A causal association between schizophrenia and bipolar disorder on rheumatoid arthritis: A two-sample Mendelian randomization study

Gonul Hazal Koc¹, Fatih Ozel², Kaan Okay³,⁴, Dogukan Koc⁵*

¹Department of Internal Medicine, Izmir Katip Celebi University Faculty of Medicine, Izmir, Turkey
²Environmental Toxicology, Department of Organismal Biology, Uppsala University, Uppsala, Sweden
³Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, Izmir, Turkey
⁴Izmir Biomedicine and Genome Center, Dokuz Eylul University Health Campus, Izmir, Turkey
⁵Department of Child and Adolescent Psychiatry, Dokuz Eylul University, Izmir, Turkey

*Corresponding Author at Department of Child and Adolescent Psychiatry, Dokuz Eylul University Faculty of Medicine, 35330 Balcova/Izmir, Turkey. Tel: +90 232 4123564
E-mail address: dogukan.koc@deu.edu.tr

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

**Background:** Schizophrenia (SCZ) and bipolar disorder (BD) are both associated with several autoimmune/inflammatory disorders including rheumatoid arthritis (RA). However, a causal association of SCZ and BD on RA is controversial and elusive. In the present study, we aimed to investigate the causal association of SCZ and BD with RA by using the Mendelian randomization (MR) approach.

**Methods:** A two-sample MR (2SMR) study including the inverse-variance weighted (IVW), weighted median, simple mode, weighted mode and MR-Egger methods were performed. We employed summary-level genome-wide association study (GWAS) data including BD and SCZ as exposure and RA as an outcome. We utilized data from the Psychiatric Genomics Consortium (PGC) for BD (n = 41,917) and SCZ (n = 33,426), whereas RA GWAS dataset (58,284 individuals) from the European ancestry.

**Results:** We obtained independent \( r^2 < 0.001 \) 48 and 52 single nucleotide polymorphisms (SNPs) from BD and SCZ data at genome-wide significance \( p < 5 \times 10^{-8} \), respectively. Next, these SNPs were utilized as instrumental variables (IVs) in 2SMR analysis to explore the causality of BD and SCZ on RA. The two out of five MR methods showed a statistically significant inverse causal association between BD and RA: weighted median method (odds ratio (OR), 0.869, [95% CI, 0.764-0.989]; \( P = 0.034 \)) and inverse-variance weighted (IVW) method (OR, 0.810, [95% CI, 0.689-0.953]; \( P = 0.011 \)). However, we did not find any significant association of SCZ with RA (OR, 1.008, [95% CI, 0.931-1.092]; \( P = 0.829 \), using the IVW method).

**Conclusions:** These results provide support for an inverse causal association between BD and RA. Further investigation is needed to explain the underlying protective mechanisms in the development of RA.

**Key words:** Bipolar disorder; Schizophrenia; Rheumatoid arthritis; Genome-wide association study, Mendelian randomization; Single nucleotide polymorphism
Introduction

Schizophrenia (SCZ) and bipolar disorder (BD) are severe psychiatric disorders that are characterized by significant emotional, cognitive and behavioral symptoms, and they cause substantial deterioration in functioning [1, 2]. Both disorders pose an increased risk of morbidity and mortality and are associated with decreased life expectancy [3-5]. Genetic factors have been investigated for SCZ [6] and the interplay between genetic and environment has also been a focus [7, 8]. Still, the etiopathogenesis of SCZ seems heterogeneous, and the existing evidence remains inconclusive. Many risk factors have been identified for BD in a similar manner [9], yet, our understanding of the biological underpinnings is limited. Notwithstanding, disturbances in inflammatory signaling mechanisms seem to play a critical role in the etiopathogenesis of SCZ and BD [10, 11]. Considerable evidence supporting the link between these disorders and inflammation emerges from the epidemiological studies that show the increased comorbidity of SCZ and BD with autoimmune diseases [12, 13]. However, not only causality is an important aspect to be investigated, but also potential confounders stand to be addressed in those relations of SCZ, BD and inflammatory diseases.

