General Psychiatry

Why caution should be applied when interpreting and promoting findings from Mendelian randomisation studies

Alice R Carter,[1,2] Abigail Fraser,[1,2] Laura D Howe,[1,2] Sian Harris ,[1] Amanda Hughes [1,2]

To cite: Carter AR, Fraser A, Howe LD, *et al*. Why caution should be applied when interpreting and promoting findings from Mendelian randomisation studies. *General Psychiatry* 2023;**36**:e101047. doi:10.1136/gpsych-2023-101047

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gpsych-2023-101047).

Received 01 March 2023
Accepted 19 July 2023

Check for updates

© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY. Published by BMJ.

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[2]Population Health Sciences, University of Bristol, Bristol, UK

Correspondence to
Dr Amanda Hughes;
amanda.hughes@bristol.ac.uk

INTRODUCTION

In their article entitled 'Investigating genetic causal relationships between blood pressure and anxiety, depressive symptoms, neuroticism and subjective well-being', Cai and colleagues[1] presented the results of a two-sample Mendelian randomisation[2] (MR) study examining associations between blood pressure traits (systolic, diastolic, hypertension and pulse pressure) and psychological traits (anxiety, depression, neuroticism and subjective well-being). After correction for multiple testing, the authors reported a small but statistically significant result for the effect of diastolic blood pressure (DBP) on neuroticism (an increase of 0.0036 standard deviation (SD) per mm Hg increase in DBP), which they argued demonstrated a causal relationship. All other results were null. However, key aspects of their approach mean that caution is warranted regarding their conclusion. Here we attempt to replicate their results, highlighting key issues with how these methods have been implemented. We also consider the wider limitations of MR approaches and discuss the need for caution when interpreting MR results.

Summary data-based MR (often also referred to as two-sample MR) uses summary statistics from genome-wide association studies (GWAS) to estimate the association between genetic variants and exposures of interest on the one hand, and between the same genetic variants and outcomes on the other, to explore whether associations between exposures and outcomes are likely to reflect a causal effect. Because genetic variants are determined before birth and do not change during a person's lifetime, MR studies are less impacted by classical confounding and reverse causation than non-genetic observational studies. An overview of MR methods and their assumptions is given in online supplemental box 1.

Crucially, Cai *et al* described their analysis as a bidirectional MR study: one that investigated two possible causal directions of an association and compared the evidence for each. However, the authors only carried out analysis for the impact of blood pressure traits on psychological outcomes, not the converse. This has important implications for the interpretation of results, which are presented as supporting a causal effect of DBP on neuroticism but not vice versa. Interestingly, the authors did not examine the reverse causal path, deciding to run analyses only where 10 or more genetic variants were independently associated with the exposure with an F-statistic of >10. This is an unusual decision given that methods exist for MR using a single genetic variant[3] (we apply these later in this commentary).

Cai and colleagues applied generalised summary data-based MR (GSMR)[4] with genetic associations drawn from both peer-reviewed GWAS and online databases whose contents may have undergone less stringent quality control (see Cai *et al*, table 1). Estimates are calculated using the heterogeneity in dependent instruments (HEIDI) outlier test to minimise the impact of horizontal pleiotropy. This refers to the situation in which the exposure-associated genetic variants are associated with the outcome via pathways independent of the exposure, which can inflate causal effects in MR. Nevertheless, they rely on a single method to infer causality rather than presenting estimates from a range of MR methods, each with different assumptions and limitations.

An absence of horizontal pleiotropy is one of the three core assumptions of MR; however, there are several further sources of bias relevant to MR studies.[2] In Cai *et al*'s study, the sample populations used to derive associations of genetic variants with exposures and

Gen Psych: first published as 10.1136/gpsych-2023-101047 on 10 August 2023. Downloaded from http://gpsych.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

Table 1 Number of SNPs and F-statistics for exposures

Phenotype	Clumping threshold (R^2)	Maximum number of SNPs associated with exposure at $p<5\times10^{-8}$*	Mean F-statistic across all SNPs
Diastolic blood pressure	<0.05	1145	60.42
	<0.001	455	79.42
Systolic blood pressure	<0.05	1063	59.11
	<0.001	456	75.13
Pulse pressure	<0.05	1445	59.13
	<0.001	484	81.79
Hypertension	<0.05	88	46.00
	<0.001	72	47.97
Anxiety	<0.001	0	NA
Depressive symptoms	<0.001	2	38.75
Neuroticism	<0.001	10	38.45
Subjective well-being	<0.001	1	27.56

