UCSF
UC San Francisco Electronic Theses and Dissertations

Title
COVID-19 Outcomes in People with Rheumatic Disease: Results from a Global Physician-Reported Registry

Permalink
https://escholarship.org/uc/item/9v17d7tc

Author
Izadi, Zahra

Publication Date
2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

COVID-19 Outcomes in People with Rheumatic Disease: Results from a Global Physician-Reported Registry

by
Zahra Izadi

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Science

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Jinoos Yazdany
E1F7BDB9E03644B...
Jinoos Yazdany
Chair

DocuSigned by:
Maria Glymour
Maria Glymour

DocuSigned by:
David V. Glidden
759D048A0C4A4F7...
Davd V. Glidden

Committee Members

Copyright 2022

by

Zahra Izadi

Dedication

I am grateful to the American College of Rheumatology and the European Alliance of Associations for Rheumatology for funding that made this research possible.

First and foremost, I would like to thank the members of my Dissertation Committee, Jinoos Yazdany, Maria Glymour, and Dave Glidden. Your mentorship throughout the doctoral program has been invaluable for my growth as an epidemiologist. Dr Yazdany, I have been incredibly lucky to have you as a mentor, advisor, and committee chair. Thank you for being so patient with me when I joined UCSF as I transitioned back into academia, and for your generosity with your time, expertise, resources, and professional network. Throughout your mentorship you always made me feel supported and created a sense of connection and belonging that encouraged me to think bigger, take risks, and explore new ideas. Your mentorship has had a significant impact not just on my academic progress and success, but also on my personal growth, and sense of fulfillment. Dr Glymour, thank you for your phenomenal mentorship as a committee member and PhD program director. I am grateful for your continued help and guidance throughout the various stages of the program. Your mentorship has taught me not only to conduct rigorous research but to always think about and reflect on the broader public health impact of my work. Your attention to validity, precision, and generalizability has impacted the way I approach research questions and will have a lasting effect on my future research. Dr Glidden, thank you for your amazing mentorship, your support as a committee member, and guidance on the statistical aspects of my research. I have learned so much from you through the biostatistics courses, the PhD seminars, and one-on-one meetings. Your expertise in conveying complex biostatistical methods using simpler core concepts has been instrumental to my research and helped me immensely as I explored novel applications of these methods.

To my mentors and colleagues at the Quality and Informatics Lab, at UCSF, especially Gabriela Schmajuk, Milena Gianfrancesco, Patti Katz, Laura Trupin, and Julia Kay, it has been a real pleasure to work with all of you. I have learned so much from you over the past years and truly appreciated being a part of this brilliant collaborative group. Thank you for supporting this dissertation and for providing opportunities to work on other exciting projects.

To all my colleagues from the Epidemiology and Translational Science PhD program, especially Kristina Dang, Michelle Roh, Francois Rerolle, Adrienne Epstein, Crystal Langlais, Erika Meza, Alice Guan, Sandeep Brar, Jean Digitale, Sarah Raifman, Eduardo Santiago-Rodriguez, and Shelly DeVost, I have been so lucky to learn with and from you all. Thank you for all your support, input, and constructive feedback. You are all fantastic researchers that I have been impressed with and valued as friends.

This dissertation was not possible without the love and support of my family. To my mother and father, Fariba Gharai and Ali Izadi, thank you for being my rock and my inspiration. I am blessed to have had your presence, your endless love, and your professional guidance during this doctoral journey. To my sister Narges Izadi and brother-in-law, Nima Jafarian, thank you for all your love, support, and generosity; for housing me and lending me your car, for energizing me by teaching me to snowboard and generally making life so fun.

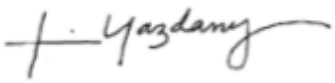
Lastly, I would like to thank the COVID-19 Global Rheumatology Alliance (GRA) Steering Committee in particular Jinoos Yazdany, Philip Robinson, and Pedro Machado, as well as the GRA regional leads and members whose tireless efforts facilitated this research; and my co-authorship and senior authorship, especially Erica Brenner, Satveer Mahil, Milena Gianfrancesco, Anja Strangfeld, Elsa Mateus, Kimme Hyrich, Laure Gossec, Loreto Carmona, Alfredo Aguirre, Leanna Wise, Ali Duarte Garcia, Rebecca Grainger, Jean Liew, Zachary

Wallace, Jeffery Sparks, Catherine Smith, and Michael Kappelman, for all their help, support, and contributions to this work and their dedicated feedback throughout the various stages of the review process.

Acknowledgement of Previously Published Materials

A version of Chapter 1 in this dissertation was published in JAMA Network Open on October 18, 2021. The Dissertation Committee Members supervised the research that forms the basis of this dissertation chapter. The published material is substantially the product of Zahra Izadi's period of study at the University of California, San Francisco and was primary conducted and written by her. The work she completed for this published manuscript is comparable to a standard dissertation chapter.

Approved:



Jinoos Yazdany, MD MPH, Dissertation Chair

COVID-19 Outcomes in People with Rheumatic Disease: Results from a Global Physician-Reported Registry

Zahra Izadi

ABSTRACT

The illness caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) – a novel coronavirus identified in Wuhan at the end of 2019 that led to a global pandemic – primarily manifests as a lung infection with symptoms ranging from those of a mild upper respiratory tract infection to severe pneumonia, acute respiratory distress syndrome (ARDS), and death. Severe illness with coronavirus disease 2019 (COVID-19) can occur in healthy individuals and also in people with underlying medical conditions, including people with rheumatic diseases. Rheumatic diseases involve the dysregulation of the immune system, lead to systemic inflammation of the joints, muscles, bones, and organs, and are more prevalent in older ages.[1,2] People with rheumatic disease have a higher prevalence of several comorbidities (such as pulmonary and kidney disease, heart disease or hypertension, obesity, and diabetes)[3-8] and may be receiving immunosuppressive or immunomodulatory medications which can increase the risk of serious or opportunistic infections.[9,10]

In general, immunosuppression and the presence of comorbidities are associated with an increased risk of serious infection in people with rheumatic diseases.[11] In addition, the impact of previous epidemics caused by coronavirus infections like severe acute respiratory syndrome 1 (SARS-1) and middle east respiratory syndrome (MERS) on patients with rheumatic diseases or other immune-mediated inflammatory diseases (IMIDs) has been scarcely reported.[12] Therefore, in the midst of the pandemic, the implications of COVID-19 for people living with rheumatic

diseases was a considerable concern. In March 2020, to address this knowledge gap, the COVID-19 Global Rheumatology Alliance (GRA) registry was developed by a global network of rheumatologists, scientists, and patient representatives.[13] The GRA registry is a physician-reported registry of people with rheumatic diseases diagnosed with COVID-19 and includes information on patient demographics, rheumatic disease characteristics, immunomodulatory medications used for the treatment of rheumatic disease, and comorbidities, as well as COVID-19 diagnoses, treatments, outcomes, and complications. The overall goal of this dissertation is to use data from the GRA registry to study COVID-19 outcomes in people with rheumatic diseases to advance rheumatology care in the COVID-19 pandemic.

The dissertation is organized into three chapters each describing a key aim of this dissertation. The first chapter pools data from three global COVID-19 registries of individuals with rheumatic diseases, inflammatory bowel disease (IBD), and psoriasis to compare the association between tumor necrosis factor inhibitor (TNFi) monotherapy and COVID-19-related hospitalization or death among individuals with IMIDs, with other commonly prescribed immunomodulatory regimens. Data from the pooled analysis found that TNFi monotherapy was associated with fewer hospitalizations or deaths compared with other immunomodulatory regimens including methotrexate (MTX), azathioprine/mercaptopurine (AZA/6MP), and janus kinase inhibitors (JAKi). In addition, TNFi combination therapy was associated with more favorable outcomes when MTX was used instead of AZA/6MP. These findings support the continued use of TNFi monotherapy during the pandemic and suggest that clinicians should weigh the risks versus benefits of de-escalating treatment or changing medications when a patient is receiving concomitant TNFi and AZA/6MP.

The second chapter describes the development and evaluation of a prediction model for COVID-19 ARDS in people with rheumatic diseases and the development of a simple risk-score calculator for use in clinical settings. The prediction model was developed using a series of supervised machine learning algorithms and information that can be easily obtained at COVID-19 exposure or onset and predicted ARDS with good discrimination in the test set and in external validation sets. Age, daily glucocorticoid dose, pulmonary hypertension, interstitial lung disease, chronic kidney disease, anti-CD20 monoclonal antibody use, diabetes, hypertension, active rheumatic disease, and morbid obesity were identified as the most influential factors in predicting the onset of ARDS. A simple and interpretable regression-based risk-score calculator also predicted ARDS with good discrimination in the test set and in external validation sets. The risk-score calculator has the potential to guide risk-stratification and the treatment of COVID-19 among people with rheumatic diseases during the pandemic.

The third chapter links data from the GRA registry to a robust array of country-level factors and uses a novel methodological approach to investigate potential mechanisms of the disparate impact of COVID-19 on people with rheumatic diseases, globally. Data from this analysis indicated that a range of factors related to geographical residence impacted COVID-19 outcomes independent of known patient-level demographic and clinical risk factors. Namely, lower country socioeconomic status, environmental exposures, higher demands on or lower capacity of health resources, and fewer government-imposed containment measures were independently associated with COVID-19-related death after controlling for patient demographics and clinical characteristics. These findings highlight the importance of environmental and societal factors as potential explanations of the observed global health disparities during the pandemic.

The study designs and analytical methods utilized in this work seek to quantify and address important biases inherent to real-world conveniently sampled data. Together, findings from the three chapters provide important evidence to advance rheumatology care in the COVID-19 pandemic and lay foundation for a new research agenda to address regional disparities in COVID-19 outcomes in people with rheumatic diseases.

Table of Contents

CHAPTER 1 : Tumor Necrosis Factor Inhibitors and the Risk of Hospitalization or Mortality Among Patients with Immune-Mediated Inflammatory Disease and COVID-19. 1

ABSTRACT 1

KEY POINTS 3

INTRODUCTION 4

METHODS 5

RESULTS 10

DISCUSSION 13

CHAPTER 2 : Development of a Prediction Model for COVID-19 Acute Respiratory Distress Syndrome in Patients with Rheumatic Diseases: Results from the Global Rheumatology Alliance Registry 26

ABSTRACT 26

SIGNIFICANCE AND INNOVATIONS 27

INTRODUCTION 28

METHODS 30

RESULTS 35

DISCUSSION 39

CHAPTER 3 : Country-level factors associated with COVID-19-related death in people with rheumatic disease: results from the COVID-19 Global Rheumatology Alliance Registry 59

ABSTRACT 59

RESEARCH IN CONTEXT 61

INTRODUCTION 62

METHODS 64

RESULTS 68

DISCUSSION 70

REFERENCES 90

List of Figures

Figure 1.1. Equation for inverse odds sampling weights and standardized difference..... 17

Figure 1.2. Adjusted odds of COVID-19 related hospitalization or death for immunomodulatory treatment regimens compared with tumor necrosis factor inhibitor monotherapy in registry-specific and pooled analyses. 18

Figure 1.3. E-value analysis of potential unmeasured confounding of the pooled estimates of association..... 19

Figure 2.1. Dataset partitioning into training, test, and validation sets..... 44

Figure 2.2. The Index of Union method for classification threshold optimization..... 45

Figure 2.3. Gradient boosting machine predictor importance plot. 46

Figure 2.4. Adjusted odds ratios obtained from the multivariable logistic regression model. 47

Figure 2.5. Calibration plot for the multivariable regression model showing level of agreement between observed and predicted risk in the test and validation sets. 48

Figure 2.6. The risk-score calculator pocket care side 1..... 49

Figure 2.7. The risk-score calculator pocket care side 2..... 50

Figure 3.1. Study timeline..... 75

Figure 3.2. Directed acyclic graph showing conceptualized relationships between outcome, individual-level factors, and the regional covariates. 76

Figure 3.3. Number of deaths reported and case fatality rate by country among patients included in the analyses. 77

Figure 3.4. Temporal trends in COVID-19 measures and containment index across six global regions..... 78

Figure 3.5. Temporal trends in population mobility across six global regions..... 79

Figure 3.6. Temporal trends in climatic factors across six global regions..... 80

Figure 3.7. Associations between regional characteristics and COVID-19-related death, independent of individual-level demographics and clinical risk factors. 81

Figure 3.8. Odds ratios of COVID-19-related death corresponding to individual-level characteristics, after controlling for regional and temporal confounding..... 82

List of Tables

Table 1.1. Details of patients excluded from the study. 20

Table 1.2. Patient and clinical characteristics of the study population and COVID-19 outcomes. 21

Table 1.3. Adjusted pooled odds of COVID-19 related hospitalization or death using data from the three registries. 23

Table 1.4. Adjusted registry-specific odds of COVID-19-related hospitalization or death. 24

Table 1.5. Sensitivity analysis to determine the magnitude of potential reporting-bias using data from the GRA registry 25

Table 2.1. Distribution of unknown or missing data across variables and datasets. 51

Table 2.2. List of baseline predictors used in machine learning classifiers. 52

Table 2.3. Demographic and clinical characteristics of the study population 54

Table 2.4. Predictive performance of machine learning classifiers in the test set. 56

Table 2.5. Predictive performance of the gradient boosting machine in the test set and across validation sets 57

Table 2.6. Predictive performance of the multivariable logistic regression model in the test set and across validation sets. 58

Table 3.1. Countries and U.S. states included in the analyses. 83

Table 3.2. Regional covariate definitions and source datasets 84

Table 3.3. Patient characteristics grouped into six global regions. 87

Table 3.4. Baseline regional characteristics grouped into six global regions. 88

Table 3.5. Inclusion of temporal and regional covariates as fixed effects and corresponding shrinkage in intraclass correlation coefficients. 89

List of Abbreviations

ACR	American college of rheumatology
API	application programming interface
ARDS	acute respiratory distress syndrome
AUC	area under curve
AZA/6MP	azathioprine or mercaptopurine
b/tsDMARDs	biologic or targeted synthetic disease modifying antirheumatic drugs
BAYESGLM	Bayesian generalized linear models
BMI	body mass index
CARET	the classification and regression training package
CD	Crohn's disease
CFR	case fatality rate
CI	confidence interval
COPD	chronic obstructive pulmonary disease
COVID-19	coronavirus disease 2019
csDMARDs	conventional synthetic disease modifying antirheumatic drugs
CSSE	center for systems science and engineering
CT scan	computed tomography scan
DAG	directed acyclic graph

DMARDs	disease modifying antirheumatic drugs
EHR	electronic health record
EULAR	European alliance of associations for rheumatology
GAM	generalized additive models
GBM	gradient boosting machine
GC	glucocorticoids
GDP	gross domestic product
GLMNET	the lasso and elastic-net regularized generalized linear models
GRA	the COVID-19 global rheumatology alliance
HIPAA	health insurance portability and accountability act
IBD	inflammatory bowel disease
ICC	intra-class correlation coefficient
ICIs	integrated calibration indices
ICU	intensive care unit
IMIDs	immune-mediated inflammatory diseases
IRB	institutional review board
IU	index of union
JAKi	janus kinase inhibitors
KNN	k-nearest neighbors

MERS	middle east respiratory syndrome
MTX	methotrexate
NNET	neural networks
OR	odds ratio
PCR	polymerase chain reaction
PM	particulate matter
PsA	psoriatic arthritis
PsoProtect	psoriasis patient registry for outcomes, therapy and epidemiology of COVID-19 infection
RA	rheumatoid arthritis
RR	risk ratio
SARS-1	severe acute respiratory syndrome 1
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SD	standard deviation
SECURE-IBD	secure epidemiology of coronavirus under research exclusion for inflammatory bowel disease
SLE	systemic lupus erythematosus
SpA	spondyloarthritis
STROBE	strengthening the reporting of observational studies in epidemiology

SVM	support vector machines
TNFi	tumor necrosis factor inhibitor
TRIPOD	transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
UC	ulcerative colitis
WHO	world health organization

CHAPTER 1 : Tumor Necrosis Factor Inhibitors and the Risk of Hospitalization or Mortality Among Patients with Immune-Mediated Inflammatory Disease and COVID-19

Zara Izadi, Erica J Brenner, Satveer K Mahil, Nick Dand, Zenas Z N Yiu, Mark Yates, Ryan C Ungaro, Xian Zhang, Manasi Agrawal, Jean-Frederic Colombel, Milena A Gianfrancesco, Kimme L Hyrich, Anja Strangfeld, Loreto Carmona, Elsa F Mateus, Saskia Lawson-Tovey, Eva Klingberg, Giovanna Cuomo, Marta Caprioli, Ana Rita Cruz-Machado, Ana Carolina Mazeda Pereira, Rebecca Hasseli, Alexander Pfeil, Hanns-Martin Lorenz, Bimba Franziska Hoyer, Laura Trupin, Stephanie Rush, Patricia Katz, Gabriela Schmajuk, Lindsay Jacobsohn, Andrea M Seet, Samar Al Emadi, Leanna Wise, Emily L Gilbert, Alí Duarte-García, Maria O Valenzuela-Almada, Carolina A Isnardi, Rosana Quintana, Enrique R Soriano, Tiffany Y-T Hsu, Kristin M D'Silva, Jeffrey A Sparks, Naomi J Patel, Ricardo Machado Xavier, Claudia Diniz Lopes Marques, Adriana Maria Kakehasi, René-Marc Flipo, Pascal Claudepierre, Alain Cantagrel, Philippe Goupille, Zachary S Wallace, Suleman Bhana, Wendy Costello, Rebecca Grainger, Jonathan S Hausmann, Jean W Liew, Emily Sirotich, Paul Sufka, Philip C Robinson, Pedro M Machado, Christopher EM Griffiths, Jonathan N Barker, Catherine H Smith, Jinoos Yazdany, Michael D Kappelman on behalf of Psoriasis Patient Registry for Outcomes, Therapy and Epidemiology of COVID-19 Infection (PsoProtect), Secure Epidemiology of Coronavirus Under Research Exclusion for Inflammatory Bowel Disease (SECURE-IBD), and COVID-19 Global Rheumatology Alliance (GRA).

ABSTRACT

Importance: While tumor necrosis factor inhibitors (TNFi) are widely prescribed globally due to their high efficacy at ameliorating shared immune pathways across immune-mediated

inflammatory diseases (IMIDs), the impact of COVID-19 on individuals with IMIDs receiving TNFi remains poorly understood.

Objective: To compare the association between TNFi monotherapy and COVID-19-related hospitalization or death among individuals with IMIDs, with other commonly prescribed immunomodulatory regimens.

Design: Pooled analysis using data from three global COVID-19 registries of individuals with rheumatic diseases, inflammatory bowel disease (IBD), and psoriasis, from March 12, 2020, to February 1, 2021.

Setting: Clinicians directly reported COVID-19 outcomes as well as demographic and clinical characteristics of individuals with IMIDs diagnosed with confirmed or suspected COVID-19 using online data entry portals.

Participants: Adults (≥18 years) with a diagnosis of inflammatory arthritis, IBD, or psoriasis.

Exposure(s): Exposure categories included: tumor necrosis factor inhibitor (TNFi) monotherapy (reference), TNFi in combination with methotrexate, TNFi in combination with azathioprine/mercaptopurine, methotrexate monotherapy, azathioprine/mercaptopurine monotherapy, janus kinase inhibitor (JAKi) monotherapy.

Main outcome(s) and Measure(s): COVID-19-related hospitalization or death. Registry-level analyses and a pooled analysis of data across the registries were conducted using multilevel multivariable logistic regression, adjusting for demographics and clinical characteristics and accounting for country, calendar month and registry-level correlations.

Results: A total of 6,077 patients from 74 countries were included. Mean (SD) age was 48.8 (16.5) years and 58.6% were female. The most common IMID diagnoses were rheumatoid arthritis (35.3%) and Crohn's disease (25.3%). A total of 1,297 patients were hospitalized and 189 died. In the pooled analysis, compared with TNFi monotherapy, higher odds of hospitalization or death were observed with TNFi in combination with azathioprine/mercaptopurine (odds ratio: 1.74, 95% CI: 1.17-2.58), azathioprine/mercaptopurine monotherapy (1.84, 1.30-2.61), methotrexate monotherapy (2.0, 1.57-2.56), and JAKi monotherapy (1.82, 1.21-2.73), but not with TNFi in combination with methotrexate (1.18, 0.85-1.63). Similar findings were obtained in analyses that accounted for potential reporting bias and after excluding COVID-19 diagnoses based on symptoms alone.

Conclusions and Relevance: In this cohort study, among individuals with IMIDs, TNFi monotherapy was associated with a lower risk of adverse COVID-19 outcomes compared with other commonly prescribed immunomodulatory regimens.

KEY POINTS

Question: Compared with other immunomodulatory regimens, is tumor necrosis factor inhibitor (TNFi) monotherapy used at the time of COVID-19 diagnosis associated with adverse COVID-19 outcomes in patients with immune-mediated inflammatory disease?

Findings: In a pooled analysis of 6,077 patients diagnosed with COVID-19, TNFi combination therapy with azathioprine/mercaptopurine (AZA/6MP), methotrexate monotherapy, AZA/6MP monotherapy, and janus kinase inhibitor monotherapy were each associated with statistically significantly higher odds of hospitalization or death, compared with TNFi monotherapy.

Meaning: These findings support the continued use of TNFi monotherapy among individuals with immune-mediated inflammatory disease during the pandemic.

INTRODUCTION

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), can have mild symptoms or present as a severe and/or life-threatening infection.[14] Comorbidities such as lung disease, diabetes, and obesity increase the risk for adverse COVID-19 outcomes.[11] Any influence of treatments for immune-mediated inflammatory diseases (IMIDs) on COVID-19 outcomes remains a topic of interest. These treatments impact the immune system and are associated with a higher risk of infections overall.[15] This raises the concern of impaired immune response to SARS-CoV2. However, many damaging effects of SARS-CoV-2 are due to a hyperinflammatory response.[16] As such, treatments that target an overactive immune response may have a protective effect against adverse COVID-19 outcomes.[14,16]

Tumor necrosis factor inhibitors (TNFi), a class of biological therapies that target the pro-inflammatory cytokine TNF, are first- or second-line treatments for many IMIDs. International registries of patients with IMIDs have provided initial information regarding COVID-19 outcomes in individuals using TNFi during the pandemic. The Secure Epidemiology of Coronavirus Under Research Exclusion for Inflammatory Bowel Disease (SECURE-IBD) registry, a database of patients with inflammatory bowel disease (IBD) who contract COVID-19, found that prevalent TNFi use at COVID-19 diagnosis, compared to no use, was not associated with severe COVID-19 (OR 0.9, 95% CI 0.4-2.2).[17] The COVID-19 Global Rheumatology Alliance (GRA) physician-reported registry of COVID-19 outcomes in people with rheumatic diseases, found that prevalent TNFi use at COVID-19 diagnosis, compared to no use, was

associated with a lower odds of COVID-19-related hospitalization (OR 0.40, 95% CI 0.19-0.81).[18] Similarly, the Psoriasis Patient Registry for Outcomes, Therapy and Epidemiology of COVID-19 Infection (PsoProtect), found higher odds of hospitalization among patients treated with non-biologic systemic therapies compared to biologics, including TNFi (OR 2.84; 95% CI 1.31-6.18).[19] Although studies from individual registries provide initial information, they are often underpowered for more granular analyses of commonly used medications, such as analyses of monotherapy versus combination immunomodulatory drug use, and for analyses of medications that are used less frequently.

Pooling data across registries offers a unique opportunity to rapidly assess any association between TNFi and COVID-19 outcomes among individuals with IMIDs, and to evaluate consistency of findings across studies and diseases. We pooled data from three international registries of patients with IBD, psoriasis, and rheumatic diseases to evaluate associations between TNFi monotherapy and COVID-19-related hospitalization or death, compared with other commonly prescribed immunomodulatory regimens across individuals with IMIDs.

METHODS

Registry designs. Details of the GRA, SECURE-IBD, and PsoProtect registry designs have been described previously[13,17,19-21]. Briefly, clinicians and trained staff directly report COVID-19 outcomes as well as demographic and clinical characteristics of individuals with IMIDs diagnosed with confirmed or suspected COVID-19 using online data entry portals. Quality is assessed by registry-specific data validation teams who remove all known or potential duplicates and address erroneous or ineligible reports. GRA and PsoProtect involve only limited data; no personal identifiers except COVID-19 diagnosis dates are included. SECURE-IBD is in accordance with HIPAA Safe Harbor De-Identification standards. GRA was determined ‘not

human subjects research' by the UK Health Research Authority and the University of Manchester, as well as under United States Federal Guidelines assessed by the University of California, San Francisco and patient consent was not required. For SECURE-IBD, the UNC-Chapel Hill Office for Human Research Ethics has determined that storage and analysis of de-identified data does not constitute human subjects research and does not require IRB approval. Voluntary ethical approval was sought by PsoProtect and granted by the Leeds Research Ethics Committee (ref 20/YH/0135) and patient consent was not required. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

COVID-19 diagnosis. In patients with rheumatic disease, COVID-19 diagnosis was based on polymerase chain reaction (PCR), antibody serology, metagenomic testing, CT scan, laboratory assay, or based on symptoms alone. In patients with IBD, COVID-19 diagnosis was based on PCR, symptoms with confirmatory antibody serology, or rapid antigen testing. In patients with psoriasis, both confirmed and suspected COVID-19 were reported, however, information on the type of diagnostic tests was not collected.

Exposure. To obtain sufficient statistical power, each exposure category was required to have more than 250 patients in the pooled analysis. Consequently, exposure was defined as a categorical variable with the following categories: TNFi (includes adalimumab, certolizumab pegol, etanercept, golimumab, and infliximab) monotherapy (reference), TNFi in combination with methotrexate (MTX), TNFi in combination with azathioprine/mercaptopurine (AZA/6MP), MTX monotherapy, AZA/6MP monotherapy, and janus kinase inhibitor (JAKi, includes tofacitinib, baricitinib, and upadacitinib) monotherapy.

Outcome. The outcome of interest was hospitalization or death due to COVID-19.

Inclusion criteria. We included adults (≥18 years) with a diagnosis of IBD, inflammatory arthritis (IA), or psoriasis reported to the SECURE-IBD, GRA, or PsoProtect registries, respectively, on or before February 1st, 2021. Our analysis included reconciled patients only. In GRA, reconciled was defined as having at least one of the following outcomes: deceased; symptoms resolved at the time of data entry; not hospitalized >30 days after initial diagnosis date; hospitalized and discharged; or not at risk of further interventions/death. In SECURE-IBD and PsoProtect, a case was defined as reconciled after a minimum of 7 days (14 days in PsoProtect) or if sufficient time had passed to observe the disease course through the resolution of acute illness or death.

Exclusion criteria. To limit confounding by other immunomodulatory medications, we excluded patients in which exposure categories were used with concomitant drugs other than sulfasalazine, mesalamine, hydroxychloroquine or chloroquine, leflunomide, oral budesonide, or glucocorticoids.

Statistical analysis. We used descriptive statistics to summarize the demographic and clinical characteristics of the study population. Continuous variables were reported as mean (SD) or median (25th and 75th percentile) as appropriate. Categorical variables were reported as number and percentage (%). We performed registry-level analyses and a pooled analysis of data across the three registries to estimate independent associations between exposure categories and COVID-19 outcomes. Registry-level effect estimates were reported for exposure categories that had ≥10 patients. Associations were estimated using multilevel multivariable mixed-effects logistic regression and reported as odds ratios (OR) with 95% confidence intervals (CIs). We chose mixed-effects regression for its ability to handle missing data using maximum-likelihood estimation and fit random effects to account for multilevel clustering[22].

Covariates included in all models were age, sex, current tobacco use, IMID activity (remission vs. active disease, as reported by the clinician), key comorbidities (cardiovascular disease [including coronary artery disease, heart failure, arrhythmia], diabetes, hypertension, obstructive lung disease [including COPD, asthma], interstitial or other chronic lung disease, kidney disease [including chronic renal insufficiency, end stage renal disease], obesity [BMI $\geq$30 kg/m^2], and cancer, each included as a dichotomous variable), and prednisone-equivalent glucocorticoid dose included as a continuous variable. For the registry-level analyses, we included other concomitant medications as follows: sulfasalazine, hydroxychloroquine or chloroquine, and leflunomide for the GRA analysis; mesalamine, sulfasalazine, and oral budesonide for the SECURE-IBD analysis. If any of these medications were significant confounders ($p<0.05$) in the registry-level analysis, we also included them as a covariate in the pooled analysis and assigned patients from registries that did not include the respective medication to non-use of that medication. Registry-level analyses also controlled for disease diagnosis (rheumatoid arthritis [RA, reference], psoriatic arthritis [PsA], spondyloarthritis [SpA], and "other or more than one type of inflammatory arthritis", and Crohn's disease [CD, reference], ulcerative colitis [UC], and unspecified IBD, for the GRA and SECURE-IBD registries, respectively).

