Evaluating smoking, chronic obstructive pulmonary disease, and race/ethnicity and interstitial lung disease in rheumatoid arthritis

Boyang Zheng

Faculty of Medicine
Department of Epidemiology, Biostatistics and Occupational Health
Montreal, Quebec, Canada
October 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science (Epidemiology)

TABLE OF CONTENTS

ABSTRACT

Background

Rheumatoid arthritis (RA) is associated with various lung diseases. The development of interstitial lung disease (ILD) is a significant cause of mortality in RA. Chronic obstructive pulmonary disease (COPD), a smoking-related disease, is also prevalent in RA and it is unclear if one lung disease is associated with the other. Given that smoking is also a potential determinant of RA-ILD, our first goal was to determine the incidence of RA-ILD and examine whether smoking and COPD is associated with RA-ILD onset using administrative health data from the United States. Another unclear determinant of RA-ILD is race/ethnicity, which is associated with other types of ILD. Our second goal was to determine whether race/ethnicity is associated with RA-ILD.

Methods

New-onset ILD in incident adult RA subjects were identified in the MarketScan Commercial Claims database (2010-2018) based on physician and/or hospitalization diagnostic claims. Smoking (current, past, and never) was available in a subset of subjects with available health risk assessment questionnaires. A second MarketScan Medicaid database (2011-2018, non-linkable to the Commercial Claims database), was used to study race/ethnicity (white, African-American, Hispanic, or other) and RA-ILD development. Kaplan-Meier analyses were performed to assess time to ILD onset stratified by co-existing COPD, smoking, and race/ethnicity (in the respective databases). Multivariate Cox regression models adjusted for age, sex, and immunosuppressant use prior to RA diagnosis.

Results

Among 373,940 new RA subjects in the Commercial Claims cohort, 1.7% developed ILD with an incidence rate of 8.1 (95% CI 7.9, 8.3) per 1000 person-year. After adjusting for age, sex, and immunosuppressants, the hazard ratio (HR) was 2.10 (95% CI 1.89, 2.32) for ILD onset in those with COPD versus those without. In the subset of subjects with smoking information, there was no clear association of ILD with smoking status and a non-statistically significant association (HR 1.64, 95% CI 0.93, 2.89) for an association between ILD onset and COPD. In the Medicaid cohort, 1016 (2.2%) subjects developed ILD, incidence rate of 6.3 (95% CI 5.6, 7.0) per 1000 person-year. Hispanic subjects had a higher risk of RA-ILD (HR 1.51, 95% CI 1.07, 2.14), adjusting for age, sex, baseline COPD, and previous immunosuppressants.

Conclusions

Co-existing COPD in incident RA subjects was associated with a higher risk of ILD but association between smoking and RA-ILD onset remains unclear. Further studies are needed to better evaluate the complex relationships between cumulative smoking exposure, COPD and RA-ILD. Finally, Hispanic subjects had an increased likelihood of developing RA-ILD than other race/ethnicity groups.

RÉSUMÉ

Contexte

La polyarthrite rhumatoïde (PAR) est associée au développement d'une maladie pulmonaire interstitielle (MPI) qui entraîne une mortalité importante. La maladie pulmonaire obstructive chronique (MPOC), une maladie liée au tabagisme, est une autre maladie pulmonaire prévalente dans la PAR. Bien que le tabagisme soit également un facteur de risque pour la PAR-MPI, nous ne savons si ces deux maladies pulmonaires sont associées une avec l'autre. Notre premier objectif était de déterminer l'incidence de la PAR-MPI et d'examiner si le tabagisme et la MPOC seraient associés à l'apparition de la PAR-MPI en utilisant des données sur la santé provenant des États-Unis. La race/ethnicité étant associée à d'autres types de MPI, est potentiellement un autre déterminant de la PAR-MPI. Notre deuxième objectif était de déterminer si la race/ethnicité était associée à la PAR-MPI.

Méthodes

Le développement de MPI chez les adultes nouvellement atteints par la PAR a été identifié dans la base de données «MarketScan Commercial Claims» (2010-2018), qui contient les codes de diagnostic du médecin et/ou lors d'une hospitalisation. Le tabagisme (actuel/passé/jamais) était disponible dans un sous-groupe de sujets. Une deuxième base de données «MarketScan Medicaid » (2011-2018), non liée aux «Commercial Claims») a été utilisée pour étudier la race/ethnicité (Blanche, Afro-américaine, Hispanique ou autre) et le développement de la PAR-MPI. Des analyses de Kaplan-Meier ont été effectuées pour évaluer le délai d'apparition de la MPI stratifiée en fonction de la présence de la MPOC, du tabagisme et de la race/ethnicité. Des modèles de régression de Cox ont contrôlés pour l'âge, le sexe et l'utilisation d'immunosuppresseurs avant le diagnostic de PAR.

Résultats

Parmi 373 940 sujets avec la PAR dans la cohorte «Commercial Claims», 1.7% ont développé une MPI, taux d'incidence de 8.1 [intervalle de confiance (IC) à 95% 7.9-8.3] par 1000 personnes-années. Après l'ajustement en fonction de l'âge, du sexe et de l'utilisation d'immunosuppresseurs, le rapport de risque (HR) était de 2.10 (IC 95% 1.89-2.32) pour l'apparition d'une MPI chez les personnes avec MPOC par rapport à celles qui n'en avaient pas. Dans le sous-groupe de sujets ayant des informations sur le tabagisme, il n'y avait pas d'association claire entre la MPI et le tabagisme. Il y avait une tendance non statistiquement significative (HR 1.64, IC 95% 0.93-2.89) entre l'apparition de la MPI et la MPOC. Dans la cohorte «Medicaid», 1016 (2.2%) sujets ont développé une MPI, taux d'incidence de 6.3 (IC 95% 5.6, 7.0)

par 1000 années-personnes. Les sujets hispaniques avaient un risque plus élevé de PAR-MPI (HR 1.51, IC95% 1.07- 2.14), ajusté pour l'âge, le sexe, la MPOC, et l'utilisation antérieure d'immunosuppresseurs.

Conclusion

La présence d'une MPOC chez les sujets atteints de PAR est associée à un risque plus élevé de MPI, mais l'association entre le tabagisme et l'apparition de MPI reste incertaine. D'autres études seront nécessaires pour mieux évaluer les relations complexes entre le tabagisme, la MPOC et la PAR-MPI. Finalement, les sujets hispaniques ont une probabilité plus élevée de développer une PAR-MPI que les autres groupes raciaux / ethniques.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr Sasha Bernatsky for her continuous guidance in helping me throughout my research project and thesis writing. I would like to thank my co-supervisor Dr Evelyne Vinet for her encouragement and insightful comments.

I would like to thank Cristiano Soares de Moura, whose support was invaluable, for his continued mentoring and patience with the computing-intensive analyses. I would like to also thank the other manuscript co-authors: Marina Amaral De Avila Machado, Dr Jeffrey Curtis, and Dr Cristian A. Pineau for their insight.

Finally, I would like to thank all members of the rheumatology division at the Montreal General Hospital who have all been so supportive throughout my training and this first part of my career.

PREFACE AND CONTRIBUTION OF AUTHORS

Cristiano Soares de Moura performed MarketScan data extraction and helped with data analyses. Marina Machado, Christian A. Pineau, Jeffery R. Curtis, and Evelyne Vinet provided insights and suggestions and reviewed the manuscript. Sasha Bernatsky, as the principal investigator, oversaw all aspects of the study. I conceptualized the project, helped analyze the data, interpreted the results, and wrote the manuscript and thesis, with the help of Sasha Bernatsky and input from each co-author member.

GENERAL OVERVIEW

My thesis consists of an introduction where I discuss rheumatoid arthritis and related lung diseases to highlight what is known and what are some key knowledge gaps. I end the introduction by stating my objectives for this thesis. This will be followed by a presentation of the methodology used for the studies. Subsequent chapters will present the results of my analyses. The results of the first study will be presented in the format of a manuscript submitted for peer-reviewed journal publication. The results of the second study will be presented afterwards. Finally, these will be followed by a discussion and summary of my overall findings.

CHAPTER 1: INTRODUCTION

1.1 Rheumatoid arthritis

Rheumatoid arthritis (RA) is a chronic autoimmune disease that affects 1% of the North American population (1). This is the most frequent rheumatic disease and the hallmark is polyarthritis (inflammation involving multiple joints). While classification criteria exist to identify patients for clinical trials, these are not diagnostic (2). The diagnosis is made based on clinical judgment by experienced physicians depending on the pattern and duration of joint disease, the presence of other extra-articular features, and the exclusion of possible mimickers. There is no blood test to definitively diagnose RA, although two RA-related autoantibodies, rheumatoid factor (RF) and anti-citrullinated peptides antibodies (ACPA), can be found in 55%–65% of patients (3).

Rheumatoid arthritis requires treatment with immunosuppression in order to prevent joint destruction and disability (4). The first line is a class of agents called disease modifying anti-rheumatic drugs (DMARDs) that aim to maintain a state of disease remission. A poor response requires further immunosuppression using biologic agents that target various specific inflammatory pathway molecules. Corticosteroids are short acting drugs that suppress inflammation and are generally used to treat acute flares. These are sometimes used in low doses to control persistently active disease.

Risk factors for RA include female sex, smoking, family history of RA, older age, and other possible environmental triggers (5, 6). Genetic factors, particularly certain human leukocyte antigen genotypes, confer an increased risk of RA(7). Although the exact pathogenesis of RA remains unclear, a key step is the formation of ACPAs due to increased protein citrullination(8). This leads to immune system abnormalities including dysregulated B and T lymphocytes, increased autoantibody production, activation of innate immune cells, and a state of chronic inflammation(9). Susceptible individuals progress through preclinical stages of local and systemic inflammatory activity prior to the onset of symptoms. There is mounting knowledge that the lungs are an important site of inflammation even early in the course of disease.

1.2 Rheumatoid arthritis pathogenesis and the lungs

Although RA primarily affects the joints, it can affect other organs. There is increasing evidence that RA pathogenesis may be initiated in the lungs, leading to the production of autoantibodies and autoimmunity (10). In fact, lung disease may precede clinical RA onset in subjects with high RF or ACPA titres (11). Protein citrullination occurs in lung alveoli and bronchial mucosa (12, 13) and smoking is a

significant contributor to this process by increasing mucosal inflammation (14). Smokers have a 20-fold higher risk of developing RA compared to never smokers, and tobacco exposure is strongly associated with the presence of RA-related antibodies (15). Other inhalational exposures such as mining dust and air pollution have also been shown to increase RA onset(7) or antibody positivity(16).

