Fall 12-20-2019

Enhancing the Identification of Rheumatoid Arthritis-Associated Lung Disease

Bryant England
University of Nebraska Medical Center

ENHANCING THE IDENTIFICATION OF RHEUMATOID ARTHRITIS-ASSOCIATED LUNG DISEASE

by

Bryant R. England

A DISSERTATION

Presented to the Faculty of

the University of Nebraska Graduate College

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

Medical Sciences Interdepartmental Area Graduate Program

(Clinical & Translational Research)

Under the Supervision of Professor Ted R. Mikuls

University of Nebraska Medical Center

Omaha, Nebraska

December, 2019

Supervisory Committee:

Ted R. Mikuls, M.D., M.S.P.H.	Geoffrey M. Thiele, Ph.D.
Fang Yu, Ph.D.	Shinobu Watanabe-Galloway, Ph.D.

ACKNOWLEDGEMENTS

I am indebted to the many individuals who have supported this work. I would like to thank Dr. Ted Mikuls who has served as my mentor over the past seven years. Completion of this research and training as well as a clinician investigator career would not have been possible without his encouragement, guidance, feedback, support, and the plethora of opportunities he has made available. I would like to thank my committee members Drs. Geoffrey Thiele, Fang Yu, and Shinobu Watanabe-Galloway who provided terrific feedback and guidance on my research projects and training.

The research that follows is a product of the work of many collaborative individuals who deserve recognition. Harlan Sayles instructed me in statistical programming and assisted with implementing the coding for analyses in Chapters 2, 3, and 4. Punyasha Roul assisted with performing the queries to generate the cohorts in national Veterans Health Administration data in Chapters 3 and 4. Drs. Tina Mahajan and Namrata Singh assisted with performing medical record reviews to validate interstitial lung disease cases for Chapters 3 and 4. Mike Duryee and the Experimental Immunology Lab members performed serum and tissue antibody testing for Chapter 3. Drs. Kevin Deane and Kristen Demoruelle provided lung tissues for Chapter 3. Drs. Brian Sauer and Grant Cannon provided national Veterans Health Administration data and comorbidity data for Chapter 2. I am appreciative of the Veterans Affairs Rheumatoid Arthritis registry investigators who have worked together over the past 16 years to build a data-rich resource to improve the lives of individuals with rheumatoid arthritis. Last, but certainly not least, thank you to the patients who generously participate in the Veterans Affairs Rheumatoid Arthritis registry.

I am extremely grateful to all the members of the University of Nebraska Medical Center Division of Rheumatology and Immunology. All have supported me through their

teaching and collaboration, but I would like to recognize those not previously mentioned who have had a tremendous influence on my development. Dr. James O'Dell has served as a role-model and mentor since I was a medical student, providing me countless career and research opportunities. Dr. Amy Cannella personalized my rheumatology fellowship training so I could focus on research and research training. Dr. Kaleb Michaud has served as a mentor in rheumatology registry research and epidemiologic methods.

Completing this research and training would not have been possible without many generous funding agencies. I would like to thank the Department of Internal Medicine, Division of Rheumatology & Immunology, College of Medicine Physician Scientist Training Program, and Mentored Scholars Program at the University of Nebraska Medical Center as well as the Rheumatology Research Foundation and the Great Plains IDeA-CTR.

Finally, I thank my family for their support. My wife, Melissa, who not only took on the additional family workload that allowed me to pursue this training despite her own career as a physician, but provided encouragement along the way (such as suggesting I try to graduate before our children!). Our children, Owen, Alice, and Emily, who are a constant source of joy, ask great questions, and are poised to be life-long learners. My parents, Bob and Sandy, who fostered a pursuit of education and excellence. And last, my late grandfather, Dr. B.J. England, who will forever be an inspiration to me as a physician and over a single lunch many summers ago set the wheels in motion for me to pursue a career as a clinician investigator.

ABSTRACT

ENHANCING THE IDENTIFICATION OF RHEUMATOID ARTHRITIS-ASSOCIATED LUNG DISEASE

Bryant R. England, M.D., Ph.D.

University of Nebraska, 2019

Supervisor: Ted R. Mikuls, M.D., M.S.P.H

Rheumatoid arthritis (RA) is a systemic autoimmune disease that predisposes afflicted individuals to reduced quality of life, physical disability, and premature mortality. While joint involvement is the primary manifestation of RA, extra-articular features including lung disease are responsible for a significant portion of the excess mortality. In this dissertation I demonstrate the contribution of chronic lung diseases to premature mortality in RA, contrasting with the more widely recognized comorbidity in RA of cardiovascular disease. Then, I establish that a novel serum biomarker, anti-malondialdehyde-acetaldehyde adduct (MAA) antibody, is associated with the presence of interstitial lung disease (ILD) in RA subjects. Further implicating its role in the pathogenesis of RA-ILD, I will demonstrate the presence of MAA as well as the co-localization of MAA with RA autoantigens and immune effectors cells in the lungs of RA-ILD subjects. Finally, I describe how biomedical informatics algorithms that incorporate multiple ILD diagnosis codes, provider specialty, and diagnostic testing can accurately classify ILD status in RA subjects. Together, these studies advance our ability to identify RA-associated lung diseases across the spectrum of clinical and translational research. These results will pave the way for future clinical and translational research studies to compose biomarker panels that aid in the screening of RA subjects for lung disease, identify pathways that could be targeted for novel therapeutics in RA-ILD, and facilitate

the completion of comparative effectiveness and outcomes research studies using real-world data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACPA	Anti-citrullinated protein antibody
ACR	American College of Rheumatology
Anti-CCP	Anti-cyclic citrullinated peptide antibody
bDMARDs	Biologic disease-modifying anti-rheumatic drugs
CI	Confidence interval
Cit-HSP90	Citrullinated heat shock protein-90
COPD	Chronic obstructive pulmonary disease
CPT	Current procedural terminology
CT	Computed tomography
DAS28	Disease Activity Score in 28 joints
DMARDs	Disease-modifying anti-rheumatic drugs
HR	Hazard ratio
HCUP-CCS	Healthcare Cost and Utilization Project-Clinical Classification Software
ICD	International Classification of Diseases
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
IP-10	Interferon-y inducible protein-10
MAA	Malondialdehyde-acetaldehyde adducts

MDHAQ	Multidimensional Health Assessment Questionnaire
MMP-7	Matrix metalloproteinase-7
NPV	Negative predictive value
OR	Odds ratio
PARC	Pulmonary and activation-regulated chemokine
PFT	Pulmonary function test
PPV	Positive predictive value
RA	Rheumatoid arthritis
RF	Rheumatoid factor
RU	Relative units
SD	Standard deviation
SP-D	Surfactant protein-D
TNFi	Tumor necrosis factor inhibitor
VARA	Veterans Affairs Rheumatoid Arthritis registry

LIST OF PUBLICATIONS

1. England BR, Sayles H, Michaud K, Thiele GM, Poole JA, Caplan L, Sauer BC, Cannon GW, Reimold A, Kerr GS, Baker JF, Mikuls TR. Chronic lung disease in U.S. Veterans with rheumatoid arthritis and the impact on survival. *Clinical Rheumatology* 2018; 37(11):2907-2915. PMID: 30280369.

2. England BR, Duryee MJ, Roul P, Mahajan TD, Singh N, Poole JA, Ascherman DP, Caplan L, Demoruelle MK, Deane KD, Klassen LW, Thiele GM, Mikuls TR. Malondialdehyde-Acetaldehyde Adducts and Antibody Responses in Rheumatoid Arthritis-Interstitial Lung Disease. *Arthritis & Rheumatology* 2019; 71(9):1483-1493. PMID: 30933423.

3. England BR, Roul P, Mahajan TD, Singh N, Yu F, Sayles H, Cannon GW, Sauer BC, Baker JF, Curtis JR, Mikuls TR. Performance of Administrative Algorithms to Identify Interstitial Lung Disease in Rheumatoid Arthritis. *Arthritis Care & Research* 2019 [Epub ahead of print]. PMID: 31421018.

CHAPTER 1: INTRODUCTION

1.1 Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a chronic, autoimmune disease that affects between 0.5 and 1.0% of the United States population.[1] Women are most commonly afflicted with RA, with the highest incidence rates occurring in those in the 6th to 7th decade of life.[2] The typical presentation of RA is characterized by swelling, pain, and stiffness in the small joints of the hands and feet, but frequently involves additional medium to large joints over time.[3] If left untreated, the articular manifestations of RA can lead to bone erosions, joint deformities, and functional impairment. While joint symptoms often herald the onset of RA and are the primary target of treatment, RA is a systemic disease with numerous extra-articular manifestations. These diverse extra-articular manifestations include cardiovascular disease, pulmonary disease, ophthalmic disease, osteoporosis, and subcutaneous nodules, among others.[4] Although RA is most common among females, extra-articular manifestations have a predilection to affect males.[5]

Establishing the diagnosis of RA has been facilitated by the identification of serum biomarkers closely associated with RA. Rheumatoid factor (RF) is an antibody targeting the Fc portion of an IgG that is present in approximately 70% of RA patients.[6] Generation of these antibodies is not unique to RA, with RF also being detected in other autoimmune or infectious diseases.[7] Anti-citrullinated protein antibodies (ACPAs) target citrullinated peptides formed through a process of citrullination, when arginine is converted to citrulline by peptidylarginine deiminase enzymes.[8] Similarly to RF, ACPAs are detected in approximately 70% of RA patients.[6] However, the specificity of ACPAs for RA is greater than RF, reported to be >95% in a large meta-analysis.[6] Several other novel autoantibodies are currently being investigated in RA. These include antibodies

targeting malondialdehyde-acetaldehyde (MAA) adducts, a product of oxidative stress,[9,10] as well as anti-carbamylated protein antibodies[11] and antibodies against peptidylarginine deiminases.[12]

While the pathophysiology of RA is not fully understood, both genetic predisposition and environmental factors drive RA risk. *HLA-DRB1* alleles encoding a five amino acid sequence at position 70-74 of the HLA-DRβ chain, termed the shared epitope, carry among the highest genetic risk for developing RA.[13] More recently, HLA-DRB1 haplotypes that include the amino acid valine at position 11 have also been associated with RA incidence and severity.[14,15] Outside of the human leukocyte antigen region, PTPN22 single nucleotide polymorphisms carry heightened RA risk.[16] While various environmental factors have been studied as risk factors for RA, tobacco smoke is the strongest environmental risk factor identified to date, an association more closely linked to seropositive RA.[17,18] Moreover, a strong gene-environment interaction between shared epitope alleles and tobacco smoke has been observed, with seropositive RA risk 21-fold higher among smokers with shared epitope alleles.[19] Recent work harnessing longitudinal biorepositories has allowed investigators to discover the “pre-clinical” period of RA when autoantibodies are detected years in advance of clinical symptoms.[20] Pro-inflammatory cytokine production may signal the imminent transition from “pre-clinical” to “clinical” RA.[21]

As a result of the articular and extra-articular manifestations of RA, affected individuals are at risk for poor long-term outcomes. Individuals with RA have reduced quality of life,[22] physical impairment,[23] work disruption,[24] and may require joint replacement surgery.[25] Furthermore, individuals with RA have higher mortality rates than the general population.[26,27] Many of these affected health domains are exacerbated by the increased comorbidity burden present in RA patients. Common comorbid conditions

afflicting RA patients include pulmonary disease, cardiovascular diseases, ophthalmic diseases, osteoporosis and fractures, mental health disorders, and chronic pain disorders, including fibromyalgia, among others.[28,29] Because comorbidities may be a manifestation of RA, a consequence of RA or its treatment, or related to a common external risk factor,[28] differentiating extra-articular manifestations of RA from an RA-related comorbidity is inherently complex.

To optimally prevent the aforementioned short- and long-term complications, RA must be aggressively managed using disease-modifying anti-rheumatic drugs (DMARDs), medications that have demonstrated the ability to slow articular disease progression. Current American College of Rheumatology (ACR) and European League Against Rheumatism guidelines recommend the early initiation of DMARDs after RA diagnosis with a treatment goal of achieving and maintaining low disease activity or remission.[30,31] RA disease activity is typically assessed by composite disease activity measures that incorporate patient and/or provider assessments such as swollen and tender joint counts, and laboratory tests.[32,33] Methotrexate is currently the cornerstone therapy in RA, but many RA patients will not reach treatment goals on methotrexate monotherapy.[34] Escalation to combination DMARDs (e.g. triple therapy – the combination of methotrexate, hydroxychloroquine, and sulfasalazine),[35] biologic DMARDs (bDMARDs), or targeted-synthetic DMARDs typically occurs when treatment goals are not met with methotrexate alone.[30,31] As a result of early aggressive treatment, treating-to-a-target, and an expanding repertoire of treatment options, more RA patients are meeting treatment goals,[36] fewer joint replacement surgeries are occurring,[37] and physical function among RA populations is improving.[38] However, these improvements have not necessarily been accompanied by improvements in survival, where results are conflicting and several studies in the modern treatment era continue to observe higher

mortality in RA patients than the general population.[39] Analyses of RA patients diagnosed between 1980 and 2007 in Olmstead County, Minnesota suggest improvement in cardiovascular mortality, but not other causes of death.[40] Recent large cohort studies in men and women with RA, including work by our group, highlight respiratory disease as a major driver of premature mortality in RA.[26,27] The reasons underlying this lag in improvement of non-cardiovascular deaths are not well elucidated, but widespread efforts to identify and target cardiovascular disease in RA have been implemented over the prior two decades.[41] In contrast, there have been fewer attempts to identify and target other causes of mortality, including lung disease.

1.2 Lung Disease in Rheumatoid Arthritis

The three initial cases of pulmonary involvement secondary to RA were recorded in 1948 by Ellman and Ball in the *British Medical Journal*.[42] These cases, 2 of which proved fatal, described findings consistent with interstitial lung disease (ILD). In 1953, Anthony Caplan described 51 cases of radiologic abnormalities in the lungs of coal miners with RA.[43] In addition to fibrosis that was present in 90% of cases, a unique presentation of well-defined multifocal opacities was noted that later became termed Caplan syndrome. These landmark reports inspired widespread efforts to characterize and evaluate pulmonary manifestations accompanying RA and were essential for recognizing the systemic nature of RA. Today, a wide array of pulmonary manifestations are recognized to complicate the natural course of RA. These include pulmonary nodules, pleural effusions/serositis, ILD, bronchiectasis, and bronchiolitis.[44] Common chronic lung diseases in the general population, such as chronic obstructive pulmonary disease (COPD), also appear to be overrepresented among RA subjects, even after accounting for tobacco use.[45,46]

In addition to being associated with established RA, the lungs appear to be a site where RA autoantibody responses are generated and lung disease may be the initial manifestation of RA.[47,48] RA autoantibodies have been detected in individuals with chronic lung disease in the absence of articular manifestations,[49] and cigarette smoking, the strongest environmental risk factor for RA, induces citrullination in the lungs.[50] Further suggesting a potential pathogenic link between the lungs and joints are the presence of airway abnormalities and sputum RA autoantibody expression in individuals at high risk for RA,[51,52] enrichment of ACPAs in the sputum and bronchoalveolar fluid from patients with RA,[52-54] and identification of shared citrullinated peptides in the lung and joint tissue from patients with RA.[55] The aforementioned findings and a growing understanding of the links between mucosal and systemic autoimmunity have sparked great interest in whether the lungs may be an originating site of RA.[56]

Clinical and translational research on lung disease in RA has largely focused on ILD, which clinically affects between 5-15% of RA patients and up to three-fold more subclinically.[48,57] The pathogenesis of RA-ILD is poorly understood, but believed to be a result of genetic predisposition and environmental exposures (e.g., tobacco use) that drive autoimmunity, pro-inflammatory responses, oxidative stress, fibrosis, and remodeling of the extracellular matrix.[44] These processes result in inflammation and fibrosis of the lungs, which cause symptoms like shortness of breath, dyspnea on exertion, and non-productive cough.[58] Risk factors for developing RA-ILD include older age, male sex, tobacco use, other extra-articular features of RA, and RA disease severity.[48,57,59,60] There are unique patterns of RA-ILD which historically were determined by surgical lung biopsy but more recently are determined by high-resolution computed tomography (CT). The most common histopathologic pattern in RA-ILD is usual interstitial pneumonia,[48,61] as in idiopathic pulmonary fibrosis (IPF). The genetic basis for

RA-ILD also appears to be shared with IPF. A *MUC5B* promoter variant previously linked to IPF was recently identified to be associated with RA-ILD, particularly the usual interstitial pneumonia pattern.[62] The prognosis is poor in RA-ILD, with a median survival after diagnosis of RA-ILD reported to be less than 3 years.[47,57] Only a marginally better prognosis has been suggested by other studies.[63,64]

Further highlighting the impact of lung diseases on long-term outcomes was a prior study of mortality we conducted among male U.S. Veterans with RA. We observed a significantly higher risk of death among the RA subjects compared to age- and sex-matched rates.[26] Moreover, respiratory-related mortality was the most overrepresented cause of death among RA subjects with rates nearly 3-fold higher than age- and sex-matched general population rates (**Figure 1**). Among these respiratory deaths, COPD was the most frequent respiratory cause of death. Sparks et al. similarly found respiratory diseases to be the most overrepresented cause of death in a 36-year study of nurses with RA.[27] In concordance with our findings, COPD was the most frequent respiratory cause of death in their study. Together, these studies demonstrate the long-term consequences of chronic lung disease and the need to study the impact of lung diseases besides RA-ILD on long-term outcomes.

A hypothesized factor contributing to poor long-term outcomes in RA-associated lung disease is the delayed identification of lung diseases, particularly for RA-ILD. Current methods to identify RA-ILD include chest radiography, high-resolution CT, and pulmonary function tests (PFTs). Typically, testing with these modalities is prompted by the development of respiratory symptoms including cough, shortness of breath, or dyspnea on exertion. Relying on the development of clinical symptoms to prompt testing inherently results in late detection, with up to 30% of RA patients having subclinical ILD on CT.[65] Ideally, a screening method to identify RA patients at highest risk for lung

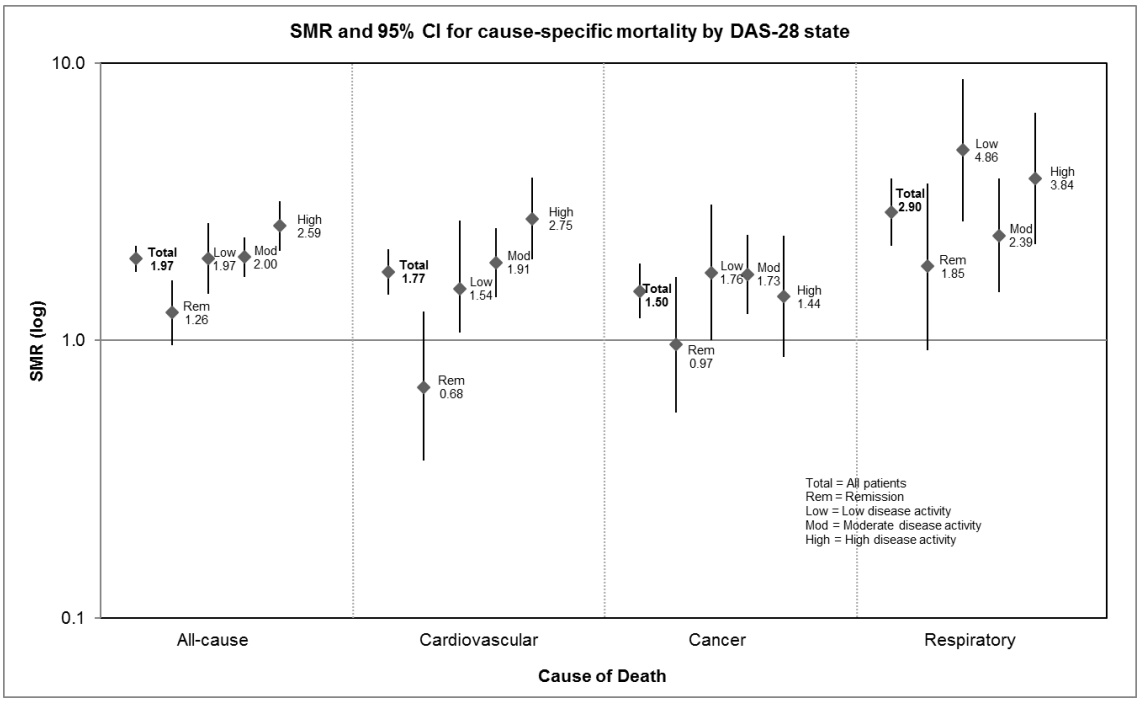
diseases could be used to tailor screening approaches. But accounting for the few established risk factors for RA-ILD,[48,57,59,66-68] is neither sensitive nor specific enough to use in the clinical setting. Better methods to identify individuals with, or at risk for, RA-ILD would allow for studying and implementing targeted therapies earlier in the disease course before irreversible fibrosis is established.

Serum biomarkers have the potential to serve this role as part of RA-ILD screening protocols. Candidate biomarkers identified in RA-ILD to date have included Krebs von den Lungen-6,[64,69] ACPA,[66] anti-citrullinated heat shock protein-90 (anti-cit HSP90 antibody),[70] matrix metalloproteinase-7 (MMP-7),[71,72] interferon-y inducible protein-10 (IP-10),[71] pulmonary and activation-regulated chemokine (PARC),[72] and surfactant protein-D (SP-D).[72] Investigation of these RA-ILD serum biomarkers has been limited by small sample sizes, failing to account for predictive clinical factors or appropriate disease controls, and minimal (or absent) external validation. The identification of serum biomarkers to enhance the detection of RA-ILD remains a critical knowledge gap.[73]

Also responsible for the poor long-term outcomes in RA-associated lung disease is uncertainty regarding optimal treatment approaches. There are no completed controlled trials, nor are there clinical practice guidelines, to inform the management of RA-associated lung disease. This is again most troublesome in RA-ILD, where there is concern that the use of many RA therapies may result in acute episodes of pneumonitis or even progression of underlying ILD. Essentially all DMARDs approved for the treatment of RA have been reported to be associated with drug-induced pneumonitis, which can greatly complicate the diagnosis and/or management of RA-related pulmonary manifestations.[74-76] In the absence of controlled trial data, providers rely on pharmacoepidemiologic evaluations of drug effectiveness and safety through

comparative effectiveness and outcomes research. However, comparative effectiveness studies in RA-ILD using large observational datasets have been limited by the lack of validated approaches to identify RA-ILD patients.[77-79] This is in contrast to RA, and other rheumatic diseases, where there has been substantial research on the validity of administrative based algorithms to build disease cohorts.[80-83] The development of validated algorithms for identifying RA-ILD would allow for comparative effectiveness and outcomes research with large real-world datasets that could immediately inform and improve RA-ILD management.

Figure 1. Standardized mortality ratios for cause-specific mortality by enrollment disease activity in men with RA



Age-adjusted standardized mortality ratios and 95% confidence intervals (CIs) for all-cause, cardiovascular, cancer, and respiratory-related mortality among men with RA using U.S. life tables from the Centers for Disease Prevention and Control. Overall mortality rates and mortality rates stratified by RA disease activity state (remission, low, moderate, and high) are shown. Reprinted with permission from: England BR et al. Cause-Specific Mortality in Male US Veterans with Rheumatoid Arthritis. *Arthritis Care Res (Hoboken)*, 2016;68(1):36-45.

Abbreviations: CI, confidence interval; DAS-28, 28-joint Disease Activity Score; SMR, standardized mortality ratio

1.3 Objectives

The overall objective of this dissertation is to identify means to enhance the identification and investigation of RA-associated lung disease.

In Chapter 2, I use freely available comorbidity classification software to assess the impact of chronic lung diseases, including RA-ILD and other chronic lung diseases, on survival in RA. Additionally, I contrast the risk of death in RA for individuals with chronic lung diseases compared to those with cardiovascular disease, a widely recognized determinant of poor long-term outcomes in RA.

In Chapter 3, I evaluate biological methods to identify RA-ILD by testing whether a novel serum autoantibody, anti-MAA, is associated with the presence of ILD among a large cohort of RA subjects. Furthermore, I investigate whether MAA modified proteins are present in lung tissues from RA-ILD subjects and whether these modified proteins colocalize with other recognized RA-related autoantigens (namely citrullinated proteins) and/or biologically relevant immune effector cells.

In Chapter 4, I leverage biomedical informatics approaches to derive administrative data algorithms that can accurately identify RA-ILD in large, real-world datasets.

Finally, in Chapter 5, I discuss how the findings from this dissertation improve our ability to identify RA-associated lung disease and pave the way for future high-impact clinical and translational research in RA-associated lung disease to improve the existing poor long-term outcomes for this high-risk patient population.

CHAPTER 2: IMPORTANCE OF LUNG DISEASE

2.1 Background

Several chronic lung diseases have been described in RA including ILD, obstructive lung diseases (COPD, bronchiectasis, bronchiolitis), pulmonary nodules, medication toxicities, and pleural diseases. Perhaps the most concerning of these pulmonary manifestations is ILD, which is clinically apparent in 5-15% of RA patients and carries a poor long-term prognosis.[44,47,57,84] In our prior work, we demonstrated that respiratory-related deaths were the most over-represented cause of death in men with RA [26] and similar findings were reported in women with RA in the Nurses' Health Study.[27] In both studies, COPD, rather than ILD, was the leading cause of respiratory-related death in RA. Despite their frequency, the prognostic importance of lung diseases in RA beyond ILD is not well established. In a population-based incident RA cohort study, obstructive lung disease (defined as an obstructive defect on spirometry and a physician diagnosis of airway or parenchymal lung disease) was associated with a 2-fold higher risk of mortality.[46] Bronchiectasis and bronchiolitis have also been reported to increase the risk of mortality in RA patients in a few small studies.[64,85,86]

In addition to RA itself, several DMARDs have been implicated in chronic lung diseases (e.g. drug-induced pneumonitis).[87-90] Moreover, increased adverse events were reported in COPD patients receiving abatacept in a randomized controlled trial.[91] Because of the potential for pulmonary toxicity with these agents, there is significant concern regarding optimal DMARD selection in patients with chronic lung disease, evidenced by epidemiologic channeling to leflunomide (away from methotrexate) in RA patients with ILD.[78] The long-term safety and best practices for the use of DMARDs in RA patients with chronic lung disease remains an important and unanswered question.

Our objective was to evaluate the risk of death among RA patients with chronic lung disease, including chronic lung diseases other than ILD. To illustrate its importance on RA outcomes, we contrasted this risk with cardiovascular disease, another overrepresented comorbid condition in RA patients that is well-established as a determinant of poor long-term outcomes.[26,41,92] Additionally, we investigated whether select DMARD use in RA patients with chronic lung disease was associated with differential impact on survival.

2.2 Methods

Participants

We utilized the Veterans Affairs Rheumatoid Arthritis (VARA) Registry, a multicenter longitudinal observational cohort study of U.S. Veterans with RA fulfilling the 1987 ACR classification criteria.[93] The VARA Registry, initiated in 2003, has been well described previously.[94] Subjects were followed from the time of enrollment until death or censoring at the end of available vital status data (December 31, 2013). All subjects provided written informed consent before enrollment, and each site received institutional review board approval. This study was approved by the VARA Scientific Ethics Advisory Committee.

Chronic lung disease assessment

Recognizing that chronic lung diseases are often characterized by an insidious onset of pulmonary symptoms leading to diagnosis, we assessed prevalent chronic lung disease by using outpatient diagnostic codes within the VA Corporate Data Warehouse collected over a 2-year period, 12 months prior to and 12 months following enrollment. Diagnostic codes were categorized using the Healthcare Cost and Utilization Project

Clinical Classification Software (HCUP-CCS, https://www.hcup-us.ahrq.gov/). HCUP-CCS is a freely available software tool developed by the Agency for Healthcare Research and Quality that categorizes International Classification of Diseases (ICD), 9th Revision, Clinical Modification codes into 295 distinct categories. Chronic lung disease categories included were 127: chronic obstructive pulmonary disease and bronchiectasis, 128: asthma, 132: lung disease due to external agents; and 133: other lower respiratory disease. Respiratory codes representing acute lung conditions were not included (122: pneumonia; 123: influenza; 125: acute bronchitis; 126: other upper respiratory infections; 129: aspiration pneumonitis, food/vomitus; 130: pleurisy, pneumothorax, pulmonary collapse, 131: respiratory failure, insufficiency, arrest; 134: other upper respiratory disease). Diagnostic codes (ICD-9-CM) corresponding to these HCUP-CCS categories are shown in **Appendix A**. Chronic lung disease categories were not mutually exclusive (i.e., patients could have multiple chronic lung diseases). To enhance the specificity of disease classification, we required at least two HCUP-CCS codes within this 24-month window and at least one of these codes to have occurred prior to VARA enrollment. Cardiovascular disease was assessed in the same manner, using HCUP-CCS categories 96-97, 100-101, 105-110, and 112-114 (**Appendix B**). Recognizing that ICD-9-CM codes commonly used for ILD are included within a HCUP-CCS category that also contains non-ILD codes (**Appendix C**), we used diagnostic codes specific to ILD (ICD-9-CM: 495, 515-517, 714.81) entered into the registry at enrollment by treating rheumatologists. An overall measure of comorbidity burden was assessed using the Rheumatic Disease Comorbidity Index.[95]

Clinical variables and vital status

In conjunction with routine rheumatology care, ACR core measures were collected by the treating rheumatologists, including erythrocyte sedimentation rate

(mm/hour), 28-joint swollen joint count, 28-joint tender joint count, patient and provider global assessment (0-100mm visual analogue scale), Multidimensional Health-Assessment Questionnaire (MDHAQ) [96] as well as calculation of the Disease Activity Score in 28 joints (DAS28).[97] DMARDs, both biologic and non-biologic, and prednisone use were similarly collected within the registry. Additional variables collected at enrollment were sex, smoking status (current, former, never), education level, and self-reported race. *HLA-DRB1* shared epitope alleles, C-reactive protein, anti-cyclic citrullinated peptide (anti-CCP, U/ml) antibody, and RF (IU/ml) were measured using banked serum and genomic DNA collected at enrollment, as previously described.[98,99] Vital status was determined by linkage with the National Death Index (Center of Excellence for Suicide Prevention, Joint Department of VA and Department of Defense Suicide Data Repository; http://vaww.virec.research.va.gov/Mortality/Overview.htm; extract through 2013).[26]

Statistical analysis

Baseline characteristics were compared between subjects with and without chronic lung disease at the time of enrollment using chi-square and independent t-tests. Multivariable Cox proportional hazards models were used to assess the association of chronic lung disease with all-cause mortality. Covariates included in the Cox models were age, sex, race, smoking status, MDHAQ, DAS28, baseline DMARDs, and baseline prednisone use. MDHAQ and DAS28 varied over time, while all other variables were fixed at enrollment values. To compare mortality risk between chronic lung disease and cardiovascular disease, a combined categorization was created: neither comorbidity (referent), chronic lung disease alone, cardiovascular disease alone, or both comorbidities occurring together.