We fitted country-level random effects to account for within-country correlations. To account for changes in COVID-19 treatment and propensity for hospitalization over time, we also fitted random effects for the calendar-month during which the case was diagnosed (GRA), reported (SECURE-IBD), or at onset (PsoProtect). The pooled model additionally included registry-level random effects accounting for within-registry correlations. The hierarchical ordering of random effects in the pooled model was country, followed by time and registry. To improve model fit, we removed influential statistical outliers identified in continuous variables (age and

glucocorticoid dose) from the analyses. As a result, two patients were removed who received a daily prednisone dose greater than 70mg. All analyses were conducted using Stata software, version 16.0 (StataCorp). The threshold for statistical significance was 2-sided $P < .05$.

Sensitivity analyses. Rheumatology clinics from two large health-systems (Mass General Brigham in Massachusetts and Mayo Clinics in Minnesota and Florida, USA) had processes in place to systematically identify all patients with COVID-19, regardless of COVID-19 severity. To assess the magnitude of potential reporting-bias arising from convenience sampling, ORs were derived after reweighting the covariate distribution of GRA patients to those of the aforementioned health-systems using the inverse-odds sampling-weight technique (see formula in Figure 1.1).[23] Reweighted log ORs were then compared with original log ORs using standardized difference (see formula in Figure 1.1).[24] In order to obtain reliable standard errors, both the original and reweighted estimates were obtained from logistic regression models that incorporated both country and calendar-month as fixed effects. To ensure country-level confounding was sufficiently accounted for, we fitted an indicator variable for each of the top 18 largest countries included in the analyses. A likelihood ratio test was performed to assess the equivalence between the logistic regression model and a mixed-effects regression model that incorporated all countries as random effects ($p=0.067$). All missing values were imputed using 50 repeats of multiple imputation by chained equations.

In order to assess bias due to potential unmeasured confounding, we derived e-values corresponding to estimates of association from our pooled analysis. Given hospitalization or death was a common outcome (>15%), square root of the OR was used to approximate risk ratio.[25]

In addition, to assess the robustness of the results, a pooled sensitivity analysis was performed after excluding COVID-19 diagnoses based on symptoms alone.

RESULTS

As of February 1, 2021, 8,268 patients were reported to be receiving an exposure regimen at COVID-19 diagnosis: GRA (5,220), SECURE-IBD (2,720) and PsoProtect (328). A total of 6,077 patients, from 74 countries, met study eligibility criteria (Table 1.1) and were included in the analyses; of these, 56.6%, 38.4%, and 4.9% were from the GRA, SECURE-IBD and PsoProtect registries, respectively. Of the 2191 patients excluded from the analyses, most were excluded because they had a rheumatic disease diagnosis other than inflammatory arthritis (827 patients), were patients that were nonreconciled (581 patients) or received concomitant medications that were listed in the exclusion criteria (551 patients) (Table 1.1).

The demographic and clinical characteristics of the 6,077 patients analyzed are shown in Table 1.2. Most patients were from Europe 3,215 (52.9%) and North America 2,015 (33.2%), mean (SD) age was 48.8 (16.5) years and 3,563 (58.6%) were female. The most common disease diagnoses were RA (2,146, 35.3%), CD (1,537, 25.3%), UC (762, 12.5%), and SpA (624, 10.3%). The most common comorbidities were hypertension (1,360, 22.4%), diabetes (541, 8.9%), obstructive lung disease (430, 7.1%), and cardiovascular disease (388, 6.4%). Current smoking and obesity were substantially more prevalent among patients in the PsoProtect registry (14.0% and 30.7%, respectively) compared with those in the GRA registry (4.4% and 19.6%, respectively) or the SECURE-IBD registry (4.3% and 16.5%, respectively).

TNFi monotherapy was reported in 1,183 (34.4%), 1,445 (61.9%) and 216 (72%) of patients from the GRA, SECURE-IBD and PsoProtect registries, respectively (Table 1.2). MTX

monotherapy was the most prevalent therapy at COVID-19 diagnosis in patients from the GRA registry (1,438, 41.8%). AZA/6MP use, alone or in combination with a TNFi, was reported in a minority of GRA patients (26, 0.8%) and in none of the PsoProtect patients. JAKi monotherapy was reported in 219 (6.4%) and 67 (2.9%) patients from the GRA and SECURE-IBD registries, respectively, and in none of the PsoProtect patients. About one-fifth (1,297, 21.3%) of all patients included in the analyses were hospitalized and 3.1% (189) died (Table 1.2). Both hospitalization and death were more common among patients included from the GRA registry than the SECURE-IBD registry or the PsoProtect registry.

Along with the prespecified covariates, sulfasalazine, leflunomide and oral budesonide were included in the pooled multivariable model because these medications were statistically significantly associated with hospitalization or death in the GRA and SECURE-IBD registry-level analyses. In the pooled analysis, compared with TNFi monotherapy, higher odds of hospitalization or death were observed with TNFi in combination with AZA/6MP (OR 1.74, 95% CI 1.17-2.58). Differences in the odds of hospitalization or death between TNFi monotherapy and TNFi in combination with MTX were not statistically significance in registry-specific or pooled analyses. Compared with TNFi monotherapy, higher odds of hospitalization or death were observed with MTX monotherapy (OR 2.0, 95% CI 1.57-2.66), AZA/6MP monotherapy (OR 1.84, 95% CI 1.30-2.61), and JAKi monotherapy (OR 1.82, 95% CI 1.21-2.73) in the pooled analyses.

Although ORs obtained from registry-specific analyses were generally in the same direction and of similar magnitude as those obtained from the pooled analysis, we observed some notable differences (Figure 1.2): ORs for MTX monotherapy compared with TNFi monotherapy were larger among patients in the PsoProtect registry than patients in the SECURE-IBD or the GRA

registries. ORs for AZA/6MP monotherapy compared with TNFi monotherapy were larger among patients in the GRA registry than patients in the SECURE-IBD registry. In addition, Jaki monotherapy was not associated with higher odds of hospitalization or death compared with TNFi monotherapy (OR, 0.60; 95% CI, 0.22-1.64) among patients in the SECURE-IBD registry.

Other factors associated with higher odds of hospitalization or death in the pooled analysis included older age (OR per 1 year increase in age, 1.04; 95% CI, 1.04-1.05); active IMID at COVID-19 diagnosis (OR, 1.27; 95% CI, 1.04-1.55); obesity (OR, 1.39; 95% CI, 1.10-1.75); lung disease (interstitial: OR, 1.81 [95% CI, 1.12-2.95]; obstructive: OR, 2.34 [95% CI, 1.69-3.24]); cardiovascular disease (OR, 1.58; 95% CI, 1.13-2.21); diabetes (OR, 1.54; 95% CI, 1.16-2.05); chronic kidney disease (OR, 3.10; 95% CI, 1.70-5.66); concomitant use of sulfasalazine (OR, 1.62; 95% CI, 1.13-2.34), leflunomide (OR, 1.89; 95% CI, 1.20-2.99), or oral budesonide (OR, 2.86; 95% CI, 1.20-6.84); and higher daily prednisone-equivalent glucocorticoid dose (OR per 1 mg increase in dose, 1.07; 95% CI, 1.05-1.08) (Table 1.3). Female sex was associated with a protective benefit (OR, 0.79; 95% CI, 0.66-0.96). The intraclass correlation coefficient was 0.27 (95% CI, 0.20-0.36), suggesting that clustering of patients within country, calendar month, and registry explained 27% of the variation in the odds of hospitalization or death. Complete results from registry-specific analyses are shown in Table 1.4.

We compared GRA registry-specific results with results obtained after reweighting covariate distribution of the GRA population to those of rheumatology clinics that systematically reported all COVID-19 patients. Standardized differences were in the acceptable range of less than <0.1[26] (-0.004 for log OR corresponding to TNFi in combination with MTX compared with TNFi monotherapy; 0.019 for log OR corresponding to MTX monotherapy compared with TNFi monotherapy; -0.007 for log OR corresponding to Jaki monotherapy compared with TNFi

monotherapy), indicating reporting-bias did not substantially impact estimates of association in the GRA registry (Table 1.5).

E-values for treatment regimens ranged from 1.39 for TNFi in combination with MTX to 2.18 for MTX monotherapy (Figure 1.3). On the risk ratio (RR) scale, kidney disease which was the measured covariate with the strongest association with hospitalization or death, had an approximate RR of 1.76. This suggests that confounding of the estimates of association for TNFi in combination with AZA/6MP, AZA/6MP monotherapy, and JAKi monotherapy is plausible but unlikely to explain away all of the associations observed; and confounding of the estimate of association for MTX monotherapy is unlikely.

Over one-third of patients (112, 37.3%) from the PsoProtect registry, 23.4% (824) of patients from the GRA registry, and none of the patients from the SECURE-IBD registry had a COVID-19 diagnosis based on symptoms alone. Our pooled results remained consistent in a sensitivity analysis that excluded these patients (Figure 1.2).

DISCUSSION

We found that TNFi monotherapy was associated with a lower risk of COVID-19 related hospitalization or death among patients with IMIDs when compared to other commonly used regimens, including MTX, AZA/6MP, and JAKi. After controlling for active disease and common comorbidities, the odds of hospitalization or death with TNFi combination therapy versus TNFi monotherapy depended on the additional medication used. Patients on TNFi and AZA/6MP had higher odds of hospitalization or death compared with TNFi monotherapy, while individuals using TNFi with MTX had similar odds of hospitalization or death compared with TNFi alone.

The lower odds of poor COVID-19 outcomes with pre-infection TNFi use has several possible explanations. While the exact mechanism of SARS-CoV-2-related hyperinflammation remains uncertain, high serum TNF concentrations at the time of COVID-19 admission have been associated with organ damage and poor COVID-19 outcomes.[27] As such, blocking TNF could inhibit this detrimental immune response. Multiple case series showing favorable outcomes of patients receiving TNFi treatment support this assertion.[14,28,29] Upcoming results from clinical trials investigating the use of TNFi enable further evaluation of the effect of TNFi on COVID-19 outcomes.[30,31]

Other possible explanations for our findings include the effects of non-TNFi immunosuppressive medications on COVID-19 outcomes. Thiopurine medications are associated with a higher risk of opportunistic viral infections.[32-34] A large registry of patients with IBD found that use of thiopurines including AZA and 6MP was associated with a higher risk of serious viral infection, specifically species of the *Herpesviridae*.[35] Although data relating to other viruses cannot be directly extrapolated to COVID-19, this highlights the potential for thiopurine use to increase risk of poor SARS-CoV-2 infection outcomes. Moreover, a recent study from the SECURE-IBD database showed that thiopurine monotherapy and combination thiopurines with TNFi was associated with worse COVID-19 outcomes compared to TNFi monotherapy.[36] In contrast, researchers have postulated that MTX may decrease the cytokine storm associated with COVID-19.[37,38] However, our results suggest worse outcomes associated with MTX monotherapy than with TNFi monotherapy. This could mean either that TNFi therapy is exerting a protective effect, or that MTX is exerting a harmful effect. Notably, the direction of effect was the same for MTX used in combination with TNFi, although effect estimate crossed the line of no effect, possibly relating to the use of lower MTX doses for combination therapy (compared to monotherapy).[39,40]

Timing of treatment initiation with JAKi may be an important factor in influencing COVID-19 outcomes. The second iteration of the Adaptive COVID-19 Treatment Trial (ACTT-2) suggests a protective effect of treatment with baricitinib with remdesivir against poor COVID-19 outcomes in some subgroups of patients with established severe COVID-19.[41] However, population-based data from patients using JAKi prior to COVID-19 suggest worse outcomes, which is consistent with the known effect of this class of medications on reducing the innate immune response leading to impaired viral clearance.[42] In our comparative analyses, we found that JAKi monotherapy was associated with higher odds of hospitalization or death than TNFi monotherapy.

Strengths of this study include the robust, worldwide collaboration between three international registries that enabled evaluation of a large, geographically diverse sample of adults with IMIDs. To our knowledge, this is the first study pooling data across registries evaluating COVID-19 outcomes in patients with IMIDs. Pooling data increased power of the study, allowed for more granular medication analyses, and improved generalizability across IMIDs. Importantly, our analyses controlled for active disease, something that is only possible with registry data as this variable is not typically available in administrative data or electronic health records. Furthermore, reporting to each registry occurred directly by clinicians or trained staff, which likely increased the accuracy of the information.

Limitations of the work include the risk of reporting-bias, as registries used convenience sampling. Lack of a global COVID-19 registration system limited the feasibility of including a control group. Results from our e-value analysis suggested that some confounding of the observed estimates of association is plausible. The threshold for hospitalization and how patients are treated for COVID-19 differs over time and across regions. Such differences have the

potential to introduce bias if insufficiently accounted for in the analyses. While we attempted to account for correlations in hospitalization or death due to unmeasured temporal and geographical factors, residual confounding may remain. Additionally, we were unable to account for patients' socioeconomic status, or the duration and previous lines of IMID therapy. Lastly, although the case report forms were similar, the data domains across registries were not entirely uniform. Specifically, time, disease activity, and certain comorbidities were recorded slightly differently across registries. Our efforts in harmonizing data across the registries may have led to loss of information and residual confounding of the analyses.

In summary, our results suggest that among patients with IMIDs, TNFi monotherapy is associated with a lower risk of COVID-19-related hospitalization or death compared with other immunomodulatory regimens. These findings support the continued use of TNFi monotherapy during the pandemic and call for further research investigating the effect of other biologics on COVID-19 outcomes. TNFi combination therapy was associated with a more favorable safety profile when MTX was used instead of AZA/6MP, suggesting that clinicians should weigh the risks versus benefits of de-escalating treatment or changing medications when a patient is receiving concomitant TNFi and AZA/6MP.

A

$$W_i(X_i, S_i) = \begin{cases} \dfrac{1 - p(Xi;\beta)}{p(Xi;\beta)}, & S_i = 1, \\ 0, & S_i = 0 \end{cases}$$

- W_i is the weight assigned to each patient,
- X_i is the vector of covariates for each patient,
- S_i represents sampling; $S_i = 0$ for patients from practices that systematically reported all COVID-19 diagnoses, and $S_i = 1$ for patients from all other practices,
- $p(Xi;\beta)$ is an estimator for $P(S_i = 1|X_i)$.

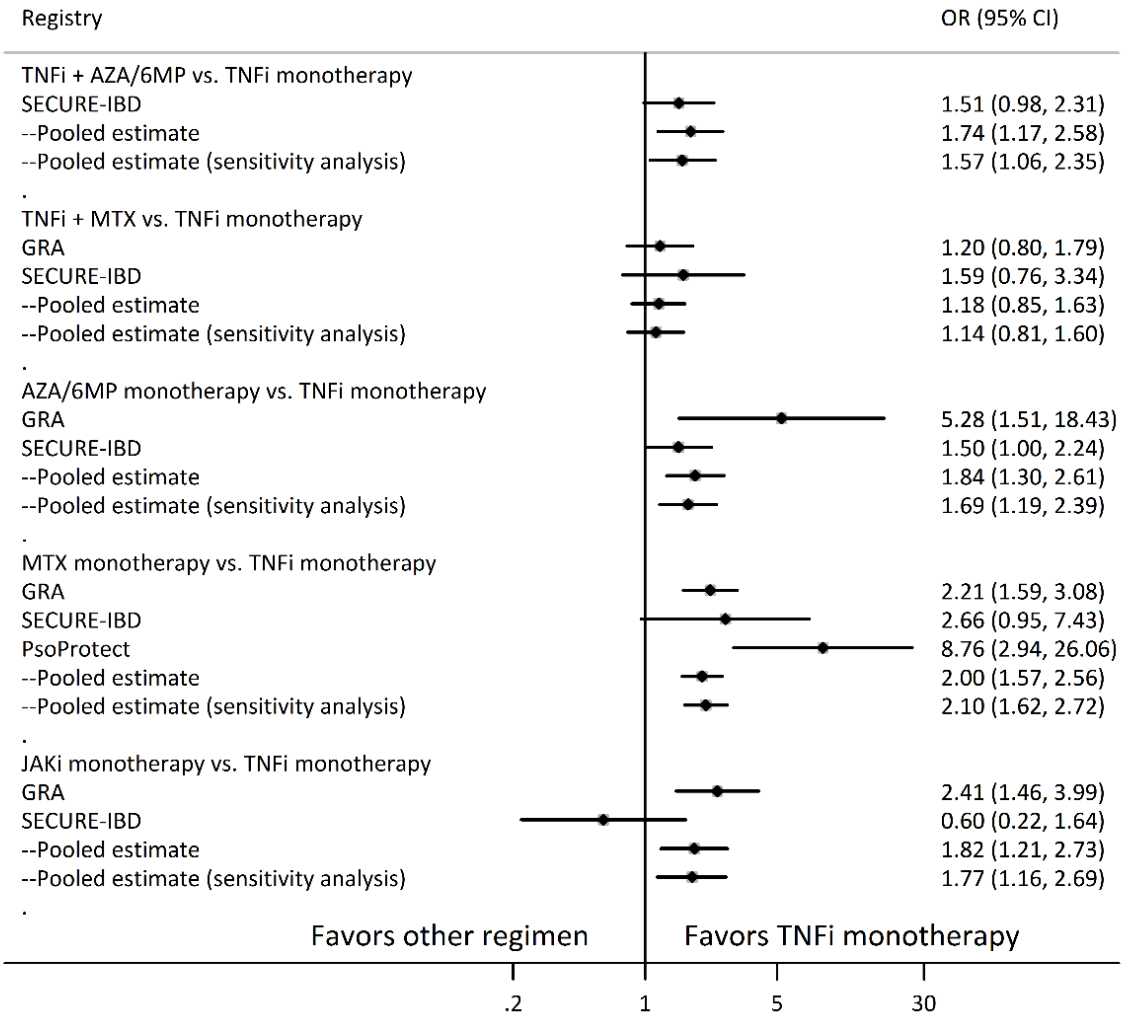
B

$$Z = \frac{\hat{\theta}_{\text{model 'a'}} - \hat{\theta}_{\text{reference model}}}{\sqrt{\hat{\sigma}^2_{\text{model 'a'}} + \hat{\sigma}^2_{\text{reference model}}}}$$

Where the $\hat{\theta}$ are log odds ratios, and the $\hat{\sigma}^2$ are associated variances.

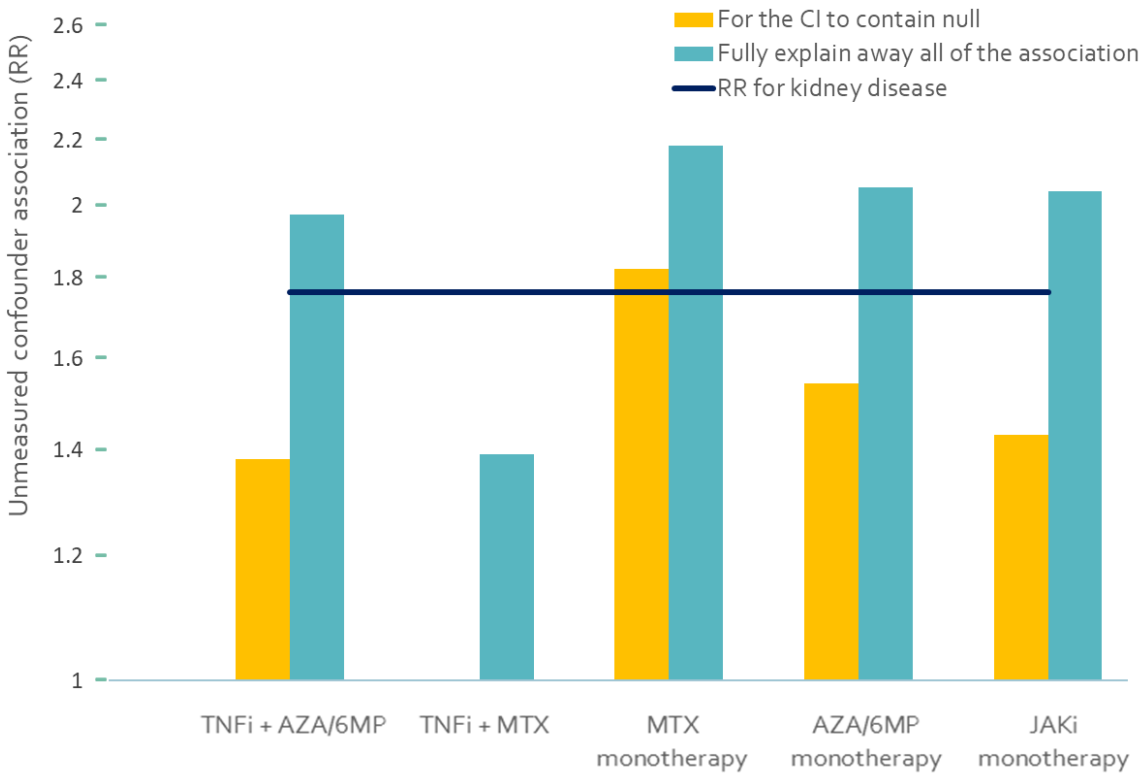
A: Inverse odds sampling weights; B: Standardized difference.

Figure 1.1. Equation for inverse odds sampling weights and standardized difference.



TNFi monotherapy is the reference category. Pooled estimates obtained using hierarchical multivariable mixed-effects logistic regression with registry, calendar-month, and country random effects. Pooled sensitivity analysis excludes patients with a presumptive diagnosis (based on symptoms alone) of COVID-19. All odds ratios adjusted for the following: age, sex, current tobacco use, immune-mediated disease activity (remission vs. active disease), key comorbidities (cardiovascular disease, diabetes, hypertension, obstructive lung disease, interstitial or other chronic lung disease, kidney disease, obesity [BMI ≥30 kg/m2], and cancer), and prednisone-equivalent glucocorticoid dose. Pooled analysis additionally adjusted for sulfasalazine, leflunomide, and oral budesonide. Pooled sensitivity analysis additionally adjusted for leflunomide and oral budesonide. GRA registry-level analyses additionally adjusted for sulfasalazine, hydroxychloroquine, or chloroquine, leflunomide, and immune-mediated disease diagnosis. SECURE-IBD registry-level analyses additionally adjusted for mesalamine, sulfasalazine, oral budesonide, and immune-mediated disease diagnosis. Abbreviations: MTX: methotrexate; TNFi: tumor necrosis factor inhibitor; AZA/6MP: azathioprine/mercaptopurine; JAKi: janus kinase inhibitor. N = 3,441 (GRA); 2,336 (SECURE-IBD); 300 (PsoProtect); 6,077 (Pooled); 5,213 (Pooled, sensitivity analysis).

Figure 1.2. Adjusted odds of COVID-19 related hospitalization or death for immunomodulatory treatment regimens compared with tumor necrosis factor inhibitor monotherapy in registry-specific and pooled analyses.



The E-value is the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to fully explain away a specific treatment–outcome association. The x-axis includes the estimates of association for exposure regimens obtained from the primary pooled analysis (with TNFi monotherapy as the reference category). The yellow bars represent magnitude of the association of the unmeasured confounder (on the RR scale) that is required to make the statistical significance of each observed association non-significant. The blue bars represent the magnitude of the association of the unmeasured confounder that is required to fully explain away each of the observed association. The solid blue line represents the approximate risk ratio for kidney disease (RR=1.76) – the measured covariate with the strongest association with hospitalization or death. RR: risk ratio – RR estimates were approximated by the square root of the odds ratio. JAKi: janus kinase inhibitors; MTX: methotrexate; AZA: azathioprine; 6MP: 6-mercaptopurine; TNFi: tumor necrosis factor inhibitor.

Figure 1.3. E-value analysis of potential unmeasured confounding of the pooled estimates of association.

Table 1.1. Details of patients excluded from the study.

	GRA	SECURE-IBD	PsoProtect	All patients
	N			
Total patients using an exposure treatment regimen* as of Feb. 1, 2021	5,220	2,720	328	8,268
Patients excluded	1,779	384	28	2,191
Reason for exclusion				
Age missing or <18 y	0	226	4	230
Nonreconciled[§]	581	0	0	581
Noninflammatory arthritis diagnosis	827	NA	NA	827
Receipt of concomitant medication listed in exclusion criteria**	370	157	24	551
Influential statistical outliers^*	1	1	0	2
Number of patients included in the study	3,441	2,336	300	6,077

* Exposure regimens included tumor necrosis factor inhibitor monotherapy; tumor necrosis factor inhibitor in combination with methotrexate; tumor necrosis factor inhibitor in combination with azathioprine/mercaptopurine; methotrexate monotherapy; azathioprine/mercaptopurine monotherapy; janus kinase inhibitor monotherapy.

[§] In the GRA registry, reconciled was defined as at least one of the following: deceased; symptoms resolved at the time of data entry; not hospitalized > 30 days after initial diagnosis date; hospitalized and discharged; or not at risk of further interventions/death. In the SECURE-IBD and PsoProtect registries, a case was defined as reconciled after a minimum of 7 days (14 days in PsoProtect) or if sufficient time had passed to observe the disease course through resolution of acute illness or death.

**Exclusion criteria concomitant medications included any medication other than sulfasalazine, mesalamine, hydroxychloroquine or chloroquine, leflunomide, oral budesonide, or glucocorticoids.

^*To improve model fit, we removed influential statistical outliers identified in continuous variables. Two patients were removed who received a daily prednisone dose >70mg.

Table 1.2. Patient and clinical characteristics of the study population and COVID-19 outcomes.