1.3 Rheumatoid arthritis associated interstitial lung disease

Lung-related complications are a leading cause of mortality behind cardiovascular disease in RA(17). The most commonly associated lung disease in RA is interstitial lung disease (ILD). ILD is the result of chronic inflammation in distal airspaces and lung parenchyma that leads to fibrosis and impaired oxygen exchange. RA-ILD is diagnosed on imaging, typically with high resolution CT scan (HRCT) and by the exclusion of certain other causes. 20-50% of RA patients have subclinical ILD, defined as interstitial abnormalities affecting ≥5% of any lung region on HRCT (18). Importantly, close to 50% of patients with subclinical ILD demonstrate radiographic progression over 1-3 years(19). A smaller number of RA patients develop a clinically significant RA-ILD causing shortness of breath and other symptoms, which would bring this complication to medical attention. Clinically significant ILD has also been defined as ILD found on follow up without systematic HRCT screening, and is often attributed to cases of RA-ILD identified by administrative databases. This is a serious condition, associated with a 50% mortality rate within 5 years of detection (20). Two well-characterized North American RA cohorts show the 30-year cumulative incidence of developing a clinically significant RA-ILD to be around 6-8% (21, 22) and other worldwide studies report a similar prevalence of 2-10% in the RA population (23-27). The only other study examining RA-ILD incidence using MarketScan found a yearly incidence rate of RA-ILD between 2.7 (95% CI 2.5–2.9) to 3.8 (95% CI 3.5–4.0) per 100,000 people in the *general* population (not limited to RA subjects) between 2003-2014. This study did not examine RA-ILD incidence among subjects with RA (28).

1.4 Risk factors for RA-ILD

Important risk factors for ILD in RA include older age and male sex. These have been consistently associated with the presence of clinically evident ILD in multiple cohorts (29). While RA predominantly affects women, men have a worse prognosis and increased extra-articular involvement(30). The reason for this sex difference is unclear and could be due to a combination of differential genetic and environmental exposures such as smoking. Another major risk factor for RA-ILD is the presence of ACPA and RF. Seropositive disease was associated with 2-3 times the odds of RA-ILD onset (31, 32). Patients with higher chronic RA disease activity also appear to have a higher risk of ILD compared to patients in

remission (33). Recent evidence suggests that genetic features similar to those in other types of ILD may also contribute to RA-ILD (34).

1.5 Smoking and RA-ILD

Smoking is regarded as a preventable risk factor for RA and potentially for RA-ILD, although findings regarding the association between smoking and RA-ILD have been inconsistent. Smoking is a risk factor in the idiopathic form of ILD called idiopathic pulmonary fibrosis (IPF), to which RA-ILD often resembles(35). The degree of cumulative smoking exposure in RA was associated with various lung imaging abnormalities, including subclinical ILD (36). Another frequently cited study showed that >25 pack years of smoking was associated with an OR=3.76 (95% CI 1.59, 8.88) of lung abnormalities compared to non-smoking (37). However the lung abnormalities in this study were not specific for ILD and were defined by a composite measure that included pulmonary function test findings, chest radiograph changes, and symptoms. This composite measure included changes found in other respiratory diseases such as COPD. Another study found that current and/or past smoking was associated with RA-ILD but this association was no longer present after adjusting for RA-related autoantibodies(32). In contrast, several studies have not found clear associations between smoking and RA-ILD prevalence (20, 38-41), RA-ILD morphology(42), or RA-ILD mortality (20, 43, 44). Some of the negative findings may partly have been related to an insufficient power to document the association between smoking and RA-ILD. Differences in study findings may also be because some studies looked at nonspecific lung outcomes and/or subclinical ILD rather than clinically evident ILD.

1.6 Rheumatoid arthritis and chronic obstructive pulmonary disease (COPD)

COPD, also colloquially called smoker's lung, is the result of airflow obstruction due to chronic inflammation in the airways (called chronic bronchitis) and/or the alveoli with alveolar wall destruction (called emphysema). The main exposure linked with COPD is smoking and about 15-20% of smokers develop COPD, in a dose dependent manner (45, 46). COPD is characterized by lung inflammation that can persist even after smoking cessation and key inflammatory cells involved (macrophages, neutrophils and T cells) are also players in RA pathogenesis(47). Importantly, RA patients have close to a 2-fold higher risk of COPD than in the general population (48). It is present in 3-10% of RA patients and occurs even at relatively low smoking exposures (48, 49). The reverse association has also been found, where RA incidence was increased among those with COPD, even after adjusting for cumulative smoking (50). Besides ILD and COPD, other inflammatory airway changes such as bronchiectasis have also been found with increased prevalence in RA (51) and it is not clear why.

Emphysematous changes seen in COPD are also found in association with ILD (52) and both share similar risk factors such as age and smoking. New evidence also shows genetic commonalities between emphysema, ILD, and RA-ILD in particular, such as the presence of shortened telomeres in alveolar fibroblasts and other lung tissues and somatic cells (53, 54). However, there is a knowledge gap regarding whether an association exists in RA between one type of inflammatory lung disease such as COPD, and another, particularly ILD.

1.7 Race/Ethnicity and RA-ILD

There is some evidence that race/ethnicity might influence the development of ILD. For example in IPF, which morphologically resembles RA-ILD, lower incidence rates have been reported in East Asian countries (ranging from 1.2 to 4.16 per 100 000 person year) compared to European and North American cohorts where incidence tended to be higher at 2.8 to 19 per 100 000 person year (55).

In other autoimmune rheumatic diseases besides RA, black (African American) race has been identified as a risk factor for ILD development in scleroderma (56). Another US cohort examining the prevalence of all-cause ILD found that blacks and Hispanics with autoimmune rheumatic diseases were more likely to have ILD compared to white patients. Black patients also had more ILD associated with isolated autoimmune disease features and a better overall survival compared to others (57).

Because each rheumatic disease is very heterogeneous, extrapolation from one to another is difficult. There have been no comparable studies of ethnic/racial differences related specifically to RA-ILD development. Many RA cohorts from Europe and North America are predominantly white (50, 58, 59). While RA-ILD prevalence has been studied in clinical cohorts worldwide, it is difficult to compare prevalence rates between countries. In countries such as Japan (32), South Korea(60), and Turkey(61), the clinically significant RA-ILD prevalence seems to be in the same range (2-10%) as European and North American cohorts (23-27). This comparison is limited by the wide range of RA-ILD prevalence rates due to different disease definitions, different detection methods, and different RA inclusion criteria. Understanding the impact of race/ethnicity on RA-ILD onset may help better personalize RA care and also provide possible insights into underlying disease mechanisms.

1.8 OBJECTIVES

The goal was to identify associations for ILD onset in incident RA using large population based data. To address particular knowledge gaps, the specific objectives were to examine:

1) The association between smoking and RA-ILD within a large RA population sample
2) The association between COPD, another frequent lung complication in RA, with RA-ILD
3) The association between race/ethnicity and RA-ILD

While discordant findings regarding smoking exist in the literature, we hypothesized that smokers would be more likely to develop RA-ILD than non-smokers. In the absence of past literature examining the association between COPD and ILD in RA, we hypothesized that subjects with existing RA-COPD would be more likely to develop RA-ILD than those without COPD. Finally, we hypothesized that, similar to what has been found with ILD risk in other autoimmune rheumatic diseases, RA-ILD risk may be associated with race/ethnicity: specifically, non-white race/ethnicity may be associated with a higher RA-ILD risk.

CHAPTER 2: METHODS

2.1 Data sources

We performed cohort analyses using three different databases owned by IBM. The first was the MarketScan Commercial Claims database (2010-2018) consisting of health plan data on millions of individuals in the United States (US) who are covered by employer sponsored health insurance. This data covers all geographical regions of the US. It contains basic demographics (age, sex, employment, and location), as well as all physician billing diagnostic and procedure claims, and hospitalization diagnostic codes. Physicians must submit a principal diagnosis code after a patient encounter to receive payment. Hospitals tabulate the principal diagnosis for each admission as well as all diagnoses relevant to the hospitalization.

Linked to this is a second database, the Health Risk Assessment (HRA) database (2010-2018). This can be considered a sub-cohort of employees enrolled in the main MarketScan Commercial Claims database who have available self-reported health risk assessment questionnaires from their health insurance plans. This provides additional data including health risk behaviors such as smoking.

The third database used was the Multi-State Medicaid database (2011-2018). This contains the medical and prescription drug data of millions of individuals enrolled in Medicaid and covers all US regions. Medicaid enrollees in the United States must meet certain criterion (variable across States) which includes having a low income threshold, having certain permanent disabilities, and/or being above 65 years of age. This includes subjects covered by Medicare who may be dually covered. In addition to the basic demographic variables, this database captures information on race/ethnicity. The Medicaid database, however, is not linkable to the above MarketScan databases, and may contain overlapping subjects. Thus, the first two sets of data (MarketScan Commercial Claims and HRA) were studied separately from the Medicaid data. The MarketScan Commercial Claims and HRA databases were used to evaluate associations between COPD, smoking and RA-ILD. The Medicaid data was used to evaluate the association between race/ethnicity and RA-ILD.

This study has been approved by the McGill University institutional review board (approval #A04-M47-12B).

2.2 Subject selection

Adult subjects > 18 years of age with new-onset RA were identified based on the presence of at least 2 claims by any physician or a hospitalization claim with a principal diagnosis of RA using International Classification of Diseases-9-Clinical Modification (ICD-9-CM) codes 714 and ICD-10-CM codes M05, M06.

We required subjects to have had at least 12 months of continuous enrolment in the coverage plan without any prior RA diagnosis to ensure that they were incident RA cases. The date of the first RA diagnosis was considered to be the index date.

2.3 Variable definitions

In all three cohorts, ILD was identified based on the presence of at least 2 physician claims at least 1 month apart. This included claim codes for fibrosis: ICD-9-CM 515, 516.3 and ICD-10-CM J84.1; and/or codes for rheumatic lung disease: ICD-9-CM 714.81 and ICD-10-CM M05.1. Pre-existing ILD before the first RA diagnostic code was excluded based on a one-year look-back period prior to index date. Co-existing COPD was identified based on previous algorithms(62, 63) using the presence of at least 1 claim using ICD-9 CM codes 491, 492, and 496 or ICD-10-CM codes J41-J44 assessed up to one year before RA onset. The full list of the ICD codes used and their descriptions are shown in Table 1.

In the HRA sub-cohort containing smoking information, smoking was classified as “currently smoking”, “previously smoked”, or never smoker. Information on race/ethnicity was available in the Medicaid cohort as “White”, “Black”, “Hispanic” or “Other”.

Other covariates available in all databases were sex and age at index date.