Associations of DMARDs with mortality in RA subjects with chronic lung disease was determined in stratified analyses (all patients, patients with chronic lung disease, and patients without lung disease) and interaction terms were tested using multivariable Cox regression models adjusting for age, sex, race, smoking status, MDHAQ, DAS28, and baseline prednisone use and clustered by enrollment site. DMARDs were modeled as time-varying and baseline use separately. Because tumor necrosis factor inhibitors (TNFi) comprised 95% of biologic DMARD use at baseline, all biologic DMARDs were modeled together. In sensitivity analyses, we assessed DMARDs using propensity score adjustment. We calculated propensity scores for receiving methotrexate, leflunomide, hydroxychloroquine, sulfasalazine, azathioprine, or biologic DMARDs at baseline using age, sex, race, education level, smoking status, chronic lung disease, Rheumatic Disease Comorbidity Index score, RA disease duration, anti-CCP antibody positivity, MDHAQ, DAS28, and prednisone use as predictors in logistic regression models. The resulting propensity scores (both as continuous values and propensity score quintiles) were entered as a covariate into Cox models with the DMARD of interest. Proportional hazards assumptions were tested in Cox models by Schoenfeld residuals, which were not significant. A P value <0.05 was considered significant in all analyses, which were completed using Stata v15 (StataCorp, College Station, TX).

2.3 Results

Baseline characteristics

Baseline characteristics of the study population (N = 2,053) stratified by chronic lung disease status are shown in **Table 1**. Those with chronic lung disease were older (P < 0.001), male predominant (P = 0.005), less likely to have a high-school education (P = 0.004), more likely to be current or former smokers (P < 0.001), and had higher

comorbidity scores (P < 0.001). Anti-CCP antibody concentrations were higher in those with chronic lung disease (mean 286 vs. 230 U/mL, P = 0.007), while frequency of anti-CCP and RF positivity and RF concentration did not differ between those with and without chronic lung disease. Similarly, race, disease duration, subcutaneous nodules, and shared epitope positivity did not differ by chronic lung disease status.

Frequency of chronic lung disease

Using HCUP-CCS categories, 27% of participants (n = 554) had chronic lung disease with other lower respiratory disease being the most common (16.1%) followed by COPD/bronchiectasis (14.7%) (**Table 2**). Using physician entered diagnostic codes, the prevalence of COPD (18.1%) was similar while ILD was recorded in 5.2% of patients.

Chronic lung disease and survival

During a total observation period of 6,682 patient-years, there were 341 deaths with 139 of these occurring in patients with chronic lung disease (81.6 per 1,000 person-years, 95% confidence interval [CI] 69.1-96.0) and 202 occurring in patients without lung disease (40.6 per 1,000 person-years, 95% CI 35.4-46.6). Deaths among those with chronic lung disease occurred on average 1.3 years earlier compared to those without (mean age at death of 73.3 ± 9.4 vs. 74.6 ± 9.4 years). Respiratory-related deaths accounted for 22.3% of deaths among those with chronic lung disease, but only for 9.9% among those without chronic lung disease. Chronic lung disease (per HCUP-CCS codes) was associated with a significantly greater mortality risk in all models (**Table 3**). In fully adjusted models, chronic lung disease was associated with a 51% increased risk of mortality (hazard ratio [HR] 1.51, 95% CI 1.26-1.81). Except for asthma, each individual HCUP-CCS chronic lung disease category was associated with a greater

mortality risk (HRs: COPD/bronchiectasis 1.61, 95% CI 1.39-1.86; other lower respiratory 1.32, 95% CI 1.27-1.36). Using physician entered diagnostic codes from the registry, COPD was associated with a 1.48-fold higher mortality risk (95% CI 1.16- 1.90) and ILD was associated with a 1.90-fold higher mortality risk (95% CI 1.23-2.96).

We then contrasted the mortality risk between chronic lung disease and cardiovascular disease using a combined lung and cardiovascular disease classification. Using HCUP-CCS codes, 55.6% of participants had neither lung nor cardiovascular disease comorbidity, 16.5% of participants had only chronic lung disease, 17.3% had only cardiovascular disease, and 10.6% had both comorbidities. Both chronic lung disease and cardiovascular disease were associated with a similar increased risk of mortality alone compared to individuals free of both conditions (**Figure 2**; cardiovascular HR 1.62, 95% CI 1.33-1.94; chronic lung HR 1.46, 95% CI 1.03-2.06). Those simultaneously afflicted with both chronic lung disease and cardiovascular disease had numerically higher mortality risk than those with lung or cardiovascular comorbidity alone (HR 2.28, 95% CI 1.80-2.89). In sub-analyses, COPD by both HCUP-CCS (HR 1.76, 95% CI 1.33-2.32) and physician entered diagnostic codes (1.61, 95% CI 1.28-2.02) were associated with a similar risk of death as comorbid cardiovascular disease (**Table 4**). Other lower respiratory comorbidities identified by HCUP-CCS had a numerically lower risk of death than comorbid cardiovascular disease (other lower respiratory HR 1.13, 95% CI 0.85-1.51; cardiovascular HR 1.56, 95% CI 1.35-1.79) while ILD by physician entered codes had a numerically higher risk of death (ILD HR 2.18, 95% CI 1.08-4.41; cardiovascular HR 1.73, 95% CI 1.58-1.90). The presence of both lung disease and cardiovascular disease was associated with the highest risk of death across all sub-analyses.

DMARDs and mortality risk

While individual DMARDs demonstrated associations with mortality in general, no DMARDs (baseline or time-varying use) were associated with an increased risk of mortality in RA patients with chronic lung disease (**Table 5**). Furthermore, there was no evidence of differential risk of mortality with baseline or time-varying methotrexate or bDMARD use in those with/without chronic lung disease (all P values for interaction ≥ 0.15). There was, however, evidence of a significant interaction between hydroxychloroquine use and chronic lung disease (P ≤ 0.04). In stratified analyses, hydroxychloroquine use was associated with a numerically more protective association with mortality in those with chronic lung disease compared to RA subjects without lung disease; however, associations were not statistically significant within each subgroup. Sub-analyses stratified by individual lung diseases (HCUP-CCS COPD/bronchiectasis, other lower respiratory disease; physician entered codes for COPD and ILD) were consistent with overall lung disease analyses (data not shown). In sensitivity analyses incorporating propensity scores for baseline medication use, most DMARDs were again not associated with an increased risk of mortality in those with chronic lung disease, except for baseline sulfasalazine use (continuous: HR 1.73, 95% CI 1.03-2.91; quintiles: HR 1.67, 95% CI 1.04-2.69; **Table 6**).

Table 1. Baseline characteristics of RA cohort by chronic lung disease status

Variable	Lung Disease (n=554)	No Lung Disease (n=1499)	P value
Demographics and comorbidities			
Age, years	65.7 (9.9)	62.7 (11.2)	<0.001
Male sex, %	93.1	89.2	0.005
White, %	79.0	76.0	0.14
High-school education, %	83.3	88.1	0.004
Smoking status, %			<0.001
Current	27.9	25.9	
Former	58.4	50.5	
Never	13.7	23.6	
Body mass index, kg/m^2	28.5 (5.6)	28.4 (5.6)	0.54
RDCI score	3.9 (1.2)	1.7 (1.3)	<0.001
RA disease status			
RA duration, years	11.5 (11.5)	11.8 (11.4)	0.60
Shared epitope positive, %	72.3	71.7	0.82
Anti-CCP positive, %	79.3	77.0	0.25
Anti-CCP, U/mL	286 (439)	230 (389)	0.007
RF positive, %	80.5	79.4	0.59
RF, IU/mL	370 (719)	328 (707)	0.25
Nodules, %	31.9	29.5	0.27
MD-HAQ, 0-3	1.0 (0.6)	0.9 (0.6)	0.002
ESR, mm/Hr	31.7 (1.2)	24.6 (0.6)	<0.001
C-reactive protein, mg/dL	1.4 (2.1)	1.1 (1.9)	0.003
DAS28	4.1 (1.5)	3.9 (1.6)	0.003
Medications			
Methotrexate, %	50.2	57.3	0.003
Leflunomide, %	15.4	9.7	<0.001
Hydroxychloroquine, %	34.8	34.0	0.72
Sulfasalazine, %	14.4	14.6	0.93
Biologic DMARD, %	28.3	27.5	0.69
Prednisone, %	47.0	38.4	<0.001

P value by independent t-test or X^2.

Abbreviations: RDCI, Rheumatic Disease Comorbidity Index; RA, rheumatoid arthritis; CCP, cyclic citrullinated peptide; RF, rheumatoid factor; ACR, American College of Rheumatology; MD-HAQ, multidimensional Health Assessment Questionnaire; ESR, erythrocyte sedimentation rate; DAS28, 28-joint Disease Activity Score; CDAI, Clinical Disease Activity Index; DMARD, disease-modifying antirheumatic drug; NSAIDs, non-steroidal anti-inflammatory drugs

Table 2. Frequency of specific chronic lung disease comorbidities

Lung disease	N	% of Patients
Any HCUP-CCS chronic lung disease	554	27.0
HCUP-CCS, COPD and bronchiectasis	301	14.7
HCUP-CCS, asthma	62	3.0
HCUP-CCS, lung disease due to external agents	2	0.1
HCUP-CCS, other lower respiratory disease†	330	16.1
Physician entered, ILD‡	106	5.2
Physician entered, COPD	371	18.1

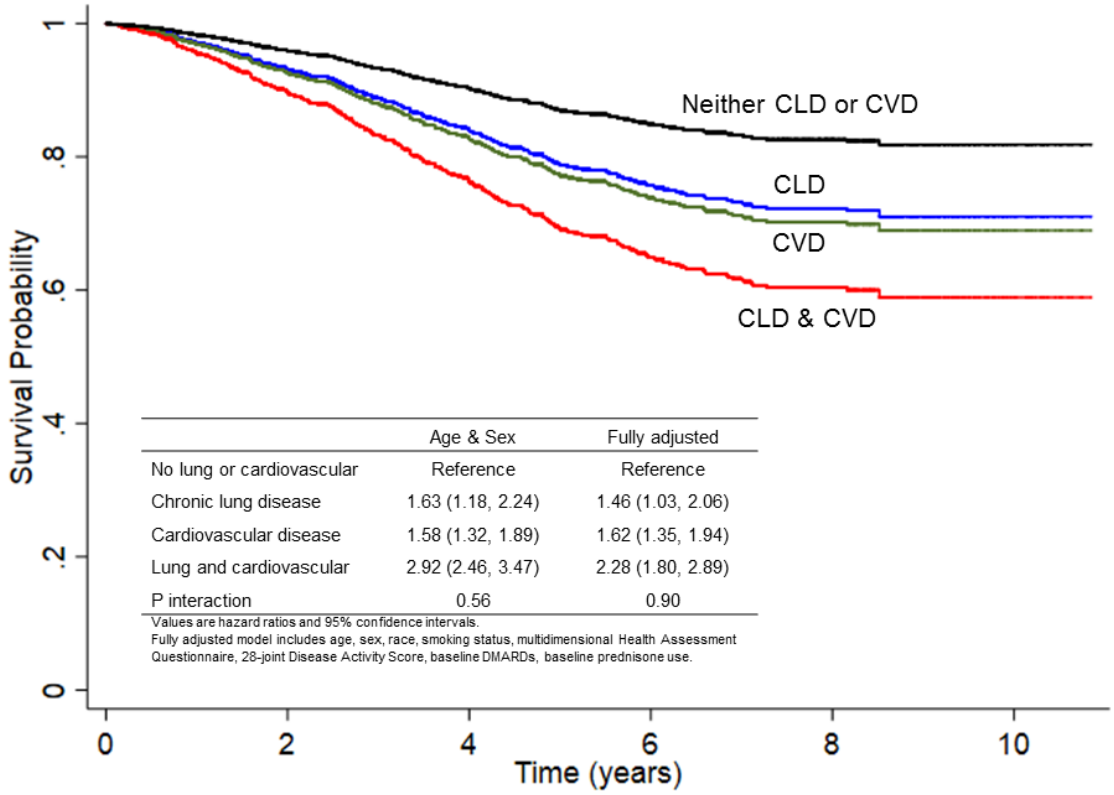
Abbreviations: HCUP-CCS, Healthcare Cost and Utilization Project-Clinical Classification Software; COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease
† Most diagnostic codes for interstitial lung disease are included in this HCUP-CCS category
‡ ICD-9-CM: 495, 515-517, 714.81

Table 3. Associations of chronic lung disease with all-cause mortality in RA

	Age & Sex	Intermediate†	Fully adjusted‡
HCUP-CCS			
Chronic lung disease	1.80 (1.54, 2.11)	1.72 (1.41, 2.10)	1.51 (1.26, 1.81)
COPD & bronchiectasis	2.02 (1.60, 2.56)	1.89 (1.50, 2.38)	1.61 (1.39, 1.86)
Asthma	1.02 (0.47, 2.17)	1.11 (0.49, 2.53)	0.77 (0.23, 2.54)
Other lower respiratory	1.52 (1.45, 1.59)	1.47 (1.35, 1.60)	1.32 (1.27, 1.36)
Physician entered			
COPD	1.82 (1.43, 2.30)	1.70 (1.34, 2.16)	1.48 (1.16, 1.90)
Interstitial lung disease	1.99 (1.35, 2.94)	1.90 (1.26, 2.86)	1.90 (1.23, 2.96)

Values are hazard ratios and 95% confidence intervals.
†Intermediate model includes age, sex, race, and smoking status.
‡Fully adjusted model includes covariates from intermediate model and multidimensional Health Assessment Questionnaire, 28-joint Disease Activity Score, baseline DMARDs, baseline prednisone use.
Abbreviations: HCUP-CCS, Health Care Cost and Utilization Project Clinical Classification Software; COPD, chronic obstructive pulmonary disease

Figure 2. Comparison of all-cause mortality risk between chronic lung and cardiovascular comorbidity



	Age & Sex	Fully adjusted
No lung or cardiovascular	Reference	Reference
Chronic lung disease	1.63 (1.18, 2.24)	1.46 (1.03, 2.06)
Cardiovascular disease	1.58 (1.32, 1.89)	1.62 (1.35, 1.94)
Lung and cardiovascular	2.92 (2.46, 3.47)	2.28 (1.80, 2.89)
P interaction	0.56	0.90

Values are hazard ratios and 95% confidence intervals.
Fully adjusted model includes age, sex, race, smoking status, multidimensional Health Assessment Questionnaire, 28-joint Disease Activity Score, baseline DMARDs, baseline prednisone use.

Probability of survival by chronic lung disease (CLD) and cardiovascular disease (CVD) status assessed during the 12 months prior to and following registry enrollment (neither CVD nor CLD [black], CVD only [green], CLD only [blue], both CVD and CLD [red]). The table contained within the figure shows hazard ratios for the association of CVD and CLD with all-cause mortality in multivariable Cox models as well as the null interaction between CVD and CLD.

Abbreviations: CLD, chronic lung disease; CVD, cardiovascular disease; DMARDs, disease-modifying anti-rheumatic drugs

Table 4. Comparison of mortality risk between specific categories of chronic lung disease and cardiovascular disease comorbidity

	HR (95% CI)
HCUP-CCS - COPD/bronchiectasis	
No COPD or cardiovascular disease	Reference
COPD	1.76 (1.33, 2.32)
Cardiovascular disease	1.75 (1.50, 2.04)
COPD and cardiovascular disease	2.38 (1.88, 3.01)
HCUP-CCS - Other lower respiratory	
No other lung disease or cardiovascular disease	Reference
Other lung disease	1.13 (0.85, 1.51)
Cardiovascular disease	1.56 (1.35, 1.79)
Other lung disease and cardiovascular disease	2.12 (1.81, 2.49)
Physician - COPD	
No COPD or cardiovascular disease	Reference
COPD	1.61 (1.28, 2.02)
Cardiovascular disease	1.78 (1.57, 2.03)
COPD and cardiovascular disease	2.16 (1.54, 3.03)
Physician - Interstitial Lung Disease	
No ILD or cardiovascular disease	Reference
ILD	2.18 (1.08, 4.41)
Cardiovascular disease	1.73 (1.58, 1.90)
ILD and cardiovascular disease	2.86 (1.76, 4.64)

Adjusted for age, sex, race, smoking status, multidimensional Health Assessment Questionnaire, 28-joint Disease Activity Score, baseline DMARDs, baseline prednisone use
Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; HCUP-CCS, Health Care Cost and Utilization Project Clinical Classification Software; HR, hazard ratio; ILD, interstitial lung disease

Table 5. Associations of DMARDs with mortality in RA subjects with and without comorbid lung disease

	All subjects	Lung disease	No lung disease	P interaction
Time-varying				
Methotrexate	0.60 (0.50, 0.71)	0.56 (0.42, 0.75)	0.62 (0.52, 0.74)	0.59
Leflunomide	0.79 (0.60, 1.04)	0.67 (0.40, 1.11)	0.94 (0.63, 1.42)	0.10
Azathioprine	1.30 (0.82, 2.06)	0.81 (0.33, 1.96)	1.72 (1.21, 2.44)	0.09
Hydroxychloroquine	1.08 (0.88, 1.33)	0.96 (0.71, 1.30)	1.20 (0.97, 1.48)	0.04
Sulfasalazine	0.81 (0.68, 0.96)	0.90 (0.72, 1.13)	0.74 (0.60, 0.91)	0.33
Biologic DMARD	0.60 (0.48, 0.75)	0.64 (0.46, 0.88)	0.54 (0.41, 0.70)	0.19
Baseline				
Methotrexate	0.79 (0.59, 1.05)	0.80 (0.53, 1.19)	0.76 (0.56, 1.04)	0.80
Leflunomide	0.86 (0.59, 1.26)	0.88 (0.45, 1.70)	0.94 (0.63, 1.41)	0.63
Azathioprine	0.96 (0.66, 1.39)	0.71 (0.28, 1.79)	1.36 (0.70, 2.63)	0.25
Hydroxychloroquine	0.93 (0.77, 1.14)	0.68 (0.44, 1.04)	1.19 (0.97, 1.47)	0.01
Sulfasalazine	1.00 (0.80, 1.24)	1.27 (0.93, 1.74)	0.82 (0.54, 1.23)	0.14
Biologic DMARD	1.05 (0.88, 1.25)	1.09 (0.87, 1.37)	0.97 (0.74, 1.28)	0.15

Values are hazard ratios and 95% confidence intervals.
Models adjusted for age, sex, race, smoking status, multidimensional Health Assessment Questionnaire, 28-joint Disease Activity Score, chronic lung comorbidity (all-subjects only), and baseline prednisone use.
Abbreviations: DMARD, disease-modifying anti-rheumatic drug

Table 6. Associations of baseline DMARDs with mortality risk in RA subjects with chronic lung comorbidity using propensity score adjustment

	Propensity Score Continuous	Propensity Score Quintiles
Methotrexate	0.86 (0.44, 1.68)	0.87 (0.45, 1.70)
Leflunomide	0.90 (0.47, 1.72)	0.91 (0.50, 1.69)
Azathioprine	0.59 (0.15, 2.29)	0.59 (0.15, 2.34)
Hydroxychloroquine	0.87 (0.58, 1.31)	0.88 (0.58, 1.34)
Sulfasalazine	1.73 (1.03, 2.91)	1.67 (1.04, 2.69)
Biologic DMARD	0.90 (0.74, 1.08)	0.88 (0.74, 1.06)

Values represent hazard ratios and 95% confidence intervals
Abbreviations: DMARDs, disease-modifying anti-rheumatic drugs

2.4 Discussion

Utilizing a cohort of U.S. Veterans with RA and adjusting for key confounders (e.g., smoking status), we have demonstrated that chronic lung disease is associated with reduced survival in RA. These findings were not limited to ILD, but included the more common pulmonary manifestation of COPD. Among the most compelling findings, the effect of chronic lung disease on mortality risk was comparable to that of comorbid cardiovascular disease. Our results emphasize the importance of targeting chronic lung diseases in RA patients with a similar urgency as has been proposed for cardiovascular disease[100] in order to achieve optimal long-term patient outcomes.

Prior studies of long-term outcomes in RA patients with lung disease have primarily focused on ILD. There has been far less investigation into the long-term outcomes for RA patients with obstructive lung diseases, despite evidence that the risk of obstructive lung disease, such as COPD, is increased in RA patients.[45,101] Only a few small studies have examined the survival of RA patients with non-ILD pulmonary manifestations, namely bronchiectasis and bronchiolitis. While the comparison group varied between these studies, findings were suggestive of an increased mortality risk.[64,85,86] Our study not only confirms the findings of increased mortality with non-ILD lung diseases, but expands on these observations. We studied over 2,000 RA patients of which 27% had chronic lung disease and found a 51% increase in mortality for those with any chronic lung disease relative to RA patients without. Those with chronic lung disease were also more likely to die from lung disease (22.3% in those with chronic lung disease vs. 9.9% in those without). While physician coding for ILD numerically carried the highest risk of death in our study (nearly 2-fold higher than no lung disease), COPD/bronchiectasis and other lower respiratory disease codes were associated with an increased risk of death by 32-61% in RA patients. These results clearly indicate that

non-ILD pulmonary manifestations are a major determinant of mortality risk in RA. Furthermore, these results are in line with those from the general population where risk of death was increased 1.6-fold for moderately severe COPD and 1.7-fold for restrictive lung disease.[102]

The heightened risk of cardiovascular disease and cardiovascular disease-related mortality in RA has been well established.[26,41,92] As a result, numerous efforts have been undertaken to enhance the identification of cardiovascular disease and its risk factors as well as treating both RA and modifiable risk factors as a means of preventing premature cardiovascular disease mortality.[100] In fact, recent reports suggest these efforts may be effectively narrowing the current gap in cardiovascular disease incidence between RA patients and the general population.[40] Building on prior studies identifying respiratory-related mortality as the most overrepresented cause of death in RA,[26,27] we have illustrated that comorbid lung disease carries a similar mortality risk as comorbid cardiovascular disease. Thus, our findings provide support for the concept that researchers and clinicians alike should aggressively target the identification and treatment of lung disease in RA as well as its risk factors, as has been previously done in cardiovascular disease. It should be noted, however, that these efforts could render fewer gains than efforts focused on cardiovascular disease because of the limited interventions that are currently available (e.g., smoking cessation and oxygen supplementation). Unfortunately, data suggesting aggressive immunomodulatory therapy improves survival in RA patients with chronic lung disease is lacking, in contrast to some findings with cardiovascular disease-related mortality in RA.[103]

Selecting optimal DMARDs in RA patients with chronic lung disease is challenging. There is concern that select DMARDs could exacerbate pre-existing lung disease or cause pulmonary toxicity. Pre-existing lung disease is a risk factor for

methotrexate pneumonitis,[104] and, as suggested in the current study, individuals with/at-risk-for lung disease are often channeled to alternative therapies, such as leflunomide.[78] However, there is limited evidence to support or refute these safety concerns, making it critically important to evaluate the risk of poor long-term health outcomes with DMARDs in RA patients with chronic lung disease. Reassuringly, our analyses suggest that chronic lung disease does not appear to differentially impact the mortality risk attributable to most DMARDs, including methotrexate and bDMARDs. These findings of similar mortality risk in those with and without lung disease were robust to alternate definitions of DMARD use (baseline vs. time-varying) as well as statistical models (Cox models with multivariable adjustment vs. propensity score adjustment). As this is an observational study design in which treatment selection was informed rather than randomly selected, our observations should be interpreted as associations and not causal evidence. Moreover, we examined mortality risk stratified by a composite chronic lung disease measure. Therefore, future work will be needed to determine if there is a differential mortality risk with DMARDs based on specific sub-types of chronic lung disease in RA.

There are limitations to our study. Because our primary objective was to evaluate survival following lung disease, we ascertained lung disease at cohort inception. Left censoring may have occurred and lung disease occurring during follow-up was not included, but these would be anticipated to bias the risk of mortality towards the null. Furthermore, the insidious onset of lung disease makes it difficult to clearly define “incident” vs. “prevalent” lung disease and will need to be the subject of future studies. We assessed lung disease using diagnostic codes and categorized specific lung diseases using the HCUP-CCS. While this is a standardized, freely accessible classification software for diagnostic codes, it provides only limited classification of

chronic lung diseases and has the potential to misclassify specific lung diseases. However, we also examined diagnoses of COPD and ILD annotated by the treating rheumatologist, with resulting COPD findings that were generally in agreement with results based on HCUP-CCS classification. Given the retrospective nature of these analyses, we were unable to apply American Thoracic Society/European Respiratory Society guidelines for ILD classification.[105] Our rates of chronic lung disease are higher than others have reported,[46] which likely reflects the male predominance and frequent smoking history of our cohort and the VA population, and thus may limit generalizability. Data regarding the severity of lung disease, imaging findings, and PFTs were not available. While we adjusted for smoking status at the time of enrollment, data was not available regarding duration, dose, or intensity of tobacco exposure as well as changes in tobacco use that may have occurred during follow-up. Due to limited use of non-TNF bDMARDs in those with chronic lung disease in our sample, we grouped all bDMARDs together. Thus, future studies are needed to adequately assess the mortality risk with individual bDMARDs.

Our study has numerous strengths including its cohort design and unique study population enriched with chronic lung disease. Moreover, these analyses leveraged robust data including longitudinal RA disease measures and functional status, autoantibodies, and enrollment smoking status, all which could be potential confounders that were adjusted for in multivariable analyses. We linked the clinical data within the VARA registry to the National Death Index to assess vital status and with administrative VA data to capture diagnostic codes related to usual care. Finally, we utilized the HCUP-CCS, readily available to other researchers for use in studying chronic lung disease in RA using administrative data.

In summary, we have demonstrated a greater risk of death for RA patients with chronic lung disease that is similar in magnitude to that of cardiovascular disease and not limited to ILD. Reassuringly, methotrexate and bDMARD use, as occurring in regular care, were not observed to impart a higher risk of mortality in RA patients with chronic lung disease. While future studies should expand on our findings with further characterization of lung disease that includes imaging findings and pulmonary function testing, our study importantly brings to attention the poor prognosis that accompanies chronic lung disease in RA.

CHAPTER 3: IDENTIFICATION OF INTERSTITIAL LUNG DISEASE WITH BIOLOGY

3.1 Background

ILD is a major determinant of poor long-term outcomes in patients with RA, a population which already suffers from premature mortality. Median survival following RA-ILD diagnosis has been reported to be as short as 3 years,[57] and trends in mortality related to RA-ILD do not appear to be declining.[84] The estimated prevalence of clinically apparent ILD is 5-15% in RA patients, with up to 30% having subclinical disease on high-resolution CT.[57,65,84,106] Contributing to the wide-ranging epidemiologic estimates is the difficulty in establishing the diagnosis of RA-ILD, which relies on a multidisciplinary evaluation that often includes pulmonary function testing, high-resolution CT of the chest, and/or lung biopsy.[44,105] With a poorly understood pathogenesis and the development of clinical symptoms well after radiologic or physiologic abnormalities have established,[65,107] delays in diagnosis of RA-ILD are commonplace. These delays in detection may be particularly harmful if substantial irreversible decline occurs before effective management or other preventative strategies are initiated.

Recognizing the diagnostic uncertainties and associated diagnostic delays, there have been efforts to identify biomarkers capable of accurately identifying patients with, or at risk of developing, RA-ILD. Candidate biomarkers have included Krebs von den Lungen-6, MMP-7, IP-10, PARC, SP-D, anti-cit-HSP90, and a *MUC5B* promoter variant.[62,69-72] While these have shown promise and have provided important insight into putative pathways driving disease, the availability of these measures has yet to be translated into clinical practice. Of the biomarkers reported to date, some appear to lack specificity for RA-ILD, while others have been subject to limited testing in RA patients

with other lung disease (such as COPD) or have not been applied more broadly to large RA patient populations. Thus, there exists a need for ongoing identification and characterization of biomarkers for RA-ILD.[73]

The pathophysiology of RA-ILD encompasses multiple complex, interrelated processes - inflammation, autoimmunity, fibrosis, and oxidative stress.[44,108] MAA adducts are highly immunogenic products of oxidative stress with the potential to facilitate tolerance loss in the absence of adjuvant.[109] Antibody responses to MAA have been described by our group in RA patients and are associated with both ACPA responses and disease activity.[9] Additionally, MAA co-localizes with citrulline and immune cells in RA synovium. Moreover, both MAA and anti-MAA antibody expression are enriched in RA synovial tissues.[9,110] Beyond its potential contributions to articular disease, MAA has been demonstrated to stimulate inflammation and fibrosis in airway epithelial cells in animal models and *in vitro.*[111,112] Recognizing the pro-inflammatory and pro-fibrotic properties of MAA and our observations of increased anti-MAA antibody responses in RA, we hypothesized that MAA expression and anti-MAA antibody concentrations would be increased in RA-ILD. We tested this hypothesis by comparing circulating anti-MAA antibody concentrations in patients with RA-ILD to other RA patients, including those with other chronic lung conditions. Additionally, we examined MAA expression in lung tissues from RA-ILD, other ILD (non-RA ILD), emphysema, and normal tissues, assessing co-localization with other RA autoantigens as well as immune cells that have been consistently implicated in RA pathogenesis.

3.2 Methods

Study population and samples

Serum analyses were conducted among participants within the VARA registry.[94] The VARA registry is a multi-center prospective observational study of US Veterans with RA fulfilling the 1987 ACR criteria[93] that includes patients from 13 sites. Participants provided informed consent prior to enrollment, all sites obtained local institutional review board approval, and this study obtained approval from the VARA Scientific Ethics and Advisory Committee. At enrollment, participants' demographics, smoking status, education, disease onset, medications, and comorbidities were recorded. At enrollment and follow-up visits, ACR core measures including the MDHAQ,[113] 28-joint tender and swollen joint counts, patient and provider global assessments were collected, acute phase reactants were measured, and composite disease activity measures were scored (DAS28).[114]

Lung tissues were obtained from the National Heart, Lung, and Blood Institute Lung Tissue Research Consortium (https://ltrcpublic.com/). Samples (n=3/group) were obtained following a standard protocol from individuals with RA-ILD, ILD (non-RA; non-specific interstitial pneumonia [n=2] and idiopathic pulmonary fibrosis [n=1]), emphysema (pathologic diagnosis), and controls who underwent transplant procedures, lung volume reduction surgery, or biopsies. The latter control samples were typically collected during evaluation of suspected malignancy and had normal surrounding tissues.