Rheumatoid arthritis (RA) is a common autoimmune disease that arises from abnormal immune activity and inflammatory responses [14]. The disease also causes devastating consequences and remarkable disability among elderly persons [15]. Individuals with RA also suffer from medical comorbidities, and these comorbid disorders have considerable impacts on disability and daily functioning; therefore, psychiatric disorders have become one of the focuses of interest. Population-based studies mainly demonstrated the lower prevalence of RA in patients with SCZ [16-18]. The possible explanations for the lower prevalence of RA and SCZ have not been examined comprehensively, and the lack of evidence raises the demand for exploring this relationship with various potent methods. On the other hand, the high comorbidity of RA has been speculated for BD in the literature. Two different case-control studies demonstrated a higher prevalence of RA among patients with BD [19, 20]. The increased risk of developing BD in persons with RA has also been established in population-based settings [21-23]. Even though, a high prevalence of BD among patients with RA has been shown in another case-control study [24], the association between BD and RA was not significant in the multivariate analysis, which posits the importance of potential confounding.
Several studies mentioned above have reported the lower RA comorbidity for SCZ, whereas the higher RA comorbidity for BD. Despite these findings, the evidence on comorbidity is still limited and the causal relationships are yet to be established. Nonetheless, inflammatory disturbances seem to be the main reason for the associations between these psychiatric disorders and RA; potential confounders need to be investigated thoroughly, to avoid biased associations. Regrettably, observational studies are limited to capture many confounders, and experimental designs carry other difficulties to conduct to scrutinize the coexistence of RA and SCZ or BD. Mendelian randomization (MR) may be a favorable solution to circumvent these shortcomings. MR is a robust epidemiological method that uses the genetic variants related to the exposure of interest to make causal inferences about the non-genetic measures of the outcome [25]. MR exploits single-nucleotide polymorphisms (SNPs) as instrumental variables (IVs) for the exposure [26]; thus, it can evade the issues caused by confounding and provide more accurate causal interpretations. To our best knowledge, any MR study has not been performed to examine the causality between RA and SCZ or BD.

In the present study, we aimed to investigate the potential causal relationship between having SCZ or BD and the risk of developing RA. In doing so, we performed a 2SMR approach using the latest GWAS summary data with SCZ, BD and RA.

**Materials and methods**

We performed a 2SMR approach to assess causality between BD, SCZ, and RA using publicly available GWAS summary datasets.

**Data sources**

We downloaded the latest GWAS summary data with BD I and II, includes 7,608,183 single nucleotide polymorphisms (SNPs) of 41,917 cases and 371,549 controls [27] and the SCZ GWAS data including 8,379,106 SNPs of 33,426 cases and 32,541 controls [28] from the Psychiatric Genomics Consortium database (https://www.med.unc.edu/pgc/). The RA GWAS summary data including 8,514,610 SNPs of 14,361 RA cases and 43,923 controls [29] was downloaded from Japanese ENcyclopedia of GEnetic associations by Riken database (http://jenger.riken.jp/en/result). All GWAS summary data are based on European ancestry.
**Instrument identification**

We utilized R package TwoSampleMR v0.5.6 to harmonise SNP information and conducted a 2SMR analysis between BD, SCZ, and RA [30]. All GWAS data had beta coefficient (β), standard error of β (SE), major and minor alleles for each SNP together with the allele frequencies, p-value for relevant association, and total sample size information. We followed four steps in the analysis: i) genome-wide significant SNPs ($p \leq 5 \times 10^{-8}$) related to exposure were selected as (IVs); ii) linkage disequilibrium (LD) pruning ($r^2 < 0.001$) was performed to obtain independent SNPs; iii) genome-wide significant and independent SNPs were extracted from outcome GWAS data and those SNPs were utilized in harmonization of exposure and outcome GWAS data to ensure their effect on those data to correspond to the same allele; iv) harmonized data were used in a 2SMR analysis. Eventually, we identified independent 48 and 52 SNPs at genome-wide significance from BD and SCZ GWAS data, respectively. These SNPs were used as IVs for harmonization, MR, heterogeneity, and sensitivity analyses in the R computation environment v4.1.0 (http://www.R-project.org).

**Two-sample Mendelian randomization**

The β and SE of β were used in the estimation of the causal relationship. We obtained MR results for the following five methods: Inverse-variance weighted (IVW) [31], MR-Egger [32], weighted median [33], weighted mode and simple mode [34]. In the one direction, we assigned BD as exposure, whereas RA as an outcome. We identified independent 48 SNPs at genome-wide significance from BD data and they were used as IVs in a 2SMR analysis. The aforementioned five MR methods were utilized to generate effect size estimates of these SNPs. In the other direction, we considered SCZ as exposure, whereas RA as an outcome. At genome-wide significance, independent 52 SNPs were obtained in SCZ data and they were used as IVs in a 2SMR analysis.

**Sensitivity analyses**

Sensitivity analyses were performed to examine the existence of horizontal pleiotropy and heterogeneity. In the case of horizontal pleiotropy, a single locus can affect an outcome through one or more biological pathways independent of that of the assumed exposure. To test the horizontal pleiotropic effect amongst BD, SCZ, and RA, we performed MR-Egger regression.
analysis. Besides, we employed the leave-one-out analysis, where one variant that is strongly associated with exposure and dominates the estimate of the causal effect is removed from the analysis to re-estimate the causal effect.