Table 1: List of the selected ICD coding and their descriptors

Rheumatoid arthritis		
ICD-9-CM	714	Rheumatoid arthritis and other inflammatory polyarthropathies
ICD-10-CM	M05	Rheumatoid arthritis with rheumatoid factor
	M06	Other rheumatoid arthritis
Interstitial lung disease		
ICD-9-CM	515	Post inflammatory pulmonary fibrosis
	516.3	idiopathic interstitial pneumonia
	714.81	Rheumatoid lung disease
ICD-10-CM	J84.1	Other interstitial pulmonary diseases with fibrosis
	M05.1	Rheumatoid lung disease
Chronic obstructive pulmonary disease		
ICD-9-CM	491	Chronic bronchitis
	492	Emphysema
	496	Chronic airway obstruction, not elsewhere classified
ICD-10-CM	J41	Simple and mucopurulent chronic bronchitis
	J42	Unspecified chronic bronchitis
	J43	Emphysema
	J44	Other chronic obstructive pulmonary disease

2.4 Statistical analysis

Time to ILD onset after RA diagnosis was estimated using Kaplan-Meier curves and compared between those with and without coexisting COPD, as well as by smoking status in the HRA sub-cohort, and by race/ethnicity in the Medicaid cohort. The Cramer's V correlation coefficient was assessed between smoking and COPD in the HRA sub cohort. Adjusted hazard ratios (HR) with 95% confidence intervals (CI) for ILD onset were obtained using Cox proportional hazards models in all cohorts. In addition to COPD, smoking, and race/ethnicity (where available in the respective cohorts), the models were also adjusted for age at index date and sex. Proportional hazards assumption was verified using Schoenfeld residuals.

In the HRA sub-cohort, subjects with missing values for smoking were excluded from those particular analyses.

As previously stated, the Medicaid cohort (with race/ethnicity information) is not linked to the Commercial Claims data. Because of this, the MarketScan Commercial Claims database and the HRA sub-cohort were studied separately from the Medicaid cohort and the findings will be presented separately.

In the next chapter, the assessment of the association between COPD, smoking and RA-ILD performed in the Commercial Claims and HRA sub-cohort will be presented in the same format as the manuscript submitted in a peer reviewed journal. The chapter immediately following will present the results of the analyses performed in the Medicaid cohort where the primary objective was to assess the association between race/ethnicity and RA-ILD. Those analyses are being prepared for a separate journal article.

CHAPTER 3: RESULTS OF ANALYSES IN THE MARKETSCAN AND HRA SUB-COHORT ON SMOKING, COPD AND RA-ILD

The following text details the analyses and results of the first study on the association between smoking, COPD and RA-ILD that was performed in the full MarketScan Commercial Claims cohort (henceforth referred to simply as the full MarketScan cohort in this chapter) and the HRA sub-cohort. This is presented in the format of the manuscript submitted to Arthritis Care and Research on date October 7, although the numeration of the tables have been modified to follow the thesis order.

Association between chronic obstructive pulmonary disease, smoking, and interstitial lung disease onset in rheumatoid arthritis

Boyang Zheng[12], Cristiano Soares de Moura[2], Marina Machado[2], Christian A. Pineau[1], Jeffery R. Curtis[3], Evelyne Vinet[12], Sasha Bernatsky[12]

Author affiliations:

1 Division of Rheumatology, Mcgill University

2 Department of Epidemiology, Biostatistics and Occupational Health, McGill University

3 Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham

Abstract

Background

In rheumatoid arthritis (RA), respiratory manifestations include chronic obstructive pulmonary disease (COPD) and interstitial lung disease (ILD). We assessed whether baseline COPD and smoking were associated with RA-ILD onset.

Methods

We identified new-onset ILD in incident RA subjects within the MarketScan Commercial Claims database, using physician and/or hospitalization diagnostic codes. Smoking data (current, past, never) were available for a subset via a health questionnaire. Kaplan Meier analyses assessed time to ILD onset, stratified by prior COPD and smoking. Multivariate Cox regression models were adjusted for age, sex, and (in the subset) smoking.

Results

Among 373,940 new RA subjects, 6343 (1.7%) developed ILD (8.1 events per 1000 person-year, 95% CI 7.9, 8.3). ILD was more common among subjects with baseline COPD. Adjusting for age and sex, the hazard ratio (HR) between baseline COPD and incident ILD was 2.15, 95% CI 1.93, 2.34. We could not establish a clear relationship between current smoking and ILD; in the subset with smoking data, the HR point estimate for COPD was similar but the 95% CI was wider (due to fewer subjects) and included the null value.

Conclusions

Pre-existing COPD in incident RA subjects was associated with higher risk of future ILD. While a possible association persisted after adjusting for smoking, we were limited by reduced sample size. Further studies are needed to assess the potentially complicated relationships between smoking, COPD, and other factors in RA-associated ILD.

Key words: Rheumatoid arthritis, interstitial lung disease, smoking, risk factors, chronic obstructive pulmonary disease

Key messages:

1) In our large population-based sample, ILD incidence in RA was 8.1/1000 person-year (95% CI 7.9, 8.3)

2) RA patients with pre-existing COPD had a higher risk of developing ILD than those without COPD

3) Further studies are needed to better assess smoking and underlying mechanisms linking COPD and RA-ILD

Background

Rheumatoid arthritis (RA) is the most frequent autoimmune rheumatic disease, affecting 1% of the adult population (64). The main feature is joint inflammation, but co-occurring lung disease represents a significant cause of morbidity and mortality. It is estimated that at least 5% of RA patients develop clinically significant ILD over the disease course, which is associated with a 50% five-year mortality rate (20). The risk of RA-associated ILD (i.e. RA-ILD) is increased in the presence of RA-related autoantibodies such as rheumatoid factor (RF) and anti-citrullinated peptide antibodies (ACPA). Other identified risk factors for RA-ILD are older age, male sex, and chronically high disease activity. Smoking, which is also associated with increased ACPA titres, has been implicated as another risk factor for RA-ILD. However, the association between smoking and clinically significant RA-ILD has been inconsistently identified in various observational cohorts (20, 38).

Smoking is also a key cause of chronic obstructive pulmonary disease (COPD), an inflammatory, destructive airway disease that can manifest to various degrees as bronchitis and/or emphysema. However, RA patients are known to have an increased risk of COPD even with low or minimal cumulative smoking exposure (48). Both ILD and COPD have an inflammatory component, although affecting different areas of the lung. Given the increased susceptibility to COPD in RA, it is unknown whether the presence of COPD associated with RA (i.e. RA-COPD) would represent an additional risk factor for ILD, or if RA patients with lung involvements might share an underlying susceptibility beyond smoking alone.

Our goal was to examine COPD and smoking as risk factors for RA-ILD in a large population-based dataset by using administrative billing data. We hypothesized that both COPD and smoking would be independently associated with ILD onset in RA patients.

Methods

Data source

This retrospective longitudinal study used the IBM MarketScan Commercial Claims database (2010-2018), which is an administrative claims database from commercial and employer-sponsored health insurance plans. This cohort includes data on health insurance enrolment and diagnostic codes from hospitalizations and physician visits. Linked to this database is a sub-cohort of subjects in the MarketScan Health Risk Assessment (HRA) database. This sub-cohort includes subjects who have provided data on self-reported health risk assessment questionnaires from health insurance plans which provided information on smoking exposure (smoking sub-cohort).

This study was approved by the McGill University ethics committee (#A04-M47-12B).

Subject selection

Adult subjects with new-onset RA were identified based on the presence of 2 or more physician claims and/or 1 or more hospitalization(s) with an RA principal diagnosis using International Classification of Diseases-9-Clinical Modification (ICD-9-CM) codes 714 and ICD-10-CM codes M05, M06. Subjects had to have 12 or more months of continuous enrolment in the coverage plan without any prior RA diagnosis to ensure that they were incident RA cases. The date of the first RA diagnosis was considered as the index date.

Variable definitions

In all cohorts, ILD was identified based on the presence of 2 or more physician claims at least 1 month apart. This included claim codes for fibrosis: ICD-9-CM 515, 516.3 and ICD-10-CM J84.1; and/or codes for rheumatic lung disease: ICD-9-CM 714.81 and ICD-10-CM M05.1. Pre-existing ILD before the first RA diagnostic code was excluded based on a one-year look-back period prior to index date. Co-existing COPD was identified based on previous algorithms(62) using the presence of at least 1 claim using ICD-9-CM codes 491, 492 and 496 or ICD-10-CM codes J41-J44 both assessed up to one year before RA onset. In the smoking sub-cohort, smoking was classified as active smoker, past smoker, or never smoker. Other covariates, available in all databases were sex and age at index date. Statistical analysis

Time to ILD onset after RA diagnosis was estimated using Kaplan-Meier curves and compared between those with and without coexisting COPD, as well as by smoking status in the smoking sub-cohort. The Cramer's V correlation coefficient was assessed between smoking and COPD in the HRA sub-cohort. Adjusted hazard ratios (HR) with 95% confidence intervals (CI) for ILD onset were obtained using Cox proportional hazards models in all cohorts. The models were also adjusted for age and sex and additionally by smoking in the sub-cohort.

Results

373 940 incident RA subjects were identified in the MarketScan database, with a mean follow up time of 2.1 [standard deviation (SD) 1.8] years. 6343 (1.7%) subjects developed ILD, with an incidence rate of 8.1 per 1000 person-year (95% CI 7.9, 8.3). 18 808 incident RA subjects were identified in the smoking sub-cohort, with a mean follow up time of 2.6 (SD 1.8) years. In this sub-cohort, 288 (1.5%) subjects developed ILD, with an incidence rate of 5.9 per 1000 person-year (95% CI 5.2, 6.6). In this sub-cohort, 3499 (18.6%) of subjects were missing smoking questionnaire responses. In both cohorts the majority of

subjects were women, with a mean age of about 50 years, shown in Table 2. Baseline COPD was present in 6.4% of those who developed ILD versus 3.0% of those who did not, with a difference between proportions of 3.4% (95% CI 2.8%, 4.0%, Table 2). In Kaplan-Meier survival analysis, subjects with COPD had significantly higher rates of ILD onset (Figure 1). When adjusted for age and sex, baseline COPD was associated with a higher likelihood of developing ILD (HR 2.15, 95% CI 1.93, 2.34, Table 3). In the HRA sub-cohort, the most recent smoking questionnaires available were on average 1.0 (SD 1.1) years before the index date. Most (76.8%) RA subjects in this subset were never smokers and there was no clear difference in baseline smoking status between those who developed ILD and those who did not (Table 2). There was also no clear difference in the time to ILD onset between RA subjects with different smoking histories on Kaplan-Meier analyses. There was a weak correlation between baseline ever versus never smoker and the presence of COPD, (Cramer's V = 0.12, 95% CI 0.10, 0.14). In multivariate analyses, we did not see a clear association between ILD and smoking; though the adjusted HR estimate for COPD continued to suggest a positive association with ILD, the 95% CI was wider (due to the smaller number of subjects) and included the null value (Table 3).