*For some analyses, not all these SNPs were used. This was where not all the SNPs were available in, or could not be harmonised with, the outcome GWAS. For the number of SNPs used in each analysis, see online supplemental table 1.
GWAS, genome-wide association study; NA, not available; SNP, single nucleotide polymorphism.

outcomes contain many of the same participants. Sample overlap between the exposure and outcome GWAS can lead to overfitting bias in MR, where MR estimates are exaggerated away from the null.[5] Although recent methodological work suggests bias due to sample overlap is minimal when strong instruments are used (ie, F-statistics >10), this potential source of bias should still be acknowledged where overlapping samples are used, particularly where methods are available to estimate the size of and correct for any bias.[6] This is especially pertinent when a single large study (eg, UK Biobank) contributes heavily to both the exposure and outcome GWAS.

In the GWAS used by Cai *et al*, blood pressure was adjusted for body mass index (BMI) to increase statistical power. However, the adjustment for a covariate can bias the GWAS results and subsequent MR estimates that are based on the covariate-adjusted GWAS estimates.[7] Moreover, for MR, genetic instruments should be strongly associated with the exposure (a genome-wide significant p value threshold of $<5\times10^{-8}$) and independent of one another. Typically, independent instruments are identified using a linkage disequilibrium (LD) clumping threshold of $R^2<0.001$, but Cai *et al* used an unusually liberal threshold of $R^2<0.05$. This can overestimate the precision of results, leading to type 1 errors (false positive results). A further technical challenge is the use of binary exposure variables formed by dichotomising an underlying continuous distribution which can violate[8] a key assumption of MR. For example, hypertension is used as an exposure; however, hypertension is a dichotomisation of two continuous variables: systolic blood pressure (SBP) and DBP.

While the authors do consider pleiotropy as a non-causal explanation for associations, they do not acknowledge or discuss the potential impact of demographic or family-level processes[9] that can bias MR estimates based on genetic associations calculated in samples of unrelated individuals.[10] Lastly, the authors draw causal conclusions from results based on a single MR method. MR approaches are imperfect, and, for this reason, conclusions are more robust when based on results from several methods considered together,[11] as recommended in the MR-Strengthening the Reporting of Observational Studies in Epidemiology guidelines.[12] If GSMR supports much stronger causal claims than other genetic approaches when used on its own, the authors do not clarify why this would be so.

REPLICATION: METHODS

Here we attempt to replicate Cai *et al*'s analyses. Unfortunately, code was not provided to allow identical replication. Increasingly, it is considered standard practice for researchers to share code used in the analysis,[13] and not doing so may compound limitations of original work by making reproducibility harder. Using the same sources as Cai *et al* means some of the same issues will impact our replication : adjustment for BMI in blood pressure GWAS, overlap of sample populations used to derive exposure and outcome associations and bias due to family-level processes. However, it allows a clearer exploration of how other decisions by the authors impact their results. For three of the four psychological traits (depressive symptoms, neuroticism, subjective well-being (SWB)), we investigate both causal directions: the impact of blood pressure on psychological traits and the impact of psychological traits on blood pressure (this is not possible for anxiety, with which no genetic variants are associated at the conventional significance threshold ($p<5\times10^{-8}$)). Where possible, we apply several widely used

Gen Psych: first published as 10.1136/gpsych-2023-101047 on 10 August 2023. Downloaded from http://gpsych.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

and complementary MR methods: the inverse variance-weighted, weighted mode and MR-Egger methods. Where there are too few independent genetic associations to apply these methods, we calculate Wald ratios. For the impact of blood pressure traits on neuroticism, we explore how the threshold used to identify independent genetic associations with blood pressure affects results. We use an LD clumping threshold of R^2<0.001 and a distance of 10 000 kb (incidentally, the default settings provided in the TwoSampleMR R package[14]). In all analyses, we had suitable instrument strength (mean F>10), which notably was higher for psychological traits when a more stringent LD clumping threshold was used (table 1) than that applied by Cai *et al*. Using the same liberal LD threshold as Cai *et al*, we had a similar number of single nucleotide polymorphisms (SNPs) for analysis of blood pressure traits as the exposure.

