Performance of Hidden Markov Models (HMMs) Using % Predicted Pulmonary Function Test (PFT) Measures in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	FVC % Predicted		DLCO % Predicted		DLCOa % Predicted	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	26.0	18.5, 33.5	19.7	12.9, 26.5	20.0	13.5, 26.4
Incident Sensitivity	18.4	12.0, 24.8	13.5	4.8, 22.3	13.7	5.2, 22.1
Prevalent Sensitivity *	11.1	4.0, 18.3	8.3	3.1, 13.4	7.9	2.5, 13.3
Specificity	82.1	75.3, 88.9	87.0	80.1, 93.8	86.7	80.0, 93.4
Overall Positive Predictive Value	26.6	21.7, 31.4	27.6	20.3, 34.8	27.4	19.6, 35.2
Incident Positive Predictive Value	18.7	12.6, 24.9	20.3	11.1, 29.4	20.0	10.6, 29.4
Negative Predictive Value	81.8	80.4, 83.2	81.5	80.0, 83.0	81.5	79.9, 83.0
Overall Positive Likelihood Ratio †	1.5	0, Inf	1.6	0, Inf	1.5	0, Inf
Incident Positive Likelihood Ratio †	1.1	0.6, 1.5	1.1	0.5, 1.6	1.0	0.5, 1.6
Overall Negative Likelihood Ratio †	0.90	0, Inf	0.92	0, Inf	0.92	0, Inf
Incident Negative Likelihood Ratio †	1.00	0.91, 1.08	1.00	0.92, 1.07	1.00	0.92, 1.07
Overall Error Rate	29.0	24.7, 33.4	26.3	21.6, 31.0	26.5	21.7, 31.3
Incident Error Rate	21.4	15.1, 27.8	17.1	10.9, 23.3	17.4	11.3, 23.4
Brier Score †	0.243	0.237, 0.249	0.222	0.215, 0.230	0.225	0.217, 0.232

**Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*

† All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; DLCOa = Diffusing Capacity for Carbon Monoxide Adjusted for Pulmonary Hypertension; FVC = Forced Vital Capacity; Inf = Infinity

Appendix Table P-2 (continued): Performance of Hidden Markov Models (HMMs) Using % Predicted Pulmonary Function Test (PFT) Measures in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	TLC % Predicted		TLC % Predicted & FEV_1/FVC	
	Mean	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	27.2	22.5, 31.9	19.3	14.7, 23.9
Incident Sensitivity	17.4	11.4, 23.4	15.1	9.3, 20.8
Prevalent Sensitivity *	10.7	4.9, 16.6	8.6	3.4, 13.9
Specificity	82.8	79.0, 86.6	88.8	85.5, 92.1
Overall Positive Predictive Value	28.1	23.7, 32.5	29.9	24.1, 35.8
Incident Positive Predictive Value	19.7	13.4, 25.9	19.2	11.8, 26.7
Negative Predictive Value	82.2	80.8, 83.6	81.7	80.3, 83.1
Overall Positive Likelihood Ratio †	1.6	0, Inf	1.7	0, Inf
Incident Positive Likelihood Ratio †	1.0	0.6, 1.4	1.4	0.8, 1.9
Overall Negative Likelihood Ratio †	0.88	0, Inf	0.91	0, Inf
Incident Negative Likelihood Ratio †	1.00	0.92, 1.08	0.96	0.89, 1.03
Overall Error Rate	28.2	25.4, 31.0	24.9	22.5, 27.4
Incident Error Rate	20.8	17.2, 24.4	15.3	12.1, 18.4
Brier Score †	0.236	0.232, 0.240	0.217	0.215, 0.220

**Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*

†All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.

Abbreviations: FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; Inf = Infinity; TLC = Total Lung Capacity

Appendix Table P-2 (continued): Performance of Hidden Markov Models (HMMs) Using % Predicted Pulmonary Function Test (PFT) Measures in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	FVC & DLCO % Predicted		FVC & TLC % Predicted		DLCO & TLC % Predicted		FVC, DLCO & TLC % Predicted	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	29.6	22.1, 37.0	30.0	26.1, 33.9	28.0	23.3, 32.8	30.6	26.3, 35.0
Incident Sensitivity	20.0	12.2, 27.9	20.3	14.3, 26.3	18.5	12.0, 25.0	22.2	15.8, 28.6
Prevalent Sensitivity [*]	14.4	8.1, 20.7	15.3	8.9, 21.8	12.1	6.3, 17.9	15.8	9.4, 22.1
Specificity	81.7	76.9, 86.6	78.5	75.5, 81.6	81.7	77.6, 85.7	78.2	75.1, 81.3
Overall Positive Predictive Value	28.6	24.5, 32.6	25.6	22.3, 29.0	27.5	23.0, 32.0	25.7	22.5, 29.0
Incident Positive Predictive Value	21.3	15.1, 27.5	18.2	13.1, 23.4	19.8	13.8, 25.7	18.4	13.3, 23.5
Negative Predictive Value	82.5	81.0, 84.0	82.0	80.6, 83.4	82.2	80.7, 83.6	82.1	80.6, 83.5
Overall Positive Likelihood Ratio [†]	1.6	0, Inf	1.4	0, Inf	1.5	0, Inf	1.4	0, Inf
Incident Positive Likelihood Ratio [†]	1.1	0.8, 1.5	1.0	0.6, 1.3	1.0	0.7, 1.4	1.0	0.7, 1.3
Overall Negative Likelihood Ratio [†]	0.86	0, Inf	0.89	0, Inf	0.88	0, Inf	0.89	0, Inf
Incident Negative Likelihood Ratio [†]	0.98	0.90, 1.06	1.01	0.93, 1.10	1.00	0.92, 1.08	1.00	0.91, 1.08
Overall Error Rate	28.6	25.6, 31.6	31.1	28.7, 33.4	28.9	25.9, 31.9	31.2	28.9, 33.5
Incident Error Rate	21.7	17.3, 26.1	24.7	21.8, 27.6	21.8	18.0, 25.6	24.9	21.9, 27.8
Brier Score [†]	0.240	0.236, 0.245	0.263	0.260, 0.266	0.242	0.238, 0.247	0.265	0.262, 0.268

[*]*Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*

[†]*All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.*

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FVC = Forced Vital Capacity; Inf = Infinity; TLC = Total Lung Capacity

Appendix Table P-3: Performance of Hidden Markov Models (HMMs) Using Absolute Pulmonary Function Test (PFT) Measures in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	FVC Absolute		DLCO Absolute		DLCOa Absolute	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	27.1	22.4, 31.8	25.1	19.5, 30.7	24.9	19.3, 30.5
Incident Sensitivity	17.9	11.5, 24.3	17.1	11.0, 23.3	16.5	10.4, 22.6
Prevalent Sensitivity *	14.2	7.2, 21.3	11.4	6.1, 16.8	11.1	5.8, 16.5
Specificity	83.1	75.6, 90.5	82.2	78.2, 86.3	82.3	78.3, 86.2
Overall Positive Predictive Value	28.8	21.1, 36.5	25.9	21.0, 30.7	25.7	20.8, 30.7
Incident Positive Predictive Value	20.7	13.7, 27.7	18.3	11.7, 24.9	17.9	11.5, 24.2
Negative Predictive Value	82.2	80.7, 83.7	81.7	80.1, 83.2	81.6	80.0, 83.2
Overall Positive Likelihood Ratio †	1.7	0, Inf	1.4	0, Inf	1.4	0, Inf
Incident Positive Likelihood Ratio †	1.1	0.6, 1.6	1.0	0.6, 1.3	0.9	0.6, 1.3
Overall Negative Likelihood Ratio †	0.88	0, Inf	0.91	0, Inf	0.91	0, Inf
Incident Negative Likelihood Ratio †	0.99	0.89, 1.09	1.01	0.93, 1.09	1.01	0.94, 1.10
Overall Error Rate	28.0	22.5, 33.5	29.1	26.0, 32.1	29.1	26.1, 32.1
Incident Error Rate	20.5	13.6, 27.5	21.4	17.6, 25.2	21.4	17.7, 25.0
Brier Score †	0.230	0.223, 0.238	0.243	0.240, 0.246	0.244	0.240, 0.247

**Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*

† All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; DLCOa = Diffusing Capacity for Carbon Monoxide Adjusted for Pulmonary Hypertension; FVC = Forced Vital Capacity; Inf = Infinity

Appendix Table P-3 (continued): Performance of Hidden Markov Models (HMMs) Using Absolute Pulmonary Function Test (PFT) Measures in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	TLC Absolute		TLC Absolute & FEV_1/FVC	
	Mean	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	24.5	20.2, 28.9	21.5	16.6, 26.5
Incident Sensitivity	15.8	9.6, 21.9	15.4	9.4, 21.4
Prevalent Sensitivity *	10.1	5.0, 15.2	11.7	6.1, 17.3
Specificity	86.0	82.3, 89.8	86.7	83.1, 90.3
Overall Positive Predictive Value	30.4	24.9, 35.9	28.6	23.8, 33.5
Incident Positive Predictive Value	21.0	14.1, 27.9	19.5	12.7, 26.3
Negative Predictive Value	82.2	80.8, 83.6	81.8	80.4, 83.1
Overall Positive Likelihood Ratio †	1.8	0, Inf	1.6	0, Inf
Incident Positive Likelihood Ratio †	1.1	0.6, 1.6	1.2	0.7, 1.6
Overall Negative Likelihood Ratio †	0.88	0, Inf	0.91	0, Inf
Incident Negative Likelihood Ratio †	0.98	0.89, 1.06	0.98	0.91, 1.05
Overall Error Rate	26.1	23.3, 29.0	26.2	23.6, 28.7
Incident Error Rate	17.8	14.2, 21.4	17.2	13.9, 20.6
Brier Score †	0.218	0.214, 0.222	0.226	0.223, 0.229

**Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*

†All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.

Abbreviations: FEV_1 = Forced Expiratory Volume in One Second; FVC = Forced Vital Capacity; Inf = Infinity; TLC = Total Lung Capacity

Appendix Table P-3 (continued): Performance of Hidden Markov Models (HMMs) Using Absolute Pulmonary Function Test (PFT) Measures in Identifying Interstitial Lung Disease (ILD) Onset in Patients Enrolled in the Canadian Scleroderma Research Group Registry (CSRG).