Discussion

This is one of the few studies to examine RA-ILD incidence and its association with COPD, and it is the only population-based study of COPD, ILD, and smoking. Despite the large population of RA subjects, the number of incident ILD cases was small. Though direct comparison with the literature is difficult (as previous studies have mostly examined the prevalence or cumulative incidence of RA-ILD) our results seem comparable. Over 10 years, ILD was identified in 2.2% of incident RA subjects in a Danish administrative database (65) and 3.5% of prevalent RA subjects in a US cohort (22) while the incidence of ILD in a clinical cohort of incident RA in the United Kingdom was found to be 4.1 (95% CI 3.0, 5.4) per 1000 patient-years (20). This is in keeping with our ILD incidence rate of 5.9 (95% CI 5.2, 6.6) per 1000 person-years in our HRA subset but lower than the incidence rate of 8.1 (95% CI 7.9, 8.3) per 1000 person-years in our full MarketScan RA cohort. Our estimate of baseline COPD prevalence in incident RA (3.1%) is also similar to that identified by other studies, ranging between 2.1%-7.5%(66-68).

We found that COPD was associated with RA-ILD. However, one major difficulty was in separating the independent effects of COPD versus smoking. Unfortunately, smoking could only be assessed in the limited sub-cohort of participants with available health questionnaires. In this sub-cohort, there remained a possible association towards increased ILD among those with co-existing COPD and the wide confidence interval, which includes the null, may be due to a lack of power. It is unknown to what extent

COPD and ILD share common etiologic pathways in RA. Both ILD and COPD are characterized by an inflammatory component, although affecting different areas of the lung. As mentioned earlier, COPD in RA may occur even at low levels of smoking exposure. In fact, emphysema (loss of alveoli), a COPD-related finding, has been found on lung imaging in association with ILD changes in 27% of RA patients who have never smoked (69) and obstructive lung changes are more frequent in ACPA positive subjects than seronegative RA subjects(70).

While few studies have examined the link between COPD and RA-ILD, various RA cohorts have tried to examine smoking and ILD. Some studies were unable to establish associations between smoking and RA-ILD (20, 32, 38), although this may have been possibly due to limited power. However, at least one study found that cumulative smoking exposure, measured in pack years, was associated with lung computed tomography scan abnormalities and subclinical ILD (36). Cumulative smoking was also associated with a composite measure of lung abnormalities on pulmonary function tests, chest radiograph, and symptoms (37). However this composite outcome included changes that are not specific to ILD. Unfortunately, we had no measure of cumulative smoking in our data.

Potential Limitations

Although use of a large population-based database is a strength, there are potential limitations inherent in using administrative data for health research. The validity of using ICD code diagnoses to detect a disease state is dependent on accurate coding. For RA identification, algorithms using physician billing and hospitalization diagnostic codes have been previously validated in Canadian health databases to have a sensitivity and specificity of 97% and 77% respectively (71). In other US patient cohorts (based on Veterans Affairs administrative data linked to chart review), different RA identification algorithms have been studied and show similar sensitivity and specificity when at least two claims were required (72). We did not require DMARD use as part of the criteria for RA identification (which presumably leads to better specificity but at the cost of sensitivity) because we aimed to capture all possible incident cases. Nevertheless, the demographics of our RA cohort correspond with the expected RA sex and age distribution.

ILD identification using ICD-9 and ICD-10 coding has not been rigorously validated and the ICD codes does not map well to the current clinical classification of ILD. The coding for RA-ILD per se has never been strictly validated. A recent abstract shows improved positive predictive value (PPV) for RA-ILD when 2 or more physician claims were required and PPV further increased to 72% if these diagnoses were made by a specialist (73). However, their definition included several ICD codes for other lung

diseases not elsewhere classified, which could have reduced specificity. We required at least 2 instances of more specific ILD coding (although not limited to a specialist) to define subject entry using definitions from previously published studies examining RA-ILD (28, 65, 74) and retain sensitivity. COPD coding has been validated and the definition we used has a sensitivity/specificity of 85.0% and 78.4% respectively for ICD 10 coding (75) and a sensitivity/specificity of 79% and 62% respectively for ICD 9 coding(62). The requirement of multiple COPD coding claims or the use of inhaler medication, did not lead to an increase in the area under the receiver operating characteristics curve(62).

Imperfect specificity for ILD ascertainment could result in a non-differential misclassification of the outcome, which causes imprecision and could be a contributor to the relatively wide confidence intervals for some of our HR estimates. A non-differential misclassification of COPD exposure could also bias our estimates towards the null. Together, these factors might make it more difficult to detect an effect of COPD on RA-ILD risk. It is also possible that cases of ILD could have initially been misdiagnosed as COPD, which could have driven (at least in part) our observed association between COPD and ILD. There may also have been a detection bias given that patients with COPD may be more closely monitored for lung diseases. Under-ascertainment of smoking also cannot be excluded, as subjects may minimize reporting of this behavior in healthcare insurance plan questionnaires. This would most likely be non-differential with respect to the future onset of ILD, thus once again it could have biased our estimates (if a true association between smoking and ILD exists) towards the null. Unfortunately, cumulative smoking exposure, which might be a stronger indicator of smoking's effect on ILD risk, could not be assessed with the available data.

The weak association between smoking and COPD in the smoking sub-cohort could reflect the fact that all subjects were insured employees. These subjects may be healthier and the majority was non-smokers, which would lower the risk for COPD. Under-reporting of smoking exposure by subjects in their health insurance questionnaires as well as the non-differential misclassification of COPD may also contribute to the weak association between smoking and COPD.

Another potential limitation is a relatively short follow up [mean of 2.1 (SD 1.8 years)]. This is because in MarketScan, subjects are no longer identifiable after they change insurance status; it has been shown that 16.8% of Americans change plans yearly(76),thus our findings could represent an underestimation of ILD occurrence in incident RA. Because of frequent plan changes, it is possible that some subjects actually had prevalent RA instead of true incident RA. On the other hand, if some of our subjects actually had prevalent RA, a few might have already had prevalent ILD but were mistakenly considered incident

ILD. Overall, the RA-ILD incidence rate in our MarketScan cohort was on the higher side of reported estimates.

While both rheumatoid factor and ACPA positivity are more frequent in smokers, and also associated with RA-ILD, RA-ILD is still seen in seronegative RA (77). In fact, it is unclear whether or not these autoantibodies are potentially part of the causal pathway linking smoking and RA-ILD; if they are, they should not be adjusted for. We did adjust for age and sex which remain two important and most consistently documented potential confounders in the relationship between smoking, COPD and RA-ILD.

Conclusions

In a large US based administrative claims database, we found that COPD was associated with ILD onset in incident RA subjects. We saw for a possible association between smoking and ILD, but due to power issues, we could not definitively establish a unique role for this potential risk factor in ILD. Further studies are needed to better evaluate the effects of smoking and other factors regarding COPD and RA-ILD onset, understand the existence of any underlying pathogenic mechanism, and ultimately assess smoking-cessation and ILD screening interventions in these high risk patients.

Tables and figures

Table 2: Baseline characteristics of RA subjects in the full MarketScan and HRA smoking sub-cohort

Full MarketScan cohort	All RA subjects (n=373,940)	RA-ILD (n=6,343)	RA no ILD (n=367,597)	Difference (95% CI)
Mean age, years (SD)	50.5 (10.9)	51.3 (11.1)	50.4 (10.9)	
Female, n (%)	280,469 (75.0)	4,655 (73.4)	275,814 (75.0)	-1.64 (-0.56, -2.75)
COPD, n (%)	11,396 (3.1)	406 (6.4)	10,990 (3.0)	3.41 (2.83, 4.04)

Smoking sub-cohort	All RA subjects (n=18,808)	RA-ILD (n=288)	RA no ILD (n=18,520)	Difference (95% CI)
Mean age, years (SD)	50.7(9.9)	51.8 (9.2)	50.7 (9.9)	
Female, n (%)	13,739 (73.1)	202 (70.1)	13,537 (73.1)	3.00 (-2.00, 8.51)
Smoking, n (%)				
• Current	1,444 (9.4)	19 (8.3)	1,425 (9.5)	1.10 (-2.40, 3.46)
• Past	2,103 (13.7)	32 (14.0)	2,071 (13.7)	0.07 (-4.11, 3.24)
• Never	11,762 (76.8)	178 (77.7)	11,584 (76.8)	0.74 (-4.72, 6.52)
COPD, n (%)	704 (3.7)	16 (5.6)	688 (3.7)	1.84 (-0.29, 5.13)

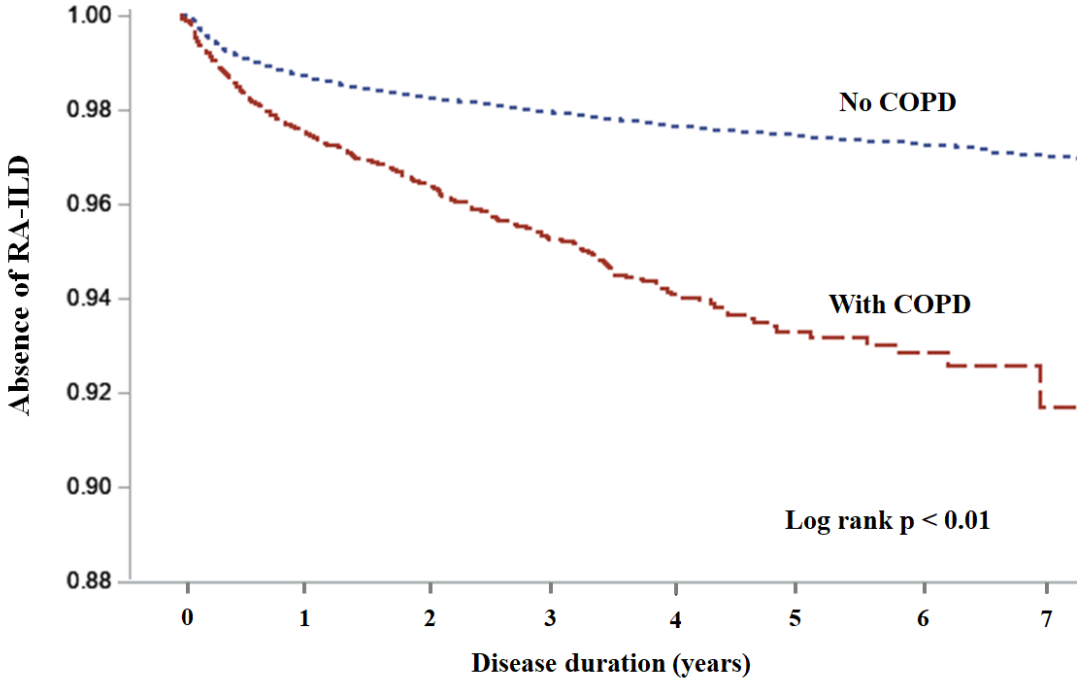
95% CI = 95% confidence interval, SD = standard deviation, ILD = Interstitial lung disease, COPD = Chronic pulmonary obstructive disease

Table 3: Multivariate Cox regression models of the likelihood of RA-ILD onset in the full MarketScan and HRA smoking sub-cohort

Full MarketScan cohort	Hazard Ratio (95% CI)
COPD	2.15 (1.93, 2.39)
Female	0.93 (0.88, 0.99)
Age at RA onset	1.00 (1.00, 1.01)
Smoking sub-cohort	**Hazard Ratio (95% CI)**
Smoking	
• Current vs never	0.80 (0.49, 1.29)
• Past vs never	0.99 (0.68, 1.44)
COPD	1.67 (0.94, 2.93)
Female	0.83 (0.62, 1.11)
Age at RA onset	1.01 (1.00, 1.03)

95% CI = 95% confidence interval, RA = Rheumatoid arthritis, COPD = Chronic pulmonary obstructive disease

Figure 1: Kaplan Meier curves for time to ILD onset stratified by co-existing COPD in the overall MarketScan RA cohort



CHAPTER 4: RESULTS OF ANALYSES IN THE MEDICAID COHORT ON RACE/ETHNICITY AND RA-ILD

Our last objective was to examine race/ethnicity as a risk factor for RA-ILD. This was examined in the Medicaid cohort. This is presented below separately because the Medicaid data contains race/ethnicity information and this database is not linkable with the MarketScan Commercial Claims and HRA data.