Analyses were conducted using R V.4.2.2. We used the TwoSampleMR package, which includes harmonisation of exposure and outcome SNPs, using default parameters. For graphs, we used Stata V.17. We have made all analysis codes available at https://github.com/alicerosecarter/BP_PsychologicalTraits_MR_Replication to allow reproducibility and replication of our analyses.

REPLICATION: RESULTS

Our results provide little evidence that any blood pressure trait affects any psychological trait (figure 1, online supplemental figures 1–3). Cai *et al* reported a 0.0036 SD increase in neuroticism (SE: 0.0008) per mm Hg increase in genetically instrumented DBP. Our primary analysis found an equivalent association of 0.001 SD (SE: 0.0014). We were unable to replicate the authors' support for the impact of DBP on neuroticism using inverse variance-weighted, weighted mode and MR-Egger methods with a conventional threshold (R^2<0.001) for independent genetic instruments for the exposure. However, applying the same liberal threshold as in the original paper (R^2<0.05) resulted in evidence of an effect of DBP on neuroticism using the inverse variance-weighted method, although this would not survive correction for multiple testing (p=0.042). For example, a stringent Bonferroni-corrected p value threshold would be 0.05/32=0.002 (where 32 is the number of exposure–outcome combinations considered).

In the second half of our bidirectional analysis, psychological traits were treated as the exposures and blood pressure traits as the outcome. Here, results of the Wald ratio method suggested an impact of lower SWB on increasing SBP and hypertension (eg, mean difference in mm Hg of SBP per unit increase in SWB=7.87 mm Hg; p<0.001). However, only a single SNP was available for SWB, and sensitivity analyses using alternative MR methods could not be conducted. We therefore urge against strong causal interpretations based on this result. Should a larger GWAS of SWB be conducted, further analyses to see if this estimate remains robust to conventional MR assumptions may become possible.

REPLICATION: INTERPRETATION

Our analysis, which applied multiple MR methods to examine the impact of blood pressure traits on psychological outcomes and vice versa, finds little evidence for an impact of blood pressure on any psychological outcome. The association reported by Cai *et al* and presented as evidence of a causal impact of DBP on neuroticism could only be replicated with one of four approaches applied and only when using a liberal threshold to identify SNPs from the GWAS to use in analyses—an approach likely to underestimate SEs. This suggests that the previously reported result should be regarded cautiously, as it may be a type 1 error (false positive). On the other hand, our results are consistent with a causal impact of SWB on both hypertension and DBP. However, we were only able to apply one method (the Wald ratio method), which limits the strength of conclusions that can be drawn.

Our analysis has important strengths relative to the original analysis: its application of multiple MR methods, use of more stringent thresholds for identifying independent instruments and bidirectional analysis. At the same time, other limitations of Cai *et al*'s study apply equally to our results because we used genetic associations from the same sources. Besides those noted by Cai and colleagues (residual effects of pleiotropy and limited generalisability beyond European populations), this includes three limitations that were not discussed in the original paper: the overlap of sample populations between exposure and outcome GWAS, the impact of using a blood pressure GWAS that was adjusted for BMI and the possibility that GWAS captures heritable genetic effects and intergenerational non-genetic ones.

INTERPRETATION OF CAUSAL EFFECTS

Considering the strong causal claims made by Cai *et al*, we were surprised by how little discussion was given to possible biological mechanisms. Instead, they discussed how psychological traits could affect blood pressure (the causal direction which their analyses did not test). Such mechanisms include sympathetic overactivity and autonomic dysfunction, blood pressure-elevating effects of medications used to treat anxiety and other coping strategies (eg, smoking and alcohol intake). The biological mechanisms leading to a potential causal effect of blood pressure on psychological traits are perhaps less clear.