	N (%) unless specified			
	GRA N = 3,441	**SECURE-IBD N = 2,336**	**PsoProtect N = 300**	**Pooled N = 6,077**
Region*				
Africa	16 (0.5)	7 (0.3)	1 (0.3)	24 (0.4)
Eastern Mediterranean	120 (3.5)	68 (2.9)	3 (1.0)	191 (3.1)
Europe	1,800 (52.3)	1,143 (48.9)	272 (90.7)	3,215 (52.9)
North America	1,066 (31.0)	942 (40.3)	7 (2.3)	2,015 (33.2)
South America	375 (10.9)	111 (4.8)	16 (5.3)	502 (8.3)
South East Asia	8 (0.2)	13 (0.6)	1 (0.3)	22 (0.4)
Western Pacific	56 (1.6)	29 (1.2)	0 (0)	85 (1.4)
Unknown	0 (0)	23 (1.0)	0 (0)	23 (0.4)
Demographics				
Age, Mean (SD)	55.0 (14.4)	39.4 (15.4)	49.9 (12.6)	48.8 (16.5)
Sex*				
Male	1,144 (33.2)	1,139 (48.8)	185 (61.7)	2,468 (40.6)
Female	2,295 (66.7)	1,153 (49.4)	115 (38.3)	3,563 (58.6)
Unknown	2 (0.1)	44 (1.9)	0 (0)	46 (0.8)
Diagnoses*				
Rheumatoid arthritis only	2,146 (62.4)	-	-	2,146 (35.3)
Spondyloarthritis only	624 (18.1)	-	-	624 (10.3)
Psoriatic arthritis only	566 (16.4)	-	-	566 (9.3)
Other IA or >1 type of IA	105 (3.1)	-	-	105 (1.7)
Crohn's disease	-	1,537 (65.8)	-	1,537 (25.3)
IBD, unspecified	-	37 (1.6)	-	37 (0.6)
Ulcerative colitis	-	762 (32.6)	-	762 (12.5)
Psoriasis	-	-	300 (100)	300 (4.9)
Disease activity*				
Remission	1,067 (31.0)	1,369 (58.6)	75 (25.0)	2,511 (41.3)
Active disease	1,829 (53.2)	864 (37.0)	225 (75.0)	2,918 (48.0)
Unknown	545 (15.8)	103 (4.4)	0 (0)	648 (10.7)
Exposure regimens*				
TNFi monotherapy	1,183 (34.4)	1,445 (61.9)	216 (72.0)	2,844 (46.8)
TNFi + methotrexate	575 (16.7)	87 (3.7)	7 (2.3)	669 (11.0)
TNFi + Azathioprine/6MP	7 (0.2)	327 (14.0)	0 (0)	334 (5.5)
Methotrexate monotherapy	1,438 (41.8)	31 (1.3)	77 (25.7)	1,546 (25.4)
Azathioprine/6MP monotherapy	19 (0.6)	379 (16.2)	0 (0)	398 (6.5)
JAKi monotherapy	219 (6.4)	67 (2.9)	0 (0)	286 (4.7)
Concomitant medications				
Sulfasalazine	246 (7.1)	48 (2.1)	-	294 (5.1)
Mesalamine	-	384 (16.4)	-	384 (16.4)

	N (%) unless specified			
	GRA N = 3,441	**SECURE-IBD N = 2,336**	**PsoProtect N = 300**	**Pooled N = 6,077**
Oral budesonide	-	39 (1.7)	-	39 (1.7)
Leflunomide	212 (6.2)	-	-	212 (6.2)
Chloroquine or HCQ	316 (9.2)	-	-	316 (9.2)
Daily glucocorticoid use*				
No glucocorticoid use	2,650 (77.0)	2,212 (94.7)	300 (100)	5,162 (84.9)
Glucocorticoid user	683 (19.8)	118 (5.0)	0 (0)	801 (13.2)
Glucocorticoid use unknown	108 (3.1)	6 (0.3)	0 (0)	114 (1.9)
Daily prednisone-equivalent GC dose in mg, Median (25th percentile, 75th percentile)	5 (5, 7.5)	20 (5, 36)	-	5 (5, 10)
Comorbidities				
Smoking status*				
Never or past smoker	2,358 (68.5)	2,236 (95.7)	197 (65.7)	4,791 (79.6)
Current smoker	153 (4.4)	100 (4.3)	42 (14.0)	295 (4.9)
Unknown	930 (27.0)	0 (0)	61 (20.3)	991 (16.3)
BMI*				
<30 kg/m^2	2,768 (80.4)	1,951 (83.5)	158 (52.7)	4,877 (80.3)
≥30 kg/m^2	673 (19.6)	385 (16.5)	92 (30.7)	1,150 (18.9)
Unknown	0 (0)	0 (0)	50 (16.7)	50 (0.8)
Interstitial lung disease	134 (3.9)	26 (1.1)	4 (1.3)	164 (2.7)
Obstructive lung disease	317 (9.2)	99 (4.2)	14 (4.7)	430 (7.1)
Cardiovascular disease	274 (8.0)	90 (3.9)	24 (8.0)	388 (6.4)
Diabetes	401 (11.7)	80 (3.4)	57 (19.0)	541 (8.9)
Hypertension	1,088 (31.6)	193 (8.3)	79 (26.3)	1,360 (22.4)
Kidney disease	93 (2.7)	24 (1)	3 (1.0)	120 (2.0)
Cancer	91 (2.6)	18 (0.8)	8 (2.7)	117 (1.9)
Hospitalization status*				
Not hospitalized	2,396 (69.6)	1,996 (85.4)	257 (85.7)	4,649 (76.5)
Hospitalized	939 (27.3)	316 (13.5)	42 (14.0)	1,297 (21.3)
Unknown	106 (3.1)	24 (1.0)	1 (0.3)	131 (2.2)
Death*				
Alive	3,266 (94.9)	2,282 (97.7)	297 (99.0)	5,845 (96.2)
Died	166 (4.8)	20 (0.9)	3 (1.0)	189 (3.1)
Unknown	9 (0.3)	34 (1.5)	0 (0)	43 (0.7)
COVID-19 diagnosis type				
Presumptive COVID-19 case**	752 (21.9)	0 (0)	112 (37.3)	864 (14.2)

*Categories are mutually exclusive [*categories that were not delineated as "mutually exclusive" were not mutually exclusive*]. **Presumptive diagnosis was based on symptoms alone.
Abbreviations: JAKi: janus kinase inhibitor; 6MP: mercaptopurine; TNFi: tumor necrosis factor inhibitor; HCQ: hydroxychloroquine; GC: glucocorticoid; IA: inflammatory arthritis; IBD: inflammatory bowel disease.

Table 1.3. Adjusted pooled odds of COVID-19 related hospitalization or death using data from the three registries.

	OR (95% CI)	P
Exposure regimens*		
TNFi monotherapy (reference)	1	--
TNFi + Methotrexate	1.18 (0.85-1.63)	0.327
TNFi + Azathioprine/6MP	1.74 (1.17-2.58)	0.006
Methotrexate monotherapy	2.00 (1.57-2.56)	<0.001
Azathioprine/6MP monotherapy	1.84 (1.30-2.61)	0.001
JAKi monotherapy	1.82 (1.21-2.73)	0.004
Concomitant medications		
Sulfasalazine	1.62 (1.13-2.34)	0.009
Leflunomide	1.89 (1.20-2.99)	0.006
Oral budesonide	2.86 (1.20-6.84)	0.018
Daily prednisone-equivalent dose (per 1mg increase)	1.07 (1.05-1.08)	<0.001
Demographics		
Female	0.79 (0.66-0.96)	0.017
Age (per year)	1.04 (1.04-1.05)	<0.001
Obesity (BMI ≥30 kg/m^2)	1.39 (1.10-1.75)	0.005
Current smoker	0.77 (0.51-1.17)	0.214
Disease activity		
Active disease	1.27 (1.04-1.55)	0.019
Comorbidities		
Interstitial lung disease	1.81 (1.12-2.95)	0.016
Obstructive lung disease	2.34 (1.69-3.24)	<0.001
Cardiovascular disease	1.58 (1.13-2.21)	0.007
Diabetes	1.54 (1.16-2.05)	0.003
Hypertension	1.19 (0.95-1.50)	0.121
Kidney disease	3.10 (1.70-5.66)	<0.001
Cancer	1.16 (0.65-2.07)	0.609

***** Categories are mutually exclusive. Odds ratios obtained using hierarchical multivariable mixed-effects logistic regression with registry, calendar-month, and country random effects. Model adjusted for all variables shown. N=6,077. Abbreviations: TNFi: tumor necrosis factor inhibitor; JAKi: janus kinase inhibitors; 6MP: mercaptopurine.

Table 1.4. Adjusted registry-specific odds of COVID-19-related hospitalization or death.

	GRA N = 3,441		SECURE-IBD N = 2,336		PsoProtect N = 300	
	OR (95% CI)	**P**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
Exposure regimens*						
TNFi monotherapy (Ref)	1	--	1	--	1	--
TNFi + Methotrexate	1.20 (0.80-1.79)	0.375	1.59 (0.76-3.34)	0.222		
TNFi + Azathioprine/6MP			1.51 (0.98-2.31)	0.061		
Methotrexate monotherapy	2.21 (1.59-3.08)	<0.001	2.66 (0.95-7.43)	0.062	8.76 (2.94-26.06)	<0.001
Azathioprine/6MP monotherapy	5.28 (1.51-18.43)	0.009	1.50 (1.00-2.24)	0.048		
JAKi monotherapy	2.41 (1.46-3.99)	0.001	0.60 (0.22-1.64)	0.322		
Diagnoses*						
Rheumatoid arthritis (Ref)	1	--				
Spondyloarthritis	1.32 (0.85-2.08)	0.220				
Psoriatic arthritis	0.89 (0.61-1.28)	0.518				
Other IA or >1 type of IA	0.71 (0.35-1.42)	0.326				
Crohn's disease (Ref)			1	--		
Ulcerative colitis			0.86 (0.24-3.01)	0.808		
IBD unspecified			1.17 (0.83-1.65)	0.377		
Disease activity						
Active disease	1.02 (0.78-1.33)	0.886	2.02 (1.45-2.80)	<0.001	0.86 (0.25-2.91)	0.806
Concomitant medications						
Sulfasalazine	1.55 (1.03-2.35)	0.037	1.69 (0.72-4.01)	0.231		
Leflunomide	1.97 (1.22-3.18)	0.005				
Hydroxychloroquine or chloroquine	0.93 (0.64-1.34)	0.684				
Mesalamine			1.24 (0.82-1.89)	0.310		
Oral budesonide			2.71 (1.11-0.60)	0.028		
Daily prednisone-equivalent GC dose (per mg)	1.07 (1.03-1.10)	<0.001	1.06 (1.04-1.08)	<0.001		
Demographics						
Female	0.79 (0.60-1.04)	0.091	0.82 (0.61-1.11)	0.209	0.37 (0.13-1.11)	0.076
Age (per year)	1.05 (1.04-1.06)	<0.001	1.04 (1.03-1.05)	<0.001	1.04 (1.00-1.09)	0.068
Comorbidities						
Current smoker	0.77 (0.44-1.36)	0.364	0.92 (0.45-1.87)	0.815	0.49 (0.10-2.39)	0.375
Obesity (BMI ≥30 kg/m^2)	1.75 (1.28-2.38)	<0.001	1.27 (0.85-1.90)	0.240	0.73 (0.23-2.33)	0.598
Interstitial lung disease	1.61 (0.92-2.81)	0.096	1.58 (0.51-4.89)	0.429		
Obstructive lung disease	2.47 (1.67-3.67)	<0.001	2.34 (1.22-4.48)	0.011	1.05 (0.11-10.02)	0.968
Cardiovascular disease	1.25 (0.81-1.93)	0.316	2.87 (1.55-5.32)	0.001	1.58 (0.34-7.37)	0.564
Diabetes	1.63 (1.17-2.29)	0.004	1.10 (0.55-2.19)	0.785	2.27 (0.75-6.82)	0.146
Hypertension	1.09 (0.83-1.44)	0.516	1.50 (0.92-2.43)	0.102	3.08 (1.01-9.45)	0.049
Kidney disease	3.92 (1.82-8.44)	<0.001	1.76 (0.54-5.76)	0.350		
Cancer	1.32 (0.67-2.59)	0.424	0.77 (0.17-3.41)	0.728	0.85 (0.06-12.85)	0.908

***** Categories are mutually exclusive. Estimates obtained using hierarchical multivariable mixed-effects logistic regression with calendar-month and country random effects. Models adjusted for all variables shown. Abbreviations: GRA: the Global Rheumatology Alliance registry; SECURE-IBD: the Secure Epidemiology of Coronavirus Under Research Exclusion for Inflammatory Bowel Disease registry; PsoProtect: the Psoriasis Patient Registry for Outcomes, Therapy and Epidemiology of COVID-19 Infection; JAKi: janus kinase inhibitors; MTX: methotrexate; AZA: azathioprine; 6MP: 6-mercaptopurine; TNFi: tumor necrosis factor inhibitor; IA: inflammatory arthritis; IBD: inflammatory bowel disease, GC: glucocorticoid.

Table 1.5. Sensitivity analysis to determine the magnitude of potential reporting-bias using data from the GRA registry.

	Original estimates from GRA registry-specific analysis			**Reweighted estimates* from GRA registry-specific analysis**			**Standardized difference****	**Regulatory agreement*^**	**Estimate agreement§**
	OR	**95% CI**	**P**	**OR**	**95% CI**	**P**			
TNFi monotherapy	1	-	-	1	-	-	-	-	-
TNFi + MTX	1.08	0.80-1.47	0.602	1.11	0.64-1.92	0.701	-0.004	yes	yes
MTX monotherapy	1.92	1.50-2.46	<0.001	1.59	1.01-2.50	0.046	0.019	yes	yes
JAKi monotherapy	1.88	1.27-2.79	0.002	1.96	1.06-3.64	0.033	-0.007	yes	yes

All odds ratios were derived using hierarchical multivariable mixed-effects logistic regression with calendar-month random effects and adjusted for the following: age, sex, current tobacco use, immune-mediated disease diagnosis, immune-mediated disease activity (remission vs. active disease), key comorbidities (cardiovascular disease, diabetes, hypertension, obstructive lung disease, interstitial or other chronic lung disease, kidney disease, obesity [BMI ≥30 kg/m2], and cancer), sulfasalazine, hydroxychloroquine or chloroquine, leflunomide and prednisone-equivalent glucocorticoid dose.

* Estimates were obtained after reweighting the covariate distribution of the GRA patients to those of rheumatology clinics from health systems that systematically reported all confirmed and suspected COVID-19 patients, using the inverse-odds sampling-weight technique.

** Standardized difference measures the magnitude of the differences between the original (potentially biased) and the reweighted estimates. Standardized differences were derived from log odds ratios, according to the methods in Franklin *et al*.[14] Values <0.1 are considered acceptable standardized differences.[15]

*^ Regulatory agreement indicates whether original estimates replicate the statistical significance and direction (when estimates are statistically significant) of reweighted estimates.

§ Estimate agreement indicates whether the original estimate lies within the 95% CI of the reweighted estimates.

Abbreviations: TNFi: tumor necrosis factor inhibitor; MTX: methotrexate; JAKi: janus kinase inhibitor.

The number of patients on azathioprine/mecaptopurine (AZA/6MP) or TNFi in combination with AZA/6MP was too small in the aforementioned rheumatology clinics to derive estimates for these exposure regimens.

CHAPTER 2 : Development of a Prediction Model for COVID-19 Acute Respiratory Distress Syndrome in Patients with Rheumatic Diseases: Results from the Global Rheumatology Alliance Registry

Zara Izadi, Milena A Gianfrancesco, Alfredo Aguirre, Anja Strangfeld, Elsa F Mateus, Kimme L Hyrich, Laure Gossec, Loreto Carmona, Saskia Lawson-Tovey, Lianne Kearsley-Fleet, Martin Schaefer, Andrea M Seet, Gabriela Schmajuk, Lindsay Jacobsohn, Patricia Katz, Stephanie Rush, Samar Al-Emadi, Jeffrey A Sparks, Tiffany Y-T Hsu, Naomi J Patel, Leanna Wise, Emily Gilbert, Alí Duarte-García, Maria O Valenzuela-Almada, Manuel F Ugarte-Gil, Sandra Lúcia Euzébio Ribeiro, Adriana de Oliveira Marinho, Lilian David de Azevedo Valadares, Daniela Di Giuseppe, Rebecca Hasseli, Jutta Richter, Alexander Pfeil, Tim Schmeiser, Carolina A Isnardi, Alvaro A Reyes Torres, Gelsomina Alle, Verónica Saurit, Anna Zanetti, Greta Carrara, Julien Labreuche, Thomas Barnetche, Muriel Herasse, Samira Plassart, Maria José Santos, Ana Maria Rodrigues, Philip C Robinson, Pedro M Machado, Emily Sirotich, Jean W Liew, Jonathan S Hausmann, Paul Sufka, Rebecca Grainger, Suleman Bhana, Wendy Costello, Zachary S Wallace, and Jinoos Yazdany, on behalf of the Global Rheumatology Alliance Registry.

ABSTRACT

Background. Some patients with rheumatic diseases might be at higher risk for COVID-19 Acute Respiratory Distress Syndrome (ARDS). We aimed to develop a prediction model for COVID-19 ARDS in this population and to create a simple risk-score calculator for use in clinical settings.

Methods. Data were derived from the COVID-19 Global Rheumatology Alliance (GRA) Registry from March 24, 2020, to May 12, 2021. Seven machine learning classifiers were trained

on ARDS outcomes using 83 variables obtained at COVID-19 diagnosis. Predictive performance was assessed in a U.S. test set and validated in patients from four countries with independent registries using area under curve (AUC), accuracy, sensitivity, and specificity. A simple risk-score calculator was developed using a regression model incorporating the most influential predictors from the best performing classifier.

Results. The study included 8,633 patients reported from 74 countries, of whom 523 (6%) had ARDS. Gradient boosting had the highest mean AUC (0.78, 95% CI: 0.67-0.88) and was considered the top performing classifier. Ten predictors were identified as key risk factors and included in a regression model. The regression model which predicted ARDS with 71% (95% CI: 61-83%) sensitivity in the test set and with sensitivities ranging from 61-80% in countries with independent registries, was used to develop the risk-score calculator.

Conclusions. We were able to predict ARDS with good sensitivity using information readily available at COVID-19 diagnosis. The proposed risk-score calculator has the potential to guide risk-stratification for treatments such as monoclonal antibodies that have potential to reduce COVID-19 disease progression.

SIGNIFICANCE AND INNOVATIONS

- The study describes the development and evaluation of a prediction model for ARDS in individuals with COVID-19 and pre-existing rheumatic diseases using information readily available at the time of COVID-19 exposure or onset.
- The prediction model had good internal and external validity and was used to develop a simple and interpretable risk-score calculator for use in clinical settings.

- Age, daily glucocorticoid dose, pulmonary hypertension, interstitial lung disease, chronic kidney disease, anti-CD20 monoclonal antibody use, diabetes, hypertension, active rheumatic disease, and morbid obesity were the most influential factors in predicting the onset of ARDS.
- The proposed risk-score calculator has the potential to guide risk-stratification and the treatment of COVID-19 in high-risk patients with rheumatic diseases.

INTRODUCTION

Acute Respiratory Distress Syndrome (ARDS), affecting about 5% of patients with Coronavirus disease 2019 (COVID-19)[43] and one-third of hospitalized patients,[44] is a life-threatening complication of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. ARDS in the setting of COVID-19 has a mortality rate of 26% to 62% in people admitted to a critical care setting and from 66% to 94% in patients who received mechanical ventilation.[45] ARDS frequently causes long-lasting effects beyond hospitalization, from cognitive impairment to physical weakness.[46] Given the high mortality and long-term consequences of ARDS, and the direct burden on the healthcare system, identification of patients at risk for this complication and use of potentially mitigating treatment strategies are important.

There is controversy regarding the existence of an increased risk of severe COVID-19 outcomes in people with rheumatic diseases.[47-50] For example, reports from a Swedish nationwide study showed that the risks of COVID-19-related hospitalization and death (but not intensive care unit (ICU) admission) were increased in rheumatoid arthritis (RA), while for other inflammatory joint diseases only the risk of COVID-19-related hospitalization was increased, compared to population referents. However, these risks were comparable to the increased risk of all-cause hospitalization in patients with rheumatic diseases, and the increased all-cause mortality risk in

RA patients, and the increased mortality risk in 2020 in RA patients was not different from that in 2015 to 2019.[47] In the U.S., a multi-institutional electronic health record (EHR) study found higher risks of hospitalization, ICU admission, acute renal failure, and venous thromboembolism (but not death) in patients with rheumatic diseases compared to matched controls.[51] Another study conducted at a multi-institutional health system among patients admitted to hospital with COVID-19 showed higher odds of admission to intensive care and mechanical ventilation in patients with rheumatic diseases compared to matched controls.[52]

The risk factors most strongly associated with ARDS, the key life-threatening organ involvement in COVID-19, are not yet identified. Predicting ARDS using information available at the time of COVID-19 diagnosis has the potential to guide clinical risk stratification and the management of COVID-19 in this population. Because ARDS is a relatively rare event in people who develop COVID-19, there are special considerations in developing statistical models predicting this outcome. Prediction using traditional regression methods can lead to overfitting, limiting the number of predictors that can reliably be used in the prediction model.[53] Common variable reduction strategies such as stepwise regression become less effective as the number of potential predictors grow, which is an important consideration when modeling numerous rheumatic diseases, medications, and comorbidities.[54] In addition, regression models typically limit the link between outcome and predictor variables to be linear and additive; as a result, regression models may fail to adequately represent complex interactions and high-dimensional relationships that may be present in patients with rheumatic diseases.[55] Machine learning algorithms provide an alternative approach with the potential to improve predictive performance, in particular sensitivity, in the setting of relatively rare events such as ARDS.

This study aimed to develop a prediction model for ARDS in individuals with COVID-19 and pre-existing rheumatic diseases using information obtained at the time of COVID-19 diagnosis and a series of machine learning algorithms for predictor selection. An additional aim was to develop a simple and interpretable risk-score calculator for potential use in clinical settings.

METHODS

Study design. This study used data from the COVID-19 Global Rheumatology Alliance (GRA) Registry,[21] from March 24, 2020, to May 12, 2021. Briefly, data from adults with rheumatic diseases diagnosed with COVID-19 are entered by rheumatology clinicians via one of two parallel international data entry portals: one[56] limited to European countries and a second[57] for the rest of the world. Five countries in Europe (France,[50,58,59] Germany,[60-62] Italy,[63] Portugal,[64,65] and Sweden[66]) and two countries in South America (Brazil[67,68] and Argentina[69]) host national registries supported by their respective national societies. National data from these countries is regularly transferred and merged into the global GRA registry. While GRA data largely depend on convenience sampling, rheumatology practices from two large health-systems within the U.S. (Mass General Brigham in Massachusetts and Mayo Clinic in Minnesota and Florida) have processes in place to systematically report all symptomatic and asymptomatic COVID-19 diagnoses, irrespective of COVID-19 severity.

Patient demographics, rheumatic disease characteristics, comorbidities, COVID-19 outcomes, and complications are entered by reporting clinicians. Methods of COVID-19 diagnoses are indicated including one or more of polymerase chain reaction, antigen testing, antibody, metagenomic testing, CT scan, laboratory assay, or a presumptive diagnosis based on symptoms or close contact alone. Quality is assessed by data validation teams who remove all known or potential duplicates and address erroneous or ineligible reports. The GRA registry was

determined ‘not human subjects research’ by the UK Health Research Authority and The University of Manchester, as well as under United States Federal Guidelines assessed by the University of California, San Francisco and patient consent was not required. We followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement for prediction model development and validation.[70]

Inclusion and exclusion criteria. We included patients with a reconciled status only, defined as the highest COVID-19 illness severity level being confirmed, and including one of the following outcomes: death; symptoms resolved at the time of data entry; not hospitalized > 30 days after initial diagnosis date; hospitalized and discharged; or not at risk of further interventions/death. Patients with a COVID-19 diagnosis date that preceded January 1st, 2020, were excluded (N=7). Additionally, we excluded patients with missing data on ARDS or any of the predictor variables (Table 2.1). Patients reported from France, Portugal, and Germany were excluded due to unavailability of data on ARDS or smoking status. No formal sample size analysis was done as all eligible patients in GRA were included in the study.

Outcome. ARDS was the outcome and the event being predicted in this study. A diagnosis of COVID-19 related ARDS was indicated by the reporting clinician at the point of data entry and in almost all cases reflected a diagnosis given to the patient by the inpatient team (e.g., pulmonologists, critical care specialists or internists directly caring for the patient).

Predictors. ARDS was predicted using 83 predictor variables related to patient demographics, rheumatic disease diagnoses and activity, immunomodulatory medications used for the treatment of rheumatic disease, and comorbidities (Table 2.2). All variables reflect data at the time of COVID-19 diagnosis.

Training, test, and validation sets. Construction of the training, test, and validation sets are depicted in Figure 2.1. Patients reported from the U.S. (except those reported from Mass General Brigham in Massachusetts and Mayo Clinic in Minnesota and Florida) and all other countries that directly reported to the GRA registry were included in the training set (N=5,673). The test set comprised all patients reported from Mass General Brigham in Massachusetts and Mayo Clinic in Minnesota and Florida (N=891). We used this approach to address any potential for provider reporting bias and to improve the generalizability of our findings by testing on a subset of data that most closely represents the underlying spectrum of COVID-19 severity among patients with pre-exiting rheumatic diseases. Additionally, patients reported from these health-systems had low rates of missing data (<7% in any variable and <10% of patients excluded due to incomplete data) permitting complete-case analyses. Patients reported from countries with independent registries were used as validation sets (N=2,069). We utilized four validation sets in total, corresponding to patients reported from Italy (N=1,060), Sweden (N=225), Brazil (N=201), and Argentina (N=583). The amount of missing data varied considerably between the validation sets (Table 2.1). Italy had the lowest rates of missing data (<7% in any variable and <10% of patients excluded due to incomplete data) and was therefore considered the primary validation set.

Prediction algorithms. Since ARDS is relatively rare and many predictors are potentially relevant to predicting this severe outcome, we used a machine learning approach for predictor selection which is suited to data with high dimensionality. In order to identify the most important predictors of ARDS, we compared predictive performance of seven supervised machine learning classifiers commonly applied in the setting of rare clinical outcomes.[71] The classifiers were trained on ARDS outcomes using three repeats of 10-fold cross-validation. Accuracy was used as

the validation metric for hyperparameter tuning. Prediction algorithms utilized instance-based learning (k-nearest neighbors [kNN], and support vector machines [SVM]), regularization (the lasso and elastic-net regularized generalized linear models [GLMNET]), Bayesian regression (Bayesian generalized linear models [BAYESGLM]), additive models (generalized additive models [GAM]), ensemble learning (gradient boosting machines [GBM]), and deep learning (neural networks [NNET]). All analyses were performed in R version 3.6.1, using the Classification And Regression Training[72] (CARET) package.

Model performance. Model performance was assessed using accuracy, and measures of discrimination including sensitivity, specificity, and area under curve (AUC). The prediction algorithm with the highest AUC in the test set was selected as the best performing classifier. AUC is an aggregate measure of the receiver operator curve and, unlike accuracy, does not depend on classification threshold value. To derive sensitivity, specificity, and accuracy, predicted ARDS probabilities were dichotomized to ascertain the event (yes, no) in the test and validation sets. For each prediction algorithm and dataset, separately, classification threshold values were selected to reduce the absolute difference between sensitivity and specificity.[73] This approach was taken to maximize both metrics (Figure 2.2), to account for differences in the distribution of predicted ARDS probabilities, and to account for potential country-level differences in health-system structure, healthcare access, and utilization. Mean classification thresholds, mean performance metrics, and corresponding 95% confidence intervals were derived from 1000 samples of 500 randomly selected patients from the test set and each validation set, using bootstrapping and sampling with replacement.

Risk-score calculator development. The risk-score calculator was derived from a multivariable logistic regression incorporating the most influential predictors from the best performing

classifier.[74] We used logistic regression to develop a risk-score calculator that was interpretable, user-friendly, and readily accessible for potential use in clinical settings across health-systems. To determine the optimum number of items in the risk-score calculator, a series of regressions with varying numbers of predictors (ranging from top 5 predictors to top N predictors, where the importance score associated with Nth predictor was >0) were trained on ARDS outcomes using 10 repeats of 10-fold cross-validation with AUC as the validation metric. To balance the calculator's ease of use in clinical settings[75,76] with predictive performance, our final regression model incorporated the lowest number of predictors associated with the highest mean AUC. To improve regression fit, we assessed linearity in the relationship between continuous predictors and the outcome and accounted for non-linear relationships using interaction terms. Direction, magnitude, and statistical significance of key risk factors associated with ARDS were reported from the final regression model using adjusted odds ratios (ORs). The predictive performance of the regression model was evaluated in the test set and validation sets using the aforementioned performance metrics and methods. Additionally, we assessed calibration of the regression model by comparing the mean predicted ARDS probabilities with the mean observed probabilities within every decile of predicted risk in the test and validation sets and reported corresponding integrated calibration indices (ICIs).[77]

To aid the interpretation of predicted probabilities, risk of ARDS development was defined as 'low' for predicted probabilities lower than the lowest country-specific mean classification threshold, 'moderate to high' for the predicted probability region between the highest and the lowest country-specific mean classification thresholds, and 'high' for predicted probabilities equal to or higher than the highest country-specific mean classification threshold.

A point-based scoring system was developed in which points were assigned to each item by multiplying each β coefficient (log OR) from the regression model by a constant arbitrary number and rounding (to the nearest integer for points 1-5 and to the nearest 5th integer for points >5) to facilitate total risk-score calculation. A total risk-score was assigned to each patient by summing the points for each item in the risk-score calculator. Mean predicted ARDS probabilities and 95% confidence intervals corresponding to each total risk-score within the 'moderate to high' category of risk are reported.

RESULTS

Characteristics of the study population. A total of 8,633 patients reported from 74 countries were included in the study. Of these, 5,673 were partitioned into the training set and 891 and 2,069 into the test set and validation sets, respectively, as described in methods. Among patients comprising the training set, mean (SD) age was 53.2 (15.2) years, 4,088 (72.1%) were female, and 4,212 (74.2%) were non-smokers. Rheumatoid arthritis (RA), reported among 2,472 (43.6%) individuals, was the most common diagnosis, followed by systemic lupus erythematosus (SLE, 12.1%), and psoriatic arthritis (10.0%). Treatment with conventional synthetic disease modifying antirheumatic drugs (csDMARDs) alone was the most common treatment modality (44.1%). A majority of individuals were in remission or low disease activity (80.3%, Table 2.3).