**Results**

After genome-wide significance filtering, LD pruning, and harmonization of potential SNPs, 48 BD and 52 SCZ independent genome-wide significant SNPs were considered as IVs. These IVs were used in a 2SMR analysis to estimate the causality of BD and SCZ on RA. The detailed information on these IVs can be found in Supplementary Table S1 and S2. The relationship is represented as odds ratios (ORs) with respective p-values in Table 1.

We used five MR methods to estimate the causality of BD and SCZ on RA. We found an inverse causal association between BD and RA, MR methods weighted median (OR= 0.869, [95% CI 0.764-0.989]; P= 0.034) and IVW (OR= 0.810, [95% CI 0.689-0.953]; P= 0.011); whereas, others methods had no statistical significance, MR Egger (OR=0.652, [95% CI 0.296-1.434]; P= 0.293), simple mode (OR= 0.905, [95% CI 0.686-1.195]; P= 0.488), and weighted mode (OR= 0.879, [95% CI 0.678-1.139]; P= 0.336). These estimates are available in Figure 1A and Table 1 in detail. Besides, we performed a leave-one-out analysis and did not find a notable effect of any single SNP that could dominate the results (Supplementary Figure S1). The funnel plot did not signify that there were heterozygous SNPs (Figure 1B). We did not find evidence for pleiotropy from the MR-Egger regression (intercept: 0.015, P= 0.583). The heterogeneity results can be found in Supplementary Table S3. On the other hand, we showed no statistical significance in the causal relationship between SCZ and RA (Table 1). More detailed results of estimation of this direction are presented as scatter and funnel plots in the Figure 1C-D, respectively.

**Discussion**

We performed 2SMR analyses to explore the causality between SCZ, BD and RA based on summary statistics of the largest available GWAS to date. Our analyses suggested that BD had a protective effect on RA, whereas SCZ had no significant effect on RA. Considering the power of the MR approach on the exclusion of confoundings, the effects of environmental factors can be decreased because genetic variants are assigned randomly at
conception, similar to randomization in a clinical trial. From this standpoint, our study can be considered as the first potential evidence for uncovering causal relationships between psychiatric disorders (SCZ, BD) and RA.

A large number of existing studies in the broader literature have shown an inverse association between SCZ and RA [17, 18, 35-38]. It has been hypothesized that SCZ is a protective factor for RA, some focusing on environmental exposures and immunity, others on misclassification biased by underdiagnosis of RA in patients with SCZ [18, 39, 40]. A recent population-based case-control study from Sweden demonstrated that patients with SCZ have a decreased risk of developing RA (hazard ratio [HR] = 0.69, 95% CI = 0.59–0.80) [18]. The similar results were obtained from a Danish Psychiatric Case Register study involving 20,495 patients with SCZ and 204,912 randomly selected age- and sex-matched controls (HR= 0.44, 95% CI = 0.24–0.81) [17]. Our finding of no evidence of a causal association between SCZ and RA thus conflict with this line of evidence. A possible explanation is that causal evidence from observational studies could be highly prone to affect by confounding factors, considering that the mean age of onset for RA is usually later in life than SCZ and comorbid medical conditions are tend to be underreporting in patients with SCZ [18, 19]. This may raise concerns about the underdiagnosis of RA in patients with SCZ. Our findings strongly suggest that there is a need to evaluate and be aware of comorbid medical conditions in patients with SCZ.

Various clinical studies have been conducted in the past to study the co-occurrence of BD and RA [19, 21, 41]. Nevertheless, the association between BD and RA remained unclear and debatable. A previous Swedish register study in a sample of 3,798 patients with BD and 6,485 controls has emphasized a greater risk of RA in BD than in controls [19]. Similar trends were observed in Danish and Taiwan register studies [21, 41]. On the other hand, a large Swedish Population Register study found a similar incidence of RA in patients with BD [18]. In this register study, the authors included 34,744 patients with BD and showed the comparable relative risks of developing RA for patients with BD (HR= 1.01, 95% CI = 0.96–1.06). Moreover, a nationally representative longitudinal study from the Netherlands showed that having any mood or any anxiety disorder did not predict new-onset arthritis [42]. However, several inevitable confounding factors in observational studies are not adequately addressed in the literature. In the present study, we aimed to clarify the causal association between BD and RA. Therefore, our findings showed novel and substantial evidence for an inverse relationship between BD and the likelihood of developing RA. Even though our analysis shows that patients
with BD had a lower chance of acquiring RA, the reasons behind this protective effect are yet unclear. However, we acknowledge that there are considerable studies to investigate the relationship between autoimmune disorders and severe psychiatric conditions, including BD and SCZ [43-46]. Immunity and gene-environment interactions represent a major risk factor for these disorders [47-50]. The previous research has highlighted the major histocompatibility complex (MHC) as one of the best validated genetic susceptibility loci for major psychiatric and autoimmune disorders [44, 51, 52]. A growing body of evidence also suggested MHC involvement in SCZ were more pronounced when compared to BD [51]. Interestingly, some MHC loci have been revealed to contribute predisposition to certain autoimmune conditions while others are protective [44]. Indeed, it is important to highlight the fact that the prevalence of comorbid autoimmune conditions in BD and SCZ differs significantly [19]. Taken together, this may be the reason why our findings demonstrated a difference in the pattern of association between BD, SCZ and RA.