Characterization of lung disease in VARA

ICD 9th and 10th revision, codes (ICD-9: 515, 516.3, 516.8, 516.9, 714.8; ICD-10: M05.1, J84.1, J84.9, J99.0) were used for initial ILD case finding within the VARA registry.[77,78,84,115] Inpatient and outpatient visit diagnoses in the Corporate Data Warehouse were queried within the VA Informatics and Computing Infrastructure.[116] Medical record review was performed within the Compensation and Pension Record

Interchange for all participants with ≥2 outpatient or ≥1 inpatient diagnostic codes for ILD. Diagnoses by provider specialty (pulmonologist, rheumatologist, and other), imaging findings (CT and chest x-ray), lung pathology, PFT results, and corresponding dates were abstracted. Participants were classified as RA-ILD if they had a pulmonologist diagnosis and imaging findings of ILD or if they had a non-pulmonologist diagnosis plus two of the following: CT or chest x-ray findings interpreted by the reading radiologist as ILD, pathology from a lung biopsy consistent with ILD, or interpretation of PFTs as restrictive by the reading pulmonologist. COPD (clinical diagnoses of chronic bronchitis and emphysema) diagnoses were extracted from medical records and recorded in the VARA registry by treating rheumatologists at the time of VARA enrollment. Patients were categorized into one of three mutually exclusive groups: 1) RA-ILD (with or without comorbid COPD), 2) those with COPD in the absence of ILD, 3) neither RA-ILD nor COPD. Recognizing that pathophysiologic processes, radiologic and physiologic abnormalities, and clinical symptoms precede a formal diagnosis of ILD (resulting in diagnostic delays), a two-year span following VARA enrollment (time of serum collection) was used for classifying prevalent ILD.[65,107] We excluded those with indeterminate ILD (physician diagnosis, CT evidence, or biopsy findings but not fulfilling the aforementioned algorithm) (**Figure 3**).

Measurement of serum and tissue analytes

Anti-MAA antibodies (IgA, IgM, and IgG isotypes) were measured by ELISA in VARA participants using banked serum from enrollment, and reported in relative units (RU) as previously described.[9] We categorized anti-MAA antibody values into quartiles to assess trends over the range of values as well as dichotomizing the anti-MAA antibody isotypes into high vs. low concentrations, with the upper three quartiles being considered high (approximating the frequency of other RA-related autoantibodies

including both anti-CCP and RF). Anti-CCP antibodies were measured using a second generation ELISA while RF was measured by nephelometry.[99]

Lung tissues were stained for MAA using an in-house MAA-specific rabbit polyclonal antibody that was labeled with a Zenon 405 reporter (Molecular Probes, Eugene, Oregon) and citrullinated proteins using a citrulline-specific mouse IgM monoclonal antibody, clone F95 (Millipore, Temecula, CA). A Cy™3-conjugated AffiniPure F(ab')2 fragment goat anti-mouse IgM, μ chain specific (Jackson Immuno Research, West Grove, PA) was used as the detection antibody for the F95. Immune cell types (macrophages, T cells, and B cells) were stained using antibodies to CD68 (polyclonal ALEXA FLUOR 594), CD3 (polyclonal ALEXA FLUOR 647), CD19 (polyclonal ALEXA FLUOR 647), and CD27 (polyclonal ALEXA FLUOR 594) (Bioss, Woburn, MA). Tissues were incubated with isotype controls using a rabbit IgG conjugated to ALEXA FLUOR 594 or 647 (Bioss). Based on prior analyses of paired lung and synovial tissues,[55] we also stained for extra-cellular matrix proteins: type II collagen (polyclonal ALEXA FLUOR 488), vimentin (polyclonal ALEXA FLUOR 647), and fibronectin (polyclonal ALEXA FLUOR 555) (Bioss). Tissues were imaged using a confocal laser scanning microscope and staining was quantified using pixel densities, as in prior studies.[9,110]

Statistical analyses

Baseline characteristics were compared between those with RA-ILD, RA+COPD, and RA alone using chi-square or ANOVA. Anti-MAA antibodies were compared between groups using Kruskal Wallis tests with Dunn's post-hoc including a Bonferroni correction. Two multivariable logistic regression models assessed the association between anti-MAA antibody and RA-ILD status (combining RA+COPD with RA alone as the comparator group because there were not significant differences in anti-MAA

antibody concentration between these groups in unadjusted comparisons) with covariates being specified *a priori*. The first (model A) adjusted for known patient characteristics associated with RA-ILD: age, sex, race, and smoking status. The second (model B) included covariates from model A in addition to RA-specific factors reported to be associated with ILD: anti-CCP antibody positivity and disease activity (DAS28).[48,57,66,67] Anti-MAA antibody isotypes were tested in separate models because of collinearity. Missing data were handled by complete-case analysis with complete data available for >98% of participants.

Tissue staining of MAA, citrulline, immune cells, and extracellular matrix proteins was compared between RA-ILD, other ILD, emphysema, and normal tissue controls via ANOVA with a post-hoc Tukey's to account for multiple comparisons. Co-localization of MAA with immune cells and extracellular matrix proteins was determined using the Fiji plugin, Coloc 2 in Image J, as previously reported.[110] To confirm the validity of this approach, we also measured co-localization between MAA and citrulline using Zen blue software (Zeiss, Thornwood, NY) in normal and RA-ILD tissues. Pearson correlations were compared across groups using ANOVA. Results were consistent between both approaches (*r* Coloc2: normal=0.12, RA-ILD=0.79, $P < 0.001$; *r* Zen blue: normal=0.19, RA-ILD=0.72, $P < 0.001$). Thus, the remainder of co-localization analyses were completed using Coloc 2 in Image J. P values < 0.05 were considered statistically significant. Analyses were completed using Stata v15 (StataCorp, College Station, TX).

3.3 Results

Study cohort derivation and characteristics

Of 2,695 patients in the VARA registry, 1,885 had anti-MAA antibody measurements from a prior study (measured on the entire cohort at that time[9]). Diagnostic code screening and subsequent chart review confirmed 90 prevalent ILD cases; an additional 63 participants were excluded because of indeterminate ILD status (**Figure 3**). Baseline characteristics of the eligible participants (n=1,823) in the VARA registry stratified by lung disease status are shown in **Table 7**. Those with RA-ILD were older, more often male, have at least a high school education, seropositive, and to have received bDMARDs or prednisone. Methotrexate use was less frequent in those with RA-ILD. RA patients with COPD were less likely to be Caucasian, to have a high-school education, and were more likely to be current smokers.

Characteristics of RA-ILD cases are shown in **Table 8**. The vast majority of cases were confirmed based on a pulmonologist diagnosis (97.8%) and CT evidence (94.4%). Restrictive PFTs were present in 60.0% and biopsy confirmation was present for 13.3%. ILD was present for a mean of 2.3 years prior to enrollment and attributed to RA in 93.3% of cases. ILD pattern was reported for only 38.9% of cases, with usual interstitial pneumonia being the most common pattern.

Serum anti-MAA antibody and RA-ILD

Median serum concentrations of IgA and IgM anti-MAA antibody were higher among those with RA-ILD than RA alone (**Table 9**; all $P < 0.05$). Additionally, median serum concentrations of IgM anti-MAA antibody were also significantly higher in RA-ILD patients (median 3,582 RU) than patients with RA+COPD (median 2,332 RU; $P = 0.01$). IgG anti-MAA antibody was not significantly different between RA-ILD, RA+COPD, and RA alone ($P = 0.09$).

After multivariable adjustment for patient characteristics and RA-related factors, higher quartiles of IgA and IgM anti-MAA antibody remained significantly associated with RA-ILD (**Table 10**). Notably, inclusion of anti-CCP antibody positivity and DAS28 in multivariable models had minimal impact on the associations between anti-MAA antibody and RA-ILD. High values of IgA anti-MAA antibody, defined by the upper three quartiles, were associated with a more than 2-fold higher odds of RA-ILD (odds ratio [OR] 2.09; 95% CI 1.11-3.90 in fully adjusted model) in the absence of a dose-dependent relationship across quartiles (P for trend = 0.07). As with IgA isotypes, higher values of IgM anti-MAA antibody were also significantly associated with RA-ILD (OR 2.23; 95% CI 1.19-4.15 in fully adjusted model) but demonstrated a dose-dependent relationship between anti-MAA antibody quartiles and prevalent ILD (P for trend = 0.004). The highest two quartiles of IgG anti-MAA antibody trended towards being associated with RA-ILD, though this did not reach statistical significance (P = 0.15 and 0.17). We assessed all three isotypes together by categorizing individuals according to the number of positive anti-MAA antibody isotypes. Individuals with 3 positive isotypes had 2.5-fold higher odds of RA-ILD than those with 0-1 positive isotype (OR 2.56; 95% CI 1.29-5.09).

Lung tissue patient characteristics

Mean (standard deviation [SD]) age of participants with tissue samples was 56.4 (11.7) years with 75.0% being female. A smoking history was present overall in 66.7% of patients (100% of other ILD and emphysema, 33.3% of normal and RA-ILD). Mean (SD) pack-years of smoking history was 17.5 (14.3). Anti-CCP antibodies and IgM RF were positive in two of three RA-ILD patients. Anti-CCP antibodies, but not IgM RF, were additionally detected in one of three other ILD patients.

MAA and citrulline expression in lung tissue

MAA expression was highest in RA-ILD lung tissues (**Figure 4A** and **4B**, P < 0.001 vs. all other groups). Citrulline was also higher in RA-ILD lung tissues (**Figure 4C** and **4D**) relative to normal and other ILD lung tissues (P < 0.001), but not significantly different than emphysematous lung tissue (P = 0.91). Expression of both MAA and citrulline was highly co-localized in RA-ILD lung tissue (**Figure 4E** and **4F**; *r*=0.79), significantly higher than in lung tissues from other patient groups (P < 0.001 vs. normal [*r*=0.12] and other ILD [*r*=0.38], P = 0.002 vs. emphysema [*r*=0.47]).

Co-localization of MAA and citrulline with immune cells in lung tissue

Staining for CD68+ macrophages and CD3+ T cells was higher in all diseased tissues relative to normal lung tissue (**Figure 5A**; all P < 0.01). Macrophage staining was higher in other ILD than in RA-ILD and emphysema (P < 0.05). In contrast, CD19+ and CD27+ (memory) B cells were more abundant in RA-ILD lung tissues than tissues from all other groups (P ≤ 0.02). There was minimal to moderate co-localization between MAA and macrophages or T cells (*r* values 0.12 to 0.54), with no significant differences between lung tissue types (**Figure 5B**; all P > 0.10). In contrast, we observed strong co-localization of MAA with CD19+ B cells, with the highest correlation identified in RA-ILD (*r*=0.78; P ≤ 0.02 vs. all other lung tissues). Co-localization of MAA with CD27+ B cells was more modest (*r* vales 0.02 to 0.30), with other ILD yielding the highest correlation (*r*=0.30; P ≤ 0.004 vs. RA-ILD and normal, P = 0.06 vs. emphysema).

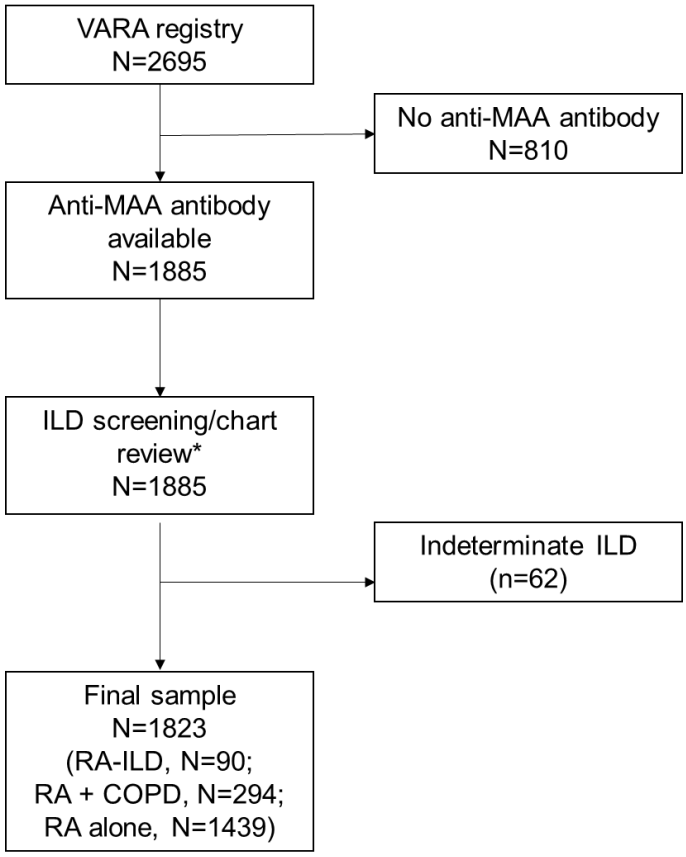
Citrulline co-localized with CD68+ macrophages to a greater degree in RA-ILD (**Figure 5C**; P = 0.04) and emphysema (P < 0.001) than in normal lung tissue. There was minimal co-localization of citrulline with T cells (*r* vales 0.07 to 0.18). There was moderate co-localization of citrulline with CD19+ B cells in both RA-ILD (*r*=0.53) and other ILD (*r*=0.44) that exceeded the degree of co-localization observed for emphysema and normal tissues (P < 0.01). Co-localization of citrulline with CD27+ (memory) B cells

was highly prevalent in diseased lung tissue (all P < 0.001 vs. normal) but not different between specific types of diseased lung tissues (all P > 0.29).

Co-localization of MAA with extracellular matrix proteins

Staining for type II collagen was higher in RA-ILD and other ILD than normal lung tissues (**Figure 6A**; P ≤ 0.002). However, co-localization of MAA with type II collagen was greater in RA-ILD (*r*=0.72) compared with other lung tissues (**Figure 6B**; *r*=0.12-0.49; all P ≤ 0.02). Fibronectin staining was higher in both RA-ILD and emphysema relative to normal lung tissues (P ≤ 0.03) with only weak co-localization of MAA and fibronectin in RA-ILD (*r*=0.21). Vimentin staining was higher in all diseased lung tissues compared to normal lung tissue (all P ≤ 0.03), although co-localization of MAA and vimentin was higher in RA-ILD than other ILD (P < 0.001) without significant differences compared to other lung tissues (all P ≥ 0.09).

Figure 3. Serum anti-MAA antibody study cohort derivation



Participants in the Veterans Affairs Rheumatoid Arthritis (VARA) registry with available anti-malondialdehyde acetaldehyde adduct (MAA) antibody measurement were screened for interstitial lung disease (ILD) using outpatient and inpatient diagnostic codes. Detailed chart review was completed to confirm ILD diagnosis if ≥2 outpatient or ≥1 inpatient diagnostic codes were identified (*), confirming 90 ILD cases.

Abbreviations: COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease; MAA, malondialdehyde-acetaldehyde adduct; VARA, Veterans Affairs Rheumatoid Arthritis registry

Table 7. Baseline characteristics of VARA participants by lung disease status

	Overall (n=1823)	RA-ILD (n=90)	RA + COPD (n=294)	RA alone (n=1439)	P value
Age, years	63.5 (11.0)	67.0 (9.9)	65.8 (9.7)	62.8 (11.3)	<0.001
Male sex	90.1	95.6	92.5	89.2	0.05
Caucasian	76.7	76.7	83.7	76.2	0.02
HS education	86.4	91.7	78.9	87.5	<0.001
Smoking status					<0.001
Current	26.1	27.8	31.0	25.1	
Former	53.4	58.9	58.8	52.0	
Never	20.4	13.3	10.2	23.0	
BMI, kg/m^2	28.4 (5.7)	27.8 (5.1)	28.3 (6.1)	28.4 (5.7)	0.67
RDCI score	1.9 (1.5)	3.2 (1.6)	3.9 (1.1)	1.4 (1.2)	<0.001
RA duration	11.1 (11.5)	13.3 (13.1)	11.1 (11.9)	10.9 (11.3)	0.17
SE positive	68.8	65.6	73.0	68.2	0.22
Anti-CCP positive	77.3	86.7	80.3	76.0	0.03
RF positive	79.8	92.2	80.6	78.9	0.009
MDHAQ	0.9 (0.6)	0.9 (0.5)	1.1 (0.6)	0.9 (0.6)	0.004
DAS28	4.0 (1.6)	4.1 (1.4)	4.4 (1.5)	3.9 (1.6)	0.003
Methotrexate	51.9	21.0	47.6	54.7	<0.001
Biologic	22.9	30.0	16.3	23.8	0.005
Prednisone	43.5	63.0	43.1	42.4	0.01

Values mean (SD) or %. P values test of group differences by ANOVA or chi-square tests
Abbreviations: RA, rheumatoid arthritis; ILD, interstitial lung disease; COPD, chronic obstructive pulmonary diseases; HS, high-school; BMI, body mass index; RDCI, rheumatic disease comorbidity index; SE, shared epitope; anti-CCP, anti-cyclic-citrullinated peptide antibody; RF, rheumatoid factor; MDHAQ, multidimensional health assessment questionnaire; DAS28, 28-joint disease activity score; VARA, Veterans Affairs Rheumatoid Arthritis Registry

Table 8. Characterization of RA-ILD cases

	Mean (SD) or N (%) of Validated ILD cases
N	90
CT evidence of ILD	88 (97.8)
Pulmonologist diagnosis	85 (94.4)
Restrictive PFTs	54 (60.0)
Biopsy confirmation	12 (13.3)
Duration of ILD, years	2.3 (3.8)
Reason for ILD includes RA	84 (93.3)
Pattern of ILD	
Usual interstitial pneumonia	26 (28.9)
Non-specific interstitial pneumonia	4 (4.4)
Other	5 (5.6)
Unknown/missing	55 (61.1)
PFT closest to enrollment (n available)	
FVC % predicted (n=62)	75.1 (17.3)
FEV1 % predicted (n=63)	74.5 (17.1)
FEV1/FVC ratio (n=60)	74.8 (9.1)
TLC % predicted (n=42)	80.8 (19.9)
DLCO % predicted (n=58)	54.9 (17.3)

Abbreviations: CT, computed tomography; ILD, interstitial lung disease; PFT, pulmonary function test; RA, rheumatoid arthritis; FVC, forced vital capacity; FEV1, forced expiratory volume in 1 second; TLC, total lung capacity; DLCO, diffusion capacity

Table 9. Anti-MAA antibody concentrations by lung disease status in RA subjects

Anti-body	RA-ILD (n=90)	RA + COPD (n=294)	RA alone (n=1439)	P value*
IgA	891 (501, 1624) †	869 (399, 1665) †	689 (323, 1440)	0.005
IgM	3582 (1302, 11141) † ±	2332 (888, 5649)	2094 (843, 5610)	0.005
IgG	2226 (1353, 3781)	1996 (1039, 3701)	1868 (943, 3415)	0.09

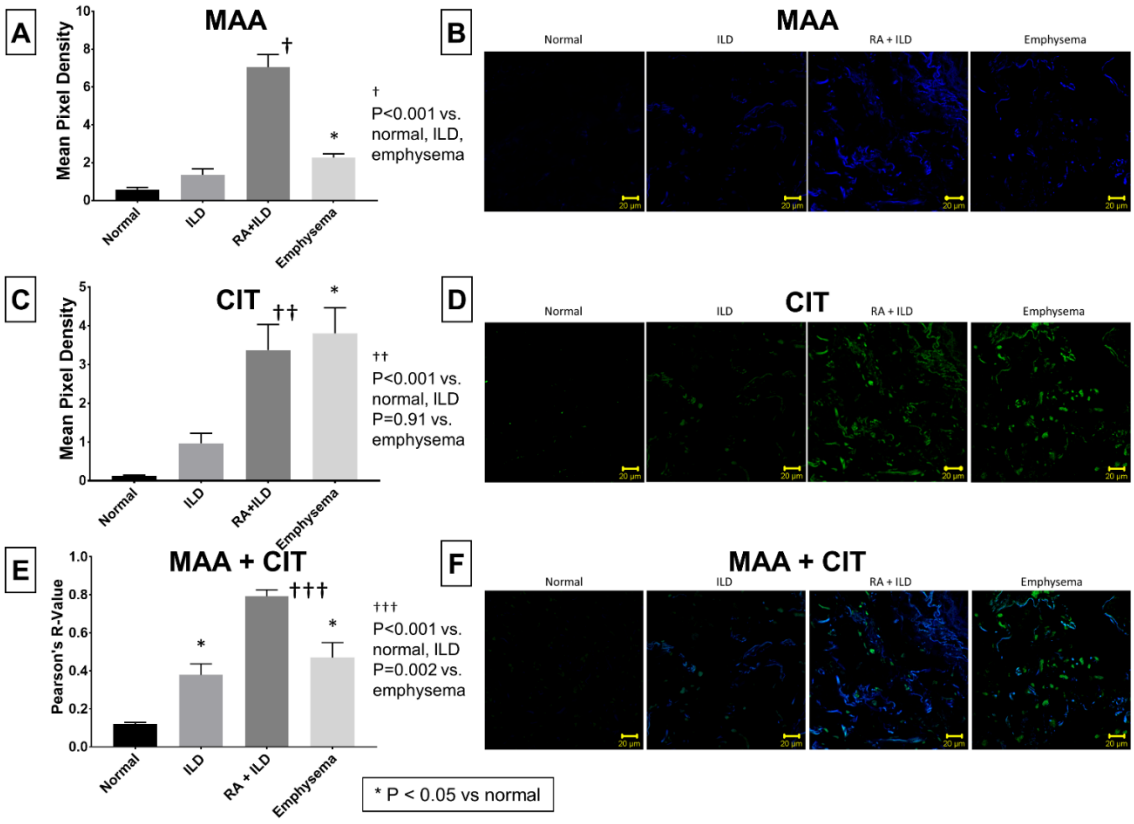
Values represent median (interquartile range) in relative units (RU)
* P value by Kruskal Wallis (unadjusted comparisons)
† $p < 0.05$ vs. RA alone (Dunn's test with Bonferroni correction)
± $p < 0.05$ vs. RA + COPD (Dunn's test with Bonferroni correction)
Abbreviations: RA, rheumatoid arthritis; ILD, interstitial lung disease; COPD, chronic obstructive pulmonary disease; anti-MAA, anti-malondialdehyde acetaldehyde adduct antibodies; RU, relative units

Table 10. Multivariable associations of anti-MAA antibody with RA-ILD

	Model A. Age, sex, race, smoking status (n=1820)		***Model B***. Model A + anti-CCP positivity and DAS28 (n=1792)	
	OR (95% CI)	P value	OR (95% CI)	P value
Quartiles				
IgA anti-MAA				
Quartile 1	Referent	-	Referent	-
Quartile 2	2.27 (1.12, 4.59)	0.02	2.09 (1.03, 4.27)	0.04
Quartile 3	2.20 (1.09, 4.43)	0.03	2.07 (1.02, 4.18)	0.04
Quartile 4	2.26 (1.12, 4.56)	0.02	2.10 (1.04, 4.25)	0.04
P trend		0.04		0.07
IgM anti-MAA				
Quartile 1	Referent	-	Referent	-
Quartile 2	1.87 (0.91, 3.86)	0.09	1.84 (0.89, 3.81)	0.10
Quartile 3	2.26 (1.11, 4.60)	0.03	2.08 (1.02, 4.27)	0.05
Quartile 4	2.93 (1.49, 5.78)	0.002	2.73 (1.38, 5.41)	0.004
P trend		0.001		0.004
IgG anti-MAA				
Quartile 1	Referent	-	Referent	-
Quartile 2	1.34 (0.69, 2.61)	0.39	1.33 (0.68, 2.59)	0.41
Quartile 3	1.73 (0.91, 3.27)	0.09	1.61 (0.84, 3.06)	0.15
Quartile 4	1.67 (0.88, 3.18)	0.12	1.58 (0.83, 3.02)	0.17
P trend		0.09		0.14
Antibody positive				
IgA anti-MAA	2.24 (1.20, 4.18)	0.01	2.09 (1.11, 3.90)	0.02
IgM anti-MAA	2.35 (1.26, 4.38)	0.007	2.23 (1.19, 4.15)	0.01
IgG anti-MAA	1.58 (0.91, 2.75)	0.11	1.50 (0.86, 2.63)	0.15

* Antibodies tested in separate logistic regression models
Abbreviations: MAA, malondialdehyde-acetaldehyde adducts; RA, rheumatoid arthritis; ILD, interstitial lung disease; anti-CCP, anti-cyclic-citrullinated peptide antibody; DAS28, 28-joint disease activity score; OR, odds ratio; CI, confidence interval

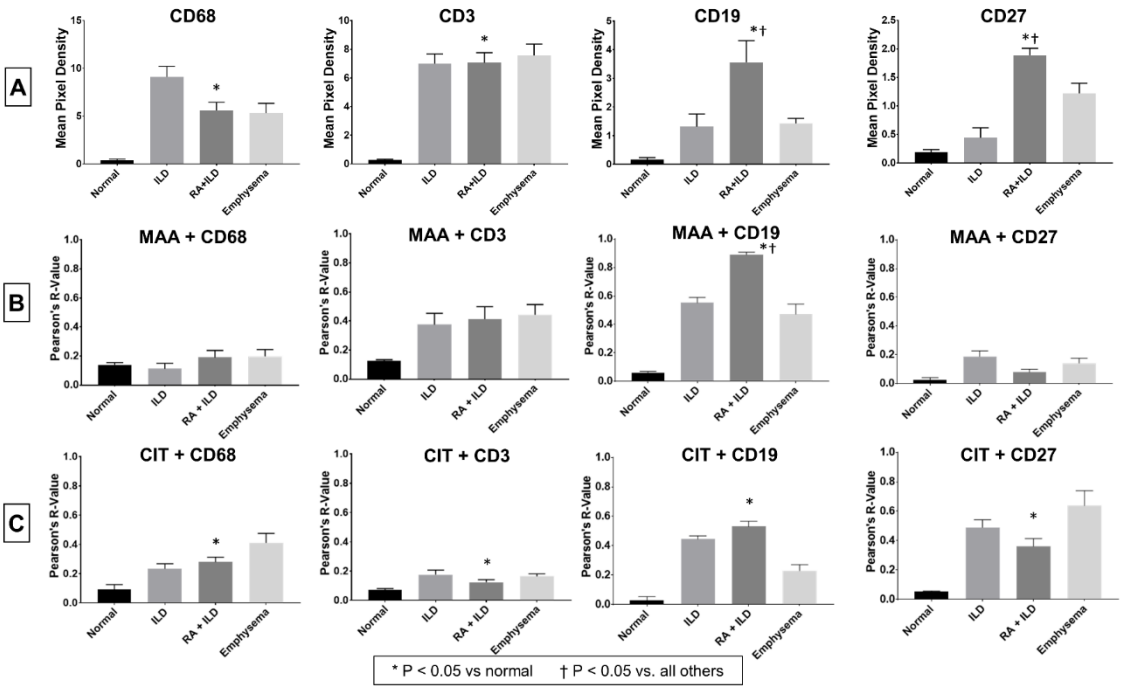
Figure 4. Lung tissue expression of MAA, citrulline, and their co-localization in RA-ILD and other lung diseases



Expression of malondialdehyde-acetaldehyde adducts (MAA) measured by pixel density (Figure 4A) and representative immunohistochemistry staining of lung tissues for MAA (Figure 4B). Tissue expression (Figure 4C) and immunohistochemistry staining of lung tissues for citrulline (Figure 4D) are also shown. The co-localization of MAA and citrulline was quantified through a correlation coefficient of their staining (Figure 4E) and overlapping immunohistochemistry staining are shown (Figure 4F).

Abbreviations: MAA, malondialdehyde-acetaldehyde adducts; CIT, citrulline; RA-ILD; rheumatoid arthritis interstitial lung disease; ILD, interstitial lung disease.

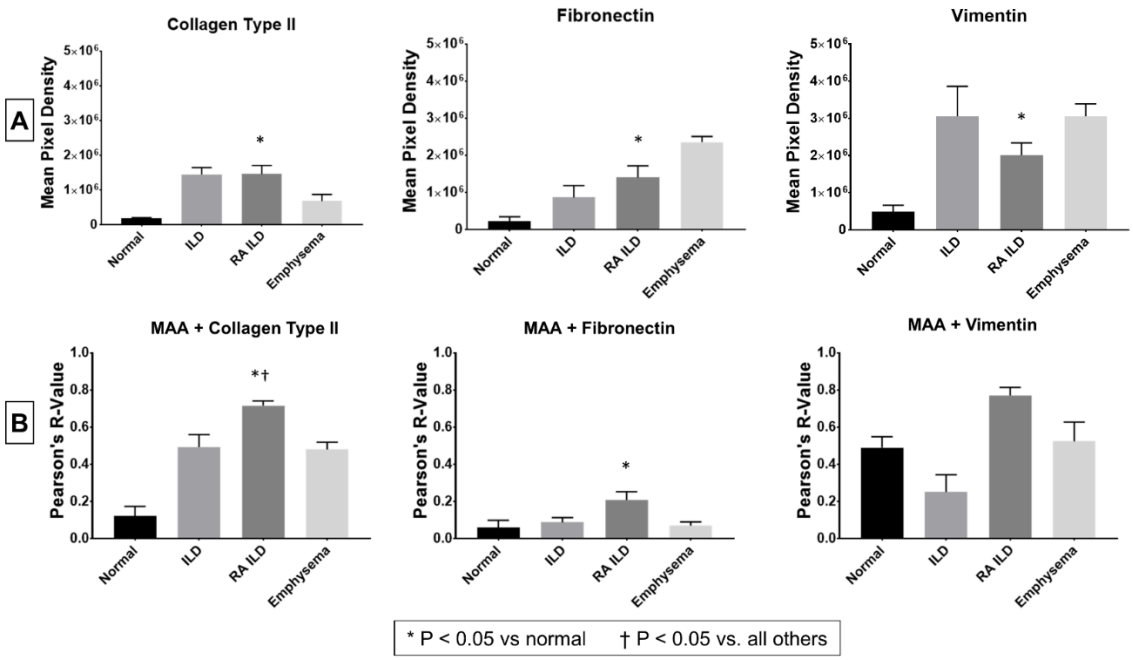
Figure 5. Co-localization of MAA and citrulline with immune cells in lung tissue from RA-ILD and other lung diseases



Tissue staining for macrophage (CD68), T cells (CD3), and B cells (CD19 and CD27) for RA-ILD, other ILD, emphysema, and healthy control lung tissues (Figure 5A). Co-localization of MAA with macrophage, T cells, and B cells in different lung tissues (Figure 5B). Co-localization of citrulline with macrophage, T cells, and B cells in different lung tissues (Figure 5C).

Abbreviations: MAA, malondialdehyde-acetaldehyde adducts; CIT, citrulline; RA-ILD; rheumatoid arthritis interstitial lung disease; ILD, interstitial lung disease.

Figure 6. Co-localization of MAA with extracellular matrix proteins in lung tissue from RA-ILD and other lung diseases



Tissue staining for extracellular matrix proteins (type II collagen, fibronectin, and vimentin) in RA-ILD, other ILD, emphysema, and healthy control lung tissues (Figure 6A). Co-localization of MAA with extracellular matrix proteins in different lung tissues (Figure 6B).

Abbreviations: MAA, malondialdehyde-acetaldehyde adducts; CIT, citrulline; RA-ILD; rheumatoid arthritis interstitial lung disease; ILD, interstitial lung disease.