	FVC & DLCO Absolute		FVC & TLC Absolute		DLCO & TLC Absolute		FVC, DLCO & TLC Absolute	
	Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Overall Sensitivity	32.5	27.1, 37.8	32.9	28.1, 37.7	31.4	26.6, 36.2	34.4	29.2, 39.7
Incident Sensitivity	20.4	13.7, 27.1	20.6	14.1, 27.0	20.4	13.3, 27.5	21.1	14.5, 27.8
Prevalent Sensitivity *	19.9	12.6, 27.2	19.1	12.1, 26.0	15.9	8.7, 23.1	22.5	14.8, 30.2
Specificity	79.6	74.9, 84.3	80.5	77.2, 83.8	81.0	77.0, 84.9	78.4	74.6, 82.2
Overall Positive Predictive Value	28.3	24.0, 32.6	29.4	24.9, 33.9	29.0	24.3, 33.6	28.2	23.9, 32.5
Incident Positive Predictive Value	21.1	14.9, 27.3	20.4	14.8, 26.0	20.0	13.4, 26.7	20.4	15.0, 25.9
Negative Predictive Value	82.7	81.2, 84.2	83.0	81.4, 84.5	82.7	81.2, 84.2	82.9	81.3, 84.5
Overall Positive Likelihood Ratio †	1.6	0, Inf	1.7	0, Inf	1.7	0, Inf	1.6	0, Inf
Incident Positive Likelihood Ratio †	1.0	0.7, 1.4	1.1	0.7, 1.4	1.1	0.7, 1.4	1.0	0.7, 1.3
Overall Negative Likelihood Ratio †	0.85	0, Inf	0.83	0, Inf	0.85	0, Inf	0.84	0, Inf
Incident Negative Likelihood Ratio †	1.00	0.91, 1.09	0.99	0.90, 1.07	0.98	0.90, 1.07	1.01	0.92, 1.10
Overall Error Rate	29.7	26.3, 33.1	28.9	26.2, 31.6	28.8	25.8, 31.9	30.3	27.3, 33.3
Incident Error Rate	23.6	19.3, 28.0	22.8	19.7, 25.9	22.4	18.7, 26.0	24.8	21.2, 28.3
Brier Score †	0.250	0.245, 0.255	0.245	0.241, 0.249	0.242	0.238, 0.246	0.260	0.255, 0.265

**Corresponds to the proportion of patients with interstitial lung disease whose disease was not detected at onset but identified subsequently during follow-up.*

†All measures are expressed as percentages except for the positive and negative likelihood ratios and Brier scores.

Abbreviations: DLCO = Diffusing Capacity for Carbon Monoxide; FVC = Forced Vital Capacity; Inf = Infinity; TLC = Total Lung Capacity

References

1. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med*. 2009;360(19):1989-2003.
2. Wigley FM, Boin F. Chapter 84: Clinical features and treatment of scleroderma. In: Firestein GS, Budd RC, Gabriel SE, McInnes IB, O'Dell JR, eds. Kelley and Firestein's Textbook of Rheumatology. 10th ed. Philadelphia, PA: Elsevier; 2017. p. 1424-60.
3. Cappelli S, Bellando Randone S, Camiciottoli G, De Paulis A, Guiducci S, Matucci-Cerinic M. Interstitial lung disease in systemic sclerosis: where do we stand? *Eur Respir Rev*. 2015;24(137):411-9.
4. Iudici M, Moroncini G, Cipriani P, Giacomelli R, Gabrielli A, Valentini G. Where are we going in the management of interstitial lung disease in patients with systemic sclerosis? *Autoimmun Rev*. 2015;14(7):575-8.
5. Silver KC, Silver RM. Management of systemic-sclerosis-associated interstitial lung disease. *Rheum Dis Clin North Am*. 2015;41(3):439-57.
6. Cappelli S, Guiducci S, Bellando Randone S, Matucci Cerinic M. Immunosuppression for interstitial lung disease in systemic sclerosis. *Eur Respir Rev*. 2013;22(129):236-43.
7. Winklehner A, Berger N, Maurer B, Distler O, Alkadhi H, Frauenfelder T. Screening for interstitial lung disease in systemic sclerosis: the diagnostic accuracy of HRCT image series with high increment and reduced number of slices. *Ann Rheum Dis*. 2012;71(4):549-52.
8. Frauenfelder T, Winklehner A, Nguyen TD, Dobrota R, Baumueller S, Maurer B, et al. Screening for interstitial lung disease in systemic sclerosis: performance of high-resolution CT with limited number of slices: a prospective study. *Ann Rheum Dis*. 2014;73(12):2069-73.
9. Nguyen-Kim TDL, Maurer B, Suliman YA, Morsbach F, Distler O, Frauenfelder T. The impact of slice-reduced computed tomography on histogram-based densitometry assessment of lung fibrosis in patients with systemic sclerosis. *J Thorac Dis*. 2018;10(4):2142-52.

10. Dakin J, Kourteli E, Winter R. Making sense of lung function tests: a hands-on guide. Koster J, Rooker W, eds. 1st ed. New York, NY: Oxford University Press Inc.; 2003.

11. Khanna D, Brown KK, Clements PJ, Elashoff R, Furst DE, Goldin J, et al. Systemic sclerosis-associated interstitial lung disease – proposed recommendations for future randomized clinical trials. *Clin Exp Rheumatol.* 2010;28(2 Suppl 58):S55-62.

12. Boin F, Hummers LK. Scleroderma-like fibrosing disorders. *Rheum Dis Clin North Am.* 2008;34(1):199-220.

13. Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. *Annu Rev Pathol.* 2011;6:509-37.

14. Denton CP, Black CM, Abraham DJ. Mechanisms and consequences of fibrosis in systemic sclerosis. *Nat Clin Pract Rheumatol.* 2006;2(3):134-44.

15. Van Den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum.* 2013;65(11):2737-47.

16. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, Jr., et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol.* 1988;15(2):202-5.

17. Young A, Khanna D. Systemic sclerosis: commonly asked questions by rheumatologists. *J Clin Rheumatol.* 2015;21(3):149-55.

18. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. *Semin Arthritis Rheum.* 2008;37(4):223-35.

19. Thompson AE, Pope JE. Increased prevalence of scleroderma in southwestern Ontario: a cluster analysis. *J Rheumatol.* 2002;29(9):1867-73.

20. Bernatsky S, Joseph L, Pineau CA, Belisle P, Hudson M, Clarke AE. Scleroderma prevalence: demographic variations in a population-based sample. *Arthritis Rheum.* 2009;61(3):400-4.

21. Scleroderma Association of British Columbia. Scleroderma overview [Internet]. [cited 4 March 2019]. Available from: https://www.sclerodermabc.ca/overview.

22. Khananian A, Hudson M, Baron M, Pope J. Mortality in Canadian systemic sclerosis patients: cause for alarm. *J Rheumatol.* 2014;41(7):1514-5.

23. Mayes MD, Lacey JV, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum.* 2003;48(8):2246-55.

24. Wilson L. Cost-of-illness of scleroderma: the case for rare diseases. *Semin Arthritis Rheum.* 1997;27(2):73-84.

25. Bernatsky S, Hudson M, Panopalis P, Clarke AE, Pope J, Leclercq S, et al. The cost of systemic sclerosis. *Arthritis Rheum.* 2009;61(1):119-23.

26. Cottin V, Brown KK. Interstitial lung disease associated with systemic sclerosis (SSc-ILD). *Respir Res.* 2019;20(1):13.

27. Bussone G, Mouthon L. Interstitial lung disease in systemic sclerosis. *Autoimmun Rev.* 2011;10(5):248-55.

28. Schoenfeld SR, Castelino FV. Interstitial lung disease in scleroderma. *Rheum Dis Clin North Am.* 2015;41(2):237-48.

29. Steele R, Hudson M, Lo E, Baron M. Clinical decision rule to predict the presence of interstitial lung disease in systemic sclerosis. *Arthritis Care Res (Hoboken).* 2012;64(4):519-24.

30. Schneider PD, Wise RA, Hochberg MC, Wigley FM. Serial pulmonary function in systemic sclerosis. *Am J Med.* 1982;73(3):385-94.

31. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med.* 2006;354(25):2655-66.

32. Hoyles RK, Ellis RW, Wellsbury J, Lees B, Newlands P, Goh NS, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum.* 2006;54(12):3962-70.

33. Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. *Lancet Respir Med.* 2016;4(9):708-19.

34. Distler O, Highland KB, Gahlemann M, Azuma A, Fischer A, Mayes MD, et al. Nintedanib for systemic sclerosis-associated interstitial lung disease. *N Engl J Med.* 2019;380(26):2518-28.

35. Bernstein EJ, Khanna D, Lederer DJ. Screening high-resolution computed tomography of the chest to detect interstitial lung disease in systemic sclerosis: a global survey of rheumatologists. *Arthritis Rheumatol.* 2018;70(6):971-2.

36. Fischer A, Strek ME, Cottin V, Dellaripa PF, Bernstein EJ, Brown KK, et al. Proceedings of the American College of Rheumatology/Association of Physicians of Great Britain and Ireland Connective Tissue Disease-Associated Interstitial Lung Disease Summit: a multidisciplinary approach to address challenges and opportunities. *Arthritis Rheumatol.* 2019;71(2):182-95.

37. Launay D, Remy-Jardin M, Michon-Pasturel U, Mastora I, Hachulla E, Lambert M, et al. High resolution computed tomography in fibrosing alveolitis associated with systemic sclerosis. *J Rheumatol.* 2006;33(9):1789-801.

38. Monnier-Cholley L, Riviere S, Barikalla AO, Mahevas T, Tiev KP, Arrive L. Systemic sclerosis: annual chest CT is not necessary for all patients. *Presse Med.* 2018;47(1):90-2.

39. Aggarwal AN, Agarwal R. The new ATS/ERS guidelines for assessing the spirometric severity of restrictive lung disease differ from previous standards. *Respirology*. 2007;12(5):759-62.

40. Cotes J, Chinn D, Quanjer PH, Roca J, Yernault J-C. Standardization of the measurement of transfer factor (diffusing capacity). *Eur Respir J*. 1993;6 Suppl 16:41-52.

41. Quanjer PH, Tammeling G, Cotes J, Pedersen O, Peslin R, Yernault J. Lung volumes and forced ventilatory flows. *Eur Respir J*. 1993;6 Suppl 16:5-40.

42. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005;26(5):948-68.

43. Suliman YA, Dobrota R, Huscher D, Nguyen-Kim TD, Maurer B, Jordan S, et al. Brief report: pulmonary function tests: high rate of false-negative results in the early detection and screening of scleroderma-related interstitial lung disease. *Arthritis Rheumatol*. 2015;67(12):3256-61.

44. Jee AS, Sahhar J, Youssef P, Bleasel J, Adelstein S, Nguyen M, et al. Review: serum biomarkers in idiopathic pulmonary fibrosis and systemic sclerosis associated interstitial lung disease – frontiers and horizons. *Pharmacol Ther*. 2019;202:40-52.

45. Wells AU, Behr J, Silver R. Outcome measures in the lung. *Rheumatology (Oxford)*. 2008;47 Suppl 5:v48-50.

46. Winstone TA, Assayag D, Wilcox PG, Dunne JV, Hague CJ, Leipsic J, et al. Predictors of mortality and progression in scleroderma-associated interstitial lung disease: a systematic review. *Chest*. 2014;146(2):422-36.

47. Kafaja S, Clements PJ, Wilhalme H, Tseng CH, Furst DE, Kim GH, et al. Reliability and minimal clinically important differences of forced vital capacity: results from the Scleroderma Lung Studies (SLS-I and SLS-II). *Am J Respir Crit Care Med*. 2018;197(5):644-52.

48. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319-38.

49. Man A, Davidyock T, Ferguson LT, Ieong M, Zhang Y, Simms RW. Changes in forced vital capacity over time in systemic sclerosis: application of group-based trajectory modelling. *Rheumatology (Oxford)*. 2015;54(8):1464-71.

50. Kim HJ, Tashkin DP, Gjertson DW, Brown MS, Kleerup E, Chong S, et al. Transitions to different patterns of interstitial lung disease in scleroderma with and without treatment. *Ann Rheum Dis*. 2016;75(7):1367-71.

