6-2-2020

An Assessment of Potential Causal Pathways Between Two Autoimmune Diseases: Mendelian Randomization Studies of Rheumatoid Arthritis and Crohn's Disease; and A Phenome-Wide Association Study to Assess Potential New Targets for Tocilizumab, an Interleukin-6 Inhibitor

Staci Abramsky Risman
CUNY School of Public Health, sarisman6@gmail.com

AN ASSESSMENT OF POTENTIAL CAUSAL PATHWAYS BETWEEN TWO AUTOIMMUNE DISEASES: MENDELIAN RANDOMIZATION STUDIES OF RHEUMATOID ARTHRITIS AND CROHN'S DISEASE; AND A PHENOME-WIDE ASSOCIATION STUDY TO ASSESS POTENTIAL NEW TARGETS FOR TOCILIZUMAB, AN INTERLEUKIN-6 INHIBITOR

A DISSERTATION

by

STACI ABRAMSKY RISMAN

Concentration: EPIDEMIOLOGY

Presented to the Faculty at the Graduate School of Public Health and Health Policy in partial fulfillment of the requirements for the degree of Doctor of Public Health

Graduate School of Public Health and Health Policy

City University of New York

New York, New York

MAY 2020

Dissertation Committee:

C. MARY SCHOOLING, PHD

KATARZYNA WYKA, PHD

SHENG LI, PHD

JAMIE GEIER , PHD

ABSTRACT

AN ASSESSMENT OF POTENTIAL CAUSAL PATHWAYS BETWEEN TWO AUTOIMMUNE DISEASES: MENDELIAN RANDOMIZATION STUDIES OF RA AND CD; AND A PHEWAS TO ASSESS POTENTIAL NEW TARGETS FOR TOCILIZUMAB, AN INTERLEUKIN-6 INHIBITOR

by

Staci Abramsky Risman

Advisor: C. Mary Schooling, PhD

Background

Rheumatoid arthritis (RA) and Crohn's disease (CD) are two among a group of immune-mediated inflammatory diseases (IMIDs). These diseases are common, collectively affecting as many as two million Americans and as many as 28 million people worldwide. They have different clinical presentations but may have a common pathogenesis. An over-expression of tumor necrosis factor (TNF) and interleukin-6 (IL-6) have been proffered as a causal factor for IMIDs. These and other IMIDs may co-occur in patients, and it has been shown that patients with one IMID may have an increased risk of another IMID. What is not clear is whether the co-occurrence of a second IMID is caused by the existence of the first IMID.

Mendelian Randomization (MR) has emerged as an important tool that facilitates causal inference from observational studies. This study design can allow for avoidance of a common problem in epidemiologic studies, confounding, because it takes advantage of the random allocation of genetic make-up at conception. MR compares disease status by genetically predicted exposure to obtain unconfounded estimates. MR implementation has been facilitated by the increasing availability of Genome-wide association studies (GWAS) genotyping millions of single nucleotide polymorphisms (SNPs).

A recent extension of Mendelian Randomization is to conduct Phenome-wide Association Studies (PheWAS) to use genetic variants that predict response to treatment, for drug target validation and repurposing studies using these genetic variants as genetic proxies to identify both on-target and off-target mechanistic effects of a pharmacologic intervention. As an example, the minor allele of rs7529229, which is on the IL6R gene and is associated with interleukin 6 (IL6) concentrations, predicts response to tocilizumab, a pharmacologic treatment approved for the treatment of RA. A further goal of the current study is to use PheWAS to explore the scope for re-purposing tocilizumab as to characterize the off-target effects of this treatment.

The hypothesis that one IMID causes another was addressed by two MR studies, in which one of the IMIDs was tested as a cause of the other. Additionally, a PheWAS evaluated whether genetic proxies of therapeutic treatments for RA can identify any new effects.

Methods

A two-sample Mendelian randomization study was conducted to assess the possible causal effect of RA on CD for the first specific aim. The primary analysis used the inverse variance weighted (IVW) method; sensitivity analyses include weighted median, MR-Egger, and MR-PRESSO

(Pleiotropy RESidual Sum and Outlier) methods. Genetic predictors of RA obtained from a GWAS including 22 studies with participants from European and Asian backgrounds, with 19,234 cases and 61,565 controls, were applied to summary genetic associations for CD, from the International IBD Genetics Consortium (IIBDGC), (17,897 cases, 33,977 controls) of European descent.

A two-sample Mendelian randomization study was conducted to assess a possible causal role of CD on RA for the second specific aim. The primary analysis used IVW to determine the effect of CD on RA. Sensitivity analyses included weighted median, MR-Egger, and MR-PRESSO. Genetic predictors of CD from summary statistics from the IIBDGC, (17,897 cases, 33,977 controls of European descent) were applied to a GWAS of RA that included 22 studies including patients of European and Asian backgrounds (19,234 cases, 61,565 controls).

One single nucleotide polymorphism (SNP) predicting response to TCZ in RA patients was identified, rs7529229. For the third specific aim, the PheWAS (Phenome-wide Association Study) function in the MR-base web-based application was used to identify traits associated with this SNP to use as potential drug targets for inclusion in the analysis. As this is an agnostic search, a p-value threshold of 2.4×10^{-6} was used based on a Bonferroni correction. The PheWAS provided effect estimates (provided as beta) and p-values which enabled evaluation of potential new targets for TCZ.

Results

For specific aim 1, all four analyses showed a protective effect of RA on CD. (IVW odds ratio [OR] 0.89, 95% confidence interval [CI] 0.77-1.00, p=0.042; MR-Egger OR 0.71, 95% CI 0.50-

0.93, p=0.004; weighted median OR 0.78, 95% CI 0.72-0.84, p=$8.4x10^{-15}$; MR-PRESSO OR 0.92, 95% CI 0.84-1.00, p= 0.047).

For specific aim 2, the primary analysis and all three sensitivity analyses showed no causal effect of CD on RA (IVW OR 0.97, 95% CI 0.88-1.07, p=0.59; weighted median OR 0.98, 95% CI 0.94-1.03, p=0.49; MR-Egger OR 0.92, 95% CI 0.67-1.18, p=0.54, MR-PRESSO OR 1.00, 95% CI 0.64-1.36, p=0.90).

For specific aim 3, the SNP rs7529229 predicting response to TCZ in RA was associated with 21,031 traits. Seventeen of these met the Bonferroni corrected p-value threshold for statistical significance. Of these, three were excluded from consideration: abdominal aortic aneurysm, due to missing beta, and two traits described as "Blood clot DVT bronchitis emphysema asthma rhinitis eczema allergy diagnosed by doctor: None of the above," as these did not indicate a trait. Among the other 10 unique traits, four showed an inverse association with the SNP predicting response to TCZ: RA, coronary heart disease (CHD), and two blood counts (red cell distribution width and granulocyte percentage of myeloid white cells). Six traits were positively associated with this SNP, meaning use of TCZ may increase the risk of the following: eczema, asthma, a less specific trait of hay fever, allergic rhinitis or eczema, tonsillectomy, and two additional blood counts (mean corpuscular hemoglobin and monocyte percentage of white cells).

Discussion

The first MR study suggests that RA protects from the development of CD. It is possible this is due to issues with the data sources, or the infrequency of co-occurrence of the two diseases. There also may be a currently unknown biological mechanism causing this effect.

The second MR study suggests that CD does not cause rheumatoid arthritis. Co-occurrence of CD and RA is fairly uncommon, and when it does occur, may be due to shared causes of both conditions such as overexpression of TNF-a or IL-6.

The PheWAS did not yield any new targets for tociluzumab. As rs7529229 predicts response to TCZ in RA patients, it was expected that there would be an inverse association between the SNP and RA. Other studies have already evaluated the inverse association of this SNP with CHD, suggesting CHD as a potential target for TCZ. The positive association with asthma was somewhat unexpected, as some studies have suggested a potential for IL-6 inhibition in asthma. The positive association with eczema were expected as this was reported in the clinical trial program for TCZ and in other PheWAS of SNPs associated with IL-6 levels. The impact of TCZ on various blood counts is difficult to interpret. The direction of the association with asthma is surprising, even though a recent PheWAS reported a similar result.

ACKNOWLEDGEMENTS

I want to express my gratitude for the help and support of my advisor and committee chair, Dr. Mary Schooling. I appreciate everything you've done to support me throughout this process. I'd also like to thank Dr. Katarzyna Wyka and Dr. Sheng Li, for jumping in for my second exam and staying for the dissertation. I am incredibly grateful for the long-term friendship and support of, Dr. Jamie Geier. I miss working together, our lunches, and random work day chats. I am so grateful that you agreed to be on my committee. Thank you so much! I also want to thank Dr. Heidi Jones for her support and guidance these past six years. Thank you to my friends and classmates for their help and advice, and especially a huge thank you to Chloe Mirzayi. I may have conquered my fear of R (or not).

I am grateful to my former employer for funding my coursework. Also, thank you to Dr. Anita Verga for being a good listener, letting me take time off when I needed to write, and being a kind and wonderful human.

A huge and loving thank you to my husband, Mark, who has kept our family and our life rolling these past six years while I was otherwise occupied. I love you and our life together. Chai/pi! Evan, thank you for being you, for your IT help, and much more. I love you so much! I know that if I don't acknowledge the dogs, Mark and Evan will wonder why, so thank you to my two trusty sidekicks, Sassy and Pickles.

As part of such a STEM-focused family, I feel I must show appreciation for the technology that allowed me to defend my dissertation online. A pandemic prevented me from finishing in person what I started six long years ago, but now the general public knows what epidemiology is (tiniest silver lining ever).

DISCLOSURE STATEMENT

This dissertation research was unfunded. Coursework for my doctoral program was funded by Pfizer, Inc. I am a former employee and a current shareholder of Pfizer, Inc. I am a former employee of Regeneron and currently work as a consultant for AbbVie, Inc, formerly Allergan.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1: An Assessment of Potential Causal Pathways Between Two Autoimmune Diseases: Mendelian Randomization Studies of Rheumatoid Arthritis and Crohn's Disease; and A Phenome-Wide Association Study to Assess Potential New Targets for Tocilizumab, an Interleukin-6 Inhibitor

SPECIFIC AIMS

Rheumatoid arthritis (RA) and Crohn's disease (CD) are two among a group of immune-mediated inflammatory diseases (IMIDs). These diseases are common, collectively affecting as many as two million Americans [1-4] and as many as 28 million people worldwide. [5-7] They have different clinical presentations but may have a common pathogenesis. An over-expression of tumor necrosis factor (TNF) has been proffered as a causal factor for IMIDs. These and other IMIDs may co-occur in patients, and it has been shown that patients with one IMID may have an increased risk of another IMID. What is not clear is whether the co-occurrence of a second IMID is caused by the existence of the first IMID. [8-10]

Mendelian Randomization (MR) has emerged as an important tool that facilitates causal inference from observational studies. This study design can allow for avoidance of a common problem in epidemiologic studies, i.e., confounding, [11] because it takes advantage of the random allocation of genetic make-up at conception. MR compares disease status by genetically predicted exposure to obtain unconfounded estimates. MR implementation has been facilitated by the increasing availability of genome-wide association studies (GWAS) genotyping millions of single nucleotide polymorphisms (SNPs).

A recent extension of Mendelian Randomization is to conduct Phenome-wide Association Studies (PheWAS) to use genetic variants that predict response to treatment, [12] for drug target validation and repurposing studies using these genetic variants as genetic proxies to identify both on-target and off-target mechanistic effects of a pharmacologic intervention. As an example, the

minor allele of rs7529229, which is on the IL6R gene and is associated with interleukin 6 (IL6) concentrations, predicts response to tocilizumab, a pharmacologic treatment approved for the treatment of RA. [13] A further goal of the current study is to use PheWAS to explore the scope for re-purposing tocilizumab as to characterize the off-target effects of this treatment.

The hypothesis that one IMID causes another was addressed by two MR studies, in which one of the IMIDs was tested as a cause of the other. Additionally, a PheWAS evaluated whether genetic proxies of therapeutic treatments for RA can identify any new effects. The three specific aims of the proposed study are listed below.

Specific aim 1: Does RA cause Crohn's disease?

Specific aim 2: Does Crohn's disease cause RA?

Specific aim 3: Can drug target validation/repurposing predict additional effects of an existing therapeutic treatment for RA?

BACKGROUND AND SIGNIFICANCE

RA and CD are important among the autoimmune diseases due to high prevalence and resulting high number of disability-adjusted life years (DALYs), a measure of early mortality and ill health. [14]. Autoimmune diseases tend to be more common among women than men, for reasons that are not well understood. One estimate suggests that while about 8% of the population lives with an autoimmune disease, 78% of those are women. A hypothesis offered for a female preponderance suggests that since women have an increased antibody response to infection, vaccination and trauma compared to men, these antibodies, which protect against infection, also cause immune disease risk. [15]

Many treatments are approved for the treatment of RA and CD. MR studies focused on drug target validation and repurposing use genetic differences as proxies to identify both on-target and off-target mechanistic effects of a pharmacologic intervention. [13] Characterizing the effects of such genetic proxies of treatments may identify additional disease targets of existing treatments as well as off-target effects.

Incidence/prevalence/mortality

RA

Rheumatoid Arthritis (RA) is a common autoimmune disease characterized by pain and inflammation in the joints, leading to joint damage and significant loss of function. [16] Age of onset of RA is typically between the age of 30 and 60 years. Estimates of the median age at disease onset range from 45-58 years. [17, 18] About 1.5 million Americans live with RA; [19] worldwide, RA may affect as many as 21 million people. [5, 7]

RA is common in the U.S. and globally. Reports of incidence vary depending on method of ascertainment. For example, a population-based cohort of adults age 18 and older in Olmstead County, Minnesota, reported age- and sex-adjusted RA incidence of 40.9 per 100,000 population (53.1 per 100,000 population in women, 27.7 per 100,000 population in men). Diagnostic criteria included the 1987 American College of Rheumatology criteria, verified via participants' medical records. [19] A retrospective cohort study using a US medical claims database (PharMetrics) reported cumulative incidence of RA in adults 18 years and older in 2006 of 73 per 100,000 population (95% confidence interval (CI) 70–75 per 100,000 population). Cumulative incidence was 100 per 100,000 in women and 44 per 100,000 in men. Claims data may reflect an under-estimate compared to population-based estimates as not all RA patients may require health care contacts meeting the criteria for inclusion in a claims cohort. [20]

Prevalence is higher in developed nations compared to developing nations. The Global Burden of Disease 2010 study estimated global prevalence of RA from age 5 to 100 to be 0.24% (95% CI 0.23% to 0.25%), 0.34% in North American men, and 0.63% in North American women. [21] In the Olmstead County, Minnesota cohort, prevalence of RA was 0.72% which, when applied to the US population (2005), suggests that about 1.5 million adults live with RA. [19] In the National Health and Nutrition Examination Survey (NHANES), a population-based study, the age-adjusted prevalence of RA was 3.8% in 2014. [1] This disparity may reflect the difference between self-report (as in NHANES) and health care provider diagnosed disease (as in the Olmstead and GBD studies).

Based on 84 unique cohorts published from 1953-2008, overall, RA patients had an increased risk of death compared to people without RA, with standardized mortality ratios (SMRs) generally between 1.2 and 1.3 in inception cohorts. Causes of death are similar in RA patients

and non-RA patients. Cardiovascular disease is the most common cause of death among RA patients, at an earlier age than among non-RA patients. Infections, gastrointestinal diseases, and pulmonary disease are more common causes of death in RA patients. Interestingly, cancer deaths are overall decreased compared to the general population, with the exception of lymphoma, which may be increased among persons treated with biologics and disease-modifying antirheumatic drugs (DMARDs). [22]

CD

Crohn's disease (CD) is an autoimmune disease characterized by abdominal pain, diarrhea, and fatigue, weight loss, fever, growth failure, anemia, and/or recurrent fistulas. [23] There are two peak age ranges for onset of CD. The first is from 15-29 years and the second is from 55-59 years. [24] In the U.S., between 593,000 to 780,000 people live with CD, with an estimated 33,000 new patients diagnosed each year, [4] and worldwide, as many as 7 million people live with CD. [6]

The estimated incidence of CD in a US cohort was 10.7 per 100,000 person-years (95% confidence interval [CI] 9.1-12.3 per 100,000 person-years). [3, 4] Other estimates of incidence in North America range from 6.3 per 100,000 person-years in the US to 23.82 per 100,000 person-years in Canada. There is variability by country in Europe, with ranges from 0.0 per 100,000 person-years in Greenland to 15.4 per 100,000 person-years in Italy. [25]

Prevalence is higher in developed nations compared to developing nations. Prevalence estimates range from 0.0963% in the US to 0.318.5% in Canada. Prevalence estimates in Europe vary widely, from 0.015% in Romania to 0.322% in Germany. [25]

Mortality risk from intestinal cancer is generally similar in CD patients compared to the general population, although one study reported a slightly increased risk among CD patients with extensive small bowel disease and among those who were aged 20–29 years at diagnosis. [26] All-cause mortality is higher in CD than in the general population; standardized mortality ratio (SMR) of 1.45 (95% CI 1.34-1.58) has been reported, and mortality from digestive conditions, all cancers, digestive cancers, colorectal cancer, lymphatic and lung cancers were significantly higher compared to the general population. [27] The Olmstead County (Minnesota) cohort study also reported a higher risk of mortality among CD patients (SMR, 1.25; 95% CI, 0.98-1.57). Specifically, in this cohort, 9% of deaths were attributed to CD complications, including intestinal fistula, perforation, or abscess. [28]

Drug target validation/repurposing

Genetic validation of potential drug treatments is increasingly used to facilitate drug development and anticipate side-effects. Mutations in the *PCSK9* gene have been found to predict familial hypercholesterolemia and coronary artery disease; leading to two drugs that target this gene: evolocumab (Repatha, marketed by Amgen) and alirocumab (Praluent, marketed by Regeneron). [29] Lovostatin, the first of the statins and approved for primary and secondary prevention of coronary artery disease, inhibits HMG-CoA reductase. Many other genetic differences that identify susceptibility to disease have been discovered using GWAS, creating many potential drug targets. [30] Interleukin 6 is implicated as a cause of RA, and tocilizumab (Actemra, marketed by Genentech) is a therapy targeting IL-6 inhibition. [13] Tumor necrosis factor alpha (TNF-a) is implicated in CD and RA; adalimumab (Humira, marketed by AbbVie) targets TNF-a. [31] Schmidt and colleagues describe four genes (*HMGCR, PCSK9, NPC1L1,* and *CETP*) with effects on lipids that have variants with specific drug targets that are approved

or are in clinical trials for coronary heart disease (CHD). Notably, the drugs with genetic validation as modulating both lipids and CHD were successful in clinical trials, but those that failed genetic validation for CHD (drugs targeting *CETP*) were not. The authors also provide guidance on the "selection of genes encoding druggable proteins" for use in drug target validation MR studies. [32] Nelson and colleagues used the Informa Pharmaprojects database to identify 22,270 drugs with 1,824 genetic targets, creating 19,085 drug-target pairs covering 705 indications, including RA and CD. Many drug-target pairs have been identified for RA and CD (651 for RA; 134 for CD). [33]

Notably, genetic variants that have been identified as predictors of treatment response for RA (and other diseases) may have mechanistic effects on other outcomes. Research has been done on tocilizumab and adalimumab to look for additional effects; for example, the Interleukin-6 Receptor Mendelian Randomisation, Consortium has shown a protective effect on coronary heart disease of the minor allele of rs7529229 allele, a proxy for tocilizumab use [13] and neurological outcomes have been identified as effects of the rs10210302 variant, a proxy of adalimumab. [34] However, whether tocilizumab, an IL-6 inhibitor, has effects on other diseases in which IL-6 is involved is unclear but important.

Public Health Impact

People with RA are at risk for falls and other injuries and may be unable to work. [35] One estimate of health care costs in the U.S. attributable to RA in 2013 was $140 billion. [36] The National Health Interview Survey (NHIS) reported that 43.5% of individuals with RA had limitations to activity related to their RA. [37] In the Global Burden of Disease 2010 study, RA was ranked as the 42nd highest contributor to global disability, and years lived with disability (YLDs) were 55 per 100,000 population (in 2010), and higher in females than in males (87.84

per 100,000 versus 22.46 per 100,000). [21]; the Global Burden of Disease 2017 study reported YLDs at 2,620,000. [14]

The Global Burden of Disease study reported YLDs at 335,400 for CD. [14] In a review of Health-Related Quality of Life (HQOL) studies, compared to healthy controls, CD patients scored worse on physical, social, and emotional function. Compared to patients with ulcerative colitis, patients with CD scored similar to or worse on general HQOL measures. [38] Estimated direct costs of CD in in the U.S. in1998 (including inpatient and outpatient visits and pharmacy) were $708 million; indirect costs were estimated at $75 million and included costs associated lost productivity. [39]

The cost of developing new molecular entities (NMEs) is in the billions of dollars, and can take, on average, over 12 years. [29, 30] Less than 10% of drugs are eventually approved by the US Food and Drug Administration (FDA) or another health authority. [29] Drug targets with genetic associations with disease have a much higher likelihood of succeeding and receiving regulatory approval. [29, 33, 40] More specifically, if causal genetic variants can be identified through genome-wide association studies (GWAS), this genetic evidence has been shown to increase drug approval more than two-fold. [29] Using genetic evidence could decrease cost of drug development, due to fewer failed clinical trials. One estimate suggests that if the proportion of drugs in development with genetic targets were increased from 15% (the current proportion) to 50%, research and development costs would decrease by anywhere from 9% to 35%. [29]

Summary of MR research of RA and CD causes, and PheWAS research on drug targets

Rheumatoid arthritis and CD are prevalent and debilitating diseases of great public health importance. Causes of these diseases are not well understood. An over-expression of tumor necrosis factor (TNF) has been proffered as one causal factor for IMIDs [10] but there are likely

other causes as well, given their multifactorial nature. A number of MR studies have evaluated various potential causes of RA while fewer have looked at CD.