3.4 Discussion

ILD complicates the disease course for 5-15% of RA patients,[57,65,84,106] resulting in potentially devastating complications of functional decline and premature mortality. Enhancing the identification of RA-ILD is an important area of translational research in RA, with serum biomarkers emerging as candidates to fulfill this need. For the first time, we investigated serum anti-MAA antibody as a potential biomarker of RA-ILD and characterized the expression of MAA in lung tissues from RA-ILD patients. We found that IgA and IgM anti-MAA antibody concentrations were higher in RA-ILD patients than in other RA patients, including those with other forms of chronic lung disease (IgM only). In parallel studies, we found MAA adduct expression to be higher in RA-ILD lung tissues than in other chronic lung diseases including other ILD. Importantly, MAA adducts demonstrated marked co-localization with citrulline, CD19+ B cells, and type II collagen that was preferential to RA-ILD lung tissues. This study is among the first to characterize a biomarker for RA-ILD that has leveraged a comparator population incorporating RA patients with other chronic lung diseases that may be overrepresented in RA.[45] Together, our findings suggest that MAA modified proteins and resulting immune responses may serve as useful biomarkers for RA-ILD and that MAA modified proteins may contribute to the pathogenesis of RA-ILD.

Serum biomarkers have been increasingly investigated for their potential role in identifying RA-ILD. Protein candidates have included widely used biomarkers in RA (anti-CCP antibody and RF),[48,66,117] novel autoantibodies (anti-cit-HSP90),[70] cytokines/chemokines (MMP-7, IP-10, PARC),[71,72] and SP-D.[72] Oxidative stress represents a potentially relevant biologic pathway that has not been harnessed in prior biomarker studies of RA-ILD. Oxidative stress, a disruption of the balance of free radicals and antioxidants, is believed to be intimately involved with the development of

diffuse lung diseases because of the continuous exposure to oxygen, high surface area, and robust blood supply in the lungs. MAA, which is generated from lipid peroxidation during oxidative stress, has the potential to link multiple pathways implicated in RA-ILD pathogenesis - oxidative stress, autoimmunity, inflammation, and fibrosis. MAA induces tolerance loss,[109] elicits robust adaptive immune responses (anti-MAA antibody), and upregulates pro-inflammatory and pro-fibrotic pathways.[111,112] Our study importantly begins to characterize lung tissue expression of MAA in different lung disease states as well as serum anti-MAA antibody responses in RA patients with and without lung diseases. Confirming our hypothesis, MAA expression in lung tissue and serum anti-MAA antibody concentrations were highest in RA-ILD patients.

Although we found over 2-fold higher odds of ILD among RA patients with serum IgA or IgM anti-MAA antibody concentrations in the upper three quartiles, it is important to note that these antibodies are not specific for RA-ILD. Anti-MAA antibodies are present in RA patients in the absence of chronic lung disease, as well as other disease states.[10] However, specificity of a candidate biomarker of RA-ILD may be less important than initial case finding, given that CT and PFTs are ultimately needed to confirm the presence and subtype of ILD (which influences prognosis). Translating these novel findings of anti-MAA antibody in RA-ILD into clinical practice will require additional work. As several other serum biomarkers have shown promise for identifying RA-ILD, biomarker panels that include anti-MAA antibody and other analytes are likely to outperform models based on a single analyte. To date, the measurement of anti-MAA antibody has leveraged the use of adducted albumin as the plating antigen, a protein that has no known pathogenic role in RA. Identification of the precise antigenic targets of anti-MAA antibody is likely to allow for improved assay performance in identifying RA patients with ILD. Finally, our current results assessed the ability of anti-MAA antibody to

identify established RA-ILD. Future study will need to assess the value of anti-MAA antibody for predicting future RA-ILD risk. This could be of even greater value than identifying prevalent RA-ILD, as it may identify patients with earlier disease that might be more amenable to therapeutic and/or preventative interventions,[118] though data specifically in RA-ILD is lacking.

Paralleling serum findings, staining for MAA adducted antigens was highest in lung tissues from RA-ILD patients. Importantly, this occurred preferentially in RA-ILD lung tissue, with significantly higher staining than in other ILD and emphysema. In contrast to MAA, citrulline was expressed in both RA-ILD and emphysema. Although the specificity of serum anti-CCP antibodies for RA approaches 96%,[119] others have similarly found citrulline and ACPA responses to accompany chronic obstructive lung diseases in the absence of RA.[49,120,121] Given the strong co-localization of MAA with citrullinated antigens in RA-ILD, we postulate that MAA could act as a “second hit” in RA pathogenesis by facilitating tolerance loss to co-localized citrullinated antigens. Although further testing will be needed to address this hypothesis, the co-localization of CD19+ B cells with MAA and citrulline would support the concept that these post-translational changes (both of which likely result from injurious stimuli) conspire in autoantibody generation. This is further supported by preliminary work in animal models suggesting immunization with co-modified (MAA+citrulline) albumin leads to greater ACPA responses than citrullinated-albumin alone.[122] Finally, vimentin is an extracellular matrix protein that has previously been shown to be a shared target of citrullination/ACPAs in the synovium and lung.[55] While we did not find vimentin expression to be increased in RA-ILD compared to other lung conditions, we observed marked co-localization of MAA with vimentin in RA-ILD lung tissues, co-localization that was significantly more robust than that seen with other ILD.

Our group has previously characterized anti-MAA antibodies in sera from RA and other rheumatic and musculoskeletal disease patients.[9,10] Circulating anti-MAA antibody concentrations are higher in RA patients than those with osteoarthritis, are associated with serum ACPAs, and are enriched within RA synovium.[9,110] As we found in RA-ILD lung tissues assessed in this study, MAA and citrulline co-localized in RA synovium.[110] Also paralleling the RA-ILD lung findings from the present study, prior work by our group has shown that MAA and citrulline both co-localize with B cells in the synovium. However, there are differences in B cells implicated by site. In the synovium, MAA and citrulline co-localized most strongly with CD27+ memory B cells.[110] In the lung tissues from RA-ILD patients, MAA co-localized most strongly with CD19+ B cells, but not with CD27+ memory B cells. While future work will be needed to elucidate the temporal evolution of immune responses to MAA, it is intriguing that immature B cells are associated most strongly with MAA adduct expression in the lung given the emerging evidence that the lungs may be a site of immune tolerance breakdown contributing to the early development of RA.[56]

There are limitations to this study. The male predominance, Veteran status, and lower prevalence of bDMARD use may affect generalizability. Collection of ILD data was obtained retrospectively and not all data were available within the medical records. This may underestimate the cross-sectional prevalence of ILD in the cohort (4.7%). However, misclassification of ILD cases as non-ILD would bias our results towards the null. Distinguishing between clinical and sub-clinical ILD cannot be definitive based on retrospective classification. By confirming physician diagnoses in the medical records, rather than relying on diagnostic codes or diagnostic testing alone, we believe the majority of ILD cases were clinically evident. Given the low frequency with which ILD pattern (usual interstitial pneumonia vs. non-specific interstitial pneumonia vs. other) was

specified, we were not able to compare anti-MAA concentrations by RA-ILD pattern. Likewise, anti-MAA antibody measurements were not available for all registry participants, which may also have reduced study power. Again, this should not have introduced bias, as antibody measurements were performed on the entire cohort at the time of the prior study without any relation to ILD status. Reflecting the prevalence of seropositivity for RF and anti-CCP antibody, we dichotomized anti-MAA antibody as being in the upper three quartiles. Only increasing IgM anti-MAA antibody quartiles were more strongly associated with the presence of ILD. Further work will be needed to determine clinically important cut-offs for these antibodies. Sample sizes were limited for lung tissue studies, with lung tissues obtained from three individuals with each lung condition, prohibiting multivariable analyses. One of the non-RA ILD patients had detectable ACPAs but was not classified as RA. Given the cross-sectional nature of the study, it is unknown if that patient later developed RA. This potential misclassification of RA-ILD as non-RA ILD would only bias our results towards the null. Lung tissue samples were not matched, so there may be unmeasured confounding.

There are important strengths to this study. We performed a detailed review of the medical records to validate ILD diagnoses in RA patients from a well characterized registry that includes robust data including many relevant covariates.[94] We evaluated not only serologic anti-MAA antibody concentrations, but also investigated tissue expression of MAA and its co-localization with citrulline, immune cells, and extracellular matrix proteins that have been consistently implicated in disease pathogenesis. Finally, we characterized MAA and anti-MAA immune responses in RA-ILD by using comparators that were free of lung disease in addition to comparators with other chronic lung diseases.

In conclusion, we found higher levels of serum IgA and IgM anti-MAA antibody to be associated with RA-ILD in a large cohort of U.S. Veterans with RA. Lung tissue expression of MAA is similarly higher in RA-ILD lung tissue where it co-localizes with citrulline, CD19+ B cells, and extracellular matrix proteins. These findings suggest that MAA immune responses could play an important role in the pathogenesis of RA-ILD and anti-MAA antibodies may be promising serum biomarkers in the identification of this extra-articular disease manifestation.

CHAPTER 4: IDENTIFICATION OF INTERSTITIAL LUNG DISEASE WITH BIOMEDICAL INFORMATICS

4.1 Background

ILD clinically affects between 5-15% of RA patients resulting in poor long-term outcomes including reduced survival and greater functional disability.[47,48,57] Given disease heterogeneity and lack of well-characterized classification criteria for RA-ILD, the case definitions used and prevalence estimates reported for RA-ILD are highly variable between studies.[47,57,65,123] Administrative data sources are increasingly being utilized in RA outcomes research, primarily to facilitate investigations examining predictors of disease-related outcomes or the safety and effectiveness of DMARDs.[124] Yet, only a few studies have begun to leverage these large administrative databases to study RA-ILD.[63,77,78,84,125]

Prior studies utilizing administrative databases for RA-ILD research have constructed RA-ILD cohorts or identified ILD outcomes in RA patients using claims databases,[63,77,78] death records,[84] or national patient registries.[125] All have used diagnostic codes for RA and ILD in combination, though additional requirements for RA diagnosis such as DMARD receipt and specific ILD diagnostic codes selected has varied between studies. To enhance specificity, authors have required ILD diagnostic tests[77] or excluded other causes of ILD (e.g., sarcoidosis, hypersensitivity pneumonitis, pneumoconioses, etc.).[63] However, the validity of these algorithms has received only limited attention,[126] hampering wider adoption of these methods for studying RA-ILD. With validated ILD algorithms, large administrative datasets could be leveraged for comparative effectiveness research and epidemiologic analyses, while deployment in

electronic health records could enhance recruitment into patient registries or clinical trials.

The objective of this study was to develop and evaluate the performance of several different administrative algorithms for the identification of ILD in a multi-center RA registry. We hypothesized that administrative algorithms that included multiple ILD diagnostic codes, a pulmonologist diagnosis, procedure codes for CT of the chest, PFTs, or lung biopsy, and exclusion of other causes of ILD would accurately classify RA-ILD compared to a comprehensive review of medical records.

4.2 Methods

Patient selection

We selected subjects enrolled in the VARA registry, a multi-center, prospective cohort study of U.S. Veterans with RA initiated in 2003.[94] All subjects fulfilled the 1987 ACR criteria for RA.[93] Participants provided informed consent prior to enrollment and all sites (n=13) obtained local institutional review board approval. This study obtained approval from the VARA Scientific Ethics and Advisory Committee.

To enrich the study sample with ILD cases, we performed stratified subsampling through initial ILD screening. We queried national VA data in the Corporate Data Warehouse to identify VARA participants with ≥1 inpatient or ≥2 outpatient ILD diagnostic codes (>30 days apart) from health care providers (physicians, physician assistants, and advanced practice nurses). ICD, 9th and 10th revision, Clinical Modification codes were selected from those previously proposed to ascertain ILD status or closely related codes (**Appendix D**).[77,78,84,115,125,126] We performed detailed, systematic medical record review on all subjects identified through initial screening (n=293) and a

random sample of all VARA subjects not identified by the ILD screening method (n=243) so as to be able to comment on the sensitivity of the selected ICD codes and ILD algorithms.

ILD data abstraction

Data was abstracted from the electronic medical records using the Compensation and Pension Record Interchange in a standardized fashion by three rheumatologists blinded to the results of the administrative algorithms using Research Electronic Data Capture (REDCap).[127,128] Regardless of ILD screening status, participants' outpatient and inpatient clinical notes, imaging reports, pathology reports, and PFT results from the earliest available date in the medical record were reviewed and recorded. Data abstracted included pulmonologist diagnoses, other physician diagnoses, chest CT results, chest x-ray results, PFT results, lung biopsy results, as well as dates corresponding to the aforementioned items. To ensure consistency between reviewers, charts were reviewed in sets of 5 in duplicate until >95% agreement on abstracted data was obtained between reviewers. As our reference standard, participants were classified as RA-ILD by medical record review using both stringent and relaxed ILD definitions. The stringent definition classified participants as RA-ILD if they had a pulmonologist diagnosis and imaging (chest CT or x-ray) findings of ILD or if they had a non-pulmonologist provider diagnosis plus two of the following: chest CT or x-ray findings interpreted by the reading radiologist as ILD, pathology from a lung biopsy consistent with ILD, or interpretation of PFTs as restrictive by the reading pulmonologist. The relaxed ILD definition additionally classified subjects as ILD who had a provider diagnosis of ILD (pulmonologist or non-pulmonologist) and either imaging findings consistent with ILD or pathology demonstrating ILD.

Algorithm development

We queried National VA data within the Corporate Data Warehouse from January 1, 1999 to August 31, 2018 for all participants to obtain the necessary components of each ILD algorithm. Data queried included inpatient and outpatient encounters in the VA, inpatient and outpatient encounters occurring outside the VA and billed to the VA, specialty of outpatient encounters, and outpatient and inpatient procedures.

We tested the characteristics of possible administrative ILD algorithms in four stages. In the first stage, we tested the performance of algorithms using different encounter types (inpatient vs. outpatient) and frequency of ILD diagnostic codes (≥1 vs. ≥2). In stage 2, we compared different ICD-9 and ICD-10 code sets (**Appendix D**). These code sets were created by removing ICD-9 and ICD-10 codes with descriptions including “unspecified” or “other”, those pertaining to rheumatoid lung, and those not consistently included in prior studies. Stage 3 testing compared algorithms that incorporated additional data available in administrative datasets that may improve algorithm specificity. These additional data were provider specialty on the ICD-9/10 diagnoses, and procedure codes for chest CT, PFTs, and lung biopsy procedures. Current Procedural Terminology (CPT) and ICD Procedural Codes for chest CT and lung biopsy were adapted from those used in IPF algorithms (**Appendix E**) [129,130]. In addition to CPT and ICD procedure codes, we also identified PFTs through the use of stop codes in National VA data, which designate clinical services provided in PFT labs. In the final stage (stage 4), we excluded other causes of ILD recorded after the final ILD diagnosis using codes for pneumoconioses, radiation pneumonitis, hypersensitivity pneumonitis, and other connective tissue diseases (**Appendix F**).

Statistical analysis

Enrollment characteristics of the VARA patients selected for these analyses were assessed descriptively and stratified by medical record review ILD classification status. Agreement between ILD algorithms and medical record review classification was assessed with percent agreement and Kappa statistics. Levels of agreement based on the Kappa statistic were interpreted as near perfect (values of 0.8-1.0), substantial (0.6-0.8), moderate (0.4-0.6), fair (0.2-0.4), or slight (0.0-0.2).[131] We also calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) along with 95% CIs for each algorithm, treating medical record classification as the reference standard. All analyses accounted for the subsampling from the overall VARA registry by the use of inverse probability weighting (R package CompareTests).[132] This ensured that the prevalence of ILD in weighted analyses was consistent with the overall cohort. Algorithm selection through each stage was based on optimal Kappa values. Several sensitivity analyses were performed testing variations of administrative ILD algorithms and using medical record ILD definitions with fewer requirements. Analyses were conducted using Stata v15 (StataCorp, College Station, TX) and R version 3.5.1 within the VA Informatics and Computing Infrastructure. We report our study in accordance with proposed reporting guidelines for assessing the quality of validation studies of health administrative data.[133]

4.3 Results

Enrollment characteristics

We identified 293 subjects in the VARA registry who met the initial ILD screening criteria and randomly selected 243 VARA participants who did not screen positive for ILD (**Figure 7**). Detailed medical record review performed on all 536 of these subjects confirmed 182 and 203 ILD cases using stringent and relaxed ILD definitions,

respectively. Patient characteristics were reflective of the overall VARA registry and the VA population with a male predominance and mean age at enrollment in the 7th decade of life (**Table 11**). Those with ILD were older, more frequently RF positive, less likely to be treated with methotrexate, and more likely to receive prednisone at enrollment.

The majority of ILD cases occurred among those who screened positive for ILD (97.3% stringent and 96.6% relaxed), had a pulmonologist diagnosis (94.5% stringent and 84.7% relaxed), and had CT evidence of ILD (98.4% stringent and 96.1% relaxed) (**Table 11**). Approximately half of the ILD cases were prevalent at the time of enrollment into the registry, and the initial date of ILD diagnosis occurred after implementation of ICD-10 in 17 cases (21 cases relaxed ILD definition). Among non-ILD cases, pulmonologist ILD diagnosis was present in 1.4-1.5%, non-pulmonologist ILD diagnosis was present in 8.5% (stringent) and 2.7% (relaxed), and CT evidence of ILD was present in 11.3% (stringent) and 7.2% (relaxed).

Stage I: Frequency of diagnosis codes and encounter types

Performance of eight different algorithms (1A to 1H) reflecting differences in frequency, encounter types, and date ranges for ILD diagnosis codes in classifying ILD is shown in **Table 12**. Kappa was greatest for algorithms 1D (0.71) and 1F (0.70). Performance was similar in classifying ILD with the relaxed definition (Kappa 0.71). Sensitivity ranged from 76.3-81.7% and specificity ranged from 96.0-97.1%, but PPV of these algorithms were modest (65.5-73.9%). Because of their equivalent performance, Algorithm 1F, which required ≥2 diagnosis codes ≥30 days apart from either inpatient or outpatient encounters, was selected for further testing.

Stage II: Diagnosis code selection

Exclusion of ICD-10 codes J84.2 and J99 did not result in any difference in ILD classification (**Table 13**). Exclusion of rheumatoid lung codes (ICD-9: 714.81; ICD-10: M05.1x) minimally attenuated sensitivity and NPV while improving specificity and PPV. Kappa was improved from the all-inclusive ICD code algorithm when rheumatoid lung codes were excluded. Medical record review identified RA-related pleural effusions and pulmonary nodules as reasons for these codes in the absence of ILD. Algorithm performance measured by Kappa worsened when “unspecified” and “other” ILD codes were excluded. Based on these performance characteristics, we constructed algorithm 2H with the following ICD codes: ICD-9 515.x, 516.3, 516.8, 516.9 and ICD-10 J84.1, J84.89, J84.9. This algorithm had the best Kappa (0.72), specificity (96.8%, 97.3% relaxed ILD definition), and PPV (69.5%, 75.3% relaxed ILD definition), with minimal attenuation of sensitivity (80.6%, 74.4% relaxed ILD definition). Algorithm 2H was thus used for further comparisons in Stage III testing.

Stage III: Provider specialty and diagnostic studies

The additional requirement of a pulmonologist diagnosis (algorithm 3A) increased the specificity from 96.8% (algorithm 2H) to 98.5%, PPV from 69.5% to 79.9%, and had substantial agreement by Kappa (0.68; 0.63 relaxed ILD definition) (**Table 14**). Requiring a rheumatologist diagnosis (algorithm 3B) modestly improved specificity but reduced sensitivity and overall algorithm performance by Kappa. Algorithms requiring a CT or PFTs between 7 and 180 days prior to ILD diagnosis (algorithms 3C and 3D) also modestly improved specificity and PPV while reducing sensitivity and NPV. Kappa for these algorithms were 0.72-0.74 (0.69-0.73 relaxed ILD definition). The requirement of a lung biopsy (algorithm 3E) was highly specific (99.9%) but poorly sensitive (9.3%), resulting in a PPV of 87.5% and only slight agreement by Kappa (0.15). Requiring a CT, PFTs, or lung biopsy in addition to a pulmonologist diagnosis (algorithm 3F and 3G)

modestly affected specificity, PPV, and Kappa. As sensitivity was reduced with the requirement of a pulmonologist diagnosis (63.7%), we tested an algorithm requiring a chest CT plus either PFTs or a lung biopsy (algorithm 3H) and an algorithm requiring either a pulmonologist diagnosis or chest CT plus either PFTs or a lung biopsy (algorithm 3I). Sensitivity improved in these algorithms to 70.1% and 76.4%. Algorithm 3H had a higher specificity and PPV, but sensitivity and agreement by Kappa were better for algorithm 3I (Kappa 0.75 vs. 0.73; 0.72 vs. 0.70 sensitive ILD definition). Algorithm performance for identifying ILD was similar by Kappa between algorithms 3A, 3C, 3D, 3G, 3H, and 3I (0.68-0.75, 0.63-0.72 relaxed ILD definition) indicating substantial agreement.

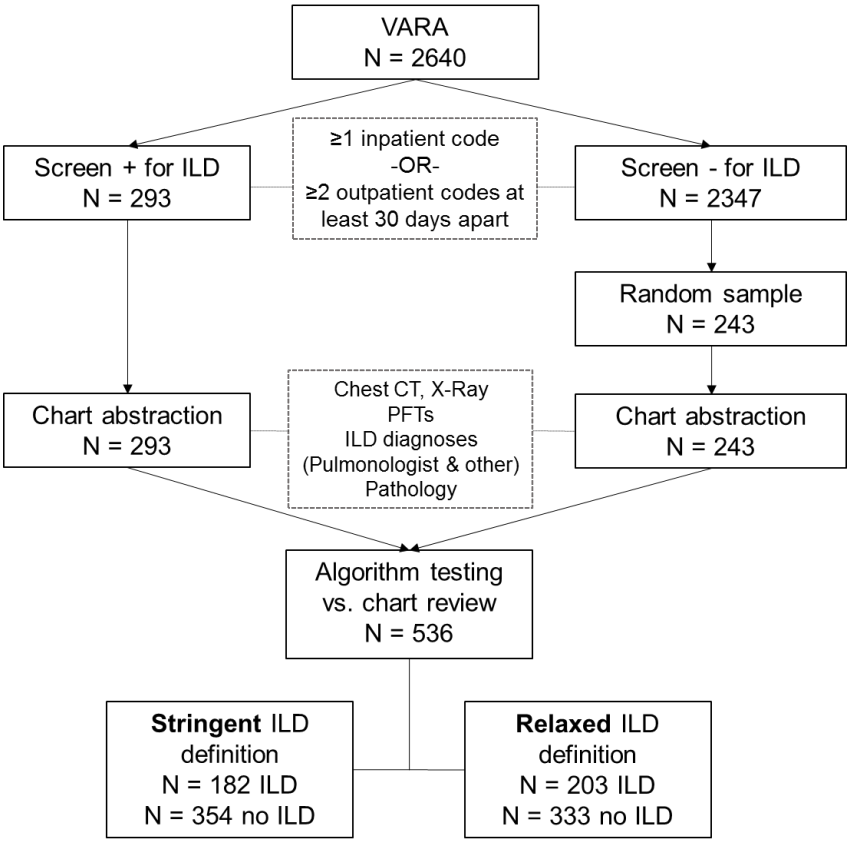
Stage IV: Exclusion of other ILD

Using the top-performing models (3A, 3C, 3D, 3G, 3H, 3I), we then excluded those with a diagnosis code for other causes of ILD that occurred on or after the date of the last ILD code recorded, following approaches in IPF.[129,130,134] After excluding other causes of ILD, there was modest improvement in specificity and PPV (range 0.7-1.2% improvement, **Table 15**) for each algorithm. Specificity for these algorithms ranged from 97.9-98.8%) and PPVs ranged from 76.0-82.9% (80.3-86.6% relaxed ILD definition). Overall performance by Kappa was similar after excluding other causes of ILD (Kappa 0.67-0.74, 0.61-0.71 relaxed ILD definition). Algorithm 4I had the best agreement with medical record review (Kappa 0.74, 0.70 relaxed ILD definition), indicating substantial agreement. Performance metrics for this algorithm were: sensitivity 73.2% (65.4% relaxed ILD definition), specificity 98.2% (98.5% relaxed ILD definition), and PPV 78.5% (82.4% relaxed ILD definition).

Sensitivity analyses

Because ILD may be detected on different types of CT scans (e.g., high-resolution, CT-angiogram, or low dose CT for lung cancer screening), we tested both broad CT codes and specific CT codes. Algorithms performed similarly regardless of the CT scan codes utilized (**Table 16**). Similarly, we tested algorithms with specific (open via thoracotomy and bronchoscopy) and broad (open, bronchoscopy, and percutaneous) lung biopsy codes. These algorithms also performed similarly, with excellent specificity but limited sensitivity. We tested algorithms that only required diagnostic testing (CT, PFT, and lung biopsy) to be completed at least 7 days prior to ILD diagnosis, rather than within a 7-180 day window. These algorithms had modestly improved sensitivity and Kappa values. We tested a broader time window for excluding other causes of ILD, excluding cases if a diagnostic code for other causes of ILD was ever recorded in national VA data. These algorithms reduced sensitivity and Kappa values. Because some non-ILD cases had clinical diagnoses or diagnostic testing for ILD but did not fulfill primary ILD definitions, we compared algorithm 4I against two additional ILD definitions with fewer requirements. Specificity was ≥98.6% and PPV improved to 83.4% and 86.3% in these models, with Kappa values still suggesting substantial agreement (Kappa 0.67).

Figure 7. Derivation of study sample and classification of ILD by medical record review



The Veterans Affairs Rheumatoid Arthritis (VARA) registry was screened for ≥2 outpatient or ≥1 inpatient discharge diagnoses of interstitial lung disease (ILD). Detailed medical record review was performed for all subjects who screened positive and a random sample of those who screened negative to validate ILD diagnoses by abstracting physician diagnoses, imaging findings, pulmonary function tests (PFTs), and pathology findings. Using the primary stringent ILD definition, 182 cases were identified, while using a relaxed definition resulted in 203 ILD cases.

Abbreviations: CT, computed tomography; ILD, interstitial lung disease; PFTs, pulmonary function tests; VARA, Veterans Affairs Rheumatoid Arthritis registry

Table 11. Characteristics of study cohort at registry enrollment by ILD status

	Stringent ILD definition		Relaxed ILD definition	
	ILD (n=182)	No ILD (n=354)	ILD (n=203)	No ILD (n=333)
Patient Characteristics				
Age, years	66.2 (9.8)	63.6 (10.4)†	66.4 (9.8)	63.4 (10.4)†
Male sex	173 (95.6)	323 (91.2)	193 (95.5)	303 (91.0)
Caucasian	138 (75.8)	273 (77.3)	154 (76.2)	257 (77.2)
Smoking status				
Current	53 (29.6)	89 (25.9)	59 (29.8)	83 (25.5)
Former	100 (55.9)	187 (54.4)	109 (55.1)	178 (54.8)
Never	26 (14.5)	68 (19.8)	30 (15.2)	64 (19.7)
High-school education	147 (86.5)	273 (86.7)	161 (85.6)	259 (87.2)
RA duration, years	11.7 (12.6)	11.1 (11.1)	12.0 (12.7)	10.9 (10.9)
Anti-CCP antibody +	137 (83.0)	246 (80.4)	152 (83.5)	231 (79.9)
Rheumatoid factor +	144 (87.8)	243 (79.2)†	158 (87.3)	229 (79.0)†
DAS28	4.2 (1.4)	4.0 (1.6)	4.2 (1.4)	4.0 (1.6)
MDHAQ	1.0 (0.6)	1.0 (0.6)	1.0 (0.6)	1.0 (0.6)
Methotrexate	51 (30.7)	163 (52.2)†	59 (32.1)	155 (52.7)†
bDMARDs	52 (28.6)	79 (22.3)	56 (27.6)	75 (22.5)
Prednisone	103 (62.1)	123 (39.4)†	113 (61.4)	113 (38.4)†
Interstitial Lung Disease Status (by medical record review)				
Screened positive for ILD	177 (97.3)	116 (32.8)†	196 (96.6)	97 (29.1)†
Pulmonologist diagnosis	172 (94.5)	5 (1.4)†	172 (84.7)	5 (1.5)†
Non-pulmonologist diagnosis	175 (96.2)	30 (8.5)†	196 (96.6)	9 (2.7)†
Imaging consistent with ILD	182 (100.0)	48 (13.6)†	202 (99.5)	28 (8.4)†
CT evidence of ILD	179 (98.4)	40 (11.3)†	195 (96.1)	24 (7.2)†
Restrictive pattern on PFTs	98 (53.9)	35 (9.9)†	99 (48.8)	34 (10.2)†
Pathology suggesting ILD	22 (12.1)	3 (0.9)†	23 (11.3)	2 (0.6)†
Prevalent at enrollment	91 (50.0)	-	101 (49.8)	-

Values mean (SD) or n (%) of non-missing, † $P < 0.05$ by independent t-test or chi-square test
Abbreviations: anti-CCP, anti-cyclic citrullinated peptide antibody; bDMARDs, biologic disease-modifying anti-rheumatic drugs; CT, computed tomography; DAS28, 28-joint disease activity score; MDHAQ, multidimensional health assessment questionnaire; ILD, interstitial lung disease; PFTs, pulmonary function tests; RA, rheumatoid arthritis; SD, standard deviation

Table 12. Performance of RA-ILD algorithms with various encounter type and diagnostic code frequency (Stage I)