MR IN THE MEDIA

Cai *et al*'s article received widespread media coverage. As of 20 July 2023, Altmetric had detected 98 news stories about this paper from 87 outlets, with accounts of this research published in French, German, Croatian and Spanish, as well as English. The article has been cited

Gen Psych: first published as 10.1136/gpsych-2023-101047 on 10 August 2023. Downloaded from http://gpsych.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

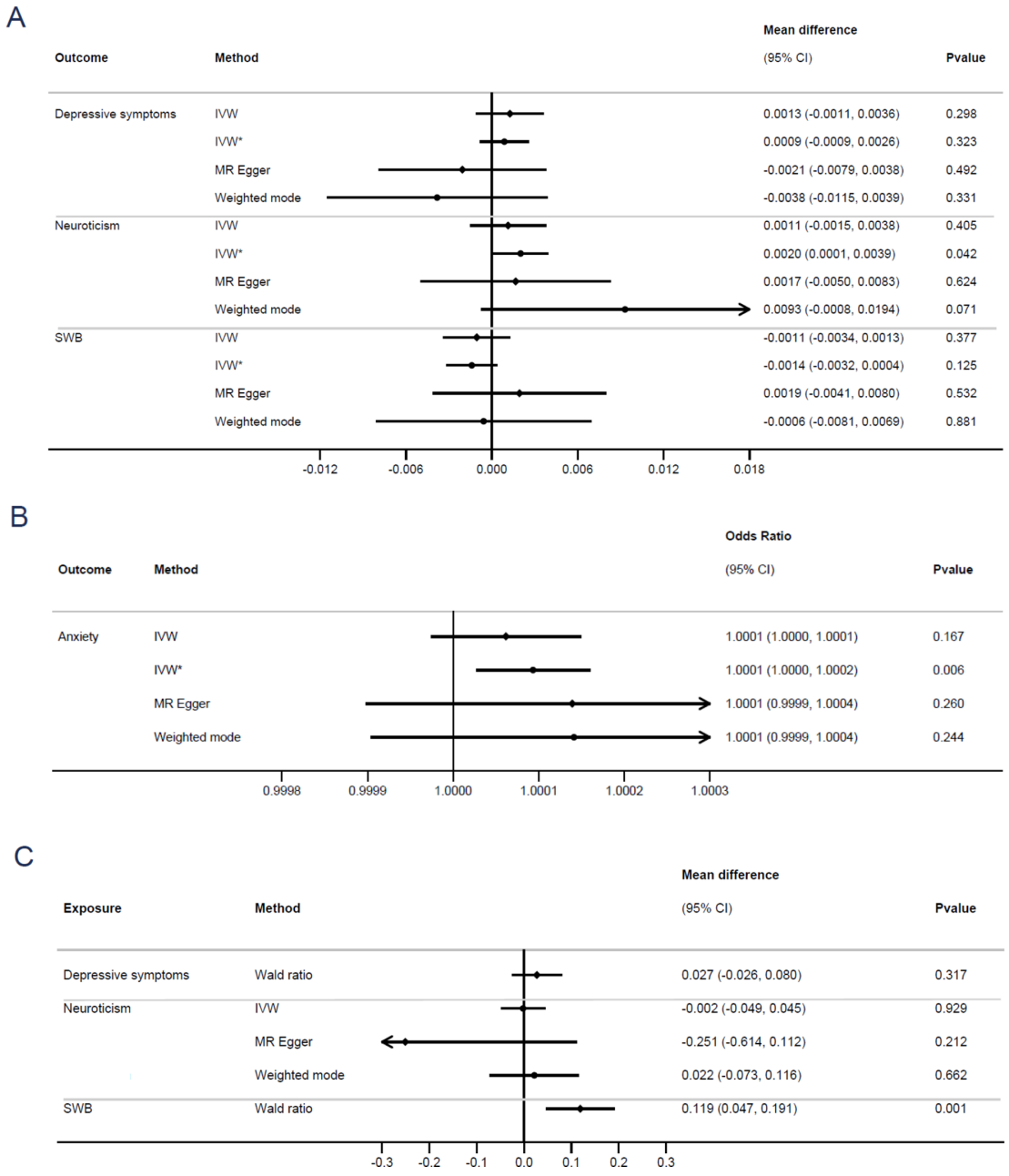


Figure 1 The bidirectional association between DBP and anxiety, depressive symptoms, neuroticism and SWB. (A) Results with DBP as the exposure for continuous outcomes on the mean difference scale. (B) Results with DBP as the exposure for dichotomous outcomes on the odds ratio scale. (C) Results with DBP as the outcome on the mean difference scale. P values have not been corrected for multiple testing. CI,confidence interval; DBP, diastolic blood pressure; IVW, inverse variance-weighted method, LD clumping threshold of <0.001; IVW*, inverse variance-weighted method, LD clumping threshold of <0.05 as used by Cai *et al*; LD, linkage disequilibrium; MR, Mendelian randomisation; SWB, subjective well-being.