Baseline characteristics of patients included in the test set were generally comparable to those of patients included in the primary validation set (patients reported from Italy), however some notable differences were observed: the prevalence of smoking, obesity, diabetes, hypertension, and chronic kidney disease was higher in the test set than the primary validation set (39.1 vs. 29.1%, 30.6 vs. 12.4%, 18.7 vs. 9.6%, 46.2 vs. 34.4%, and 12.8 vs. 6.2%, respectively); the prevalence of psoriatic arthritis and psoriasis was higher in the primary validation set than the

test set (20.8 vs. 9.1%, 17.4 vs. 6.1%, respectively); both DMARD therapy and glucocorticoid use were more prevalent in the primary validation set (83.5% and 37.8%, respectively) than the test set (70.3% and 28.7%, respectively).

Similarly, there were differences in baseline characteristics between the test set and other validation sets: The prevalence of smoking varied significantly and was lowest in Brazil (6.5%) and highest in Sweden (43.6%); SLE was more prevalent in Argentina (18.9%) and less prevalent in Sweden (2.2%) than the test set (11.1%); use of biologic or targeted synthetic DMARDs, alone or in combination with conventional synthetic DMARDs, was more prevalent in Sweden (60.0%) and Brazil (47.7%) than the test set (32.4%); and the prevalence of glucocorticoid use ranged from 10.4% in Brazil to 42.5% in Argentina. ARDS was reported among 355 (6.3%) patients in the training set, 35 (3.9%) patients in the test set, and 57 (5.4%) patients in the primary validation set (Italy). In the other validation sets, the prevalence of ARDS ranged from 3.3% (Sweden) to 8.5% (Brazil).

Predictive performance of machine learning algorithms. Among the seven machine learning classifiers, GBM had the highest AUC in the test set (mean: 0.78, 95% CI: 0.67-0.88) and was considered the top performing model (Table 2.4). In the test set, at the optimum classification threshold, GBM had a mean accuracy, sensitivity, and specificity of 0.70. In the primary validation set, GBM had a mean AUC of 0.79 (95% CI: 0.70-0.87) and a mean accuracy, sensitivity, and specificity of 0.73 at the optimum classification threshold (Table 2.5). In other validation sets, GBM's mean AUC ranged from 0.74 to 0.85, with mean sensitivity and mean specificity ranging from 0.65 to 0.78 and 0.66 to 0.78, respectively. In order of predictor importance, age, average daily prednisone-equivalent glucocorticoid dose, and pulmonary hypertension were the most influential predictors identified by GBM (Figure 2.3).

Important risk factors and risk-score calculator. As described in methods, to determine the optimum number of predictors in the risk calculator we compared the predictive accuracy of 10 regression models each including a different number of predictors, from a minimum of 5 predictors to a maximum of 14 predictors. Thus, the first regression model included the top 5 predictors identified by the GBM classifier. The second regression model included the top 6 predictors and so on, such that the tenth regression model included the top 14 predictors identified by the GBM classifier. We stopped at 14 predictors because the 15th predictor identified by GBM had an importance score of 0. Each of the ten regression models was trained on ARDS outcomes using 10 repeats of 10-fold cross-validation. AUC was used as the validation metric to identify the best performing regression model which included the optimum number of predictors. The model including the top 5 predictors had a mean AUC of 0.74. Mean AUC increased with increasing number of predictors and reached a maximum value of 0.77 in the model that included the top 10 predictors identified by GBM. Mean AUC stayed at a constant value of 0.77 in the remaining four models that included the top 11, 12, 13 and 14 predictors. We therefore chose 10 as the optimum number of items in the risk calculator to facilitate clinical utility (by including fewer items) without compromising predictive accuracy. The risk-score calculator was therefore derived from a multivariable logistic regression model incorporating the ten most influential predictors from GBM. Average daily prednisone-equivalent glucocorticoid doses >60mg were considered clinically high doses. We fitted an interaction term to account for the potential effect-modification in dose-response in patients receiving glucocorticoid doses >60mg. The resulting regression was equivalent to a simpler regression that Winsorized glucocorticoid doses >60mg to 60mg (calibration slope: 0.99 [1.00 indicating perfect calibration]; calibration intercept: 0.00; correlation coefficient: 0.99, $p<0.0001$). We therefore

opted for the simpler regression model in creating the risk-score calculator. All ten predictors were independently and statistically significantly associated with the development of ARDS (Figure 2.4): older age (OR 1.45, 95% CI 1.33-1.57, per decade increase in age), higher average daily prednisone-equivalent glucocorticoid doses (1.17, 1.11-1.23, per 5mg increase in dose), pulmonary hypertension (3.97, 2.13-7.42), interstitial lung disease (2.49, 1.74-3.57), chronic renal insufficiency or end stage renal disease (2.05, 1.43-2.93), anti-CD20 monoclonal antibody use (3.00, 1.95-4.63), diabetes (1.42, 1.08-1.87), hypertension (1.40, 1.10-1.80), moderate or high rheumatic disease activity (1.57, 1.21-2.03), and morbid obesity (1.92, 1.26-2.92).

Predictive performance of the final regression model was assessed in the test set and each validation set from countries with independent registries. In the test set, the model had a mean AUC of 0.79 (95% CI: 0.68-0.88) and a mean accuracy, sensitivity, and specificity of 0.71 at the optimum classification threshold (Table 2.6). In the primary validation set, the model had a mean AUC of 0.77 (95% CI: 0.68-0.86) and a mean accuracy, sensitivity, and specificity of 0.73 at the optimum classification threshold. In other validation sets, mean AUCs ranged from 0.71 to 0.85, with both mean sensitivity and mean specificity ranging from 0.61 to 0.80. The calibration plot showed relatively poor agreement between the observed and predicted ARDS risk in the test set (calibration slope: 0.43, intercept: 0.00, ICI: 0.056), and good agreement in the primary validation set (calibration slope: 0.80, intercept: 0.00, ICI: 0.024). The model had relatively poor to moderate calibration in other validation sets with calibration slopes, intercepts and ICIs ranging from 1.38 to 1.91, -0.03 to 0.01, and 0.029 to 0.049, respectively (Figure 2.5).

Figures 2.6 and 2.7 provide details of the ARDS risk-score calculator developed from the multivariable regression model. Predicted ARDS probabilities <4% (corresponding to total scores ≤60) were defined as ‘low’ risk, predicted ARDS probabilities between 4-9%

(corresponding to total scores 61-89) were defined as 'moderate to high' risk, and predicted ARDS probabilities >9% were defined as 'high' risk. As described in methods, these thresholds were not quantitatively derived but instead reflect probabilities that were felt to be clinically meaningful.

DISCUSSION

In this global sample of patients with rheumatic diseases, we developed a simple ARDS risk-score calculator which has the potential for risk-stratification and to guide management of COVID-19 among individuals with rheumatic diseases in routine clinical settings. A machine learning classifier, GBM, predicted the onset of ARDS with 70% sensitivity in the test set, and with 73% sensitivity in the primary validation set, using information obtained at COVID-19 diagnosis. A multivariable regression model using the ten most influential predictors from GBM predicted ARDS with 71% sensitivity in the test set and with 73% sensitivity in the primary validation set. Rheumatic disease characteristics and medications identified as important risk factors in predicting COVID-19 ARDS align with previously reported factors associated with COVID-19 hospitalization or death in patients with immune-mediated inflammatory diseases.[12,17,19,78-80] Other risk factors including older age, obesity, chronic lung disease and chronic kidney disease were also consistent with risk factors identified by a recently published prognostic model for adverse COVID-19 outcomes using information obtained at diagnosis in a general population-based cohort from Iceland.[81]

Our study findings help identify patients with underlying rheumatic diseases who may be at a higher risk for ARDS from COVID-19. Use of baseline information at COVID-19 symptom onset or at COVID-19 diagnosis in asymptomatic patients facilitates early triage of high-risk patients for monitoring, prophylaxis, or treatment interventions. For example, with the recent

Food and Drug Administration Emergency Use Authorizations[82,83] for the use of monoclonal antibodies for the treatment of ambulatory patients with COVID-19, or as post-exposure prophylaxis for high-risk individuals exposed to the virus, a risk calculator coupled with clinical judgment can prioritize which patients are most likely to derive benefit from this therapy. Our findings also identify potentially modifiable risk factors that rheumatologists can consider when making patient care decisions to minimize the risk of adverse COVID-19 outcomes, namely, glucocorticoid dose, rituximab use,[84] and rheumatic disease activity.

With only 10 items, the proposed calculator is simple to use and can be easily implemented in clinical settings. Additionally, information required for the scoring system is available in both outpatient and inpatient settings, or even remotely without the need for close contact, which is not the case with existing ARDS prediction models that require physical examination, laboratory measurements and imaging data.[85-87] In classification, there is typically an inverse relationship between sensitivity and specificity. In this study, we selected classification thresholds that maximized both sensitivity and specificity by minimizing the absolute difference between them. This choice is somewhat arbitrary; in practice the trade-off between specificity and sensitivity must account for the underlying population risk for ARDS, health gains from available treatment or monitoring interventions, and the regional health-system structure that governs the availability and access to health resources.

With the exception of Brazil, both GBM and the GBM-based regression model performed slightly better in validation sets than the test set. This may be explained by the fact the training set was more similar in nature to the validation sets than the test set, such that the provider reporting bias affecting the training set was of a similar magnitude of the bias affecting the validation sets. It is plausible that rheumatology practices that systematically reported all

COVID-19 diagnoses and comprised the test set also captured information on important risk factors, such as comorbidities, more comprehensively than practices that comprised the training set and validation sets. Calibration plots support this hypothesis: predicted probabilities of ARDS were higher than the observed risk in the test set whilst they were largely comparable in Italy, Sweden, and Argentina and lower than the observed risk in Brazil. Without processes in place to systematically report all COVID-19 diagnoses and capture complete information on baseline characteristics, it is possible that provider reporting patterns were influenced by COVID-19 severity, provider perceptions of factors related to COVID-19 severity, and availability of information through direct interactions between the patient and their rheumatologist during the pandemic. Furthermore, patients may underreport important social and behavioral factors such as smoking. This social desirability bias can vary across countries and cultures[88] and may additionally explain the discordances observed in predictive performance.

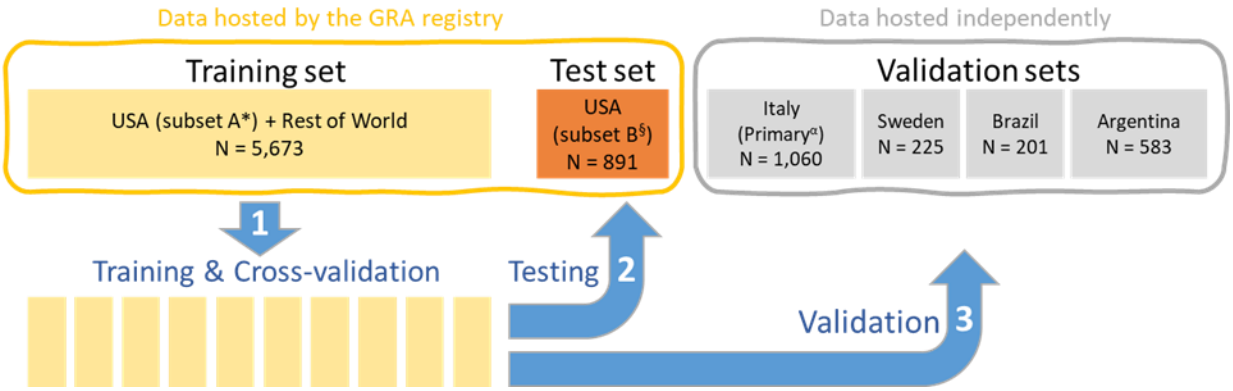
This study has important strengths. First, to our knowledge this is the first study predicting COVID-19 ARDS among individuals with rheumatic diseases. Second, the prediction models were trained on a global sample of individuals with rheumatic disease, thus increasing the heterogeneity and likely generalizability of patient characteristics. This has the potential to improve prediction accuracy by increasing the number of potential predictors and accounting for complex high-dimensional relationships between them. Importantly, active rheumatic disease status was captured as a predictor. The registry is unique among other data sources in rheumatic diseases in being able to capture data on disease activity that is not typically available in administrative data or EHRs. Furthermore, reporting occurred directly by rheumatology clinicians, which likely increased the accuracy of the information. Third, we tested the performance of prediction models in a subset of practices that had processes in place to minimize

potential provider reporting bias. Maximizing the heterogeneity of COVID-19 outcomes in the test set improves the generalizability of our findings to the target population of individuals with pre-existing rheumatic diseases with COVID-19. Fourth, the external validity of our prediction models was assessed using external datasets from Europe and Latin America.

Limitations of this work include potential provider reporting bias and missing data in the training set and external validation sets; the tool should therefore be used with caution outside of the United States. Assessments of calibration showed relatively poor agreement between observed and predicted probabilities of ARDS in the test set and in external validation sets; we therefore recommend that the tool is used as a guide for COVID-19 prognosis and in conjunction with clinical judgement. While we attempted to account for country-level differences in health-system structure, healthcare access and utilization through optimizing ARDS classification thresholds at the country level, a residual impact by these factors may remain. Additionally, we were unable to account for other important sociodemographic or environmental factors such as race/ethnicity, alcohol consumption, occupation, poverty, housing conditions, or air pollution, all of which may influence the outcomes of COVID-19, including the development of ARDS. Much of the data were obtained prior to the wide availability of COVID-19 vaccines, which may lower risk of developing severe COVID-19 outcomes such as ARDS. However, globally only a minority of people are vaccinated. Conversely, the more contagious and virulent SARS-CoV-2 delta variant means many people may be at risk of ARDS. Vaccinated COVID-19 patients with rheumatic diseases have been reported to experience breakthrough infection possibly due to inadequate humoral vaccine immune response associated with some immunosuppressors.[89]

In summary, a GBM-based regression model predicted COVID-19 ARDS with good sensitivity in patients with pre-existing rheumatic diseases using demographics and basic clinical

characteristics that can be easily obtained at COVID-19 diagnosis. Prediction accuracies were largely comparable or better in external datasets from four countries that hosted independent COVID-19 registries. Age, daily glucocorticoid dose, pulmonary hypertension, interstitial lung disease, chronic kidney disease, anti-CD20 monoclonal antibody use, diabetes, hypertension, active rheumatic disease, and morbid obesity were the most influential factors in predicting the onset of ARDS. Further studies are needed to prospectively evaluate the clinical utility of the proposed risk-score calculator for its potential to guide risk-stratification, prophylaxis with monoclonal antibodies and treatment of COVID-19 in high-risk patients with rheumatic diseases.



*Subset A included all patients reported from the U.S. except patients reported from Partners in Massachusetts and Mayo Clinics in Minnesota and Florida. §Subset B included all patients reported from Partners in Massachusetts and Mayo Clinics in Minnesota and Florida. These health systems systematically reported all COVID-19 diagnoses, irrespective of severity. α Italy had the lowest rates of unknown data (<7% in any variable and <10% of patients excluded due to incomplete data) among all validation sets and was therefore considered the primary validation set. (1) Seven supervised machine learning algorithms were trained on ARDS outcomes using three repeats of 10-fold cross-validation. (2) Predictive performance was assessed in the test set. (3) Predictive performance was further assessed in the validation sets.

Figure 2.1. Dataset partitioning into training, test, and validation sets.

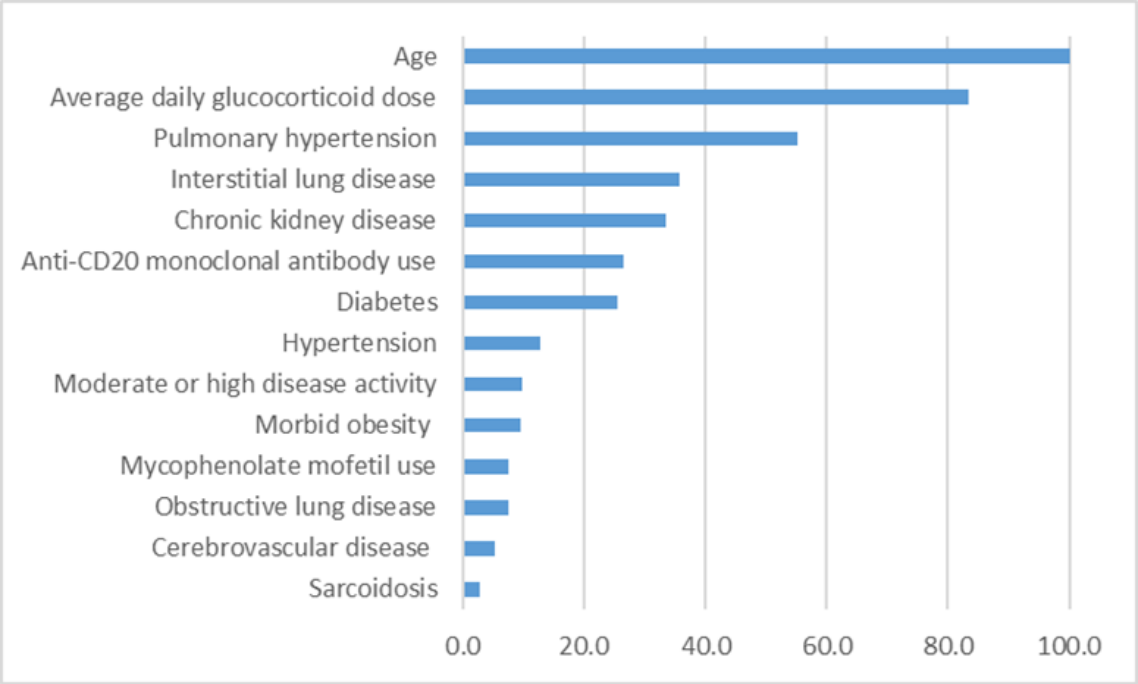
$$\boldsymbol{IU(c) = (|Se(c) - AUC| + |Sp(c) - AUC|)}$$

The cutpoint $\hat{c}_{IU}$, which minimizes the $IU(c)$ function and the $|Se(c) - Sp(c)|$ difference, will be the optimal cutpoint value.

The cutpoint $\hat{c}_{IU}$ defined by the IU method should therefore satisfy two conditions: (1) sensitivity and specificity obtained at this cutpoint should be simultaneously close to the AUC value; (2) the difference between sensitivity and specificity obtained at this cutpoint should be minimum.

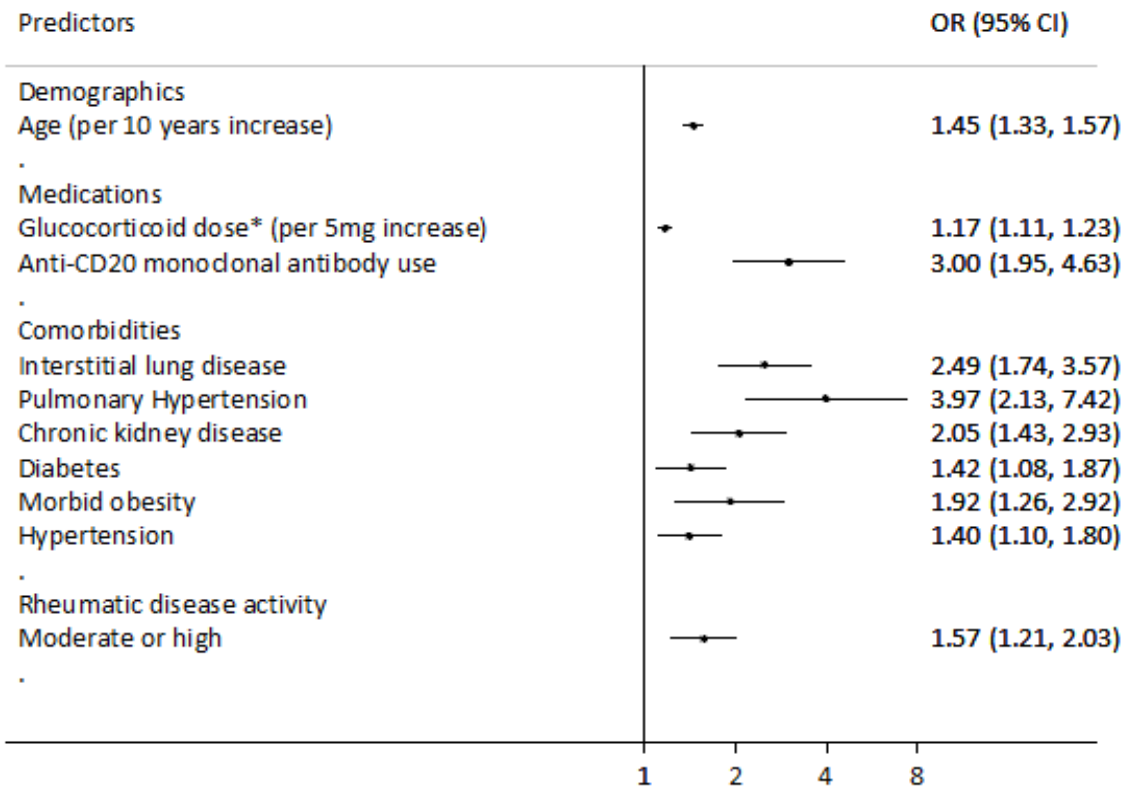
IU: index of union; c: cutpoint; Se: sensitivity; Sp: specificity; AUC: area under curve.

Figure 2.2. The Index of Union method for classification threshold optimization.



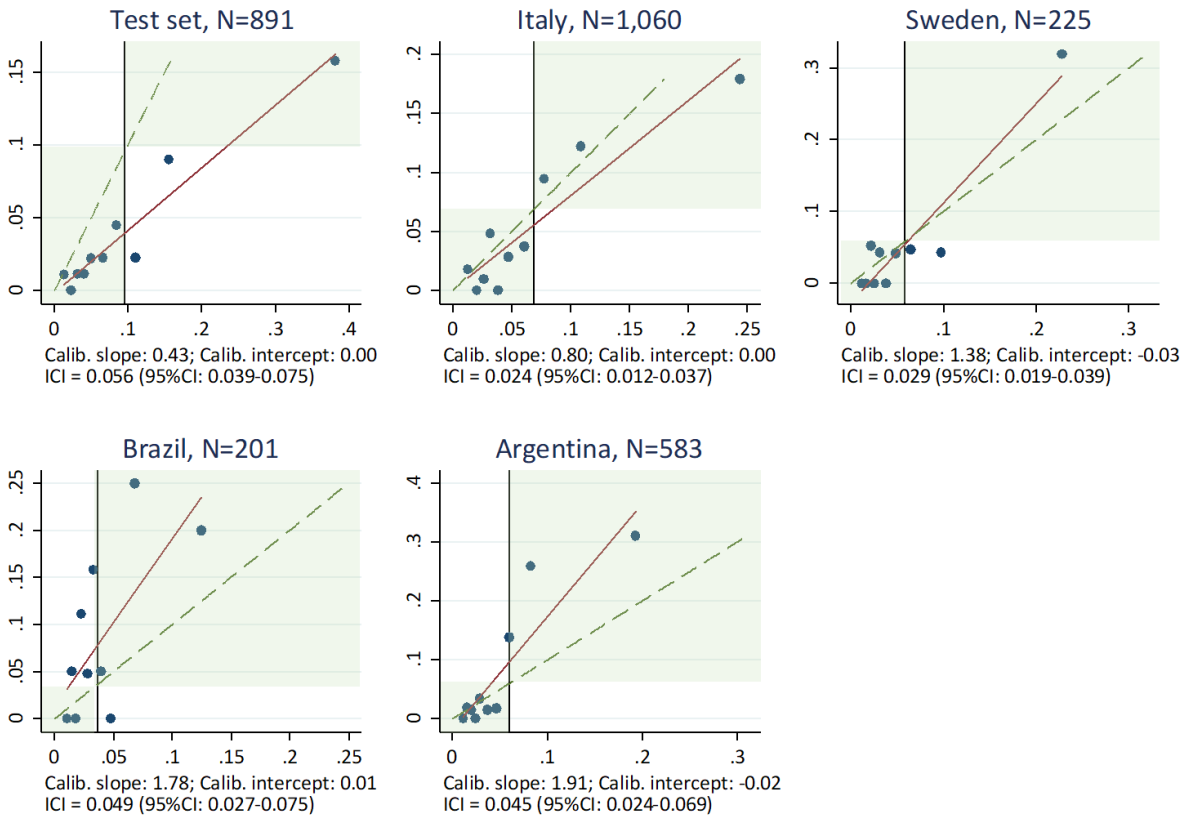
Importance scores are derived by permuting each predictor variable at a time and computing the associated reduction in predictive performance. To normalize and allow comparison across algorithms, importance scores are scaled between 0 and 100. Includes predictors with importance score >0.

Figure 2.3. Gradient boosting machine predictor importance plot.



Top ten most influential predictors identified by the gradient boosting machine.
*Average daily prednisone-equivalent dose.

Figure 2.4. Adjusted odds ratios obtained from the multivariable logistic regression model.



Plots show mean predicted probabilities (within every decile of predicted risk) on the x-axis and mean observed probabilities on the y-axis. Dashed green line represents perfect calibration with a slope of 1. Red line represents the fitted slope. Vertical black line represents the optimum classification threshold. ICI: Integrated Calibration Index. ICI is a weighted average of the absolute difference between the calibration curve and the diagonal line of perfect calibration, where the absolute differences are weighted by the density function of the weights. This is equivalent to integrating f (x) over the distribution of the predicted probabilities. Mean ICIs and corresponding 95% confidence intervals were derived from 1000 random samples of 500 patients from each dataset using bootstrapping and sampling with replacement.

Figure 2.5. Calibration plot for the multivariable regression model showing level of agreement between observed and predicted risk in the test and validation sets.

COVID-19 Acute Respiratory Distress Syndrome (ARDS) Risk Calculator

For use in adult patients with rheumatic disease and a suspected or confirmed diagnosis of COVID-19.

COVID-19 Global Rheum Alliance

Add up points to calculate total score

Age in years
+ Average daily prednisone-equivalent glucocorticoid dose in mg*
+ 35 if patient has pulmonary hypertension
+ 30 if patient is on an anti-CD20 monoclonal antibody**
+ 25 if patient has interstitial lung disease
+ 20 if patient has chronic kidney insufficiency or end stage kidney disease
+ 15 if patient is morbidly obese (BMI ≥40)
+ 10 if patient has diabetes
+ 10 if patient has hypertension
+ 10 if patient has moderate or high rheumatic disease activity

* Up to a maximum of 60mg; ** Including use within the past 12 months. All information to be obtained at COVID-19 symptom onset or diagnosis. Much of the data used in the development of this tool were obtained prior to the wide availability of COVID-19 vaccines. The tool should therefore be used with caution in people who have been vaccinated.

This tool was created with the support of the American College of Rheumatology (ACR) and European Alliance of Associations for Rheumatology (EULAR). However, its content is strictly the work of its authors and has no affiliation with any organization or institution. A printable version is available at: https://rheum-covid.org/

	Total score	Mean Probability of ARDS
Low risk	≤60	<4%
Moderate to High risk	61-76	4-6%
	77-84	6-8%
	85-89	8-9%
High risk	≥90	>9%

Turn over for more detailed information.

Figure 2.6. The risk-score calculator pocket care side 1.

COVID-19 Acute Respiratory Distress Syndrome (ARDS) Risk Calculator

For use in adult patients with rheumatic disease and a suspected or confirmed diagnosis of COVID-19.

Total score	Probability (%) of ARDS, Mean (95% CI)	Total score	Probability (%) of ARDS, Mean (95% CI)
60	3.4 (3.4-3.4)	76	5.9 (5.9-6.0)
61	3.5 (3.5-3.6)	77	6.1 (6.1-6.2)
62	3.7 (3.6-3.7)	78	6.4 (6.3-6.4)
63	3.8 (3.8-3.8)	79	6.6 (6.5-6.6)
64	3.9 (3.9-4.0)	80	6.9 (6.8-6.9)
65	4.1 (4.1-4.1)	81	7.0 (7.0-7.1)
66	4.3 (4.2-4.3)	82	7.3 (7.2-7.4)
67	4.3 (4.3-4.4)	83	7.5 (7.5-7.6)
68	4.5 (4.5-4.6)	84	7.8 (7.7-7.9)
69	4.7 (4.6-4.7)	85	8.0 (8.0-8.1)
70	4.8 (4.8-4.9)	86	8.4 (8.3-8.5)
71	5.1 (5.0-5.1)	87	8.6 (8.5-8.7)
72	5.2 (5.1-5.2)	88	8.9 (8.8-9.0)
73	5.3 (5.3-5.4)	89	9.3 (9.2-9.4)
74	5.6 (5.5-5.6)	90	9.5 (9.3-9.6)
75	5.7 (5.7-5.8)	91	10.0 (9.8-10.1)

This calculator was developed in 5,673 individuals with rheumatic diseases and COVID-19 from 72 countries across 4 continents (mean age 53, 72% female, 44% with a diagnosis of rheumatoid arthritis, 80% in remission or low disease activity, and an ARDS prevalence of 6%).