Results

48, 695 incident RA subjects were identified in the Medicaid cohort, with a mean follow-up time of 3.3 (SD 2.0) years. 1016 (2.2%) subjects developed ILD, with an incidence rate of 6.3 (95%CI, 5.7, 7.1) per 1000 person-year. Table 4 shows baseline features of those who developed ILD and those who did not. 416 (0.85%) subjects had missing information on race/ethnicity. There was a higher proportion of Hispanic subjects among those who developed ILD (3.4%) compared to those who did not (2.3%), with an absolute difference of 1.06 (95%CI, CI 0.11, 2.36). The presence of COPD was also more common in the ILD (versus no-ILD) group.

Table 4: Medicaid cohort baseline characteristics

Medicaid cohort	All RA subjects (n=48,695)	RA-ILD (n=1,016)	RA no-ILD (n=47,679)	Difference (95% CI)
Mean follow-up duration, years (SD)	1208.4 (733.8)	577.23 (600.0)	1221.86 (730.5)	
Mean age, years (SD)	55.5 (15.6)	55.44 (15.1)	55.49 (15.7)	
Female, n (%)	39524 (81.2)	826 (81.3)	38698 (81.16)	0.14 (-2.40, 2.44)
Race/ethnicity, n (%)*				
• White	26439 (54.8)	575 (56.9)	25864 (54.7)	2.35 (-0.75, 5.4)
• Black	15179 (31.4)	294 (29.1)	14885 (31.5)	-2.28 (-5.02, 0.61)
• Hispanic	1124 (2.3)	34 (3.4)	1090 (2.3)	1.06 (0.11, 2.36)
• Other	5537 (11.5)	108 (10.7)	5429 (11.5)	-0.76(-2.53, 1.31)
COPD, n (%)	8357 (17.2)	285 (28.1)	8072 (16.9)	11.12 (8.42, 13.98)

*416 (0.85%) of subjects had missing information on race/ethnicity
95% CI = 95% confidence interval, SD = standard deviation, ILD = Interstitial lung disease, COPD = Chronic pulmonary obstructive disease

The incidence rate of ILD in the cohort based on race/ethnicity is shown in Table 5. Kaplan-Meier curves for ILD onset shows that Hispanic subjects developed more ILD during follow-up (log rank p=0.03) (Figure 2). On multivariate analyses adjusted for age, sex, and presence of COPD, Hispanic subjects had a higher risk of developing ILD than white subjects (HR 1.49, 95% CI 1.06, 2.12) (Table 6).

In secondary analyses, the incidence rate of ILD among those with baseline COPD (up to the time of RA diagnosis) was significantly higher than those without COPD (Table 5). This association remained evident in multivariate analyses, where the presence of COPD was associated with a higher risk of developing ILD, HR 2.01, 95% CI 1.74, 2.31 (Table 6).

Figure 2: Kaplan Meier curves for time to ILD onset stratified by race in the Medicaid cohort

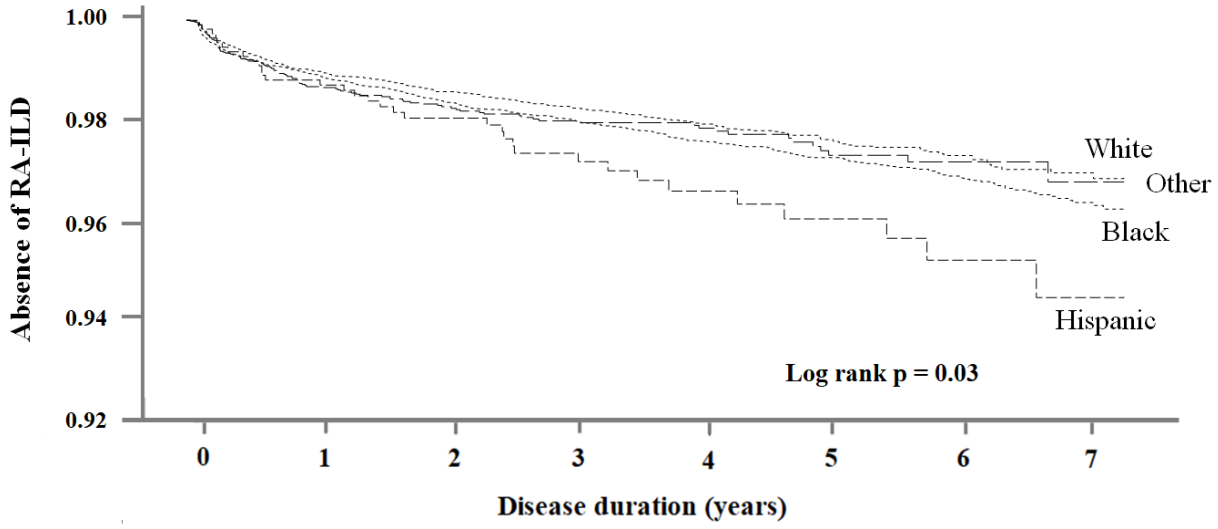


Table 5: Incidence rate of ILD in the Medicaid cohort

	Incidence rate (per 1000 PY)	95% CI	
Medicaid cohort	6.34	5.67	7.13
Race/Ethnicity			
• White	6.53	6.02	7.09
• Black	5.57	4.97	6.25
• Hispanic	9.14	6.53	12.79
• Other	6.59	5.46	7.96
Presence vs absence of COPD			
• No COPD	5.36	4.99	5.77
• COPD	11.37	10.13	12.78

PY = person-years, COPD = chronic pulmonary obstructive disease

Table 6: Cox regression of the likelihood of RA-ILD onset in the Medicaid cohort

Medicaid cohort	Univariate Cox model Hazard Ratio (95% CI)	Multivariate Cox model Hazard Ratio (95% CI)
Race/ethnicity (compared to White)		
• Black	0.87 (0.75, 1.00)	0.92 (0.80, 1.06)
• Hispanic	1.40 (0.99, 1.97)	1.49 (1.06, 2.12)
• Other	0.97 (0.79, 1.19)	0.99 (0.80, 1.22)
Female	0.99 (0.84, 1.16)	1.01 (0.86, 1.19)
Age	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
COPD presence	1.97 (1.71, 2.25)	2.01 (1.74, 2.31)

COPD = chronic pulmonary obstructive disease

CHAPTER 5: DISCUSSION OF OVERALL RESULTS

The following chapter will discuss the results of all the analyses performed and their limitations. Because we have already discussed in Chapter 3 our findings in the Commercial Claims/HRA dataset, regarding smoking, COPD, and RA-ILD, there will be a particular focus in this chapter on the association between race/ethnicity and RA-ILD in the Medicaid cohort. We will also discuss how findings in this cohort compare to the full MarketScan Commercial Claims cohort as well as to existing literature.

Prevalence/Incidence of RA-ILD

The majority of RA-ILD studies published to date have examined the prevalence or the cumulative incidence rather than annual incidence rates. We identified both cumulative incidence and annual incidence rates which can be compared to existing literature.

In the Medicaid cohort, the cumulative incidence of RA- ILD over the follow up period was 2.1%. This is similar to our results in the Commercial Claims cohort, as well as earlier studies. For example, in a Danish population based study using the same ICD 10 codes we used for both RA and ILD: the prevalence of new onset ILD in incident RA subjects between 2004-2015 was almost identical, at 2.2% (65). In two large United Kingdom (UK) RA inception cohorts, the cumulative incidence of RA-ILD was slightly higher (3.7%) but patients were followed longer (a maximum follow-up of 25 years) than in our study (78). Similarly various other cohorts have reported a 10-year cumulative incidence of RA-ILD ranging from 3.5%-6.6%(79) but the higher rate is expected given the relatively long follow-up.

Our RA-ILD incidence rates were 8.1 (95% CI 7.9, 8.3) per 1000 person-years in in the Medicaid cohort and 6.3 (95%CI 5.67, 7.13) per 1000 person-years in the Commercial Claims cohort. The reason for the slight difference is unclear and could be related to several unmeasured variables, including race/ethnicity, socioeconomic status, smoking, or differences in health care access and diagnostic testing. In fact, there is evidence that access for low income US residents may actually be better with Medicaid than with private insurance(80, 81). There have been few other studies which reported incidence rates of RA-ILD. In one UK cohort of patients with early RA (less than 2 years disease duration) with a median follow up duration of 10 years, the incidence rate of RA-ILD was 4.1 per 1000 patient-years (95% CI 3.0, 5.4)(20).

Race/Ethnicity and RA-ILD

Despite having few Hispanic subjects (representing only 2.3% of the Medicaid cohort), we found that they were more likely to develop ILD than white RA subjects. This is a novel finding in RA-ILD and is

supported by a previous study which found that Hispanic subjects had an increased risk of any extra-articular RA manifestation (OR 3.43, 95% CI 1.53-7.70) compared to white subjects after adjusting for age, sex, disease duration, and RF status (82). Indirect support for our finding was also seen in a large US cohort showing Hispanics and blacks with autoimmune rheumatic diseases were more likely to develop ILD than whites (57). The proportion of Hispanic and black subjects are similar to those found in other Medicaid studies in rheumatic disease (83). We were unable to study other race/ethnicity groups (e.g. Asians) in the current study because they were categorized as "Other". Detailed assessments of the effects of race/ethnicity is also highly challenging because of its complex relations with other characteristics including access to care, smoking, occupational and environmental exposures, and other potential confounders or effect modifiers of RA-ILD risk that were unaccounted for.