RA

Li and colleagues conducted a meta-analysis of MR studies to assess the role of interleukin-6 in RA; they determined that reduced IL-6 may be causally linked to increased risk of RA overall and especially in the Asian population. [41] The Interleukin 1 Genetics Consortium also evaluated the causal effect of genetic inhibition of IL-6 on RA and type 2 diabetes, coronary heart disease, ischemic stroke, and abdominal aortic aneurysm (n=453,411). Results suggested causal associations of IL-6 inhibition with RA (reduced risk) as well as cardiovascular disease (increased risk). [42] Vitamin D has also been considered as a risk factor for RA, with variable results. One MR study reported vitamin D-associated SNPs were not associated with RA. [43] Yarwood and colleagues evaluated 6 SNPs associated with vitamin D to determine response to anti-TNF-alpha inhibitor therapy in RA patients; one SNP showed a poorer response to TNF-alpha inhibitor therapy. [44] However, observational studies in patients can be difficult to interpret, because they are open to selection bias. An MR study reported that linoleic acid was inversely associated with RA (i.e., protective) (OR 0.97, 95% CI 0.95-0.98). [45]

CD

A literature search yielded only abstracts, describing MR studies of CD; no full published papers were found. Three abstracts, all by Parisinos and colleagues, describe the role of interleukein-6 in the pathogenesis of CD. [46-48] Two additional abstracts by Porcu and colleagues describe pleiotropic effects of the *GSDMB* gene on diseases that are "suggestive of mechanistic connections" including RA and CD. [49, 50]

Drug target validation/repurposing

Studies have identified SNPs associated with soluble fraction of the IL6 receptor (sIL6R) and reviewed associated traits. PheWAS have then evaluated traits associated with such SNPs across the phenome. For example, rs4129267, which is an sIL6R variant, was found to have positive, statistically significant associations with monocyte count, mean platelet volume, eczema, and asthma. Mendelian randomization using 34 SNPs associated with sIL6R circulating levels showed causal inverse associations with various types of stroke, atrial fibrillation, coronary artery disease, abdominal aortic aneurism, and RA, suggesting protective effects of IL-6 inhibition, and positive associations with eczema and asthma, suggesting IL-6 inhibition increases risk of these diseases. Also included in these analyses were heritable longevity; sIL6R was positively associated with parental age at death, and various blood parameters, of which only C-reactive protein was significantly, inversely associated with sIL6R levels. [51] Another PheWAS reviewed other IL6R SNPs (rs2228145; rs4129267), in patients in the Million Veteran Program (MVP) with results replicated in data from the UK Biobank and Vanderbilt University Biobank. Significant inverse associations between these SNPs and the following traits were found: aortic aneurysm, ischemic heart disease (including coronary atherosclerosis, ischemic heart disease, and myocardial infarction), and vascular diseases such as peripheral vascular disease and atherosclerosis of the extremities. Significant positive associations were identified for skin conditions such as atopic dermatitis (eczema), seborrheic dermatitis, and erythematosquamous dermatosis. Several pulmonary, renal, and eye conditions were also positively associated with these IL6R SNPs, as were two musculoskeletal conditions (acquired deformities of finger and gouty arthropathy). [52] While these two PheWAS identified traits

associated with SNPs that are IL-6 inhibitors, a search of the literature did not find any PheWAS that specifically evaluated traits associated with the specific SNP rs7529229.

Innovation

This study will fill an important gap in the literature by evaluating the causal associations between two IMIDs. It is important to determine whether a second IMID is caused by having a first IMID. RA and CD all have a substantial public health impact, causing significant pain, lost time and productivity at work and physical limitations. These are chronic illnesses patients may live with for decades and any one of them may be debilitating, leading to significant pain, lost productivity, and health care costs; a second IMID will significantly compound these problems. Long-term, this information will be helpful for health care providers who care for patients with one diagnosed IMID; it can encourage close surveillance with regard to subsequent development of another IMID, if appropriate. This information may be helpful in the pharmaceutical sector, where already treatments being developed for one IMID are tested and health authority approvals are received for the treatment of other IMIDs, though posology may be different (e.g., tofacitinib [Xeljanz, Pfizer],which is approved for three IMIDs: RA, ulcerative colitis, psoriatic arthritis; etanercept [Enbrel, Pfizer/Amgen], approved for five IMIDs: RA, plaque psoriasis, psoriatic arthritis, ankylosing spondylitis, and polyarticular juvenile idiopathic arthritis; adalimumab, which is approved for 10 IMIDs, including both RA and CD [Humira, AbbVie]). [31, 53, 54] Drug target repurposing can identify novel disease areas for existing therapies to bring new treatment options to patients while cutting down on time and cost in clinical trials. [29, 30, 33]

These studies address the study questions using novel design. MR is an epidemiologic study design that has some similarities with randomized controlled trials and avoids confounding. MR takes advantage of the natural randomization of genetic material at conception. MR's ability to

provide unconfounded estimates is important.[55] Other designs with similar properties exist, such as sibling or twin studies, but are not always feasible, whereas the investment in genetic studies has provided resources for MR. PheWAS examine the associations of specific genotypes with all phenotypes, providing an important tool in the search for new drug targets for existing therapies. The growth of the field of genetics and the ongoing publication of new GWAS is enabling the expansion of the use of MR and PheWAS to answer more questions using these methods. Mendelian Randomization and PheWAS are now also relatively easy to implement using an online tool, MR-Base, along with available R code. [56, 57]

STATISTICAL APPROACH

Conceptual/Theoretical Framework

Mendel's law of independent assortment states inheritance of traits is independent of inheritance of other traits. Davey Smith and Ebrahim described Mendelian randomization as taking advantage of "the random assortment of alleles at the time of gamete formation" [58, 59] to obtain unconfounded proxies of exposure. Genetic variants are independent of typical behavioral and environmental confounders; for this reason, Mendelian randomization is similar to a randomized study design in avoiding confounding but remains open to selection bias, because the randomization takes place before rather than after recruitment.

MR is an instrumental variable analysis with genetic variants as the instrument variable (IV). An instrumental variable is used to measure the effect of an exposure on an outcome when the exposure is open to confounding. An instrumental variable is typically described as needing to meet three conditions: relevance, exclusion-restriction and independence. Relevance indicates that the instrument is associated with the exposure. Exclusion-restriction indicates that the instrument affects the outcome only via effects on the exposure. Independence indicates that the

instrument does not share common causes with the outcome. [60-62] These assumptions are applied as follows in an MR study: the genotype (as the IV) must be associated with the phenotype/exposure disease (relevance), the genotype affects the outcome disease only through the phenotype (exclusion-restriction), and there are no common causes (confounders) of the association between genotype and the outcome disease of interest. Estimates for an MR study are obtained by combining the Wald estimates (genetic variant on outcome divided by genetic variant on exposure).

The Directed Analytic Graph (DAG) in **Figure 1.1** describes MR as an instrumental variable. The assumptions are indicated by the arrow from the IV to exposure (relevance), the lack of arrows from instrument to outcome (exclusion-restriction), and the lack of arrows from confounder to instrument (independence).

Preliminary data

A review of the literature provided background data on RA, CD, and drug target validation. The literature also provided background, on available MR studies on the two diseases and the use of drug target validation, as described above. The literature also yielded data on available GWAS that may be appropriate for use in the analysis (see below). However, no data were found describing the possible causal associations between the two IMIDs under study.

Study design

For the first two specific aims, a two sample MR study was implemented using a GWAS that provides genetic associations with the exposure and another GWAS that provides genetic associations with the outcome, because this approach enables research questions to be answered even if one study including both exposure and outcome does not exist. [63] Estimates of the

effect of exposure on outcome can then be made using a well-established technique, separate sample instrumental variable analysis, which only requires summary statistics from the GWAS, which are available in MR-base. [57]

For the third specific aim, the Phenome-Wide Association Study (PheWAS) function in the MR-base web-based application [64] was used to identify traits associated with this SNP to use as potential drug targets for tocilizumab.

Study population and eligibility criteria

The study populations were those included in published, publicly available, curated GWAS available in MR-base, [57] giving genetic associations with disease onset. Typically, to date most large GWAS have largely been conducted in middle-aged and older people of European descent. For the first two specific aims, the exposure was the disease of interest (RA or CD); the outcomes included the other autoimmune disease of interest. The third specific aim used the SNP rs7529229, which predicts response to tocilizumab in RA, as the exposure.

Data sources, collection, and management

MR-base [57] is a web-based platform that contains a curated catalog of summary statistics from many GWAS covering many diseases. Using the search function in this utility, it is possible to select instruments for an exposure and assess its effects on the disease areas of interest using summary statistics. The data are aggregated, deidentified, and publicly available.

The specific GWAS chosen for the three specific aims are described below.

RA: Specific aim 1 (exposure); Specific aim 2 (outcome)

A large, curated GWAS of RA which includes 22 combined GWAS (18 studies including participants of European background, 4 studies including those of Asian background). The

original data sources contained RA cases, of which 88.1% were seropositive for anti-citrullinated peptide antibody (ACPA), 9.3% were seronegative, and 2.6% had unknown antibody status. All RA cases had confirmed diagnosis according to the 1987 American College of Rheumatology criteria [65] or were diagnosed by a rheumatologist. The sample includes 19234 cases and 61565 controls. Geographically matched controls were drawn from multiple sources, including an age-related macular degeneration GWAS, a myocardial infarction GWAS, the UK 1958 birth cohort, the UK National Blood Services cohort, and non-autoimmune disease cases from the Wellcome Trust Case Control Consortium (WTCCC). [66, 67]

CD: Specific aim 1 (outcome); Specific aim 2 (exposure)

A large, curated GWAS of Crohn's disease is the International IBD Genetics Consortium (IIBDGC), which combined data from six CD GWAS, all of European descent, and identified a total of 71 loci including 17897 cases and 33977 controls. Patients in the IIBDGC-consortium GWAS were enrolled in medical centers in the U.S. and Canada (Cedars-Sinai Medical Center; Johns Hopkins University; University of Chicago; University of Montreal; University of Pittsburgh; University of Toronto Genetics Research Centers); CD was physician diagnosed using x-rays, endoscopy reports, pathology reports, and other diagnostic tools. The GWAS included geographically matched controls as well as additional age- and ethnicity-match controls from the New York Health Project. [67-71]

rs7529229, predictor of response to tocilizumab: Specific aim 3 (exposure)

Forty studies from the US, UK, and Europe contributed to the identification of the IL-6R SNP rs7529229 as associated with increased circulating IL-6 concentration, consistent with IL-6 inhibition from TCZ infusions for RA every four weeks (posology 4-8mg/kg). Total participants in these studies was 133,44. Mean age of participants of contributing studies ranged from late

40s to mid 70s, with one outlier (mean age 25.6), and with one study not reporting mean age. [13]

Sampling/Recruitment procedures

This study is a secondary analysis of publicly available data. The sampling/recruitment procedures for the underlying studies were reported, where available, and their implications for validity of the estimates considered.

Primary and secondary outcome definitions

RA: All RA cases in the combined GWAS described above had a clinical diagnosis of RA by a rheumatologist, and/or met the 1987 American College of Rheumatology diagnostic criteria. [65, 66]

CD: All CD cases in the IIBDGC-consortium GWAS were diagnosed by physicians using x-rays, endoscopy reports, pathology reports, or other diagnostic tools. [67-71]

Data analysis plan

Two-sample MR analyses were conducted for all three specific aims. For the first two specific aims, the primary analysis used the inverse variance weighted (IVW) method; sensitivity analyses were weighted median, MR-Egger, and MR-PRESSO (Pleiotropy RESidual Sum and Outlier) methods, which essentially compare the estimates from different ways of combining the genetic variant specific Wald estimates. Publicly available summary statistics from GWAS included in MR-base were used; the GWAS for each exposure and outcome is described above in the data sources section.

The analyses for the first two specific aims was conducted in the MR-base web application, [57] except for the MR-PRESSO sensitivity analyses, which was conducted in R. [56] A two sample

MR in MR-base is implemented in several steps. The first step is to select exposures from among GWAS available in the data catalog. The second step is to choose outcomes, from among those available in the catalog. Next, in the web-based application, the "Run MR" selection provides the inverse variance weighted, weighted median, and MR-Egger analyses. This utility matches allele direction for the SNPs across data sources, replaces palindromic SNPs as necessary and finds proxies for SNPs predicting the exposure but unavailable for the outcome. MR-PRESSO was conducted separately in R. [56, 72, 73]

Wald estimates show the effect of the exposure on the outcome for each included SNP. [74] For the first two specific aims, these Wald estimates were combined in a meta-analysis of the effect of the exposure on the outcome for each included SNP (ratio of SNP on outcome to SNP on exposure); this is the inverse variance weighted method and is the primary analysis for the first two specific aims. [74] This method uses the first-term of Fieller's theorem (variance of SNP on outcome divided by SNP on exposure) [75] to approximate the SNP-specific variance. This method assumes each included instrument fulfills the assumptions of an IV. Inclusion of pleiotropic SNPs may decrease the validity of this method.

The additional methods were used as sensitivity analyses to determine whether IV assumptions are violated or if an outlier SNP is driving an effect estimate. These sensitivity analyses use different assumptions. Two that were run in the MR-base platform include weighted median and MR-Egger. A weighted median of the SNP-specific Wald estimates is valid as long as >50% of the weight is contributed by valid instruments. [76] MR-Egger relaxes the assumption that there are no pleiotropic effects; this method is robust to invalid instruments, but assumes there no genetic confounders, i.e., the genetic instruments do cause a confounder of exposure and outcome. This is usually termed the Instrument strength independent of direct effects (InSIDE)

assumption. [77] MR-PRESSO identifies and removes outliers due to horizontal pleiotropy and tests differences in the estimates before and after this correction. MR-PRESSO requires that >50% are valid instruments. Removing outliers in this way leads to an unbiased estimate, [72, 73] but all these methods assume the InSIDE assumption.

When selecting the GWAS representing the disease area, SNPs with p-values below $5x10^{-8}$ were extracted and used in the MR analysis, to ensure that the SNPs strongly predicted the exposure, i.e., at genome wide significance, rather than being chance associations. Different SNPs are often very close together on the human genome, so their associations with a phenotype can be highly correlated. The r^2 (correlation coefficient) threshold for identifying correlations between SNPs (linkage disequilibrium) was 0.001, only keeping SNPs with correlation below this threshold gives independence between genetic variants predicting the exposure.[78, 79]

The F-statistic was calculated for each SNP and overall. The F-statistics for each SNP were calculated by dividing the beta of the exposure squared by the standard error of the exposure squared; the overall F-statistic is the average of all the F-statistics of the SNPs. [76] The F-statistic shows the strength of the instrument (F>10 suggests an instrument is not weak).

Associations are presented as odds ratios (ORs) with 95% confidence intervals (CIs) and p-values.

For the third specific aim, the SNP of interest was searched using the PheWAS function in MR-base [57] to identify other potential drug targets for TCZ, based on traits noted as associated with this SNP. The Bonferroni correction was used to account for multiple comparisons. The traits were aligned so that the minor allele was C for all traits to ensure appropriate direction of

association with the SNP among traits. PhenoScanner [80]was searched to identify the effect allele and its effect allele frequency (EAF).

Statistical power and sample size

The sample size for each aim were based on the available GWAS for exposure and outcome. Based on the large number of patients in each GWAS, it was expected that the results would be robust. MR is a powerful technique, and like all instrumental variable analyses, it requires large sample sizes. The sample size for an MR analysis is the sample size required for exposure and outcome divided by the r^2 for genetic variants on exposure. The minimum detectable effect size based on available sample sizes was provided for the first two specific aims. [81]

Study timeline

The goal was to complete one final paper for each specific aim by mid-March 2020, with anticipated defense in May 2020. Specific timelines are as follows:

Date	Milestone	Notes
January 3, 2020	Post-defense dissertation proposal revision circulated to committee	
January 15, 2020	Draft of paper 1 circulated to committee for review	Paper 1: Causal effect of RA on CD
January 31, 2020	Comments expected from committee on paper 1	
January 31, 2020	Draft of paper 2 circulated to committee for review	Paper 2: Causal effect of CD on RA

February 15, 2020	Comments expected from committee on paper 2	
February 29, 2020	Draft of paper 3 circulated to committee for review	Paper 3: A Phenome-Wide Association Study to Assess Potential New Targets for Tocilizumab, an Interleukin-6 Inhibitor
March 15, 2020	Comments expected from committee on paper 3	
January 31-April 15, 2020	Revision to papers 1-3; submission of revised full dissertation draft to committee for review	
May 7, 2020	Final comments received from committee	
May 15, 2020	Dissertation defense	

Generalizability

The results of these MR analyses are expected to have external validity, because there is no reason to think the causes of these diseases act via mechanisms that are population specific, although specific causes may be more relevant in some populations that others. [82]

Limitations

There are limitations to the MR method. First, the method relies on strong assumptions which can be difficult to verify empirically. This method required genetic instruments that predict the outcomes; GWAS have provided such instruments for both IMIDs and for a proxy of

tocilizumab. Second, genetic variants may have only a small effect on these diseases, which could introduce bias due to weak instruments, which were checked via the F-statistic. [83, 84] Additionally, effect size (small or otherwise) may not correspond to clinical significance, nor indicate effects of a clinical intervention at a future time, either due to the importance of cumulative exposure or time-dependent exposure. [64] Third, confounders of genetic variants on exposure may exist, typically as a result of population stratification; therefore, a homogeneous population is generally necessary for MR studies, with correction for population stratification. The current study used GWAS that primarily include people of European descent. Fourth, pleiotropy is another possibility; the instrument(s) should only impact the outcome through the exposure. The potential for pleiotropy requires sensitivity analyses. Fifth, RA and Crohn's are more common in women than men, however sex-specific analysis is not usually possible using summary data. Sixth, MR studies usually require extremely large sample sizes; this is typically addressed by the very large sample sizes in GWAS, though for some studies it can be challenging to include sufficient numbers of cases, depending on the prevalence of the disease and method of disease ascertainment. Finally, an important current limitation of MR studies is that they address causal factors of disease incidence, but not issues related to disease progression; studies of disease progression and studies of patients are open to selection bias because they are missing those who have already died because of the exposure or because of competing risk of the outcome. [83, 84]

PheWAS also have limitations. They are hypothesis-generating only; other studies such as randomized controlled trials are still required to confirm the effect suggested in a PheWAS. A Bonferroni correction for statistical significance may not be appropriate if the traits are not independent from each other. Differences across populations could affect both the expected

allele frequency for the minor allele (which can vary by population) and also access of different populations to medical care from which diagnosis and ultimately identification of traits in a PheWAS are possible. [85]

Human subjects protections

Not applicable; GWAS provide deidentified and aggregated data.

REFERENCES

1. Park, J., A. Mendy, and E.R. Vieira, Various Types of Arthritis in the United States: Prevalence and Age-Related Trends From 1999 to 2014. Am J Public Health, 2018. **108**(2): p. 256-258.
2. Sheriff, M.Z., et al., Opportunistic Infections Are More Prevalent in Crohn's Disease and Ulcerative Colitis: A Large Population-Based Study. Inflamm Bowel Dis, 2019.
3. Shivashankar, R., et al., Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. Clin Gastroenterol Hepatol, 2017. **15**(6): p. 857-863.
4. Ganz, M.L., et al., The Economic and Health-related Impact of Crohn's Disease in the United States: Evidence from a Nationally Representative Survey. Inflamm Bowel Dis, 2016. **22**(5): p. 1032-41.
5. Safiri, S., et al., Global, regional and national burden of rheumatoid arthritis 1990-2017: a systematic analysis of the Global Burden of Disease study 2017. Ann Rheum Dis, 2019. **78**(11): p. 1463-1471.
6. Collaborators, G.B.D.I.B.D., The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol, 2020. **5**(1): p. 17-30.
7. Disease, G.B.D., I. Injury, and C. Prevalence, Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet, 2017. **390**(10100): p. 1211-1259.
8. Robinson, D., Jr., et al., Co-occurrence and comorbidities in patients with immune-mediated inflammatory disorders: an exploration using US healthcare claims data, 2001-2002. Curr Med Res Opin, 2006. **22**(5): p. 989-1000.
9. Cohen, R., et al., Autoimmune disease concomitance among inflammatory bowel disease patients in the United States, 2001-2002. Inflamm Bowel Dis, 2008. **14**(6): p. 738-43.
10. Blandizzi, C., et al., The role of tumour necrosis factor in the pathogenesis of immune-mediated diseases. Int J Immunopathol Pharmacol, 2014. **27**(1 Suppl): p. 1-10.
11. Lawlor, D.A., et al., Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med, 2008. **27**(8): p. 1133-63.
12. Collins, F.S. and H. Varmus, A new initiative on precision medicine. N Engl J Med, 2015. **372**(9): p. 793-5.
13. Interleukin-6 Receptor Mendelian Randomisation Analysis, C., et al., The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet, 2012. **379**(9822): p. 1214-24.
14. Disease, G.B.D., I. Injury, and C. Prevalence, Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet, 2018. **392**(10159): p. 1789-1858.
15. Fairweather, D., S. Frisancho-Kiss, and N.R. Rose, Sex differences in autoimmune disease from a pathological perspective. Am J Pathol, 2008. **173**(3): p. 600-9.
16. **https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html.** . January 9, 2019]; Available from: https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html. .

17. Innala, L., et al., Age at onset determines severity and choice of treatment in early rheumatoid arthritis: a prospective study. Arthritis Res Ther, 2014. **16**(2): p. R94.
18. Goemaere, S., et al., Onset of symptoms of rheumatoid arthritis in relation to age, sex and menopausal transition. J Rheumatol, 1990. **17**(12): p. 1620-2.
19. Myasoedova, E., et al., Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. Arthritis Rheum, 2010. **62**(6): p. 1576-82.
20. Crane, M.M., et al., Epidemiology and Treatment of New-Onset and Established Rheumatoid Arthritis in an Insured US Population. Arthritis Care Res (Hoboken), 2015. **67**(12): p. 1646-55.
21. Cross, M., et al., The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. Ann Rheum Dis, 2014. **73**(7): p. 1316-22.
22. Sokka, T., B. Abelson, and T. Pincus, Mortality in rheumatoid arthritis: 2008 update. Clin Exp Rheumatol, 2008. **26**(5 Suppl 51): p. S35-61.
23. Lichtenstein, G.R., et al., ACG Clinical Guideline: Management of Crohn's Disease in Adults. Am J Gastroenterol, 2018. **113**(4): p. 481-517.
24. Quezada, S.M., E.K. Steinberger, and R.K. Cross, Association of age at diagnosis and Crohn's disease phenotype. Age Ageing, 2013. **42**(1): p. 102-6.
25. Ng, S.C., et al., Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet, 2018. **390**(10114): p. 2769-2778.
26. Munkholm, P., et al., Intestinal cancer risk and mortality in patients with Crohn's disease. Gastroenterology, 1993. **105**(6): p. 1716-23.
27. Bitton, A., et al., Mortality Trends in Crohn's Disease and Ulcerative Colitis: A Population-based Study in Quebec, Canada. Inflamm Bowel Dis, 2016. **22**(2): p. 416-23.
28. Aniwan, S., et al., Overall and Cause-Specific Mortality of Inflammatory Bowel Disease in Olmsted County, Minnesota, From 1970 Through 2016. Mayo Clin Proc, 2018. **93**(10): p. 1415-1422.
29. King, E.A., J.W. Davis, and J.F. Degner, Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet, 2019. **15**(12): p. e1008489.
30. Roberts, R., Mendelian Randomization Studies Promise to Shorten the Journey to FDA Approval. JACC Basic Transl Sci, 2018. **3**(5): p. 690-703.
31. Humira prescribing information. December 13, 2019]; Available from: https://www.rxabbvie.com/pdf/humira.pdf.
32. Schmidt, A.F.F., C.; Gordillo-Marañón M, et al., Genetic drug target validation using Mendelian randomization. bioRxiv 2019.
33. Nelson, M.R., et al., The support of human genetic evidence for approved drug indications. Nat Genet, 2015. **47**(8): p. 856-60.
34. Tan, C.Y., J.T.; Tan, L., Genome-wide Association Study Based on A/T/N System Identifies New Susceptibility Loci for Alzheimer's Disease. 2018. **90**(15 (Supplement)): p. P2.194.
35. Theis, K.A., et al., Employment exit and entry among U.S. adults with and without arthritis during the Great Recession. A longitudinal study: 2007-2009, NHIS/MEPS. Work, 2018. **60**(2): p. 303-318.
36. Murphy, L.B., et al., Medical Expenditures and Earnings Losses Among US Adults With Arthritis in 2013. Arthritis Care Res (Hoboken), 2018. **70**(6): p. 869-876.