This risk calculator sorted patients who developed ARDS from patients who did not develop ARDS correctly on average 79% of the time in a sample of patients from the U.S., 77% of the time in a sample of patients from Italy, 82% of the time in a sample of patients from Sweden, 71% of the time in a sample of patients from Brazil, and 85% of the time in a sample of patients from Argentina.

Figure 2.7. The risk-score calculator pocket care side 2.

Table 2.1. Distribution of unknown or missing data across variables and datasets.

	Training set	**Test set**	**Validation sets**			
			Primary	**Other**		
			Italy	**Sweden**	**Brazil**	**Argentina**
N to start with	9346	968	1176	370	700	1025
N (%) unknown or missing						
Outcome (ARDS)	1,716 (18.4)	0 (0)	76 (6.5)	9 (2.4)	0 (0)	0 (0)
Age	0 (0)					
Sex	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.1)	0 (0)
Smoking status	2,656 (28.4)	3 (0.3)	35 (3.0)	0 (0)	6 (0.9)	127 (12.4)
Rheumatic disease diagnoses	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Rheumatic disease activity	1,901 (20.3)	63 (6.5)	9 (0.8)	16 (4.3)	379 (54.1)	59 (5.8)
Comorbidities*	551 (5.9)	3 (0.3)	0 (0)	112 (30.3)	113 (16.1)	138 (13.5)
Immunomodulatory medications^	284 (3.0)	0 (0)	0 (0)	21 (5.7)	46 (6.6)	241 (23.5)
Daily Glucocorticoid dose	1,379 (14.8)	11 (1.1)	8 (0.7)	17 (4.6)	233 (33.3)	7 (0.7)
N (%) excluded due to incomplete data	3,673 (39.3)	77 (8.0)	116 (9.9)	145 (39.2)	499 (71.3)	442 (43.1)
N included in analyses	5,673	891	1,060	225	201	583

* Comorbidities included: Interstitial lung disease; Obstructive lung disease; Diabetes; Morbid obesity (BMI >=40); Obesity (BMI >=30); Hypertension; Cardiovascular disease; Cerebrovascular disease; Pulmonary hypertension; Chronic renal insufficiency or end stage renal disease; Cancer; Organ transplant recipient; Immunodeficiency; Inflammatory bowel disease; Liver disease; Chronic neurological or neuromuscular disease; Trisomy 21; Psychiatric condition; Macrophage activation syndrome (prior to COVID-19 diagnosis); Psoriasis; Pregnancy; Post-partum (< 6 weeks); Eczema; Congenital heart disease; Obstructive sleep apnea; and Lymphopenia.

^ Immunomodulatory medications included: Abatacept; Antifibrotics (pirfenidone; nintedinib); Antimalarials (including hydroxychloroquine; chloroquine); Apremilast; Azathioprine/6-MP; Belimumab; CD-20 inhibitors (including rituximab within last 12 months; ofatumumab); Cyclophosphamide; Cyclosporine; Denosumab; IL-1 inhibitors (including anakinra; canakinumab; rilonacept); IL-6 inhibitors (including tocilizumab; sarilumab); IL-12/23 inhibitors (ustekinumab); IL-17 inhibitors (including secukinumab; ixekizumab); IVIG; JAK inhibitors (including tofacitinib; baricitinib; upadacitinib); Leflunomide; Methotrexate; Mycophenolate mofetil/mycophenolic acid; Sulfasalazine; Tacrolimus; Thalidomide/lenalidomide; TNF-inhibitors (including infliximab; etanercept; adalimumab; golimumab; certolizumab; and biosimilars); Steroid eye drops; IL-23 inhibitors (guselkumab; risankizumab; tildrakizumab); and Colchicine.

Table 2.2. List of baseline predictors used in machine learning classifiers.

Demographics	Age	Continuous (years)
	Male gender	Dichotomous (yes, no)
	Current or former smoker	Dichotomous (yes, no)
	Current smoker	Dichotomous (yes, no)
Rheumatic Disease Diagnoses	Vasculitis	Dichotomous (yes, no)
	Anti-phospholipid antibody syndrome	Dichotomous (yes, no)
	Autoinflammatory syndrome	Dichotomous (yes, no)
	Spondyloarthritis	Dichotomous (yes, no)
	Behcet's	Dichotomous (yes, no)
	Chronic recurrent multifocal osteomyelitis	Dichotomous (yes, no)
	Giant cell arteritis	Dichotomous (yes, no)
	IgG4-related disease	Dichotomous (yes, no)
	Inflammatory myopathy	Dichotomous (yes, no)
	Juvenile idiopathic arthritis	Dichotomous (yes, no)
	Mixed connective tissue disease	Dichotomous (yes, no)
	Ocular inflammation	Dichotomous (yes, no)
	Polymyalgia rheumatica	Dichotomous (yes, no)
	Psoriatic arthritis	Dichotomous (yes, no)
	Rheumatoid arthritis	Dichotomous (yes, no)
	Other inflammatory arthritis	Dichotomous (yes, no)
	Sarcoidosis	Dichotomous (yes, no)
	Sjogren's syndrome	Dichotomous (yes, no)
	Systemic lupus erythematosus	Dichotomous (yes, no)
	Systemic sclerosis	Dichotomous (yes, no)
	Undifferentiated connective tissue disease	Dichotomous (yes, no)
	Gout	Dichotomous (yes, no)
	Calcium pyrophosphate deposition disease (CPPD)	Dichotomous (yes, no)
	Inclusion body myositis	Dichotomous (yes, no)
	Localized scleroderma	Dichotomous (yes, no)
Rheumatic Disease Activity	Moderate or high rheumatic disease activity	Dichotomous (yes, no)
Comorbidities	Interstitial lung disease	Dichotomous (yes, no)
	Obstructive lung disease	Dichotomous (yes, no)
	Diabetes	Dichotomous (yes, no)
	Morbid obesity (BMI >=40)	Dichotomous (yes, no)
	Obesity (BMI >=30)	Dichotomous (yes, no)
	Hypertension	Dichotomous (yes, no)
	Cardiovascular disease	Dichotomous (yes, no)
	Cerebrovascular disease	Dichotomous (yes, no)
	Pulmonary hypertension	Dichotomous (yes, no)
	Chronic renal insufficiency or end stage renal disease	Dichotomous (yes, no)
	Cancer	Dichotomous (yes, no)
	Organ transplant recipient	Dichotomous (yes, no)
	Immunodeficiency	Dichotomous (yes, no)

Comorbidities	Inflammatory bowel disease	Dichotomous (yes, no)
	Liver disease	Dichotomous (yes, no)
	Chronic neurological or neuromuscular disease	Dichotomous (yes, no)
	Trisomy 21	Dichotomous (yes, no)
	Psychiatric condition	Dichotomous (yes, no)
	Macrophage activation syndrome (prior to COVID-19 diagnosis)	Dichotomous (yes, no)
	Psoriasis	Dichotomous (yes, no)
	Pregnancy	Dichotomous (yes, no)
	Post-partum (< 6 weeks)	Dichotomous (yes, no)
	Eczema	Dichotomous (yes, no)
	Congenital heart disease	Dichotomous (yes, no)
	Obstructive sleep apnea	Dichotomous (yes, no)
	Lymphopenia	Dichotomous (yes, no)
Immunomodulatory Medications Used for Treating Rheumatic Disease	Abatacept	Dichotomous (yes, no)
	Antifibrotics (pirfenidone, nintedinib)	Dichotomous (yes, no)
	Antimalarials (including hydroxychloroquine, chloroquine)	Dichotomous (yes, no)
	Apremilast	Dichotomous (yes, no)
	Azathioprine / 6-MP	Dichotomous (yes, no)
	Belimumab	Dichotomous (yes, no)
	CD-20 inhibitors (including rituximab within last 12 months, ofatumumab)	Dichotomous (yes, no)
	Cyclophosphamide	Dichotomous (yes, no)
	Cyclosporine	Dichotomous (yes, no)
	Denosumab	Dichotomous (yes, no)
	IL-1 inhibitors (including anakinra, canakinumab, rilonacept)	Dichotomous (yes, no)
	IL-6 inhibitors (including tocilizumab, sarilumab)	Dichotomous (yes, no)
	IL-12/23 inhibitors (ustekinumab)	Dichotomous (yes, no)
	IL-17 inhibitors (including secukinumab, ixekizumab)	Dichotomous (yes, no)
	IVIG	Dichotomous (yes, no)
	JAK inhibitors (including tofacitinib, baricitinib, upadacitinib)	Dichotomous (yes, no)
	Leflunomide	Dichotomous (yes, no)
	Methotrexate	Dichotomous (yes, no)
	Mycophenolate mofetil / mycophenolic acid	Dichotomous (yes, no)
	Sulfasalazine	Dichotomous (yes, no)
	Tacrolimus	Dichotomous (yes, no)
	Thalidomide / lenalidomide	Dichotomous (yes, no)
	TNF-inhibitors (infliximab, etanercept, adalimumab, golimumab, certolizumab)	Dichotomous (yes, no)
	Steroid eye drops	Dichotomous (yes, no)
	IL-23 inhibitors (guselkumab, risankizumab, tildrakizumab)	Dichotomous (yes, no)
	Colchicine	Dichotomous (yes, no)
	Average daily prednisone-equivalent glucocorticoid dose	Continuous (mg)

Table 2.3. Demographic and clinical characteristics of the study population.

	Training set N = 5,673	Test set N = 891	Validation sets			
			Primary	Other		
			Italy N = 1,060	Sweden N = 225	Brazil N = 201	Argentina N = 583
Age, years, Mean (SD)	53.2 (15.2)	58.0 (17.1)	56.6 (14.6)	53.5 (14.7)	47.8 (14.1)	49.2 (14.2)
Sex						
Male	1,585 (27.9)	236 (26.5)	311 (29.3)	88 (39.1)	57 (28.4)	126 (21.6)
Female	4,088 (72.1)	655 (73.5)	749 (70.7)	137 (60.9)	144 (71.6)	457 (78.4)
Smoking status						
Never smoker	4,212 (74.2)	543 (60.9)	752 (70.9)	127 (56.4)	188 (93.5)	395 (67.8)
Former smoker	1,152 (20.3)	307 (34.5)	199 (18.8)	88 (39.1)	5 (2.5)	154 (26.4)
Current smoker	309 (5.4)	41 (4.6)	109 (10.3)	10 (4.4)	8 (4)	34 (5.8)
Most common diagnoses						
Rheumatoid arthritis	2,472 (43.6)	322 (36.1)	360 (34)	100 (44.4)	60 (29.9)	299 (51.3)
Psoriatic arthritis	569 (10)	81 (9.1)	220 (20.8)	46 (20.4)	23 (11.4)	47 (8.1)
Spondyloarthritis	554 (9.8)	45 (5.1)	108 (10.2)	40 (17.8)	54 (26.9)	48 (8.2)
Other inflammatory arthritis	145 (2.6)	63 (7.1)	12 (1.1)	12 (5.3)	0 (0)	0 (0)
Systemic lupus erythematosus	689 (12.1)	99 (11.1)	80 (7.5)	5 (2.2)	25 (12.4)	110 (18.9)
Vasculitis	171 (3)	49 (5.5)	40 (3.8)	8 (3.6)	1 (.5)	23 (3.9)
Sjorgen's	195 (3.4)	34 (3.8)	29 (2.7)	0 (0)	9 (4.5)	31 (5.3)
Polymyalgia rheumatica	102 (1.8)	47 (5.3)	25 (2.4)	0 (0)	0 (0)	3 (.5)
Systemic sclerosis	165 (2.9)	23 (2.6)	63 (5.9)	1 (.4)	11 (5.5)	20 (3.4)
Disease activity						
Remission or low	4,554 (80.3)	695 (78)	894 (84.3)	194 (86.2)	166 (82.6)	457 (78.4)
Moderate or high	1,119 (19.7)	196 (22)	166 (15.7)	31 (13.8)	35 (17.4)	126 (21.6)
Most common comorbidities						
None	2,040 (36)	182 (20.4)	315 (29.7)	102 (45.3)	81 (40.3)	317 (54.4)
At least 1 comorbidity	3,633 (64)	709 (79.6)	745 (70.3)	123 (54.7)	120 (59.7)	266 (45.6)
Interstitial lung disease	288 (5.1)	42 (4.7)	70 (6.6)	5 (2.2)	6 (3)	33 (5.7)
Obstructive lung disease	433 (7.6)	145 (16.3)	69 (6.5)	28 (12.4)	6 (3)	9 (1.5)
Obesity	926 (16.3)	273 (30.6)	131 (12.4)	16 (7.1)	26 (12.9)	93 (16)
Diabetes	786 (13.9)	167 (18.7)	102 (9.6)	15 (6.7)	20 (10)	52 (8.9)
Hypertension	1,921 (33.9)	412 (46.2)	365 (34.4)	56 (24.9)	67 (33.3)	161 (27.6)
CVD	463 (8.2)	129 (14.5)	169 (15.9)	21 (9.3)	13 (6.5)	19 (3.3)
Chronic kidney disease	274 (4.8)	114 (12.8)	66 (6.2)	3 (1.3)	8 (4)	17 (2.9)
Cancer	191 (3.4)	70 (7.9)	64 (6)	4 (1.8)	4 (2)	12 (2.1)
Liver disease	156 (2.7)	24 (2.7)	66 (6.2)	1 (.4)	0 (0)	8 (1.4)
Neurological or neuromuscular disease	77 (1.4)	40 (4.5)	53 (5)	6 (2.7)	0 (0)	5 (.9)

	Training set N = 5,673	Test set N = 891	Validation sets			
			Primary	Other		
			Italy N = 1,060	Sweden N = 225	Brazil N = 201	Argentina N = 583
Psychiatric disease	91 (1.6)	44 (4.9)	27 (2.5)	2 (.9)	2 (1)	22 (3.8)
Psoriasis	291 (5.1)	54 (6.1)	184 (17.4)	13 (5.8)	6 (3)	28 (4.8)
Medications						
No DMARDs	939 (16.6)	265 (29.7)	175 (16.5)	13 (5.8)	25 (12.4)	5 (.9)
csDMARDs alone	2,501 (44.1)	338 (37.9)	396 (37.4)	77 (34.2)	80 (39.8)	405 (69.5)
b/tsDMARDs alone	1,196 (21.1)	193 (21.7)	278 (26.2)	85 (37.8)	64 (31.8)	91 (15.6)
csDMARDs + b/tsDMARDs	1,037 (18.3)	95 (10.7)	211 (19.9)	50 (22.2)	32 (15.9)	82 (14.1)
Glucocorticoid (GC) use						
No use	3,942 (69.5)	635 (71.3)	659 (62.2)	172 (76.4)	180 (89.6)	335 (57.5)
GC user	1,731 (30.5)	256 (28.7)	401 (37.8)	53 (23.6)	21 (10.4)	248 (42.5)
GC dose*, mg, Median (IQR)	5 (5)	5 (5)	5 (0)	5 (2.5)	10 (5)	5 (5)
ARDS						
Yes	355 (6.3)	35 (3.9)	57 (5.4)	12 (5.3)	17 (8.5)	47 (8.1)
No	5,318 (93.7)	856 (96.1)	1,003 (94.6)	213 (94.7)	184 (91.5)	536 (91.9)

*Average daily prednisone-equivalent dose among GC users. CVD: cardiovascular disease; DMARD: disease modifying antirheumatic drug; csDMARDs: conventional systemic DMARDs; b/tsDMARDs: biologic or targeted synthetic DMARDs.

Table 2.4. Predictive performance of machine learning classifiers in the test set.

	Performance metric, Mean (95% CI)			
	AUC	**Accuracy***	**Sensitivity***	**Specificity***
KNN	0.69 (0.57-0.81)	0.66 (0.54-0.77)	0.66 (0.53-0.79)	0.66 (0.54-0.77)
SVM	0.55 (0.46-0.67)	0.55 (0.50-0.64)	0.55 (0.50-0.64)	0.55 (0.50-0.65)
GLMNET	0.76 (0.66-0.88)	0.70 (0.58-0.84)	0.70 (0.58-0.85)	0.70 (0.58-0.84)
BAYESGLM	0.74 (0.62-0.86)	0.67 (0.56-0.81)	0.67 (0.55-0.82)	0.67 (0.56-0.81)
GAM	0.76 (0.65-0.86)	0.71 (0.59-0.81)	0.71 (0.58-0.83)	0.71 (0.59-0.81)
GBM	0.78 (0.67-0.88)	0.70 (0.60-0.82)	0.70 (0.58-0.82)	0.70 (0.60-0.82)
NN	0.77 (0.64-0.87)	0.70 (0.55-0.83)	0.70 (0.55-0.83)	0.70 (0.55-0.83)

*Metrics obtained at the optimum classification threshold. Means and corresponding 95% confidence intervals were derived from 1000 random samples of 500 patients from the test set using bootstrapping and sampling with replacement. KNN: k-nearest neighbors; SVM: support vector machines; GLMNET: the lasso and elastic-net regularized generalized linear models; BAYESGLM: Bayesian generalized linear models; GAM: generalized additive models; GBM: gradient boosting machines; NN: neural networks.

Table 2.5. Predictive performance of the gradient boosting machine in the test set and across validation sets.

		Classification threshold, Mean (95% CI)	Percentile of predicted risk*	Performance metric, Mean (95% CI)			
				AUC	Accuracy	Sensitivity	Specificity
Test set		0.088 (0.071-0.116)	69.7 (59.3-80.8)	0.78 (0.67-0.88)	0.70 (0.61-0.82)	0.70 (0.58-0.83)	0.70 (0.61-0.82)
Primary validation set	Italy	0.073 (0.062-0.085)	71.1 (60.5-77.5)	0.79 (0.70-0.87)	0.73 (0.63-0.81)	0.73 (0.62-0.81)	0.73 (0.63-0.81)
Other validation sets	Sweden	0.055 (0.041-0.068)	66.2 (54.7-76.4)	0.79 (0.68-0.89)	0.69 (0.58-0.79)	0.69 (0.56-0.82)	0.69 (0.58-0.79)
	Brazil	0.042 (0.034-0.056)	63.7 (53.2-73.1)	0.74 (0.65-0.82)	0.66 (0.55-0.76)	0.65 (0.55-0.77)	0.66 (0.55-0.76)
	Argentina	0.063 (0.056-0.070)	73.9 (65.2-77.9)	0.85 (0.80-0.90)	0.78 (0.70-0.83)	0.78 (0.69-0.84)	0.78 (0.70-0.84)

*Percentiles of predicted risk correspond to the mean (95% CI) classification thresholds. Mean classification thresholds, mean performance metrics, and corresponding 95% confidence intervals were derived from 1000 random samples of 500 patients from each dataset using bootstrapping and sampling with replacement.

Table 2.6. Predictive performance of the multivariable logistic regression model in the test set and across validation sets.

		Classification threshold, Mean (95% CI)	Percentile of predicted risk*	Performance metrics, Mean (95% CI)			
				AUC	Accuracy	Sensitivity	Specificity
Test set		0.096 (0.074-0.128)	71.1 (58.8-80.4)	0.79 (0.68-0.88)	0.71 (0.62-0.82)	0.71 (0.61-0.83)	0.71 (0.62-0.82)
Primary validation set	Italy	0.069 (0.056-0.085)	70.7 (61.9-77.8)	0.77 (0.68-0.86)	0.73 (0.65-0.80)	0.73 (0.64-0.81)	0.73 (0.65-0.80)
Other validation sets	Sweden	0.058 (0.044-0.081)	70.7 (60.9-81.8)	0.82 (0.72-0.92)	0.74 (0.62-0.84)	0.74 (0.59-0.85)	0.74 (0.62-0.84)
	Brazil	0.036 (0.033-0.041)	60.2 (54.2-67.7)	0.71 (0.63-0.78)	0.61 (0.55-0.70)	0.61 (0.52-0.71)	0.61 (0.55-0.70)
	Argentina	0.060 (0.053-0.069)	75.5 (70.5-79.9)	0.85 (0.79-0.91)	0.80 (0.75-0.85)	0.80 (0.74-0.86)	0.80 (0.75-0.85)

*Percentiles of predicted risk correspond to the mean (95% CI) classification thresholds. Classification thresholds, performance metrics, and corresponding 95% confidence intervals were derived from 1000 random samples of 500 patients from each dataset using bootstrapping and sampling with replacement.

CHAPTER 3 : Country-level factors associated with COVID-19-related death in people with rheumatic disease: results from the COVID-19 Global Rheumatology Alliance Registry

Zara Izadi, Milena A Gianfrancesco, Gabriela Schmajuk, Lindsay Jacobsohn, Patricia Katz, Stephanie Rush, Clairissa Ja, Tiffany Taylor, Kie Shidara, Maria I Danila, Katherine D Wysham, Anja Strangfeld, Elsa F Mateus, Kimme L Hyrich, Laure Gossec, Loreto Carmona, Saskia Lawson-Tovey, Lianne Kearsley-Fleet, Martin Schaefer, Samar Al-Emadi, Jeffrey A Sparks, Tiffany Y-T Hsu, Naomi J Patel, Leanna Wise, Emily Gilbert, Alí Duarte-García, Maria O Valenzuela-Almada, Manuel F Ugarte-Gil, Lotta Ljung, Carlo A Scirè, Greta Carrara, Eric Hachulla, Christophe Richez, Patrice Cacoub, Thierry Thomas, Maria J Santos, Miguel Bernardes, Rebecca Hasseli, Anne Regierer, Hendrik Schulze-Koops, Ulf Müller-Ladner, Guillermo Pons-Estel, Romina Tanten, Romina E Nieto, Cecilia N Pisoni, Yohana S Tissera, Ricardo Xavier, Claudia D Lopes Marques, Gecilmara C S Pileggi, Philip C Robinson, Pedro M Machado, Emily Sirotich, Jean W Liew, Jonathan S Hausmann, Paul Sufka, Rebecca Grainger, Suleman Bhana, Monique Gore-Massy, Zachary S Wallace, and Jinoos Yazdany, on behalf of the Global Rheumatology Alliance Registry.

ABSTRACT

Background. Although clinical risk factors for poor COVID-19 outcomes have been identified in people with rheumatic disease, associations between environmental and societal factors and poor outcomes remain unknown. We investigated whether country-level environmental and societal factors were associated with COVID-19-related death in people with rheumatic disease.

Methods. Individual-level data on adults with rheumatic disease and COVID-19 were derived from the COVID-19 Global Rheumatology Alliance Registry from March 2020 to August 2021. Country-level covariates potentially associated with death were obtained from publicly available sources. Multivariable logistic regression was used to evaluate independent associations between regional characteristics and death, after controlling for known individual-level risk factors. A series of nested mixed-effects regression models with country as random effects were used to determine whether the regional covariates identified sufficiently explained country-level variations in death.

Findings. We included 14,044 patients, from 23 countries. Mean (SD) age was 54.4 (15.6) years and 10,178 (72.5%) were female. Factors positively associated with death included air pollution (odds ratio: 1.10 per 10-micrograms increase in fine particulate matter per cubic meter, 95% CI: 1.01-1.17), the share of population aged 65 or older (1.19 per 1% increase, 1.10-1.30) and population mobility (1.03 per 1% increase in number of visits to grocery and pharmacy stores, 1.02-1.05). Number of hospital beds (0.94 per unit increase per 1,000 population, 0.88-1.00), human development index (0.65 per 0.1-unit increase, 0.44-0.96), and government response stringency (0.83 per 10-unit increase in containment index, 0.74-0.93) were associated with fewer deaths. These factors sufficiently explained country-level variations in death (Intraclass correlation coefficient = 1.2%, 95% CI: 0.1-9.5%, p = 0.14) .

Interpretation. Our findings highlight the importance of environmental and societal factors as potential explanations of the disparate impact of COVID-19 on people with rheumatic disease, globally.

Funding. American College of Rheumatology and European Alliance of Associations for Rheumatology.

RESEARCH IN CONTEXT

Evidence before this study. PubMed was searched for articles published up to November 1, 2021, studying the association between country-level policies and socioeconomic resources and COVID-19 outcomes in people with rheumatic disease. We first included the Medical Subject Heading "COVID-19" AND the Medical Subject Heading Major Topic "Global Health/statistics and numerical data". This search was not restricted by language or type of publication. We found multiple ecological studies investigating the association between country-level factors (such as population-level burden of comorbidities, socioeconomic factors, environmental factors, and containment policies) and COVID-19 mortality rates. However, results were inconclusive across studies and reliable individual-level inferences may not be deducible from ecological study designs. We then performed a further unrestricted Medical Subject Headings search with "COVID-19" AND "rheumatology" OR "rheumatic disease" AND Title or Abstract search with "disparity" OR "disparities" to identify relevant studies among people with rheumatic disease. A small number of articles were identified including one study that highlighted regional disparities in the availability of national rheumatology society COVID-19 recommendations, including locally agreed protocols on disease-modifying antirheumatic drug use, in Africa, and another study that showed disparate COVID-19 outcomes among racial/ethnic minorities with rheumatic disease in the United States. However, in people with rheumatic disease, the impact of country-level policies and socioeconomic resources on global disparities in COVID-19 outcomes has not been characterized.

Added value of this study. In people with rheumatic disease, country-level characteristics including exposure to air pollutants, lower country socioeconomic status, higher demands on or lower capacity of health resources, fewer government-imposed containment policies, and

increased population mobility were associated with higher odds of COVID-19-related death independent of individual-level risk factors including age, rheumatic disease activity, immunosuppression, and comorbidities. Importantly, the inclusion of individuals as units of analysis in our study permits more reliable inferences about individuals' levels of risk than those obtained from ecological study designs.

Implications of all the available evidence. Study findings highlight the impact of societal policies and resources on COVID-19 outcomes in people with rheumatic disease, globally. These findings lay foundation for a new research agenda to address global disparities in COVID-19 outcomes in people with rheumatic disease.

INTRODUCTION

The current evidence has identified demographic and clinical risk factors associated with poor outcomes of SARS-CoV-2 infection in people with rheumatic disease.[12,78,79,84,90-97] While this has facilitated individual risk-stratification and guided rheumatic disease management decisions, neither the temporal dynamics of the pandemic nor the potential capacities of countries' healthcare systems have been assessed as additional factors of importance. In the general population, country-level estimates of COVID-19 case fatality rate have ranged from 0.5% to 20%.[98,99] Similarly, rates of excess death (from any cause) have varied significantly across countries, during the pandemic.[100] However, the underlying causes of global disparities in COVID-19 outcomes are not fully understood.

In people with rheumatic disease, in addition to regional differences in the distribution of individual-level risk factors associated with a poor COVID-19 prognosis, such as age, comorbidities, rheumatic disease activity, and treatments, several other factors may explain

global disparities in the risk of poor outcomes after SARS-CoV-2 infection. First, waves of SARS-CoV-2 infection occurred earlier in some countries and therefore observed differences in outcomes might reflect differences in quickly evolving management strategies in the first months of the pandemic. Second, differences in outcomes might reflect country-specific healthcare capacity to handle surges in the number of patients needing intensive care and other resources. Third, variations might reflect country-level differences in wealth or governmental response with mitigation strategies. Given the heightened interest in global health equity, examining diverse country-level factors including environmental and socioeconomic factors, healthcare resources, population health and demographics, COVID-19 containment policies, and individual behaviors, is important to unravel potential mechanisms of disparate COVID-19 outcomes in people with rheumatic disease, across nations.

A number of ecological studies have investigated regional variations in poor COVID-19 outcomes in the general population. However, consistent with the ecological fallacy,[101,102] inferences on individual-level risk may not be deducible from population-level measures of association due to loss of information. Individual-level databases, such as insurance claims or electronic health records, often lack readily accessible information on important clinical parameters (e.g., rheumatic disease activity or glucocorticoid dose). In addition, such databases usually operate nationally or sub-nationally, and thus cannot be used to evaluate the influence of country-level characteristics on health outcomes. The COVID-19 Global Rheumatology Alliance (GRA) registry, in contrast, is unique in its inclusion of cases of COVID-19 in people with prevalent rheumatic disease from around the world and comprehensive data regarding each individual's rheumatic disease characteristics as well as COVID-19 diagnosis and outcomes. We used data from the GRA registry linked to a robust array of country-level factors to investigate

potential mechanisms of the disparate impact of COVID-19 on people with rheumatic disease globally.