51. Ciancio N, Pavone M, Torrisi SE, Vancheri A, Sambataro D, Palmucci S, et al. Contribution of pulmonary function tests (PFTs) to the diagnosis and follow up of connective tissue diseases. *Multidiscip Respir Med*. 2019;14:17.

52. Canadian Scleroderma Research Group. Who we are – Qui sommes-nous [Internet]. 2015 [cited 15 April 2019]. Available from: http://www.canadiansclerodermaresearchgroup.org/.

53. Alhajeri H, Hudson M, Fritzler M, Pope J, Tatibouet S, Markland J, et al. 2013 American College of Rheumatology/European League Against Rheumatism classification criteria for systemic sclerosis outperform the 1980 criteria: data from the Canadian Scleroderma Research Group. *Arthritis Care Res (Hoboken)*. 2015;67(4):582-7.

54. American Heart Association. Classes of heart failure [Internet]. 2017 [cited 6 August 2019]. Available from: https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure/classes-of-heart-failure.

55. O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. *BMJ Evid Based Med*. 2018;23(2):54-5.

56. Sellares J, Hernandez-Gonzalez F, Lucena CM, Paradela M, Brito-Zeron P, Prieto-Gonzalez S, et al. Auscultation of velcro crackles is associated with usual interstitial pneumonia. *Medicine (Baltimore)*. 2016;95(5):e2573.

57. Sgalla G, Walsh SLF, Sverzellati N, Fletcher S, Cerri S, Dimitrov B, et al. "Velcro-type" crackles predict specific radiologic features of fibrotic interstitial lung disease. *BMC Pulm Med.* 2018;18(1):103.

58. Hax V, Bredemeier M, Didonet Moro AL, Pavan TR, Vieira MV, Pitrez EH, et al. Clinical algorithms for the diagnosis and prognosis of interstitial lung disease in systemic sclerosis. *Semin Arthritis Rheum.* 2017;47(2):228-34.

59. Ibe OC. Markov processes for stochastic modeling. 2nd ed. Oxford: Elsevier, 2013. Chapter 2: Basic concepts in stochastic processes; p.29-48.

60. Ibe OC. Markov processes for stochastic modeling. 2nd ed. Oxford: Elsevier, 2013. Chapter 3: Introduction to Markov processes; p.49-57.

61. Ibe OC. Markov processes for stochastic modeling. 2nd ed. Oxford: Elsevier, 2013. Chapter 4: Discrete-time Markov Chains; p.59-84.

62. Ibe OC. Markov processes for stochastic modeling. 2nd ed. Oxford: Elsevier, 2013. Chapter 5: Continuous-time Markov chains; p.85-102.

63. Jackson CH, Sharples LD, Thompson SG, Duffy SW, Couto E. Multistate Markov models for disease progression with classification error. *J R Stat Soc, Ser D Stat.* 2003;52(2):193-209.

64. Jackson CH, Sharples LD. Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Stat Med.* 2002;21(1):113-28.

65. Jackson C. Multi-state modelling with R: the msm package. Cambridge, UK. 2007 Oct 1.

66. Ibe OC. Markov processes for stochastic modeling. 2nd ed. Oxford: Elsevier, 2013. Chapter 14: Hidden Markov models; p.417-51.

67. Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Stat.* 1966;37(6):1554-63.

68. Seibold J. Scleroderma. In: Harris ED, Budd RC, Firestein GS, Genovese MC, Sergent JS, Ruddy S, Sledge CB, eds. Kelley's Textbook of Rheumatology. 7th ed. Philadelphia PA, Elsevier; 2005. p.1279-1308.

69. Khanna D, Denton CP. Evidence-based management of rapidly progressing systemic sclerosis. *Best Pract Res Clin Rheumatol.* 2010;24(3):387-400.

70. Baldwin EDeF, Cournand A, Richards DW Jr. Pulmonary insufficiency; a study of 39 cases of pulmonary fibrosis. *Medicine (Baltimore).* 1949;28(1):1-25.

71. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-36.

72. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.

73. Konig G, Luderschmidt C, Hammer C. Lung involvement in scleroderma. *Chest.* 1984;85(3):318-24.

74. Peters-Golden M, Wise RA, Schneider P. Clinical and demographic predictors of loss of pulmonary function in systemic sclerosis. *Medicine (Baltimore).* 1984;63(4):221-31.

75. Steen VD, Owens GR, Redmond C, Rodnan GP, Medsger TA Jr. The effect of D-penicillamine on pulmonary findings in systemic sclerosis. *Arthritis Rheum.* 1985;28(8):882-8.

76. De Clerck LS, Dequeker J, Francx L, Demedts M. D-Penicillamine therapy and interstitial lung disease in scleroderma. A long-term followup study. *Arthritis Rheum.* 1987;30(6):643-50.

77. Greenwald GI, Tashkin DP, Gong H, Simmons M, Duann S, Furst DE, et al. Longitudinal changes in lung function and respiratory symptoms in progressive systemic sclerosis. Prospective study. *Am J Med.* 1987;83(1):83-92.

78. McCarthy DS, Baragar FD, Dhingra S, Sigurdson M, Sutherland JB, Rigby M, et al. The lungs in systemic sclerosis (scleroderma): a review and new information. *Semin Arthritis Rheum*. 1988;17(4):271-83.

79. Zarafonetis CJD, Dabich L, Devol EB, Skovronski JJ, Negri D, Yuan W. Retrospective studies in scleroderma: pulmonary findings and effect of potassium p-aminobenzoate on vital capacity. *Respiration*. 1989;56(1-2):22-33.

80. Silver RM, Miller KS, Kinsella MB, Smith EA, Schabel SI. Evaluation and management of scleroderma lung disease using bronchoalveolar lavage. *Am J Med*. 1990;88(5):470-6.

81. Abramson MJ, Barnett AJ, Littlejohn GO, Smith MM, Hall S. Lung function abnormalities and decline of spirometry in scleroderma: an overrated danger? *Postgrad Med J*. 1991;67(789):632-7.

82. Wells AU, Hansell DM, Rubens MB, Cullinan P, Black CM, du Bois RM. The predictive value of appearances on thin-section computed tomography in fibrosing alveolitis. *Am Rev Respir Dis*. 1993;148(4):1076-82.

83. Wells AU, Hansell DM, Harrison NK, Lawrence R, Black CM, du Bois RM. Clearance of inhaled 99mTc-DTPA predicts the clinical course of fibrosing alveolitis. *Eur Respir J*. 1993;6(6):797-802.

84. Wells AU, Rubens MB, du Bois RM, Hansell DM. Serial CT in fibrosing alveolitis: prognostic significance of the initial pattern. *AJR Am J Roentgenol*. 1993;161(6):1159-65.

85. Dujic Z, Eterovic D, Tocilj J, Andjelinovic D. Increase of pulmonary diffusing capacity in systemic sclerosis. *Br J Rheumatol*. 1994;33(5):437-41.

86. Steen VD, Conte C, Owens GR, Medsger TA Jr. Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum*. 1994;37(9):1283-89.

87. Steen VD, Lanz JK Jr, Conte C, Owens GR, Medsger TA Jr. Therapy for severe interstitial lung disease in systemic sclerosis: a retrospective study. *Arthritis Rheum*. 1994;37(9):1290-6.

88. Tashkin DP, Clements PJ, Wright RS, Gong H Jr, Simmons MS, Lachenbruch PA, et al. Interrelationships between pulmonary and extrapulmonary involvement in systemic sclerosis. A longitudinal analysis. *Chest*. 1994;105(2):489-95.

89. Behr J, Adelmann-Grill BC, Hein R, Beinert T, Schwaiblmair M, Krombach F, Fruhmann G. Pathogenetic and clinical significance of fibroblast activation in scleroderma lung disease. *Respiration*. 1995;62(4):209-16.

90. Behr J, Vogelmeier C, Beinert T, Meurer T, Krombach F, Konig G, Fruhmann G. Bronchoalveolar lavage for evaluation and management of scleroderma disease of the lung. *Am J Respir Crit Care Med*. 1996;154(2):400-6.

91. Jacobsen S, Halberg P, Ullman S, Hoier-Madsen M, Petersen J, Mortensen J, et al. A longitudinal study of pulmonary function in Danish patients with systemic sclerosis. *Clin Rheumatol*. 1997;16(4):384-90.

92. Greidinger EL, Flaherty KT, White B, Rosen A, Wigley FM, Wise RA. African-American race and antibodies to topoisomerase I are associated with increased severity of scleroderma lung disease. *Chest*. 1998;114(3):801-7.

93. Atamas SP, Yurovsky VV, Wise R, Wigley FM, Goter Robinson CJ, Henry P, et al. Production of type 2, cytokines by CD8+ lung cells is associated with greater decline in pulmonary function in patients with systemic sclerosis. *Arthritis Rheum*. 1999;42(6):1168-78.

94. Kon OM, Daniil Z, Black CM, du Bois RM. Clearance of inhaled technetium-99m-DTPA as a clinical index of pulmonary vascular disease in systemic sclerosis. *Eur Respir J*. 1999;13(1):133-6.

95. Witt C, Borges AC, John M, Fietze I, Baumann G, Krause A. Pulmonary involvement in diffuse cutaneous systemic sclerosis: broncheoalveolar fluid granulocytosis predicts progression of fibrosing alveolitis. *Ann Rheum Dis*. 1999;58(10):635-40.

96. White B, Moore WC, Wigley FM, Xiao, HQ, Wise RA. Cyclophosphamide is associated with pulmonary function and survival benefit in patients with scleroderma and alveolitis. *Ann Intern Med*. 2000;132(12):947-54.

97. Yuhara T, Takemura H, Akama T, Yamane K, Sumida T. The relationship between serum immunoglobulin levels and pulmonary involvement in systemic sclerosis. *J Rheumatol*. 2000;27(5):1207-14.

98. Marie I, Dominique S, Levesque H, Ducrott P, Denis P, Hellot MF, et al. Esophageal involvement and pulmonary manifestations in systemic sclerosis. *Arthritis Care Res (Hoboken)*. 2001;45(4):346-54.

99. Scorza R, Caronni M, Mascagni B, Berruti V, Bazzi S, Micallef E, et al. Effects of long-term cyclic iloprost therapy in systemic sclerosis with Raynaud's phenomenon. A randomized, controlled study. *Clin Exp Rheumatol*. 2001;19(5):503-8.

100. Bouros D, Wells AU, Nicholson AG, Colby TV, Polychronopoulos V, Pantelidis P, et al. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am J Respir Crit Care Med*. 2002;165(12):1581-6.

101. Giacomelli R, Valentini G, Salsano F, Cipriani P, Sambo P, Conforti M, et al. Cyclophosphamide pulse regimen in the treatment of alveolitis in systemic sclerosis. *J Rheumatol*. 2002;29(4):731-6.

102. Pakas I, Ioannidis JPA, Malagari K, Skopouli FN, Moutsopoulos HM, Vlachoyiannopoulos PG. Cyclophosphamide with low or high dose prednisolone for systemic sclerosis lung disease. *J Rheumatol*. 2002;29(2):298-304.

103. Kowal-Bielecka O, Distler O, Kowal K, Siergiejko Z, Chwiecko J, Sulik A, et al. Elevated levels of leukotriene B4 and leukotriene E4 in bronchoalveolar lavage fluid from patients with scleroderma lung disease. *Arthritis Rheum*. 2003;48(6):1639-46.