37. Barbour, K.E., et al., Vital Signs: Prevalence of Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation - United States, 2013-2015. MMWR Morb Mortal Wkly Rep, 2017. **66**(9): p. 246-253.
38. Cohen, R.D., The quality of life in patients with Crohn's disease. Aliment Pharmacol Ther, 2002. **16**(9): p. 1603-9.
39. Sandler, R.S., et al., The burden of selected digestive diseases in the United States. Gastroenterology, 2002. **122**(5): p. 1500-11.
40. Ryaboshapkina, M. and M. Hammar, Tissue-specific genes as an underutilized resource in drug discovery. Sci Rep, 2019. **9**(1): p. 7233.
41. Li, B., et al., Circulating interleukin-6 and rheumatoid arthritis: A Mendelian randomization meta-analysis. Medicine (Baltimore), 2016. **95**(23): p. e3855.
42. Interleukin 1 Genetics, C., Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. Lancet Diabetes Endocrinol, 2015. **3**(4): p. 243-53.
43. Viatte, S., et al., The role of genetic polymorphisms regulating vitamin D levels in rheumatoid arthritis outcome: a Mendelian randomisation approach. Ann Rheum Dis, 2014. **73**(7): p. 1430-3.
44. Yarwood, A., et al., Testing the role of vitamin D in response to antitumour necrosis factor alpha therapy in a UK cohort: a Mendelian randomisation approach. Ann Rheum Dis, 2014. **73**(5): p. 938-40.
45. Zhao, J.V. and C.M. Schooling, Role of linoleic acid in autoimmune disorders: a Mendelian randomisation study. Ann Rheum Dis, 2019. **78**(5): p. 711-713.
46. Parisinos C, S.S., Patel R, et al, OTU-006 The interleukin-6 receptor as a drug target in inflammatory bowel disease; a mendelian randomisation study. Gut, 2018. **67**: p. A55.
47. Parisinos C, S.S., Katsoulis M, et al, The interleukin-6 receptor as a drug target in inflammatory bowel disease; a mendelian randomization study. . United Eruopean Gastroenterology Journal., 2018. **6**(A42-3).
48. Parisinos CA, S.S., Katsoulis M, et al., The interleukin-6 receptor as a target for prevention of Crohn's disease and ulcerative colitis; A Mendelian randomisation study. Journal of Crohn's and Colitis. Conference: 13th Congress of European Crohn's and Colitis Organisation, ECCO 2018. Austria. 12 (Supplement 1) (pp S530-S531), 2018. Date of Publication: February 2018. . Journal of Crohn’s & Colitis, 2018. **12**(S530-1).
49. Porcu E, L.K., Reuger S, et al., Mendelian randomization integrating GWAS and EQTL data reveals genetic determinants of complex and clinical traits. . Human Heredity, 2018. **83**(242-3).
50. Porcu E, R.S., Santoni FA, et al., Mendelian randomization combining GWAS and eQTL data reveals new loci, extensive pleiotropy and genetic determinants of complex and clinical traits. Human Heredity. . Human Heredity, 2018. **83**: p. 20.
51. Rosa, M., et al., A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. NPJ Genom Med, 2019. **4**: p. 23.
52. Cai, T., et al., Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy: A Phenome-Wide Association Study. JAMA Cardiol, 2018. **3**(9): p. 849-857.
53. **Xeljanz prescribing information.** November 2, 2019]; Available from: http://labeling.pfizer.com/ShowLabeling.aspx?id=959

54. Enbrel prescribing information. November 2, 2019]; ≈]. Available from: https://www.enbrel.com/how-enbrel-works?isipaid=true&gclsrc=aw.ds&&gclid=CjwKCAjw0vTtBRBREiwA3URt7kkmPen8DLOORE5kjSt9wpvycsKUURcqExQdbwBjffRRZZa_FJFxFRoCYOsQAvD_BwE.
55. DiPrete, T.A., C.A.P. Burik, and P.D. Koellinger, Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data. Proc Natl Acad Sci U S A, 2018. **115**(22): p. E4970-E4979.
56. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA
57. Hemani, G., et al., The MR-Base platform supports systematic causal inference across the human phenome. Elife, 2018. **7**.
58. Davey Smith, G., Ebrahim S., 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?* Int J Epidemiol, 2003. **32**: p. 1-22.
59. Davey Smith, G.a.E.S., Mendelian randomization: prospects, potentials, and limitations. I. Int J Epidemiol, 2004. **33**: p. 30-42.
60. Lousdal, M.L., An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol, 2018. **15**: p. 1.
61. Baser, O., Too much ado about instrumental variable approach: is the cure worse than the disease? Value Health, 2009. **12**(8): p. 1201-9.
62. Swanson, S.A. and M.A. Hernan, Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology, 2013. **24**(3): p. 370-4.
63. Burgess, S., et al., Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol, 2015. **30**(7): p. 543-52.
64. Walker, V.M., et al., Using the MR-Base platform to investigate risk factors and drug targets for thousands of phenotypes. Wellcome Open Res, 2019. **4**: p. 113.
65. Arnett, F.C., et al., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum, 1988. **31**(3): p. 315-24.
66. Okada, Y., et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature, 2014. **506**(7488): p. 376-81.
67. Wellcome Trust Case Control, C., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. **447**(7145): p. 661-78.
68. Liu, J.Z. and C.A. Anderson, Genetic studies of Crohn's disease: past, present and future. Best Pract Res Clin Gastroenterol, 2014. **28**(3): p. 373-86.
69. Franke, A., et al., Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet, 2010. **42**(12): p. 1118-25.
70. Jostins, L., et al., Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature, 2012. **491**(7422): p. 119-24.
71. Barrett, J.C., et al., Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet, 2008. **40**(8): p. 955-62.
72. Verbanck, M., et al., Publisher Correction: Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet, 2018. **50**(8): p. 1196.
73. Verbanck, M., et al., Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet, 2018. **50**(5): p. 693-698.

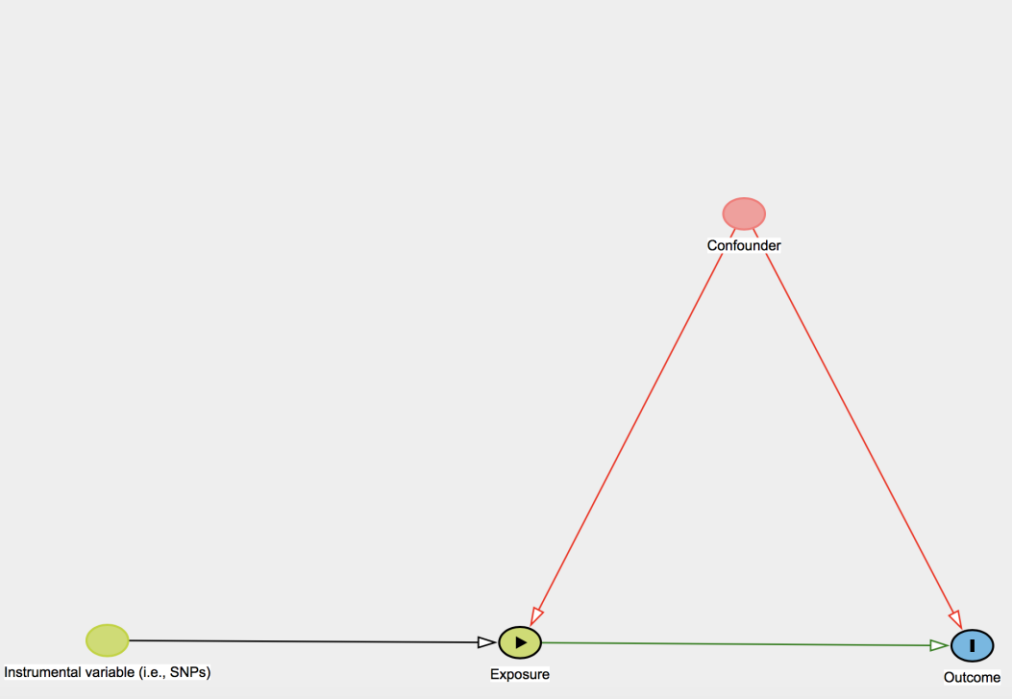
74. Wald, A., The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics, 1940. **11**: p. 284-300.
75. Fieller, E.C., Some problems in interval estimation. Journal of the Royal Statistical Society, Series B, 1954. **16**(2): p. 175-85.
76. Bowden, J., et al., Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol, 2016. **40**(4): p. 304-14.
77. Burgess, S. and S.G. Thompson, Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol, 2017. **32**(5): p. 377-389.
78. VanLiere, J.M. and N.A. Rosenberg, Mathematical properties of the r2 measure of linkage disequilibrium. Theor Popul Biol, 2008. **74**(1): p. 130-7.
79. Hill, W.G. and A. Robertson, Linkage disequilibrium in finite populations. Theor Appl Genet, 1968. **38**(6): p. 226-31.
80. Phenoscanner. December, 4, 2019]; Available from: http://www.phenoscanner.medschl.cam.ac.uk.
81. Freeman, G., B.J. Cowling, and C.M. Schooling, Power and sample size calculations for Mendelian randomization studies using one genetic instrument. Int J Epidemiol, 2013. **42**(4): p. 1157-63.
82. Lopez, P.M., S.V. Subramanian, and C.M. Schooling, Effect measure modification conceptualized using selection diagrams as mediation by mechanisms of varying population-level relevance. J Clin Epidemiol, 2019. **113**: p. 123-128.
83. Schooling, C.M., G. Freeman, and B.J. Cowling, Mendelian randomization and estimation of treatment efficacy for chronic diseases. Am J Epidemiol, 2013. **177**(10): p. 1128-33.
84. Schooling, C.M.L., P.; Au Yeung, S. L., et. al., Survival bias and competing risk can severely bias Mendelian Randomization studies of specific conditions. https://www.biorxiv.org/content/biorxiv/early/2019/10/21/716621.full.pdf, 2019.
85. Hebbring, S.J., The challenges, advantages and future of phenome-wide association studies. Immunology, 2014. **141**(2): p. 157-65.
86. Textor, J., et al., Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol, 2016. **45**(6): p. 1887-1894.
87. Eaton, W.W., et al., Epidemiology of autoimmune diseases in Denmark. J Autoimmun, 2007. **29**(1): p. 1-9.
88. Weng, X., et al., Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern california-managed care organization. Am J Gastroenterol, 2007. **102**(7): p. 1429-35.
89. Srirangan, S. and E.H. Choy, The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. Ther Adv Musculoskelet Dis, 2010. **2**(5): p. 247-56.
90. Parkes, M., et al., Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet, 2007. **39**(7): p. 830-2.
91. Jordan, D.M., M. Verbanck, and R. Do, HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. Genome Biol, 2019. **20**(1): p. 222.
92. Higgins, J.P., et al., Measuring inconsistency in meta-analyses. BMJ, 2003. **327**(7414): p. 557-60.

93. Burgess, S., Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol, 2014. **43**(3): p. 922-9.
94. Lang, J., et al., Diaphragm disease: pathology of disease of the small intestine induced by non-steroidal anti-inflammatory drugs. J Clin Pathol, 1988. **41**(5): p. 516-26.
95. Olsen, N.J., Drug-induced autoimmunity. Best Pract Res Clin Rheumatol, 2004. **18**(5): p. 677-88.
96. Parikh-Patel, A., et al., Risk of cancer among rheumatoid arthritis patients in California. Cancer Causes Control, 2009. **20**(6): p. 1001-10.
97. Garg, S.K., F.S. Velayos, and J.B. Kisiel, Intestinal and Nonintestinal Cancer Risks for Patients with Crohn's Disease. Gastroenterol Clin North Am, 2017. **46**(3): p. 515-529.
98. Doria, A., et al., Autoinflammation and autoimmunity: bridging the divide. Autoimmun Rev, 2012. **12**(1): p. 22-30.
99. Duerr, R.H., et al., A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science, 2006. **314**(5804): p. 1461-3.
100. Rioux, J.D., et al., Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet, 2007. **39**(5): p. 596-604.
101. Halling, M.L., et al., Patients with inflammatory bowel disease have increased risk of autoimmune and inflammatory diseases. World J Gastroenterol, 2017. **23**(33): p. 6137-6146.
102. Gross, V., et al., Evidence for continuous stimulation of interleukin-6 production in Crohn's disease. Gastroenterology, 1992. **102**(2): p. 514-9.
103. Gheita, T.A., et al., Involvement of IL-23 in enteropathic arthritis patients with inflammatory bowel disease: preliminary results. Clin Rheumatol, 2014. **33**(5): p. 713-7.
104. Rothman, K.J., Causes. Am J Epidemiol, 1976. **104**(6): p. 587-92.
105. Remicade prescribing information. March 21, 2020]; Available from: http://www.janssenlabels.com/package-insert/product-monograph/prescribing-information/REMICADE-pi.pdf.
106. Actemra prescribing information. December 23, 2019]; Available from: https://www.gene.com/download/pdf/actemra_prescribing.pdf.
107. Kevzara prescribing information.
108. Bowden, J., G. Davey Smith, and S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol, 2015. **44**(2): p. 512-25.
109. Orchard, T.R., Management of arthritis in patients with inflammatory bowel disease. Gastroenterol Hepatol (N Y), 2012. **8**(5): p. 327-9.
110. Vavricka, S.R., et al., Frequency and risk factors for extraintestinal manifestations in the Swiss inflammatory bowel disease cohort. Am J Gastroenterol, 2011. **106**(1): p. 110-9.
111. Zippi, M., et al., Extraintestinal manifestations in a large series of Italian inflammatory bowel disease patients. World J Gastroenterol, 2014. **20**(46): p. 17463-7.
112. Hsu, Y.C., et al., Gastrointestinal complications and extraintestinal manifestations of inflammatory bowel disease in Taiwan: A population-based study. J Chin Med Assoc, 2017. **80**(2): p. 56-62.
113. Harbord, M., et al., The First European Evidence-based Consensus on Extra-intestinal Manifestations in Inflammatory Bowel Disease. J Crohns Colitis, 2016. **10**(3): p. 239-54.

114. Cotsapas, C. and D.A. Hafler, Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol, 2013. **34**(1): p. 22-6.
115. Brennan, F.M., et al., Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. Lancet, 1989. **2**(8657): p. 244-7.
116. Knight, D.M., et al., Construction and initial characterization of a mouse-human chimeric anti-TNF antibody. Mol Immunol, 1993. **30**(16): p. 1443-53.
117. Thorbecke, G.J., et al., Involvement of endogenous tumor necrosis factor alpha and transforming growth factor beta during induction of collagen type II arthritis in mice. Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7375-9.
118. Williams, R.O., M. Feldmann, and R.N. Maini, Anti-tumor necrosis factor ameliorates joint disease in murine collagen-induced arthritis. Proc Natl Acad Sci U S A, 1992. **89**(20): p. 9784-8.
119. Papadakis, K.A. and S.R. Targan, Tumor necrosis factor: biology and therapeutic inhibitors. Gastroenterology, 2000. **119**(4): p. 1148-57.
120. Corazza, N., et al., Transmembrane tumor necrosis factor is a potent inducer of colitis even in the absence of its secreted form. Gastroenterology, 2004. **127**(3): p. 816-25.
121. Nakai, M., et al., The role of the tumor necrosis factor receptor in 2,4,6-trinitrobenzene sulfonic acid (TNBS)-induced colitis in mice. Dig Dis Sci, 2005. **50**(9): p. 1669-76.
122. Tilg, H., A. Moschen, and A. Kaser, Mode of function of biological anti-TNF agents in the treatment of inflammatory bowel diseases. Expert Opin Biol Ther, 2007. **7**(7): p. 1051-9.
123. Beyazal, M.S., Devrimsel, G., Cure, M.C., et al., Serum Levels of IL-17, IL-6, TNF-α and Disease Activity in Patients with Rheumatoid Arthritis. Akt Rheumatol, 2017. **42**: p. 323-8.
124. Mokry, L.E., et al., Interleukin-18 as a drug repositioning opportunity for inflammatory bowel disease: A Mendelian randomization study. Sci Rep, 2019. **9**(1): p. 9386.
125. Zheng, J., et al., Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep, 2017. **4**(4): p. 330-345.
126. Wurtz, P., et al., Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. J Am Coll Cardiol, 2016. **67**(10): p. 1200-1210.
127. Ference, B.A., et al., Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. J Am Coll Cardiol, 2015. **65**(15): p. 1552-61.
128. Myocardial Infarction Genetics Consortium, I., et al., Inactivating mutations in NPC1L1 and protection from coronary heart disease. N Engl J Med, 2014. **371**(22): p. 2072-82.
129. He, S., et al., A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. J Biol Chem, 2010. **285**(9): p. 6706-15.
130. Cohen, J.C., et al., Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med, 2006. **354**(12): p. 1264-72.
131. Navarese, E.P., et al., Effects of Proprotein Convertase Subtilisin/Kexin Type 9 Antibodies in Adults With Hypercholesterolemia: A Systematic Review and Meta-analysis. Ann Intern Med, 2015. **163**(1): p. 40-51.

132. Collaboration, C.R.P.C.H.D.G., et al., Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. BMJ, 2011. **342**: p. d548.
133. Zacho, J., et al., Genetically elevated C-reactive protein and ischemic vascular disease. N Engl J Med, 2008. **359**(18): p. 1897-908.
134. Polfus, L.M., R.A. Gibbs, and E. Boerwinkle, Coronary heart disease and genetic variants with low phospholipase A2 activity. N Engl J Med, 2015. **372**(3): p. 295-6.
135. Millwood, I.Y., et al., Lipoprotein-Associated Phospholipase A2 Loss-of-Function Variant and Risk of Vascular Diseases in 90,000 Chinese Adults. J Am Coll Cardiol, 2016. **67**(2): p. 230-231.
136. Pendergrass, S.A., et al., Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS Genet, 2013. **9**(1): p. e1003087.
137. Welter, D., et al., The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.
138. Denny, J.C., et al., Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol, 2013. **31**(12): p. 1102-10.
139. Coronary Artery Disease Genetics, C., A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet, 2011. **43**(4): p. 339-44.
140. Eyre, S., et al., High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet, 2012. **44**(12): p. 1336-40.
141. Agency, E.M., Assessment Report, RoActemra, C.f.M.P.f.H.U. (CHMP), Editor. 2018.
142. Neveu, W.A., et al., Elevation of IL-6 in the allergic asthmatic airway is independent of inflammation but associates with loss of central airway function. Respir Res, 2010. **11**: p. 28.
143. Rincon, M. and C.G. Irvin, Role of IL-6 in asthma and other inflammatory pulmonary diseases. Int J Biol Sci, 2012. **8**(9): p. 1281-90.
144. Yokoyama, A., et al., Circulating interleukin-6 levels in patients with bronchial asthma. Am J Respir Crit Care Med, 1995. **151**(5): p. 1354-8.
145. Tillie-Leblond, I., et al., Balance between proinflammatory cytokines and their inhibitors in bronchial lavage from patients with status asthmaticus. Am J Respir Crit Care Med, 1999. **159**(2): p. 487-94.
146. Hirano, T., Interleukin 6 and its receptor: ten years later. Int Rev Immunol, 1998. **16**(3-4): p. 249-84.
147. Wang, J., et al., Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. Pharmacogenomics J, 2013. **13**(3): p. 235-41.

Figure 1.1. Directed Analytic Graph (DAG) showing Mendelian Randomization as an Instrumental Variable. DAG made using DAGgity v.2.3, daggity.net. [86]



Chapter 2: An Assessment of Potential Causal Pathways Between Two Autoimmune Diseases: a Mendelian Randomization Study of Rheumatoid Arthritis as an antecedent and possible cause of Crohn's Disease

ABSTRACT

Background

Rheumatoid arthritis (RA) and Crohn's disease (CD) are two among a group of immune-mediated inflammatory diseases (IMIDs). These diseases are common, collectively affecting as many as two million Americans, and cause significant pain and disability. These and other IMIDs may co-occur in patients, and patients with one IMID may have an increased risk of another IMID. Although these diseases may share at least one common cause (e.g., tumor necrosis factor-alpha overexpression), it is not clear is whether the co-occurrence of a second IMID is caused by the existence of the first IMID.

Methods

A two-sample Mendelian randomization study was conducted to assess the possible causal effect of RA on CD. The primary analysis used the inverse variance weighted (IVW); sensitivity analyses include weighted median, MR-Egger, and MR-PRESSO (Pleiotropy RESidual Sum and Outlier) methods. Genetic predictors of RA obtained from a GWAS including 22 studies with participants from European and Asian backgrounds, with 19,234 cases and 61,565 controls, were applied to summary genetic associations for CD, from the International IBD Genetics Consortium (IIBDGC), (17,897 cases, 33,977 controls) of European descent.

Results

All four analyses showed a protective effect of RA on CD. (IVW odds ratio [OR] 0.89, 95% confidence interval [CI] 0.77-1.00, p=0.042; MR-Egger OR 0.71, 95% CI 0.50-0.93, p=0.004;

weighted median OR 0.78, 95% CI 0.72-0.84, $p=8.4x10^{-15}$; MR-PRESSO OR 0.92, 95% CI 0.84-1.00, p= 0.047).

Conclusions

These analyses suggest that RA protects from the development of Crohn's disease. It is possible this is due to issues with the data sources, or the infrequency of co-occurrence of the two diseases.

Keywords

Mendelian randomization; rheumatoid arthritis; Crohn's disease

INTRODUCTION

Rheumatoid arthritis (RA) and Crohn's disease (CD) are two among a large group of immune-mediated inflammatory diseases (IMIDs). Rheumatoid Arthritis (RA) is a common autoimmune disease characterized by pain and inflammation in the joints, leading to joint damage and significant loss of function. [16] Age of onset of RA is typically between 30 and 60 years. Estimates of the median age at disease onset range from 45-58 years. [17, 18] About 1.5 million Americans live with RA; [19] globally, 21 million people are affected. [5, 7] Crohn's disease (CD) is an autoimmune disease characterized by abdominal pain, diarrhea, and fatigue, weight loss, fever, growth failure, anemia, or recurrent fistulas. [23] There are two peak age ranges for onset of CD. The first is from 15-29 years and the second is from 55-59 years. [24] In the U.S., between 593,000 to 780,000 people live with CD, with an estimated 33,000 new patients diagnosed each year; [4] globally, as many as 7 million people live with CD. [6]

These IMIDs may co-occur, and it has been shown that patients with one IMID may have an increased risk of another IMID. The Danish National Patient Registry reported that the prevalence of co-occurrence of these two diseases was 0.0027%. [87] A US-based commercial insurance claims database reported that co-occurrence of RA in CD patients was 1.7%. [88] What is not clear is whether the co-occurrence of a second IMID is caused by the existence of the first IMID. [8-10, 89] RA and CD have different clinical presentations but may have a common pathogenesis. Causes of these diseases are not well understood. An over-expression of tumor necrosis factor (TNF) may be one causal factor for IMIDs; [10] interleukin 6 (IL-6) has also been implicated, but there are likely other causes as well, and causality may be multifactorial in nature.

Mendelian Randomization (MR) has emerged as an important tool that facilitates causal inference from observational studies. This study design can allow for avoidance of a common problem in epidemiologic studies, confounding, [11] because it takes advantage of the random allocation of genetic make-up at conception. MR compares disease status by genetically predicted exposure to obtain unconfounded estimates. MR implementation has been facilitated by the increasing availability of genome wide association studies (GWAS) genotyping millions of single nucleotide polymorphisms (SNPs). Other designs with similar properties exist, such as sibling or twin studies, but are not always feasible, whereas the investment in genetic studies has provided resources for MR. A number of MR studies have evaluated various potential causes of RA (e.g., IL-5 [41, 42]; Vitamin D [43] [44]; linoleic acid [45]), while fewer have looked at CD (only abstracts were found, looking at IL-6 [46-48] and the *GSDMB* gene[49, 50]*).* No published MR studies were found that looked at these two IMIDs as potentially causally related. This study addresses the question of whether RA causes CD using a MR, which is an instrumental variable analysis with genetic variants as the instrument variable (IV).