Algorithm	Description[a]	Sensitivity	Specificity	PPV	NPV	%Agreement	Kappa
Stringent ILD definition							
1A	≥1 outpatient diagnosis	85.8 (68.8, 94.3)	91.2 (89.9, 92.4)	48.2 (40.1, 56.3)	98.5 (96.1, 99.5)	90.8	0.57 (0.48, 0.65)
1B	≥1 discharge diagnosis	48.4 (38.9, 58.0)	96.9 (96.3, 97.3)	58.9 (52.4, 65.2)	95.3 (93.2, 96.7)	92.7	0.49 (0.41, 0.57)
1C	≥1 outpatient or discharge diagnosis	87.3 (68.9, 95.6)	89.8 (88.6, 91.0)	44.7 (37.7, 52.0)	98.7 (96.0, 99.6)	89.6	0.54 (0.46, 0.52)
1D	≥2 outpatient diagnosis, >30 days apart	80.7 (65.7, 90.1)	96.5 (95.6, 97.3)	68.2 (59.6, 75.7)	98.2 (96.1, 99.2)	95.2	0.71 (0.62, 0.79)
1E	≥1 discharge diagnosis or ≥2 outpatient diagnosis, >30 days apart	85.2 (67.8, 94.0)	94.5 (93.4, 95.4)	58.7 (51.0, 66.0)	98.6 (96.2, 99.5)	93.7	0.66 (0.58, 0.74)
1F	≥2 diagnoses[b], >30 days apart	81.7 (66.4, 91.0)	96.0 (95.0, 96.8)	65.5 (56.9, 73.3)	98.3 (96.1, 99.2)	94.8	0.70 (0.61, 0.78)
1G	≥2 diagnoses[b], >30 days and ≤365 days apart	66.3 (55.0, 76.1)	97.0 (96.2, 97.7)	67.0 (58.0, 74.9)	96.9 (95.2, 98.1)	94.5	0.64 (0.55, 0.72)
1H	≥2 diagnoses[b], >30 days and ≤730 days apart	72.1 (59.6, 81.9)	96.8 (95.9, 97.5)	67.2 (58.2, 75.1)	97.4 (95.6, 98.5)	94.7	0.67 (0.58, 0.75)
Relaxed ILD definition							
1A	≥1 outpatient diagnosis	87.3 (72.5, 94.7)	92.8 (91.0, 94.2)	58.1 (47.9, 67.6)	98.4 (96.1, 99.4)	92.2	0.65 (0.56, 0.74)
1B	≥1 discharge diagnosis	44.8 (36.1, 53.7)	97.2 (96.7, 97.6)	63.8 (57.4, 69.7)	94.1 (91.6, 95.8)	91.9	0.48 (0.41, 0.56)
1C	≥1 outpatient or discharge diagnosis	89.3 (72.7, 96.3)	91.4 (89.7, 92.8)	53.8 (45.0, 62.4)	98.7 (96.0, 99.6)	91.1	0.62 (0.53, 0.71)
1D	≥2 outpatient diagnosis, >30 days apart	74.6 (60.5, 84.9)	97.1 (96.1, 97.8)	73.9 (65.3, 81.0)	97.2 (94.7, 98.5)	94.9	0.71 (0.62, 0.79)

1E	≥1 discharge diagnosis or ≥2 outpatient diagnosis, >30 days apart	80.0 (63.8, 90.0)	95.2 (94.1, 96.1)	64.6 (56.9, 71.6)	97.7 (95.0, 99.0)	93.7	0.68 (0.59, 0.76)
1F	≥2 diagnoses[b], >30 days apart	76.3 (61.8, 86.5)	96.7 (95.6, 97.5)	71.6 (63.0, 78.9)	97.4 (94.8, 98.7)	94.6	0.71 (0.61, 0.79)
1G	≥2 diagnoses[b], >30 days and ≤365 days apart	62.7 (51.4, 72.7)	97.6 (96.8, 98.3)	74.3 (65.4, 81.6)	96.0 (93.8, 97.5)	94.2	0.65 (0.56, 0.73)
1H	≥2 diagnoses[b], >30 days and ≤730 days apart	68.0 (55.6, 78.2)	97.4 (96.5, 98.1)	74.4 (65.4, 81.7)	96.5 (94.2, 97.9)	94.5	0.68 (0.59, 0.76)

[a] ICD-9: 515.x, 516.3, 516.8, 516.9, 714.81; ICD-10: M05.1x, J84.1, J84.2, J84.89, J84.9, J99

[b] outpatient or discharge diagnoses

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; ILD, interstitial lung disease; CT, computed tomography; PFT, pulmonary function test

Table 13. Performance of RA-ILD algorithms with various ILD diagnostic codes (Stage II)

Algorithm	Description	Sensitivity	Specificity	PPV	NPV	%Agreement	Kappa
Stringent ILD definition							
2A	All ILD codes*	81.7 (66.4, 91.0)	96.0 (95.0, 96.8)	65.5 (56.9, 73.3)	98.3 (96.1, 99.2)	94.8	0.70 (0.61, 0.78)
2B	Exclude M05.1x (*n=6 excluded*)	81.0 (65.7, 90.5)	96.3 (95.4, 97.1)	67.0 (58.9, 74.2)	98.2 (96.1, 99.2)	95.1	0.71 (0.62, 0.78)
2C	Exclude J84.2 (*n=0 excluded*)	81.7 (66.4, 91.0)	96.0 (95.0, 96.8)	65.5 (56.9, 73.3)	98.3 (96.1, 99.2)	94.8	0.70 (0.61, 0.78)
2D	Exclude J84.89 & J84.9 (*n=3 excluded*)	76.9 (61.2, 87.5)	96.2 (95.3, 96.8)	64.9 (57.4, 71.7)	97.8 (95.5, 99.0)	94.5	0.67 (0.59, 0.75)
2E	Exclude J99 (*n=0 excluded*)	81.7 (66.4, 91.0)	96.0 (95.0, 96.8)	65.5 (56.9, 73.3)	98.3 (96.1, 99.2)	94.8	0.70 (0.61, 0.78)
2F	Exclude 714.81 (*n=8 excluded*)	81.3 (66.1, 90.6)	96.3 (95.3, 97.1)	67.1 (58.3, 74.9)	98.2 (96.1, 99.2)	95.0	0.71 (0.62, 0.78)
2G	Exclude 516.8, 516.9 (*n=31 excluded*)	74.6 (61.6, 84.3)	96.6 (95.6, 97.4)	67.3 (57.8, 75.5)	97.6 (95.7, 98.7)	94.8	0.68 (0.59, 0.76)
2H	ICD-9 515.x, 516.3, 516.8, 516.9; ICD-10 J84.1, J84.89, J84.9 (*n=17 excluded*)	80.6 (65.5, 90.1)	96.8 (95.8, 97.5)	69.5 (61.1, 76.7)	98.2 (96.1, 99.2)	95.4	0.72 (0.63, 0.80)
Relaxed ILD definition							
2A	All ILD codes*	76.3 (61.8, 86.5)	96.7 (95.6, 97.5)	71.6 (63.0, 78.9)	97.4 (94.8, 98.7)	94.6	0.71 (0.61, 0.79)
2B	Exclude M05.1x (*n=6 excluded*)	75.2 (60.9, 85.5)	97.0 (96.0, 97.7)	72.9 (64.9, 79.7)	97.3 (94.8, 98.6)	94.8	0.71 (0.62, 0.79)
2C	Exclude J84.2 (*n=0 excluded*)	76.3 (61.8, 86.5)	96.7 (95.6, 97.5)	71.6 (63.0, 78.9)	97.4 (94.8, 98.7)	94.6	0.71 (0.61, 0.79)
2D	Exclude J84.89 & J84.9 (*n=3 excluded*)	72.1 (57.7, 83.1)	96.8 (96.0, 97.5)	71.3 (63.9, 77.7)	96.9 (94.2, 98.4)	94.4	0.69 (0.59, 0.77)

2E	Exclude J99 (*n=0 excluded*)	76.3 (61.8, 86.5)	96.7 (95.6, 97.5)	71.6 (63.0, 78.9)	97.4 (94.8, 98.7)	94.6	0.71 (0.61, 0.79)
2F	Exclude 714.81 (*n=8 excluded*)	75.5 (61.2, 85.8)	96.9 (95.9, 97.7)	73.0 (64.1, 80.3)	97.3 (94.8, 98.6)	94.8	0.71 (0.62, 0.79)
2G	Exclude 516.8, 516.9 (*n=31 excluded*)	68.3 (56.0, 78.5)	97.1 (96.0, 97.8)	72.1 (62.5, 80.0)	96.5 (94.2, 97.9)	94.2	0.67 (0.57, 0.75)
2H	ICD-9 515.x, 516.3, 516.8, 516.9; ICD-10 J84.1, J84.89, J84.9 (*n=17 excluded*)	74.4 (60.3, 84.8)	97.3 (96.4, 98.0)	75.3 (67.0, 82.1)	97.2 (94.7, 98.5)	95.1	0.72 (0.63, 0.80)

*ICD-9: 515.x, 516.3, 516.8, 516.9, 714.81; ICD-10: M05.1x, J84.1, J84.2, J84.89, J84.9, J99

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; ILD, interstitial lung disease; CT, computed tomography; PFT, pulmonary function test

Table 14. Performance of RA-ILD algorithms with various provider specialties and diagnostic testing (Stage III)

Algorithm	Description*	Sensitivity	Specificity	PPV	NPV	%Agreement	Kappa
Stringent ILD definition							
3A	≥1 pulmonologist ILD diagnosis	63.7 (51.8, 74.1)	98.5 (97.7, 99.0)	79.9 (69.8, 87.2)	96.7 (94.7, 97.9)	95.5	0.68 (0.59, 0.77)
3B	≥1 rheumatologist ILD diagnosis	55.7 (45.3, 65.5)	98.0 (97.3, 98.5)	72.1 (62.9, 79.7)	96.0 (94.1, 97.3)	94.4	0.60 (0.51, 0.68)
3C	CT 7-180 days prior to ILD diagnosis	72.9 (58.8, 83.5)	97.7 (96.8, 98.4)	75.2 (65.3, 83.0)	97.5 (95.3, 98.6)	95.6	0.72 (0.62, 0.80)
3D	PFT 7-180 days prior to ILD diagnosis	75.0 (62.1, 84.6)	98.1 (97.5, 98.5)	78.1 (72.2, 83.1)	97.7 (95.8, 98.7)	96.1	0.74 (0.66, 0.81)
3E	Lung biopsy 7-180 days prior to ILD diagnosis	9.3 (7.3, 11.8)	99.9 (99.6, 99.9)	87.5 (67.6, 95.9)	92.2 (90.3, 93.7)	92.1	0.15 (0.12, 0.20)
3F	≥1 pulmonologist diagnosis and CT or lung biopsy 7-180 days prior to ILD diagnosis	58.2 (46.7, 68.9)	98.4 (98.0, 98.7)	77.5 (72.0, 82.2)	96.2 (94.0, 97.6)	95.0	0.64 (0.55, 0.72)
3G	≥1 pulmonologist diagnosis and CT or lung biopsy or PFTs 7-180 days prior to ILD diagnosis	63.0 (51.3, 73.4)	98.6 (97.9, 99.1)	81.0 (71.5, 87.9)	96.6 (94.6, 97.9)	95.6	0.69 (0.59, 0.77)
3H	CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis	70.1 (57.3, 80.4)	98.5 (98.0, 98.9)	81.9 (76.0, 86.6)	97.2 (95.1, 98.4)	96.0	0.73 (0.64, 0.81)
3I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis	76.4 (62.7, 86.2)	98.0 (97.2, 98.5)	77.4 (69.1, 84.0)	97.9 (96.0, 98.9)	96.2	0.75 (0.66, 0.82)
Relaxed ILD definition							
3A	≥1 pulmonologist ILD diagnosis	55.5 (45.1, 65.4)	98.6 (97.8, 99.1)	81.6 (71.5, 88.6)	95.2 (92.9, 96.8)	94.3	0.63 (0.53, 0.72)
3B	≥1 rheumatologist ILD diagnosis	52.5 (42.7, 62.1)	98.5 (97.8, 99.0)	79.7 (70.7, 86.4)	95.0 (92.7, 96.6)	94.0	0.60 (0.51, 0.69)

3C	CT 7-180 days prior to ILD diagnosis	65.8 (53.2, 76.4)	98.1 (97.1, 98.7)	79.3 (69.4, 86.7)	96.2 (93.8, 97.8)	94.8	0.69 (0.59, 0.78)
3D	PFT 7-180 days prior to ILD diagnosis	68.2 (56.1, 78.3)	98.5 (98.0, 98.9)	83.3 (77.7, 87.6)	96.6 (94.3, 98.0)	95.5	0.73 (0.64, 0.80)
3E	Lung biopsy ≥7 days prior to ILD diagnosis	8.3 (6.7, 10.4)	99.9 (99.7, 99.9)	91.7 (72.1, 97.9)	90.8 (88.6, 92.6)	90.8	0.14 (0.11, 0.18)
3F	≥1 pulmonologist diagnosis and CT or lung biopsy 7-180 days prior to ILD diagnosis	50.9 (41.1, 60.6)	98.5 (98.1, 98.8)	79.3 (74.0, 83.7)	94.8 (92.3, 96.5)	93.8	0.59 (0.50, 0.67)
3G	≥1 pulmonologist diagnosis and CT or lung biopsy or PFTs 7-180 days prior to ILD diagnosis	55.0 (44.7, 64.8)	98.7 (98.0, 99.2)	82.8 (73.2, 89.4)	95.2 (92.9, 96.8)	94.4	0.63 (0.54, 0.72)
3H	CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis	62.8 (51.4, 72.9)	98.8 (98.3, 99.1)	85.4 (79.9, 89.7)	95.9 (93.5, 97.4)	95.1	0.70 (0.61, 0.78)
3I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis	68.2 (55.8, 78.4)	98.3 (97.5, 98.8)	81.0 (72.8, 87.2)	96.6 (94.3, 98.0)	95.3	0.72 (0.62, 0.79)

*ICD-9: 515.x, 516.3, 516.8, 516.9; ICD-10: J84.1, J84.89, J84.9

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; ILD, interstitial lung disease; CT, computed tomography; PFT, pulmonary function test

Table 15. Performance of RA-ILD algorithms with exclusion of other ILD (Stage IV)

Algorithm	Description	Sensitivity	Specificity	PPV	NPV	%Agreement	Kappa
Stringent ILD definition							
4A	≥1 pulmonologist ILD diagnosis and exclusion other ILD[a]	60.4 (49.2, 70.7)	98.7 (97.9, 99.2)	81.1 (70.8, 88.4)	96.4 (94.4, 97.7)	95.4	0.67 (0.57, 0.75)
4C	CT 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	69.7 (56.5, 80.4)	97.9 (97.0, 98.6)	76.0 (65.9, 83.8)	97.2 (95.1, 98.4)	95.5	0.70 (0.60, 0.79)
4D	PFT 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	71.7 (59.6, 81.4)	98.2 (97.7, 98.6)	78.8 (72.7, 83.8)	97.4 (95.6, 98.5)	96.0	0.73 (0.65, 0.80)
4G	≥1 pulmonologist diagnosis and CT or lung biopsy or PFTs 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	59.8 (48.7, 70.0)	98.8 (98.1, 99.2)	81.9 (72.2, 88.7)	96.4 (94.4, 97.7)	95.5	0.67 (0.57, 0.75)
4H	CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	66.9 (54.9, 77.1)	98.7 (98.2, 99.0)	82.9 (76.9, 87.6)	96.9 (94.8, 98.1)	95.9	0.72 (0.63, 0.79)
4I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	73.2 (60.3, 83.0)	98.2 (97.4, 98.7)	78.5 (70.1, 85.1)	97.6 (95.7, 98.6)	96.1	0.74 (0.65, 0.81)
Relaxed ILD definition							
4A	≥1 pulmonologist ILD diagnosis and exclusion other ILD[a]	52.8 (42.9, 62.4)	98.8 (98.0, 99.3)	82.9 (72.6, 90.0)	95.0 (92.7, 96.6)	94.2	0.62 (0.52, 0.70)
4C	CT 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	63.0 (51.1, 73.6)	98.3 (97.3, 98.9)	80.3 (70.1, 87.6)	96.0 (93.5, 97.5)	94.7	0.68 (0.58, 0.77)
4D	PFT 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	65.4 (53.9, 75.4)	98.7 (98.2, 99.0)	84.2 (78.6, 88.6)	96.3 (94.1, 97.7)	95.4	0.71 (0.62, 0.79)

4G	≥1 pulmonologist diagnosis and CT or lung biopsy or PFTs 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	52.2 (42.5, 61.7)	98.9 (98.2, 99.3)	83.7 (74.1, 90.3)	95.0 (92.7, 96.6)	94.3	0.61 (0.52, 0.70)
4H	CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	60.0 (49.2, 70.0)	98.9 (98.5, 99.3)	86.6 (81.0, 90.8)	95.6 (93.2, 97.2)	95.0	0.68 (0.59, 0.76)
4I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD[a]	65.4 (53.6, 75.5)	98.5 (97.7, 99.0)	82.4 (74.0, 88.5)	96.3 (94.1, 97.7)	95.2	0.70 (0.61, 0.78)

[a] exclusion of other ILD using diagnostic codes for pneumoconioses, radiation, hypersensitivity pneumonitis, other connective tissue diseases on or after the last ILD diagnosis code date (see Appendix F)

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; ILD, interstitial lung disease; CT, computed tomography; PFT, pulmonary function test

Table 16. Performance of modified RA-ILD algorithms (sensitivity analyses)

Algorithm	Description/modification[a]	Sensitivity	Specificity	PPV	NPV	%Agreement	Kappa
Stringent ILD definition							
3C	CT ≥7 days prior to ILD diagnosis	78.1 (63.8, 87.9)	97.2 (96.4, 97.9)	72.0 (63.6, 79.0)	98.0 (96.0, 99.0)	95.6	0.73 (0.64, 0.80)
3C	CT ≥7 days prior to ILD diagnosis with broad codes[b]	78.2 (63.9, 87.9)	97.2 (96.3, 97.8)	71.6 (63.1, 78.7)	98.0 (96.1, 99.0)	95.6	0.72 (0.63, 0.80)
3D	PFT ≥7 days prior to ILD diagnosis	80.0 (66.2, 89.1)	97.5 (96.9, 97.9)	75.4 (69.8, 80.3)	98.0 (96.0, 99.0)	95.9	0.75 (0.68, 0.82)
3E	Lung biopsy ≥7 days prior to ILD diagnosis	11.5 (9.2, 14.4)	99.9 (99.6, 99.9)	89.7 (72.4, 96.6)	92.4 (90.5, 93.9)	92.3	0.19 (0.15, 0.24)
3E	Lung biopsy ≥7 days prior to ILD diagnosis with broad codes	12.4 (9.8, 15.7)	99.8 (99.5, 99.9)	84.8 (68.4, 93.5)	92.4 (90.5, 94.0)	92.3	0.20 (0.16, 0.25)
3F	≥1 pulmonologist diagnosis and CT or lung biopsy ≥7 days prior to ILD diagnosis	62.8 (51.1, 73.1)	98.5 (97.7, 99.0)	79.7 (69.5, 87.1)	96.6 (94.6, 97.9)	95.5	0.68 (0.58, 0.76)
3G	≥1 pulmonologist diagnosis and CT or lung biopsy or PFTs ≥7 days prior to ILD diagnosis	63.2 (51.4, 73.6)	98.5 (97.7, 99.0)	79.8 (69.7, 87.1)	96.6 (94.6, 97.9)	95.5	0.68 (0.58, 0.77)
4I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD ever	68.5 (56.8, 78.3)	98.4 (97.6, 98.9)	79.1 (70.4, 85.8)	97.2 (95.4, 98.3)	95.9	0.71 (0.62, 0.79)
Relaxed ILD definition							
3C	CT ≥7 days prior to ILD diagnosis	70.8 (57.7, 81.2)	97.6 (96.8, 98.3)	76.6 (68.3, 83.2)	96.9 (94.5, 98.2)	95.0	0.71 (0.61, 0.79)
3C	CT ≥7 days prior to ILD diagnosis with broad codes[b]	70.9 (57.7, 81.2)	97.6 (96.7, 98.2)	76.1 (67.8, 82.9)	96.9 (94.5, 98.2)	95.0	0.71 (0.61, 0.79)
3D	PFT ≥7 days prior to ILD diagnosis	73.9	98.0	81.0	97.0	95.5	0.75

		(60.8, 83.7)	(97.5, 98.5)	(75.8, 85.4)	(94.6, 98.4)		(0.66, 0.82)
3E	Lung biopsy ≥7 days prior to ILD diagnosis	10.2 (8.2, 12.7)	99.9 (99.7, 99.9)	93.1 (76.2, 98.3)	90.9 (88.8, 92.7)	91.0	0.17 (0.13, 0.21)
3E	Lung biopsy ≥7 days prior to ILD diagnosis with broad codes	11.0 (8.8, 13.7)	99.8 (99.6, 99.9)	87.9 (71.8, 95.4)	91.0 (88.8, 92.8)	91.0	0.18 (0.14, 0.22)
3F	≥1 pulmonologist diagnosis and CT or lung biopsy ≥7 days prior to ILD diagnosis	54.8 (44.5, 64.6)	98.6 (97.8, 99.1)	81.4 (71.2, 88.5)	95.2 (92.9, 96.8)	94.2	0.62 (0.53, 0.71)
3G	≥1 pulmonologist diagnosis and CT or lung biopsy or PFTs ≥7 days prior to ILD diagnosis	55.1 (44.8, 65.0)	98.6 (97.8, 99.1)	81.5 (71.3, 88.6)	95.2 (92.9, 96.8)	94.3	0.63 (0.53, 0.72)
4I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD ever[c]	61.4 (50.5, 71.3)	98.7 (97.9, 99.1)	83.3 (74.6, 89.5)	95.9 (93.7, 97.4)	95.0	0.68 (0.59, 0.76)
Two of clinical diagnosis of ILD, CT, restrictive PFTs, or biopsy (n=210 ILD)							
4I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD[c]	60.5 (49.1, 70.8)	98.6 (97.8, 99.1)	83.4 (75.0, 89.4)	95.4 (92.8, 97.1)	94.5	0.67 (0.57, 0.76)
Clinical diagnosis of ILD or two of CT, restrictive PFTs, or biopsy (n=220 ILD)							
4I	Pulmonologist diagnosis or CT and PFTs or lung biopsy 7-180 days prior to ILD diagnosis and exclusion other ILD[c]	58.7 (47.8, 68.8)	98.8 (98.0, 99.3)	86.3 (0.78, 0.92)	94.9 (92.2, 96.7)	94.2	0.67 (0.57, 0.75)

[a] ICD-9: 515.x, 516.3, 516.8, 516.9; ICD-10: J84.1, J84.89, J84.9; [b] broad CT codes (includes CT-A and low dose CT); [c] Excluding when diagnostic codes for pneumoconioses, radiation, hypersensitivity pneumonitis, or other connective tissue diseases were present (Appendix F)
Abbreviations: PPV, positive predictive value; NPV, negative predictive value; ILD, interstitial lung disease; CT, computed tomography; PFT, pulmonary function test

4.4 Discussion

To facilitate the use of administrative data for RA-ILD research, we have characterized the performance of administrative algorithms for identifying ILD among RA patients compared to detailed medical record review. Algorithms including specific ICD-9 and ICD-10 ILD codes attributed to multiple encounters, pulmonologist diagnosis or diagnostic testing, and exclusion of other causes of ILD were able to accurately classify ILD. The best performing algorithm (algorithm 4I, **Table 15**) requiring ≥2 ILD diagnosis codes at least 30 days apart, a single pulmonologist diagnosis for ILD or CT and either PFTs or lung biopsy 7-180 days prior to ILD diagnosis, and exclusion of other ILD causes after the last ILD diagnosis yielded substantial agreement to medical record review by Kappa (0.74). PPVs for this algorithm ranged from 78.5-86.3% depending on the requirements of the ILD reference-standard definition. Because there is a trade-off between sensitivity and specificity with these ILD algorithms and differences in availability of algorithm components within different datasets, the choice of algorithm will depend on the purpose of the study and available data. Our results provide detailed data on the performance of several ILD algorithms that will support investigator selection of ILD case finding approaches in future studies.

Similar to administrative algorithms developed to identify RA[80] and other rheumatic conditions,[81-83] the requirement of multiple diagnosis codes for ILD separated over time enhanced the specificity and PPV of administrative ILD algorithms. Importantly, these results demonstrate that some diagnosis codes incorporated into prior RA-ILD algorithms lack specificity for ILD.[63,77,78,84,125] Most notably among these were ICD-9 (714.81) and ICD-10 (M05.1x) codes for “rheumatoid lung”. Because there are numerous pulmonary manifestations of RA including ILD, obstructive lung disease, nodules, and pleural effusions,[44] these codes may be used for these other entities in the

setting of RA. Indeed, pleural effusions and pulmonary nodules were reasons for these codes occurring in the absence of ILD. We recommend the following ICD-9 and ICD-10 codes for identifying ILD in RA patients: ICD-9 515.x, 516.3, 516.8 and ICD-10 J84.1, J84.89, J84.9 (bolded in **Appendix D**).

Our results illustrate that requiring a pulmonologist diagnosis of ILD achieves excellent specificity (algorithm 3A, 98.5%), but at the expense of sensitivity (63.7%). This reduction in sensitivity may be exaggerated in our cohort because pulmonologist diagnoses outside the VA health care system would not be captured by our algorithms. Therefore, algorithms with pulmonologist diagnosis may actually perform better in other settings. At least in our sample, requiring further diagnostic testing such as chest CT, PFTs, and lung biopsy did not significantly improve the PPV or Kappa from algorithms that already included a pulmonologist diagnosis of ILD. Eliminating the requirement of a pulmonologist diagnosis, we found that requiring a recent CT plus PFTs or lung biopsy in the prior 6 months achieved a similar PPV (81.9% vs. 79.9%). Using broad vs. specific CPT codes for these diagnostic tests rendered little impact on model performance. Combining either a pulmonologist diagnosis or the aforementioned diagnostic tests (algorithm 3I) optimized the sensitivity while preserving a reasonable specificity, leading to optimal algorithm performance by Kappa. Further refining this algorithm with exclusion of other ILD causes maintained overall algorithm performance while modestly increasing PPV (algorithm 4I).

Because we performed detailed medical record review on a random sample of VARA participants who did not screen positive for ILD, we were able to assess not only the specificity and PPV but also the sensitivity and NPV. Algorithms that incorporated pulmonologist diagnosis or diagnostic testing obtained the highest PPV (≥78.5%), but algorithms without these additional criteria had similar Kappa values (0.72), reflecting

improvements in specificity at the expense of sensitivity. As the overall performance (measured by Kappa) did not significantly differ for several algorithms, choice should be directed by specific study needs for either greater sensitivity or specificity and data availability. For example, completion of RA-ILD comparative effectiveness and outcomes research in large administrative datasets will require specific ILD algorithms, such as algorithm 4I (PPV 78.5-86.3%). Epidemiologic studies of RA-ILD, rather, may implement both a specific (algorithm 4I) and more sensitive algorithm (algorithm 2H, NPV 98.2% and 97.9%), recognizing the "truth" lies between the estimates from the specific and sensitive algorithms.

The generalizability of our findings may be limited by male predominance of the VARA registry, unique exposures of the Veteran population, as well as the coding practices represented by the 13 VARA-associated VA medical centers at which this work was conducted. However, the Veterans Health Administration represents the largest integrated health care system in the US with reduced barriers to access among its beneficiaries and a single electronic health record. Patients may receive care outside the VA, which affects capture by administrative algorithms and medical record review. To mitigate this, we reviewed the clinical notes for mention of outside care and selected claims originating from non-VA care when constructing our administrative ILD algorithms. Supporting the validity of our findings is that limited testing in a prior study of Kaiser Permanente Northern California found a PPV of 63% for ≥2 ILD diagnosis codes using imaging reports for ILD validation.[126] This is in agreement with the PPV of 65.5% for a similar algorithm (algorithm 1F) in our study. The derived ILD algorithms are also currently being externally validated in additional non-VA datasets.

Validation of ILD diagnoses through medical records was retrospective, with diagnostic testing dictated through regular clinical care and interpreted by the treating

providers. Additionally, there are currently no widely accepted classification criteria for RA-ILD. Because of the potential of misclassification in our reference standard, we used stringent and relaxed primary ILD definitions for each stage of algorithm development and testing. Furthermore, because some participants who did not fulfill the stringent or relaxed primary ILD definitions had clinical diagnoses or diagnostic testing suggestive of ILD, we performed sensitivity analyses comparing optimal algorithms to ILD definitions with fewer requirements. Overall algorithm performance was consistent between ILD definitions, with increased PPV using ILD definitions with fewer requirements. Because only approximately 10% of ILD cases were initially diagnosed after ICD-10 implementation, our findings may underestimate ICD-10 code contribution to ILD classification, a possibility that will need to be addressed in future research. Finally, we assessed the accuracy of ILD algorithms within a cohort fulfilling 1987 ACR criteria,[93] and the performance of these algorithms may vary if applied outside of this setting (e.g., in combination with administrative algorithms to identify RA). However, the results from our study will serve as a valuable benchmark for future efforts focused on external validation. Given the high specificity of administrative algorithms for RA,[80] we would anticipate to observe minimal reductions in the specificity of these ILD algorithms.

In conclusion, we have demonstrated that administrative algorithms can be used to accurately identify ILD in a RA cohort. Our results detail the performance metrics of these different algorithms for ILD, which can be applied to large administrative data sources to perform further clinical and epidemiologic study of RA-ILD.

CHAPTER 5: DISCUSSION

5.1 Summary

RA causes significant morbidity and mortality, with respiratory-related deaths being the most overrepresented cause of death in RA.[26,27] In this dissertation, investigation into the prognosis of chronic lung diseases in RA was assessed and compared to cardiovascular disease. Subsequently, a novel serum biomarker for the most fatal RA-associated lung disease, RA-ILD, was evaluated. Finally, accurate algorithms for classifying RA-ILD in large, real-world datasets were derived. With these results, we now have biological and biomedical informatics tools to more optimally investigate RA-associated lung diseases and enhance its identification.

The importance of considering, studying, and managing chronic lung disease in RA was illustrated by our findings in Chapter 2 that comorbid lung disease carried a prognosis as poor as cardiovascular disease in patients with RA. The risk of death was 1.5-fold higher in RA subjects with comorbid chronic lung disease and 1.6-fold higher in RA subjects with cardiovascular disease compared to RA subjects with neither comorbidity. This poor prognosis was not limited to RA-ILD but also present for COPD, bronchiectasis, and other lower respiratory diseases. We also confirmed that RA-ILD carried the greatest mortality risk among the RA-associated lung diseases. Challenging the common therapeutic dogma that methotrexate should be avoided in RA patients with chronic lung disease, we did not find a higher risk of death in RA patients with lung disease receiving methotrexate than RA patients without lung disease.

In Chapter 3, we tested the hypothesis that higher serum concentrations of anti-MAA antibody would be significantly associated with RA-ILD independent of established RA-ILD risk factors. Consistent with our hypothesis, IgA and IgM anti-MAA antibody

isotype concentrations in the upper three quartiles were associated with a more than 2-fold higher odds of ILD among our RA population. Moreover, we assembled a diseased control group, demonstrating that anti-MAA antibody (specifically the IgM isotype) is uniquely higher in ILD than RA patients with COPD (a chronic condition also more common in RA with overlapping manifestations). Beyond the potential role of anti-MAA antibody as a specific serum biomarker of RA-ILD based on these results, our investigation of lung tissues from RA-ILD subjects and controls (diseased and healthy) demonstrated the enhanced presence of MAA-modified proteins in the lungs from RA-ILD subjects. In RA-ILD, MAA modified proteins co-localized with other RA autoantigens (citrulline), relevant immune effector cells, and extracellular matrix proteins. These findings implicate MAA and immune responses to MAA in the pathogenesis of RA-ILD.