104. Yanaba K, Hasegawa M, Hamaguchi Y, Fujimoto M, Takehara K, Sato S. Longitudinal analysis of serum KL-6 levels in patients with systemic sclerosis: association with the activity of pulmonary fibrosis. *Clin Exp Rheumatol*. 2003;21(4):429-36.

105. Airo P, Danieli E, Parrinello G, Antonioli CM, Cavazzana I, Toniati P, et al. Intravenous cyclophosphamide therapy for systemic sclerosis. A single-center experience and review of the literature with pooled analysis of lung function test results. *Clin Exp Rheumatol*. 2004;22(5):573-78.

106. Yanaba K, Hasegawa M, Takehara K, Sato S. Comparative study of serum surfactant protein-D and KL-6 concentrations in patients with systemic sclerosis as markers for monitoring the activity of pulmonary fibrosis. *J Rheumatol*. 2004;31(6):1112-20.

107. De Santis M, Bosello S, La Torre G, Capuano A, Tolusso B, Pagliari G, et al. Functional, radiological and biological markers of alveolitis and infections of the lower respiratory tract in patients with systemic sclerosis. *Respir Res*. 2005;6(1):96.

108. Kodera M, Hasegawa M, Komura K, Yanaba K, Takehara K, Sato S. Serum pulmonary and activation-regulated chemokine/CCL18 levels in patients with systemic sclerosis: a sensitive indicator of active pulmonary fibrosis. *Arthritis Rheum*. 2005;52(9):2889-96.

109. Kowal-Bielecka O, Kowal K, Rojewska J, Bodzenta-Lukaszyk A, Siergiejko Z, Sierakowska M, et al. Cyclophosphamide reduces neutrophilic alveolitis in patients with scleroderma lung disease: a retrospective analysis of serial bronchoalveolar lavage investigations. *Ann Rheum Dis*. 2005;64(9):1343-6.

110. Plastiras SC, Karadimitrakis SP, Ziakas PD, Vlachoyiannopoulos PG, Moutsopoulos HM, Tzelepis GE. Scleroderma lung: initial forced vital capacity as predictor of pulmonary function decline. *Arthritis Care Res (Hoboken)*. 2006;55(4):598-602.

111. Beretta L, Bertolotti F, Cappiello F, Barili M, Masciocchi M, Toussoun K, et al. Interleukin-1 gene complex polymorphisms in systemic sclerosis patients with severe restrictive lung physiology. *Hum Immunol*. 2007;68(7):603-9.

112. Beretta L, Caronni M, Raimondi M, Ponti A, Viscuso T, Origgi L, et al. Oral cyclophosphamide improves pulmonary function in scleroderma patients with fibrosing alveolitis: experience in one centre. *Clin Rheumatol*. 2007;26(2):168-72.

113. Clements PJ, Roth MD, Elashoff R, Tashkin DP, Goldin J, Silver RM, et al. Scleroderma lung study (SLS): differences in the presentation and course of patients with limited versus diffuse systemic sclerosis. *Ann Rheum Dis*. 2007;66(12):1641-7.

114. Goh NSL, Veeraraghavan S, Desai SR, Cramer D, Hansell DM, Denton CP, et al. Bronchoalveolar lavage cellular profiles in patients with systemic sclerosis-associated interstitial lung disease are not predictive of disease progression. *Arthritis Rheum*. 2007;56(6):2005-12.

115. Mittoo S, Wigley FM, Wise R, Xiao H, Hummers L. Persistence of abnormal bronchoalveolar lavage findings after cyclophosphamide treatment in scleroderma patients with interstitial lung disease. *Arthritis Rheum*. 2007;56(12):4195-202.

116. Tashkin DP, Elashoff R, Clements PJ, Roth MD, Furst DE, Silver RM, et al. Effects of 1-year treatment with cyclophosphamide on outcomes at 2 years in scleroderma lung disease. *Am J Respir Crit Care Med*. 2007;176(10):1026-34.

117. Tzelepis GE, Plastiras SC, Karadimitrakis SP, Vlachoyiannopoulos PG. Determinants of pulmonary function improvement in patients with scleroderma and intestitial lung disease. *Clin Exp Rheumatol*. 2007;25(5):734-9.

118. Berezne A, Ranque B, Valeyre D, Brauner M, Allanore Y, Launay D, et al. Therapeutic strategy combining intravenous cyclophosphamide followed by oral azathioprine to treat worsening interstitial lung disease associated with systemic sclerosis: a retrospective multicenter open-label study. *J Rheumatol*. 2008;35(6):1064-72.

119. Boin F, De Fanis U, Bartlett SJ, Wigley FM, Rosen A, Casolaro V. T cell polarization identifies distinct clinical phenotypes in scleroderma lung disease. *Arthritis Rheum*. 2008;58(4):1165-74.

120. Goh NSL, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med*. 2008;177(11):1248-54.

121. Strange C, Bolster MB, Roth MD, Silver RM, Theodore A, Goldin J, et al. Bronchoalveolar lavage and response to cyclophosphamide in scleroderma interstitial lung disease. *Am J Respir Crit Care Med*. 2008;177(1):91-8.

122. Assassi S, Tan FK, McNearney TA, Estrada-Y-Martin R, Draeger HT, Anderson J, et al. The whole blood interferon score predicts progression of interstitial lung disease in systemic sclerosis. *Arthritis Rheum*. 2009;60 Suppl 10:443.

123. De Souza RBC, Borges CTL, Capelozzi VL, Parra ER, Jatene FB, Kavakama J, et al. Centrilobular fibrosis: an underrecognized pattern in systemic sclerosis. *Respiration*. 2009;77(4):389-97.

124. Gordon J, Mersten J, Lyman S, Kloiber SA, Wildman HF, Crow MK, et al. Imatinib mesylate (Gleevec) in the treatment of systemic sclerosis: interim results of a phase IIa, one year, open label clinical trial. *Arthritis Rheum*. 2009;60:606.

125. Khanna D, Tseng CH, Furst DE, Clements PJ, Elashoff R, Roth M, et al. Minimally important differences in the Mahler's Transition Dyspnoea Index in a large randomized controlled trial – results from the scleroderma lung study. *Rheumatology (Oxford)*. 2009;48(12):1537-40.

126. Ottewell L, Walker K, Griffiths B. Long-term outcome of a targeted treatment regimen with IV cyclophosphamide (CYC) in patients with systemic sclerosis (SSC) and interstitial lung disease (ILD). *Rheumatology (Oxford)*. 2009;48 Suppl 1:i76.

127. Schmidt K, Martinez-Gamboa L, Meier S, Witt C, Meisel C, Hanitsch LG, et al. Bronchoalveoloar lavage fluid cytokines and chemokines as markers and predictors for the outcome of interstitial lung disease in systemic sclerosis patients. *Arthritis Res Ther*. 2009;11(4):R111.

128. Wanchu A, Suryanaryana BS, Sharma S, Sharma A, Bambery P. High-dose prednisolone and bolus cyclophosphamide in interstitial lung disease associated with systemic sclerosis: a prospective open study. *Int J Rheum Dis*. 2009;12(3):239-42.

129. Assassi S, Sharif R, Lasky RE, McNearney TA, Estrada-Y-Martin RM, Draeger H, et al. Predictors of interstitial lung disease in early systemic sclerosis: a prospective longitudinal study of the GENISOS cohort. *Arthritis Res Ther*. 2010;12(5):R166.

130. Boin F, Wigley F, Wise R, Rosen A. Circulating T cell polarization is associated with respiratory decline in scleroderma patients with active interstitial lung disease. *Arthritis Rheum*. 2010;62 Suppl 10:1989.

131. Colaci M, Sebastiani M, Giuggioli D, Manfredi A, Spagnolo P, Luppi F, et al. Bronchoalveolar lavage and response to cyclophosphamide in scleroderma alveolitis. *Scand J Rheumatol*. 2010;39(2):155-9.

132. Cuomo G, Abignano G, Iudici M, Petrillo A, Valentini G. Mycophenolate mophetil in the treatment of systemic sclerosis-interstitial lung disease (SSc-ILD) in patients unresponsive to cyclophosphamide (CYC). *Clin Exp Rheumatol*. 2010;28(2):S167.

133. Gilson M, Zerkak D, Wipff J, Dusser D, Dinh-Xuan AT, Abitbol V, et al. Prognostic factors for lung function in systemic sclerosis: prospective study of 105 cases. *Eur Respir J*. 2010;35(1):112-7.

134. Mittoo S, Robinson D, Hudson M, Baron M. Predicting lung function decline in systemic sclerosis (SSC). *Ann Rheum Dis*. 2010;69 Suppl 2:A27.

135. Schorr M, Wise RA, Wigley FM, Hummers LK. Long-term outcome of patients with an isolated low diffusing capacity. *Clin Exp Rheumatol*. 2010;28(5):S76.

136. Seibold JR, Denton CP, Furst DE, Guillevin L, Rubin LJ, Wells A, et al. Randomized, prospective, placebo-controlled trial of bosentan in interstitial lung disease secondary to systemic sclerosis. *Arthritis Rheum*. 2010;62(7):2101-8.

137. Shahane A, Parambil J, Xu M, Chatterjee S. Scleroderma lung disease: effect of co-existent pulmonary hypertension on progression of interstitial lung disease. *Arthritis Rheum*. 2010;62 Suppl 10:592.

138. Steen VD, Domsic RT, Medsger T. Severe pulmonary fibrosis is uncommon in scleroderma patients with RNA polymerase 3 antibody. *Arthritis Rheum*. 2010;62 Suppl 10:1231.

139. Theodore AC, Simms RW, Lafyatis RA, Kissin EY, Merkel PA, Cruikshank WW, et al. A preliminary investigation of the predictive value and response to thereapy with cyclophosphamide of interleukin-16 in bronchoalveolar lavaage from patients with interstitial lung disease in systemic sclerosis. *Clin Exp Rheumatol*. 2010;28(5):S78-9.

140. Abhishek A, Yazdani R, Pearce F, Regan M, Lim K, Hubbard R, et al. Outcome of systemic sclerosis associated interstitial lung disease treated with intravenous cyclophosphamide. *Clin Rheumatol.* 2011;30(8):1099-104.

141. De Santis M, Inzitari R, Bosello SL, Peluso G, Fanali C, Iavarone F, et al. Beta-thymosins and interstitial lung disease: study of a scleroderma cohort with a one-year follow-up. *Respir Res.* 2011;12(1):22.

142. Espinosa G, Simeon CP, Plasin MA, Xaubet A, Munoz X, Fonollosa V, et al. Efficacy of cyclophospamide in the treatment of interstitial lung disease associated with systemic sclerosis. *Arch Bronconeumol.* 2011;47(5):239-45.

143. Goh NSL, Desai SR, Anagnostopoulos C, Hansell DM, Hoyles RK, Sato H, et al. Increased epithelial permeability in pulmonary fibrosis in relation to disease progression. *Eur Respir J.* 2011;38(1):184-90.

144. Hasegawa M, Fujimoto M, Hamaguchi Y, Matsushita T, Inoue K, Sato S, et al. Use of serum Clara cell 16-kDa (CC16) levels as a potential indicator of active pulmonary fibrosis in systemic sclerosis. *J Rheumatol.* 2011;38(5):877-84.