METHODS

MR as instrumental variable analysis

An instrumental variable should meet three conditions: relevance, exclusion-restriction and independence. Relevance indicates that the instrument is associated with the exposure. Exclusion-restriction indicates that the instrument affects the outcome only via effects on the exposure. Independence indicates that the instrument does not share common causes with the outcome. [60-62] The assumptions are applied as follows in an MR study: the genotype (as the IV) must be associated with the phenotype/exposure disease (relevance), the genotype affects the

outcome disease only through the phenotype (exclusion-restriction), and there are no common causes (confounders) of the association between genotype and the outcome disease of interest.

Data sources

Genetic predictors of RA

A large, curated GWAS of RA which includes 22 combined GWAS (18 studies including participants of European background, 4 studies including those of Asian background). The original data sources contained RA cases, of which 88.1% were seropositive for anti-citrullinated peptide antibody (ACPA), 9.3% were seronegative, and 2.6% had unknown antibody status. All RA cases had confirmed diagnosis according to the 1987 American College of Rheumatology criteria [65] or were diagnosed by a rheumatologist. The sample includes 19234 cases and 61565 controls. Geographically matched controls were drawn from multiple sources, including an age-related macular degeneration GWAS, a myocardial infarction GWAS, the UK 1958 birth cohort, the UK National Blood Services cohort, and non-autoimmune disease cases from the Wellcome Trust Case Control Consortium (WTCCC). [66, 67]

Genetic associations with Crohn's disease

A large, curated GWAS of Crohn's disease is the International IBD Genetics Consortium (IIBDGC), which combines data from six CD GWAS, all of European descent, including 17897 cases and 33977 controls. Patients in the IIBDGC-consortium GWAS were enrolled in medical centers in the U.S. and Canada (Cedars-Sinai Medical Center; Johns Hopkins University; University of Chicago; University of Montreal; University of Pittsburgh; University of Toronto Genetics Research Centers); CD was physician diagnosed using x-rays, endoscopy reports, pathology reports, and other diagnostic tools. The GWAS included geographically matched

controls as well as additional age- and ethnicity-match controls from the New York Health Project. Comprehensive demographic data for all component GWAS were not found, but one component GWAS (Welcome Trust Case Control Consortium) reported median age of participants as 45.7 years among the WTCCC panel and 43.9 years among the replication panel, and 60% of participants were <age 50; however these participants represent only a small proportion of the total participants in the IIBDGC. [67-71, 90]

Statistical analysis

All SNPs identified as predictors of RA with significance $p<5x10^{-8}$ were reviewed in PhenoScanner to ascertain potential pleiotropy. There are two types of pleiotropy: vertical and horizontal. Vertical pleiotropy occurs when a SNP has an effect on the outcome via a downstream effect. This in essence captures what MR does: it determines, with the SNP as the IV, whether an exposure has an effect on an outcome. Horizontal pleiotropy occurs when the SNP, as the IV, influences multiple diseases or traits, and thereby affects the disease independent of the exposure. [91] The PhenoScanner review was conducted to determine whether the SNPs predictive of RA are also associated with CD directly; this would suggest horizontal pleiotropy.

A two-sample MR analysis was conducted. SNPs to predict RA were selected as having p-values below $5x10^{-8}$. Independent SNPs were selected as having r^2 (correlation coefficient) threshold for identifying linkage disequilibrium as <0.001, suggesting independence between genetic variants predicting the exposure.[78, 79] The F-statistic was calculated for each SNP and overall [76] to show the strength of the instrument (F>10 suggests a strong instrument).

The primary analysis used the inverse variance weighted (IVW) method with multiplicative random effects. Sensitivity analyses included weighted median, MR-Egger, and MR-PRESSO methods. The inverse variance weighted method combines Wald estimates for each SNP (ratio of

SNP on outcome to SNP on exposure) [74] using the first-term of Fieller's theorem (variance of SNP on outcome divided by SNP on exposure) [75] to approximate the SNP-specific variance. This method assumes each included instrument fulfills the assumptions of an IV. Inclusion of pleiotropic SNPs may decrease the validity of this method.

The additional methods were used as sensitivity analyses to determine whether the IV assumptions are violated or if one or more outlier SNPs is driving an effect estimate. These sensitivity analyses use different assumptions. A weighted median of the SNP-specific Wald estimates is valid as long as >50% of the weight is contributed by valid instruments. [76] MR-Egger relaxes the assumption that there are no pleiotropic effects, but does not identify or correct for them. This method is robust to invalid instruments, but relies on the InSIDE assumption (INstrument Strength Independent of Direct Effect). [77] MR-PRESSO (Pleiotropy RESidual Sum and Outlier) identifies and removes outliers due to horizontal pleiotropy and tests differences in the estimates before and after this correction. The MR-Egger intercept p-value was reported because a significant intercept indicates that the IVW estimate may not be valid, and the MR-Egger I^2 value was calculated to determine how much variation across the meta-analysis is due to heterogeneity. [92] MR-PRESSO requires that >50% are valid instruments. Removing outliers in this way leads to an unbiased estimate. [72, 73]

The minimum effect size detectable with the current sample size, based on 80% power and significance 0.05, was calculated using code from Burgess, provided in Appendix 2. [93]

The MR-base platform [57]and R version 3.6.1 (2019-07-05) [56] were used in this analysis.

Protection of human subjects

Deidentified, publicly available summary data were used in this study; therefore Institutional Review Board approval was not required.

RESULTS

Forty-one independent SNPs were identified as predictors of RA (significance $p<5x10^{-8}$). These SNPs were reviewed in PhenoScanner [80]; all 41 SNPs are located on different genes, and horizontal pleiotropy was found, as five SNPs were also predictive of Crohn's disease: rs168962, rs212389, rs34536443, rs6679677, and rs9747973. On additional SNP was also associated with inflammatory bowel disease (rs9296009). Most of the RA-associated SNPs were associated with multiple other traits. Many were associated with other autoimmune diseases (e.g., allergic and celiac diseases) as well as an assortment of other diseases and traits (e.g., height, coronary artery disease, hypothyroidism). **Table 2.1** lists characteristics of these 41 SNPs.

The inverse variance weighted analysis indicated that RA is associated with a reduced risk of Crohn's disease (OR 0.89, 95% CI 0.77-1.00, p=0.042). Both the MR-Egger analysis and the weighted median analyses also suggest reduced risk of CD in RA patients (MR-Egger OR 0.71, 95% CI 0.50-0.93, p=0.004; weighted median OR 0.78, 95% CI 0.72-0.84, $p=8.4x10^{-15}$). The scatter plot (**Figure 2.1**) shows that the MR-Egger line missed the origin, and the intercept was significant, suggesting that the IVW estimate is not valid. For this reason, an MR-PRESSO analysis was conducted to evaluate the effect of eliminating outliers. The MR-PRESSO analysis also suggested a reduced risk of CD in RA patients (OR 0.92, 95% CI 0.84-1.00, p= 0.047). MR-PRESSO identified 15 SNPs as outliers potentially causing horizontal pleiotropy; this included the six identified via PhenoScanner with known associations with CD or inflammatory bowel disease, plus an additional 9 SNPs. The MR-PRESSO result suggests that even after

eliminating these 15 outlier SNPs, this decreased risk remains statistically significant. **Table 2.2** summarizes the results of the four analyses.

The F-statistic was calculated by dividing the squared beta of the exposure by the squared standard error of the exposure; the F-statistic average was 129, suggesting strong instruments (all F-statistics for individual SNPs were >10).

A calculation was performed using R [56] to determine the minimum effect size detectable given the available sample size. Assumptions in this calculation were as follows: 80% power, significance (alpha)=0.05, and the ratio of cases to controls (17,897 and 33,977, respectively) is 0.53. The minimum effect detectable with the available sample size and 80% power was 0.108, corresponding to an odds ratio of 1.11, which would represent an increased odds of CD in RA, or 0.90, which would represent a decreased odds of CD in RA. The R code for this calculation is provided in **Appendix 1.**

DISCUSSION

This MR study suggests that a decreased risk of CD in RA patients. With strong instruments as shown by high F-statistics, the assumptions for the primary and sensitivity analyses were met; even the removal of a substantial (but less than 50%) proportion of SNPs as potential outliers in the MR-PRESSO analysis yielded a significant result.

Co-occurrence of RA and CD vs co-occurrence of IMIDs

It is surprising that one IMID would be associated with a lower risk of another IMID, particularly in light of claims databases showing co-occurrence of IMIDs A study of two large commercially insured patient populations (US-based) reported co-occurrence of (any) IMIDs in 9.1-9.8% of patients. These datasets excluded Medicare and pension plan enrollees, suggesting

this may be a low estimate, based on the age distribution of some IMIDs, which have older ages of onset. [8] However, contrary to an almost 10% co-occurrence of *any* two IMIDs, the Danish National Patient Registry reported very low of co-occurrence of RA and CD specifically (0.0027%) [12], and a US-based commercial insurance claims database reported substantially higher (but still low) co-occurrence of RA and CD (1.7%.) [13] It seems that other IMIDs may co-occur, but not specifically RA with CD. There are, however, biases inherent in US-based commercial insurance database studies that could suggest underreporting of RA and CD. The age distribution of participants in these plans may be younger than the general population, and healthier (healthy enough to work, though immediate family members are typically also covered). Older patients (typically covered by Medicare rather than commercial insurance) are generally more likely to have an RA diagnosis, so these data sources may miss some overlap between RA and CD populations.

RA treatments potentially causing CD, in opposition to current MR results

Some treatments for RA could actually cause Crohn's disease. For example, there have been reports of Crohn's disease following use of non-steroidal anti-inflammatory drugs (NSAIDs), which used to be the first-line treatments for RA. [94] There are also biological mechanisms by which more recent drugs may cause autoimmunity. A receptor antagonist could lead to increases in other cytokines, which could cause another IMID in an RA patient taking a TNF-a inhibitor such as infliximab or etanercept. Treatment of an autoimmune disease may also unmask another untreated, subclinical disease. This has been seen infrequently in RA and CD patients whose treatment with a TNF-a inhibitor has led to a demyelinating disorder that mimics early multiple sclerosis, but such an association has not been shown with RA patients developing CD after this treatment. [95]

Role of interleukin-6 (IL-6)]and tumor necrosis factor-alpha (TNF-a) inhibition

It is difficult to identify reasons that RA should be protective for CD. There are thought to be multiple common contributing causes of these diseases, including IL-6 and TNF-a overexpression. A possibility is that treating this overexpression of a causal cytokine could simultaneously treat diagnosed RA and preemptively treat subclinical CD, which could prevent CD from being clinically diagnosed. As some medications are approved for the treatment of both RA and CD (e.g., adalimumab, marketed as Humira by Abbvie), [31] a shared causal pathway makes sense. This might suggest that the SNPs identified as causal for RA were actually causal for response to treatment for RA and not directly causal for RA itself. This was not assessed in this analysis.

Further differences between RA and CD

Finally, RA and CD are both disorders of the immune system, but they have different relations with cancer. Both women and men with RA have lower incidence of some cancers compared to the general population (e.g., in women, breast, ovary, uterus, cervix, colon/rectum, and melanoma and in men, prostate and colon/rectum) and similar incidence of many other cancers (e.g., in both women and men, stomach, pancreas, bladder, kidney, thyroid, brain/CNS, and myeloma; in women additionally, esophagus, liver, and Hodgkin's disease and in men additionally melanoma), compared to the general population. Incidence of a few cancers were increased (in women and men, lung, non-Hodgkin's lymphoma, and leukemia, and in men only, esophagus and liver), compared to the general population. [96] Patients with CD have increased risk of multiple types of colorectal cancers, Hodgkin's disease, lymphoma, melanoma, non-melanoma skin cancer, and cervical cancer in women. [97] Mortality from cancer is similar in RA patients compared to the general population; [22] mortality from cancer is higher in CD

patients compared to the general population. [27] These findings suggest that RA is more the result of a very vigilant immune system, whereas CD this vigilance does not appear to be evident for CD either because RA and CD represent different disorders of the immune system or because CD specifically pre-disposes to certain cancers. A distinction has been suggested between autoinflammatory diseases (AIDs) and autoimmune diseases (AD), of which CD is classified as one of the former and RA as one of the later. [98] Different pathogenic triggers and inflammatory processes may exist, but it is not clear whether these are clinically (or otherwise) useful distinctions.

Data sources

The most feasible reason for this surprising result is the role of bias related to the data source(s). The age distributions of the two data sources were not fully provided. Since RA is often diagnosed at an older age than CD, [17, 18, 24] if patients enrolled in the GWAS with diagnosed RA but without having been diagnosed with CD, then those participants may have been unlikely ever to develop CD. This could lead to the lower risk of CD in those patients as seen in this MR study; that it is not a mechanistic protective effect of RA, but a function of people diagnosed with RA but not CD having enrolled in the GWAS.

It is also unclear whether controls in each data source were "healthy controls" (e.g., free from the "other" disease), or if they represented the underlying cohort form which cases were drawn. In the former case, such an omission of RA patients in the CD data set and CD patients in the RA data set would underestimate the association seem between the diseases and could affect the direction of the association.

Protective effect of RA on CD

It is possible that these results are not due to issues related to the data sources and that having RA is truly protective for CD. A biological mechanism for this is not clear and would require further research.

Strengths and limitations

There are limitations to the MR method. First, the method relies on strong assumptions which can be difficult to verify empirically. This method requires a genetic instrument that predicts the outcome. GWAS have provided such instruments for RA. Genetic variants may have only a small effect on these diseases, which could introduce bias due to weak instruments. [83, 84] The 41 instruments were strongly associated with the exposure, based on F-statistics well over 10 (individually and overall). Second, confounders of genetic variants on exposure may exist, typically as a result of population stratification; therefore, a homogeneous population is generally necessary for MR studies, with correction for population stratification, [83, 84] as used here. Third, with regard to the exclusion-restriction assumption, the MR-PRESSO analysis suggests some horizontal pleiotropy. However, estimates were the same after removing 15 outliers. The low MR-Egger I^2 value of 8.30% suggests low heterogeneity across the SNPs included in the MR-Egger analysis. Fourth, effect size (small or otherwise) may not correspond to clinical significance, nor indicate effects of a clinical intervention at a future time, either due to the importance of cumulative exposure or time-dependent exposure. [64] Fifth, RA and Crohn's are more common in women than men, however sex-specific analysis is not publicly available. Sixth, MR studies usually require extremely large sample sizes; this is typically addressed by the very large sample sizes in the GWAS used in this study; there were 19234 cases and 61565 controls in the RA dataset and 17897 cases and 33977 controls in the CD dataset, and the study was well-powered to detect the protective effect of RA on CD (power=0.88).

Finally, an important current limitation of MR studies is that they can only address causal factors of disease incidence, but not issues related to disease progression; studies of disease progression and studies of patients are open to selection bias because they are missing those who have already died because of the exposure or because of competing risk of the outcome. [83, 84]

CONCLUSION

This Mendelian randomization study determined that rheumatoid arthritis is associated with a reduced risk of Crohn's disease. Further study would be required to better understand several this result, in particular, a biological mechanism that would cause patients with RA to be less likely to develop CD. However, it is highly possible that this result was driven by bias due to the age of the exposure population; participants who developed RA without having first developed CD may be unlikely ever to do so.

KEY MESSAGES

- Mendelian randomization is a powerful method that can determine causal relationships between exposures and outcomes, if appropriate instruments can be identified.
- Although IMIDs often co-occur, RA and CD co-occur infrequently.
- This MR study showed that RA is protective for CD; however, an underlying mechanism for this has not been described, and selection bias may be a factor.

REFERNCES

1. **https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html.** . January 9, 2019]; Available from: https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html. .
2. Innala, L., et al., Age at onset determines severity and choice of treatment in early rheumatoid arthritis: a prospective study. Arthritis Res Ther, 2014. **16**(2): p. R94.
3. Goemaere, S., et al., Onset of symptoms of rheumatoid arthritis in relation to age, sex and menopausal transition. J Rheumatol, 1990. **17**(12): p. 1620-2.
4. Myasoedova, E., et al., Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. Arthritis Rheum, 2010. **62**(6): p. 1576-82.
5. Disease, G.B.D., I. Injury, and C. Prevalence, Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet, 2017. **390**(10100): p. 1211-1259.
6. Safiri, S., et al., Global, regional and national burden of rheumatoid arthritis 1990-2017: a systematic analysis of the Global Burden of Disease study 2017. Ann Rheum Dis, 2019. **78**(11): p. 1463-1471.
7. Lichtenstein, G.R., et al., ACG Clinical Guideline: Management of Crohn's Disease in Adults. Am J Gastroenterol, 2018. **113**(4): p. 481-517.
8. Quezada, S.M., E.K. Steinberger, and R.K. Cross, Association of age at diagnosis and Crohn's disease phenotype. Age Ageing, 2013. **42**(1): p. 102-6.
9. Ganz, M.L., et al., The Economic and Health-related Impact of Crohn's Disease in the United States: Evidence from a Nationally Representative Survey. Inflamm Bowel Dis, 2016. **22**(5): p. 1032-41.
10. Collaborators, G.B.D.I.B.D., The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol, 2020. **5**(1): p. 17-30.
11. Eaton, W.W., et al., Epidemiology of autoimmune diseases in Denmark. J Autoimmun, 2007. **29**(1): p. 1-9.
12. Weng, X., et al., Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern california-managed care organization. Am J Gastroenterol, 2007. **102**(7): p. 1429-35.
13. Cohen, R., et al., Autoimmune disease concomitance among inflammatory bowel disease patients in the United States, 2001-2002. Inflamm Bowel Dis, 2008. **14**(6): p. 738-43.
14. Robinson, D., Jr., et al., Co-occurrence and comorbidities in patients with immune-mediated inflammatory disorders: an exploration using US healthcare claims data, 2001-2002. Curr Med Res Opin, 2006. **22**(5): p. 989-1000.
15. Blandizzi, C., et al., The role of tumour necrosis factor in the pathogenesis of immune-mediated diseases. Int J Immunopathol Pharmacol, 2014. **27**(1 Suppl): p. 1-10.
16. Srirangan, S. and E.H. Choy, The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. Ther Adv Musculoskelet Dis, 2010. **2**(5): p. 247-56.
17. Lawlor, D.A., et al., Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med, 2008. **27**(8): p. 1133-63.
18. Li, B., et al., Circulating interleukin-6 and rheumatoid arthritis: A Mendelian randomization meta-analysis. Medicine (Baltimore), 2016. **95**(23): p. e3855.

19. Interleukin 1 Genetics, C., Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. Lancet Diabetes Endocrinol, 2015. **3**(4): p. 243-53.
20. Viatte, S., et al., The role of genetic polymorphisms regulating vitamin D levels in rheumatoid arthritis outcome: a Mendelian randomisation approach. Ann Rheum Dis, 2014. **73**(7): p. 1430-3.
21. Yarwood, A., et al., Testing the role of vitamin D in response to antitumour necrosis factor alpha therapy in a UK cohort: a Mendelian randomisation approach. Ann Rheum Dis, 2014. **73**(5): p. 938-40.
22. Zhao, J.V. and C.M. Schooling, Role of linoleic acid in autoimmune disorders: a Mendelian randomisation study. Ann Rheum Dis, 2019. **78**(5): p. 711-713.
23. Parisinos C, S.S., Patel R, et al, OTU-006 The interleukin-6 receptor as a drug target in inflammatory bowel disease; a mendelian randomisation study. Gut, 2018. **67**: p. A55.
24. Parisinos C, S.S., Katsoulis M, et al, The interleukin-6 receptor as a drug target in inflammatory bowel disease; a mendelian randomization study. . United Eruopean Gastroenterology Journal., 2018. **6**(A42-3).
25. Parisinos CA, S.S., Katsoulis M, et al., The interleukin-6 receptor as a target for prevention of Crohn's disease and ulcerative colitis; A Mendelian randomisation study. Journal of Crohn's and Colitis. Conference: 13th Congress of European Crohn's and Colitis Organisation, ECCO 2018. Austria. 12 (Supplement 1) (pp S530-S531), 2018. Date of Publication: February 2018. . Journal of Crohn's & Colitis, 2018. **12**(S530-1).
26. Porcu E, L.K., Reuger S, et al., Mendelian randomization integrating GWAS and EQTL data reveals genetic determinants of complex and clinical traits. . Human Heredity, 2018. **83**(242-3).
27. Porcu E, R.S., Santoni FA, et al., Mendelian randomization combining GWAS and eQTL data reveals new loci, extensive pleiotropy and genetic determinants of complex and clinical traits. Human Heredity. . Human Heredity, 2018. **83**: p. 20.
28. Lousdal, M.L., An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol, 2018. **15**: p. 1.
29. Baser, O., Too much ado about instrumental variable approach: is the cure worse than the disease? Value Health, 2009. **12**(8): p. 1201-9.
30. Swanson, S.A. and M.A. Hernan, Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology, 2013. **24**(3): p. 370-4.
31. Arnett, F.C., et al., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum, 1988. **31**(3): p. 315-24.
32. Okada, Y., et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature, 2014. **506**(7488): p. 376-81.
33. Wellcome Trust Case Control, C., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. **447**(7145): p. 661-78.
34. Liu, J.Z. and C.A. Anderson, Genetic studies of Crohn's disease: past, present and future. Best Pract Res Clin Gastroenterol, 2014. **28**(3): p. 373-86.
35. Franke, A., et al., Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet, 2010. **42**(12): p. 1118-25.
36. Jostins, L., et al., Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature, 2012. **491**(7422): p. 119-24.

37. Barrett, J.C., et al., Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet, 2008. **40**(8): p. 955-62.
38. Parkes, M., et al., Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet, 2007. **39**(7): p. 830-2.
39. Jordan, D.M., M. Verbanck, and R. Do, HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. Genome Biol, 2019. **20**(1): p. 222.
40. VanLiere, J.M. and N.A. Rosenberg, Mathematical properties of the r2 measure of linkage disequilibrium. Theor Popul Biol, 2008. **74**(1): p. 130-7.
41. Hill, W.G. and A. Robertson, Linkage disequilibrium in finite populations. Theor Appl Genet, 1968. **38**(6): p. 226-31.
42. Bowden, J., et al., Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol, 2016. **40**(4): p. 304-14.
43. Wald, A., The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics, 1940. **11**: p. 284-300.
44. Fieller, E.C., Some problems in interval estimation. Journal of the Royal Statistical Society, Series B, 1954. **16**(2): p. 175-85.
45. Burgess, S. and S.G. Thompson, Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol, 2017. **32**(5): p. 377-389.
46. Higgins, J.P., et al., Measuring inconsistency in meta-analyses. BMJ, 2003. **327**(7414): p. 557-60.
47. Verbanck, M., et al., Publisher Correction: Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet, 2018. **50**(8): p. 1196.
48. Verbanck, M., et al., Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet, 2018. **50**(5): p. 693-698.
49. Burgess, S., Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol, 2014. **43**(3): p. 922-9.
50. Hemani, G., et al., The MR-Base platform supports systematic causal inference across the human phenome. Elife, 2018. **7**.
51. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA
52. Phenoscanner. December, 4, 2019]; Available from: http://www.phenoscanner.medschl.cam.ac.uk.
53. Lang, J., et al., Diaphragm disease: pathology of disease of the small intestine induced by non-steroidal anti-inflammatory drugs. J Clin Pathol, 1988. **41**(5): p. 516-26.
54. Olsen, N.J., Drug-induced autoimmunity. Best Pract Res Clin Rheumatol, 2004. **18**(5): p. 677-88.
55. Humira prescribing information. December 13, 2019]; Available from: https://www.rxabbvie.com/pdf/humira.pdf.
56. Parikh-Patel, A., et al., Risk of cancer among rheumatoid arthritis patients in California. Cancer Causes Control, 2009. **20**(6): p. 1001-10.