COPD and RA-ILD

This is one of the first studies of COPD and RA-ILD and this association was examined in both the Commercial Claims and Medicaid cohorts. The prevalence of COPD among incident RA subjects in the Medicaid cohort (17.2%) was significantly higher than the prevalence of COPD in our Commercial Claims cohort. This may be due to differences between subjects with private health plans and subjects on Medicaid as COPD prevalence varies greatly across regions, particularly in areas of different social-economic status in the US (84, 85).

In the Medicaid cohort, there was a twofold increase in the risk for ILD related to COPD (HR 2.01, 95% CI 1.74, 2.31) after adjusting for race/ethnicity, age, and sex. This is comparable to the Commercial Claims cohort, where the HR associated with COPD and ILD was 2.15 (95% CI 1.93, 2.39). In the HRA sub-cohort of patients with smoking information, the HR for increased ILD risk related to COPD had a wider confidence interval that included the null value (HR 1.67, 95% CI 0.94, 2.93). This could be due to a much lower power in the HRA sub-cohort. The additional adjustment for smoking is less likely to be the main cause of this non-statistically significant finding because there was only a weak association between smoking status and COPD in the HRA sub-cohort. While we accounted for current or past smoking, we were unable to account for cumulative smoking which is more strongly associated with COPD. Furthermore, given the rarity of ILD, we could not definitively establish if COPD is linked to ILD or whether smoking is a confounder.

Potential Limitations

The analyses in the Commercial Claims, HRA, and Medicaid cohorts share similar limitations which will be discussed in more detail below; cohort specific limitations, where they apply, will also be pointed out.

Accuracy of disease identification

The major potential limitation inherent in using administrative data for health research is that the validity of billing and hospitalization diagnoses is dependent on accurate coding. For RA identification, the coding used has been previously validated in Canadian health databases and was reported to have a sensitivity and specificity of 97% and 77% respectively (71). This definition was further validated in Swedish health databases using our ICD 9 and ICD 10 diagnostic codes(86). The addition of further requirements to include at least 3 RA claims or claims by a rheumatologist did not improve accuracy (71). In various administrative-data-based US cohorts, different RA case definition algorithms (including additional medication and/or blood test requirements) showed globally similar sensitivity and specificity, as long as at least two physician claims were required (72). We did not require that subjects be on DMARDs at diagnosis as part of the RA identification criteria because we aimed to capture all possible incident cases. While this additional requirement improved positive predictive values of RA diagnoses in a validation study of elderly onset RA, where the mean age was 80 years old(87), it did not impact the accuracy in other validation studies (71, 88) particularly when incident RA was required. Regardless, the patient demographics of our RA cohorts correspond with the expected age and sex distribution.

Contrary to RA, ILD identification using ICD-9 and ICD-10 coding has been less extensively validated and is less robust. ILD coding maps poorly to the current clinical classification as the field has evolved significantly since the inception of ICD-10 coding in 1990. In IPF, the prototypical ILD for which there must not be any underlying cause such as autoimmune disease, validation studies of coding algorithms have shown positive predictive values (PPV) of 44-46% in two US cohorts when one IPF related claim was required(89). When at least 2 instances of ILD claims were required along with a procedural claim for lung imaging, the PPV increased to 66%(89). The coding for RA-ILD has never been strictly validated, but can be thought to have less specific requirements than IPF because it requires much less stringent diagnostic criteria. A recent unpublished abstract shows PPV for RA-ILD improved from 43% to 52% when requiring at least 2 instances of ICD-9 claims, and further improvement to 72% if these diagnoses were made by a specialist (73). They found no benefit by requiring a diagnostic test to be performed after ILD diagnosis. However, those authors included several codes for various other lung diseases not elsewhere classified, which may have reduced specificity (73). Thus, we required at least 2 instances of ILD coding to define subject entry and this used a definition that was in keeping with several published

studies examining RA-ILD in particular (28, 65, 74) even though this definition has not been validated against another reference standard.

COPD identification using administrative claims coding has been validated in different cohorts. The ICD-10 coding definition we used had a sensitivity and specificity of 85.0% and 78.4% respectively (75) and the ICD-9 coding definition used had a sensitivity and specificity of 79% and 62% respectively(62). The addition of other coding requirements (including at least 2 COPD claim codes and/or the use of inhaler medication) into an identification algorithm did not significantly improve accuracy. There were small increases in specificity to the detriment of sensitivity, however the area under the receiver operating characteristics curve remained the same when validated in a large US cohort(62).

Due to the inaccuracy of coding identification, the misclassification of ILD outcome would most likely to be non-differential when different race/ethnicities and different smoking status are considered the exposure. This would favor the finding of a null association between the exposure and the outcome, which would bias against finding any significant result. Because COPD and ILD are respiratory diseases that can both manifest with dyspnea and cough, it is conceivable that COPD may have been misdiagnosed as ILD or vice-versa, leading to a differential misclassification between COPD status (exposure) and ILD (outcome). While this may bias towards a false association between COPD and ILD, this scenario would be unlikely given the very different diagnostic modalities between the two diseases. Further, the requirement of at least 2 ILD codes, and the gap in time between a COPD diagnosis preceding RA with an ILD diagnosis after RA onset would also make this scenario less likely.

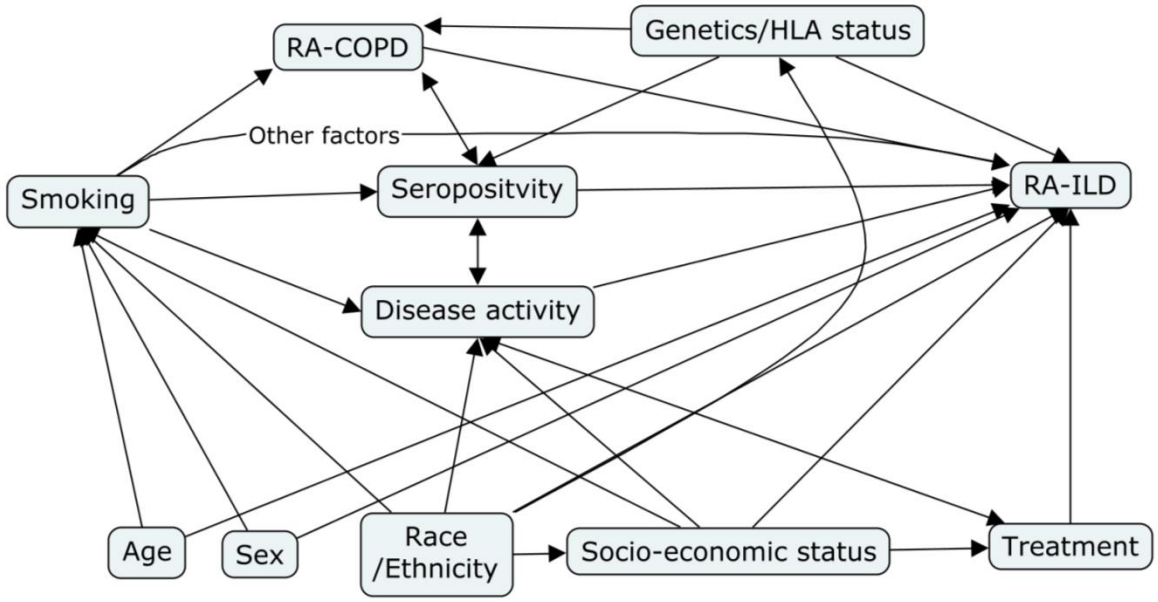
Another potential limitation with administrative data is that some RA subjects may not represent true incident cases of RA. This is because subjects may drop in and out of administrative health records because of changes in health plans. While a 1 year look back period was required to ensure that no previous RA or ILD diagnoses were present before the index date, patients with very stable RA who did not consult any physician during that period could have been included as incident patients. Regardless, this may potentially affect the incidence rate of RA-ILD (possibly inflating the 'incident' ILD cases with prevalent cases). Although as noted earlier, with a relatively short average follow-up, we may have a more conservative estimate of RA-ILD incidence.

Potential residual confounding

We were unable to account for multiple risk factors of RA-ILD such as RF or ACPA seropositivity, RA disease activity, and HLA phenotypes. These have complex relationships with RA-ILD, smoking, COPD, and race/ethnicity. A non-exhaustive directed acyclic graph of these relationships is shown in Figure 3.

Both RF and ACPA seropositivity are associated with increased RA, RA-ILD, and RA-COPD. Smokers also tend to have more seropositive disease. Because seropositivity may be implicated in the causal pathway between smoking and RA-ILD, seropositivity may not need to be adjusted for as a confounder. However seropositivity is unlikely to be the sole cause of increased RA-ILD as both RA and RA-ILD can be seen in sero-negative patients. Also unclear is how genetic factors might influence RA-COPD and RA-ILD beyond seropositivity. Racial differences may also affect RA-ILD onset through genetic factors and other factors related to differential exposures stemming from social economic status and health care access. We were only able to account for age and sex which nevertheless remain two of the most important and consistently documented confounders in the relationship between smoking, COPD and RA-ILD. We were also unable to adjust for disease activity which has been associated with RA-ILD development (90).

Figure 3: Directed acyclic graph of the complex relationships between smoking, COPD, race/ethnicity, and other factors associated with RA-ILD



Caption: This directed acyclic graph is non-exhaustive and shows presumed causal relationships denoted by an arrow. RA = rheumatoid arthritis, COPD = chronic obstructive pulmonary disease, HLA = human leukocyte antigen, ILD = interstitial lung disease, seropositivity = either rheumatoid factor or anti-citrullinated peptide positive rheumatoid arthritis.

Other limitations

In spite of the large population based databases there were relatively few cases of RA-ILD identified over the follow up intervals. Further stratification by smoking status and race/ethnicity limited the power of our analyses, particularly with regards to smoking.

Inherent in our assessment of smoking (by a self-reported measure) was the possibility of misclassification of exposure. Reporting bias cannot be excluded as subjects may minimize reported smoking in healthcare plan questionnaires. We believe this would be non-differential with respect to the outcome of ILD and would bias any true association between smoking and ILD towards the null. Furthermore, cumulative smoking exposure, which is a stronger indicator of smoking risk, could not be assessed with the available data.

Other potential limitations include a short follow up duration with a mean of 2.1±1.8 years after the first RA diagnosis in the MarketScan cohort and 3.3±2.0 years in the Medicaid cohort. This short disease duration may be because subjects are no longer identifiable when they change insurance status and it has been shown that close to 17% of Americans change plans each year (76). In longitudinal RA cohorts of more than 20 years duration, close to 50% of RA-ILD occurs within the first 5 years after RA onset (22, 78). Thus, since RA-ILD can still develop later in the disease, our findings are likely an underestimate of the lifetime incidence of RA-ILD given the short follow up duration. We also did not examine patients who already developed ILD before RA diagnosis given that our goal was to investigate risk factors for RA-ILD incidence.