The present study has several strengths. We used the recently accessible and analyzable latest BD GWAS data providing confidence in MR assumptions to estimate the causality. All participants in GWAS data were of European ancestry, which facilitated the elimination of confoundings stemming from different races. Furthermore, we observed no horizontal pleiotropy, that is, the outcome directly affected by IVs of exposure without confounding. This provides robust causality of BD on RA and eases the interpretation of the genetic basis of the relationship. A potential limitation of our study is that in the direction of BD to RA, effect size estimates had borderline significance, which could be due to relatively limited sample size and/or low causal relationship. Also, we postulated a linear relationship between exposure and outcome, but using summary data and excluding other interactions (e.g., gene-gene and gene-environment) may lead to deviations from this linearity. Eventually, the clumping algorithm identifying correlated genetic variants depended primarily on European samples. This is due to the lack of corresponding data for other different populations. Namely, we highlight that as this study examines solely upon effects within a specific population, the results cannot be immediately generalized to other populations without further justification.

In conclusion, we did not find an inverse causal association between SCZ and RA contrary to the findings from observational studies. At this stage of understanding, we emphasize that there is a still need of assessing and being aware of symptoms of arthritis in patients with SCZ. On the other hand, our study demonstrated that BD reduces the risk of RA
in the European-ancestry population. Further studies are necessary to explore the biological mechanisms underlying this genetically predicted relationship.

**Ethics approval and consent to participate**
The present study used only publicly available summary-level statistics. No individual-level data was analyzed. Ethical approval is therefore not required.

**Data accessibility and code availability**
Only publicly available data were used in this study. Data sources and handling of these data are described in the Materials and methods. The codes can be found in R package TwoSampleMR.

**Competing interests**
The authors declare that they have no conflicts of interest with the contents of this article.

**Funding**
Not applicable

**Acknowledgements**
The authors thank all individuals who shared GWAS summary statistics.
References


Bergen S, O'dushlaine C, Ripke S, et al. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Molecular psychiatry 2012;17(9):880-6.


Figure legend

Figure 1.

A. Scatter plot of single nucleotide polymorphism (SNP) potential effects on bipolar disorder versus rheumatoid arthritis. The 95% CI for the effect size on rheumatoid arthritis is shown as vertical lines, while the 95% CI for the effect size on bipolar disorder is shown as horizontal lines. The slope of fitted lines represents the estimated Mendelian randomization effect per method. B. Funnel plot for bipolar disorder shows the estimation using the inverse of the standard error of the causal estimate with each individual SNP as a tool. The vertical line represents the estimated causal effect obtained using IVW and MR-Egger methods. C. Scatter plot of SNP potential effects on schizophrenia versus rheumatoid arthritis D. Funnel plot for schizophrenia shows the estimation using the inverse of the standard error of the causal estimate with each individual SNP as a tool.
Table 1. Two-sample Mendelian Randomization analysis of the effects of bipolar disorder, schizophrenia, and rheumatoid arthritis using summary-level data.

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Outcome</th>
<th>SNPs, n&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Method</th>
<th>OR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar disorder</td>
<td>Rheumatoid arthritis</td>
<td>48</td>
<td>MR Egger</td>
<td>0.652 (0.296-1.434)</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted median</td>
<td>0.869 (0.764-0.989)</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IVW</td>
<td>0.810 (0.689-0.953)</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Simple mode</td>
<td>0.905 (0.686-1.195)</td>
<td>0.488</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted mode</td>
<td>0.879 (0.678-1.139)</td>
<td>0.336</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Rheumatoid arthritis</td>
<td>52</td>
<td>MR Egger</td>
<td>1.122 (0.746-1.687)</td>
<td>0.581</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted median</td>
<td>1.007 (0.908-1.115)</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IVW</td>
<td>1.008 (0.931-1.092)</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Simple mode</td>
<td>1.008 (0.778-1.305)</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weighted mode</td>
<td>1.031 (0.797-1.334)</td>
<td>0.814</td>
</tr>
</tbody>
</table>

SNP, single nucleotide polymorphism; IVW, inverse-variance weighted; OR, odds ratio; CI, confidence interval.

<sup>a</sup> The number of instrumental SNPs used in each analysis after harmonization of data filtering palindromic SNPs.
Figure 1.