57. Garg, S.K., F.S. Velayos, and J.B. Kisiel, Intestinal and Nonintestinal Cancer Risks for Patients with Crohn's Disease. Gastroenterol Clin North Am, 2017. **46**(3): p. 515-529.
58. Sokka, T., B. Abelson, and T. Pincus, Mortality in rheumatoid arthritis: 2008 update. Clin Exp Rheumatol, 2008. **26**(5 Suppl 51): p. S35-61.
59. Bitton, A., et al., Mortality Trends in Crohn's Disease and Ulcerative Colitis: A Population-based Study in Quebec, Canada. Inflamm Bowel Dis, 2016. **22**(2): p. 416-23.
60. Doria, A., et al., Autoinflammation and autoimmunity: bridging the divide. Autoimmun Rev, 2012. **12**(1): p. 22-30.
61. Schooling, C.M., G. Freeman, and B.J. Cowling, Mendelian randomization and estimation of treatment efficacy for chronic diseases. Am J Epidemiol, 2013. **177**(10): p. 1128-33.
62. Schooling, C.M.L., P.; Au Yeung, S. L., et. al., Survival bias and competing risk can severely bias Mendelian Randomization studies of specific conditions. https://www.biorxiv.org/content/biorxiv/early/2019/10/21/716621.full.pdf, 2019.
63. Walker, V.M., et al., Using the MR-Base platform to investigate risk factors and drug targets for thousands of phenotypes. Wellcome Open Res, 2019. **4**: p. 113.

Table 2.1. SNPs identified in MR-base as associated with RA (p<5x10-8), and their location as determined in PhenoScanner. [80] Rows highlighted in yellow were identified as outliers and removed in the MR-PRESSO analysis; red text indicates SNPs identified as associated with Crohn's disease or inflammatory bowel disease in Phenoscanner, suggesting possible horizontal pleiotropy.

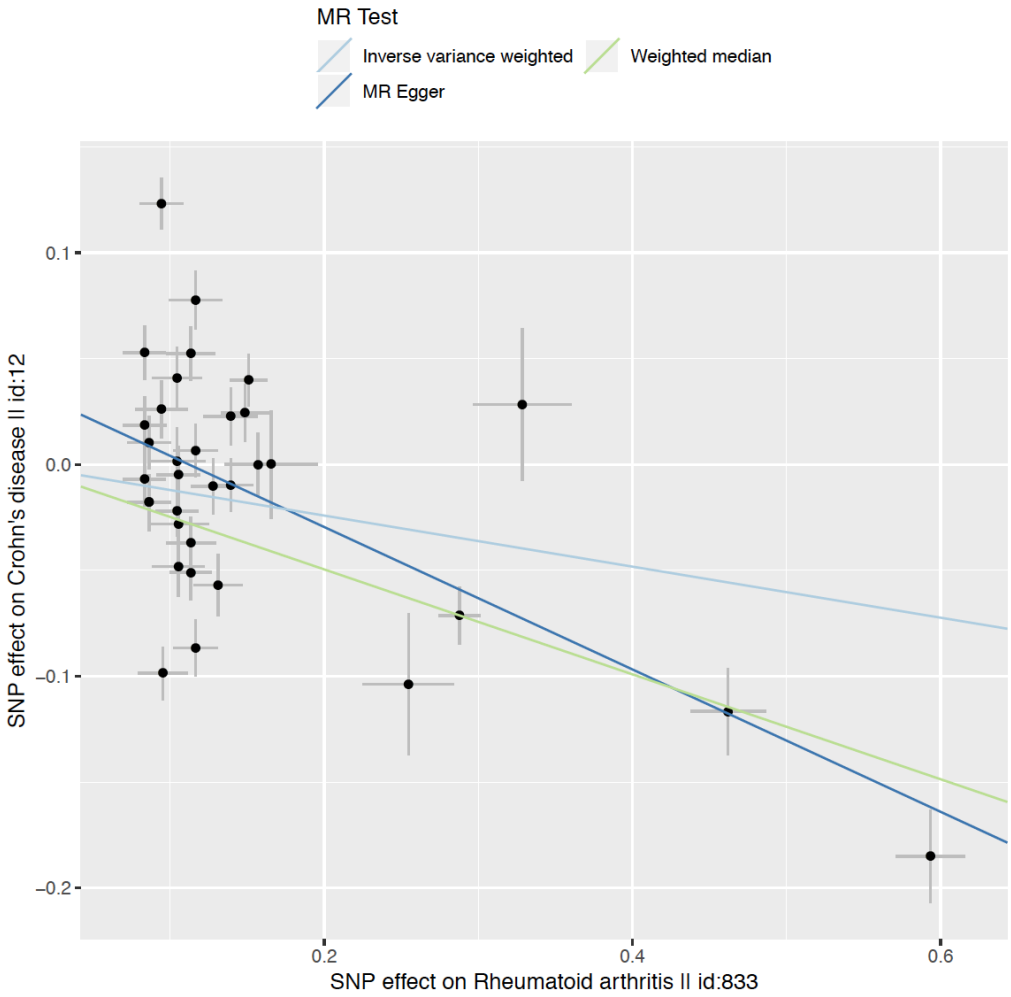
Variant:	Gene
rs11574914	*CCL21*
rs11889341	*stat4*
rs11933540	*RP11-324H7.1*
rs12126142	*IL6R*
rs12466919	*LINC01185*
rs13330176	*RP11-542M13.2*
rs1571878	*CCR6*
rs168962	*ZFP36L1*
rs1858036	*SPRED2*
rs1893592	*UBASH3A*
rs1953126	*PHF19*
rs2105325	*RP11-296O14.2*
rs212389	*RP1-111C20.3*
rs2301888	*PADI4*
rs2561477	*C5orf30*
rs2736337	*BLK*
rs3087243	*CTLA4*
rs34536443	*TYK2*
rs3757387	*IRF5*
rs3784099	*RAD51B*
rs3806624	*EOMES*
rs4239702	*CD40*
rs4409785	*RP11-338H14.1*
rs6679677	*PHTF1*
rs6712515	*LINC01104*
rs706778	*IL2RA*
rs71508903	*ARID5B*
rs73081554	*P11-80H18.3*
rs773125	*SUOX*
rs7752903	*AL356739.1*
rs7754520	*XXbac-BPG116M5.17*
rs8026898	*PCAT29*
rs8032939	*RASGRP1*
rs876938	*MMEL1*
rs909685	*SYNGR1*
rs9277411	*HLA-DPB1*
rs9277956	*HTATSF1P1*
rs9296009	*PRRT1*
rs9348832	*MAS1L*
rs947474	*RP11-563J2.3*
rs9747973	*GRB7*

Table 1.2. Mendelian randomization association of RA on CD, based on 34 independent SNPs* (p-values all <5x10^{-8}).

Method	OR	95% CI	Pval	Cochran's Q-statistic (p-value)	MR-Egger intercept p-value	MR-Egger I^2
MR-Egger	0.71	0.50-0.93	0.004	386.7 (0)	0.028	8.30%
Weighted median	0.78	0.72-0.84	8.4E-15			
Inverse variance weighted	0.89	0.77-1.003	0.042	451.0 (0)		
MR-PRESSO	0.92	0.84-1.00	0.047			

*34 of 41 SNPs were included in the IVW, weighted median, and MR-Egger analyses; 7 SNPs were palindromic and therefore excluded. The MR-PRESSO analysis based on 26 SNPs after elimination of 15 outliers.

Figure 2.1. Scatter plot of IVW, MR Egger, and weighted median analyses showing effect of SNPs on RA and CD.



Chapter 3: An Assessment of Potential Causal Pathways Between Two Autoimmune Diseases: a Mendelian Randomization Study of whether Crohn's Disease as an antecedent and possible cause of Rheumatoid Arthritis

ABSTRACT

Background

Rheumatoid arthritis (RA) and Crohn's disease (CD) are two immune-mediated inflammatory diseases (IMIDs), together affecting as many as two million Americans. These diseases cause significant pain and disability. Studies show that two or more IMIDs may co-occur in patients. RA and CD may share several common causes (e.g., tumor necrosis factor-alpha overexpression, excess interleukin 6), but it is not clear if CD itself causes rheumatoid arthritis.

Methods

A two-sample Mendelian randomization study was conducted to assess a possible causal role of CD on RA. The primary analysis used inverse variance weighting (IVW) to determine the effect of CD on RA. Sensitivity analyses included weighted median, MR-Egger, and MR-PRESSO (Pleiotropy RESidual Sum and Outlier). Genetic predictors of CD from summary statistics from the International IBD Genetics Consortium (IIBDGC), (17,897 cases, 33,977 controls of European descent) were applied to a GWAS of RA that included 22 studies including patients of European and Asian backgrounds (19,234 cases, 61,565 controls).

Results

The primary analysis and all three sensitivity analyses showed no causal effect of CD on RA (IVW odds ratio [OR] 0.97, 95% confidence interval [CI] 0.88-1.07, p=0.59; weighted median OR 0.98, 95% CI 0.94-1.03, p=0.49; MR-Egger OR 0.92, 95% CI 0.67-1.18, p=0.54, MR-PRESSO OR 1.00, 95% CI 0.64-1.36, p=0.90).

Conclusions

These analyses suggest that CD does not cause rheumatoid arthritis. Co-occurrence of CD and RA is fairly uncommon, and when it does occur, may be due to shared causes of both conditions such as overexpression of TNF-a or IL-6.

Keywords

Mendelian randomization; Crohn’s disease; rheumatoid arthritis

INTRODUCTION

Crohn's disease (CD) and Rheumatoid Arthritis (RA) are two among a large group of immune-mediated inflammatory diseases (IMIDs). Crohn's disease (CD) is characterized by abdominal pain, diarrhea, fatigue, weight loss, fever, growth failure, anemia, or recurrent fistulas. [23] Age of onset of CD peaks at age 15-29 years and again at 55-59 years. [24] In the U.S., it is estimated that 593,000 to 780,000 people live with CD, with an estimated 33,000 new cases per year. [4], and prevalence estimates range from 0.1%-0.15%. [99, 100] Rheumatoid Arthritis (RA) is a common autoimmune disease characterized by pain and inflammation in the joints. [16] Estimates of median age of onset in the US range from 45 years in women/50 years in men [18] to 58 years overall. [17] Estimated prevalence of RA in the US is 0.72% (95% CI 0.66–0.77), with RA affecting about 1.5 million Americans. [19]

IMIDs often co-occur, patients with one IMID may have a higher risk of developing another IMID. A study using two claims databases, the IMS Health Integrated Administration Claims Database and the Market Scan Commercial Claims and Encounters Database, evaluated the co-occurrence of CD and other IMIDs between 2001 and 2002. In both datasets, there was an increased risk of RA in CD patients (in MarketScan, OR 2.37, 95% CI 2.02-2.77; in IMS Health, OR 2.76, 95% CI 2.40-3.18). [9] The Danish National Patient Registry reported increased risk of rheumatoid arthritis in CD patients (OR 2.1; 95% CI 1.6-2.8) [101] but also reported a very low prevalence of co-occurrence of these two diseases (0.0027%) [87] A US-based commercial insurance claims database also reported increased risk of RA in CD patients (OR 1.9, 95% CI 1.5–2.3); prevalence of RA in CD patients was 1.7%. [88]

RA and CD have different clinical presentations but may share several common causes. Various cytokines have been implicated in the development of CD, RA, and other autoimmune diseases;

an over-expression of tumor necrosis factor (TNF) may be one causal factor; [10] interleukin 6 (IL-6) [41, 89, 102] and interleukin 23 (IL-23) [103] have also been implicated. Causality is likely multifactorial, due to multiple risk factors, of which none appears to be a necessary cause. [104] Different cytokine targets of different treatments also suggests multiple causal pathways. Tumor necrosis factor alpha (TNF-a) inhibitors include etanercept (Enbrel, marketed by Pfizer), [54] infliximab (Remicade, marketed by Janssen), [105]and adalimumab (Humira, marketed by AbbVie); [31] interleukin-6 (IL-6) inhibitors include tocilizumab (Actemra, marketed by Genentech) [106] and sarilumab (Kevzara, marketed by Regeneron and Sanofi). [107]

The public health impact of each of these two diseases is substantial The Global Burden of Disease study reported YLDs (years lived with disability) at 335,400 for CD. [14] In a review of Health-Related Quality of Life (HQOL) studies, compared to healthy controls, CD patients scored worse on physical, social, and emotional function. Compared to patients with ulcerative colitis, patients with CD scored similar to or worse on general HQOL measures. [40] Estimated direct costs of CD in 1998 (including inpatient and outpatient visits and pharmacy) were $708 million; indirect costs were estimated at $75 million and included costs associated lost productivity. [41] People with RA are at risk for falls and other injuries and may be unable to work. [37] One estimate of health care costs attributable to RA in 2013 was $140 billion. [38] The National Health Interview Survey (NHIS) reported that 43.5% of individuals with RA had limitations to activity related to their RA. [39] In the Global Burden of Disease 2010 study, RA was ranked as the 42nd highest contributor to global disability, and YLDs were 55 per 100,000 population (in 2010), and higher in females than in males (87.84 per 100,000 versus 22.46 per 100,000). [21]; the Global Burden of Disease 2017 study reported YLDs at 2,620,000. [14]

If CD is a causal factor in RA, then these impacts are magnified. To date, it has not been established whether CD is a causal factor in developing RA. If it is determined that CD is a cause of RA, then the importance of adequately treating CD to prevent the additional pain, disability, and costs associated with RA is clear.

Mendelian Randomization (MR) is a method that facilitates causal inference from observational studies. This study design bypasses confounding, which is a common problem in epidemiologic studies, [11] by taking advantage of the random allocation of genetic material at conception. MR compares disease status by genetically predicted exposure to obtain unconfounded estimates of association. The method is supported by available genome wide association studies (GWAS) genotyping millions of single nucleotide polymorphisms (SNPs). Other study designs which take advantage of genetics to avoid confounding, including sibling or twin studies, are not always feasible, so the increasing availability of genetic studies has supported the growth of the MR method.

Few MR studies have evaluated causes of CD. A review of the literature identified only five abstracts, which looked at IL-6 [46-48] and the *GSDMB* gene [49, 50] as potential causal factors of CD. On the other hand, a search for MR studies and RA yielded 61 results, many of which have evaluated various potential causes of RA. For example, decreased levels of IL-6 may be linked to an increased risk of RA. [41, 42] One study found no effect of vitamin-D associated SNPs on risk of RA. Another found one vitamin D SNP associated with a poorer response to TNF-alpha inhibitor therapy in RA patients. [43, 44] An MR study found an inverse effect of linoleic acid and RA. [45]

No published MR studies were found that looked at these two IMIDs as potentially causally related. Given the limited data available on the causal relationship between CD and RA, and the public health importance with respect to causality, this study seeks to address the question of whether CD causes RA using an MR study.

METHODS

MR as instrumental variable analysis

MR is an instrumental variable analysis with genetic variants as the instrument variable (IV). An instrumental variable is used to assess the effect of an exposure on an outcome when the exposure is open to confounding. Three conditions are necessary to satisfy an instrumental variable. Relevance means that the chosen instrument is associated with the exposure. Exclusion-restriction means that the instrument's effect on the outcome work only through its effects on the exposure. Independence means a lack of confounding between the instrument and the outcome. [60-62]

Data sources

Genetic predictors of Crohn's disease

The International IBD Genetics Consortium (IIBDGC) is a large, curated GWAS of CD. This GWAS combines data from six CD GWAS, with 17,897 cases and 33,977 controls. All GWAS participants are of European descent. Participants were enrolled in medical centers in the U.S. and Canada, and Europe. Disease ascertainment was by physicians using typical diagnostic tools (e.g., x-rays, endoscopy reports, pathology reports). Geographically matched controls were included, as were age- and ethnicity-match controls from the New York Health Project. [67-71] Comprehensive demographic data for all component GWAS were not found, but one component

GWAS (Welcome Trust Case Control Consortium) reported median age of participants as 45.7 years among the WTCCC panel and 43.9 years among the replication panel, and 60% of participants were <age 50; however these participants represent only a small proportion of the total participants in the IIBDGC. [67, 90]

Genetic associations with RA

The data source for RA included 22 combined GWAS, of which 18 studies included participants of European background and 4 studies included those of Asian background. RA cases were physician diagnosed according to the 1987 American College of Rheumatology criteria or similar criteria as ascertained by a rheumatologist. The sample includes 19,234 cases and 61,565 controls. [65, 66]

Statistical analysis

All SNPs identified as predictors of CD with significance $p<5x10^{-8}$ were reviewed in PhenoScanner to ascertain potential pleiotropy. There are two types of pleiotropy: vertical and horizontal. Vertical pleiotropy occurs when a SNP has an effect on the outcome that has an effect downstream of the exposure. This in essence captures what MR does: it determines, with the SNP as the IV, whether an exposure has an effect on an outcome. Horizontal pleiotropy occurs when the SNP, as the IV, influences multiple diseases or traits independent of the exposure. [91] The PhenoScanner review was conducted to determine whether the SNPs predictive of CD are also associated with RA; this would suggest horizontal pleiotropy.

A two-sample MR analysis was conducted. SNPs to predict CD were selected with p-values below $5x10^{-8}$. Independent SNPs were selected as having r^2 (correlation coefficient) threshold for identifying linkage disequilibrium as <0.001, to obtain independent genetic variants predicting

the exposure.[78, 79] The F-statistic was calculated for each SNP and overall [76] to show the strength of the instrument (F>10 suggests a strong instrument). All SNPs identified as predictive of CD were searched using PhenoScanner [80], a comprehensive, curated genotype to phenotype cross-reference, to assess the potential for horizontal pleiotropy.

The primary analysis used the inverse variance weighted (IVW) method with multiplicative random effects. Sensitivity analyses included weighted median, MR-Egger, and MR-PRESSO methods. The inverse variance weighted method meta-analyzes Wald estimates for each SNP (ratio of SNP on outcome to SNP on exposure) [74] using the first-term of Fieller's theorem (variance of SNP on outcome divided by SNP on exposure) [75] to approximate the SNP-specific variance. This method assumes each included instrument fulfills the assumptions of an IV. Inclusion of pleiotropic SNPs may decrease the validity of this method.

Three additional methods were used as sensitivity analyses to determine whether the IV assumptions were violated or if outlier SNPs were driving an effect estimate. These sensitivity analyses use different assumptions. A weighted median of the SNP-specific Wald estimates is valid as long as >50% of the weight is contributed by valid instruments. [76] MR-Egger relaxes the assumption that there are no pleiotropic effects, but does not identify or correct for them. This method is robust to invalid instruments, but relies on the InSIDE assumption (Instrument Strength Independent of Direct Effect) [77], which states that the strength of the association between the IV (SNP) and the exposure should be independent of pleiotropic effects. This can be considered relaxing the exclusion-restriction IV assumption. [108] The MR-Egger intercept p-value was reported, because a non-zero intercept indicates that the IVW estimate may not be valid. The MR-Egger I^2 value was calculated to identify the percentage of total variation across the meta-analysis that is due to heterogeneity. [92] The third sensitivity analysis, MR-PRESSO

identifies and removes outliers due to horizontal pleiotropy and tests differences in the estimates before and after this correction. MR-PRESSO requires that >50% are valid instruments. Removing outliers in this way leads to an unbiased estimate. [72, 73]

The minimum effect size detectable with the current sample size, based on 80% power and significance 0.05, was calculated using code from Burgess, provided in Appendix 2. [93]

The MR-base platform [57] and R version 3.6.1 (2019-07-05) [56] were used in this analysis.

Protection of human subjects

Deidentified, publicly available summary data were used in this study. Institutional Review Board approval was not required.

RESULTS

Table 3.1 shows 121 independent SNPs identified as predictors of Crohn's disease (significance $p<5x10^{-8}$). These SNPs were reviewed in PhenoScanner [80] for potential pleiotropy. Horizontal pleiotropy is suggested, as 16 SNPs identified as causal for CD were also predictive of RA. Of these, 15 were among the 21 excluded in the MR-PRESSO sensitivity analysis. Most of the CD-associated SNPs were associated with multiple other traits. Many were associated with other autoimmune diseases (e.g., allergic and celiac diseases, psoriasis, type 1 diabetes mellitus) as well as an assortment of other diseases and traits (e.g., height, body mass index, and various blood cell counts). However, these factors are not though to cause RA.

The inverse variance weighted analysis and all sensitivity analyses were null (IVW OR 0.97, 95% CI 0.88-1.07, p=0.59; weighted median OR 0.98, 95% CI 0.94-1.03, p=0.49; MR-Egger OR 0.92, 95% CI 0.67-1.18, p=0.54; MR-PRESSO OR 1.00; 95% CI 0.64-1.36, p=0.90). The MR-

Egger intercept was not significant. The MR-PRESSO analysis identified and eliminated 21 outliers out of 121 SNPs.

Table 3.2 summarizes the results of the four analyses.

The F-statistic was calculated by dividing the squared beta of the exposure by the squared standard error of the exposure; the F-statistic average was 90 and all F-statistics for individual SNPs were >10.

A power calculation was performed using R [47] to determine the minimum effect size detectable given the available sample size. Assumptions in this calculation were as follows: 80% power, significance (alpha)=0.05, and the ratio of cases to controls (19,234 and 61,565, respectively) is 0.312. The minimum effect detectable with the available sample size and 80% power was calculated to be 0.05, corresponding to an odds ratio of 1.05. The R code for this calculation is provided in **Appendix 2**.

DISCUSSION

The review of 121 SNPs predictive of CD in PhenoScanner [80] suggested some amount of horizontal pleiotropy, as 16 SNPs identified as causal for CD were also predictive of RA. However, 15 of these 16 pleiotropic SNPs were removed in the MR-PRESSO analysis, yet the null result persisted. The assumption of >50% valid instruments was met for all three sensitivity analyses, based on these MR-PRESSO findings. The F-statistic suggested strong instruments. The observed effect sizes were below the minimum effect size detectable based on sample sizes.

Selection bias

Selection bias may be a factor that has prevented this MR study from identifying a causal relationship between CD and RA. CD is has two peak age ranges for diagnosis, and the first is

several decades younger than the second, which is closer to the average age of onset for RA. [17, 18, 24] The average age of all participants in the full CD GWAS was not found, but one of the component GWAS reported that 60% of participants were under age 50. [67] Therefore, is possible that patients were enrolled in the CD GWAS prior to an age of RA diagnosis. This would suggest that RA could not be detected in the analysis because there were too few GWAS participants with an RA diagnosis simply due to their age. A different dataset with an age distribution skewed older may have yielded a different result.