SUMMARY

Using a US population based administrative database, we assessed the incidence of ILD in new onset RA subjects and its association with certain potential risk factors. Despite the large number of RA subjects, incident RA-ILD identified using administrative claim diagnoses were rare. We found that RA-ILD onset was associated with pre-existing COPD at RA diagnosis, but not with current or past smoking. However the assessment of smoking was limited by the absence of cumulative exposure data and low statistical power. We also found an association between Hispanic race/ethnicity and increased RA-ILD. These findings may help target high risk subjects that may benefit from ILD screening. Further studies are needed to better evaluate these factors in well characterized clinical cohorts to understand the existence of any underlying pathogenic mechanism, to assess the role of interventions or increased monitoring in these high risk patients.

REFERENCES

1. Hunter TM, Boytsov NN, Zhang X, Schroeder K, Michaud K, Araujo AB. Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004-2014. Rheumatol Int. 2017;37(9):1551-7.
2. Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatology. 2012;51(suppl_6):vi5-vi9.
3. Taylor P, Gartemann J, Hsieh J, Creeden J. A Systematic Review of Serum Biomarkers Anti-Cyclic Citrullinated Peptide and Rheumatoid Factor as Tests for Rheumatoid Arthritis. Autoimmune Diseases. 2011;2011:815038.
4. Mian A, Ibrahim F, Scott DL. A systematic review of guidelines for managing rheumatoid arthritis. BMC Rheumatology. 2019;3(1):42.
5. Deane KD, Demoruelle MK, Kelmenson LB, Kuhn KA, Norris JM, Holers VM. Genetic and environmental risk factors for rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2017;31(1):3-18.
6. Lahiri M, Morgan C, Symmons DPM, Bruce IN. Modifiable risk factors for RA: prevention, better than cure? Rheumatology. 2011;51(3):499-512.
7. Karlson EW, Deane K. Environmental and gene-environment interactions and risk of rheumatoid arthritis. Rheum Dis Clin North Am. 2012;38(2):405-26.
8. Makrygiannakis D, af Klint E, Lundberg IE, Löfberg R, Ulfgren AK, Klareskog L, et al. Citrullination is an inflammation-dependent process. Ann Rheum Dis. 2006;65(9):1219-22.
9. Vossenaar ER, van Venrooij WJ. Citrullinated proteins: sparks that may ignite the fire in rheumatoid arthritis. Arthritis Res Ther. 2004;6(3):107-11.
10. Rangel-Moreno J, Hartson L, Navarro C, Gaxiola M, Selman M, Randall TD. Inducible bronchus-associated lymphoid tissue (iBALT) in patients with pulmonary complications of rheumatoid arthritis. The Journal of clinical investigation. 2006;116(12):3183-94.
11. Demoruelle MK, Weisman MH, Simonian PL, Lynch DA, Sachs PB, Pedraza IF, et al. Brief report: airways abnormalities and rheumatoid arthritis-related autoantibodies in subjects without arthritis: early injury or initiating site of autoimmunity? Arthritis Rheum. 2012;64(6):1756-61.
12. Samara KD, Trachalaki A, Tsitoura E, Koutsopoulos AV, Lagoudaki ED, Lasithiotaki I, et al. Upregulation of citrullination pathway: From Autoimmune to Idiopathic Lung Fibrosis. Respir Res. 2017;18(1):218-.
13. Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, Eklund A, et al. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Ann Rheum Dis. 2008;67(10):1488-92.
14. Lugli EB, Correia RE, Fischer R, Lundberg K, Bracke KR, Montgomery AB, et al. Expression of citrulline and homocitrulline residues in the lungs of non-smokers and smokers: implications for autoimmunity in rheumatoid arthritis. Arthritis Res Ther. 2015;17(1):9.
15. Klareskog L, Stolt P, Lundberg K, Källberg H, Bengtsson C, Grunewald J, et al. 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. 2006;54(1):38-46.
16. Bernatsky S, Smargiassi A, Joseph L, Awadalla P, Colmegna I, Hudson M, et al. Industrial air emissions, and proximity to major industrial emitters, are associated with anti-citrullinated protein antibodies. Environmental research. 2017;157:60-3.
17. Shaw M, Collins BF, Ho LA, Raghu G. Rheumatoid arthritis-associated lung disease. European Respiratory Review. 2015;24(135):1-16.

18. Salaffi F, Carotti M, Di Carlo M, Tardella M, Giovagnoni A. High-resolution computed tomography of the lung in patients with rheumatoid arthritis: Prevalence of interstitial lung disease involvement and determinants of abnormalities. Medicine (Baltimore). 2019;98(38):e17088.
19. Dawson JK, Fewins HE, Desmond J, Lynch MP, Graham DR. Predictors of progression of HRCT diagnosed fibrosing alveolitis in patients with rheumatoid arthritis. Annals of the Rheumatic Diseases. 2002;61(6):517-21.
20. Koduri G, Norton S, Young A, Cox N, Davies P, Devlin J, et al. Interstitial lung disease has a poor prognosis in rheumatoid arthritis: results from an inception cohort. Rheumatology (Oxford). 2010;49(8):1483-9.
21. Turesson C, O'Fallon WM, Crowson CS, Gabriel SE, Matteson EL. Extra-articular disease manifestations in rheumatoid arthritis: incidence trends and risk factors over 46 years. Ann Rheum Dis. 2003;62(8):722-7.
22. Bongartz T, Nannini C, Medina-Velasquez YF, Achenbach SJ, Crowson CS, Ryu JH, et al. Incidence and mortality of interstitial lung disease in rheumatoid arthritis: a population-based study. Arthritis Rheum. 2010;62(6):1583-91.
23. Olson AL, Swigris JJ, Sprunger DB, Fischer A, Fernandez-Perez ER, Solomon J, et al. Rheumatoid Arthritis–Interstitial Lung Disease–associated Mortality. Am J Resp Crit Care Med. 2011;183:372-8.
24. Brown KK. Rheumatoid Lung Disease. Proc Am Thorac Soc. 2007;4:443-8.
25. Bongartz T, Nannini C, Medina-Velasquez Y, Achenbach S, Crowson C, Ryu J, et al. Incidence and mortality of interstitial lung disease in rheumatoid arithritis: a population-based study. Arthritis Rheum. 2010;62:1583-91.
26. Doyle TJ, Lee JS, Dellaripa PF, Lederer JA, Matteson EL, Fischer A, et al. A roadmap to promote clinical and translational research in rheumatoid arthritis-associated interstitial lung disease. Chest. 2014;145(3):454-63.
27. Crestani B. The respiratory system in connective tissue disorders. Allergy. 2005;60(6):715-34.
28. Raimundo K, Solomon JJ, Olson AL, Kong AM, Cole AL, Fischer A, et al. Rheumatoid Arthritis-Interstitial Lung Disease in the United States: Prevalence, Incidence, and Healthcare Costs and Mortality. J Rheumatol. 2018.
29. Spagnolo P, Lee JS, Sverzellati N, Rossi G, Cottin V. The Lung in Rheumatoid Arthritis: Focus on Interstitial Lung Disease. Arthritis Rheumatol. 2018;70(10):1544-54.
30. Cojocaru M, Cojocaru IM, Silosi I, Vrabie CD, Tanasescu R. Extra-articular Manifestations in Rheumatoid Arthritis. Maedica (Buchar). 2010;5(4):286-91.
31. Kelly CA, Saravanan V, Nisar M, Arthanari S, Woodhead FA, Price-Forbes AN, et al. Rheumatoid arthritis-related interstitial lung disease: associations, prognostic factors and physiological and radiological characteristics--a large multicentre UK study. Rheumatology (Oxford). 2014;53(9):1676-82.
32. Mori S, Koga Y, Sugimoto M. Different risk factors between interstitial lung disease and airway disease in rheumatoid arthritis. Respir Med. 2012;106(11):1591-9.
33. Sparks JA, He X, Huang J, Fletcher EA, Zaccardelli A, Friedlander HM, et al. Rheumatoid Arthritis Disease Activity Predicting Incident Clinically Apparent Rheumatoid Arthritis-Associated Interstitial Lung Disease: A Prospective Cohort Study. Arthritis Rheumatol. 2019;71(9):1472-82.
34. Juge P-A, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, et al. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. New England Journal of Medicine. 2018;379(23):2209-19.
35. Bellou V, Belbasis L, Konstantinidis A, Evangelou E. Tobacco smoking and risk for idiopathic pulmonary fibrosis: a prospective cohort study in UK Biobank. European Respiratory Journal. 2017;50(suppl 61):PA4887.
36. Johnson C, Giles JT, Bathon J, Lederer D, Hoffman EA, Barr RG, et al. Smoking and Subclinical ILD in RA versus the Multi-Ethnic Study of Atherosclerosis. PLoS One. 2016;11(4):e0153024.