145. Hoshino K, Satoh T, Kawaguchi Y, Kuwana M. Association of hepatocyte growth factor promoter polymorphism with severity of interstitial lung disease in Japanese patients with systemic sclerosis. *Arthritis Rheum.* 2011;63(8):2465-72.

146. Jayaweera S, Sahhar J, Englert H, Proudman S, Hill C, Youssef P. Mycophenolate mofetil (MMF) in scleroderma (SSC) associated interstitial lung disease (ILD) and skin disease – the Australian experience. *Intern Med J.* 2011;41:11.

147. Khanna D, Mayes MD, Abtin F, Clements PJ, Maranian P, Assassi S, et al. A one-year, phase I/IIa, open-label pilot trial of imatinib mesylate in the treatment of systemic sclerosis-associated active interstitial lung disease. *Arthritis Rheum.* 2011;63(11):3540-6.

148. Khanna D, Tseng CH, Farmani N, Steen V, Furst DE, Clements PJ, et al. Clinical course of lung physiology in patients with scleroderma and interstitial lung disease: analysis of the Scleroderma Lung Study placebo group. *Arthritis Rheum*. 2011;63(10):3078-85.

149. Mittoo S, Wigley FM, Wise RA, Woods A, Xiao H, Hummers LK. Long term effects of cyclophosphamide treatment on lung function and survival in scleroderma patients with interstitial lung disease. *Open Rheumatol J*. 2011;5(1):1-6.

150. Poormoghim H, Lakeh MM, Mohammadipour M, Talehy-Moineddin S, Sodagari F. Pulmonary survival study in 91 patients with systemic sclerosis. *Rheumatol Int*. 2011;31(12):1577-82.

151. Rosato E, Rossi C, Molinaro I, Giovannetti A, Pisarri S, Salsano F. Long-term N-acetylcysteine therapy in systemic sclerosis interstitial lung disease: a retrospective study. *Int J Immunopathol Pharmacol*. 2011;24(3):727-33.

152. Roth MD, Tseng CH, Clements PJ, Furst DE, Tashkin DP, Goldin JG, et al. Predicting treatment outcomes and responder subsets in scleroderma-related interstitial lung disease. *Arthritis Rheum*. 2011;63(9):2797-808.

153. Tiev KP, Hua-Huy T, Kettaneh A, Gain M, Duong-Quy S, Toledano C, et al. Serum CC chemokine ligand-18 predicts lung disease worsening in systemic sclerosis. *Eur Respir J*. 2011;38(6):1355-60.

154. Volpinari S, Corte RL, Bighi S, Ravenna F, Prandini N, Monaco AL, et al. Bronchoalveolar lavage in systemic sclerosis with lung involvement: role and correlations with functional, radiological and scintigraphic parameters. *Rheumatol Int*. 2011;31(9):1183-8.

155. Abignano G, Del Galdo F, Emery P, Buch M. Extended course cyclophosphamide and methylprednisolone pulse therapy can stabilize initially refractory interstitial lung disease in patients with SSc: a single-centre experience. *Rheumatology (Oxford)*. 2012;51:ii114.

156. De Santis M, Bosello SL, Peluso G, Pinnelli M, Alivernini S, Zizzo G, et al. Bronchoalveolar lavage fluid and progression of scleroderma interstitial lung disease. *Clin Respir J*. 2012;6(1):9-17.

157. Hesselstrand R, Andreasson KA, Wuttge DM, Bozovic G, Scheja A, Saxne T. Increased serum COMP predicts mortality in ssc: results from a longitudinal study of interstitial lung disease. *Rheumatology (Oxford)*. 2012;51(5):915-20.

158. Kishore Babu KV, Srinivasa C, Sirisha K, Naidu ARK, Rajasekhar L, Suresh K, et al. Outcome of immunosuppressive therapy in patients with scleroderma and interstitial lung disease, 2 years follow up data, a retrospective study. *Indian J Rheumatol*. 2012;7:S27.

159. Kuwana M, Kaburaki J. Natural history of pulmonary function in patients with SSc and interstitial lung disease. *Rheumatology (Oxford)*. 2012;51:ii83-4.

160. Kuwana M, Takeuchi T, Kaburaki J. Elevation of KL-6 at early disease course predicts subsequent deterioration of pulmonary function in patients with systemic sclerosis and interstitial lung disease. *Arthritis Rheum*. 2012;64 Suppl 10:S371.

161. Le Gouellec N, Faivre JB, Hachulla AL, Hachulla E, Perez T, Morell-Dubois S, et al. Prognosis factors for survival and progression-free survival in systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis*. 2012;71 Suppl 3:400.

162. Schupp J, Riemekasten G, Kollert C, Jager B, Muller-Quernheim J, Prasse A. CCL18 as marker of disease progression in systemic sclerosis. *Eur Respir J*. 2012;40 Suppl 56:P692.

163. Sfriso P, Cozzi F, Oliviero F, Caso F, Cardarelli S, Facco M, et al. CXCL11 in bronchoalveolar lavage fluid and pulmonary function decline in systemic sclerosis. *Clin Exp Rheumatol*. 2012;30 Suppl 71:S71-5.

164. Soriano A, Margiotta D, Marigliano B, Alemanno P, Vadacca M, Afeltra A. Pulmonary fibrosis and connective tissue diseases: follow-up of lung involvement in rheumatoid arthritis and systemic sclerosis. *Ann Rheum Dis*. 2012;71 Suppl 3:343.

165. Tiev KP, Hua-Huy T, Kettaneh A, Allanore Y, Le-Dong NN, Duong-Quy S, et al. Alveolar concentration of nitric oxide predicts pulmonary function deterioration in scleroderma. *Thorax*. 2012;67(2):157-63.

166. Ananyeva L, Desinova O, Starovoytova M, Koneva O, Yutkina N, Onsyannikova O, et al. Rituximab for the treatment of systemic sclerosis associated interstitial lung disease: a case series. *Ann Rheum Dis*. 2013;72 Suppl 3:648.

167. Ando K, Motojima S, Doi T, Nagaoka T, Kaneko N, Aoshima M, et al. Effect of glucocorticoid monotherapy on pulmonary function and survival in Japanese patients with scleroderma-related interstitial lung disease. *Respir Investig*. 2013;51(2):69-75.

168. Burt RK, Oliveira MC, Shah SJ, Moraes DA, Simoes B, Gheorghiade M, et al. Cardiac involvement and treatment-related mortality after non-myeloablative haemopoietic stem-cell transplantation with unselected autologous peripheral blood for patients with systemic sclerosis: a retrospective analysis. *Lancet*. 2013;381(9872):1116-24.

169. Celeste S, Santaniello A, Caronni M, Franchi J, Severino A, Scorza R, et al. Carbohydrate antigen 15.3 as a serum biomarker of interstitial lung disease in systemic sclerosis patients. *Eur J Intern Med*. 2013;24(7):671-6.

170. De Lauretis A, Sestini P, Pantelidis P, Hoyles R, Hansell DM, Goh NSL, et al. Serum interleukin 6 is predictive of early functional decline and mortality in interstitial lung disease associated with systemic sclerosis. *J Rheumatol*. 2013;40(4):435-46.

171. Elhaj M, Charles J, Pedroza C, Liu X, Zhou X, Estrada-Y-Martin RM, et al. Can serum surfactant protein D or CC-chemokine ligand 18 predict outcome of interstitial lung disease in patients with early systemic sclerosis? *J Rheumatol*. 2013;40(7):1114-20.

172. Enghelmayer JI, Giovini V, Altube S, Craviotto F, Cuenca E, Legarreta CG, et al. Does the immunosuppressive therapy modify the pulmonary function parameters in interstitial lung disease (ILD) associated with scleroderma?: our experience. *Am J Respir Crit Care Med*. 2013;187:A2919.

173. Koneva O, Ovsjannikova O, Ananjeva L, Lesnyak V. Isolated reduction of diffusing lung capacity in patients with systemic sclerosis (SSC): evaluation covering a 5 year period. *Ann Rheum Dis*. 2013;72 Suppl 3:A651-2.

174. Liu X, Mayes MD, Pedroza C, Draeger HT, Gonzalez EB, Harper BE, et al. Does C-reactive protein predict the long-term progression of interstitial lung disease and survival in patients with early systemic sclerosis? *Arthritis Care Res (Hoboken)*. 2013;65(8):1375-80.

175. Panopoulos ST, Bournia VK, Trakada G, Giavri I, Kostopoulos C, Sfikakis PP. Mycophenolate versus cyclophosphamide for progressive interstitial lung disease associated with systemic sclerosis: a 2-year case control study. *Lung*. 2013;191(5):483-9.

176. Radic M, Becker MO, Adler S, Schmidt K, Burmester GR, Riemekasten G. Cytokines in bronchoalveolar lavage fluid and serum predict deterioration of lung function and mortality in systemic sclerosis patients. *Ann Rheum Dis*. 2013;72 Suppl 3:A501.

177. Stock CJ, Sato H, Fonseca C, Banya WAS, Molyneaux PL, Adamali H, et al. Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. *Thorax*. 2013;68(5):436-41.

178. Vacca A, Garau P, Porru G, Calvisi S, Ibba V, Piga M, et al. Safety and efficacy of oral cyclophosphamide long-term therapy in systemic sclerosis: experience of a single-centre. *Ann Rheum Dis*. 2013;72 Suppl 3:A503.

179. Wu M, Pedroza C, Salazar G, Zhou X, Reveille JD, Mayes MD, et al. Plasma MCP-1 and IL-10 levels predict long-term progression of interstitial lung disease in patients with early systemic sclerosis. *Arthritis Rheum*. 2013;65 Suppl 10:S742.

180. Zhang XJ, Bonner A, Hudson M, Baron M, Pope J. Association of gastroesophageal factors and worsening of forced vital capacity in systemic sclerosis. *J Rheumatol*. 2013;40(6):850-8.

181. Ananyeva L, Ovsyannikova O, Lesnyak V, Koneva O, Desinova O, Starovoytova M. Slow progressive interstitial lung disease associated with systemic sclerosis: distinct disease phenotype? *Ann Rheum Dis*. 2014;73 Suppl 2:1013.

182. Chakr R, Gasparin A, Neves M, Rangel G, Hendler J, Moro AL, et al. Long-term effectiveness and safety of cyclophosphamide in systemic sclerosis lung disease. *Ann Rheum Dis*. 2014;73 Suppl 2:565-6.

183. Christmann RB, Sampaio-Barros P, Stifano G, Borges CL, De Carvalho CR, Kairalla R, et al. Association of interferon- and transforming growth factor beta-regulated genes and macrophage activation with systemic sclerosis-related progressive lung fibrosis. *Arthritis Rheumatol*. 2014;66(3):714-25.

184. Cottrell TR, Wise RA, Wigley FM, Boin F. The degree of skin involvement identifies distinct lung disease outcomes and survival in systemic sclerosis. *Ann Rheum Dis*. 2014;73(6):1060-6.

185. Fraticelli P, Gabrielli B, Pomponio G, Valentini G, Bosello S, Riboldi P, et al. Low-dose oral imatinib in the treatment of systemic sclerosis interstitial lung disease unresponsive to cyclophosphamide: a phase II pilot study. *Arthritis Res Ther*. 2014;16(4):R144.