METHODS

Registry design. This study used data from the COVID-19 Global Rheumatology Alliance (GRA) Registry from March 12, 2020, to August 27, 2021. Details of the GRA registry have been described previously.[12,21] Briefly, data from adults with rheumatic disease diagnosed with confirmed or suspected COVID-19 are entered by rheumatology clinicians via two parallel international data entry portals.[56,57] Data entered include patient demographics, rheumatic disease characteristics, immunomodulatory medications used for the treatment of rheumatic disease, comorbidities, COVID-19 outcomes, and complications. Methods of COVID-19 diagnoses are indicated including one or more of polymerase chain reaction, antigen testing, antibody, metagenomic testing, CT scan, laboratory assay, or a presumptive diagnosis based on symptoms or close contact alone. Quality is assessed by data validation teams who remove all known or potential duplicates and address erroneous reports. The GRA registry contains only limited data; no personal identifiers, with the exception of COVID-19 diagnosis dates, are included. Due to the limited data and the non-interventional nature of the study, the GRA registry was determined to be nonhuman subjects research by the United Kingdom Health Research Authority, the University of Manchester (United Kingdom), and the University of California, San Francisco. An institutional review board or ethics committee approval or informed consent was therefore not required. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for reporting cohort studies.[103]

Study timelines. For each country, index date was defined as the first date a COVID-19 diagnosis was reported to the COVID-19 Data Repository by the Center for Systems Science and

Engineering (CSSE) at Johns Hopkins University.[104] The end of follow up was August 27, 2021, or the most recent COVID-19 diagnosis date reported to the GRA registry by the respective country, whichever occurred earlier (Figure 3.1).

Inclusion criteria. We included adults from countries that reported 100 or more adults to the GRA registry. This cluster size was chosen to increase statistical power to detect country-level variations in COVID-19 related death as identified by a non-null variance of the random effects.[105] For inclusion, patients were required to have a confirmed status of their highest COVID-19 illness severity-level which included one of the following: death; symptoms resolved at the time of data entry; not hospitalized > 30 days after initial diagnosis date; hospitalized and discharged; or not at risk of further interventions/death. In total, 14,044 people from 23 countries were included in the analyses (Table 3.1).

Outcome and covariates. The outcome, COVID-19-related death, was documented by the reporting clinician. Individual-level demographics, rheumatic disease characteristics, and comorbidities were obtained from the GRA registry and reflected data at the time of COVID-19 diagnosis. A country-specific variable for follow-up time was generated, defined as time (in months) between the date of an individual's COVID-19 diagnosis (as entered in the GRA registry) and their respective country's index date (Figure 3.1). Follow-up time reflected how health system gained experience over time treating COVID-19. Country-level covariates considered to be associated with an individual's odds of death were obtained from publicly available sources (Table 3.2). Variance-inflation-factor analysis was used to remove country-level covariates that were highly intercorrelated in their associations with death. The GRA registry included data on reporting clinicians' country, as well as state, if the reporting clinician was based in the U.S. We therefore retrieved state-level data for a subset of regional covariates

where this data was available (Table 3.2). Individual-level and regional data were merged such that dates of all time-varying regional covariates corresponded with each patient's COVID-19 diagnosis date or month, as appropriate. Conceptualized relationships between outcome, individual-level, and regional covariates are depicted using a simplified directed acyclic graph (DAG, Figure 3.2).

Missing data. All individual-level covariates had less than 10% missing (or unknown) data and all country-level covariates had less than 5% missing data. Missing values in individual-level covariates were imputed using predictive mean matching across 5 nearest neighbors and multiple imputation by chained equations. Record entry date was used to impute the date of COVID-19 diagnosis in 1,300/14,044 (9%) patients with a missing COVID-19 diagnosis date.

Statistical analysis. We used descriptive statistics to summarize the characteristics of the study population as well as baseline regional characteristics, stratified by World Health Organization (WHO) regions (Eastern Mediterranean, Europe, Pan America [further subdivided into Latin America and North America], South-East Asia, and Western Pacific). Changes in time-varying regional characteristics over the follow-up period were summarized using time-series scatter plots.

Independent associations between regional characteristics and an individual's odds of death, reported as odds ratios (OR) with 95% confidence intervals (CIs), were estimated using multivariable logistic regression after accounting for individual-level demographics, rheumatic disease characteristics, comorbidities, and follow-up time. Regional covariates included all variables listed in Table 3.2; briefly: population density, precipitation, temperature, air pollutants (as measured by fine particulate matter, $PM_{2.5}$), median age (as a proxy for the country's socioeconomic status)[106], human development index (a composite measure of life expectancy,

mean years of education, and gross national income per capita), number of hospital beds, share of the elderly population (as a proxy for population burden of comorbidities and burden on health resources), cardiovascular mortality, diabetes prevalence, mortality attributed to air pollution, SARS-CoV-2 reproduction rate (R), cumulative and incident COVID-19 death rates (reflecting multi-wave COVID-19 dynamics over time), government response (measured by containment index, a composite index based on thirteen response indicators such as closures, travel controls, stay-at-home requirements, public information campaigns, face covering requirements, contact tracing, testing and vaccination policies), and population mobility (measured by Google maps API tracking country-level movement of individuals). Individual-level demographics included age and sex. Individual-level rheumatic disease characteristics included diagnosis (rheumatoid arthritis [reference], psoriatic arthritis, spondyloarthritis, other inflammatory arthritis, systemic lupus erythematosus [SLE], vasculitis, and other diagnoses); rheumatic disease activity (remission [reference], low, moderate, and high); level of immunosuppression inferred by use of disease modifying antirheumatic drugs (DMARDs; conventional synthetic (cs)DMARDs only [reference], biologic or targeted synthetic (b/ts)DMARDs only, csDMARDs in combination with b/tsDMARDs, and no use of DMARDs); and average daily prednisone-equivalent glucocorticoid dose; average daily doses >60mg were Winsorized to 60mg as they were considered clinically high doses. Individual-level comorbidities included morbid obesity (defined as a BMI $\geq$40 kg/m^2), cardiovascular disease (coronary artery disease and congestive heart failure) or hypertension, lung disease (pulmonary hypertension, interstitial, obstructive or other lung disease), diabetes, and kidney disease (chronic kidney insufficiency or end stage kidney disease), each included as a dichotomous variable.

Follow-up time was included both as a continuous variable and as a quadratic term to improve model fit.

We used the shrinkage[107,108] property of mixed-effects regression to determine whether the regional covariates included in the model sufficiently explained the proportion of the total variance in an individual's odds of death that was accounted for by country-level clustering. A series of nested multivariable mixed-effects logistic regression models were fitted with country included as random effects and all covariates successively added as fixed effects. To sequentially capture any residual random-effects variance, we used our DAG to guide the order in which covariates were incorporated in nested mixed-effects models. A likelihood ratio test was performed to compare each model with an equivalent logistic regression model that did not include country random effects. An intra-class correlation coefficient (ICC) that approached zero and a statistically non-significant likelihood ratio test indicated sufficient explanation of the observed country-level variations in death.

All analyses were conducted using Stata software, version 16.0 (StataCorp). The threshold for statistical significance was 2-sided $P < .05$.

RESULTS

Of the 14,044 people included in the study, the majority were from Europe (6,369 [45.4%]) and North America (3,506 [25.0%]). Mean (SD) age was 54.4 (15.6) years and 10,178 (72.5%) were female. Rheumatoid arthritis, reported in 5,696 (40.6%) people, was the most common diagnosis, followed by SLE (1,650, 11.7%), and psoriatic arthritis (1,430, 10.2%). The majority of people were in remission or low disease activity (11,372, 81.0%) and csDMARDs, used in 6,491 (46.2%) people with or without glucocorticoids, were the most common treatment modality.

Common comorbidities included cardiovascular disease or hypertension (in 5,254 people, 37.4%), lung disease (2,018, 14.4%), diabetes (1,763, 12.6%), and kidney disease (823, 5.9%, Table 3.3). In total, 865 deaths were reported. Number of deaths and case fatality rate (%) by country among patients included in the analyses is provided in Figure 3.3.

The distribution of baseline country characteristics varied considerably across WHO regions (Table 3.4). The minimum country-specific follow-up time was 17 months. We therefore provide time-series plots over a period of 17 months from the index date. Latin America saw the fastest increases in cumulative COVID-19 death rates over time, followed by North America and Europe (Figure 3.4). Temporal trends for containment index, estimates of R (SARS-CoV-2 reproduction), and population mobility were similar across the WHO regions (Figures 3.4 and 3.5). Temporal trends in climatic factors (including temperature, precipitation, and $PM_{2.5}$) between March 2020 to July 2021 are provided in the supplement (Figure 3.6).

The multivariable logistic regression showed independent associations between $PM_{2.5}$ (OR: 1.10 per 10-micrograms increase per cubic meter, 95% CI: 1.01-1.17), proportion of population aged 65 or older (1.19 per % increase, 1.10-1.30), visits to grocery and pharmacy stores (1.03 per % increase in number of visits, 1.02-1.05), and visits to workplaces (1.02 per % increase in number of visits, 1.00-1.03) and higher odds of COVID-19 related death in people with rheumatic disease, after controlling for individual demographics and clinical characteristics. Conversely, median population age (OR: 0.83 per year, 95% CI: 0.78-0.89), number of hospital beds (0.94 per unit increase per 1,000 population, 0.88-1.00), human development index (0.65 per 0.1-unit increase, 0.44-0.96), containment index (0.83 per 10-unit increase, 0.74-0.93), and follow-up time (0.78 per month, 0.69-0.88) were independently associated with lower odds of death (Figure 3.7).

To determine whether the identified regional factors sufficiently explained the observed country-level variations in death, ICCs were derived from nested mixed-effects logistic regression models that sequentially incorporated regional covariates as fixed effects (Table 3.5). A base model including only patient demographics as fixed effects had an ICC of 14.2% (95% CI: 7.5-25.2%). Addition of individual-level rheumatic disease characteristics and comorbidities reduced the ICC to 10.1% (5.1-19.1%). Addition of follow-up time and regional factors further shrunk the ICC to 1.2% (0.1-9.5%). The resulting model was no longer favorable over a logistic regression model without country random effects (P for likelihood ratio test = 0.14) indicating that the identified regional factors had sufficiently explained the observed country-level variations in death.

Estimated associations between individual-level characteristics and death were consistent with previously published reports (Figure 3.8).[12,47-50,61,78,84]

DISCUSSION

To our knowledge, this is the first study that combines individual-level and regional data to investigate the independent association between regional parameters such as a country's climatic, societal, and economic factors, burden on healthcare resources, or pandemic policies, and an individual's odds of COVID-19-related death. We found that a range of factors related to geographical residence impacted COVID-19 outcomes. Lower country socioeconomic status, environmental exposures, higher demands on or lower capacity of health resources, and fewer government-imposed containment measures were independently associated with COVID-19-related death. In particular, specific regional factors associated with death included exposure to air pollutants, share of elderly population (aged 65 years or older), and population mobility (as proxied by visits to workplaces, grocery, and pharmacy stores). Factors associated with fewer

deaths included human development index (a composite measure of life-expectancy, education level, and income), number of hospital beds per 1,000 population, and containment index (a composite of 13 measures reflecting strictness of government response). These identified factors point to some potential mechanisms of the observed disparities in COVID-19 outcomes for people with rheumatic disease globally.

Poverty, limited health resources, and challenges in coordination of health and other social policies are existing obstacles to achieving global health equity in rheumatic disease burden and outcomes.[109] In the developing world, rheumatic conditions can cause significant morbidity and mortality due to physician shortages, poor health literacy, and reduced access to healthcare as well as mental, social, and emotional support systems.[110-112] Treatment of rheumatic disease in low-income countries is limited by the cost of immunosuppressive drugs, unavailability, and unaffordability of health insurance.[113] In addition, high costs and limited availability of laboratory and diagnostic tests result in delayed diagnoses which contribute to worse rheumatic disease prognoses.[114] Nonetheless, our ICC analysis showed that differences in the distribution of rheumatic disease characteristics and comorbidities associated with poor COVID-19 outcomes (such as higher rheumatic disease activity, high-dose glucocorticoid use, and pre-existing lung disease), accounted for a relatively small proportion (~29%, as determined by a reduction in ICC from 14.2% to 10.1%) of the country-level variation in COVID-19-related deaths. Our findings provide evidence for the importance of environmental and societal factors in mitigating risk for severe COVID-19 for people with rheumatic disease during the pandemic. These findings lay foundation for initiatives that seek to address global disparities in COVID-19 outcomes such as the American College of Rheumatology's Global Health Task Force.

Although a number of studies have investigated sources of regional variation in adverse COVID-19 outcomes using national[98,99,115-119] and subnational[120-122] data, results have been inconclusive, largely due to study design limitations and methodological inconsistencies. A recent study examined regional variation in temporal trajectories of COVID-19 case fatality rate (CFR) using country as the unit of analysis.[98] The study found CFRs to be positively associated with share of the elderly population (aged 70 years or older) and negatively associated with number of hospital beds per 1,000 population. In contrast to our findings however, the study reported higher CFRs with increasing levels of government response stringency, and also gross domestic product (GDP) per capita and total health expenditure as share of GDP. Despite controlling for differences in testing strategies among countries, the study also found that share of death due to lower respiratory infections was negatively associated with COVID-19 CFRs. These findings may be biased from a lack of accounting for individual-level risk factors. Another study that used U.S. states and boroughs as the unit of analysis, identified population density and population mobility as important factors contributing to COVID-19 infection and mortality rates, although this study did not account for measures of population health or burden of comorbidities.[121] Our findings are therefore largely consistent with other studies to date despite methodological differences. Importantly, the inclusion of individuals as units of analysis in our study permits more reliable inferences about individual-level risk than those obtained from ecological study designs and provides insight into inconsistencies observed in previously published research.

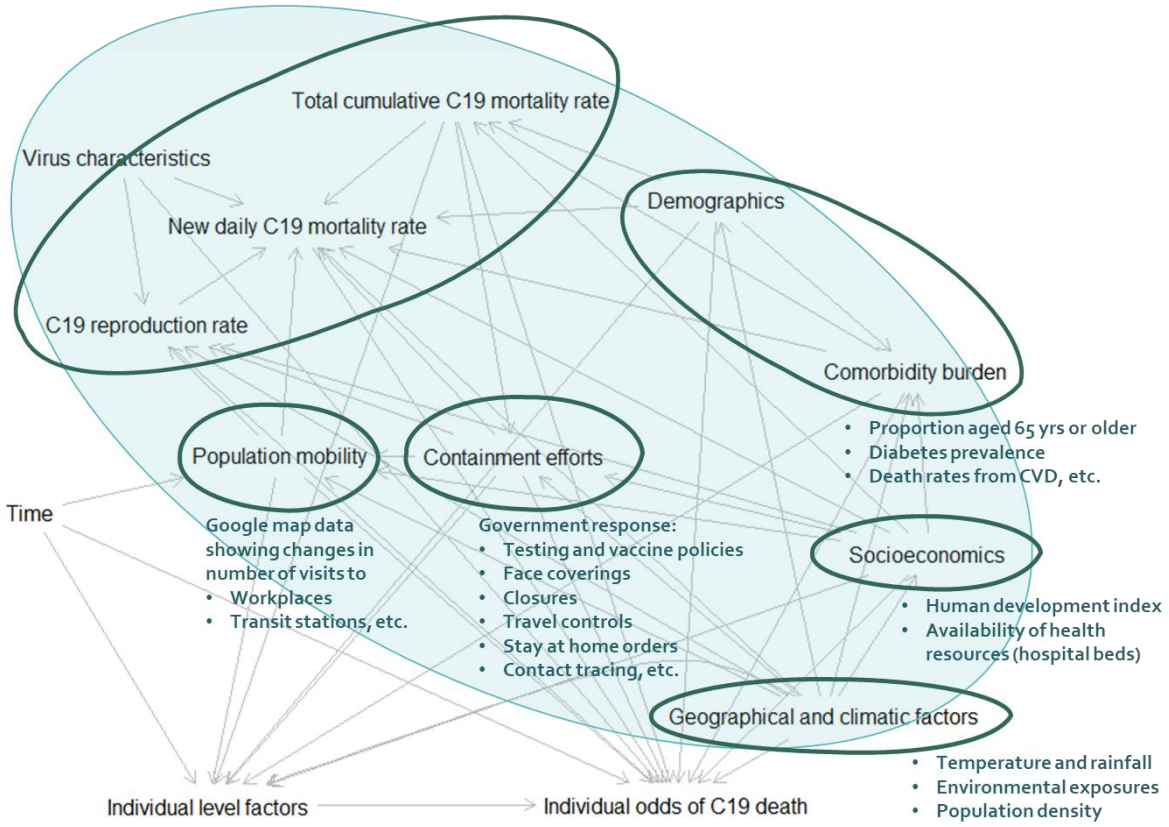
Strengths of this study include the large size of the study population with individuals from 23 countries and four continents. This captured a wide diversity in the distribution of regional covariates and powered the study to detect associations between regional characteristics and COVID-19-related death. Furthermore, our methodological approach enabled us to demonstrate

that the identified regional characteristics almost entirely explained the observed country-level variations in death. Limitations of the work include the potential for provider reporting-bias, as the GRA registry used convenience sampling. However, recently published results from the registry suggested reporting-bias did not substantially impact estimates of association.[123] Reporting biases could partially explain the country-level case fatality rates reported in this study; importantly, these rates should not be taken as an estimate of the overall death rate among patients with rheumatic diseases and COVID-19. While all-cause mortality rates facilitate a more accurate assessment of global disparities by additionally capturing the indirect impact of COVID-19 during the pandemic, we were not able to study all-cause death as the GRA registry did not include individuals without a diagnosis of COVID-19. Further studies would be fruitful in determining whether findings remain consistent for all-cause death during the pandemic. Misclassification of regional covariates, in particular mobility trends, is plausible as individuals can opt not to use location services, and likely differential with respect to region. While we suspect this misclassification to be non-differential with respect to an individual's risk for COVID-19-related death, it remains a potential source of bias. We relied on geocoding of the reporting clinician to determine U.S. state which may not match the patient, particularly for those who live on the borders or in areas where regions are relatively closely aggregated (e.g., U.S. Northeast). This study may have limited generalizability as regional variations generally increase with number, size, and granularity of the regions. Thus, with larger study populations (inclusion of more countries and more patients from each country) and more granular data (such as county- or health-system-level data), additional regional parameters and further considerations of their complex high-dimensional relationships will be needed to account for regional variations in COVID-19 outcomes. Finally, it was not possible to account for additional important individual-

level risk factors such as socioeconomic status, vaccination status, severity, and treatment of COVID-19.

In conclusion, our results suggest that among people with rheumatic disease, time-period of the epidemic wave, exposure to air pollutants, regional socioeconomic factors, availability and burden on health resources, government response stringency, and population mobility are associated with COVID-19-related death, independent of individuals' demographics, rheumatic disease characteristics, and comorbidities. These findings highlight the importance of environmental and societal factors as potential explanations of the observed global health disparities during the pandemic and lay foundation for a new research agenda to address regional disparities in COVID-19 outcomes in people with rheumatic disease. The novel multilayer epidemiological infrastructure and methods exemplified by this study have broad applications not only to research that seeks to understand modifiable determinants of human health but also will be key to addressing other pressing global issues, such as climate change.

Figure 3.1. Study timeline.



Shaded green area represents regional covariates. To simplify the DAG, regional covariates were grouped into domains represented by dark green circles. For domains that included more than one regional covariate, some examples are provided in green text. C19: COVID-19.

Figure 3.2. Directed acyclic graph showing conceptualized relationships between outcome, individual-level factors, and the regional covariates.

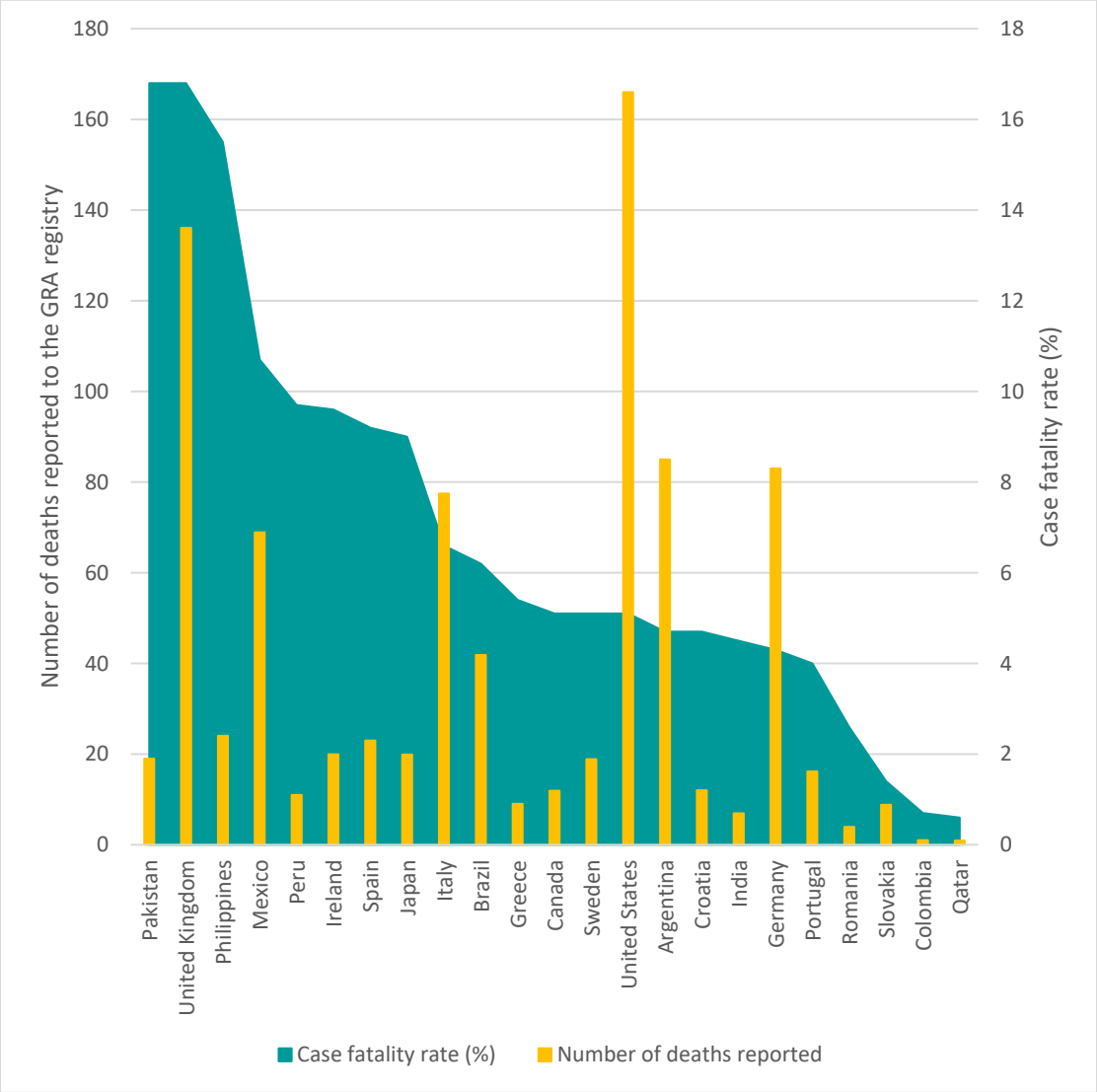
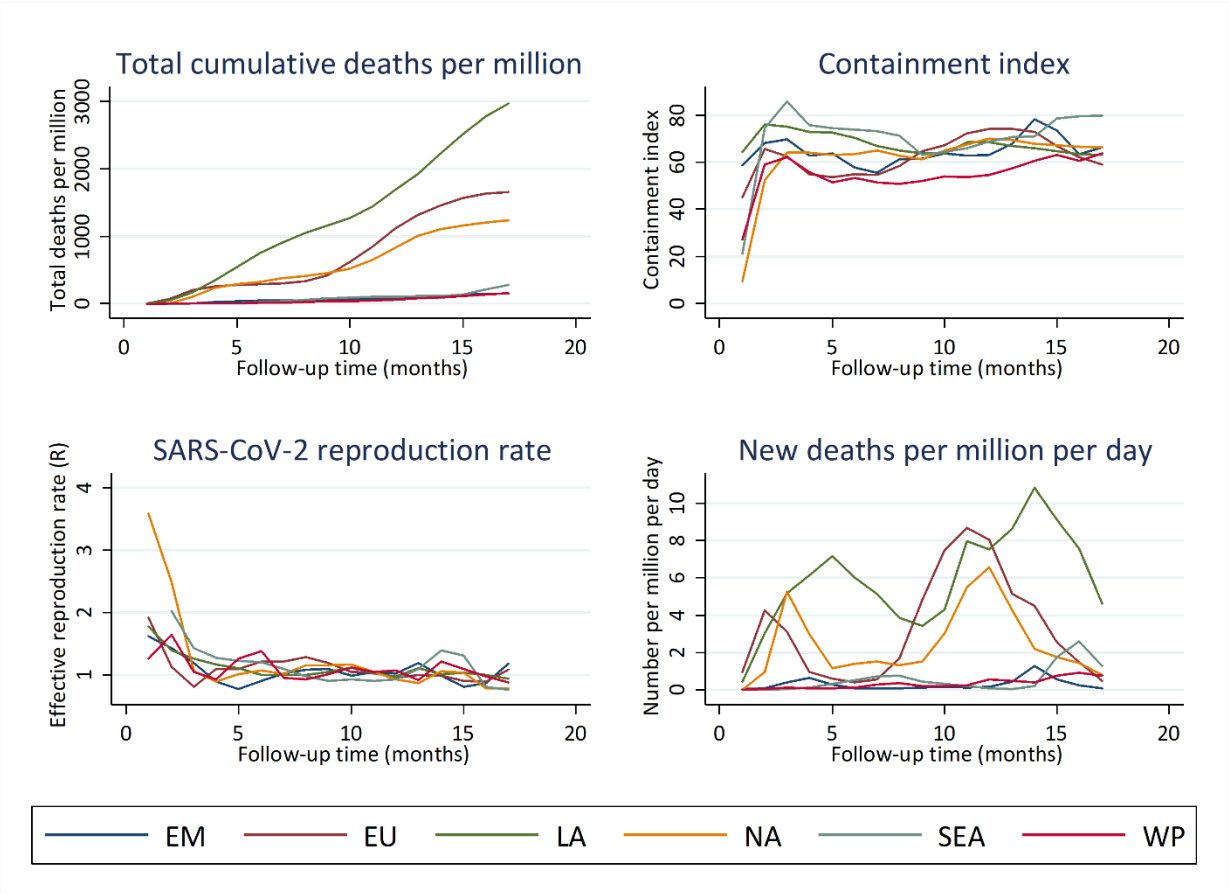
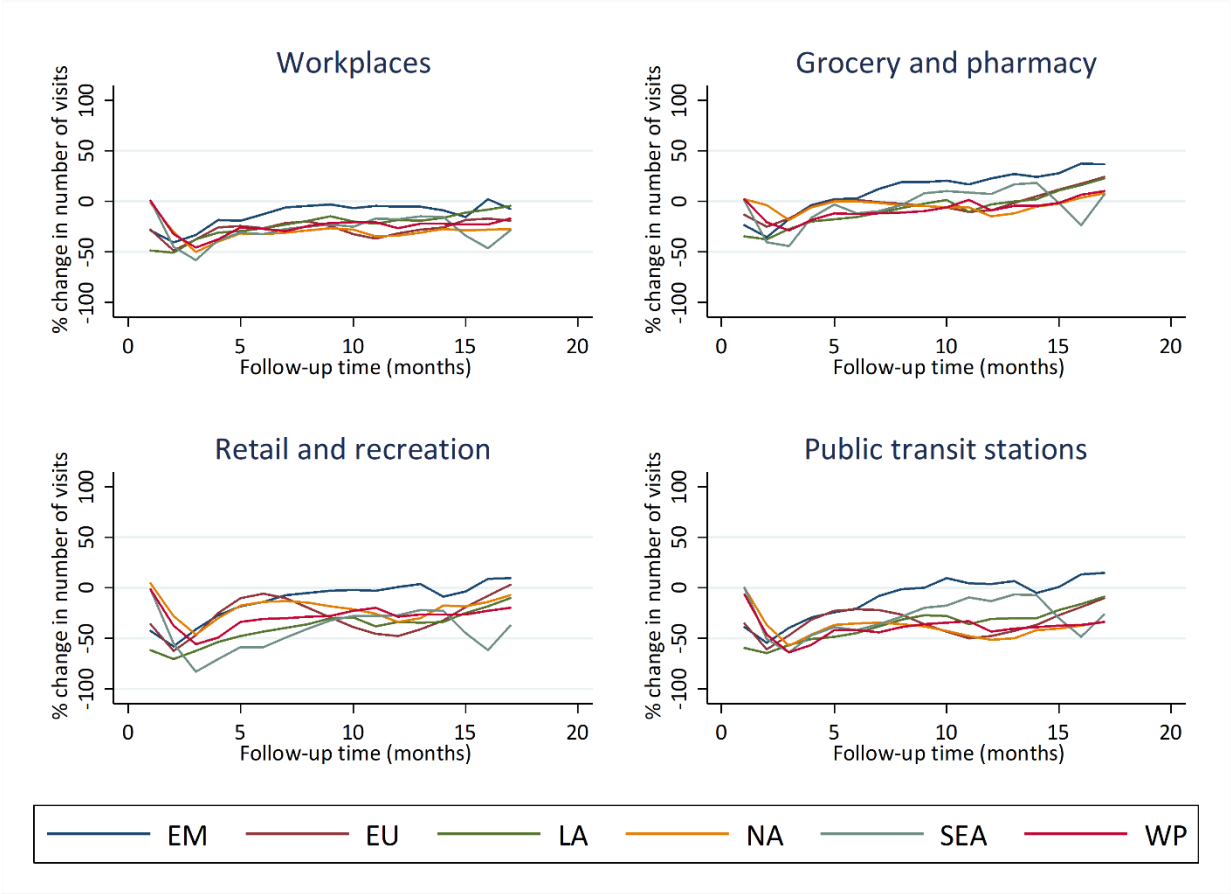


Figure 3.3. Number of deaths reported and case fatality rate by country among patients included in the analyses.