Finally, to enable much needed comparative effectiveness and outcomes research in RA-ILD we leveraged a large RA registry and linkage to national administrative data to develop algorithms for RA-ILD as part of efforts outlined in Chapter 4. Administrative based algorithms for RA-ILD with multiple ILD diagnostic codes plus either a pulmonologist diagnosis of ILD or testing for ILD (CT, PFTs, or lung biopsy) achieved >98% specificity and had substantial agreement with the reference standard of medical record review. With the understanding that varying the components of a proposed ILD algorithm would naturally result in a trade-off between sensitivity and specificity, we provided detailed performance characteristics for several different RA-ILD algorithms that could serve as a resource for different types of clinical and epidemiologic research in RA-ILD using a variety of data sources.

5.2 Future work

Serum biomarkers hold substantial promise for case-finding approaches, including their integration into ILD screening protocols that could be widely deployed in patients with RA. In this situation, sensitivity is the diagnostic characteristic of the biomarker (or screening protocol) that is most critical. A highly sensitive biomarker could provide for efficient case finding, identifying high-risk patients that would benefit from additional, more resource-intensive testing needed to confirm the diagnosis. In contrast, if a biomarker is highly specific for RA-ILD, confirmatory testing with a high-resolution CT and PFTs are likely to still be needed as these tests provide information that will ultimately guide treatment selection and predict prognosis. Findings on CT may suggest the histopathologic pattern of RA-ILD that has implications for treatment selection with the belief that anti-fibrotic therapies have more efficacy in usual interstitial pneumonia (as in IPF with its typical “honeycombing” appearance on CT) while immunomodulatory therapies have greater efficacy in non-specific interstitial pneumonia (with its typical CT findings of “ground glass” and absence of “honeycombing”). The degree of lung involvement on CT and physiologic impairment on PFTs provide prognostic information to providers and patients.[135-137] Furthermore, the results of these tests will serve as baseline values that will be followed longitudinally to determine if treatments are effective or if the underlying disease is progressing.[138,139]

In Chapter 3, we showed that anti-MAA antibody was independently associated with RA-ILD. In contrast, patients with RA and the lowest values of anti-MAA antibody were substantially less likely to have ILD, suggesting anti-MAA antibody could serve as a biomarker in an RA-ILD screening model to identify patients in need of further testing. Pairing anti-MAA antibody with other serum biomarkers may further improve the sensitivity and specificity for detecting RA-ILD and such work is underway. A proposed

model for the enhanced identification of RA-ILD through a two-stage process incorporating serum biomarkers is shown in **Figure 8**. In this proposed model, individuals with RA are screened for RA-ILD with a panel of serum biomarkers that includes anti-MAA antibody and that encompass the different pathophysiologic pathways implicated in RA-ILD (autoimmunity, inflammation, fibrosis, and oxidative stress). Those with a normal screening biomarker are then followed expectantly or with serial biomarker testing, while an abnormal screening biomarker panel prompts further testing with a high-resolution CT scan and PFTs. If RA-ILD is confirmed, then patients are referred for management of ILD, ideally in a multi-specialty treatment center. Patients with indeterminate test results (e.g. abnormal serum biomarkers with equivocal CT/PFT findings) or normal test results are considered high-risk, monitored serially, and targeted with preventive strategies (e.g. smoking cessation or avoidance of other airway irritants).

This screening and confirmation approach to the identification is in contrast to a universal testing approach where all patients with RA undergo high-resolution CT scanning and PFTs. This would maximize the sensitivity for detecting RA-associated lung diseases, provide prognostic information when lung disease is present, and could identify other lung pathologies like lung cancer, which is increased approximately 60% in RA patients.[140] However, universal CT and PFT testing would be an expensive approach, adding to the economic burden of RA on the health care system which already imposes a total annual cost of over $19 billion dollars.[141] This also imposes significant costs to patients, who have greatly benefited from the efficacy of bDMARDs but also economically impacted by their costs.[142] Other limitations of the universal testing approach include exposing RA patients to unnecessary radiation and the detection of incidental findings, a phenomenon that leads to additional medical testing, medical costs, unnecessary procedures, and anxiety for patients.[143]

Which biomarkers to include in the proposed RA-ILD screening model is an area for future study. Candidate biomarkers evaluated to date in RA-ILD and their performance are listed in **Table 17**. Many of these biomarkers have also been studied in other connective tissue disease related ILD or IPF.[144-147] Since the primary use of these biomarkers will be screening for ILD in a RA cohort, this lack of specificity for the systemic disease causing the ILD is of little consequence. Most of these RA-ILD biomarkers have been tested in isolation, with the exception of the combination of MMP-7, PARC, and SP-D studied by Doyle et al.[72] Future studies should test whether other combinations of these and other newly identified RA-ILD biomarkers have value in combination to further develop and refine this ILD screening model. Because of the potential for multicollinearity between serum biomarkers, analytic techniques such as principal component analysis should be considered in future efforts to identify biomarker signatures.[148] Although serum and sputum/bronchoalveolar biomarkers have been investigated in RA-ILD,[149] serum biomarkers are greatly preferred because of the added feasibility of collecting the samples.

Additional questions arise from this proposed RA-ILD screening model. When and how often should ILD screening biomarkers be tested? Immediately at the time of RA diagnosis or later in the disease course? RA-ILD is often thought to be a feature of established RA. Supporting this, Kelly et al. found the median duration of RA at ILD diagnosis was 9 years.[48] However, other studies have highlighted the early appearance of ILD in RA patients. Koduri et al. reported that over 50% of ILD cases occurred within 3 years of RA diagnosis.[47] In a small cohort of 37 subjects with RA for <2 years, Gabbay et al. detected evidence of ILD on CT in 58% of subjects.[65] Finally, ILD may even present before RA in approximately 10-15% of cases.[47,48] Other pulmonary abnormalities, including involvement of the airways, may also predate the appearance of RA.[51] Based

on these results, we propose that the initial ILD screening should occur early after RA diagnosis. This timing would be further justified if these candidate biomarkers are demonstrated to be predictive of incident ILD. Whether anti-MAA antibody, as well as other candidate RA-ILD biomarkers, can predict the incidence or progression of ILD or prognosticate survival are areas of planned investigation.

The RA-ILD screening protocol proposed is intended for the identification of ILD in previously diagnosed RA patients. When the ILD predates articular symptoms, there is a need for a diagnostic evaluation of unspecified ILD to establish a diagnosis of RA. In this situation, it is expected that established RA autoantibodies, RF and ACPAs, will be more helpful than the ILD biomarkers listed in **Table 17**. With a specificity of >95%,[119] the presence of ACPAs would be highly suggestive of RA as the underlying connective tissue disease. Although ACPAs and RF may be absent in up to 30-40% of RA patients, the sensitivity for RA is likely to be greater in the setting of this extraarticular manifestation because these antibodies are strongly associated with the presence of RA-ILD.[48,66,71,72] In contrast, the specificity of ACPAs and RF may be lower in this setting because other chronic lung diseases are associated with the presence of ACPAs in the absence of RA.[49] These hypotheses on the diagnostic performance of ACPAs and RF for RA in the setting of ILD warrant future investigation.

While the association of serum anti-MAA antibody with RA-ILD in Chapter 3 support its integration into future RA-ILD screening efforts, the enhanced staining for MAA modified proteins in the lungs from RA-ILD subjects and co-localization with other autoantigens, immune effector cells, and extracellular matrix proteins also implicates this post-translational modification in the pathogenesis of RA-ILD. Animal and *in vitro* studies building upon these findings could begin to elucidate the mechanisms by which MAA modification and the immune responses targeting MAA-modified proteins may facilitate

the loss of tolerance to citrullinated antigens and drive pro-inflammatory and/or pro-fibrotic responses within the lungs. For example, differentiation of fibroblasts into activated myofibroblasts as a result of epithelial injury and inflammatory responses is a central process in the development of pulmonary fibrosis.[150] MAA could serve as a pro-inflammatory mediator between oxidative stress related epithelial injury and activation of myofibroblasts in RA-ILD. In IPF, anti-oxidant therapies (N-acetylcysteine) have been tested in several randomized controlled trials with mixed results.[151-153] Further evaluation of the role of MAA in this pathway may elucidate novel targets for RA-ILD therapies. Moreover, to date, anti-MAA antibodies have been detected using MAA modified albumin, which is not a relevant antigen in the pathogenesis of RA and RA-ILD. Anti-MAA antibodies targeting relevant RA and RA-ILD antigens may have better specificity for disease detection as well as further our understanding of the role of MAA modification in RA and RA-ILD pathogenesis.

The development of administrative algorithms that can accurately classify RA-ILD in Chapter 4 provides for the first time a tool that can be readily implemented in large, real-world datasets to enhance the identification of RA-associated lung disease. Recognizing that we tested our algorithms in a cohort of RA subjects who had definitive RA meeting formal disease classification criteria, the application of these algorithms in datasets where RA case identification is conducted solely with administrative data could impact their performance. To better understand this potential threat to external validity, we are currently collaborating on an effort to test these algorithms for ILD in combination with administrative algorithms for RA in large non-VA data sources. Given the high specificity of administrative algorithms for RA[80] and the vast overrepresentation of ILD in RA,[57] we expect similar ILD algorithm performance in combination with administrative algorithms for RA.

The algorithms developed in Chapter 4 used data that is nearly universally available in administrative claims. Because the adoption of electronic medical records has become widespread,[154] leveraging the additional data available in electronic health records to improve these algorithms represents an exciting and rapidly emerging opportunity. Electronic medical records not only contain the orders for diagnostic testing and billing diagnoses, but have diagnostic testing results and clinical notes containing provider assessments, diagnoses, and management plans. Tools such as natural language processing could be used to build on these administrative algorithms by facilitating the development of electronic medical record computable phenotypes.[155] Leveraging state-of-the-art text mining approaches, characteristic imaging findings such as “reticular opacities”, “ground glass”, “honeycombing”, and “fibrosis” could be searched for, as could RA-ILD patterns including “usual interstitial pneumonia” and “non-specific interstitial pneumonia”. PFT results could be incorporated using a similar approach, allowing for identification of reductions in forced vital capacity or diffusion capacity that result from RA-ILD.[138] While the potential availability of such robust data sources is exciting, the methods by which these different data components are stored and accessed varies between health care systems and electronic medical record vendors. Thus, widespread adoption of these algorithms will likely pose an ongoing challenging until interoperability between health care systems and vendors is achieved.[156]

While better disease activity, functional status, and a reduced need for joint replacement surgeries have resulted from RA treatment advances,[36-38] it is uncertain if the incidence of RA-associated lung disease has declined. Cohorts constructed from these algorithms could address this question, as well as investigate determinants and trends of survival in RA-ILD. These prevailing research questions illustrate the potential near-term uses of these algorithms for epidemiologic and outcomes research. Other

uses for these algorithms include clinical trial planning and enrollment. Deploying these algorithms will assist with determining if an adequate number of subjects exist to support a clinical trial and then to facilitate clinical trial recruitment. A similar model was successfully implemented to recruit for a large pragmatic trial of aspirin for the prevention of cardiovascular events.[157]

The algorithms developed in this dissertation focused on RA-ILD, given that RA-ILD is the most fatal RA-associated lung disease. Deriving and validating other chronic lung disease algorithms in RA is an area for future work. Administrative algorithms for COPD have been developed in the general population through the use of multiple diagnostic codes from outpatient or inpatient encounters,[158] though they may not appropriately distinguish COPD from other obstructive lung diseases, namely asthma.[159] Algorithms using electronic health record data, which contains data elements not available in administrative health databases, have also been proposed for COPD. The addition of smoking status, COPD related medications, and other medical record data elements (e.g., the past medical history and problem list) may improve algorithm performance.[160] Since the evaluation and management of COPD in RA does not differ substantially from the general population, it is anticipated that these algorithms would perform similarly in RA cohorts. Algorithms for other RA-related pulmonary manifestations such as bronchiectasis, bronchiolitis, RA nodulosis (of the lungs), and RA pleural effusions have not been studied in RA or more broadly the general population. Results detailed in Chapter 4 illustrate that diagnostic codes pertaining to “rheumatoid lung” are often used to designate these conditions. Therefore, use of this non-specific ICD code will likely make it challenging to derive accurate algorithms for these conditions using only administrative data.

While this dissertation focused on the identification of RA-associated lung disease, the poor prognosis observed in these individuals may relate, in part, to the fact that optimal treatments for RA-associated lung disease are not known. To date, there have been no completed randomized controlled trials of therapies in RA-associated lung diseases or clinical guidelines for the management of RA-associated lung diseases. Complicating therapy selection further, most RA therapies have been linked with drug-induced pneumonitis, including methotrexate, leflunomide, sulfasalazine, TNFi, rituximab, abatacept, and tocilizumab.[74-76,161]

Of the RA therapies linked with pneumonitis, this phenomenon has been most widely described as a complication of methotrexate.[162,163] Providers often avoid methotrexate in RA-ILD patients,[78] opting for alternative RA therapies that have also been linked to pneumonitis or fatal ILD exacerbations.[74] In a meta-analysis of 22 randomized controlled trials with 8,584 RA patients, methotrexate pneumonitis was exceedingly rare, occurring in <0.3% of subjects.[163] Moreover, the authors did not find methotrexate to be associated with non-infectious respiratory events or respiratory mortality. Findings detailed in Chapter 2 evaluating mortality in RA patients with and without lung disease suggest that methotrexate should not be routinely avoided in RA patients with chronic lung disease. Using several models with various analytic approaches, methotrexate was either associated with a lower risk of mortality or not significantly associated with mortality in those with lung disease. These analyses did not study methotrexate in RA-ILD specifically since validated algorithms for classifying ILD were not available at that time, a limitation that can now be addressed through our work deriving RA-ILD algorithms (Chapter 4). Whether methotrexate leads to adverse outcomes specifically in RA-ILD is an important knowledge gap and an area of planned

study, given methotrexate treatment results in improvements in disease activity and survival.[164]

Advanced RA therapies, such as bDMARDs, have led to substantial improvements in RA outcomes, although data specific to patients with ILD is sparse. A single, small, open-label uncontrolled study of rituximab in RA-ILD showed stability of PFTs in completers, but two died over the course of a 48-week study.[165] The remainder of studies of RA therapies in RA-ILD have been observational and hampered by a number of methodological limitations. Dixon et al. found numerically more ILD-related deaths among RA-ILD subjects treated with TNFi compared to conventional DMARDs.[115] In a later study from the same registry, rituximab treatment was associated with a lower risk of mortality than TNFi in RA-ILD.[166] These studies relied on a single questionnaire response by the treating provider to establish the diagnosis of RA-ILD, an approach prone to misclassification. Curtis et al. evaluated the risk of hospitalization for respiratory events in RA-ILD using a large administrative data source, finding no significant differences different between bDMARDs.[77] The algorithms employed likely suffered from poor specificity leading to misclassificatoin. As the selection of advanced RA therapies in RA-ILD remains challenging, we have planned a study leveraging our RA-ILD algorithms and large, real-world data sources to compare the effectiveness and safety of advanced RA therapies in RA-ILD. Our proposed study uses a new-user, active-comparator design to reduce confounding and selection bias.[167] To further reduce confounding bias, we will use inverse probability of treatment weighting to balance pre-specified covariates related to the prognosis in RA-ILD between treatment groups.[168]

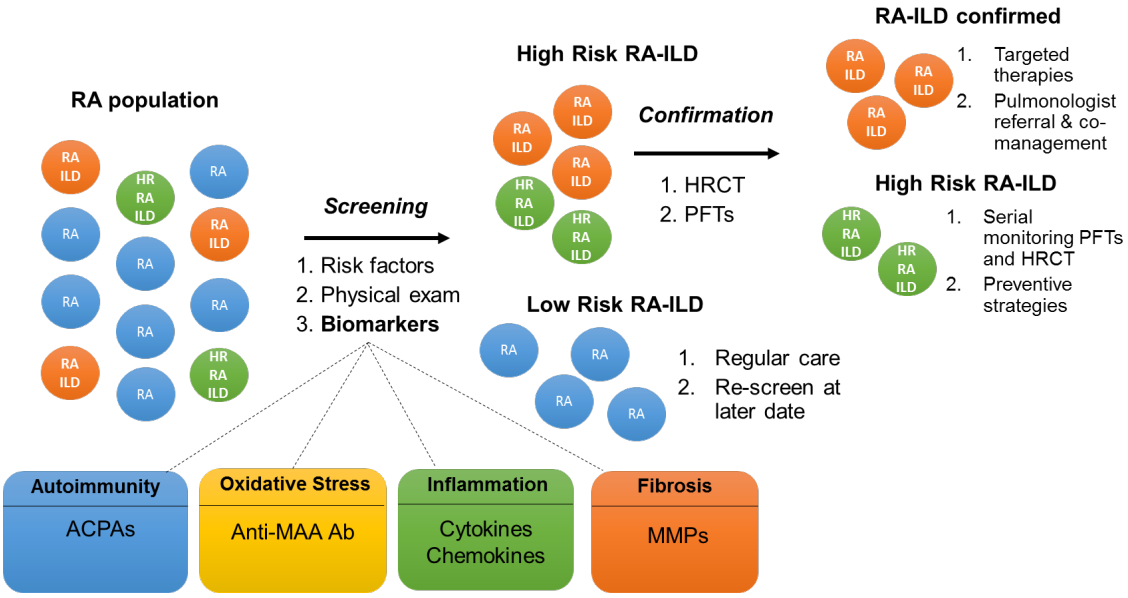
Anti-fibrotics have become standard of care for the treatment of IPF, slowing the rate of progression and reducing mortality by 30-50%.[169-171] Owing to the similarities of RA-ILD to IPF, randomized controlled trials of anti-fibrotics (pirfenidone [NCT02808871]

and nintedanib [NCT02999178]) in RA-ILD are currently enrolling. Recently, a randomized controlled trial of nintedanib in systemic sclerosis-ILD was completed finding less decline in PFTs with nintedanib compared to placebo.[172] However, there was no improvement in other systemic sclerosis or ILD outcomes (e.g., modified Rodnan skin score and St. George's Respiratory Questionnaire) and background mycophenolate treatment appeared to be most influential on PFT progression. Extrapolating from this study, it seems unlikely that anti-fibrotics will replace the need for DMARDs in RA. Other lung focused therapies such as azathioprine, mycophenolate, cyclophosphamide, and tacrolimus (agents that are not routinely used to treat articular disease in RA) have been evaluated in small, uncontrolled studies that included RA-ILD.[173-176] However, the precise role of these therapies in RA-ILD is unclear. With the anticipated increased use of lung targeted therapies in RA-ILD, pharmacoepidemiologic studies will be needed to assess the efficacy and safety of these agents, particularly when used in combination with RA DMARDs used to manage articular manifestations.

Optimal treatments in other RA-lung diseases are also unknown. There is concern about the safety of using abatacept in RA patients with COPD based on an increased frequency of COPD exacerbations reported in a single randomized controlled trial of abatacept in RA.[91] However, a recent pharmacoepidemiologic study with a prevalent new-user design and time-conditional propensity score matching compared abatacept to other bDMARDs in RA subjects with COPD found no increased risk of respiratory events.[177] Bronchiectasis, another form of chronic lung disease that can complicate the course of RA, is characterized by bacterial colonization of the lower respiratory tract that increases the risk of pulmonary infections.[178] Treatment with biologic therapies is worrisome in these patients because of the increased risk for serious infection.[179] Although not limited to bronchiectasis, findings detailed in Chapter 2

reassuringly did not identify conventional or bDMARDs to be associated with a significantly higher mortality risk in RA patients with chronic lung disease. Much more pharmacoepidemiologic evaluation of RA therapies in those with chronic lung diseases are needed, but development of valid algorithms for identifying these conditions (as we have done for RA-ILD) will be an important first step to facilitate this line of investigation moving forward.

Figure 8. Incorporation of serum biomarkers into RA-ILD identification



Proposed model for the incorporation of serum biomarkers into the identification of rheumatoid arthritis interstitial lung disease (RA-ILD). Multiple biomarkers are used to screen RA cohorts for ILD or individuals at high-risk for ILD. Confirmatory testing is obtained through a high-resolution computed tomography (HRCT), and pulmonary function tests (PFTs) to establish the diagnosis, prognosis, and treatment plan.

Abbreviations: ACPAs, anti-citrullinated protein antibodies; HR, high risk; HRCT, high-resolution computed tomography; ILD, interstitial lung disease; MAA, malondialdehyde-acetaldehyde adducts; MMPs, matrix metalloproteinases; PFTs, pulmonary function tests; RA, rheumatoid arthritis

Table 17. Candidate serum biomarkers in RA-ILD

Biomarker	Description	Performance in RA-ILD
Anti-citrullinated protein antibodies (ACPAs)	Antibodies to citrullinated peptides/proteins	AUC 0.46-0.75 for discriminating RA-ILD from RA.[71,72] Number of high level ACPAs independently associated with ILD in RA.[66]
Anti-Cit-HSP90	Antibody targeting citrullinated heat shock protein 90, chaperone proteins that regulate protein folding	High specificity (>95%) but low sensitivity (<30%) for discriminating RA-ILD from RA, mixed connective tissue disease, and IPF.[70]
Anti-MAA antibody	Antibody to product of oxidative stress	Higher quartiles of IgM antibody with >2-fold higher odds of ILD in a large RA cohort independent of traditional RA-ILD risk factors. Higher values in RA-ILD than RA with COPD.[180]
Anti-PAD3/4	Antibodies to PAD enzyme isoforms 3 & 4	Presence of cross-reactive antibody independently associated with 7-fold higher odds of ILD in RA cohort[181]
Interferon-γ-inducible protein 10	CXC family cytokine involved in chemotaxis	AUC 0.71-0.74 for discriminating RA-ILD from RA.[71]
Krebs von den Lungen-6	Glycoprotein expressed on pulmonary epithelial cells	Correlates with ILD severity on CT.[69]
Matrix metalloproteinase-7	Enzyme involved in the remodeling of extracellular matrix	AUC 0.68-0.86 for discriminating RA-ILD from RA.[71,72]
Pulmonary and activation-regulated chemokine	Chemotactic factor for T cells expressed in the lungs	AUC 0.70-0.80 for discriminating RA-ILD from RA.[72]
Rheumatoid factor	Antibody targeting the Fc portion of an IgG	AUC 0.59-0.67 for discriminating RA-ILD from RA.[71,72]
Surfactant protein-D	Collectin secreted from epithelial cells that primarily mediates host-defense function of surfactant	AUC 0.75-0.91 for discriminating RA-ILD from RA.[72]

Abbreviations: AUC, area under the curve; Cit-HSP90, citrullinated heat shock protein 90; ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; MAA, malondialdehyde-acetaldehyde adducts; PAD, peptidyl-arginine deiminase; RA, rheumatoid arthritis

5.3 Conclusion

RA-associated lung diseases are responsible for a significant proportion of the morbidity and premature mortality in RA patients. Despite recognizing their existence for over 60 years, there remains a poor understanding of their pathophysiology, optimal methods for identification, and the best medications to select for treatment when present. This dissertation advances our understanding of the impact that chronic lung diseases have on mortality in RA patients, raising awareness and encouraging further research to target pulmonary manifestations as aggressively as cardiovascular disease has been targeted. To facilitate the earlier identification of RA-ILD, we identified a novel autoantibody that is independently associated with RA-ILD that could be used as part of a screening process for this extra-articular manifestation. Finally, we have developed administrative algorithms that accurately classify RA-ILD status in RA cohorts that can be used to leverage large, real-world datasets to perform high-impact comparative effectiveness and outcomes research. Together, these studies improve our ability to identify RA-associated lung diseases and the tools developed will facilitate the completion of the proposed clinical and translational research in RA-associated lung diseases.

BIBLIOGRAPHY

1. Helmick, C.G., Felson, D.T., Lawrence, R.C., Gabriel, S., Hirsch, R., Kwoh, C.K., Liang, M.H., Kremers, H.M., Mayes, M.D., Merkel, P.A., Pillemer, S.R., Reveille, J.D. & Stone, J.H. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum* **58**, 15-25 (2008).
2. Myasoedova, E., Crowson, C.S., Kremers, H.M., Therneau, T.M. & Gabriel, S.E. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. *Arthritis Rheum* **62**, 1576-1582 (2010).
3. Erickson, A.R., Cannella, A.C. & Mikuls, T.R. Clinical features of rheumatoid arthritis. in *Kelley and Firestein's textbook of rheumatology* 1167-1186 (Elsevier, 2017).
4. Young, A. & Koduri, G. Extra-articular manifestations and complications of rheumatoid arthritis. *Best Pract Res Clin Rheumatol* **21**, 907-927 (2007).
5. Weyand, C.M., Schmidt, D., Wagner, U. & Goronzy, J.J. The influence of sex on the phenotype of rheumatoid arthritis. *Arthritis Rheum* **41**, 817-822 (1998).
6. Nishimura, K., Sugiyama, D., Kogata, Y., Tsuji, G., Nakazawa, T., Kawano, S., Saigo, K., Morinobu, A., Koshiba, M., Kuntz, K.M., Kamae, I. & Kumagai, S. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Ann Intern Med* **146**, 797-808 (2007).
7. Ingegnoli, F., Castelli, R. & Gualtierotti, R. Rheumatoid factors: clinical applications. *Dis Markers* **35**, 727-734 (2013).
8. England, B.R., Thiele, G.M. & Mikuls, T.R. Anticitrullinated protein antibodies: origin and role in the pathogenesis of rheumatoid arthritis. *Curr Opin Rheumatol* **29**, 57-64 (2017).

9. Thiele, G.M., Duryee, M.J., Anderson, D.R., Klassen, L.W., Mohring, S.M., Young, K.A., Benissan-Messan, D., Sayles, H., Dusad, A., Hunter, C.D., Sokolove, J., Robinson, W.H., O'Dell, J.R., Nicholas, A.P., Tuma, D.J. & Mikuls, T.R. Malondialdehyde-acetaldehyde adducts and anti-malondialdehyde-acetaldehyde antibodies in rheumatoid arthritis. *Arthritis Rheumatol* **67**, 645-655 (2015).

10. Mikuls, T.R., Duryee, M.J., England, B.R., Anderson, D.R., Hearth-Holmes, M., Su, K., Michaud, K., Payne, J.B., Sayles, H., Hunter, C., McGowan, J.D., Klassen, L.W. & Thiele, G.M. Malondialdehyde-acetaldehyde antibody concentrations in rheumatoid arthritis and other rheumatic conditions. *Int Immunopharmacol* **56**, 113-118 (2018).

11. Shi, J., Knevel, R., Suwannalai, P., van der Linden, M.P., Janssen, G.M., van Veelen, P.A., Levarht, N.E., van der Helm-van Mil, A.H., Cerami, A., Huizinga, T.W., Toes, R.E. & Trouw, L.A. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc Natl Acad Sci U S A* **108**, 17372-17377 (2011).

12. Harris, M.L., Darrah, E., Lam, G.K., Bartlett, S.J., Giles, J.T., Grant, A.V., Gao, P., Scott, W.W., Jr., El-Gabalawy, H., Casciola-Rosen, L., Barnes, K.C., Bathon, J.M. & Rosen, A. Association of autoimmunity to peptidyl arginine deiminase type 4 with genotype and disease severity in rheumatoid arthritis. *Arthritis Rheum* **58**, 1958-1967 (2008).

13. Holoshitz, J. The rheumatoid arthritis HLA-DRB1 shared epitope. *Curr Opin Rheumatol* **22**, 293-298 (2010).

14. Viatte, S., Plant, D., Han, B., Fu, B., Yarwood, A., Thomson, W., Symmons, D.P., Worthington, J., Young, A., Hyrich, K.L., Morgan, A.W., Wilson, A.G., Isaacs, J.D., Raychaudhuri, S. & Barton, A. Association of HLA-DRB1 haplotypes with

rheumatoid arthritis severity, mortality, and treatment response. *JAMA* **313**, 1645-1656 (2015).

15. Zhao, M., Mauer, L., Sayles, H., Cannon, G.W., Reimold, A., Kerr, G.S., Baker, J.F., Thiele, G.M., England, B.R. & Mikuls, T.R. HLA-DRB1 Haplotypes, Shared Epitope, and Disease Outcomes in US Veterans with Rheumatoid Arthritis. *J Rheumatol* **46**, 685-693 (2019).
16. Begovich, A.B., Carlton, V.E., Honigberg, L.A., Schrodi, S.J., Chokkalingam, A.P., Alexander, H.C., Ardlie, K.G., Huang, Q., Smith, A.M., Spoerke, J.M., Conn, M.T., Chang, M., Chang, S.Y., Saiki, R.K., Catanese, J.J., Leong, D.U., Garcia, V.E., McAllister, L.B., Jeffery, D.A., Lee, A.T., Batliwalla, F., Remmers, E., Criswell, L.A., Seldin, M.F., Kastner, D.L., Amos, C.I., Sninsky, J.J. & Gregersen, P.K. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* **75**, 330-337 (2004).
17. Di Giuseppe, D., Discacciati, A., Orsini, N. & Wolk, A. Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. *Arthritis Res Ther* **16**, R61 (2014).
18. Sugiyama, D., Nishimura, K., Tamaki, K., Tsuji, G., Nakazawa, T., Morinobu, A. & Kumagai, S. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis* **69**, 70-81 (2010).
19. Klareskog, L., Stolt, P., Lundberg, K., Kallberg, H., Bengtsson, C., Grunewald, J., Ronnelid, J., Harris, H.E., Ulfgren, A.K., Rantapaa-Dahlqvist, S., Eklund, A., Padyukov, L. & Alfredsson, L. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum* **54**, 38-46 (2006).

20. Sokolove, J., Bromberg, R., Deane, K.D., Lahey, L.J., Derber, L.A., Chandra, P.E., Edison, J.D., Gilliland, W.R., Tibshirani, R.J., Norris, J.M., Holers, V.M. & Robinson, W.H. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. *PLoS One* **7**, e35296 (2012).

21. Deane, K.D., O'Donnell, C.I., Hueber, W., Majka, D.S., Lazar, A.A., Derber, L.A., Gilliland, W.R., Edison, J.D., Norris, J.M., Robinson, W.H. & Holers, V.M. The number of elevated cytokines and chemokines in preclinical seropositive rheumatoid arthritis predicts time to diagnosis in an age-dependent manner. *Arthritis Rheum* **62**, 3161-3172 (2010).

22. Kosinski, M., Kujawski, S.C., Martin, R., Wanke, L.A., Buatti, M.C., Ware, J.E., Jr. & Perfetto, E.M. Health-related quality of life in early rheumatoid arthritis: impact of disease and treatment response. *Am J Manag Care* **8**, 231-240 (2002).

23. Drossaers-Bakker, K.W., de Buck, M., van Zeben, D., Zwinderman, A.H., Breedveld, F.C. & Hazes, J.M. Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. *Arthritis Rheum* **42**, 1854-1860 (1999).

24. Wolfe, F. & Hawley, D.J. The longterm outcomes of rheumatoid arthritis: Work disability: a prospective 18 year study of 823 patients. *J Rheumatol* **25**, 2108-2117 (1998).