186. Guillen-Del Castillo A, Pilar Simeon-Aznar C, Fonollosa-Pla V, Alonso-Vila S, Reverte-Vinaixa MM, Munoz X, et al. Good outcome of interstitial lung disease in patients with scleroderma associated to anti-PM/Scl antibody. *Semin Arthritis Rheum*. 2014;44(3):331-7.

187. Hoffmann-Vold AM, Midtvedt O, Garen T, Lund MB, Aalokken TM, Gran JT, et al. Moderate decline in forced vital capacity is associated with a poor outcome in systemic sclerosis patients. *Arthritis Rheumatol.* 2014;66 Suppl 10:S316-7.

188. Kumanovics G, Gorbe E, Minier T, Simon D, Berki T, Czirjak L. Follow-up of serum KL-6 lung fibrosis biomarker levels in 173 patients with systemic sclerosis. *Clin Exp Rheumatol.* 2014;32(6 Suppl 86):S138-44.

189. Kwon HM, Kang EH, Go DJ, Park JK, Lee EY, Song YW, et al. Efficacy and safety of long term cyclophosphamide treatment for interstitial lung disease in systemic sclerosis. *Ann Rheum Dis*. 2014;73 Suppl 2:1011.

190. Lambrecht S, Smith V, De Wilde K, Coudenys J, Decuman S, Deforce D, et al. Growth differentiation factor 15, a marker of lung involvement in systemic sclerosis, is involved in fibrosis development but is not indispensable for fibrosis development. *Arthritis Rheumatol.* 2014;66(2):418-27.

191. Le Gouellec N, Duhamel A, Faivre JB, Guillaume L, Sobanski V, Perez T, et al. Prognostic factors of functional outcome in systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis*. 2014;73 Suppl 2:1008-9.

192. Narvaez J, Sancho JJA, Castellvi I, Herrera S, Molina MM, Castillo D, et al. Long-term efficacy of rituximab in systemic sclerosis. *Arthritis Rheumatol.* 2014;66 Suppl 10: S737.

193. Nihtyanova SI, Schreiber BE, Ong VH, Rosenberg D, Moinzadeh P, Coghlan JG, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. *Arthritis Rheumatol.* 2014;66(6):1625-35.

194. Parida J, Nath A, Neyaz Z, Agarwal V. A double blind randomized control trial of oral tadalafil in interstitial lung disease of scleroderma. *Arthritis Rheumatol.* 2014;66 Suppl 10:S739.

195. Pham M, Griffing WL. Effects of mycophenolate mofetil on pulmonary lung function in interstitial lung disease of systemic sclerosis. *Arthritis Rheumatol.* 2014;66 Suppl 10:S742.

196. Poormoghim H, Rezaei N, Sheidaie Z, Almasi AR, Moradi-Lakeh M, Almasi S, et al. Systemic sclerosis: comparison of efficacy of oral cyclophosphamide and azathioprine on skin score and pulmonary involvement – a retrospective study. *Rheumatol Int.* 2014;34(12):1691-9.

197. Rotondo C, Praino E, Lanciano E, Fornaro M, Lopalco G, Nivuori MG, et al. Residual volume: a candidate as early marker of interstitial lung disease insystemic sclerosis patients? *Clin Exp Rheumatol.* 2014;32(2):S83.

198. Ariani A, Silva M, Parisi S, Saracco M, Bravi E, De Gennaro F, et al. Can quantitative chest CT predict interstitial lung disease worsening in systemic sclerosis? Results from a multi-centre prospective cohort study. *Ann Rheum Dis.* 2015;74 Suppl 2:587-8.

199. Balbir-Gurman A, Yigla M, Guralnik L, Hardak E, Solomonov A, Rozin AP, et al. Long-term follow-up of patients with scleroderma interstitial lung disease treated with intravenous cyclophosphamide pulse therapy: a single-center experience. *Isr Med Assoc J.* 2015;17(3):150-6.

200. Bosello SL, De Luca G, Rucco M, Berardi G, Falcione M, Danza FM, et al. Long-term efficacy of B cell depletion therapy on lung and skin involvement in diffuse systemic sclerosis. *Semin Arthritis Rheum.* 2015;44(4):428-36.

201. De Luca G, Bosello SL, Berardi G, Rucco M, Canestrari G, Correra M, et al. Tumour-associated antigens in systemic sclerosis patients with interstitial lung disease: association with lung involvement and cancer risk. *Rheumatology (Oxford).* 2015;54(11):1991-9.

202. Hoffmann-Vold AM, Aalokken TM, Lund MB, Garen T, Midtvedt O, Brunborg C, et al. Predictive value of serial high-resolution computed tomography analyses and concurrent lung function tests in systemic sclerosis. *Arthritis Rheumatol.* 2015;67(8):2205-12.

203. Iudici M, Cuomo G, Vettori S, Bocchino M, Sanduzzi Zamparelli A, et al. Low-dose pulse cyclophosphamide in interstitial lung disease associated with systemic sclerosis (SSc-ILD): efficacy of maintenance immunosuppression in responders and non-responders. *Semin Arthritis Rheum*. 2015;44(4):437-44.

204. Jordan S, Distler JHW, Maurer B, Huscher D, Van Laar JM, Allanore Y, et al. Effects and safety of rituximab in systemic sclerosis: an analysis from the European Scleroderma Trial and Research (EUSTAR) group. *Ann Rheum Dis*. 2015;74(6):1188-94.

205. Khanna D, Albera C, Fischer A, Seibold JR, Khalidi NA, Raghu G, et al. Safety and tolerability of pirfenidone in patients with systemic sclerosis interstitial lung disease. *Arthritis Rheumatol.* 2015;67 Suppl 10:816.

206. Khanna D, Nagaraja V, Tseng CH, Abtin F, Suh R, Kim G, et al. Predictors of lung function decline in scleroderma-related interstitial lung disease based on high-resolution computed tomography: implications for cohort enrichment in systemic sclerosis-associated interstitial lung disease trials. *Arthritis Res Ther*. 2015;17(1):372.

207. Koneva O, Desinova O, Ovsyannikova O, Starovoytova M, Glukchova S, Ananieva L. Impact of anti-B-cell therapy with rituximab on pulmonary function of the patients with systemic sclerosis and interstitial lung disease. *Ann Rheum Dis*. 2015;74 Suppl 2:603-4.

208. Lepri G, Avouac J, Airo P, Anguita Santos F, Bellando Randone S, Blagojevic J, et al. Mid-term effects of rituximab in connective tissue disorders related interstitial lung disease (ILD). *Ann Rheum Dis*. 2015;74 Suppl 2:1131.

209. Mani M, Sriram S, Saranya S, Saravanan M, Euphrasia L, Sankaralingam R. Rituximab in systemic sclerosis with ILD-two year outcome from a tertiary care hospital in South India. *Int J Rheum Dis*. 2015;18:114.

210. Mateos-Toledo H, Sarmiento-Padilla G, Mejia M, Castillo-Pedroza J, Buendia-Roldan I, Estrada A, et al. Functional follow-up of diffuse interstitial lung disease (ILD) in phenotypes of systemic sclerosis (SSC). *Am J Respir Crit Care Med*. 2015;191:A1158.

211. Narvaez J, Heredia S, Panos HB, Armengol E, De Lama E, Narvaez JA. Is the presence of esophageal dilation a poor prognostic factor in dilated interstitial lung disease associated with systemic sclerosis? *Arthritis Rheumatol*. 2015;67 Suppl 10:2267-8.

212. Ninaber MK, Stoilk J, Smit J, Le Roy EJ, Kroft LJM, Els Bakker M, et al. Lung structure and function relation in systemic sclerosis: application of lung densitometry. *Eur J Radiol*. 2015;84(5):975-9.

213. Radic M, Becker MO, Distler O, Huscher D, Riemekasten G. Does angiotensin and endothelin receptor blockade have an impact on lung function? An analysis from the EUSTAR data base. *Ann Rheum Dis*. 2015;74 Suppl 2:817.

214. Sakamoto N, Kakugawa T, Hara A, Nakashima S, Yura H, Harada T, et al. Association of elevated alpha-defensin levels with interstitial pneumonia in patients with systemic sclerosis. *Respir Res*. 2015;16:148.

215. Saketkoo LA, Lammi MR, Fischer A, Molitor J, Steen VD. Mycophenolate mofetil (MMF) use in scleroderma patients with pulmonary hypertension: FVC, outcomes and survival-observations from the Pulmonary Hypertension Recognition and Outcomes in Scleroderma (PHAROS) cohort. *Ann Rheum Dis*. 2015;74 Suppl 2:820.

216. Schulam P, Ligon C, Wise R, Hummers LK, Wigley FM, Saria S. A computational tool for individualized prognosis of percent of predicted forced vital capacity trajectories in systemic sclerosis. *Arthritis Rheumatol.* 2015;67 Suppl 10:1123-4.

217. Shirai Y, Takeuchi T, Kuwana M. Clinical utility of serial KL-6 measurement in interstitial lung disease associated with systemic sclerosis. *Arthritis Rheumatol.* 2015;67 Suppl 10:115.

218. Tanaseanu CM, Tiglea IA, Marta DS, Dumitrascu AL, Popescu M, Tanaseanu S, et al. Lactate dehydrogenase a possible marker of progressive microvasculopathy and interstitial lung disease in systemic sclerosis. *Acta Med Mediterr.* 2015;31(5):941-6.

219. Tashkin D, Roth M, Clements P, Furst D, Khanna D, Goldin J, et al. Efficacy and safety of mycophenolate (MMF) vs oral cyclophosphamide (CYC) for treatment of scleroderma-interstitial lung disease (SscILD): results of scleroderma lung study II. *Chest.* 2015;148(4):637A-B.

220. Volkmann E, Tashkin D, Li N, Furst D, Clements P, Elashoff R. Development of a composite outcome measure for systemic sclerosis related interstitial lung disease. *Rheumatology (Sunnyvale).* 2015;5(2):1000154.

221. Volkmann ER, Tashkin DP, Roth M, Tseng CH, LeClair H, Clements PJ, et al. CXCL4 does not predict extent or progression of interstitial lung disease in systemic sclerosis. *Arthritis Rheumatol.* 2015;67 Suppl 10:1406-7.

222. Wallace B, Kafaja S, Furst DE, Berrocal VJ, Merkel PA, Seibold JR, et al. Reliability, validity and responsiveness to change of the Saint George's Respiratory Questionnaire in early diffuse cutaneous systemic sclerosis. *Rheumatology (Oxford).* 2015;54(8):1369-79.

223. Fava A, Cimbro R, Wigley FM, Liu QR, Rosen A, Boin F. Frequency of circulating topoisomerase-I-specific CD4 T cells predicts presence and progression of interstitial lung disease in scleroderma. *Arthritis Res Ther.* 2016;18(1):99.

224. Hoffmann-Vold AM, Tennoe AH, Garen T, Midtvedt O, Abraityte A, Aalokken TM, et al. High level of chemokine CCL18 is associated with pulmonary function deterioration, lung fibrosis progression and reduced survival in systemic sclerosis. *Chest*. 2016;150(2):299-306.

225. Kloth C, Maximilian Thaiss W, Preibsch H, Mark K, Kotter I, Hetzel J, et al. Quantitative chest computed tomography analysis in patients with systemic sclerosis before and after autologous stem cell transplantation: comparison of results with those of pulmonary function tests and clinical tests. *Rheumatology (Oxford)*. 2016;55(10):1763-70.