Non-RA extra-intestinal manifestations of CD

This MR study suggests that CD does not cause RA. While CD patients commonly experience joint pain and arthropathies, these extra-intestinal manifestations of Crohn's disease differ from RA. The arthritis experienced by CD patients is primarily peripheral oligoarthritis, affecting fewer than 5 joints and most commonly affecting large joints such as hips, knees, wrists, ankles, and elbows. About 1-6% of patients develop ankylosing spondylitis, which usually affects sacroiliac joints and spine and may lead to spinal fusion. Peripheral arthritis differs from rheumatoid arthritis; RA is progressive and deforming, ultimately destroying joints, while arthritis associated with CD does not cause long-term joint damage. [109] Peripheral arthritis is commonly described in CD patients; studies suggest it may occur in nearly a third of CD patients, though estimates were lower in Norway. [110-112] Onset of arthritic symptoms may precede bowel symptoms in patients with CD. Prognosis is generally good in CD patients due to the non-progressive/non-erosive nature of this type of arthritis. [113]

Shared etiology

Different clinical presentation of CD and RA does not preclude a shared etiology. A number of issues point toward shared etiology: familial clustering, co-occurrence of IMIDs, and efficacy of

treatments across these diseases. GWAS, including those used in the current MR study, show some overlap in SNPs predicting both CD and RA, which also suggests a shared pathogenesis. While each SNP may have a small effect individually, each SNP disrupts the immune system to alter risk of disease without individually causing disease alone. [114]

TNF-alpha

Dysregulation of the immune system caused by inflammatory cytokines including TNF-a has been shown to cause CD and RA as well as other IMIDs. [10] TNF is a cytokine that regulates inflammatory reactions and immune functions by controlling cell proliferation, differentiation, and cell death, and is important in the pathogenesis and pathology of both RA and CD. [10] Preclinical studies in RA have shown that the concentration of TNF at inflammation sites is associated with disease activity, which suggested the early hypothesis that removing excess TNF would provide clinical benefit, [115, 116] and further preclinical studies identified the benefit of TNF-a inhibition specifically. [117, 118] Subsequent regulatory approval of TNF-alpha inhibitors such as etanercept, infliximab, and adalimumab to treat RA [31, 54, 105] have supported the clinical therapeutic benefit of TNF-a inhibition in many of these patients. In CD, high TNF levels are seen in inflamed mucosa of patients, [119] and animal studies have shown TNF levels to correspond to disease activity. [120, 121] As with RA, inhibition of TNF can reduce disease activity or severity, [122] and approval of two of the TNF-a inhibitors, infliximab and adalimumab, [31, 105] support the clinical benefit of inhibition of TNF-a in CD patients.

Interleukin-6

IL-6 is a cytokine that induces protein synthesis in hepatocytes, and increased serum IL-6 has been seen in acute conditions including surgery, burns, and infections, as well as in chronic conditions, such as CD and RA. Serum IL-6 concentrations were found to be higher in CD

patients than in healthy controls, and a nonsignificant trend was found toward higher serum IL-6 in CD patients with active disease versus inactive disease. [102] In RA patients, the IL-6 -174G/C variant has been associated with RA in both European and Asian populations. [41] IL-6 promotes joint inflammation through vascular endothelial growth factor (VEGF), levels of which are associated with RA disease activity. IL-6 is found in synovial fluid and serum of RA patients, with levels corresponding to disease activity and joint damage. Serum IL-6 levels were elevated in RA patients compared to controls, and higher levels were associated with more severe disease activity as measured by the DAS28. [123]

Co-occurrence of disease

While as many as 9.1-9.8% of patients with one IMID may also have a second IMID, [8] CD and RA co-occur far less frequently than that. An increased risk of RA in CD patients has been reported, which supports the idea of shared etiology, though co-occurrence of CD and RA is still low. The increased risk of RA in CD patients reported in the Danish National Patient Registry (OR 2.1; 95% CI 1.6-2.8) and the co-occurrence of CD and RA was seen in 0.0027% of patients. The authors suggested that the microbiome of the gut may have triggered RA, and they suggest a shared causal role of T-helper cells TH1 and TH17 in both CD and RA. [87, 101] The increased risk of RA in CD patients reported in a US-based commercial insurance claims database (OR 1.9, 95% CI 1.5–2.3; RA seen in 1.7% of CD patients) was attributed to possible common genetic or environmental causes, but the mechanism was not described. [88] Another study using US claims data also found increased risk of RA in CD in both databases (in MarketScan, OR 2.37, 95% CI 2.02-2.77; in IMS Health, OR 2.76, 95% CI 2.40-3.18), and the authors suggest a shared underlying pathogenic mechanism due to immune system dysfunction, without providing additional clarification on that mechanism. There are, however, biases inherent in US-based

commercial insurance database studies that could suggest underreporting of RA and CD. The participants are generally healthier than the general population (healthy enough to work). The age distribution of participants in these plans may be younger than the general population, since commercial insurance plans are generally tied to employment; older patients, typically covered by Medicare rather than commercial insurance, are generally more likely to have an RA diagnosis. Therefore, commercial insurance claims data may miss some overlap between RA and CD populations. [9]

Strengths and limitations

There are limitations to the MR method, which may have impacted the validity of these results. First, the method relies on strong assumptions ; for example, that the genetic instruments strongly predict the exposure [83, 84] However, all 121 instruments were strongly associated with the exposure, based on F-statistics well over 10 (individually and overall). Second, confounders of genetic variants on exposure may exist, typically as a result of population stratification; therefore, a homogeneous population is generally necessary for MR studies, with correction for population stratification, [83, 84] as used here. Third, with regard to the exclusion-restriction assumption, review of the SNPs in PhenoScanner suggested some horizontal pleiotropy. However, estimates remained nonsignificant after removing outliers statistically (in the MR-PRESSO analysis). The low MR-Egger I^2 value of 5.42% suggests low heterogeneity across the SNPs included in the MR-Egger analysis. Fourth, effect sizes, even if nonsignificant, may not correspond to clinical significance, nor indicate effects of a clinical intervention at a future time, either due to the importance of cumulative exposure or time-dependent exposure. Even with a nonsignificant effect size, these two IMIDs do co-occur, if infrequently, and this can be clinically significant. [64] Fifth, CD and RA are more common in women than men, but sex-

specific analysis is not publicly available. Sixth, MR studies usually require extremely large sample sizes. The GWAS used in this study were both modestly large (17,897 cases and 33,977 controls for CD, 19234 cases and 61565 controls for RA); in order for the effect size to reach statistical significance, a sample size of nearly 300,000 (over 3.5 times what was available) would have been necessary; this would represent about 20% of people with RA in the U.S. [93] (Calculation provided in **Appendix 2**). Finally, an important current limitation of all observational studies is that they can only address causal factors of disease incidence, but not issues related to disease progression; studies of disease progression (and any studies of patients) are open to selection bias because they are missing those who have already died because of the exposure or because of competing risk of the outcome, [83, 84] of if participation in the exposure GWAS takes place before development of the outcome can occur.

CONCLUSION

This Mendelian randomization study suggests that there is not a direct causal link between Crohn's Disease and rheumatoid arthritis. Other causal pathways were proffered, including shared causal partners including TNF-alpha and IL-6 overexpression. It is also possible that no causal association was seen due to selection bias, due to the expected age distribution of the exposure GWAS.

KEY MESSAGES

- Mendelian randomization is a powerful method that can determine causal relationships between exposures and outcomes, if appropriate instruments can be identified.
- Although some IMIDs co-occur, CD does not cause RA; co-occurrence of CD and RA may be due to common causes.

- It is also possible that selection bias related to the expected age distribution of the exposure GWAS prevented identification of a causal role of CD on RA.

REFERENCES

1. Lichtenstein, G.R., et al., ACG Clinical Guideline: Management of Crohn's Disease in Adults. Am J Gastroenterol, 2018. **113**(4): p. 481-517.
2. Quezada, S.M., E.K. Steinberger, and R.K. Cross, Association of age at diagnosis and Crohn's disease phenotype. Age Ageing, 2013. **42**(1): p. 102-6.
3. Ganz, M.L., et al., The Economic and Health-related Impact of Crohn's Disease in the United States: Evidence from a Nationally Representative Survey. Inflamm Bowel Dis, 2016. **22**(5): p. 1032-41.
4. Duerr, R.H., et al., A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science, 2006. **314**(5804): p. 1461-3.
5. Rioux, J.D., et al., Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet, 2007. **39**(5): p. 596-604.
6. **https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html.** . January 9, 2019]; Available from: https://www.cdc.gov/arthritis/basics/rheumatoid-arthritis.html. .
7. Goemaere, S., et al., Onset of symptoms of rheumatoid arthritis in relation to age, sex and menopausal transition. J Rheumatol, 1990. **17**(12): p. 1620-2.
8. Innala, L., et al., Age at onset determines severity and choice of treatment in early rheumatoid arthritis: a prospective study. Arthritis Res Ther, 2014. **16**(2): p. R94.
9. Myasoedova, E., et al., Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. Arthritis Rheum, 2010. **62**(6): p. 1576-82.
10. Cohen, R., et al., Autoimmune disease concomitance among inflammatory bowel disease patients in the United States, 2001-2002. Inflamm Bowel Dis, 2008. **14**(6): p. 738-43.
11. Halling, M.L., et al., Patients with inflammatory bowel disease have increased risk of autoimmune and inflammatory diseases. World J Gastroenterol, 2017. **23**(33): p. 6137-6146.
12. Eaton, W.W., et al., Epidemiology of autoimmune diseases in Denmark. J Autoimmun, 2007. **29**(1): p. 1-9.
13. Weng, X., et al., Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern california-managed care organization. Am J Gastroenterol, 2007. **102**(7): p. 1429-35.
14. Blandizzi, C., et al., The role of tumour necrosis factor in the pathogenesis of immune-mediated diseases. Int J Immunopathol Pharmacol, 2014. **27**(1 Suppl): p. 1-10.
15. Li, B., et al., Circulating interleukin-6 and rheumatoid arthritis: A Mendelian randomization meta-analysis. Medicine (Baltimore), 2016. **95**(23): p. e3855.
16. Gross, V., et al., Evidence for continuous stimulation of interleukin-6 production in Crohn's disease. Gastroenterology, 1992. **102**(2): p. 514-9.
17. Srirangan, S. and E.H. Choy, The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. Ther Adv Musculoskelet Dis, 2010. **2**(5): p. 247-56.
18. Gheita, T.A., et al., Involvement of IL-23 in enteropathic arthritis patients with inflammatory bowel disease: preliminary results. Clin Rheumatol, 2014. **33**(5): p. 713-7.
19. Rothman, K.J., Causes. Am J Epidemiol, 1976. **104**(6): p. 587-92.
20. Enbrel prescribing information. November 2, 2019]; ≈]. Available from: https://www.enbrel.com/how-enbrel-

works?isipaid=true&gclsrc=aw.ds&&gclid=CjwKCAjw0vTtBRBREiwA3URt7kkmPen8DLOORE5kjSt9wpvycsKUURcqExQdbwBjffRRZZa_FJFxFRoCYOsQAvD_BwE.
21. Remicade prescribing information. March 21, 2020]; Available from: http://www.janssenlabels.com/package-insert/product-monograph/prescribing-information/REMICADE-pi.pdf.
22. Humira prescribing information. December 13, 2019]; Available from: https://www.rxabbvie.com/pdf/humira.pdf.
23. Actemra prescribing information. December 23, 2019]; Available from: https://www.gene.com/download/pdf/actemra_prescribing.pdf.
24. Kevzara prescribing information.
25. Lawlor, D.A., et al., Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med, 2008. **27**(8): p. 1133-63.
26. Parisinos C, S.S., Patel R, et al, OTU-006 The interleukin-6 receptor as a drug target in inflammatory bowel disease; a mendelian randomisation study. Gut, 2018. **67**: p. A55.
27. Parisinos C, S.S., Katsoulis M, et al, The interleukin-6 receptor as a drug target in inflammatory bowel disease; a mendelian randomization study. . United Eruopean Gastroenterology Journal., 2018. **6**(A42-3).
28. Parisinos CA, S.S., Katsoulis M, et al., The interleukin-6 receptor as a target for prevention of Crohn's disease and ulcerative colitis; A Mendelian randomisation study. Journal of Crohn's and Colitis. Conference: 13th Congress of European Crohn's and Colitis Organisation, ECCO 2018. Austria. 12 (Supplement 1) (pp S530-S531), 2018. Date of Publication: February 2018. . Journal of Crohn's & Colitis, 2018. **12**(S530-1).
29. Porcu E, L.K., Reuger S, et al., Mendelian randomization integrating GWAS and EQTL data reveals genetic determinants of complex and clinical traits. . Human Heredity, 2018. **83**(242-3).
30. Porcu E, R.S., Santoni FA, et al., Mendelian randomization combining GWAS and eQTL data reveals new loci, extensive pleiotropy and genetic determinants of complex and clinical traits. Human Heredity. . Human Heredity, 2018. **83**: p. 20.
31. Interleukin 1 Genetics, C., Cardiometabolic effects of genetic upregulation of the interleukin 1 receptor antagonist: a Mendelian randomisation analysis. Lancet Diabetes Endocrinol, 2015. **3**(4): p. 243-53.
32. Viatte, S., et al., The role of genetic polymorphisms regulating vitamin D levels in rheumatoid arthritis outcome: a Mendelian randomisation approach. Ann Rheum Dis, 2014. **73**(7): p. 1430-3.
33. Yarwood, A., et al., Testing the role of vitamin D in response to antitumour necrosis factor alpha therapy in a UK cohort: a Mendelian randomisation approach. Ann Rheum Dis, 2014. **73**(5): p. 938-40.
34. Zhao, J.V. and C.M. Schooling, Role of linoleic acid in autoimmune disorders: a Mendelian randomisation study. Ann Rheum Dis, 2019. **78**(5): p. 711-713.
35. Lousdal, M.L., An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol, 2018. **15**: p. 1.
36. Baser, O., Too much ado about instrumental variable approach: is the cure worse than the disease? Value Health, 2009. **12**(8): p. 1201-9.
37. Swanson, S.A. and M.A. Hernan, Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology, 2013. **24**(3): p. 370-4.

38. Wellcome Trust Case Control, C., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. **447**(7145): p. 661-78.
39. Liu, J.Z. and C.A. Anderson, Genetic studies of Crohn's disease: past, present and future. Best Pract Res Clin Gastroenterol, 2014. **28**(3): p. 373-86.
40. Franke, A., et al., Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet, 2010. **42**(12): p. 1118-25.
41. Jostins, L., et al., Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature, 2012. **491**(7422): p. 119-24.
42. Barrett, J.C., et al., Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet, 2008. **40**(8): p. 955-62.
43. Parkes, M., et al., Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet, 2007. **39**(7): p. 830-2.
44. Okada, Y., et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature, 2014. **506**(7488): p. 376-81.
45. Arnett, F.C., et al., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum, 1988. **31**(3): p. 315-24.
46. Jordan, D.M., M. Verbanck, and R. Do, HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. Genome Biol, 2019. **20**(1): p. 222.
47. VanLiere, J.M. and N.A. Rosenberg, Mathematical properties of the r2 measure of linkage disequilibrium. Theor Popul Biol, 2008. **74**(1): p. 130-7.
48. Hill, W.G. and A. Robertson, Linkage disequilibrium in finite populations. Theor Appl Genet, 1968. **38**(6): p. 226-31.
49. Bowden, J., et al., Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol, 2016. **40**(4): p. 304-14.
50. Phenoscanner. December, 4, 2019]; Available from: http://www.phenoscanner.medschl.cam.ac.uk.
51. Wald, A., The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics, 1940. **11**: p. 284-300.
52. Fieller, E.C., Some problems in interval estimation. Journal of the Royal Statistical Society, Series B, 1954. **16**(2): p. 175-85.
53. Burgess, S. and S.G. Thompson, Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol, 2017. **32**(5): p. 377-389.
54. Bowden, J., G. Davey Smith, and S. Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol, 2015. **44**(2): p. 512-25.
55. Higgins, J.P., et al., Measuring inconsistency in meta-analyses. BMJ, 2003. **327**(7414): p. 557-60.
56. Verbanck, M., et al., Publisher Correction: Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet, 2018. **50**(8): p. 1196.
57. Verbanck, M., et al., Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet, 2018. **50**(5): p. 693-698.

58. Burgess, S., Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol, 2014. **43**(3): p. 922-9.
59. Hemani, G., et al., The MR-Base platform supports systematic causal inference across the human phenome. Elife, 2018. **7**.
60. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA
61. Orchard, T.R., Management of arthritis in patients with inflammatory bowel disease. Gastroenterol Hepatol (N Y), 2012. **8**(5): p. 327-9.
62. Vavricka, S.R., et al., Frequency and risk factors for extraintestinal manifestations in the Swiss inflammatory bowel disease cohort. Am J Gastroenterol, 2011. **106**(1): p. 110-9.
63. Zippi, M., et al., Extraintestinal manifestations in a large series of Italian inflammatory bowel disease patients. World J Gastroenterol, 2014. **20**(46): p. 17463-7.
64. Hsu, Y.C., et al., Gastrointestinal complications and extraintestinal manifestations of inflammatory bowel disease in Taiwan: A population-based study. J Chin Med Assoc, 2017. **80**(2): p. 56-62.
65. Harbord, M., et al., The First European Evidence-based Consensus on Extra-intestinal Manifestations in Inflammatory Bowel Disease. J Crohns Colitis, 2016. **10**(3): p. 239-54.
66. Cotsapas, C. and D.A. Hafler, Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol, 2013. **34**(1): p. 22-6.
67. Brennan, F.M., et al., Inhibitory effect of TNF alpha antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. Lancet, 1989. **2**(8657): p. 244-7.
68. Knight, D.M., et al., Construction and initial characterization of a mouse-human chimeric anti-TNF antibody. Mol Immunol, 1993. **30**(16): p. 1443-53.
69. Thorbecke, G.J., et al., Involvement of endogenous tumor necrosis factor alpha and transforming growth factor beta during induction of collagen type II arthritis in mice. Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7375-9.
70. Williams, R.O., M. Feldmann, and R.N. Maini, Anti-tumor necrosis factor ameliorates joint disease in murine collagen-induced arthritis. Proc Natl Acad Sci U S A, 1992. **89**(20): p. 9784-8.
71. Papadakis, K.A. and S.R. Targan, Tumor necrosis factor: biology and therapeutic inhibitors. Gastroenterology, 2000. **119**(4): p. 1148-57.
72. Corazza, N., et al., Transmembrane tumor necrosis factor is a potent inducer of colitis even in the absence of its secreted form. Gastroenterology, 2004. **127**(3): p. 816-25.
73. Nakai, M., et al., The role of the tumor necrosis factor receptor in 2,4,6-trinitrobenzene sulfonic acid (TNBS)-induced colitis in mice. Dig Dis Sci, 2005. **50**(9): p. 1669-76.
74. Tilg, H., A. Moschen, and A. Kaser, Mode of function of biological anti-TNF agents in the treatment of inflammatory bowel diseases. Expert Opin Biol Ther, 2007. **7**(7): p. 1051-9.
75. Beyazal, M.S., Devrimsel, G., Cure, M.C., et al., Serum Levels of IL-17, IL-6, TNF-α and Disease Activity in Patients with Rheumatoid Arthritis. Akt Rheumatol, 2017. **42**: p. 323-8.
76. Robinson, D., Jr., et al., Co-occurrence and comorbidities in patients with immune-mediated inflammatory disorders: an exploration using US healthcare claims data, 2001-2002. Curr Med Res Opin, 2006. **22**(5): p. 989-1000.

77. Schooling, C.M., G. Freeman, and B.J. Cowling, Mendelian randomization and estimation of treatment efficacy for chronic diseases. Am J Epidemiol, 2013. **177**(10): p. 1128-33.
78. Schooling, C.M.L., P.; Au Yeung, S. L., et. al., Survival bias and competing risk can severely bias Mendelian Randomization studies of specific conditions. https://www.biorxiv.org/content/biorxiv/early/2019/10/21/716621.full.pdf, 2019.
79. Walker, V.M., et al., Using the MR-Base platform to investigate risk factors and drug targets for thousands of phenotypes. Wellcome Open Res, 2019. **4**: p. 113.

Table 3.1. 121 SNPs identified in MR-base as associated with CD ($p<5x10^{-8}$), and their location as determined in PhenoScanner. Red text indicates possible pleiotropy (SNPs also associated with RA); yellow highlighting indicates 21 SNPs eliminated in MR-PRESSO analysis as possible outliers. [80]

SNP	***Gene***
rs10758669	*JAK2*
rs10798069	*PLA2G4A*
rs10800309	*FCGR2A*
rs10878302	*LRRK2*
rs10956252	*RP11-136O12.2*
rs10995271	*ZNF365*
rs11117431	*RP11-542M13.2*
rs11152949	*RP11-282C5.1*
rs11159833	*GPR65*
rs11167518	*not found*
rs11185982	*SH2D4B*
rs11236797	*RP11-672A2.7*
rs11691685	*TEX41*
rs11713774	*SATB1-AS1*
rs11768997	*not found*
rs11793497	*SNAPC4*
rs12411259	*RP1-15D23.2*
rs1250573	*ZMIZ1*
rs1267501	*RP11-146I2.1*
rs12694846	*SP140*
rs12796489	*CARS*
rs1292053	*TUBD1*
rs12949918	*STAT3*

rs1297258	*AJ006998.2*
rs13001325	*IL1RL1*
rs13407913	*ADCY3*
rs1363907	*ERAP2*
rs140143	*not found*
rs144004051	*NOS2*
rs1456896	*AC020743.3*
rs1517352	*STAT4*
rs1569328	*FOS*
rs1646019	*RMI2*
rs17129991	*IL12RB2*
rs17293632	*SMAD3*
rs17388425	*AC008697.1*
rs17391694	*RNFT1P2*
rs17622378	*C5orf56*
rs17694108	*SLC7A10*
rs181826	*NDFIP1*
rs1847472	*BACH2*
rs2024092	*SBNO2*
rs212388	*RP1-111C20.3*
rs2153283	*IPMK*
rs2227551	*PLAU*
rs2270395	*LINC02168*
rs2284553	*IFNGR2*
rs2395022	*SMURF1*
rs2413583	*AL031590.1*
rs2538470	*RP5-958B11.2*
rs259964	*ZNF831*

rs2641348	*ADAM30*
rs26528	*IL27*
rs2847293	*PTPN2*
rs28999107	*LTBR*
rs2974935	*MTX1*
rs3024505	*IL10*
rs303429	*MAP3K8*
rs3129871	*HLA-DRA*
rs3184504	*SH2B3*
rs3197999	*MST1*
rs34592089	*not found*
rs34779708	*CREM*
rs34787213	*CD6*
rs34804116	*RP11-60A8.1*
rs35164067	*CDC37*
rs35320439	*GAL3ST2*
rs35730213	*C1orf106*
rs36016881	*PARK7*
rs3776414	*DAP*
rs3801810	*SKAP2*
rs3853824	*C17orf67*
rs438475	*NOTCH4*
rs4703855	*JMY*
rs4795397	*IKZF3*
rs516246	*FUT2*
rs559928	*RPS6KA4*
rs56163845	*CPEB4*
rs6062496	*TNFRSF6B*

rs6074022	*LINC01754*
rs6111031	*SIRPB3P*
rs61839660	*IL2RA*
rs640466	*LSM14A*
rs6456426	*ZFP57*
rs6500315	*RP11-21B23.3*
rs6561151	*LINC00284*
rs6651252	*LINC00824*
rs6679677	*PHTF1*
rs6702421	*DENND1B*
rs6738394	*ARPC2*
rs6738490	*ATG16L1*
rs6740462	*AC074391.1*
rs6827756	*KIAA1109*
rs6908425	*CDKAL1*
rs7015630	*SCb-64M4.1*
rs7085798	*NKX2-3*
rs71624119	*ANKRD55*
rs7194886	*NOD2*
rs7236492	*NFATC1*
rs72727394	*RASGRP1*
rs727563	*ACO2*
rs7438704	*SLAIN2*
rs7517847	*IL23R*
rs7608910	*PUS10*
rs76906269	*LRRK2*
rs7711427	*RP11-386E5.1*
rs7773324	*RP11-157J24.2*

rs7786444	*JAZF1*
rs780094	*GCKR*
rs7848647	*TNFSF15*
rs7969592	*IFNG-AS1*
rs79980175	*TTC33*
rs8127691	*not reported*
rs915286	*LINC00598*
rs9264942	*HLA-B*
rs9457247	*RP1-167A14.2*
rs9491892	*RP11-394G3.2*
rs9494844	*AL356739.1*
rs9554587	*UBAC2*
rs9594766	*RP11-413N19.2*
rs9889296	*AC005549.3*

Table 3.2. Mendelian randomization association of CD on RA, based on 113 independent SNPs* (p-values all <5x10^{-8}).