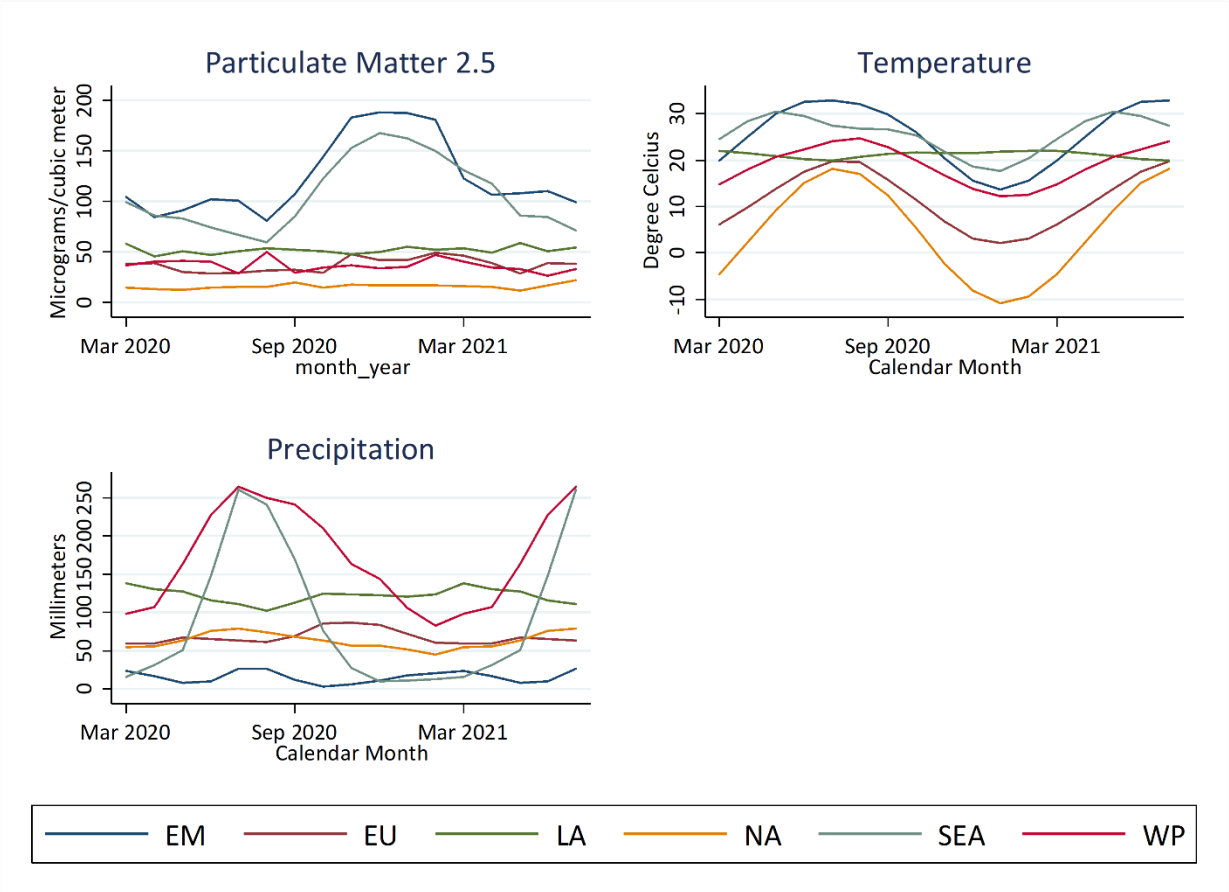
Data sourced from Our World in Data. Global regions were adapted from WHO regions. EM: Eastern Mediterranean; EU: Europe; LA: Latin America; NA: North America; SEA: South-East Asia; WP: Western Pacific.

Figure 3.4. Temporal trends in COVID-19 measures and containment index across six global regions.



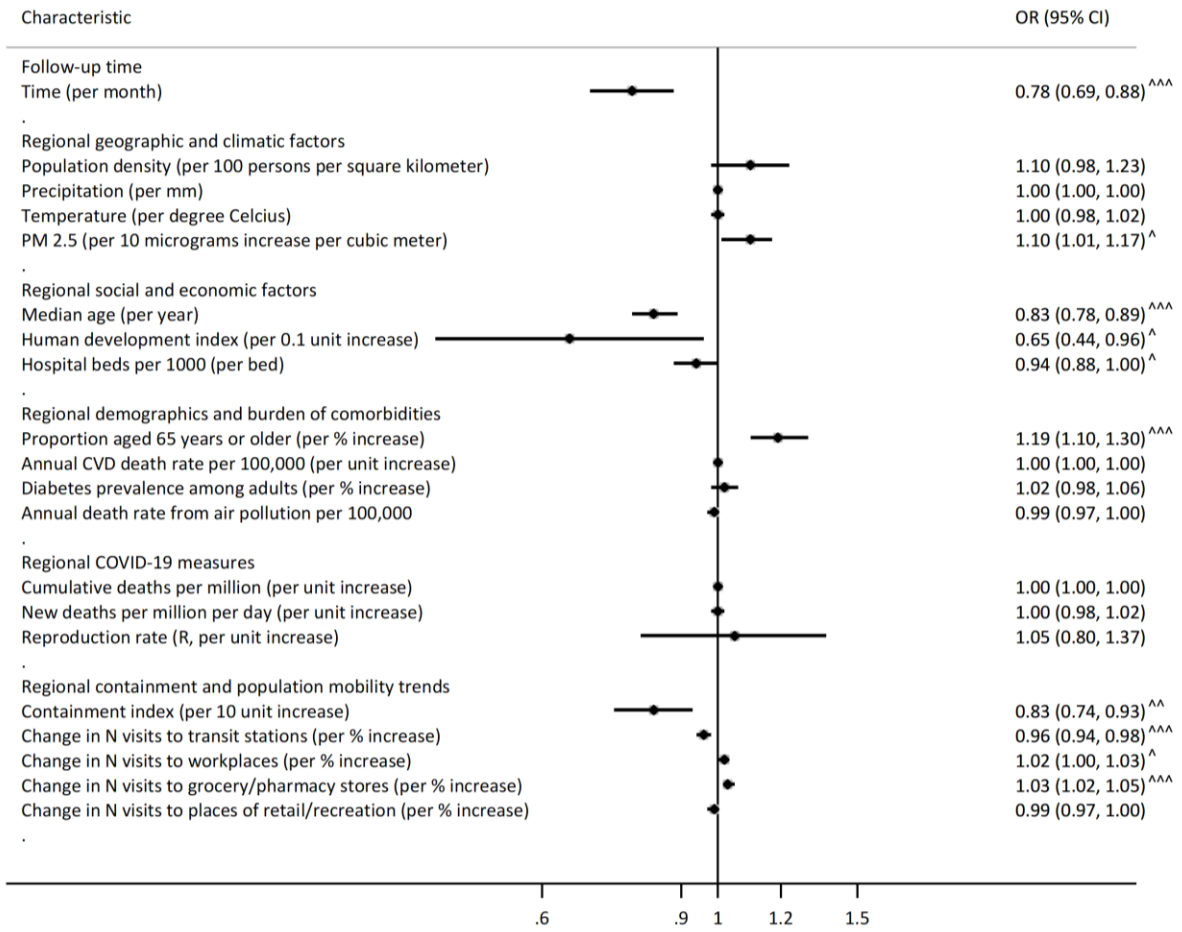
Google data sourced from Our World in Data. Global regions were adapted from WHO regions. EM: Eastern Mediterranean; EU: Europe; LA: Latin America; NA: North America; SEA: South-East Asia; WP: Western Pacific.

Figure 3.5. Temporal trends in population mobility across six global regions.



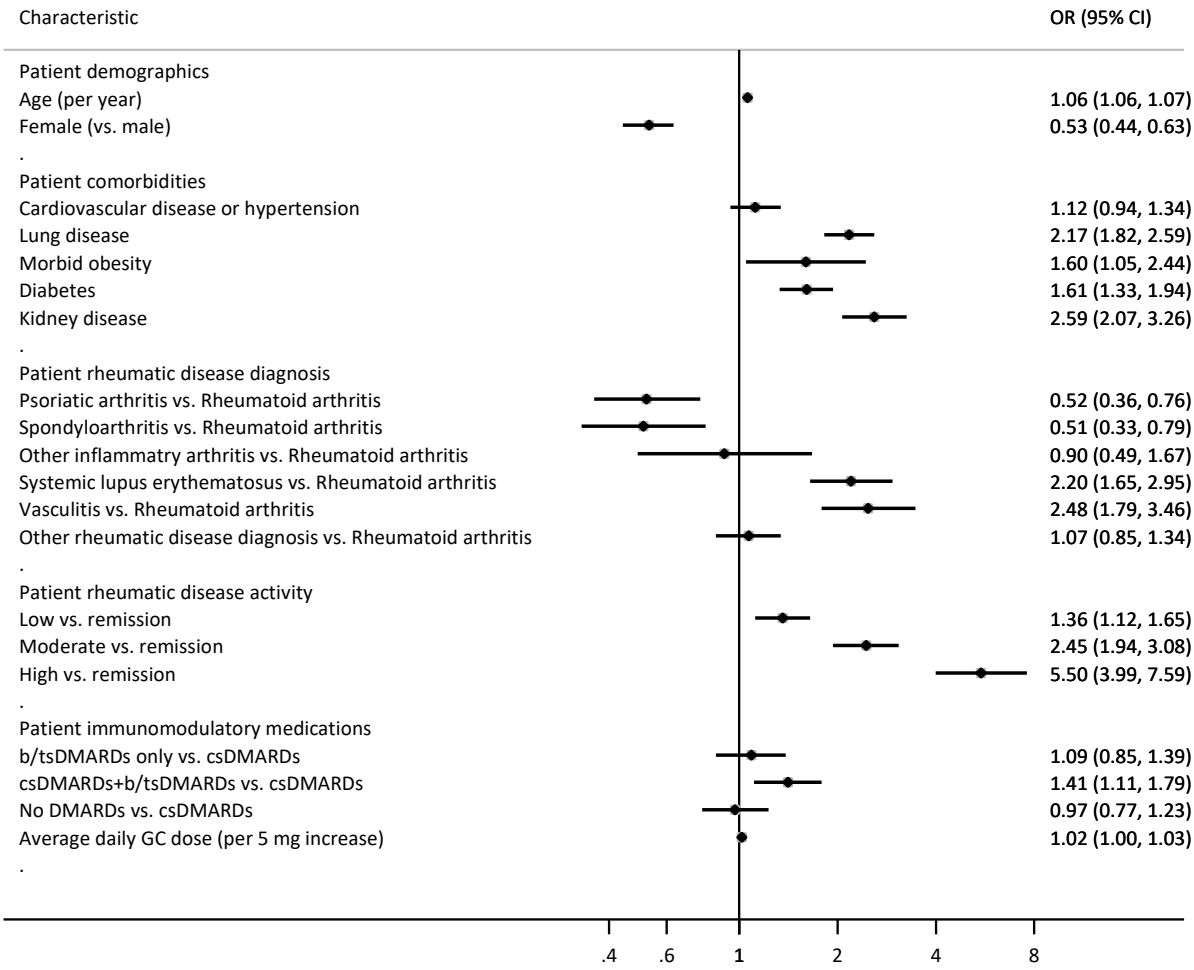
Data on precipitation and temperature sourced from World Bank Climate Knowledge Portal and projected using data from 1991-2020. Data on particulate matter (PM) 2.5 sourced from Air Quality Open Data Platform by the World Air Quality Project for countries, and United States Environmental Protection Agency for U.S. states. Global regions were adapted from WHO regions. EM: Eastern Mediterranean; EU: Europe; LA: Latin America; NA: North America; SEA: South-East Asia; WP: Western Pacific.

Figure 3.6. Temporal trends in climatic factors across six global regions.



Odds ratios derived from a multivariable logistic regression model including all covariates shown as well as individual-level demographics and clinical characteristics and follow-up time as a polynomial term. Individual-level demographics included age and sex. Individual-level clinical characteristics included rheumatic disease diagnosis (rheumatoid arthritis, psoriatic arthritis, spondyloarthritis, other inflammatory arthritis, systemic lupus erythematosus, vasculitis, and other diagnoses); rheumatic disease activity (remission, low, moderate, and high); important comorbidities including cardiovascular disease or hypertension, lung disease, morbid obesity, diabetes, and kidney disease; immunomodulatory medications (conventional systemic disease modifying antirheumatic drugs [DMARDs] only, biologic or targeted synthetic DMARDs only, conventional synthetic DMARDs in combination with biologic or targeted synthetic DMARDs, and no use of DMARDs); and average daily prednisone-equivalent glucocorticoid dose. Regional characteristics include country-level and U.S. state-level characteristics. *One potential explanation for this inverse association is that travel restrictions were more frequently in place during periods of COVID-19 peak than restrictions that governed visits to grocery and pharmacy stores or workplaces. Another potential explanation may be residual confounding by country-level socioeconomic status, i.e., populations from countries with higher socioeconomic status were more likely to travel nationally and internationally but less likely to experience adverse COVID-19 outcomes.^^^P<0.001. ^^P<0.01. ^P<0.05. PM2.5: fine particulate matter air pollutants; CVD: cardiovascular disease.

Figure 3.7. Associations between regional characteristics and COVID-19-related death, independent of individual-level demographics and clinical risk factors.



Multivariable logistic regression models adjusting for all covariates shown as well as follow-up time (as a main term and a quadratic term) and regional covariates including population density, precipitation, temperature, particulate matter 2.5 (PM 2.5), median age, human development index, number of hospital beds, proportion aged 65 years or older, cardiovascular mortality, diabetes prevalence, mortality attributed to air pollution, SARS-CoV-2 reproduction rate (R), cumulative and incident COVID-19 death rates, and population mobility trends. DMARDs: disease modifying antirheumatic drugs; b/tsDMARDs: biologic/targeted synthetic DMARDs; csDAMRDs: conventional synthetic DMARDs; GC: glucocorticoid. Regional characteristics include country-level and U.S. state-level characteristics.

Figure 3.8. Odds ratios of COVID-19-related death corresponding to individual-level characteristics, after controlling for regional and temporal confounding.

Table 3.1. Countries and U.S. states included in the analyses.

Country	N (%)	U.S. States	N (%)	U.S. States	N (%)
United States	3272 (23.3)	Massachusetts	811 (24.8)	Nebraska	31 (1.0)
Germany	1948 (13.9)	California	421 (12.9)	Washington	28 (0.9)
Argentina	1791 (12.8)	Florida	215 (6.6)	Arizona	26 (0.8)
Italy	1174 (8.4)	Minnesota	181 (5.5)	Georgia	17 (0.5)
United Kingdom	810 (5.8)	New Jersey	181 (5.5)	New Mexico	8 (0.2)
Brazil	675 (4.8)	North Carolina	178 (5.4)	Arkansas	7 (0.2)
Mexico	644 (4.6)	Texas	151 (4.6)	Idaho	5 (0.2)
Slovakia	630 (4.5)	Illinois	125 (3.8)	Oregon	4 (0.1)
Portugal	404 (2.9)	Pennsylvania	94 (2.9)	South Carolina	4 (0.1)
Sweden	370 (2.6)	Kentucky	84 (2.6)	Maine	3 (0.1)
Croatia	256 (1.8)	Ohio	82 (2.5)	Connecticut	2 (0.1)
Spain	250 (1.8)	New York	78 (2.4)	Indiana	2 (0.1)
Canada	234 (1.7)	Tennessee	74 (2.3)	South Dakota	2 (0.1)
Japan	221 (1.6)	Iowa	63 (1.9)	New Hampshire	1 (0.0)
Ireland	208 (1.5)	Colorado	49 (1.5)	Oklahoma	1 (0.0)
Greece	167 (1.2)	Louisiana	49 (1.5)	Puerto Rico	1 (0.0)
Philippines	155 (1.1)	Virginia	48 (1.5)	Utah	1 (0.0)
Qatar	155 (1.1)	Alabama	46 (1.4)		
India	154 (1.1)	Maryland	46 (1.4)		
Romania	152 (1.1)	Michigan	43 (1.3)		
Colombia	148 (1.1)	Missouri	41 (1.3)		
Pakistan	113 (0.8)	District of Columbia	36 (1.1)		
Peru	113 (0.8)	Wisconsin	33 (1.0)		

Table 3.2. Regional covariate definitions and source datasets.

Construct	Variable	Definition	Type	Regional level	Source	Period
Geographic	Population density	Population divided by land area, in square kilometers	Numeric, Baseline	Country and U.S. State	Countries: World Bank World Development Indicators, sourced from Food and Agriculture Organization and World Bank estimates; U.S. States: United States Census Bureau.	Countries: Most recent year available; U.S. States: 2020
Climatic	Precipitation	Average monthly precipitation in millimeters	Numeric, Time-dependent	Country and U.S. State	Countries: World Bank Climate Change Knowledge Portal; U.S. States: World Bank Climate Change Knowledge Portal	Countries: 1991-2020; U.S. States: 1991-2020
	Temperature	Average monthly temperature in degree Celsius	Numeric, Time-dependent	Country and U.S. State	Countries: World Bank Climate Change Knowledge Portal; U.S. States: World Bank Climate Change Knowledge Portal	Countries: 1991-2020; U.S. States: 1991-2020
	Fine particulate matter ($PM_{2.5}$)	Average monthly particulate matter 2.5 in micrograms per cubic meter of air	Numeric, Time-dependent	Country and U.S. State	Countries: Air Quality Open Data Platform by the World Air Quality Project; U.S. States: United States Environmental Protection Agency	Countries: Current, Average monthly: States: Current, Average monthly
Social and economic measures of development	Median age	Median age of the population in years	Numeric, Baseline	Country and U.S. State	Countries: UN Population Division, World Population Prospects, 2017 Revision; U.S. States: United States Census Bureau.	Countries: UN projection for 2020; U.S. States: 2019
	Life expectancy*	Life expectancy at birth defined as the average number of years that a newborn could expect to live if he or she were to pass through life subject to the age-specific mortality rates of a given period	Numeric, Baseline	Country and U.S. State	Countries: Our World in Data based on estimates by James C. Riley, Clio Infra, and the United Nations Population Division. U.S. States: County Health Rankings and Roadmaps by the University of Wisconsin Population Health Institute.	Countries: 2019; U.S. States: 2018
	Human Development Index	A composite index defined as the geometric mean of normalized indices in three dimensions including life expectancy at birth, mean of years of schooling for adults aged 25 years and more and expected years of schooling for children of school entering age, and gross national income per capita; HDI is ranked on a scale from 0 to 1.0, with 1.0 being the highest HDI.	Numeric, Baseline	Country and U.S. State	Countries: United Nations Development Programme; U.S. States: Global Data Lab by the Institute for Management Research at Radbound University	Countries: 2019; U.S. States: 2019
	Hospital beds	Number of hospital beds per 1,000 population	Numeric, Baseline	Country and U.S. State	Countries: Our World In Data, sourced from the Organisation for Economic Co-operation and Development, Eurostat, World Bank, national government records and other sources; U.S. States: Global Health Data Exchange by The Institute for Health Metrics and Evaluation	Countries: Most recent year available since 2010; U.S. States: 2019

Construct	Variable	Definition	Type	Regional level	Source	Period
Population demographic	Proportion aged 65 or older	Proportion of the population that is 65 years and older	Numeric, Baseline	Country and U.S. State	Countries: World Bank World Development Indicators based on age/sex distributions of United Nations World Population Prospects, 2017 Revision; U.S. States: Population Reference Bureau	Countries: Most recent year available; U.S. States: 2018
Population burden of comorbidities	Death rate from cardiovascular disease	Annual number of deaths per 100,000 population	Numeric, Baseline	Country and U.S. State	Countries: Global Burden of Disease Collaborative Network by the Institute for Health Metrics and Evaluation; U.S. States: Centers for Disease Control and Prevention, National Center for Health Statistics	Countries: 2017; U.S. States: 2019
	Diabetes prevalence	Proportion of adult population with diabetes	Numeric, Baseline	Country and U.S. State	Countries: World Bank World Development Indicators, sourced from International Diabetes Federation, Diabetes Atlas; U.S. States: Centers for Disease Control and Prevention, Diagnosed Diabetes	Countries: 2017; U.S. States: 2018
	Death rate from air pollution	Annual number of deaths per 100,000 population from both outdoor and indoor air pollution	Numeric, Baseline	Country	Global Burden of Disease Collaborative Network by the Institute for Health Metrics and Evaluation (IHME)	2017
COVID-19 measures	Total death rate	Total cumulative deaths attributed to COVID-19 per day per million population	Numeric, Time-dependent	Country	Our World In Data, sourced from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University	Current, Daily
	New death rate	New deaths attributed to COVID-19 per day per million population	Numeric, Time-dependent	Country	Our World In Data, sourced from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University	Current, Daily
	Reproduction rate (R)	Daily estimates of the effective SARS-CoV-2 reproduction rate (R)	Numeric, Time-dependent	Country	Our World In Data, based on Arroyo Marioli et al. (2020). https://doi.org/10.2139/ssrn.3581633	Current, Daily
Government response	Containment index	A composite index recorded daily based on thirteen government response indicators including school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, international travel controls, testing policy, the extent of contact tracing, requirements to wear face coverings, and policies around vaccine rollout, rescaled to a value from 0 to 100, with 100 representing strictest response	Numeric, Time-dependent	Country	Oxford COVID-19 Government Response Tracker, Blavatnik School of Government	Current, Daily

Construct	Variable	Definition	Type	Regional level	Source	Period
Population mobility trends	Places of retail and recreation	Percentage change in the number of visitors per day (calculated as a rolling 7-day average) to places of retail and recreation compared to the median value for the 5-week period from January 3 to February 6, 2020	Numeric, Time-dependent	Country	Our World In Data, sourced from Google	Current, Daily
	Grocery and pharmacy stores	Percentage change in the number of visitors per day (calculated as a rolling 7-day average) to grocery and pharmacy stores compared to the median value for the 5-week period from January 3 to February 6, 2020	Numeric, Time-dependent	Country	Our World In Data, sourced from Google	Current, Daily
	Transit stations	Percentage change in the number of visitors per day (calculated as a rolling 7-day average) to transit stations compared to the median value for the 5-week period from January 3 to February 6, 2020	Numeric, Time-dependent	Country	Our World In Data, sourced from Google	Current, Daily
	Workplaces	Percentage change in the number of visitors per day (calculated as a rolling 7-day average) to workplaces compared to the median value for the 5-week period from January 3 to February 6, 2020	Numeric, Time-dependent	Country	Our World In Data, sourced from Google	Current, Daily

*Life expectancy was not used as a covariate due to strong collinearity with human development index.

Table 3.3. Patient characteristics grouped into six global regions.

	EM	EU	SEA	WP	NA	LA	Total
	N = 268	**N = 6369**	**N = 154**	**N = 376**	**N = 3506**	**N = 3371**	**N = 14044**
Demographics							
Age, Mean (SD)	46.5 (13.1)	56.2 (15.5)	47.0 (13.5)	55.1 (17.6)	56.0 (15.8)	50.2 (14.4)	54.4 (15.6)
Sex, N (%)							
Male	89 (33.2)	2,080 (32.7)	38 (24.7)	144 (38.3)	905 (25.8)	610 (18.1)	3,866 (27.5)
Female	179 (66.8)	4,289 (67.3)	116 (75.3)	232 (61.7)	2,601 (74.2)	2,761 (81.9)	10,178 (72.5)
Rheumatic Disease Characteristics							
Diagnosis, N (%)							
RA	125 (46.6)	2,607 (40.9)	68 (44.2)	124 (33.0)	1,413 (40.3)	1,359 (40.3)	5,696 (40.6)
PsA	16 (6.0)	960 (15.1)	9 (5.8)	9 (2.4)	359 (10.2)	77 (2.3)	1,430 (10.2)
SpA	20 (7.5)	850 (13.3)	14 (9.1)	6 (1.6)	178 (5.1)	293 (8.7)	1,361 (9.7)
Other IA	9 (3.4)	109 (1.7)	2 (1.3)	3 (0.8)	161 (4.6)	8 (0.2)	292 (2.1)
SLE	45 (16.8)	380 (6.0)	15 (9.7)	82 (21.8)	434 (12.4)	694 (20.6)	1,650 (11.7)
Vasculitis	6 (2.2)	195 (3.1)	6 (3.9)	26 (6.9)	135 (3.9)	95 (2.8)	463 (3.3)
Other diagnoses	47 (17.5)	1,268 (19.9)	40 (26.0)	126 (33.5)	826 (23.6)	845 (25.1)	3,152 (22.4)
Disease activity, N (%)							
Remission	140 (52.2)	2,704 (42.5)	54 (35.1)	154 (41.0)	837 (23.9)	1,443 (42.8)	5,332 (38.0)
Low	72 (26.9)	2,644 (41.5)	78 (50.6)	150 (39.9)	1,885 (53.8)	1,211 (35.9)	6,040 (43.0)
Moderate	44 (16.4)	839 (13.2)	13 (8.4)	46 (12.2)	669 (19.1)	573 (17.0)	2,184 (15.6)
High	12 (4.5)	182 (2.9)	9 (5.8)	26 (6.9)	115 (3.3)	144 (4.3)	488 (3.5)
Medications, N (%)							
csDMARDs only	180 (67.2)	2,608 (40.9)	132 (85.7)	185 (49.2)	1,335 (38.1)	2,051 (60.8)	6,491 (46.2)
b/tsDMARDs only	23 (8.6)	1,687 (26.5)	0 (0.0)	20 (5.3)	768 (21.9)	506 (15.0)	3,004 (21.4)
cs+b/ts DMARDs	17 (6.3)	1,094 (17.2)	16 (10.4)	20 (5.3)	689 (19.7)	447 (13.3)	2,283 (16.3)
No DMARDs	48 (17.9)	980 (15.4)	6 (3.9)	151 (40.2)	714 (20.4)	367 (10.9)	2,266 (16.1)
GC use, N (%)	59 (0.22)	1,952 (0.31)	82 (0.53)	174 (0.46)	879 (0.25)	1,201 (0.36)	4,347 (0.310)
Comorbidities[§], N (%)							
Morbid obesity^, N (%)	3 (1.2)	50 (0.8)	2 (1.4)	2 (0.5)	281 (8.4)	51 (1.7)	389 (3.0)
CVD or Hypertension, N (%)	93 (34.7)	2,417 (37.9)	41 (26.6)	137 (36.4)	1,504 (42.9)	1,062 (31.5)	5,254 (37.4)
Lung disease, N (%)	35 (13.1)	916 (14.4)	10 (6.5)	69 (18.4)	691 (19.7)	297 (8.8)	2,018 (14.4)
Diabetes, N (%)	71 (26.5)	675 (10.6)	25 (16.2)	70 (18.6)	596 (17.0)	326 (9.7)	1,763 (12.6)
Kidney disease, N (%)	22 (8.2)	335 (5.3)	0 (0.0)	30 (8.0)	306 (8.7)	130 (3.9)	823 (5.9)
Cancer, N (%)	4 (1.6)	244 (4.0)	0 (0.0)	11 (3.0)	194 (5.8)	61 (2.1)	514 (3.9)
Death, N (%)	20 (7.5)	408 (6.4)	7 (4.5)	44 (11.7)	178 (5.1)	208 (6.2)	865 (6.2)

Global regions were adapted from WHO regions. [§]Categories are not mutually exclusive. ^Body mass index greater than or equal to 40 kg/m^2.
EM: Eastern Mediterranean; EU: Europe; LA: Latin America; NA: North America; SEA: South-East Asia; WP: Western Pacific; RA: rheumatoid arthritis; PsA: psoriatic arthritis; SLE: systemic lupus erythematosus; SpA: spondyloarthritis; IA: inflammatory arthritis; CVD: cardiovascular disease; DMARDs: disease modifying antirheumatic drugs; csDMARDs: conventional synthetic DMARDs; b/ts DMARDs: biologic or targeted synthetic DMARDs; GC: glucocorticoid.