25. Wolfe, F. & Zwillich, S.H. The long-term outcomes of rheumatoid arthritis: a 23-year prospective, longitudinal study of total joint replacement and its predictors in 1,600 patients with rheumatoid arthritis. *Arthritis Rheum* **41**, 1072-1082 (1998).

26. England, B.R., Sayles, H., Michaud, K., Caplan, L., Davis, L.A., Cannon, G.W., Sauer, B.C., Solow, E.B., Reimold, A.M., Kerr, G.S., Schwab, P., Baker, J.F. & Mikuls, T.R. Cause-Specific Mortality in Male US Veterans With Rheumatoid Arthritis. *Arthritis Care Res (Hoboken)* **68**, 36-45 (2016).

27. Sparks, J.A., Chang, S.C., Liao, K.P., Lu, B., Fine, A.R., Solomon, D.H., Costenbader, K.H. & Karlson, E.W. Rheumatoid Arthritis and Mortality Among Women During 36 Years of Prospective Follow-Up: Results From the Nurses' Health Study. *Arthritis Care Res (Hoboken)* **68**, 753-762 (2016).

28. Michaud, K. & Wolfe, F. Comorbidities in rheumatoid arthritis. *Best Pract Res Clin Rheumatol* **21**, 885-906 (2007).

29. Mikuls, T.R. & Saag, K.G. Comorbidity in rheumatoid arthritis. *Rheum Dis Clin North Am* **27**, 283-303 (2001).

30. Singh, J.A., Saag, K.G., Bridges, S.L., Jr., Akl, E.A., Bannuru, R.R., Sullivan, M.C., Vaysbrot, E., McNaughton, C., Osani, M., Shmerling, R.H., Curtis, J.R., Furst, D.E., Parks, D., Kavanaugh, A., O'Dell, J., King, C., Leong, A., Matteson, E.L., Schousboe, J.T., Drevlow, B., Ginsberg, S., Grober, J., St Clair, E.W., Tindall, E., Miller, A.S. & McAlindon, T. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol* **68**, 1-26 (2016).

31. Smolen, J.S., Landewe, R., Bijlsma, J., Burmester, G., Chatzidionysiou, K., Dougados, M., Nam, J., Ramiro, S., Voshaar, M., van Vollenhoven, R., Aletaha, D., Aringer, M., Boers, M., Buckley, C.D., Buttgereit, F., Bykerk, V., Cardiel, M., Combe, B., Cutolo, M., van Eijk-Hustings, Y., Emery, P., Finckh, A., Gabay, C., Gomez-Reino, J., Gossec, L., Gottenberg, J.E., Hazes, J.M.W., Huizinga, T., Jani, M., Karateev, D., Kouloumas, M., Kvien, T., Li, Z., Mariette, X., McInnes, I., Mysler, E., Nash, P., Pavelka, K., Poor, G., Richez, C., van Riel, P., Rubbert-Roth, A., Saag, K., da Silva, J., Stamm, T., Takeuchi, T., Westhovens, R., de Wit, M. & van der Heijde, D. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* **76**, 960-977 (2017).

32. Anderson, J., Caplan, L., Yazdany, J., Robbins, M.L., Neogi, T., Michaud, K., Saag, K.G., O'Dell, J.R. & Kazi, S. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res (Hoboken)* **64**, 640-647 (2012).

33. Anderson, J.K., Zimmerman, L., Caplan, L. & Michaud, K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score with 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Patient Activity Score (PAS) and Patient Activity Score-II (PASII), Routine Assessment of Patient Index Data (RAPID), Rheumatoid Arthritis Disease Activity Index (RADAI) and Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5), Chronic Arthritis Systemic Index (CASI), Patient-Based Disease Activity Score With ESR (PDAS1) and Patient-Based Disease Activity Score without ESR (PDAS2), and Mean Overall Index for Rheumatoid Arthritis (MOI-RA). *Arthritis Care Res (Hoboken)* **63 Suppl 11**, S14-36 (2011).

34. O'Dell, J.R., Curtis, J.R., Mikuls, T.R., Cofield, S.S., Bridges, S.L., Jr., Ranganath, V.K., Moreland, L.W. & Investigators, T.T. Validation of the methotrexate-first strategy in patients with early, poor-prognosis rheumatoid arthritis: results from a two-year randomized, double-blind trial. *Arthritis Rheum* **65**, 1985-1994 (2013).

35. O'Dell, J.R., Mikuls, T.R., Taylor, T.H., Ahluwalia, V., Brophy, M., Warren, S.R., Lew, R.A., Cannella, A.C., Kunkel, G., Phibbs, C.S., Anis, A.H., Leatherman, S., Keystone, E. & Investigators, C.R. Therapies for active rheumatoid arthritis after methotrexate failure. *N Engl J Med* **369**, 307-318 (2013).

36. Aga, A.B., Lie, E., Uhlig, T., Olsen, I.C., Wierod, A., Kalstad, S., Rodevand, E., Mikkelsen, K., Kvien, T.K. & Haavardsholm, E.A. Time trends in disease activity, response and remission rates in rheumatoid arthritis during the past decade: results from the NOR-DMARD study 2000-2010. *Ann Rheum Dis* **74**, 381-388 (2015).

37. Louie, G.H. & Ward, M.M. Changes in the rates of joint surgery among patients with rheumatoid arthritis in California, 1983-2007. *Ann Rheum Dis* **69**, 868-871 (2010).

38. Krishnan, E., Lingala, B., Bruce, B. & Fries, J.F. Disability in rheumatoid arthritis in the era of biological treatments. *Ann Rheum Dis* **71**, 213-218 (2012).

39. Dadoun, S., Zeboulon-Ktorza, N., Combescure, C., Elhai, M., Rozenberg, S., Gossec, L. & Fautrel, B. Mortality in rheumatoid arthritis over the last fifty years: systematic review and meta-analysis. *Joint Bone Spine* **80**, 29-33 (2013).

40. Myasoedova, E., Gabriel, S.E., Matteson, E.L., Davis, J.M., 3rd, Therneau, T.M. & Crowson, C.S. Decreased Cardiovascular Mortality in Patients with Incident Rheumatoid Arthritis (RA) in Recent Years: Dawn of a New Era in Cardiovascular Disease in RA? *J Rheumatol* **44**, 732-739 (2017).

41. England, B.R., Thiele, G.M., Anderson, D.R. & Mikuls, T.R. Increased cardiovascular risk in rheumatoid arthritis: mechanisms and implications. *BMJ* **361**, k1036 (2018).

42. Ellman, P. & Ball, R.E. Rheumatoid disease with joint and pulmonary manifestations. *Br Med J* **2**, 816-820 (1948).

43. Caplan, A. Certain unusual radiological appearances in the chest of coal-miners suffering from rheumatoid arthritis. *Thorax* **8**, 29-37 (1953).

44. Shaw, M., Collins, B.F., Ho, L.A. & Raghu, G. Rheumatoid arthritis-associated lung disease. *Eur Respir Rev* **24**, 1-16 (2015).

45. Sparks, J.A., Lin, T.C., Camargo, C.A., Jr., Barbhaiya, M., Tedeschi, S.K., Costenbader, K.H., Raby, B.A., Choi, H.K. & Karlson, E.W. Rheumatoid arthritis and risk of chronic obstructive pulmonary disease or asthma among women: A marginal structural model analysis in the Nurses' Health Study. *Semin Arthritis Rheum* **47**, 639-648 (2018).

46. Nannini, C., Medina-Velasquez, Y.F., Achenbach, S.J., Crowson, C.S., Ryu, J.H., Vassallo, R., Gabriel, S.E., Matteson, E.L. & Bongartz, T. Incidence and mortality of obstructive lung disease in rheumatoid arthritis: a population-based study. *Arthritis Care Res (Hoboken)* **65**, 1243-1250 (2013).

47. Koduri, G., Norton, S., Young, A., Cox, N., Davies, P., Devlin, J., Dixey, J., Gough, A., Prouse, P. & Winfield, J. Interstitial lung disease has a poor prognosis in rheumatoid arthritis: results from an inception cohort. *Rheumatology* **49**, 1483-1489 (2010).

48. Kelly, C.A., Saravanan, V., Nisar, M., Arthanari, S., Woodhead, F.A., Price-Forbes, A.N., Dawson, J., Sathi, N., Ahmad, Y. & Koduri, G. Rheumatoid arthritis-related interstitial lung disease: associations, prognostic factors and physiological and radiological characteristics—a large multicentre UK study. *Rheumatology* **53**, 1676-1682 (2014).

49. Quirke, A.M., Perry, E., Cartwright, A., Kelly, C., De Soyza, A., Eggleton, P., Hutchinson, D. & Venables, P.J. Bronchiectasis is a Model for Chronic Bacterial Infection Inducing Autoimmunity in Rheumatoid Arthritis. *Arthritis Rheumatol* **67**, 2335-2342 (2015).

50. Makrygiannakis, D., Hermansson, M., Ulfgren, A.K., Nicholas, A.P., Zendman, A.J., Eklund, A., Grunewald, J., Skold, C.M., Klareskog, L. & Catrina, A.I. Smoking increases peptidylarginine deiminase 2 enzyme expression in human

lungs and increases citrullination in BAL cells. *Ann Rheum Dis* **67**, 1488-1492 (2008).

51. Demoruelle, M.K., Weisman, M.H., Simonian, P.L., Lynch, D.A., Sachs, P.B., Pedraza, I.F., Harrington, A.R., Kolfenbach, J.R., Striebich, C.C., Pham, Q.N., Strickland, C.D., Petersen, B.D., Parish, M.C., Derber, L.A., Norris, J.M., Holers, V.M. & Deane, K.D. Brief report: airways abnormalities and rheumatoid arthritis-related autoantibodies in subjects without arthritis: early injury or initiating site of autoimmunity? *Arthritis Rheum* **64**, 1756-1761 (2012).
52. Willis, V.C., Demoruelle, M.K., Derber, L.A., Chartier-Logan, C.J., Parish, M.C., Pedraza, I.F., Weisman, M.H., Norris, J.M., Holers, V.M. & Deane, K.D. Sputum autoantibodies in patients with established rheumatoid arthritis and subjects at risk of future clinically apparent disease. *Arthritis Rheum* **65**, 2545-2554 (2013).
53. Reynisdottir, G., Karimi, R., Joshua, V., Olsen, H., Hensvold, A.H., Harju, A., Engstrom, M., Grunewald, J., Nyren, S., Eklund, A., Klareskog, L., Skold, C.M. & Catrina, A.I. Structural changes and antibody enrichment in the lungs are early features of anti-citrullinated protein antibody-positive rheumatoid arthritis. *Arthritis Rheumatol* **66**, 31-39 (2014).
54. Rangel-Moreno, J., Hartson, L., Navarro, C., Gaxiola, M., Selman, M. & Randall, T.D. Inducible bronchus-associated lymphoid tissue (iBALT) in patients with pulmonary complications of rheumatoid arthritis. *J Clin Invest* **116**, 3183-3194 (2006).
55. Ytterberg, A.J., Joshua, V., Reynisdottir, G., Tarasova, N.K., Rutishauser, D., Ossipova, E., Haj Hensvold, A., Eklund, A., Skold, C.M., Grunewald, J., Malmstrom, V., Jakobsson, P.J., Ronnelid, J., Padyukov, L., Zubarev, R.A., Klareskog, L. & Catrina, A.I. Shared immunological targets in the lungs and joints

of patients with rheumatoid arthritis: identification and validation. *Ann Rheum Dis* **74**, 1772-1777 (2015).

56. Mikuls, T.R., Payne, J.B., Deane, K.D. & Thiele, G.M. Autoimmunity of the lung and oral mucosa in a multisystem inflammatory disease: The spark that lights the fire in rheumatoid arthritis? *J Allergy Clin Immunol* **137**, 28-34 (2016).

57. Bongartz, T., Nannini, C., Medina-Velasquez, Y.F., Achenbach, S.J., Crowson, C.S., Ryu, J.H., Vassallo, R., Gabriel, S.E. & Matteson, E.L. Incidence and mortality of interstitial lung disease in rheumatoid arthritis: a population-based study. *Arthritis Rheum* **62**, 1583-1591 (2010).

58. Spagnolo, P., Lee, J.S., Sverzellati, N., Rossi, G. & Cottin, V. The Lung in Rheumatoid Arthritis: Focus on Interstitial Lung Disease. *Arthritis Rheumatol* **70**, 1544-1554 (2018).

59. Sparks, J.A., He, X., Huang, J., Fletcher, E.A., Zaccardelli, A., Friedlander, H.M., Gill, R.R., Hatabu, H., Nishino, M., Murphy, D.J., Iannaccone, C.K., Mahmoud, T.G., Frits, M.L., Lu, B., Rosas, I.O., Dellaripa, P.F., Weinblatt, M.E., Karlson, E.W., Shadick, N.A. & Doyle, T.J. Rheumatoid arthritis disease activity predicting incident clinically-apparent RA-associated interstitial lung disease: A prospective cohort study. *Arthritis Rheumatol* (2019).

60. Kiely, P., Busby, A.D., Nikiphorou, E., Sullivan, K., Walsh, D.A., Creamer, P., Dixey, J. & Young, A. Is incident rheumatoid arthritis interstitial lung disease associated with methotrexate treatment? Results from a multivariate analysis in the ERAS and ERAN inception cohorts. *BMJ Open* **9**, e028466 (2019).

61. Lee, H.K., Kim, D.S., Yoo, B., Seo, J.B., Rho, J.Y., Colby, T.V. & Kitaichi, M. Histopathologic pattern and clinical features of rheumatoid arthritis-associated interstitial lung disease. *Chest* **127**, 2019-2027 (2005).

62. Juge, P.A., Lee, J.S., Ebstein, E., Furukawa, H., Dobrinskikh, E., Gazal, S., Kannengiesser, C., Ottaviani, S., Oka, S., Tohma, S., Tsuchiya, N., Rojas-Serrano, J., Gonzalez-Perez, M.I., Mejia, M., Buendia-Roldan, I., Falfan-Valencia, R., Ambrocio-Ortiz, E., Manali, E., Papiris, S.A., Karageorgas, T., Boumpas, D., Antoniou, K., van Moorsel, C.H.M., van der Vis, J., de Man, Y.A., Grutters, J.C., Wang, Y., Borie, R., Wemeau-Stervinou, L., Wallaert, B., Flipo, R.M., Nunes, H., Valeyre, D., Saidenberg-Kermanac'h, N., Boissier, M.C., Marchand-Adam, S., Frazier, A., Richette, P., Allanore, Y., Sibilia, J., Dromer, C., Richez, C., Schaeverbeke, T., Liote, H., Thabut, G., Nathan, N., Amselem, S., Soubrier, M., Cottin, V., Clement, A., Deane, K., Walts, A.D., Fingerlin, T., Fischer, A., Ryu, J.H., Matteson, E.L., Niewold, T.B., Assayag, D., Gross, A., Wolters, P., Schwarz, M.I., Holers, M., Solomon, J.J., Doyle, T., Rosas, I.O., Blauwendraat, C., Nalls, M.A., Debray, M.P., Boileau, C., Crestani, B., Schwartz, D.A. & Dieude, P. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. *N Engl J Med* **379**, 2209-2219 (2018).

63. Raimundo, K., Solomon, J.J., Olson, A.L., Kong, A.M., Cole, A.L., Fischer, A. & Swigris, J.J. Rheumatoid Arthritis-Interstitial Lung Disease in the United States: Prevalence, Incidence, and Healthcare Costs and Mortality. *J Rheumatol* **46**, 360-369 (2019).

64. Tsuchiya, Y., Takayanagi, N., Sugiura, H., Miyahara, Y., Tokunaga, D., Kawabata, Y. & Sugita, Y. Lung diseases directly associated with rheumatoid arthritis and their relationship to outcome. *Eur Respir J* **37**, 1411-1417 (2011).

65. Gabbay, E., Tarala, R., Will, R., Carroll, G., Adler, B., Cameron, D. & Lake, F.R. Interstitial lung disease in recent onset rheumatoid arthritis. *Am J Respir Crit Care Med* **156**, 528-535 (1997).

66. Giles, J.T., Danoff, S.K., Sokolove, J., Wagner, C.A., Winchester, R., Pappas, D.A., Siegelman, S., Connors, G., Robinson, W.H. & Bathon, J.M. Association of fine specificity and repertoire expansion of anticitrullinated peptide antibodies with rheumatoid arthritis associated interstitial lung disease. *Ann Rheum Dis* **73**, 1487-1494 (2014).

67. Saag, K.G., Kolluri, S., Koehnke, R.K., Georgou, T.A., Rachow, J.W., Hunninghake, G.W. & Schwartz, D.A. Rheumatoid arthritis lung disease. Determinants of radiographic and physiologic abnormalities. *Arthritis Rheum* **39**, 1711-1719 (1996).

68. Aubart, F., Crestani, B., Nicaise-Roland, P., Tubach, F., Bollet, C., Dawidowicz, K., Quintin, E., Hayem, G., Palazzo, E., Meyer, O., Chollet-Martin, S. & Dieude, P. High levels of anti-cyclic citrullinated peptide autoantibodies are associated with co-occurrence of pulmonary diseases with rheumatoid arthritis. *J Rheumatol* **38**, 979-982 (2011).

69. Kinoshita, F., Hamano, H., Harada, H., Kinoshita, T., Igishi, T., Hagino, H. & Ogawa, T. Role of KL-6 in evaluating the disease severity of rheumatoid lung disease: comparison with HRCT. *Respir Med* **98**, 1131-1137 (2004).

70. Harlow, L., Rosas, I.O., Gochuico, B.R., Mikuls, T.R., Dellaripa, P.F., Oddis, C.V. & Ascherman, D.P. Identification of citrullinated hsp90 isoforms as novel autoantigens in rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheum* **65**, 869-879 (2013).

71. Chen, J., Doyle, T.J., Liu, Y., Aggarwal, R., Wang, X., Shi, Y., Ge, S.X., Huang, H., Lin, Q., Liu, W., Cai, Y., Koontz, D., Fuhrman, C.R., Golzarri, M.F., Liu, Y., Hatabu, H., Nishino, M., Araki, T., Dellaripa, P.F., Oddis, C.V., Rosas, I.O. & Ascherman, D.P. Biomarkers of rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheumatol* **67**, 28-38 (2015).

72. Doyle, T.J., Patel, A.S., Hatabu, H., Nishino, M., Wu, G., Osorio, J.C., Golzarri, M.F., Traslosheros, A., Chu, S.G., Frits, M.L., Iannaccone, C.K., Koontz, D., Fuhrman, C., Weinblatt, M.E., El-Chemaly, S.Y., Washko, G.R., Hunninghake, G.M., Choi, A.M., Dellaripa, P.F., Oddis, C.V., Shadick, N.A., Ascherman, D.P. & Rosas, I.O. Detection of Rheumatoid Arthritis-Interstitial Lung Disease Is Enhanced by Serum Biomarkers. *Am J Respir Crit Care Med* **191**, 1403-1412 (2015).

73. Doyle, T.J., Lee, J.S., Dellaripa, P.F., Lederer, J.A., Matteson, E.L., Fischer, A., Ascherman, D.P., Glassberg, M.K., Ryu, J.H., Danoff, S.K., Brown, K.K., Collard, H.R. & Rosas, I.O. A roadmap to promote clinical and translational research in rheumatoid arthritis-associated interstitial lung disease. *Chest* **145**, 454-463 (2014).

74. Roubille, C. & Haraoui, B. Interstitial lung diseases induced or exacerbated by DMARDS and biologic agents in rheumatoid arthritis: a systematic literature review. *Semin Arthritis Rheum* **43**, 613-626 (2014).

75. Perez-Alvarez, R., Perez-de-Lis, M., Diaz-Lagares, C., Pego-Reigosa, J.M., Retamozo, S., Bove, A., Brito-Zeron, P., Bosch, X. & Ramos-Casals, M. Interstitial lung disease induced or exacerbated by TNF-targeted therapies: analysis of 122 cases. *Semin Arthritis Rheum* **41**, 256-264 (2011).

76. Hadjinicolaou, A.V., Nisar, M.K., Bhagat, S., Parfrey, H., Chilvers, E.R. & Ostor, A.J. Non-infectious pulmonary complications of newer biological agents for rheumatic diseases--a systematic literature review. *Rheumatology (Oxford)* **50**, 2297-2305 (2011).

77. Curtis, J.R., Sarsour, K., Napalkov, P., Costa, L.A. & Schulman, K.L. Incidence and complications of interstitial lung disease in users of tocilizumab, rituximab,

abatacept and anti-tumor necrosis factor alpha agents, a retrospective cohort study. *Arthritis Res Ther* **17**, 319 (2015).

78. Suissa, S., Hudson, M. & Ernst, P. Leflunomide use and the risk of interstitial lung disease in rheumatoid arthritis. *Arthritis Rheum* **54**, 1435-1439 (2006).

79. Jones, N., Schneider, G., Kachroo, S., Rotella, P., Avetisyan, R. & Reynolds, M.W. A systematic review of validated methods for identifying pulmonary fibrosis and interstitial lung disease using administrative and claims data. *Pharmacoepidemiol Drug Saf* **21 Suppl 1**, 256-260 (2012).

80. Chung, C.P., Rohan, P., Krishnaswami, S. & McPheeters, M.L. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. *Vaccine* **31 Suppl 10**, K41-61 (2013).

81. Moores, K.G. & Sathe, N.A. A systematic review of validated methods for identifying systemic lupus erythematosus (SLE) using administrative or claims data. *Vaccine* **31 Suppl 10**, K62-73 (2013).

82. Harrold, L.R., Saag, K.G., Yood, R.A., Mikuls, T.R., Andrade, S.E., Fouayzi, H., Davis, J., Chan, K.A., Raebel, M.A., Von Worley, A. & Platt, R. Validity of gout diagnoses in administrative data. *Arthritis Rheum* **57**, 103-108 (2007).

83. Shrestha, S., Dave, A.J., Losina, E. & Katz, J.N. Diagnostic accuracy of administrative data algorithms in the diagnosis of osteoarthritis: a systematic review. *BMC Med Inform Decis Mak* **16**, 82 (2016).

84. Olson, A.L., Swigris, J.J., Sprunger, D.B., Fischer, A., Fernandez-Perez, E.R., Solomon, J., Murphy, J., Cohen, M., Raghu, G. & Brown, K.K. Rheumatoid arthritis-interstitial lung disease-associated mortality. *Am J Respir Crit Care Med* **183**, 372-378 (2011).

85. Swinson, D.R., Symmons, D., Suresh, U., Jones, M. & Booth, J. Decreased survival in patients with co-existent rheumatoid arthritis and bronchiectasis. *Br J Rheumatol* **36**, 689-691 (1997).

86. Puechal, X., Genin, E., Bienvenu, T., Le Jeunne, C. & Dusser, D.J. Poor survival in rheumatoid arthritis associated with bronchiectasis: a family-based cohort study. *PLoS One* **9**, e110066 (2014).

87. Kremer, J.M., Alarcon, G.S., Weinblatt, M.E., Kaymakcian, M.V., Macaluso, M., Cannon, G.W., Palmer, W.R., Sundy, J.S., St Clair, E.W., Alexander, R.W., Smith, G.J. & Axiotis, C.A. Clinical, laboratory, radiographic, and histopathologic features of methotrexate-associated lung injury in patients with rheumatoid arthritis: a multicenter study with literature review. *Arthritis Rheum* **40**, 1829-1837 (1997).

88. Chikura, B., Lane, S. & Dawson, J.K. Clinical expression of leflunomide-induced pneumonitis. *Rheumatology (Oxford)* **48**, 1065-1068 (2009).

89. Lioté, H., Lioté, F., Séroussi, B., Mayaud, C. & Cadranel, J. Rituximab-induced lung disease: a systematic literature review. *European Respiratory Journal* **35**, 681-687 (2010).

90. Khasnis, A.A. & Calabrese, L.H. Tumor necrosis factor inhibitors and lung disease: a paradox of efficacy and risk. *Semin Arthritis Rheum* **40**, 147-163 (2010).

91. Weinblatt, M., Combe, B., Covucci, A., Aranda, R., Becker, J.C. & Keystone, E. Safety of the selective costimulation modulator abatacept in rheumatoid arthritis patients receiving background biologic and nonbiologic disease-modifying antirheumatic drugs: A one-year randomized, placebo-controlled study. *Arthritis Rheum* **54**, 2807-2816 (2006).

92. Avina-Zubieta, J.A., Thomas, J., Sadatsafavi, M., Lehman, A.J. & Lacaille, D. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis* **71**, 1524-1529 (2012).

93. Arnett, F.C., Edworthy, S.M., Bloch, D.A., McShane, D.J., Fries, J.F., Cooper, N.S., Healey, L.A., Kaplan, S.R., Liang, M.H., Luthra, H.S. & et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* **31**, 315-324 (1988).

94. Mikuls, T.R., Reimold, A., Kerr, G.S. & Cannon, G.W. Insights and Implications of the VA Rheumatoid Arthritis Registry. *Fed Pract* **32**, 24-29 (2015).

95. England, B.R., Sayles, H., Mikuls, T.R., Johnson, D.S. & Michaud, K. Validation of the rheumatic disease comorbidity index. *Arthritis Care Res (Hoboken)* **67**, 865-872 (2015).

96. Pincus, T., Sokka, T. & Kautiainen, H. Further development of a physical function scale on a MDHAQ [corrected] for standard care of patients with rheumatic diseases. *J Rheumatol* **32**, 1432-1439 (2005).

97. van der Heijde, D.M., van 't Hof, M., van Riel, P.L. & van de Putte, L.B. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* **20**, 579-581 (1993).

98. Mikuls, T.R., Gould, K.A., Bynote, K.K., Yu, F., Levan, T.D., Thiele, G.M., Michaud, K.D., O'Dell, J.R., Reimold, A.M., Hooker, R., Caplan, L., Johnson, D.S., Kerr, G., Richards, J.S., Cannon, G.W., Criswell, L.A., Noble, J.A., Bridges, S.L., Jr., Hughes, L. & Gregersen, P.K. Anticitrullinated protein antibody (ACPA) in rheumatoid arthritis: influence of an interaction between HLA-DRB1 shared epitope and a deletion polymorphism in glutathione S-transferase in a cross-sectional study. *Arthritis Res Ther* **12**, R213 (2010).

99. Miriovsky, B.J., Michaud, K., Thiele, G.M., O'Dell, J.R., Cannon, G.W., Kerr, G., Richards, J.S., Johnson, D., Caplan, L., Reimold, A., Hooker, R. & Mikuls, T.R. Anti-CCP antibody and rheumatoid factor concentrations predict greater disease activity in men with rheumatoid arthritis. *Ann Rheum Dis* **69**, 1292-1297 (2010).

100. Agca, R., Heslinga, S.C., Rollefstad, S., Heslinga, M., McInnes, I.B., Peters, M.J., Kvien, T.K., Dougados, M., Radner, H., Atzeni, F., Primdahl, J., Sodergren, A., Wallberg Jonsson, S., van Rompay, J., Zabalan, C., Pedersen, T.R., Jacobsson, L., de Vlam, K., Gonzalez-Gay, M.A., Semb, A.G., Kitas, G.D., Smulders, Y.M., Szekanecz, Z., Sattar, N., Symmons, D.P. & Nurmohamed, M.T. EULAR recommendations for cardiovascular disease risk management in patients with rheumatoid arthritis and other forms of inflammatory joint disorders: 2015/2016 update. *Ann Rheum Dis* **76**, 17-28 (2017).

101. McGuire, K., Avina-Zubieta, J.A., Esdaile, J.M., Sadatsafavi, M., Sayre, E.C., Abrahamowicz, M. & Lacaille, D. Risk of Incident Chronic Obstructive Pulmonary Disease in Rheumatoid Arthritis: A Population-Based Cohort Study. *Arthritis Care Res (Hoboken)* **71**, 602-610 (2019).

102. Mannino, D.M., Buist, A.S., Petty, T.L., Enright, P.L. & Redd, S.C. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. *Thorax* **58**, 388-393 (2003).

103. Roubille, C., Richer, V., Starnino, T., McCourt, C., McFarlane, A., Fleming, P., Siu, S., Kraft, J., Lynde, C., Pope, J., Gulliver, W., Keeling, S., Dutz, J., Bessette, L., Bissonnette, R. & Haraoui, B. The effects of tumour necrosis factor inhibitors, methotrexate, non-steroidal anti-inflammatory drugs and corticosteroids on cardiovascular events in rheumatoid arthritis, psoriasis and psoriatic arthritis: a systematic review and meta-analysis. *Ann Rheum Dis* **74**, 480-489 (2015).

104. Saravanan, V. & Kelly, C.A. Reducing the risk of methotrexate pneumonitis in rheumatoid arthritis. *Rheumatology (Oxford)* **43**, 143-147 (2004).

105. Travis, W.D., Costabel, U., Hansell, D.M., King, T.E., Jr., Lynch, D.A., Nicholson, A.G., Ryerson, C.J., Ryu, J.H., Selman, M., Wells, A.U., Behr, J., Bouros, D., Brown, K.K., Colby, T.V., Collard, H.R., Cordeiro, C.R., Cottin, V., Crestani, B., Drent, M., Dudden, R.F., Egan, J., Flaherty, K., Hogaboam, C., Inoue, Y., Johkoh, T., Kim, D.S., Kitaichi, M., Loyd, J., Martinez, F.J., Myers, J., Protzko, S., Raghu, G., Richeldi, L., Sverzellati, N., Swigris, J., Valeyre, D. & Pneumonias, A.E.C.o.I.I. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* **188**, 733-748 (2013).

106. Dawson, J.K., Fewins, H.E., Desmond, J., Lynch, M.P. & Graham, D.R. Fibrosing alveolitis in patients with rheumatoid arthritis as assessed by high resolution computed tomography, chest radiography, and pulmonary function tests. *Thorax* **56**, 622-627 (2001).

107. Chen, J., Shi, Y., Wang, X., Huang, H. & Ascherman, D. Asymptomatic preclinical rheumatoid arthritis-associated interstitial lung disease. *Clin Dev Immunol* **2013**, 406927 (2013).

108. Bargagli, E., Olivieri, C., Bennett, D., Prasse, A., Muller-Quernheim, J. & Rottoli, P. Oxidative stress in the pathogenesis of diffuse lung diseases: a review. *Respir Med* **103**, 1245-1256 (2009).

109. Thiele, G.M., Tuma, D.J., Willis, M.S., Miller, J.A., McDonald, T.L., Sorrell, M.F. & Klassen, L.W. Soluble proteins modified with acetaldehyde and malondialdehyde are immunogenic in the absence of adjuvant. *Alcohol Clin Exp Res* **22**, 1731-1739 (1998).

110. Mikuls, T.R., Duryee, M.J., Rahman, R., Anderson, D.R., Sayles, H.R., Hollins, A., Michaud, K., Wolfe, F., Thiele, G.E., Sokolove, J., Robinson, W.H., Lingampalli, N., Nicholas, A.P., Talmon, G.A., Su, K., Zimmerman, M.C., Klassen, L.W. & Thiele, G.M. Enrichment of malondialdehyde-acetaldehyde antibody in the rheumatoid arthritis joint. *Rheumatology (Oxford)* **56**, 1794-1803 (2017).