Method	OR	95% CI	**pval**	Cochran's Q-statistic (p-value)	MR-Egger intercept p-value	MR-Egger I2
MR Egger	0.92	0.67-1.18	0.54	2048.35 (0)	0.65	5.42%
Weighted median	0.98	0.94-1.03	0.49			
Inverse variance weighted	0.97	0.88-1.07	0.59	2052.11 (0)		
MR-PRESSO	1.00	0.64-1.36	0.90			

*Palindromic SNPs were excluded from analyses via MR-base web-based application. [57]

Chapter 4: A Phenome-Wide Association Study to Assess Potential New Targets for Tocilizumab, an Interleukin-6 Inhibitor

ABSTRACT

Background

A recent use of Mendelian randomization (MR) studies and Phenome-Wide Association Studies (PheWAS) involves using genetic variants that predict response to treatment to identify both on-target and off-target mechanistic effects of a pharmacologic intervention. The goal of drug repurposing is to identify new therapeutic areas as potential new targets for existing treatments. Interleukin-6 is a cytokine that is implicated as a cause of rheumatoid arthritis (RA), and tocilizumab (TCZ) is a therapy targeting IL-6 inhibition that is approved for treatment of RA. The goal of the current study is to use MR to explore the scope for repurposing TCZ to characterize possible additional uses of and off-target effects of this treatment.

Methods

One single nucleotide polymorphism (SNP) predicting response to TCZ in RA patients was identified, rs7529229. The PheWAS (Phenome-wide Association Study) function in the MR-base web-based application was used to identify traits associated with this SNP to use as potential drug targets for inclusion in the analysis. As this is an agnostic search, a p-value threshold of 2.4×10^{-6} was used based on a Bonferroni correction. The PheWAS provided effect estimates (provided as beta) and p-values which enabled evaluation of potential new targets for TCZ.

Results

The SNP rs7529229 predicting response to TCZ in RA was associated with 21,031 traits. Seventeen of these met the Bonferroni corrected p-value threshold for statistical significance. Of these, three were excluded from consideration: abdominal aortic aneurysm, due to missing associations, and two traits described as “Blood clot DVT bronchitis emphysema asthma rhinitis eczema allergy diagnosed by doctor: None of the above,” as these did not indicate a trait. Among the other 10 unique traits, four showed an inverse association with the SNP predicting response to TCZ: RA, coronary heart disease (CHD), and two blood counts (red cell distribution width and granulocyte percentage of myeloid white cells). Six traits were positively associated with this SNP, meaning use of TCZ may increase the risk of the following: eczema, asthma, a less specific trait of hay fever, allergic rhinitis or eczema, tonsillectomy, and two additional blood counts (mean corpuscular hemoglobin and monocyte percentage of white cells).

Conclusions

As rs7529229 predicts response to TCZ in RA patients, it was expected that there would be an inverse association between the SNP and RA. Other studies have already evaluated the inverse association between this SNP and CHD, suggesting CHD as a potential target for TCZ. The positive associations with asthma was somewhat unexpected, as some studies have suggested a potential for IL-6 inhibition in asthma. The positive associations with eczema were expected as this was reported in the clinical trial program for TCZ and in other PheWAS of SNPs associated with IL-6 levels. The impact of TCZ on various blood counts are difficult to interpret. The direction of the association with asthma is surprising, even though a recent PheWAS reported a similar result. No new targets for TCZ were identified in this PheWAS.

Keywords

Phenome-wide association study; drug target validation; drug target repurposing; tocilizumab; interleukin-6; rheumatoid arthritis, asthma, Crohn's disease

INTRODUCTION

Mendelian Randomization (MR) is an important tool that facilitates causal inference from observational studies. This study design can allow for avoidance of a common problem in epidemiologic studies, confounding, [11] because it takes advantage of the random allocation of genetic make-up at conception. MR compares disease status by genetically predicted exposure to obtain unconfounded estimates. MR implementation has been facilitated by the increasing availability of GWAS genotyping millions of single nucleotide polymorphisms (SNPs). A recent extension of MR is to use genetic variants that predict response to treatment both as validation of efficacy of the treatment in the target population and to identify appropriate patient groups as new potential targets for specific existing treatments. [12] Drug target validation and repurposing studies can use these genetic variants as proxies to identify both on-target and off-target mechanistic effects of a pharmacologic intervention. As an example, a GWAS has identified a SNP, rs7529229, that predicts response to tocilizumab (TCZ), an interleukin-6 (IL-6) inhibitor, in rheumatoid arthritis (RA) patients. [13] Drug repurposing seeks to identify new therapeutic areas as potential new targets for existing treatments.

The goal of the current study is to use MR to explore the scope for re-purposing TCZ, an existing and approved treatment for RA, as well as to characterize possible off-target effects of this treatment.

Interleukin-6 in RA

IL-6 is a cytokine that induces protein synthesis in hepatocytes, and increased serum IL-6 has been seen in acute conditions including surgery, burns, and infections, as well as in chronic conditions, such as RA. In RA patients, the IL-6 -174G/C variant has been associated with RA in both European and Asian populations. [41] IL-6 promotes joint inflammation through vascular

endothelial growth factor (VEGF), levels of which are associated with RA disease activity. IL-6 is found in synovial fluid and serum of RA patients, with levels corresponding to disease activity and joint damage. [89]

Tocilizumab

Tocilizumab (TCZ) is an IL-6 receptor antagonist marketed by Genentech under the brand name Actemra. [106] It is approved for the following indications: moderate to severe RA in adults with inadequate response to disease-modifying antirheumatic drugs (DMARDs); giant cell arteritis in adults; active polyarticular juvenile idiopathic arthritis (JIA) in children age 2 and older; active systematic JIA in children age 2 and older; chimeric antigen receptor T-cell induced severe or life-threatening cytokine release syndrome in adults and children age 2 and older. TCZ binds to IL-6 receptors sIL-6R and mIL-6R (soluble and membrane-bound) to inhibit signaling through these receptors. IL-6 is a pro-inflammatory cytokine which is produced by T-cells, B-cells, lymphocytes monocytes and fibroblasts. It is also produced by synovial and endothelial cells, which causes local IL-6 production in joints affected by RA. [106]

Drug target validation and identifying new targets

Genetic validation of potential drug treatments is increasingly used to facilitate drug development and anticipate side-effects. Mutations in the *PCSK9* gene have been found to predict familial hypercholesterolemia and coronary artery disease; leading to two drugs that target this gene: evolocumab (Repatha, marketed by Amgen) and alirocumab (Praluent, marketed by Regeneron). [29] Lovostatin, the first of the statins and approved for primary and secondary prevention of coronary artery disease, inhibits HMG-CoA reductase. Many other genetic differences that identify susceptibility to disease have been discovered using GWAS, creating many potential drug targets. [30] Schmidt and colleagues describe four genes (*HMGCR, PCSK9,*

NPC1L1, and *CETP*) with effects on lipids and/or coronary heart disease where genetic validation using Mendelian randomization has confirmed the findings of clinical trials. The authors also provide guidance on the "selection of genes encoding druggable proteins" for use in drug target validation MR studies. [32] Nelson and colleagues used the Informa Pharmaprojects database to identify 22,270 drugs with 1,824 genetic targets, creating 19,085 drug-target pairs covering 705 indications, including 651drug-target pairs for RA. [33] Tumor necrosis factor alpha (TNF-a) is implicated in both CD and RA; adalimumab (Humira, marketed by AbbVie) targets TNF-a. [31] Interleukin 6 is implicated as a cause of RA, and tocilizumab (Actemra, marketed by Genentech) is a therapy targeting IL-6 inhibition. [13, 106]

MR studies have become a method for identifying new target indications for approved drugs. An MR study confirmed that three SNPs associated with interleukin-18 (IL18) were associated with an increase in susceptibility for Crohn's disease (OR 1.15, 95% CI = 1.01–1.31, p-value = 0.041); the authors note that an anti-IL18 monoclonal antibody recently failed efficacy goals in a phase II trials for type 2 diabetes mellitus, but these MR results suggest another indication target. [124] Zheng and colleagues provide a summary of recent genetic target drug validation and repurposing MR studies with varying exposures and the outcome of CHD in most cases [13, 125-135]. However, MR has not been used extensively to investigate drug validation and repurposing for RA treatments.

Using MR to identify new target indications can be cost-effective and time-saving. The cost of developing new molecular entities (NMEs) is in the billions of dollars, and the process can take, on average, over 12 years. [29, 30] Less than 10% of drugs are eventually approved by the FDA or another health authority. [29] Drug targets with genetic associations with disease have a much higher likelihood of succeeding and receiving regulatory approval. [29, 33, 40] When causal

genetic variants can be identified through GWAS, this genetic evidence has been shown to increase drug approval by more than two-fold. [29] Using genetic evidence could decrease cost of drug development, due to fewer failed clinical trials. One estimate suggests that if the proportion of drugs in development with genetic targets were increased from 15% (the current proportion) to 50%, research and development costs would decrease by anywhere from 9% to 35%. [29]

PheWas

PheWAS is a method to assess genotype-phenotype associations consistently and comprehensively. The method looks for associations of specific SNPs with many potential clinical outcomes or phenotypes. It is a complement to the GWAS method. Whereas GWAS typically use a set of SNPs to identify associations of all genotypes with a single or small number of phenotypes, PheWAS examines the associations of specific genotypes with all phenotypes. PheWAS can be conducted by analyzing genetic data from large, population-based cohort studies. The National Human Genome Research Institute (NHGRI) has created a repository for US-based cohorts called the Population Architecture using Genomics and Epidemiology (PAGE) network. [136] Other GWAS are available in the MR-base web-based platform. [57] The GWAS Catalog available on the MR-base platform is provided jointly by the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI), and can be found at https://www.ebi.ac.uk/gwas/. [137]

PheWAS can also be conducted using of electronic medical records (EMRs). The first large-scale PheWAS using EMRs looked through 3,144 SNPs and 1,358 phenotypes from EMRs of people of European descent, and the method confirmed 66% of previously identified associations

and suggested 63 new associations of outcomes including autoimmune, hematologic, neoplastic, circulatory, endocrine, and other diseases at a significance level of $p<5.6x10^{-6}$. [138] [34-36]

Studies have identified SNPs associated with soluble fraction of the IL6 receptor (sIL6R) and reviewed associated traits. PheWAS have evaluated traits associated with rs4129267, which is an sIL6R variant, and found positive, statistically significant associations with monocyte count, mean platelet volume, eczema, and asthma. Mendelian randomization using 34 SNPs associated with sIL6R circulating levels showed causal inverse associations with various types of stroke, atrial fibrillation, coronary artery disease, abdominal aortic aneurism, and RA, suggesting protective effects of IL-6 inhibition, and positive associations with eczema and asthma, suggesting IL-6 inhibition increases risk of these diseases. Also included in these analyses were heritable longevity; sIL6R was associated with parental age at death, and various blood parameters, of which only C-reactive protein was significantly, inversely associated with sIL6R levels. [51] Another PheWAS reviewed other IL6R SNPs (rs2228145; rs4129267), in patients in the Million Veteran Program (MVP) with results replicated in data from the UK Biobank and Vanderbilt University Biobank. Significant inverse associations between these SNPs and the following traits were found: aortic aneurysm, ischemic heart disease (including coronary atherosclerosis, ischemic heart disease, and myocardial infarction), and vascular diseases such as peripheral vascular disease and atherosclerosis of the extremities. Significant positive associations were identified for skin conditions such as atopic dermatitis (eczema), seborrheic dermatitis, and erythematosquamous dermatosis. Several pulmonary, renal, and eye conditions were also positively associated with these IL6R SNPs, as were two musculoskeletal conditions (acquired deformities of finger and gouty arthropathy). [52] While these two PheWAS identified

traits associated with SNPs that are IL-6 inhibitors, a search of the literature did not find any PheWAS that specifically evaluated traits associated with the specific SNP rs7529229.

METHODS

Data sources

rs7529229 as a predictor of response to TCZ

Forty studies from the US, UK, and Europe with 133,499 participants contributed to a biomarker study that evaluated IL-6, C-reactive protein (CRP), total cholesterol, HDL-cholesterol, LDL cholesterol, and fibrinogen. Seventeen of these 40 studies, with 29,978 participants, contributed IL-6 data. These data enabled identification of SNP rs7529229 as associated with increased log circulating IL-6 concentration, consistent with IL-6 inhibition from TCZ infusions for RA every four weeks (posology 4-8mg/kg). For each minor allele, the mean log IL-6 concentration increase was 9.45% (95% CI 8.34–10.57; $p=8{\cdot}41\times10^{-68}$). Mean ages of participants in the contributing studies ranged from 46 years to 75 years, with one outlier (mean age 26) and one study which did not report age. [13]

Genetic predictors of traits associated with rs7529229

A wide range of traits from the PheWAS were obtained from GWAS that are available in the MR-base web-based platform. [57] [137] Nearly all of the data for these traits came from the UK Biobank. The UK Biobank enrolled participants in the UK from 2006-2010. All participants were intended to be aged 40 to 69 years at enrollment. Participant data were captured in different ways; some by self-report, others by physical assessment. Genetic data are available for approximately 488,000 UK Biobank participants.

The CARDIoGRAMplusC4D GWAS provided data for CHD, one of the traits included in the PheWAS; this GWAS is a meta-analysis of multiple GWAS including participants of European and South Asian descent, with 15,420 CHD cases and 15,065 controls; [139]

RA data came from two sources. One was a meta-analysis of 11,475 cases, of whom 7,222 were ACPA positive, 3,297 were ACPA negative, and 957 were unassigned), and 15,870 controls in participants of European descent. [140] The other RA data source combined 22 GWAS of which 18 studies included participants of European background and 4 studies included those of Asian background. Among participants, 88.1% were seropositive for anti-citrullinated peptide antibody (ACPA), 9.3% were seronegative, and 2.6% had unknown antibody status. The sample includes 19234 cases and 61565 controls. [65-67]

Statistical analysis

The SNP of interest was searched using the PheWAS function in MR-base [57] to identify other potential drug targets for TCZ, based on traits noted as associated with this SNP. The Bonferroni correction was used to account for multiple comparisons. The traits were aligned so that the minor allele was C for all traits to ensure appropriate direction of association with the SNP among traits.

PhenoScanner [80] was searched to identify the effect allele and its effect allele frequency (EAF).

Protection of human subjects

Deidentified, publicly available summary data were used in this study. Institutional Review Board approval was not required.

RESULTS

The single SNP rs7529229 identified as a predictor of response to TCZ in RA [13] was reviewed in the MR-base PheWAS function [57] to determine other traits associated with the SNPs. This PheWAS identified 21,031 traits associated with this SNP. The Bonferroni correction yielded a p-value threshold of $2.4x10^{-6}$, which 17 of the 21,031 traits met.

PhenoScanner [80] showed that the effect allele is C and has an effect allele frequency (EAF) of 0.3738 in people of European descent. The 17 traits that were statistically significant following the Bonferroni correction were aligned on the minor allele (C) to allow for comparisons and to understand the direction of associations with rs7529229.

Among these 17 traits, three were excluded from further review as follows: abdominal aortic aneurism, due to missing estimate, which made the association uninterpretable, and two traits described as "Blood clot DVT bronchitis emphysema asthma rhinitis eczema allergy diagnosed by doctor: None of the above," as these did not indicate a trait.

Several of these traits appear in the PheWAS more than once due to the availability of multiple data sources in the GWAS catalog for some traits. [137] Therefore, these 14 listed traits actually represent 10 unique traits.

Four traits showed an inverse association with the SNP predicting response to TCZ, suggesting use of TCZ may lower risk of the trait: RA (p-values $1.1430x10^{-07}$ and $1.8000x10^{-08}$, from two different data sets); coronary heart disease (CHD; p-value $2.5x10^{-07}$), and two blood counts (red cell distribution width, p-value $4.1350x10^{-07}$ and granulocyte percentage of myeloid white cells, p-value $1.8330x10^{-06}$).

Six traits were positively associated with this SNP, suggesting use of TCZ may increase the risk of the following: asthma (p-values both $1.6x10^{-6}$ from two different data sets), eczema (p-values

$6.6x10^{-8}$ and $3.4x10^{-8}$, from two different data sets), a less specific trait described as hay fever/allergic rhinitis/eczema (p-values $2.2x10^{-14}$ and $3.4x10^{-13}$, from two different data sets), tonsillectomy (p-value $1.10x10^{-06}$), and two additional blood counts (mean corpuscular hemoglobin, p-value $3.95x10^{-07}$ and monocyte percentage of white cells, p-value $6.93x20^{-09}$).

This PheWAS can be found in **Table 4.1.**

DISCUSSION

These results are essentially unsurprising. An inverse association between rs7529229 and RA is to be expected, as RA is an approved indication for TCZ. [106] The inverse association between rs7529229 and CHD in particular has also been described, [13] as has this relation between other IL-6 inhibiting SNPs and CHD. [51, 52] The effects on blood counts, whether positive or negative, are difficult to interpret, and have been reported; [51, 52] additionally, changes to blood counts are described as potential adverse events in the TCZ product information. [106] Eczema has also been reported in the clinical trial program, [141] which reinforces the positive association between rs7529229 and eczema; this too has been reported in other IL-6 inhibitor PheWAS. [51, 52]

The positive association between rs7529229 and asthma is more interesting and possibly unexpected. A negative association would be more expected, because IL-6 has a plausible role in the pathogenesis of asthma. Lung epithelial cells produce IL-6 in response to respiratory viruses, allergens, or exercise. [142, 143] Circulating IL-6 is elevated in patients with asthma, [144] including in the bronchoalveolar lavage fluid (BALF) of asthmatic patients; [143, 145] this suggests that the role of IL-6 is more than just pro-inflammatory as this elevated level is seen in both allergic and intrinsic asthma. IL-6 is associated with both viral infections, which can lead to exacerbations of asthma, and obesity, which is associated with severity of asthma. [143, 146] IL-

6 levels have been found to be associated with lung function; forced expiratory volume in 1 second (FEV1) was shown to be inversely associated with IL-6 levels in patients in several studies, and in obese asthmatic patients, a correlation was seen between serum IL-6 levels and impaired lung function. These results suggest that IL-6 may be involved in the pathogenesis of asthma, and suggest that IL-6 may be a target for asthma treatment. [142, 143] However, this PheWAS suggests that an IL-6 inhibitor would not provide benefit for asthma patients and may actually cause increased risk of asthma. This is reinforced by a similar finding in the PheWAS for another IL-6 inhibiting SNP, rs4129267. [51] Therefore, it is possible that the IL-6 processes in asthma are not actually pathogenic, but instead are involved in the disease process in other pro-inflammatory ways.

Strengths and limitations

There are strengths and limitations to the PheWAS method. A strength of PheWAS is the availability of large GWAS data sets; the size of these studies enables identification of many traits associated with specific SNPs. PheWAS allow evaluation of some diseases that may not be considered to have genetic effects, such as specific blood counts. A major benefit of PheWAS is that it provides a method for searching for other drug targets, which can save substantial time and money. However, any PheWAS, but particularly in the context of drug target repurposing, is hypothesis-generating only; other studies such as randomized controlled trials are still required to confirm the effect suggested in a PheWAS. A Bonferroni correction for statistical significance may not be appropriate if the traits are not independent from each other. Differences across populations could affect both the expected allele frequency for the minor allele (which can vary by population) and also access of different populations to medical care from which diagnosis and ultimately identification of traits in a PheWAS are possible. [85]

Given the ongoing identification of new SNPs that are associated with IL-6 levels, and given the importance of IL-6 in the pathology of so many diseases, it will be important to continue to consider whether any of these SNPs besides rs7529229 are specifically associated with response to an IL-6 inhibitor such as TCZ. Other SNPs associated with TCZ response have been identified in a small number of patients, but to date these have not been validated. [147]

CONCUSION

TCZ is already known to be inversely associated with RA and CHD and positively associated with eczema and asthma. This study did not identify new targets for TCZ. As more SNPs are identified as associated with IL-6 levels, it may be possible to determine additional SNPs associated with TCZ or other IL-6 inhibitors, from which point new drug targets may be identified.

KEY MESSAGES

- Phenome-wide association studies can be used to identify potential new drug targets.
- The current PheWAS did not identify any new targets for TCZ, an IL-6 inhibitor.
- More SNPs associated with IL-6 levels, and IL-6 inhibiting drug therapies, may enable identification of new drug targets in the future.

REFERENCES

1. Lawlor, D.A., et al., Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med, 2008. **27**(8): p. 1133-63.
2. Collins, F.S. and H. Varmus, A new initiative on precision medicine. N Engl J Med, 2015. **372**(9): p. 793-5.
3. Interleukin-6 Receptor Mendelian Randomisation Analysis, C., et al., The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet, 2012. **379**(9822): p. 1214-24.
4. Li, B., et al., Circulating interleukin-6 and rheumatoid arthritis: A Mendelian randomization meta-analysis. Medicine (Baltimore), 2016. **95**(23): p. e3855.
5. Srirangan, S. and E.H. Choy, The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. Ther Adv Musculoskelet Dis, 2010. **2**(5): p. 247-56.
6. Actemra prescribing information. December 23, 2019]; Available from: https://www.gene.com/download/pdf/actemra_prescribing.pdf.
7. King, E.A., J.W. Davis, and J.F. Degner, Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet, 2019. **15**(12): p. e1008489.
8. Roberts, R., Mendelian Randomization Studies Promise to Shorten the Journey to FDA Approval. JACC Basic Transl Sci, 2018. **3**(5): p. 690-703.
9. Schmidt, A.F.F., C.; Gordillo-Marañón M, et al., Genetic drug target validation using Mendelian randomization. bioRxiv 2019.
10. Nelson, M.R., et al., The support of human genetic evidence for approved drug indications. Nat Genet, 2015. **47**(8): p. 856-60.
11. Humira prescribing information. December 13, 2019]; Available from: https://www.rxabbvie.com/pdf/humira.pdf.
12. Mokry, L.E., et al., Interleukin-18 as a drug repositioning opportunity for inflammatory bowel disease: A Mendelian randomization study. Sci Rep, 2019. **9**(1): p. 9386.
13. Zheng, J., et al., Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep, 2017. **4**(4): p. 330-345.
14. Wurtz, P., et al., Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. J Am Coll Cardiol, 2016. **67**(10): p. 1200-1210.
15. Ference, B.A., et al., Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. J Am Coll Cardiol, 2015. **65**(15): p. 1552-61.
16. Myocardial Infarction Genetics Consortium, I., et al., Inactivating mutations in NPC1L1 and protection from coronary heart disease. N Engl J Med, 2014. **371**(22): p. 2072-82.
17. He, S., et al., A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. J Biol Chem, 2010. **285**(9): p. 6706-15.
18. Cohen, J.C., et al., Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med, 2006. **354**(12): p. 1264-72.
19. Navarese, E.P., et al., Effects of Proprotein Convertase Subtilisin/Kexin Type 9 Antibodies in Adults With Hypercholesterolemia: A Systematic Review and Meta-analysis. Ann Intern Med, 2015. **163**(1): p. 40-51.