Table 3.4. Baseline regional characteristics grouped into six global regions.

	EM	EU	SEA	WP	NA	LA	Total
	N = 2	N = 11	N = 1	N = 2	N = 2	N = 5	N = 23
	Mean (SD)						
Average annual temperature (degree Celsius)	24 (8)	10 (7)	25 (4)	19 (10)	13 (10)	21 (5)	19 (6)
Average annual precipitation (mm)	15 (15)	71 (27)	88 (93)	172 (77)	85 (35)	118 (82)	91 (52)
Population density (persons/km^2)	241 (20)	125 (78)	450 (-)	350 (3)	193 (675)	35 (20)	179 (552)
Median age (year)	28 (6)	44 (3)	28 (-)	37 (16)	39 (2)	31 (2)	38 (5)
Human development index[§]	0.70 (0.21)	0.90 (0.04)	0.65 (-)	0.82 (0.14)	0.92 (0.02)	0.79 (0.03)	0.89 (0.08)
Hospital beds per 1,000 population	0.9 (0.4)	4.3 (1.9)	0.5 (-)	7.0 (8.5)	2.5 (0.7)	2.4 (1.5)	2.9 (1.9)
Proportion aged 65 years or older (%)	3 (2)	19 (3)	6 (-)	16 (16)	16 (2)	8 (2)	16 (5)
Annual CVS death rate per 100,000 population	300 (174)	179 (87)	282 (-)	225 (206)	162 (30)	146 (42)	173 (65)
Diabetes prevalence among adults (%)	12 (6)	6 (2)	10 (-)	6 (1)	11 (2)	8 (3)	10 (3)
Annual death rate from air pollution per 100,000 population	93 (43)	22 (10)	132 (-)	59 (68)	19 (1)	32 (8)	26 (23)

Regional characteristics include country-level and U.S. state-level characteristics; ^Global regions were adapted from WHO regions. [§]Human development index is a composite index defined as the geometric mean of normalized indices in three dimensions including life expectancy at birth, mean of years of schooling for adults aged 25 years and more and expected years of schooling for children of school entering age, and gross national income per capita; HDI is ranked on a scale from 0 to 1.0, with 1.0 being the highest HDI.
EM: Eastern Mediterranean; EU: Europe; LA: Latin America; NA: North America; SEA: South-East Asia; WP: Western Pacific; CVS: cardiovascular.

Table 3.5. Inclusion of temporal and regional covariates as fixed effects and corresponding shrinkage in intraclass correlation coefficients.

Mixed-effected model description	**ICC (95% CI)**	**P^**
Base model with individual-level demographics as fixed effects	14.2% (7.5-25.2%)	<0.0001
Add individual-level rheumatic disease characteristics and comorbidities	10.1% (5.1-19.1%)	<0.0001
Add follow-up time (as main term and quadratic term)	8.6% (4.2-16.7%)	<0.0001
Add regional geographical and climatic covariates	6.9% (3.0-15.0%)	<0.0001
Add regional social and economic covariates	4.2% (1.7-9.8%)	<0.0001
Add regional demographics and burden of comorbidities	3.8% (1.4-10.2%)	<0.0001
Add regional cumulative COVID-19 deaths, containment efforts and population mobility trends	1.2% (0.1-9.5%)	0.14

Regional characteristics include country-level and U.S. state-level characteristics; ^P for a likelihood ratio test comparing mixed-effects model to a logistic regression model without country random effects. Covariates were added successively to nested mixed-effects models. All models include country random effects. ICC: intraclass correlation coefficient.

REFERENCES

1. 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):1576-1582.

2. Crowson CS, Matteson EL, Myasoedova E, et al. The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. *Arthritis and rheumatism.* 2011;63(3):633-639.

3. Mikuls TR, Johnson SR, Gravallese EM, Saag KG. COVID-19: Care of adult patients with systemic rheumatic disease. In: Clair EWS, Goldenberg DL, Romain PL, Curtis MR, eds. *UpToDate*: UpToDate; 2021.

4. Mikuls TR. Co-morbidity in rheumatoid arthritis. *Best practice & research. Clinical rheumatology.* 2003;17(5):729-752.

5. Bichile T, Petri M. Prevention and management of co-morbidities in SLE. *Presse medicale (Paris, France : 1983).* 2014;43(6 Pt 2):e187-195.

6. Perez-Chada LM, Merola JF. Comorbidities associated with psoriatic arthritis: Review and update. *Clinical immunology (Orlando, Fla.).* 2020;214:108397.

7. Burner TW, Rosenthal AK. Diabetes and rheumatic diseases. *Current opinion in rheumatology.* 2009;21(1):50-54.

8. Queiro R, Lorenzo A, Tejón P, Coto P, Pardo E. Obesity in psoriatic arthritis: Comparative prevalence and associated factors. *Medicine.* 2019;98(28):e16400.

9. Atzeni F, Masala IF, di Franco M, Sarzi-Puttini P. Infections in rheumatoid arthritis. *Current opinion in rheumatology.* 2017;29(4):323-330.

10. Sepriano A, Kerschbaumer A, Smolen JS, et al. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2019 update of the EULAR recommendations for the management of rheumatoid arthritis. *Annals of the rheumatic diseases.* 2020;79(6):760-770.

11. Listing J, Gerhold K, Zink A. The risk of infections associated with rheumatoid arthritis, with its comorbidity and treatment. *Rheumatology (Oxford, England).* 2013;52(1):53-61.

12. Gianfrancesco M, Hyrich KL, Al-Adely S, et al. Characteristics associated with hospitalisation for COVID-19 in people with rheumatic disease: data from the COVID-19 Global Rheumatology Alliance physician-reported registry. *Annals of the rheumatic diseases.* 2020;79(7):859-866.

13. Robinson PC, Yazdany J. The COVID-19 Global Rheumatology Alliance: collecting data in a pandemic. *Nature Reviews Rheumatology.* 2020;16(6):293-294.

14. Stallmach A, Kortgen A, Gonnert F, Coldewey SM, Reuken P, Bauer M. Infliximab against severe COVID-19-induced cytokine storm syndrome with organ failure-a cautionary case series. *Crit Care.* 2020;24(1):444.

15. Kuek A, Hazleman BL, Ostör AJ. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. *Postgraduate medical journal.* 2007;83(978):251-260.

16. Robinson PC, Liew DFL, Liew JW, et al. The Potential for Repurposing Anti-TNF as a Therapy for the Treatment of COVID-19. *Med (N Y).* 2020;1(1):90-102.

17. Brenner EJ, Ungaro RC, Gearry RB, et al. Corticosteroids, But Not TNF Antagonists, Are Associated With Adverse COVID-19 Outcomes in Patients With Inflammatory

Bowel Diseases: Results From an International Registry. *Gastroenterology.* 2020;159(2):481-491.e483.

18. Gianfrancesco M, Hyrich KL, Al-Adely S, et al. Characteristics associated with hospitalisation for COVID-19 in people with rheumatic disease: data from the COVID-19 Global Rheumatology Alliance physician-reported registry. *Annals of the rheumatic diseases.* 2020;79(7):859-866.

19. Mahil SK, Dand N, Mason KJ, et al. Factors associated with adverse COVID-19 outcomes in patients with psoriasis-insights from a global registry-based study. *The Journal of allergy and clinical immunology.* 2021;147(1):60-71.

20. Wallace ZS, Bhana S, Hausmann JS, et al. The Rheumatology Community responds to the COVID-19 pandemic: the establishment of the COVID-19 global rheumatology alliance. *Rheumatology.* 2020;59(6):1204-1206.

21. Gianfrancesco MA, Hyrich KL, Gossec L, et al. Rheumatic disease and COVID-19: initial data from the COVID-19 Global Rheumatology Alliance provider registries. *The Lancet. Rheumatology.* 2020;2(5):e250-e253.

22. Vittinghoff E. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models.* 2nd ed. New York: Springer; 2012.

23. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. *American journal of epidemiology.* 2017;186(8):1010-1014.

24. Franklin JM, Pawar A, Martin D, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clinical pharmacology and therapeutics.* 2020;107(4):817-826.

25. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of internal medicine.* 2017;167(4):268-274.

26. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine.* 2009;28(25):3083-3107.

27. Del Valle DM, Kim-Schulze S, Huang HH, et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat Med.* 2020;26(10):1636-1643.

28. Dolinger MT, Person H, Smith R, et al. Pediatric Crohn's Disease and Multisystem Inflammatory Syndrome in Children (MIS-C) and COVID-19 Treated with Infliximab. *J Pediatr Gastroenterol Nutr.* 2020.

29. Rodríguez-Lago I, Ramírez de la Piscina P, Elorza A, Merino O, Ortiz de Zárate J, Cabriada JL. Characteristics and Prognosis of Patients With Inflammatory Bowel Disease During the SARS-CoV-2 Pandemic in the Basque Country (Spain). *Gastroenterology.* 2020;159(2):781-783.

30. OCTRU Trials. AVID-CC: Adalimumab for Coronavirus in Community Care. [Internet]. 2020 [Accessed January 7, 2020]. Available from: https://avid-cc.octru.ox.ac.uk/.

31. Clinicaltrials.gov. NCT04425538. A Phase 2 Trial of Infliximab in Coronavirus Disease 2019 (COVID-19). [Internet]. 2020 [Accessed January 7, 2020]. Available from: https://clinicaltrials.gov/ct2/show/NCT04425538.

32. Goodsall TM, Costello SP, Bryant RV. COVID-19 and implications for thiopurine use. *Med J Aust.* 2020;212(10):490-490.e491.

33. Kirchgesner J, Lemaitre M, Carrat F, Zureik M, Carbonnel F, Dray-Spira R. Risk of Serious and Opportunistic Infections Associated With Treatment of Inflammatory Bowel Diseases. *Gastroenterology.* 2018;155(2):337-346.e310.

34. Toruner M, Loftus EV, Jr., Harmsen WS, et al. Risk factors for opportunistic infections in patients with inflammatory bowel disease. *Gastroenterology.* 2008;134(4):929-936.

35. Wisniewski A, Kirchgesner J, Seksik P, et al. Increased incidence of systemic serious viral infections in patients with inflammatory bowel disease associates with active disease and use of thiopurines. *United European Gastroenterol J.* 2020;8(3):303-313.

36. Ungaro RC, Brenner EJ, Gearry RB, et al. Effect of IBD medications on COVID-19 outcomes: results from an international registry. *Gut.* 2021;70(4):725-732.

37. Frohman EM, Villemarette-Pittman NR, Cruz RA, et al. Part II. high-dose methotrexate with leucovorin rescue for severe COVID-19: An immune stabilization strategy for SARS-CoV-2 induced 'PANIC' attack. *J Neurol Sci.* 2020;415:116935.

38. Safavi F, Nath A. Silencing of immune activation with methotrexate in patients with COVID-19. *J Neurol Sci.* 2020;415:116942.

39. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE. ACG Clinical Guideline: Management of Crohn's Disease in Adults. *The American journal of gastroenterology.* 2018;113(4):481-517.

40. Maini RN, Breedveld FC, Kalden JR, et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor alpha monoclonal antibody combined with low-

dose weekly methotrexate in rheumatoid arthritis. *Arthritis and rheumatism.* 1998;41(9):1552-1563.

41. Kalil AC, Patterson TF, Mehta AK, et al. Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *The New England journal of medicine.* 2021;384(9):795-807.

42. Kilian A, Chock YP, Huang IJ, et al. Acute respiratory viral adverse events during use of antirheumatic disease therapies: A scoping review. *Seminars in arthritis and rheumatism.* 2020;50(5):1191-1201.

43. World Health O. *Clinical management of COVID-19: interim guidance, 27 May 2020.* Geneva: World Health Organization; 2020 2020.

44. Tzotzos SJ, Fischer B, Fischer H, Zeitlinger M. Incidence of ARDS and outcomes in hospitalized patients with COVID-19: a global literature survey. *Critical care (London, England).* 2020;24(1):516.

45. Wu C, Chen X, Cai Y, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med.* 2020;180(7):934-943.

46. Mart MF, Ware LB. The long-lasting effects of the acute respiratory distress syndrome. *Expert review of respiratory medicine.* 2020;14(6):577-586.

47. Bower H, Frisell T, Di Giuseppe D, et al. Impact of the COVID-19 pandemic on morbidity and mortality in patients with inflammatory joint diseases and in the general population: a nationwide Swedish cohort study. *Annals of the rheumatic diseases.* 2021.

48. Cordtz R, Lindhardsen J, Soussi BG, et al. Incidence and severeness of COVID-19 hospitalisation in patients with inflammatory rheumatic disease: a nationwide cohort study from Denmark. *Rheumatology (Oxford, England).* 2020.

49. Peach E, Rutter M, Lanyon P, et al. Risk of death among people with rare autoimmune diseases compared with the general population in England during the 2020 COVID-19 pandemic. *Rheumatology (Oxford, England).* 2021;60(4):1902-1909.

50. Severity of COVID-19 and survival in patients with rheumatic and inflammatory diseases: data from the French RMD COVID-19 cohort of 694 patients. *Annals of the rheumatic diseases.* 2020;80(4):527-538.

51. D'Silva KM, Jorge A, Cohen A, et al. COVID-19 Outcomes in Patients With Systemic Autoimmune Rheumatic Diseases Compared to the General Population: A US Multicenter, Comparative Cohort Study. *Arthritis & rheumatology (Hoboken, N.J.).* 2021;73(6):914-920.

52. Hsu TY, D'Silva KM, Patel NJ, et al. Laboratory trends, hyperinflammation, and clinical outcomes for patients with a systemic rheumatic disease admitted to hospital for COVID-19: a retrospective, comparative cohort study. *The Lancet. Rheumatology.* 2021;3(9):e638-e647.

53. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology.* 1996;49(12):1373-1379.

54. Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for clinical predictive modeling. *International journal of medical informatics.* 2018;116:10-17.

55. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevska O, written on behalf of AMEB-DCTCG. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med.* 2019;7(7):152-152.

56. EULAR COVID-19 Registry. 2020; https://www.eular.org/eular_covid_19_registry.cfm. Accessed September 14th, 2021.

57. COVID-19 Global Rheumatology Alliance 2020 https://rheum-covid.org/. Accessed September 14th, 2021.

58. Avouac J, Drumez E, Hachulla E, et al. COVID-19 outcomes in patients with inflammatory rheumatic and musculoskeletal diseases treated with rituximab: a cohort study. *The Lancet. Rheumatology.* 2021;3(6):e419-e426.

59. Trefond L, Drumez E, Andre M, et al. Impact of hydroxychloroquine used as DMARD on SARS-CoV-2 tests and infection evolution in a population of 871 patients with inflammatory rheumatic and musculoskeletal diseases. *Joint bone spine.* 2021;88(6):105226.

60. Hasseli R, Mueller-Ladner U, Schmeiser T, et al. National registry for patients with inflammatory rheumatic diseases (IRD) infected with SARS-CoV-2 in Germany (ReCoVery): a valuable mean to gain rapid and reliable knowledge of the clinical course of SARS-CoV-2 infections in patients with IRD. *RMD open.* 2020;6(2):e001332.

61. Hasseli R, Mueller-Ladner U, Hoyer BF, et al. Older age, comorbidity, glucocorticoid use and disease activity are risk factors for COVID-19 hospitalisation in patients with inflammatory rheumatic and musculoskeletal diseases. *RMD open.* 2021;7(1):e001464.

62. Hasseli R, Pfeil A, Hoyer BF, et al. Do patients with rheumatoid arthritis show a different course of COVID-19 compared to patients with spondyloarthritis? *Clinical and experimental rheumatology.* 2021;39(3):639-647.

63. Scirè CA, Carrara G, Zanetti A, et al. COVID-19 in rheumatic diseases in Italy: first results from the Italian registry of the Italian Society for Rheumatology (CONTROL-19). *Clinical and experimental rheumatology.* 2020;38(4):748-753.

64. Canhão H, Faustino A, Martins F, Fonseca JE. Reuma.pt - the rheumatic diseases portuguese register. *Acta reumatologica portuguesa.* 2011;36(1):45-56.

65. Santos M, Canhão H, Mourão A, et al. Reuma.pt contribution to the knowledge of immune-mediated systemic rheumatic diseases. *Acta reumatologica portuguesa.* 2017;2017.

66. The Swedish Rheumatology Quality Register. https://srq.nu/en/welcome-health-professional/. Accessed August 20, 2021.

67. Marques CDL, Kakehasi AM, Pinheiro MM, et al. High levels of immunosuppression are related to unfavourable outcomes in hospitalised patients with rheumatic diseases and COVID-19: first results of ReumaCoV Brasil registry. *RMD open.* 2021;7(1).

68. Marques C, Kakehasi AM, Gomides APM, et al. A Brazilian Cohort of Patients With Immuno-Mediated Chronic Inflammatory Diseases Infected by SARS-CoV-2

(ReumaCoV-Brasil Registry): Protocol for a Prospective, Observational Study. *JMIR research protocols.* 2020;9(12):e24357.

69. Isnardi CA, Gómez G, Quintana R, et al. Características epidemiológicas y desenlaces de la infección por SARS-CoV-2 en pacientes con patologías reumáticas: Primer reporte del registro argentino SAR-COVID. *Rev Arg Reumatol.* 2021;32(1):7-15.

70. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine.* 2015;162(1):W1-73.

71. Schaefer J, Lehne M, Schepers J, Prasser F, Thun S. The use of machine learning in rare diseases: a scoping review. *Orphanet Journal of Rare Diseases.* 2020;15(1):145.

72. Kuhn M. Building Predictive Models in R Using the caret Package. *2008.* 2008;28(5):26.

73. Unal I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Computational and Mathematical Methods in Medicine.* 2017;2017:3762651.

74. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis.* 2019;11(Suppl 4):S574-S584.

75. Richter AN, Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. *Artificial Intelligence in Medicine.* 2018;90:1-14.

76. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health technology assessment (Winchester, England).* 2004;8(36):iii-iv, ix-xi, 1-158.

77. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Statistics in medicine.* 2019;38(21):4051-4065.

78. Strangfeld A, Schäfer M, Gianfrancesco MA, et al. Factors associated with COVID-19-related death in people with rheumatic diseases: results from the COVID-19 Global Rheumatology Alliance physician-reported registry. *Annals of the rheumatic diseases.* 2021.

79. Bachiller-Corral J, Boteanu A, Garcia-Villanueva MJ, et al. Risk of Severe COVID-19 Infection in Patients With Inflammatory Rheumatic Diseases. *The Journal of rheumatology.* 2021.

80. Pablos JL, Abasolo L, Alvaro-Gracia JM, et al. Prevalence of hospital PCR-confirmed COVID-19 cases in patients with chronic inflammatory and autoimmune rheumatic diseases. *Annals of the rheumatic diseases.* 2020;79(9):1170-1173.

81. Eythorsson E, Bjarnadottir V, Runolfsdottir HL, et al. Development and validation of a prognostic model for COVID-19: a population-based cohort study in Iceland. *medRxiv.* 2021:2021.2007.2019.21260759.

82. FDA. FDA authorizes REGEN-COV monoclonal antibody therapy for post-exposure prophylaxis (prevention) for COVID-19. . 2021; https://www.fda.gov/drugs/drug-safety-and-availability/fda-authorizes-regen-cov-monoclonal-antibody-therapy-post-exposure-prophylaxis-prevention-covid-19. Accessed August 20, 2021.

83. FDA. EUA Regeneron LOA Treatment and Prophylaxsis of COVID-19. 2021; https://www.fda.gov/media/145610/download. Accessed August 9, 2021.

84. Sparks JA, Wallace ZS, Seet AM, et al. Associations of baseline use of biologic or targeted synthetic DMARDs with COVID-19 severity in rheumatoid arthritis: Results from the COVID-19 Global Rheumatology Alliance physician registry. *Annals of the rheumatic diseases.* 2021;80(9):1137-1146.

85. Zeiberg D, Prahlad T, Nallamothu BK, Iwashyna TJ, Wiens J, Sjoding MW. Machine learning for patient risk stratification for acute respiratory distress syndrome. *PLoS One.* 2019;14(3):e0214465.

86. Ding XF, Li JB, Liang HY, et al. Predictive model for acute respiratory distress syndrome events in ICU patients in China using machine learning algorithms: a secondary analysis of a cohort study. *Journal of translational medicine.* 2019;17(1):326.

87. Xu W, Sun N-N, Gao H-N, et al. Risk factors analysis of COVID-19 patients with ARDS and prediction based on machine learning. *Scientific Reports.* 2021;11(1):2933.

88. Middleton KL, Jones JL. Socially desirable response sets: The impact of country culture. *Psychology & Marketing.* 2000;17(2):149-163.

89. Cook C, Patel NJ, D'Silva KM, et al. Clinical characteristics and outcomes of COVID-19 breakthrough infections among vaccinated patients with systemic autoimmune rheumatic diseases. *medRxiv.* 2021:2021.2008.2004.21261618.

90. Mehta B, Pedro S, Ozen G, et al. Serious infection risk in rheumatoid arthritis compared with non-inflammatory rheumatic and musculoskeletal diseases: a US national cohort study. *RMD open.* 2019;5(1):e000935.

91. Pimentel-Quiroz VR, Ugarte-Gil MF, Harvey GB, et al. Factors predictive of serious infections over time in systemic lupus erythematosus patients: data from a multi-ethnic, multi-national, Latin American lupus cohort. *Lupus.* 2019;28(9):1101-1110.

92. Au K, Reed G, Curtis JR, et al. High disease activity is associated with an increased risk of infection in patients with rheumatoid arthritis. *Annals of the rheumatic diseases.* 2011;70(5):785-791.

93. England BR, Roul P, Yang Y, et al. Risk of COVID-19 in Rheumatoid Arthritis: A National Veterans Affairs Matched Cohort Study in At-Risk Individuals. *Arthritis & rheumatology (Hoboken, N.J.).* 2021;73(12):2179-2188.

94. Akiyama S, Hamdeh S, Micic D, Sakuraba A. Prevalence and clinical outcomes of COVID-19 in patients with autoimmune diseases: a systematic review and meta-analysis. *Annals of the rheumatic diseases.* 2020.

95. Freites Nuñez DD, Leon L, Mucientes A, et al. Risk factors for hospital admissions related to COVID-19 in patients with autoimmune inflammatory rheumatic diseases. *Annals of the rheumatic diseases.* 2020;79(11):1393-1399.

96. Ye C, Cai S, Shen G, et al. Clinical features of rheumatic patients infected with COVID-19 in Wuhan, China. *Annals of the rheumatic diseases.* 2020;79(8):1007-1013.

97. Damiani G, Pacifico A, Bragazzi NL, Malagoli P. Biologics increase the risk of SARS-CoV-2 infection and hospitalization, but not ICU admission and death: Real-life data from a large cohort during red-zone declaration. *Dermatologic therapy.* 2020;33(5):e13475.

98. Sorci G, Faivre B, Morand S. Explaining among-country variation in COVID-19 case fatality rate. *Scientific Reports.* 2020;10(1):18909.

99. Pana TA, Bhattacharya S, Gamble DT, et al. Country-level determinants of the severity of the first global wave of the COVID-19 pandemic: an ecological study. *BMJ open.* 2021;11(2):e042034.

100. Karlinsky A, Kobak D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *Elife.* 2021;10:e69336.

101. Firebaugh G. Ecological Fallacy, Statistics of. In: Smelser NJ, Baltes PB, eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon; 2001:4023-4026.

102. de Munck VC. Units of Analysis. In: Kempf-Leonard K, ed. *Encyclopedia of Social Measurement*. New York: Elsevier; 2005:899-907.

103. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of clinical epidemiology.* 2008;61(4):344-349.

104. Coronavirus Resource Center. John Hopkins University & Medicine. . 2021; https://coronavirus.jhu.edu/. Accessed October 4, 2021.

105. Austin PC, Leckie G. The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models. *Journal of Statistical Computation and Simulation.* 2018;88(16):3151-3163.

106. Hayward MD, Majmundar MK. *Future directions for the demography of aging : proceedings of a workshop.* Washington, DC: The National Academies Press; 2018.

107. Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models.* Cambridge ;: Cambridge University Press; 2007.

108. Midway S. Data Analysis in R. Chapter 9 Random Effects. . 2021; https://bookdown.org/steve_midway/DAR/random-effects.html. Accessed October 4, 2021.

109. Dey D, Sciascia S, Pons-Estel GJ, Ding H, Shen N. Health Disparities in Rheumatic Diseases: Understanding Global Challenges in Africa, Europe, Latin America, and Asia and Proposing Strategies for Improvement. *Rheumatic diseases clinics of North America.* 2021;47(1):119-132.

110. Phuti A, Schneider M, Tikly M, Hodkinson B. Living with systemic lupus erythematosus in the developing world. *Rheumatology international.* 2018;38(9):1601-1613.

111. Al Maini M, Adelowo F, Al Saleh J, et al. The global challenges and opportunities in the practice of rheumatology: White paper by the World Forum on Rheumatic and Musculoskeletal Diseases. *Clinical rheumatology.* 2015;34(5):819-829.

112. Gibson T. Rheumatology in India and Pakistan today. *Rheumatology.* 2015;54(5):753-754.

113. Samson Aregbeshola B, Mohsin Khan S. Out-of-Pocket Payments, Catastrophic Health Expenditure and Poverty Among Households in Nigeria 2010. *International Journal of Health Policy and Management.* 2018;7(9):798-806.

114. Dey D. Juggling Art: Making Critical Clinical Decisions without Vital Laboratory Support in Autoimmune Rheumatic Patients in a Resource Poor Setting. *Ghana Med J.* 2017;51(2):47-49.

115. Pistor M, Hoepner AGF, Lin Y, et al. Immunotherapies and COVID-19 mortality: a multidisciplinary open data analysis based on FDA's Adverse Event Reporting System. *Annals of the rheumatic diseases.* 2021:annrheumdis-2021-220679.

116. Kenyon C. Flattening-the-curve associated with reduced COVID-19 case fatality rates-an ecological analysis of 65 countries. *J Infect.* 2020;81(1):e98-e99.

117. Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EClinicalMedicine.* 2020;25:100464.

118. Banik A, Nag T, Chowdhury SR, Chatterjee R. Why Do COVID-19 Fatality Rates Differ Across Countries? An Explorative Cross-country Study Based on Select Indicators. *Global Business Review.* 2020;21(3):607-625.

119. Sabawoon W. Differences by country-level income in COVID-19 cases, deaths, case-fatality rates, and rates per million population in the first five months of the pandemic. *medRxiv.* 2020:2020.2007.2013.20153064.

120. Huang Q, Jackson S, Derakhshan S, et al. Urban-rural differences in COVID-19 exposures and outcomes in the South: A preliminary analysis of South Carolina. *PLoS One.* 2021;16(2):e0246548.

121. Roy S, Ghosh P. Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. *PLoS One.* 2020;15(10):e0241165-e0241165.

122. Pozzer A, Dominici F, Haines A, Witt C, Münzel T, Lelieveld J. Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovascular Research.* 2020;116(14):2247-2253.

123. Izadi Z, Brenner EJ, Mahil SK, et al. Association Between Tumor Necrosis Factor Inhibitors and the Risk of Hospitalization or Death Among Patients With Immune-Mediated Inflammatory Disease and COVID-19. *JAMA network open.* 2021;4(10):e2129639.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Zahra Izadi

56FA40829645408...

Author Signature

3/7/2022

Date