111. Wyatt, T.A., Kharbanda, K.K., McCaskill, M.L., Tuma, D.J., Yanov, D., DeVasure, J. & Sisson, J.H. Malondialdehyde-acetaldehyde-adducted protein inhalation causes lung injury. *Alcohol* **46**, 51-59 (2012).

112. Wyatt, T.A., Kharbanda, K.K., Tuma, D.J. & Sisson, J.H. Malondialdehyde-acetaldehyde-adducted bovine serum albumin activates protein kinase C and stimulates interleukin-8 release in bovine bronchial epithelial cells. *Alcohol* **25**, 159-166 (2001).

113. Pincus, T., Swearingen, C. & Wolfe, F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. *Arthritis Rheum* **42**, 2220-2230 (1999).

114. Prevoo, M.L., van 't Hof, M.A., Kuper, H.H., van Leeuwen, M.A., van de Putte, L.B. & van Riel, P.L. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* **38**, 44-48 (1995).

115. Dixon, W.G., Hyrich, K.L., Watson, K.D., Lunt, M., Consortium, B.C.C., Symmons, D.P. & British Society for Rheumatology Biologics, R. Influence of anti-TNF therapy on mortality in patients with rheumatoid arthritis-associated interstitial lung disease: results from the British Society for Rheumatology Biologics Register. *Ann Rheum Dis* **69**, 1086-1091 (2010).

116. U.S. Department of Veterans Affairs: VHA Corporate Data Warehouse. (2018).

117. Yin, Y., Liang, D., Zhao, L., Li, Y., Liu, W., Ren, Y., Li, Y., Zeng, X., Zhang, F., Tang, F., Shan, G. & Zhang, X. Anti-cyclic citrullinated Peptide antibody is associated with interstitial lung disease in patients with rheumatoid arthritis. *PLoS One* **9**, e92449 (2014).

118. Steen, V.D., Lanz, J.K., Jr., Conte, C., Owens, G.R. & Medsger, T.A., Jr. Therapy for severe interstitial lung disease in systemic sclerosis. A retrospective study. *Arthritis Rheum* **37**, 1290-1296 (1994).

119. Whiting, P.F., Smidt, N., Sterne, J.A., Harbord, R., Burton, A., Burke, M., Beynon, R., Ben-Shlomo, Y., Axford, J. & Dieppe, P. Systematic review: accuracy of anti-citrullinated Peptide antibodies for diagnosing rheumatoid arthritis. *Ann Intern Med* **152**, 456-464; W155-466 (2010).

120. Ruiz-Esquide, V., Gomara, M.J., Peinado, V.I., Gomez Puerta, J.A., Barbera, J.A., Canete Jde, D., Haro, I. & Sanmarti, R. Anti-citrullinated peptide antibodies in the serum of heavy smokers without rheumatoid arthritis. A differential effect of chronic obstructive pulmonary disease? *Clin Rheumatol* **31**, 1047-1050 (2012).

121. Sigari, N., Moghimi, N., Shahraki, F.S., Mohammadi, S. & Roshani, D. Anti-cyclic citrullinated peptide (CCP) antibody in patients with wood-smoke-induced chronic obstructive pulmonary disease (COPD) without rheumatoid arthritis. *Rheumatol Int* **35**, 85-91 (2015).

122. Maloley, P.M., Duryee, M.J., Hunter, C.D., O'Dell, J.R., Anderson, D.R., Mikuls, T.R., Thiele, G.M. & Klassen, L.W. Modification of Proteins with Malondialdehyde-Acetaldehyde and Citrulline Elicit Antibody Responses in DBA/1J Mice [abstract]. *Arthritis Rheum* **68**, 1824-1825 (2016).

123. Carmona, L., Gonzalez-Alvaro, I., Balsa, A., Angel Belmonte, M., Tena, X. & Sanmarti, R. Rheumatoid arthritis in Spain: occurrence of extra-articular

manifestations and estimates of disease severity. *Ann Rheum Dis* **62**, 897-900 (2003).

124. Hudson, M., Tascilar, K. & Suissa, S. Comparative effectiveness research with administrative health data in rheumatoid arthritis. *Nat Rev Rheumatol* **12**, 358-366 (2016).

125. Hyldgaard, C., Hilberg, O., Pedersen, A.B., Ulrichsen, S.P., Lokke, A., Bendstrup, E. & Ellingsen, T. A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. *Ann Rheum Dis* **76**, 1700-1706 (2017).

126. Herrinton, L.J., Harrold, L.R., Liu, L., Raebel, M.A., Taharka, A., Winthrop, K.L., Solomon, D.H., Curtis, J.R., Lewis, J.D. & Saag, K.G. Association between anti-TNF-alpha therapy and interstitial lung disease. *Pharmacoepidemiol Drug Saf* **22**, 394-402 (2013).

127. Harris, P.A., Taylor, R., Minor, B.L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S.N. & Consortium, R.E. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform* **95**, 103208 (2019).

128. Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N. & Conde, J.G. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* **42**, 377-381 (2009).

129. Ley, B., Urbania, T., Husson, G., Vittinghoff, E., Brush, D.R., Eisner, M.D., Iribarren, C. & Collard, H.R. Code-based Diagnostic Algorithms for Idiopathic Pulmonary Fibrosis. Case Validation and Improvement. *Ann Am Thorac Soc* **14**, 880-887 (2017).

130. Raghu, G., Weycker, D., Edelsberg, J., Bradford, W.Z. & Oster, G. Incidence and prevalence of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* **174**, 810-816 (2006).

131. Landis, J.R. & Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174 (1977).

132. Katki, H.A., Li, Y., Edelstein, D.W. & Castle, P.E. Estimating the agreement and diagnostic accuracy of two diagnostic tests when one test is conducted on only a subsample of specimens. *Stat Med* **31**, 436-448 (2012).

133. Benchimol, E.I., Manuel, D.G., To, T., Griffiths, A.M., Rabeneck, L. & Guttmann, A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* **64**, 821-829 (2011).

134. Esposito, D.B., Lanes, S., Donneyong, M., Holick, C.N., Lasky, J.A., Lederer, D., Nathan, S.D., O'Quinn, S., Parker, J. & Tran, T.N. Idiopathic Pulmonary Fibrosis in United States Automated Claims. Incidence, Prevalence, and Algorithm Validation. *Am J Respir Crit Care Med* **192**, 1200-1207 (2015).

135. Matsumoto, T., Iwano, S., Takahashi, N., Asai, S., Watanabe, T., Asai, N., Sobue, Y., Ito, S., Ishiguro, N. & Kojima, T. Association between chest computed tomography findings and respiratory adverse events in rheumatoid arthritis patients undergoing long-term biological therapy. *Int J Rheum Dis* (2018).

136. Jacob, J., Hirani, N., van Moorsel, C.H.M., Rajagopalan, S., Murchison, J.T., van Es, H.W., Bartholmai, B.J., van Beek, F.T., Struik, M.H.L., Stewart, G.A., Kokosi, M., Egashira, R., Brun, A.L., Cross, G., Barnett, J., Devaraj, A., Margaritopoulos, G., Karwoski, R., Renzoni, E., Maher, T.M. & Wells, A.U. Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J* **53**(2019).

137. Solomon, J.J., Chung, J.H., Cosgrove, G.P., Demoruelle, M.K., Fernandez-Perez, E.R., Fischer, A., Frankel, S.K., Hobbs, S.B., Huie, T.J., Ketzer, J., Mannina, A., Olson, A.L., Russell, G., Tsuchiya, Y., Yunt, Z.X., Zelarney, P.T., Brown, K.K. & Swigris, J.J. Predictors of mortality in rheumatoid arthritis-associated interstitial lung disease. *Eur Respir J* **47**, 588-596 (2016).

138. Zamora-Legoff, J.A., Krause, M.L., Crowson, C.S., Ryu, J.H. & Matteson, E.L. Progressive Decline of Lung Function in Rheumatoid Arthritis-Associated Interstitial Lung Disease. *Arthritis Rheumatol* **69**, 542-549 (2017).

139. Lammi, M.R., Baughman, R.P., Birring, S.S., Russell, A.M., Ryu, J.H., Scholand, M., Distler, O., LeSage, D., Sarver, C., Antoniou, K., Highland, K.B., Kowal-Bielecka, O., Lasky, J.A., Wells, A.U. & Saketkoo, L.A. Outcome Measures for Clinical Trials in Interstitial Lung Diseases. *Curr Respir Med Rev* **11**, 163-174 (2015).

140. Simon, T.A., Thompson, A., Gandhi, K.K., Hochberg, M.C. & Suissa, S. Incidence of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. *Arthritis Res Ther* **17**, 212 (2015).

141. Birnbaum, H., Pike, C., Kaufman, R., Marynchenko, M., Kidolezi, Y. & Cifaldi, M. Societal cost of rheumatoid arthritis patients in the US. *Curr Med Res Opin* **26**, 77-90 (2010).

142. Michaud, K., Messer, J., Choi, H.K. & Wolfe, F. Direct medical costs and their predictors in patients with rheumatoid arthritis: a three-year study of 7,527 patients. *Arthritis Rheum* **48**, 2750-2762 (2003).

143. Lumbreras, B., Donat, L. & Hernandez-Aguado, I. Incidental findings in imaging diagnostic tests: a systematic review. *Br J Radiol* **83**, 276-289 (2010).

144. Rosas, I.O., Richards, T.J., Konishi, K., Zhang, Y., Gibson, K., Lokshin, A.E., Lindell, K.O., Cisneros, J., Macdonald, S.D., Pardo, A., Sciurba, F., Dauber, J.,

Selman, M., Gochuico, B.R. & Kaminski, N. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med* **5**, e93 (2008).

145. Greene, K.E., King, T.E., Jr., Kuroki, Y., Bucher-Bartelson, B., Hunninghake, G.W., Newman, L.S., Nagae, H. & Mason, R.J. Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *Eur Respir J* **19**, 439-446 (2002).
146. Jee, A.S., Sahhar, J., Youssef, P., Bleasel, J., Adelstein, S., Nguyen, M. & Corte, T.J. Review: Serum biomarkers in idiopathic pulmonary fibrosis and systemic sclerosis associated interstitial lung disease - frontiers and horizons. *Pharmacol Ther* (2019).
147. Bonhomme, O., Andre, B., Gester, F., de Seny, D., Moermans, C., Struman, I., Louis, R., Malaise, M. & Guiot, J. Biomarkers in systemic sclerosis-associated interstitial lung disease: review of the literature. *Rheumatology (Oxford)* (2019).
148. Jolliffe, I.T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci* **374**, 20150202 (2016).
149. Harlow, L., Gochuico, B.R., Rosas, I.O., Doyle, T.J., Osorio, J.C., Travers, T.S., Camacho, C.C., Oddis, C.V. & Ascherman, D.P. Anti-citrullinated heat shock protein 90 antibodies identified in bronchoalveolar lavage fluid are a marker of lung-specific immune responses. *Clin Immunol* **155**, 60-70 (2014).
150. Moore, M.W. & Herzog, E.L. Regulation and Relevance of Myofibroblast Responses in Idiopathic Pulmonary Fibrosis. *Curr Pathobiol Rep* **1**, 199-208 (2013).
151. Demedts, M., Behr, J., Buhl, R., Costabel, U., Dekhuijzen, R., Jansen, H.M., MacNee, W., Thomeer, M., Wallaert, B., Laurent, F., Nicholson, A.G., Verbeken, E.K., Verschakelen, J., Flower, C.D., Capron, F., Petruzzelli, S., De Vuyst, P.,

van den Bosch, J.M., Rodriguez-Becerra, E., Corvasce, G., Lankhorst, I., Sardina, M., Montanari, M. & Group, I.S. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* **353**, 2229-2242 (2005).

152. Idiopathic Pulmonary Fibrosis Clinical Research, N., Raghu, G., Anstrom, K.J., King, T.E., Jr., Lasky, J.A. & Martinez, F.J. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med* **366**, 1968-1977 (2012).

153. Network, I.P.F.C.R. Randomized trial of acetylcysteine in idiopathic pulmonary fibrosis. *New England Journal of Medicine* **370**, 2093-2101 (2014).

154. Adler-Milstein, J., DesRoches, C.M., Kralovec, P., Foster, G., Worzala, C., Charles, D., Searcy, T. & Jha, A.K. Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. *Health Aff (Millwood)* **34**, 2174-2180 (2015).

155. Liao, K.P., Cai, T., Savova, G.K., Murphy, S.N., Karlson, E.W., Ananthakrishnan, A.N., Gainer, V.S., Shaw, S.Y., Xia, Z., Szolovits, P., Churchill, S. & Kohane, I. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ* **350**, h1885 (2015).

156. Coorevits, P., Sundgren, M., Klein, G.O., Bahr, A., Claerhout, B., Daniel, C., Dugas, M., Dupont, D., Schmidt, A., Singleton, P., De Moor, G. & Kalra, D. Electronic health records: new opportunities for clinical research. *J Intern Med* **274**, 547-560 (2013).

157. Pfaff, E., Lee, A., Bradford, R., Pae, J., Potter, C., Blue, P., Knoepp, P., Thompson, K., Roumie, C.L., Crenshaw, D., Servis, R. & DeWalt, D.A. Recruiting for a pragmatic trial using the electronic health record and patient portal: successes and lessons learned. *J Am Med Inform Assoc* **26**, 44-49 (2019).

158. Gershon, A.S., Wang, C., Guan, J., Vasilevska-Ristovska, J., Cicutto, L. & To, T. Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD* **6**, 388-394 (2009).

159. Smidth, M., Sokolowski, I., Kaersvang, L. & Vedsted, P. Developing an algorithm to identify people with Chronic Obstructive Pulmonary Disease (COPD) using administrative data. *BMC Med Inform Decis Mak* **12**, 38 (2012).

160. Lee, T.M., Tu, K., Wing, L.L. & Gershon, A.S. Identifying individuals with physician-diagnosed chronic obstructive pulmonary disease in primary care electronic medical records: a retrospective chart abstraction study. *NPJ Prim Care Respir Med* **27**, 34 (2017).

161. Hadjinicolaou, A.V., Nisar, M.K., Parfrey, H., Chilvers, E.R. & Ostor, A.J. Non-infectious pulmonary toxicity of rituximab: a systematic review. *Rheumatology (Oxford)* **51**, 653-662 (2012).

162. Salliot, C. & van der Heijde, D. Long-term safety of methotrexate monotherapy in patients with rheumatoid arthritis: a systematic literature research. *Ann Rheum Dis* **68**, 1100-1104 (2009).

163. Conway, R., Low, C., Coughlan, R.J., O'Donnell, M.J. & Carey, J.J. Methotrexate and lung disease in rheumatoid arthritis: a meta-analysis of randomized controlled trials. *Arthritis Rheumatol* **66**, 803-812 (2014).

164. Choi, H.K., Hernan, M.A., Seeger, J.D., Robins, J.M. & Wolfe, F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *Lancet* **359**, 1173-1177 (2002).

165. Matteson, E.L., Bongartz, T., Ryu, J.H., Crowson, C.S., Hartman, T.E. & Dellaripa, P.F. Open-label, pilot study of the safety and clinical effects of rituximab in patients with rheumatoid arthritis-associated interstitial pneumonia. *Open Journal of Rheumatology and Autoimmune Diseases* **2**, 53 (2012).

166. Druce, K.L., Iqbal, K., Watson, K.D., Symmons, D.P.M., Hyrich, K.L. & Kelly, C. Mortality in patients with interstitial lung disease treated with rituximab or TNFi as a first biologic. *RMD Open* **3**, e000473 (2017).

167. Yoshida, K., Solomon, D.H. & Kim, S.C. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol* **11**, 437-441 (2015).

168. Austin, P.C. & Stuart, E.A. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* **34**, 3661-3679 (2015).

169. Raghu, G., Rochwerg, B., Zhang, Y., Garcia, C.A., Azuma, A., Behr, J., Brozek, J.L., Collard, H.R., Cunningham, W., Homma, S., Johkoh, T., Martinez, F.J., Myers, J., Protzko, S.L., Richeldi, L., Rind, D., Selman, M., Theodore, A., Wells, A.U., Hoogsteden, H., Schunemann, H.J., American Thoracic, S., European Respiratory, s., Japanese Respiratory, S. & Latin American Thoracic, A. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med* **192**, e3-19 (2015).

170. Nathan, S.D., Albera, C., Bradford, W.Z., Costabel, U., Glaspole, I., Glassberg, M.K., Kardatzke, D.R., Daigl, M., Kirchgaessler, K.U., Lancaster, L.H., Lederer, D.J., Pereira, C.A., Swigris, J.J., Valeyre, D. & Noble, P.W. Effect of pirfenidone on mortality: pooled analyses and meta-analyses of clinical trials in idiopathic pulmonary fibrosis. *Lancet Respir Med* **5**, 33-41 (2017).

171. Richeldi, L., Cottin, V., du Bois, R.M., Selman, M., Kimura, T., Bailes, Z., Schlenker-Herceg, R., Stowasser, S. & Brown, K.K. Nintedanib in patients with idiopathic pulmonary fibrosis: Combined evidence from the TOMORROW and INPULSIS((R)) trials. *Respir Med* **113**, 74-79 (2016).

172. Distler, O., Highland, K.B., Gahlemann, M., Azuma, A., Fischer, A., Mayes, M.D., Raghu, G., Sauter, W., Girard, M., Alves, M., Clerisme-Beaty, E., Stowasser, S., Tetzlaff, K., Kuwana, M., Maher, T.M. & Investigators, S.T. Nintedanib for Systemic Sclerosis-Associated Interstitial Lung Disease. *N Engl J Med* **380**, 2518-2528 (2019).

173. Barnes, H., Holland, A.E., Westall, G.P., Goh, N.S. & Glaspole, I.N. Cyclophosphamide for connective tissue disease-associated interstitial lung disease. *Cochrane Database Syst Rev* **1**, CD010908 (2018).

174. Fischer, A., Brown, K.K., Du Bois, R.M., Frankel, S.K., Cosgrove, G.P., Fernandez-Perez, E.R., Huie, T.J., Krishnamoorthy, M., Meehan, R.T., Olson, A.L., Solomon, J.J. & Swigris, J.J. Mycophenolate mofetil improves lung function in connective tissue disease-associated interstitial lung disease. *J Rheumatol* **40**, 640-646 (2013).

175. Oldham, J.M., Lee, C., Valenzi, E., Witt, L.J., Adegunsoye, A., Hsu, S., Chen, L., Montner, S., Chung, J.H., Noth, I., Vij, R. & Strek, M.E. Azathioprine response in patients with fibrotic connective tissue disease-associated interstitial lung disease. *Respir Med* **121**, 117-122 (2016).

176. Witt, L.J., Demchuk, C., Curran, J.J. & Strek, M.E. Benefit of adjunctive tacrolimus in connective tissue disease-interstitial lung disease. *Pulm Pharmacol Ther* **36**, 46-52 (2016).

177. Suissa, S., Hudson, M., Dell'Aniello, S., Shen, S., Simon, T.A. & Ernst, P. Comparative safety of abatacept in rheumatoid arthritis with COPD: A real-world population-based observational study. *Semin Arthritis Rheum* (2019).

178. Wilczynska, M.M., Condliffe, A.M. & McKeon, D.J. Coexistence of bronchiectasis and rheumatoid arthritis: revisited. *Respir Care* **58**, 694-701 (2013).

179. Singh, J.A., Cameron, C., Noorbaloochi, S., Cullis, T., Tucker, M., Christensen, R., Ghogomu, E.T., Coyle, D., Clifford, T., Tugwell, P. & Wells, G.A. Risk of serious infection in biological treatment of patients with rheumatoid arthritis: a systematic review and meta-analysis. *Lancet* **386**, 258-265 (2015).

180. England, B.R., Duryee, M.J., Roul, P., Mahajan, T.D., Singh, N., Poole, J.A., Ascherman, D.P., Caplan, L., Demoruelle, M.K., Deane, K.D., Klassen, L.W., Thiele, G.M. & Mikuls, T.R. Malondialdehyde-Acetaldehyde Adducts and Antibody Responses in Rheumatoid Arthritis-Associated Interstitial Lung Disease. *Arthritis Rheumatol* (2019).

181. Giles, J.T., Darrah, E., Danoff, S., Johnson, C., Andrade, F., Rosen, A. & Bathon, J.M. Association of cross-reactive antibodies targeting peptidyl-arginine deiminase 3 and 4 with rheumatoid arthritis-associated interstitial lung disease. *PLoS One* **9**, e98794 (2014).

APPENDICES

APPENDIX A: Diagnostic Codes for Lung Disease Categories

Diagnostic codes corresponding to HCUP-CCS categories for lung diseases

HCUP-CCS category	ICD-9-CM codes
Included	
127. Chronic obstructive pulmonary disease and bronchiectasis	490 4910 4911 4912 49120 49121 49122 4918 4919 4920 4928 494 4940 4941 496
128. Asthma	49300 49301 49302 49310 49311 49312 49320 49321 49322 49381 49382 49390 49391 49392
132. Lung disease due to external agents	4950 4951 4952 4953 4954 4955 4956 4957 4958 4959 500 501 502 503 504 505 5060 5061 5062 5063 5064 5069 5071 5078 5080 5081 5082 5088 5089
133. Other lower respiratory disease	5131 514 515 5160 5161 5162 5163 51630 51631 51632 51633 51634 51635 51636 51637 5164 5165 51661 51662 51663 51664 51669 5168 5169 5172 5178 5183 5184 51889 5194 5198 5199 7825 78600 78601 78602 78603 78604 78605 78606 78607 78609 7862 7863 78630 78631 78639 7864 78652 7866 7867 7868 7869 7931 79311 79319 7942 V126 V1260 V1261 V1269 V426
Not included	
122. Pneumonia	00322 0203 0204 0205 0212 0221 0310 0391 0521 0551 0730 0830 1124 1140 1144 1145 11505 11515 11595 1304 1363 4800 4801 4802 4803 4808 4809 481 4820 4821 4822 4823 48230 48231

	48232 48239 4824 48240 48241 48242 48249 4828 48281 48282 48283 48284 48289 4829 483 4830 4831 4838 4841 4843 4845 4846 4847 4848 485 486 5130 5171
123. Influenza	4870 4871 4878 488 4880 48801 48802 48809 4881 48811 48812 48819 48881 48882 48889
125. Acute bronchitis	4660 4661 46611 46619
126. Other upper respiratory infections	0320 0321 0322 0323 0340 460 4610 4611 4612 4613 4618 4619 462 4640 46400 46401 46410 46411 46420 46421 46430 46431 4644 46450 46451 4650 4658 4659 4730 4731 4732 4733 4738 4739 78491
129. Aspiration pneumonitis; food/vomitus	5070
130. Pleurisy; pneumothorax; pulmonary collapse	5100 5109 5110 5111 5118 51189 5119 5120 5128 51281 51282 51283 51284 51289 5180 5181 5182
131. Respiratory failure; insufficiency; arrest	5173 5185 51851 51852 51853 51881 51882 51883 51884 7991 V461 V4611 V4612 V4613 V4614 V462

HCUP-CCS: Healthcare Cost and Utilization Project Clinical Classification Software: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp
ICD-9-CM: International Classification of Disease, 9th edition, Clinical Modification

APPENDIX B: Diagnostic Codes for Cardiovascular Categories

Diagnostic codes corresponding to HCUP-CCS categories for cardiovascular disease

HCUP-CCS category	ICD-9-CM codes
96. Heart valve disorders	3940 3941 3942 3949 3950 3951 3952 3959 3960 3961 3962 3963 3968 3969 3970 3971 3979 4240 4241 4242 4243 42490 42491 42499 7852 7853 V422 V433
97. Peri-, endo-, and myocarditis; cardiomyopathy	03282 03640 03641 03642 03643 07420 07421 07422 07423 11281 11503 11504 11513 11514 11593 11594 1303 3910 3911 3912 3918 3919 3920 393 3980 39890 39899 4200 42090 42091 42099 4210 4211 4219 4220 42290 42291 42292 42293 42299 4230 4231 4232 4233 4238 4239 4250 4251 42511 42518 4252 4253 4254 4257 4258 4259 4290
100. Acute myocardial infarction	4100 41000 41001 41002 4101 41010 41011 41012 4102 41020 41021 41022 4103 41030 41031 41032 4104 41040 41041 41042 4105 41050 41051 41052 4106 41060 41061 41062 4107 41070 41071 41072 4108 41080 41081 41082 4109 41090 41091 41092
101. Coronary atherosclerosis and other heart disease	4110 4111 4118 41181 41189 412 4130 4131 4139 4140 41400 41401 41406 4142 4143 4144 4148 4149 V4581 V4582
105. Conduction disorders	4260 42610 42611 42612 42613 4262 4263 4264 42650 42651 42652 42653 42654 4266 4267 42681 42682 42689

	4269 V450 V4500 V4501 V4502 V4509 V533 V5331 V5332 V5339
106. Cardiac dysrhythmias	4270 4271 4272 42731 42732 42760 42761 42769 42781 42789 4279 7850 7851
107. Cardiac arrest and ventricular fibrillation	42741 42742 4275
108. Congestive heart failure, non-hypertensive	39891 4280 4281 42820 42821 42822 42823 42830 42831 42832 42833 42840 42841 42842 42843 4289
109. Acute cerebrovascular disease	34660 34661 34662 34663 430 431 4320 4321 4329 43301 43311 43321 43331 43381 43391 4340 43400 43401 4341 43410 43411 4349 43490 43491 436
110. Occlusion or stenosis of precerebral arteries	4330 43300 4331 43310 4332 43320 4333 43330 4338 43380 4339 43390
112. Transient cerebral ischemia	4350 4351 4352 4353 4358 4359
113. Late effects of cerebrovascular disease	438 4380 43810 43811 43812 43813 43814 43819 43820 43821 43822 43830 43831 43832 43840 43841 43842 43850 43851 43852 43853 4386 4387 43881 43882 43883 43884 43885 43889 4389
114. Peripheral and visceral atherosclerosis	4400 4401 4402 44020 44021 44022 44023 44029 4404 4408 4409 4439 5570 5571 5579

HCUP-CCS: Healthcare Cost and Utilization Project Clinical Classification Software (HCUP-CCS): https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp
ICD-9-CM: International Classification of Diseases, 9th edition, Clinical Modification

APPENDIX C: Descriptions of "Other" Lung Codes

Descriptions of ICD-9-CM codes included in other lower respiratory disease (HCUP-CCS category 133)

ICD-9-CM	Description
513.1	Abscess of the mediastinum
514	Pulmonary congestion
515	Post-inflammatory pulmonary fibrosis
516.0	Other alveolar and parietoalveolar pneumonopathy
516.1	Idiopathic pulmonary hemosiderosis
516.2	Pulmonary alveolar microlithiasis
516.3x	Idiopathic interstitial pneumonia
516.4	Lymphangioleiomyomatosis
516.5	Adult pulmonary Langerhans cell histiocytosis
516.6	Interstitial lung diseases of childhood
516.8	Other specified alveolar and parietoalveolar pneumonopathies
516.9	Unspecified alveolar and parietoalveolar pneumonopathy
517.2	Lung involvement in systemic sclerosis
517.8	Lung involvement in other diseases classified elsewhere
518.3	Pulmonary eosinophilia
518.4	Acute edema of lung, unspecified
518.89	Other diseases of lung, not elsewhere classified
519.4	Disorders of the diaphragm
519.8	Other diseases of respiratory system, not elsewhere classified
519.9	Unspecified disease of respiratory system
782.5	Cyanosis
786	Symptoms involving respiratory system and other chest symptoms
793.1	Nonspecific findings on radiological and other examination of lung field
794.2	Nonspecific abnormal results of pulmonary function study
V12.6	Personal history of diseases of respiratory system
V42.6	Lung transplant status

HCUP-CCS: Healthcare Cost and Utilization Project Clinical Classification Software
ICD-9-CM: International Classification of Diseases, 9th edition, Clinical Modification

APPENDIX D: Diagnostic Codes for Interstitial Lung Disease

Diagnostic codes used for interstitial lung disease algorithms

Codes	Description
ICD-9	
515	Postinflammatory pulmonary fibrosis
516.3x (516.30-516.37)	Idiopathic interstitial pneumonia
516.8	Other specified alveolar and parietoalveolar pneumonopathies
516.9	Unspecified alveolar and parietoalveolar pneumonopathy
714.81	Rheumatoid lung
ICD-10	
M05.1x (M05.10-M05.19)	Rheumatoid lung disease with rheumatoid arthritis
J84.1x (J84.10-J84.17)	Other interstitial pulmonary diseases with fibrosis
J84.2	Lymphoid interstitial pneumonia
J84.89	Other specified interstitial pulmonary diseases
J84.9	Interstitial pulmonary disease, unspecified
J99	Respiratory disorders in diseases classified elsewhere

Abbreviations: ICD, International Classification of Diseases
Bolded are recommended for future use based on study results.

APPENDIX E: Procedure Codes for Interstitial Lung Disease

Procedure codes used in interstitial lung disease algorithms.

Procedure	CPT	ICD-9-CM Procedure	ICD-10-PCS[a]
Lung biopsy			
Surgical	32095-32097, 32602, 32607-32608	33.20, 33.28, 34.21	0BB30*X-0BB90*X, 0BBB0*X-0BBM0*X
Transbronchial	31628, 31629, 31632	33.27	0BB38*X-0BB98*X 0BBB8*X-0BBM8*X
Percutaneous	32405	33.26	0BB33*X-0BB93*X 0BBB3*X-0BBM3*X
Chest computed tomography (CT)			
Chest CT	71250, 71260, 71270	87.41	BB24***
Low dose CT chest	G0297		
CT-angiogram	71275		
Pulmonary function tests[b]			
Spirometry	94010, 94060, 94070, 94150, 94200, 94375	89.37, 89.38	4A09***
Lung volume	94250, 94726-94727		
Diffusion capacity	94729		

[a] * Denotes any code in this position

[b] PFTs additionally identified through pulmonary function test lab stop code (104)

Abbreviations: CPT, current procedural terminology; ICD-9-CM, International Classification of Diseases, 9th revision, Clinical Modification; ICD-10-PCS, International Classification of Diseases, 10th revision, Procedure Coding System

APPENDIX F: Diagnostic Codes for Other Interstitial Lung Disease

Diagnostic codes for other causes of interstitial lung disease.

Condition	ICD-9-CM	ICD-10-CM
Sarcoidosis	135.x	D86.x
Systemic sclerosis	517.2, 710.1	M34.x
Myositis	710.3-710.4	M33.x
Systemic lupus erythematosus	710.0	M32.x
Hypersensitivity pneumonitis	495.x	J67.x
Pneumoconioses (including asbestos)	500.x-505.x	J60.x-J64.x
Radiation pneumonitis	508.1	J70.0-J70.1

Abbreviations: ICD, International Classification of Diseases; CM, Clinical Modification