226. Owen C, Ngian GS, Elford K, Moore O, Stevens W, Nikpour M, et al. Mycophenolate mofetil is an effective and safe option for the management of systemic sclerosis-associated interstitial lung disease: results from the Australian Scleroderma Cohort Study. *Clin Exp Rheumatol*. 2016;34(5):170-6.

227. Shenoy PD, Bavaliya M, Sashidharan S, Nalianda K, Sreenath S. Cyclophosphamide versus mycophenolate mofetil in scleroderma interstitial lung disease (SSc-ILD) as induction therapy: a single-centre, retrospective analysis. *Arthritis Res Ther*. 2016;18(1):123.

228. U.S. National Library of Medicine (ClinicalTrials.gov). Open-label study with bosentan in interstitial lung disease (BUILD 2 OL) [Internet]. 2015 [cited 23 February 2017] Available from: http://clinicaltrials.gov/show/NCT00319033.

229. EU Clinical Trials Register. Systemic sclerosis associated interstitial lung disease: a longitudinal observational study assessing optimum treatment regimens [Internet]. [cited 23 February 2017]. Available from : https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2008-000224-27.

230. U.S. National Library of Medicine (ClinicalTrials.gov). Intravenous cyclophosphamide for the treatment of systemic sclerosis associated interstitial lung disease [Internet]. 2016 [cited 23 February 2017]. Available from: https://clinicaltrials.gov/show/NCT01570764.

231. U.S. National Library of Medicine (ClinicalTrials.gov). Study of pomalidomide (CC-4047) to evaluate safety, tolerability, pharmacokinetics, pharmacodynamics and effectiveness for subjects with systemic sclerosis with interstitial lung disease [Internet]. 2017 [cited 23 February 2017]. Available from: https://clinicaltrials.gov/show/NCT01559129.

232. U.S. National Library of Medicine (ClinicalTrials.gov). Treatment and prevention of progression of interstitial lung disease in systemic sclerosis [Internet]. 2014 [cited 24 February 2017]. Available from: http://clinicaltrials.gov/show/NCT01858259.

233. U.S. National Library of Medicine (ClinicalTrials.gov). A trial to compare nintedanib with placebo for patients with scleroderma related lung fibrosis [Internet]. 2017 [cited 23 February 2017]. Available from: https://clinicaltrials.gov/show/NCT02597933.

234. U.S. National Library of Medicine (ClinicalTrials.gov). A double-blinded study to evaluate the safety, tolerability, and efficacy of BMS-986020 versus placebo in diffuse cutaneous systemic sclerosis (dcSSc) [Internet]. 2016 [cited 23 February 2017]. Available from: https://clinicaltrials.gov/show/NCT02588625.

235. U.S. National Library of Medicine (ClinicalTrials.gov). Comparing and combining bortezomib and mycophenolate in SSc pulmonary fibrosis [Internet]. 2017 [cited 24 February 2017]. Available from: https://clinicaltrials.gov/show/NCT02370693.

236. U.S. National Library of Medicine (ClinicalTrials.gov). Abituzumab in SSc-ILD [Internet]. 2017 [cited 24 February 2017]. Avalable from: https://clinicaltrials.gov/show/NCT02745145.

237. Harrison NK, Myers AR, Corrin B, Soosay G, Dewar A, Black CM, et al. Structural features of interstitial lung disease in systemic sclerosis. *Am Rev Respir Dis*. 1991;144(3):706-13.

238. Wells AU, Hansell DM, Rubens MB, King AD, Cramer D, Black CM, et al. Fibrosing alveolitis in systemic sclerosis: indices of lung function in relation to extent of disease on computed tomography. *Arthritis Rheum*. 1997;40(7):1229-36.

239. Wells AU, Rubens MB, Du Bois RM, Hansell DM. Functional impairment in fibrosing alveolitis: relationship to reversible disease on thin section computed tomography. *Eur Respir J*. 1997;10(2):280-5.

240. Diot E, Boissinot E, Asquier E, Guilmot JL, Lemarie E, Valat C, et al. Relationship between abnormalities on high-resolution CT and pulmonary function in systemic sclerosis. *Chest*. 1998;114(6):1623-9.

241. Kim EA, Johkoh T, Lee KS, Ichikado K, Koh EM, Kim TS, et al. Interstitial pneumonia in progressive systemic sclerosis: serial high-resolution CT findings with functional correlation. *J Comput Assist Tomogr*. 2001;25(5):757-63.

242. Shahin AA, Sabri YY, Mostafa HA, Sabry EY, Hamid MA, Gamal H, et al. Pulmonary function tests, high-resolution computerized tomography, alpha-1 antitrypsin measurement, and early detection of pulmonary involvement in patients with systemic sclerosis. *Rheumatol Int*. 2001;20(3):95-100.

243. Han S, Kim S, Kang Y. DLCO as a predictor of pulmonary fibrosis on high resolution CT in systemic sclerosis. *Ann Rheum Dis*. 2003;62:230.

244. Ooi GC, Mok MY, Tsang KW, Wong Y, Khong PL, Fung PC, et al. Interstitial lung disease in systemic sclerosis. *Acta Radiol*. 2003;44(3):258-64.

245. Orlandi I, Camiciottoli G, Diciotti S, Bartolucci M, Cavigli E, Nacci F, et al. Thin-section and low-dose volumetric computed tomographic densitometry of the lung in systemic sclerosis. *J Comput Assist Tomogr*. 2006;30(5):823-7.

246. Beretta L, Santaniello A, Lemos A, Masciocchi M, Scorza R. Validity of the Saint George's Respiratory Questionnaire in the evaluation of the health-related quality of life in patients with interstitial lung disease secondary to systemic sclerosis. *Rheumatology (Oxford)*. 2007;46(2):296-301.

247. Camiciottoli G, Orlandi I, Bartolucci M, Meoni E, Nacci F, Diciotti S, et al. Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. *Chest*. 2007;131(3):672-81.

248. Goldin JG, Lynch DA, Strollo DC, Suh RD, Schraufnagel DE, Clements PJ, et al. High-resolution CT scan findings in patients with symptomatic scleroderma-related interstitial lung disease. *Chest*. 2008;134(2):358-67.

249. Bellia M, Cannizzaro F, Scichilone N, Riili M, Triolo G, Midiri M, et al. HRCT and scleroderma: semiquantitative evaluation of lung damage and functional abnormalities. *Radiol Med*. 2009;114(2):190-203.

250. Goldin J, Elashoff R, Kim HJ, Yan X, Lynch D, Strollo D, et al. Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. *Chest*. 2009;136(5):1333-40.

251. Vonk MC, Broers B, Heijdra YF, Ton E, Snijder R, van Dijk AP, et al. Systemic sclerosis and its pulmonary complications in The Netherlands: an epidemiological study. *Ann Rheum Dis*. 2009;68(6):961-5.

252. Ananyeva L, Teplova LV, Lesnyak VN, Starovojtova MN, Desinova OV, Nevskaya TA. Relationship between computed tomography and lung function parameters in patients with different duration of systemic sclerosis. *Clin Exp Rheumatol*. 2010;28(2):S151.

253. Peng M, Xu W, Wang Q, Zhao J, Li M, Xu D, et al. Pulmonary involvement in Chinese patients with SSc and correlations between the imaging and functional abnormalities. *Clin Exp Rheumatol*. 2010;28(2):S150.

254. Kim H, Brown M, Abtin F, Lu P, Chong D, Goldin J. Association of texture-based quantitative fibrotic patterns and pulmonary function test in a new validation set. *Eur Respir J*. 2011;38:1440.

255. Kim HJ, Brown MS, Elashoff R, Li G, Gjertson DW, Lynch DA, et al. Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. *Eur Radiol*. 2011;21(12):2455-65.

256. Moghadam KG, Gharibdoost F, Parastandechehr G, Salehian P. Assessments of pulmonary involvement in patients with systemic sclerosis. *Arch Iran Med*. 2011;14(1):22-6.

257. Parra ER, Oliveira e. Silva L, Rangel M, Capelozzi VL. Increase alpha-smooth muscle actin, telomerase, interleukin-4 and endothelin-1 expression in progressive pulmonary fibrosis of systemic sclerosis (SSc). *Virchows Archiv*. 2011;459:S144-5.

258. Assayag D, Hudson M, Tatibouet S, Steele R, Hirsch A, Baron M. Clinical correlates of lung scan abnormalities in patients with systemic sclerosis-associated interstitial lung disease. *Am J Respir Crit Care Med*. 2012;185:A6616.

259. Mantero M, Aliberti S, Erba G, Allevi E, Ricci M, Galbiati S, et al. Cytokines panel in exhaled breath condensate in SSc. *Rheumatology (Oxford)*. 2012;51:ii69.

260. Mittal S, Kumar U, Guleria R, Seith Bhalla A, Mohan A, Mathur S, et al. HRCT chest score and bronchoalveolar lavage fluid cytology in assessment of disease activity of systemic sclerosis associated interstitial lung disease. *Indian J Rheumatol*. 2012;7:S5.

261. Pernot J, Puzenat E, Magy-Bertrand N, Manzoni P, Gondouin A, Bourdin H, et al. Detection of interstitial lung disease in systemic sclerosis through partitioning of lung transfer for carbon monoxide. *Respiration*. 2012;84(6):461-8.

262. Perrin F, Chambellan A, Hardouin JB, Mourrain Langlois E, Neel A, Masseanu A, et al. Membrane diffusion and capillary blood volume measurements in patients with SSc. *Rheumatology (Oxford)*. 2012;51:ii71-2.

263. Wilsher M, Good N, Hopkins R, Young P, Milne D, Gibson A, et al. The six-minute walk test using forehead oximetry is reliable in the assessment of scleroderma lung disease. *Respirology*. 2012;17(4):647-52.

264. Zimmermann A, Pizzichini E, Nobre LF, Pizzichini MMM. Semiquantitative evaluation of highresolution CT in SSc: preliminary applicability of a simple score in clinical practice. *Rheumatology (Oxford)*. 2012;51:ii46-7.

265. Gatta G, Di Grezia G, Iacomino A, Russo A, Petrillo M, Feragalli B, et al. HRCT in systemic sclerosis: correlation between respiratory functional indexes and extension of lung failure. *J Biol Regul Homeost Agents*. 2013;27(2):579-87.

266. Nguyen-Kim TDL, Maurer B, Suliman YA, Morsbach F, Distler O, Frauenfelder T. Histogram-based quantification of fibrosis in systemic sclerosis on sequential low dose HRCT of 9 CT slices. *J Thorac Imaging*. 2013;28(5):W74-5.

267. Piorunek T, Kuznar-Kaminska B, Cofta S, Batura-Gabryel H, Poplawski D, Majewski D, et al. Lung impairment in scleroderma. In: Pokorski M, ed. Respiratory Regulation – Clinical Advances. Advances in Experimental Medicine and Biology. New York, NY: Springer; 2013. p.149-54.

268. Zamora FD, Kim HJ, Wang Q. Prevalence of pulmonary function test abnormalities and their correlation to high resolution computer tomography in a large scleroderma population. *Am J Respir Crit Care Med*. 2013;187:A2920.