20. Collaboration, C.R.P.C.H.D.G., et al., Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. BMJ, 2011. **342**: p. d548.
21. Zacho, J., et al., Genetically elevated C-reactive protein and ischemic vascular disease. N Engl J Med, 2008. **359**(18): p. 1897-908.
22. Polfus, L.M., R.A. Gibbs, and E. Boerwinkle, Coronary heart disease and genetic variants with low phospholipase A2 activity. N Engl J Med, 2015. **372**(3): p. 295-6.
23. Millwood, I.Y., et al., Lipoprotein-Associated Phospholipase A2 Loss-of-Function Variant and Risk of Vascular Diseases in 90,000 Chinese Adults. J Am Coll Cardiol, 2016. **67**(2): p. 230-231.
24. Ryaboshapkina, M. and M. Hammar, Tissue-specific genes as an underutilized resource in drug discovery. Sci Rep, 2019. **9**(1): p. 7233.
25. Pendergrass, S.A., et al., Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS Genet, 2013. **9**(1): p. e1003087.
26. Hemani, G., et al., The MR-Base platform supports systematic causal inference across the human phenome. Elife, 2018. **7**.
27. Welter, D., et al., The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.
28. Denny, J.C., et al., Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol, 2013. **31**(12): p. 1102-10.
29. Rosa, M., et al., A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. NPJ Genom Med, 2019. **4**: p. 23.
30. Cai, T., et al., Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy: A Phenome-Wide Association Study. JAMA Cardiol, 2018. **3**(9): p. 849-857.
31. Coronary Artery Disease Genetics, C., A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet, 2011. **43**(4): p. 339-44.
32. Eyre, S., et al., High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet, 2012. **44**(12): p. 1336-40.
33. Arnett, F.C., et al., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum, 1988. **31**(3): p. 315-24.
34. Okada, Y., et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature, 2014. **506**(7488): p. 376-81.
35. Wellcome Trust Case Control, C., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. **447**(7145): p. 661-78.
36. Phenoscanner. December, 4, 2019]; Available from: http://www.phenoscanner.medschl.cam.ac.uk.
37. Agency, E.M., Assessment Report, RoActemra, C.f.M.P.f.H.U. (CHMP), Editor. 2018.
38. Neveu, W.A., et al., Elevation of IL-6 in the allergic asthmatic airway is independent of inflammation but associates with loss of central airway function. Respir Res, 2010. **11**: p. 28.
39. Rincon, M. and C.G. Irvin, Role of IL-6 in asthma and other inflammatory pulmonary diseases. Int J Biol Sci, 2012. **8**(9): p. 1281-90.

40. Yokoyama, A., et al., Circulating interleukin-6 levels in patients with bronchial asthma. Am J Respir Crit Care Med, 1995. **151**(5): p. 1354-8.
41. Tillie-Leblond, I., et al., Balance between proinflammatory cytokines and their inhibitors in bronchial lavage from patients with status asthmaticus. Am J Respir Crit Care Med, 1999. **159**(2): p. 487-94.
42. Hirano, T., Interleukin 6 and its receptor: ten years later. Int Rev Immunol, 1998. **16**(3-4): p. 249-84.
43. Hebbring, S.J., The challenges, advantages and future of phenome-wide association studies. Immunology, 2014. **141**(2): p. 157-65.
44. Wang, J., et al., Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. Pharmacogenomics J, 2013. **13**(3): p. 235-41.

Table 4.1. PheWAS for rs7529229, including 10 unique traits (listed as 14 traits due to multiple data sources). Negative beta indicates inverse association between rs7529229 and trait (decreased risk); positive beta indicates positive association between rs7529229 and trait (increased risk).

ID	Trait	Consortium	Sample Size	Beta	EA	OA	Beta	Pval	SE	Cases	Controls
283	Rheumatoid arthritis		27345	-0.1009259	C	T	-0.1009259	1.1430E-07		11475	15870
833	Rheumatoid arthritis		80799	0.07696104	C	T	-0.076961	1.8000E-08	0.01417598	19234	61565
7	Coronary heart disease	CARDIoGRAMplusC4D	184305	0.048757	C	T	-0.048757	2.5000E-07	0.0094539	60801	123504
1269	Red cell distribution width	UK Biobank+INTERVAL+UK BiLEVE	171529	-0.0180799	C	T	-0.0180799	4.1350E-07	0.00357124	NA	NA
1256	Granulocyte percentage of myeloid white cells	UK Biobank+INTERVAL+UK BiLEVE	169545	-0.0171947	C	T	-0.0171947	1.8330E-06	0.00360404	NA	NA
UKB-b:20141	Non-cancer illness code, self-reported: eczema/dermatitis	MRC-IEU	462933	-0.001829	C	T	0.00182903	3.6000E-08	0.00033209	11819	451114
UKB-b:15096	Operation code: tonsillectomy / tonsil surgery	MRC-IEU	462933	-0.0021255	C	T	0.00212549	1.1000E-06	0.00043618	20848	442085
UKB-a:99	Non-cancer illness code self-reported: eczema/dermatitis	Neale Lab	337159	0.00216192	C	T	0.00216192	3.3846E-08	0.00039162	8718	328441
UKB-b:20296	Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: Asthma	MRC-IEU	462013	-0.0032119	C	T	0.00321186	1.6000E-06	0.00066934	53257	408756
UKB-b:18113	Non-cancer illness code, self-reported: asthma	MRC-IEU	462933	-0.0032168	C	T	0.00321677	1.6000E-06	0.00066982	53598	409335

UKB-b:17241	Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: Hayfever, allergic rhinitis or eczema	MRC-IEU	462013	-0.0067188	C	T	0.00671876	2.2000E-14	0.00087962	106201	355812
UKB-a:447	Blood clot DVT bronchitis emphysema asthma rhinitis eczema allergy diagnosed by doctor: Hayfever allergic rhinitis or eczema	Neale Lab	336782	0.00756788	C	T	0.00756788	3.3372E-13	0.0010395	77891	258891
1278	Mean corpuscular hemoglobin	UK Biobank+INTERVAL+UK BiLEVE	172332	0.01811742	C	T	0.01811742	3.9520E-07	0.00357257	NA	NA
1257	Monocyte percentage of white cells	UK Biobank+INTERVAL+UK BiLEVE	170494	0.02081604	C	T	0.02081604	6.9300E-09	0.00359355	NA	NA

EA: Effect allele; OA: other allele

CHAPTER 5: DISCUSSION

SUMMARY AND DISCUSSION

These Mendelian randomization studies have shown a decreased risk of CD in RA patients and no effect of CD on RA, and the PheWAS for rs7529229 did not suggest new targets for tocilizumab, an IL-6 inhibitor currently approved for RA and several other indications.

Co-occurrence of RA and CD vs co-occurrence of IMIDs

While IMIDs have been shown to co-occur in 9-10% of patients, [8] RA and CD specifically co-occur far less frequently, with estimates ranging from 0.0027%-1.7%. [12][13] It seems that other IMIDs may co-occur, but not specifically RA with CD. This is one factor that could have contributed to the failure to find any causal association between CD and RA, but it does not explain the decreased risk of CD in RA patients found in this MR study.

Shared etiology: the role of interleukin-6 (IL-6)]inhibition in RA, CD

Pre-treatment of subclinical CD in RA patients

There are thought to be multiple common contributing causes of RA and CD, including IL-6 and TNF-a overexpression. A possible explanation for the decreased risk of CD in RA patients is that treating this overexpression of a causal cytokine could simultaneously treat diagnosed RA and preemptively treat subclinical CD, which could prevent CD from being clinically diagnosed. As some medications are approved for the treatment of both RA and CD (e.g., adalimumab, marketed as Humira by Abbvie), [31] a shared causal pathway makes sense. However, this explanation would suggest the opposite to be true: that in CD patients, pre-treatment of subclinical RA would lead to the decreased risk of RA in CD patients, but this result was not seen in the second MR study.

IL-6 in RA and CD

IL-6 is a cytokine that induces protein synthesis in hepatocytes, and increased serum IL-6 has been seen in acute conditions including surgery, burns, and infections, as well as in chronic conditions, such as CD and RA. Serum IL-6 concentrations were found to be higher in CD patients than in healthy controls, and a nonsignificant trend was found toward higher serum IL-6 in CD patients with active disease versus inactive disease. [102] In RA patients, the IL-6 -174G/C variant has been associated with RA in both European and Asian populations. [41] IL-6 promotes joint inflammation through vascular endothelial growth factor (VEGF), levels of which are associated with RA disease activity. IL-6 is found in synovial fluid and serum of RA patients, with levels corresponding to disease activity and joint damage. Serum IL-6 levels were elevated in RA patients compared to controls, and higher levels were associated with more severe disease activity as measured by the DAS28. [123]

IL-6 in asthma

The positive association of rs7529229 with asthma is more interesting and somewhat unexpected. A negative association was expected, because IL-6 has a plausible role in the pathogenesis of asthma. Lung epithelial cells produce IL-6 in response to respiratory viruses, allergens, or exercise. [142, 143] Circulating IL-6 is elevated in patients with asthma, [144] including in the bronchoalveolar lavage fluid (BALF) of asthmatic patients; [143, 145] this suggests that the role of IL-6 is more than just pro-inflammatory as this elevated level is seen in both allergic and intrinsic asthma. IL-6 measured in BALF could aid in identifying patients with for treatment with an IL-6 inhibitor. IL-6 is associated with both viral infections, which can lead to exacerbations of asthma, and obesity, which is associated with severity of asthma. [143, 146] IL-6 levels have been found to be associated with lung function; forced expiratory volume in 1

second (FEV1) was shown to be inversely associated with IL-6 levels in patients in several studies, and in obese asthmatic patients, a correlation was seen between serum IL-6 levels and impaired lung function. These results suggest that IL-6 may be involved in the pathogenesis of asthma, and that IL-6 may be a target for asthma treatment. [142, 143] However, this PheWAS suggests that an IL-6 inhibitor would not provide benefit for asthma patients and may actually cause increased risk of asthma. This is reinforced by a similar finding in the PheWAS for another IL-6 inhibiting SNP, rs4129267. [51] Therefore, it is possible that the IL-6 processes in asthma are not actually pathogenic, but instead are involved in the disease process in other, pro-inflammatory ways.

RA as protective for CD

It is possible that RA is in fact protective for CD. A mechanism for this has not been described. Further research would be necessary to understand a protective mechanistic effect of RA on CD.

Problems with data sources for RA and CD

A feasible reason for the results showing a reduced risk of CD in RA patients and no causal relationship between CD and RA is bias related to the age distribution of the RA and CD GWAS used in these MR studies.

The first MR study evaluated the causal effect of RA on CD. Since RA is often diagnosed at an older age than CD, [17, 18, 24] if patients enrolled in the GWAS with diagnosed RA but without having been diagnosed with CD, then those participants may have been unlikely ever to develop CD. This makes it seem like RA is protective for CD. This is not a mechanistic protective effect of RA, but a function of people diagnosed with RA but not CD enrolled in the GWAS. This bias is likely a strong factor in these results.

The second MR study evaluated the causal effect of CD on RA. CD is has two peak age ranges for diagnosis, and the first is several decades younger than the second, which essentially overlaps with age of diagnosis of RA. [17, 18, 24] The average age of all participants in the full CD GWAS was not found, but one of the component GWAS reported that 60% of participants were under age 50. [67] Therefore, is possible that patients were enrolled in the CD GWAS prior to an age of RA diagnosis. This would suggest that RA could not be detected in the analysis because there were too few GWAS participants with an RA diagnosis simple due to their age. A different study with an age distribution skewed older may have yielded a different result.

A further problem with the GWAS for RA and CD is the lack of clarity regarding the control groups. It is not clear whether each study had "healthy controls"—those lacking RA in the CD GWAS or those lacking CD in the RA GWAS. A control group in a case-control study should represent the underlying cohort from which the cases were drawn, suggesting that there should have been RA patients in the controls for CD and vice versa. This could not be ascertained from descriptions of the RA and CD studies. Use of healthy controls would likely impair an ability to detect a true association between the two diseases, including direction of such an association.

Strengths and limitations

There are many strengths to the MR method. The F-statistics suggested strong instruments, and large sample sizes were available for both the MR studies and the PheWAS. Sample size allows identification of many traits in PheWAS; this method is also useful because it allows evaluation of traits not always considered genetic. This PheWAS was consistent with two other PheWAS of other SNPs implicated in IL-6.

There are also limitations to the MR method. First, the method relies on strong assumptions which can be difficult to verify empirically. This method requires genetic instruments that

predict the outcome. GWAS have provided such instruments for RA, CD, and response to TCZ in RA patients. However, genetic variants may have only a small effect on these diseases, which could introduce bias due to weak instruments. [83, 84]

Second, confounders of genetic variants on exposure may exist, typically as a result of population stratification; therefore, a homogeneous population is generally necessary for MR studies, with correction for population stratification; the current studies use fairly homogeneous populations, although the RA GWAS, which includes participants of both European and Asian descent, may be less so than the others [83, 84].

Small effect sizes may not correspond to clinical significance, nor indicate effects of a clinical intervention at a future time, either due to the importance of cumulative exposure or time-dependent exposure. [64] Additionally, low co-occurrence of RA and CD does not indicate lack of importance in those patients.

MR studies are only as good as the data sources available. The RA and CD data sources lacked clarity on several important issues, including age range and composition of control groups. This information would have enabled better interpretation of these MR results.

Finally, an important current limitation of MR studies is that they only address causal factors of disease incidence, but not issues related to disease progression; studies of disease progression and studies of patients are open to selection bias because they are missing those who have already died because of the exposure or because of competing risk of the outcome. [83, 84]

There are also strengths and weaknesses to PheWAS. A strength of PheWAS is the availability of large GWAS data sets; the size of these data sets enables identification of many traits associated with specific SNPs. PheWAS allow evaluation of some diseases that may not be

considered to have genetic effects, such as specific blood counts. A major benefit of PheWAS is that it provides a method for searching for other drug targets, which can save substantial time and money. However, PheWAS are hypothesis-generating only; other studies such as randomized controlled trials are still required to confirm the effect seen in a PheWAS. A Bonferroni correction for statistical significance may not be appropriate if the traits are not independent from each other. Differences across populations could affect both the expected allele frequency for the minor allele (which can vary by population) and also access of different populations to medical care from which diagnosis and ultimately identification of traits in a PheWAS are possible. [85]

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

These Mendelian randomization studies determined that rheumatoid arthritis is associated with a decreased risk of Crohn's disease and that Crohn's disease has no causal effect on rheumatoid arthritis. It is likely that these results were driven by bias due to the age distributions of the GWAS populations, though a protective effect of RA on development of CD cannot be excluded.

In the first MR study, GWAS participants who have made it to the age of developing RA without ever having developed CD are individuals who are unlikely to have ever developed CD. In the second MR study, GWAS participants with CD may have been too young to have developed RA, meaning it was not possible to determine an association between the two disease. Control groups selected for each GWAS may have also affected these results, and the lack of clarity on these control groups to some extent impairs interpretation of the results.

The results of the PheWAS are essentially unsurprising. An inverse association between rs7529229 and RA is to be expected, as RA is an approved indication for TCZ. [106] The inverse association between rs7529229 and CHD in particular has also been described, [13] as has this

relation between other IL-6 inhibiting SNPs and CHD. [51, 52] The effects on blood counts, whether positive or negative, are difficult to interpret, and have been reported; [51, 52] additional, changes to blood counts are described as potential adverse events in the TCZ product information. [106] Eczema has also been reported in the clinical trial program, [141] which reinforces the positive association between rs7529229 and eczema, which has been reported in other IL-6 inhibitor PheWAS. [51, 52] The positive association between rs7529229 and asthma is more interesting and possibly unexpected. A negative association would be more expected, because IL-6 has a plausible role in the pathogenesis of asthma; however, this PheWAS suggests that an IL-6 inhibitor would not provide benefit for asthma patients and may actually cause increased risk of asthma. This is reinforced by a similar finding in the PheWAS for another IL-6 inhibiting SNP, rs4129267. [51] Therefore, it is likely that the IL-6 processes in asthma are not actually pathogenic but are involved in the disease process in other ways.

KEY MESSAGES

- Mendelian randomization can determine causal associations between exposures and outcomes, if appropriate instruments are identified and adequate sample sizes are available.
- These MR studies showed a lower risk for CD in RA patients and no causal effect of CD on RA, which may be due to bias in the underlying studies.
- However, a protective effect of RA on CD cannot be ruled out; further research is required to determine a biological mechanism for this effect.
- PheWAS can be used to identify potential new drug targets, but the current PheWAS did not identify any new targets for TCZ, an IL-6 inhibitor.

- More SNPs associated with IL-6 levels, and IL-6 inhibiting drug therapies, may enable identification of new drug targets in the future.

REFERENCES

1. Robinson, D., Jr., et al., Co-occurrence and comorbidities in patients with immune-mediated inflammatory disorders: an exploration using US healthcare claims data, 2001-2002. Curr Med Res Opin, 2006. **22**(5): p. 989-1000.
2. Humira prescribing information. December 13, 2019]; Available from: https://www.rxabbvie.com/pdf/humira.pdf.
3. Gross, V., et al., Evidence for continuous stimulation of interleukin-6 production in Crohn's disease. Gastroenterology, 1992. **102**(2): p. 514-9.
4. Li, B., et al., Circulating interleukin-6 and rheumatoid arthritis: A Mendelian randomization meta-analysis. Medicine (Baltimore), 2016. **95**(23): p. e3855.
5. Beyazal, M.S., Devrimsel, G., Cure, M.C., et al., Serum Levels of IL-17, IL-6, TNF-α and Disease Activity in Patients with Rheumatoid Arthritis. Akt Rheumatol, 2017. **42**: p. 323-8.
6. Neveu, W.A., et al., Elevation of IL-6 in the allergic asthmatic airway is independent of inflammation but associates with loss of central airway function. Respir Res, 2010. **11**: p. 28.
7. Rincon, M. and C.G. Irvin, Role of IL-6 in asthma and other inflammatory pulmonary diseases. Int J Biol Sci, 2012. **8**(9): p. 1281-90.
8. Yokoyama, A., et al., Circulating interleukin-6 levels in patients with bronchial asthma. Am J Respir Crit Care Med, 1995. **151**(5): p. 1354-8.
9. Tillie-Leblond, I., et al., Balance between proinflammatory cytokines and their inhibitors in bronchial lavage from patients with status asthmaticus. Am J Respir Crit Care Med, 1999. **159**(2): p. 487-94.
10. Hirano, T., Interleukin 6 and its receptor: ten years later. Int Rev Immunol, 1998. **16**(3-4): p. 249-84.
11. Rosa, M., et al., A Mendelian randomization study of IL6 signaling in cardiovascular diseases, immune-related disorders and longevity. NPJ Genom Med, 2019. **4**: p. 23.
12. Innala, L., et al., Age at onset determines severity and choice of treatment in early rheumatoid arthritis: a prospective study. Arthritis Res Ther, 2014. **16**(2): p. R94.
13. Goemaere, S., et al., Onset of symptoms of rheumatoid arthritis in relation to age, sex and menopausal transition. J Rheumatol, 1990. **17**(12): p. 1620-2.
14. Quezada, S.M., E.K. Steinberger, and R.K. Cross, Association of age at diagnosis and Crohn's disease phenotype. Age Ageing, 2013. **42**(1): p. 102-6.
15. Wellcome Trust Case Control, C., Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 2007. **447**(7145): p. 661-78.
16. Schooling, C.M., G. Freeman, and B.J. Cowling, Mendelian randomization and estimation of treatment efficacy for chronic diseases. Am J Epidemiol, 2013. **177**(10): p. 1128-33.
17. Schooling, C.M.L., P.; Au Yeung, S. L., et. al., Survival bias and competing risk can severely bias Mendelian Randomization studies of specific conditions. https://www.biorxiv.org/content/biorxiv/early/2019/10/21/716621.full.pdf, 2019.
18. Walker, V.M., et al., Using the MR-Base platform to investigate risk factors and drug targets for thousands of phenotypes. Wellcome Open Res, 2019. **4**: p. 113.
19. Hebbring, S.J., The challenges, advantages and future of phenome-wide association studies. Immunology, 2014. **141**(2): p. 157-65.

20. Actemra prescribing information. December 23, 2019]; Available from: https://www.gene.com/download/pdf/actemra_prescribing.pdf.
21. Interleukin-6 Receptor Mendelian Randomisation Analysis, C., et al., The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet, 2012. **379**(9822): p. 1214-24.
22. Cai, T., et al., Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy: A Phenome-Wide Association Study. JAMA Cardiol, 2018. **3**(9): p. 849-857.
23. Agency, E.M., Assessment Report, RoActemra, C.f.M.P.f.H.U. (CHMP), Editor. 2018.

Appendix 1. Power calculation for Specific Aim 1, Chapter 2

Code for the calculation was provided by Burgess [1] with #notes added.

```
#rsq from MR-base output
#b1 from IVW analysis beta
#17897 CD cases and 33977 controls
#total sample 51874
#ratio 0.52673868
expit <- function(x) { return(exp(x)/(1+exp(x))) }
rsq = 0.0578485 # squared correlation
b1 = 0.1206 # causal effect
sig = 0.05 # significance level (alpha)
ratio = 0.526738676 # ratio of cases:controls = 1:ratio
#minimum effect size to detect given sample size with 80% power is 0.108

#power calculation
cat("Power of analysis with ", 51874, "participants: ",
   pnorm(sqrt(51874*rsq*(ratio/(1+ratio))*(1/(1+ratio)))*b1-qnorm(1-sig/2)))
Power= 0.803
```

Appendix 2: Power and sample size calculations for Specific Aim 2, Chapter 3

Code for these calculation was provided by Burgess [1] with #notes added as follows:

```
#power calculation
# beta for IVW 0.02637
# r2 exposure from MR- base report 0.2097407
#outcome case and control counts for total n and for ratio
#RA cases 19234 RA controls 61565 ratio 0.31241777 total 80799
#determined minimum effect size to provide 80% power
#beta to produce 80% power=0.05


expit <- function(x) { return(exp(x)/(1+exp(x))) }
rsq = 0.81582 # squared correlation
b1 = 0.0505 # causal effect estimate to reach 80% power
sig = 0.05 # significance level (alpha)
pow = 0.8 # power level (1-beta)
ratio = 0.31241777 # ratio of cases:controls = 1:ratio
#power calc SA2


cat("Power of analysis with ", 80799, "participants: ",
   pnorm(sqrt(80799*rsq*(ratio/(1+ratio))*(1/(1+ratio)))*b1-qnorm(1-sig/2)))
#sample size calculation for effect size to reach significance
cat("Sample size required for ", pow*100, "% power: ",
   (qnorm(1-sig/2)+qnorm(pow))^2/b1^2/rsq/(ratio/(1+ratio))/(1/(1+ratio)))
Sample size required for  80 % power:  296697
```

REFERENCES

1. Burgess, S., Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol, 2014. 43(3): p. 922-9.