37. Saag KG, Kolluri S, Koehnke RK, Georgou TA, Rachow JW, Hunninghake GW, et al. Rheumatoid arthritis lung disease. Determinants of radiographic and physiologic abnormalities. Arthritis Rheum. 1996;39(10):1711-9.
38. Nurmi HM, Purokivi MK, Karkkainen MS, Kettunen HP, Selander TA, Kaarteenaho RL. Variable course of disease of rheumatoid arthritis-associated usual interstitial pneumonia compared to other subtypes. BMC pulmonary medicine. 2016;16(1):107.
39. Bilgici A, Ulusoy H, Kuru O, Celenk C, Unsal M, Danaci M. Pulmonary involvement in rheumatoid arthritis. Rheumatology international. 2005;25(6):429-35.
40. Dawson JK, Fewins HE, Desmond J, Lynch MP, Graham DR. Fibrosing alveolitis in patients with rheumatoid arthritis as assessed by high resolution computed tomography, chest radiography, and pulmonary function tests. Thorax. 2001;56(8):622-7.
41. McDonagh J, Greaves M, Wright AR, Heycock C, Owen JP, Kelly C. High resolution computed tomography of the lungs in patients with rheumatoid arthritis and interstitial lung disease. British journal of rheumatology. 1994;33(2):118-22.
42. Biederer J, Schnabel A, Muhle C, Gross WL, Heller M, Reuter M. Correlation between HRCT findings, pulmonary function tests and bronchoalveolar lavage cytology in interstitial lung disease associated with rheumatoid arthritis. European radiology. 2004;14(2):272-80.
43. Dixon WG, Hyrich KL, Watson KD, Lunt M, Symmons DP. 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. 2010;69(6):1086-91.
44. Assayag D, Lubin M, Lee JS, King TE, Collard HR, Ryerson CJ. Predictors of mortality in rheumatoid arthritis-related interstitial lung disease. Respirology. 2014;19(4):493-500.
45. Marsh S, Aldington S, Shirtcliffe P, Weatherall M, Beasley R. Smoking and COPD: what really are the risks? European Respiratory Journal. 2006;28(4):883-4.
46. Rycroft CE, Heyes A, Lanza L, Becker K. Epidemiology of chronic obstructive pulmonary disease: a literature review. Int J Chron Obstruct Pulmon Dis. 2012;7:457-94.
47. King PT. Inflammation in chronic obstructive pulmonary disease and its role in cardiovascular disease and lung cancer. Clin Transl Med. 2015;4(1):68.
48. Ma Y, Tong H, Zhang X, Wang M, Yang J, Wu M, et al. Chronic obstructive pulmonary disease in rheumatoid arthritis: a systematic review and meta-analysis. Respir Res. 2019;20(1):144-.
49. Ungprasert P, Srivali N, Cheungpasitporn W, Davis Iii JM. Risk of incident chronic obstructive pulmonary disease in patients with rheumatoid arthritis: A systematic review and meta-analysis. Joint Bone Spine. 2016;83(3):290-4.
50. Ford JA, Liu X, Chu SH, Lu B, Cho MH, Silverman EK, et al. Asthma, Chronic Obstructive Pulmonary Disease, and Subsequent Risk for Incident Rheumatoid Arthritis Among Women: A Prospective Cohort Study. Arthritis Rheumatol. 2020;72(5):704-13.
51. Hassan WU, Keaney NP, Holland CD, Kelly CA. High resolution computed tomography of the lung in lifelong non-smoking patients with rheumatoid arthritis. Annals of the rheumatic diseases. 1995;54(4):308-10.
52. Kusko RL, Brothers JF, 2nd, Tedrow J, Pandit K, Huleihel L, Perdomo C, et al. Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2016;194(8):948-60.
53. Nunes H, Monnet I, Kannengiesser C, Uzunhan Y, Valeyre D, Kambouchner M, et al. Is telomeropathy the explanation for combined pulmonary fibrosis and emphysema syndrome?: report of a family with TERT mutation. American journal of respiratory and critical care medicine. 2014;189(6):753-4.

54. Stanley SE, Merck SJ, Armanios M. Telomerase and the Genetics of Emphysema Susceptibility. Implications for Pathogenesis Paradigms and Patient Care. Annals of the American Thoracic Society. 2016;13 Suppl 5(Suppl 5):S447-S51.
55. Olson AL, Gifford AH, Inase N, Fernandez Perez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. Eur Respir Rev. 2018;27(150).
56. Gelber AC, Manno RL, Shah AA, Woods A, Le EN, Boin F, et al. Race and association with disease manifestations and mortality in scleroderma: a 20-year experience at the Johns Hopkins Scleroderma Center and review of the literature. Medicine (Baltimore). 2013;92(4):191-205.
57. Adegunsoye A, Oldham JM, Bellam SK, Chung JH, Chung PA, Biblowitz KM, et al. African-American race and mortality in interstitial lung disease: a multicentre propensity-matched analysis. Eur Respir J. 2018;51(6).
58. Doyle TJ, Patel AS, Hatabu H, Nishino M, Wu G, Osorio JC, et al. Detection of Rheumatoid Arthritis-Interstitial Lung Disease Is Enhanced by Serum Biomarkers. Am J Respir Crit Care Med. 2015;191(12):1403-12.
59. Morisset J, Mageto Y, Raghu G. Mortality in interstitial lung disease: do race and skin colour matter? Eur Respir J. 2018;51(6).
60. Kim SK, Park SH, Shin IH, Choe JY. Anti-cyclic citrullinated peptide antibody, smoking, alcohol consumption, and disease duration as risk factors for extraarticular manifestations in Korean patients with rheumatoid arthritis. J Rheumatol. 2008;35(6):995-1001.
61. Eser F, Garip Y, Bodur H. Extraarticular manifestations in Turkish patients with rheumatoid arthritis: impact of EAMs on the health-related quality of life in terms of disease activity, functional status, severity of pain, and social and emotional functioning. Rheumatology international. 2012;32(6):1521-5.
62. Cooke CR, Joo MJ, Anderson SM, Lee TA, Udris EM, Johnson E, et al. The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. BMC Health Serv Res. 2011;11:37.
63. Gothe H, Rajsic S, Vukicevic D, Schoenfelder T, Jahn B, Geiger-Gritsch S, et al. Algorithms to identify COPD in health systems with and without access to ICD coding: a systematic review. BMC Health Serv Res. 2019;19(1):737.
64. Myasoedova E, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. Arthritis and rheumatism. 2010;62(6):1576-82.
65. Hyldgaard C, Hilberg O, Pedersen AB, Ulrichsen SP, Lokke A, Bendstrup E, et al. A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. Ann Rheum Dis. 2017;76(10):1700-6.
66. Gergianaki I, Tsiligianni I. Chronic obstructive pulmonary disease and rheumatic diseases: A systematic review on a neglected comorbidity. J Comorb. 2019;9:2235042X18820209.
67. Dougados M, Soubrier M, Antunez A, Balint P, Balsa A, Buch MH, et al. Prevalence of comorbidities in rheumatoid arthritis and evaluation of their monitoring: results of an international, cross-sectional study (COMORA). Ann Rheum Dis. 2014;73(1):62-8.
68. Norton S, Koduri G, Nikiphorou E, Dixey J, Williams P, Young A. A study of baseline prevalence and cumulative incidence of comorbidity and extra-articular manifestations in RA and their impact on outcome. Rheumatology (Oxford). 2013;52(1):99-110.
69. Jacob J, Song JW, Yoon HY, Cross G, Barnett J, Woo WL, et al. Prevalence and Effects of Emphysema in Never-Smokers with Rheumatoid Arthritis Interstitial Lung Disease. EBioMedicine. 2018;28:303-10.

70. Park WH, Kim SS, Shim SC, Song ST, Jung SS, Kim JH, et al. Visual Assessment of Chest Computed Tomography Findings in Anti-cyclic Citrullinated Peptide Antibody Positive Rheumatoid Arthritis: Is it Associated with Airway Abnormalities? Lung. 2016;194(1):97-105.
71. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JC, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. Arthritis Care Res (Hoboken). 2013;65(10):1582-91.
72. Chung CP, Rohan P, Krishnaswami S, McPheeters ML. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. Vaccine. 2013;31 Suppl 10:K41-61.
73. Cho S DT, Lee H, Jin Y, Tong A, Santiago Ortiz A, Sparks J, Kim S. Validation of Claims-based Algorithms to Identify Interstitial Lung Disease in Patients with Rheumatoid Arthritis [abstract]. Arthritis Rheumatol. 2019;71(suppl 10).
74. Olson AL, Swigris JJ, Sprunger DB, Fischer A, Fernandez-Perez ER, Solomon J, et al. Rheumatoid arthritis-interstitial lung disease-associated mortality. Am J Respir Crit Care Med. 2011;183(3):372-8.
75. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying Individuals with Physcian Diagnosed COPD in Health Administrative Databases. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2009;6(5):388-94.
76. Health Plan Switching: Choice Or Circumstance? Health Affairs. 2000;19(3):158-64.
77. Balbir-Gurman A, Guralnik L, Yigla M, Braun-Moscovici Y, Hardak E. Imaging aspects of interstitial lung disease in patients with rheumatoid arthritis: Literature review. Autoimmun Rev. 2018;17(2):87-93.
78. Kiely P, Busby AD, Nikiphorou E, Sullivan K, Walsh DA, Creamer P, et al. 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. 2019;9(5):e028466.
79. Myasoedova E, Crowson CS, Turesson C, Gabriel SE, Matteson EL. Incidence of extraarticular rheumatoid arthritis in Olmsted County, Minnesota, in 1995-2007 versus 1985-1994: a population-based study. J Rheumatol. 2011;38(6):983-9.
80. Three-Year Impacts Of The Affordable Care Act: Improved Medical Care And Health Among Low-Income Adults. Health Affairs. 2017;36(6):1119-28.
81. Haviland A, Sood N, McDevitt R, Marquis M. How Do Consumer-Directed Health Plans Affect Vulnerable Populations? Forum for Health Economics & Policy. 2011;14:3-.
82. Richman NC, Yazdany J, Graf J, Chernitskiy V, Imboden JB. Extraarticular manifestations of rheumatoid arthritis in a multiethnic cohort of predominantly Hispanic and Asian patients. Medicine (Baltimore). 2013;92(2):92-7.
83. Smoyer-Tomic KE, Amato AA, Fernandes AW. Incidence and prevalence of idiopathic inflammatory myopathies among commercially insured, Medicare supplemental insured, and Medicaid enrolled populations: an administrative claims analysis. BMC Musculoskelet Disord. 2012;13:103-.
84. Zhang X, Holt JB, Lu H, Wheaton AG, Ford ES, Greenlund KJ, et al. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. American journal of epidemiology. 2014;179(8):1025-33.
85. CDC. COPD Prevalence in the United States 2018 [updated June 5, 2018. Available from: https://www.cdc.gov/copd/data.html.
86. Waldenlind K, Eriksson JK, Grewin B, Askling J. Validation of the rheumatoid arthritis diagnosis in the Swedish National Patient Register: a cohort study from Stockholm County. BMC Musculoskelet Disord. 2014;15:432-.
87. Kim SY, Servi A, Polinski JM, Mogun H, Weinblatt ME, Katz JN, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. Arthritis Research & Therapy. 2011;13(1):R32.

88. Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. J Rheumatol. 2011;38(8):1612-6.
89. Ley B, Urbania T, Husson G, Vittinghoff E, Brush DR, Eisner MD, et al. Code-based Diagnostic Algorithms for Idiopathic Pulmonary Fibrosis. Case Validation and Improvement. Ann Am Thorac Soc. 2017;14(6):880-7.
90. Rojas-Serrano J, Herrera-Bringas D, Perez-Roman DI, Perez-Dorame R, Mateos-Toledo H, Mejia M. Rheumatoid arthritis-related interstitial lung disease (RA-ILD): methotrexate and the severity of lung disease are associated to prognosis. Clin Rheumatol. 2017;36(7):1493-500.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ACPA	Anti-citrullinated peptides antibodies
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
DMARD	Disease modifying agent
HLA	Human leukocyte antigen
HR	Hazard ratio
HRA	Health Risk Assessment (MarketScan Cohort)
HRCT	High resolution CT scan
IBM	International Business Machines Corporation
ICD-9-CM	International Classification of Diseases-9-Clinical Modification
ICD-10-CM	International Classification of Diseases-10-Clinical Modification
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
OR	Odds ratio
PPV	Positive predictive value
PY	Person-years
RA	Rheumatoid arthritis
RF	Rheumatoid factor
SD	Standard deviation
US	United States
UK	United Kingdom