269. Colaci M, Sebastiani M, Manfredi A, Giuggioli D, Cassone G, Manzini CU, et al. Lung involvement in systemic sclerosis: role of high resolution computed tomography and its

relationship with other pulmonary and clinico-serological features. *J Biol Regul Homeost Agents*. 2014;28(3):481-8.

270. Ariani A, Silva M, Bravi E, Saracco M, Parisi S, De Gennaro F, et al. Operator-independent quantitative chest computed tomography versus standard assessment of interstitial lung disease related to systemic sclerosis: a multi-centric study. *Mod Rheumatol*. 2015;25(5):724-30.

271. Bernstein EJ, Berrocal VJ, Steen VD, Shanmugam VK, Shah AA, Hinchcliff ME, et al. The predictive value of pulmonary function tests to diagnose interstitial lung disease in adults with early diffuse cutaneous systemic sclerosis. *Arthritis Rheumatol*. 2015;67 Suppl 10:2268-9.

272. Ghandour AM, Gamal RM, Seifeldein GS, Galal MMM, El-Hakeim EM. MDCT staging of disease extent in patients with systemic sclerosis: a pilot study. *Ann Rheum Dis*. 2015;74 Suppl 2:601-2.

273. Guarnieri G, Zanatta E, Mason P, Scarpa MC, Pigatto E, Maestrelli P, et al. Determinants of impairment in lung diffusing capacity in patients with systemic sclerosis. *Clin Exp Rheumatol* 2015; 33(4 Suppl 91):S80-6.

274. Kim HJG, Tashkin DP, Brown MS, Kleerup EC, Goldin JG. Systemic sclerosis interstitial lung disease evaluation: comparison between two quantitative computed tomography for the change assessments. *Am J Respir Crit Care Med*. 2015;191:A1161.

275. Salaffi F, Carotti M, Bosello S, Ciapetti A, Gutierrez M, Bichisecchi E, et al. Computer-aided quantification of interstitial lung disease from high resolution computed tomography images in systemic sclerosis: correlation with visual reader-based score and physiologic tests. *Biomed Res Int*. 2015;834262.

276. Antoniou KM, Margaritopoulos GA, Goh NS, Karagiannis K, Desai SR, Nicholson AG, et al. Combined pulmonary fibrosis and emphysema in scleroderma-related lung disease has a

major confounding effect on lung physiology and screening for pulmonary hypertension. *Arthritis Rheumatol.* 2016;68(4):1004-12.

277. Cetincakmak MG, Goya C, Hamidi C, Tekbas G, Abakay O, Batmaz I, et al. Quantitative volumetric assessment of pulmonary involvement in patients with systemic sclerosis. *Quant Imaging Med Surg.* 2016;6(1):50-6.

278. Salaffi F, Carotti M, Di Donato E, Di Carlo M, Ceccarelli L, Giuseppetti G. Computer-aided tomographic analysis of interstitial lung disease (ILD) in patients with systemic sclerosis (SSc). Correlation with pulmonary physiologic tests and patient-centred measures of erceived dyspnea and functional disability. *PLoS One.* 2016;11(3):e0149240.

279. Tashkin DP, Volkmann ER, Tseng CH, Kim HJ Goldin J, Clements P, et al. Relationship between quantitative radiographic assessments of interstitial lung disease and physiological and clinical features of systemic sclerosis. *Ann Rheum Dis.* 2016;75(2):374-81.

280. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet.* 2011;377(9779):1760-9.

281. Richeldi L, Du Bois R, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med.* 2014;370(22):2071-2082.

282. Wells AU. Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow's ear. *Thorax.* 2013;68(4):309-10.

283. Raghu G, Collard HR, Anstrom KJ, Flaherty KR, Fleming TR, King TE Jr, et al. Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. *Am J Respir Crit Care Med.* 2012;185(10):1044-8.

284. Herzog EL, Mathur A, Tager AM, Feghali-Bostwick C, Schneider F, Varga J. Review: interstitial lung disease associated with systemic sclerosis and idiopathic pulmonary fibrosis: how similar and distinct? *Arthritis Rheumatol.* 2014;66(8):1967-78.

285. Khanna D, Seibold JR, Wells A, Distler O, Allanore Y, Denton C, et al. Systemic sclerosis-associated interstitial lung disease: lessons from clinical trials, outcome measures, and future study design. *Curr Rheumatol Rev.* 2010;6(2):138-44.

286. Furst D, Khanna D, Matucci-Cerinic M, Clements P, Steen V, Pope J, et al. Systemic sclerosis – continuing progress in developing clinical measures of response. *J Rheumatol.* 2007;34(5):1194-1200.

287. Solomon JJ, Olson AL, Fischer A, Bull T, Brown KK, Raghu G. Scleroderma lung disease. *Eur Respir Rev.* 2013;22(127):6-19.

288. Khanna, D, Mittoo S, Aggarwal R, Proudman SM, Dalbeth N, Matteson EL, et al. Connective tissue disease-associated interstitial lung diseases (CTD-ILD) – report from OMERACT CTD-ILD working group. *J Rheumatol.* 2015;42(11):2168-71.

289. Moore OA, Proudman SM, Goh N, Corte TJ, Rouse H, Hennessy O, et al. Quantifying change in pulmonary function as a prognostic marker in systemic sclerosis-related interstitial lung disease. *Clin Exp Rheumatol.* 2015;33 Suppl 91:S111-6.

290. Goh NS, Hoyles RK, Denton CP, Hansell DM, Renzoni EA, Maher TM, et al. Short-term pulmonary function trends are predictive of mortality in interstitial lung disease associated with systemic sclerosis. *Arthritis Rheumatol.* 2017;69(8):1670-8.

291. Jack C, Slomkowski M, Gracon S, Hoover T, Felmlee J, Stewart K, et al. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology.* 2003;60(2):253-60.

292. Trabattoni D, Saresella M, Biasin M, Boasso A, Piacentini L, Ferrante P, et al. B7-H1 is up-regulated in HIV infection and is a novel surrogate marker of disease progression. *Blood*. 2003;101(7):2514-20.

293. Nadarajah V, Van Putten M, Chaouch A, Garrood P, Straub V, Lochmüller H, et al. Serum matrix metalloproteinase-9 (MMP-9) as a biomarker for monitoring disease progression in Duchenne muscular dystrophy (DMD). *Neuromuscul Disord*. 2011;21(8):569-78.

294. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, et al. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med*. 2012;18(11):1711-5.

295. Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models. *Stat Med*. 1999;18(24):3463-78.

296. Watkins RE, Eagleson S, Veenendaal B, Wright G, Plant AJ. Disease surveillance using a hidden Markov model. *BMC Med Inform Decis Mak*. 2009;9(1):39.

297. Morris JF, Koski A, Johnson LC. Spirometric standards for healthy nonsmoking adults. *Am Rev Respir Dis*. 1971;103(1):57-67.

298. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2018. URL https://www.R-project.org/.

299. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *Journal Stat Softw*. 2011;45(3):1-68.

300. Sanchez-Cano D, Ortego-Centeno N, Callejas JL, Fonollosa Pla V, Rios-Fernandez R, Tolosa-Vilella C, et al. Interstitial lung disease in systemic sclerosis: data from the spanish scleroderma study group. *Rheumatol Int*. 2018;38(3):363-74.

301. Wangkaew S, Euathrongchit J, Wattanawittawas P, Kasitanon N, Louthrenoo W. Incidence and predictors of interstitial lung disease (ILD) in Thai patients with early systemic sclerosis: inception cohort study. *Mod Rheumatol*. 2016;26(4):588-93.

302. Ashmore P, Tikly M, Wong M, Ickinger C. Interstitial lung disease in South Africans with systemic sclerosis. *Rheumatol Int*. 2018;38(4):657-62.

303. Kneebone RL, Scott W, Darzi A, Horrocks M. Simulation and clinical practice: strengthening the relationship. *Med Educ*. 2004;38(10):1095-102.

304. Kneebone R. Evaluating clinical simulations for learning procedural skills: a theory-based approach. *Acad Med*. 2005;80(6):549-53.

305. Caron M, Hoa S, Hudson M, Schwartzman K, Steele R. Pulmonary function tests as outcomes for systemic sclerosis interstitial lung disease. *Eur Respir Rev*. 2018;27(148).

306. Molberg O, Hoffmann-Vold AM. Interstitial lung disease in systemic sclerosis: progress in screening and early diagnosis. *Curr Opin Rheumatol*. 2016;28(6):613-8.

307. Hoa S, Bernatsky S, Steele RJ, Baron M, Hudson M. Association between immunosuppressive therapy and course of mild interstitial lung disease in systemic sclerosis. *Rheumatology (Oxford)*. 2019; doi:10.1093/rheumatology/kez407.

308. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277(3):826-32.

309. Powell GA, Verma A, Luo Y, Stephens D, Buckeridge D. Modeling Chronic obstructive pulmonary disease progression using continuous-time hidden Markov models. *Stud Health Technol Inform*. 2019;264:920-4.

310. Aalen OO, Farewell VT, de Angelis D, Day NE, Nöel Gill O. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Stat Med*. 1997;16(19):2191-210.

311. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol*. 2010;63(8):938-9.

312. Kim GHJ, Tashkin DP, Lo P, Brown MS, Volkmann ER, Gjertson DW, et al. Using transitional changes on HRCT to monitor the impact of cyclophosphamide or mycophenolate on systemic sclerosis-related interstitial lung disease. *Arthritis Rheumatol*. 2019; doi:10.1002/art.41085.

313. Elhai M, Hoffmann-Vold AM, Avouac J, Pezet S, Cauvet A, Leblond A, et al. Performance of candidate serum biomarkers for systemic sclerosis-associated interstitial lung disease. *Arthritis Rheumatol*. 2019;71(6):972-82.

314. Affandi AJ, Radstake TR, Marut W. Update on biomarkers in systemic sclerosis: tools for diagnosis and treatment. *Semin Immunopathol*. 2015;37(5):475-87.

315. Bonhomme O, Andre B, Gester F, de Seny D, Moermans C, Struman I, et al. Biomarkers in systemic sclerosis-associated interstitial lung disease: review of the literature. *Rheumatology (Oxford)*. 2019;58(9):1534-46.

316. Assayag D, Kaduri S, Hudson M, Hirsch A, Baron M. High resolution computed tomography scoring systems for evaluating interstitial lung disease in systemic sclerosis patients. *Rheumatology*. 2012; doi:10.4172/2161-1149.S1-003

317. Bacher A, Mittoo S, Hudson M, Tatibouet S, Baron M. Systemic sclerosis in Canada's North American Native population: assessment of clinical and serological manifestations. *J Rheumatol*. 2013;40(7):1121-6.

318. Keane MP. The fibrosis burden of systemic sclerosis. *Am J Respir Crit Care Med*. 2019; doi:10.1164/rccm.201907-1438ED.

319. American Thoracic Society. Standardization of spirometry – 1987 update. *Am Rev Respir Dis*. 1987;136(5):1285-98.

320. American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis*. 1991;144(5):1202-18.

321. American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med*. 1995;152(3):1107-36.

322. Jackson CH. Multi-state models for panel data: the msm package for R. *J Stat Softw*. 2011;38(8):1-29.

323. Devroye L. Non-Uniform Random Variate Generation. New York, NY: Springer, 1986. General principles in random variate generation; p.27-82.

324. Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol*. 2009;62(8):797-806.