in mainstream newspapers, lifestyle publications and specialist publications aimed at clinicians.

The strength of MR studies is that they are less vulnerable to the types of confounding and reverse causality bias which affect non-genetic studies, but they also have limitations which means strong causal conclusions are not always warranted. This can be difficult to convey to non-specialist audiences, but doing so is crucial. Media reports of the study repeated the claim that DBP has a causal effect on neuroticism and that actions to control blood pressure could reduce neuroticism and self-criticism. Of course, controlling blood pressure has other benefits, and such advice is perhaps unlikely to be harmful. But this may not always be the case.

For this reason, it is crucial that MR studies clearly acknowledge and responsibly communicate the nuance of their findings and their limitations. It is also important that researchers and press teams work together to communicate not only the findings of studies but also their limitations and potential sources of false positives and other errors when deciding whether or not to produce a press release or its specific wording. This is particularly

Gen Psych: first published as 10.1136/gpsych-2023-101047 on 10 August 2023. Downloaded from http://gpsych.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

important in communicating health-related information, for in addition to the public-facing coverage, articles in specialist magazines and blogs are important tools to provide updates to busy clinicians.

Twitter Alice R Carter @alicerosecarter, Sian Harris @sianharris8 and Amanda Hughes @ammhughes

Contributors ARC and AH conceptualised the study and wrote the manuscript. ARC conducted all analyses. AF, LDH and SH made critical revisions to the manuscript.

Funding This study was funded by the Medical Research Council (MC_UU_00011/6, MC_UU_00011/1).

Competing interests ARC is employed by Novo Nordisk Research Centre Oxford (NNRCO). The work presented in this manuscript was conducted prior to employment by NNRCO, and NNRCO were in no way involved in the analysis or decision to publish the manuscript.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs
Sian Harris http://orcid.org/0000-0001-5737-3551
Amanda Hughes http://orcid.org/0000-0001-5896-7650

REFERENCES

1 Cai L, Liu Y, He L. Investigating genetic causal relationships between blood pressure and anxiety, depressive symptoms, neuroticism and subjective well-being. *Gen Psychiatr* 2022;35:e100877.
2 Sanderson E, Glymour MM, Holmes MV, *et al*. Mendelian randomization. *Nat Rev Methods Primers* 2022;2:6.
3 Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for mendelian randomization. *Stat Methods Med Res* 2017;26:2333–55.
4 Zhu Z, Zheng Z, Zhang F, *et al*. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;9:224.
5 Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample mendelian randomization. *Genet Epidemiol* 2016;40:597–608.
6 Mounier N, Kutalik Z. Bias correction for inverse variance weighting mendelian randomization. *Genet Epidemiol* 2023;47:314–31.
7 Gilbody J, Borges MC, Smith GD, *et al*. Multivariable MR can mitigate bias in two-sample MR using covariable-adjusted summary associations. *Epidemiology* [Preprint].
8 Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018;33:947–52.
9 Morris TT, Davies NM, Hemani G, *et al*. Population phenomena inflate genetic associations of complex social traits. *Sci Adv* 2020;6:eaay0328.
10 Brumpton B, Sanderson E, Heilbron K, *et al*. Avoiding dynastic, Assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun* 2020;11:3519.
11 Gagliano Taliun SA, Evans DM. Ten simple rules for conducting a mendelian randomization study. *PLoS Comput Biol* 2021;17:e1009238.
12 Skrivankova VW, Richmond RC, Woolf BAR, *et al*. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* 2021;375:n2233.
13 Baker M. Why scientists must share their research code. *Nature* 2016. 10.1038/nature.2016.20504 Available: https://doi.org/10.1038/nature.2016.20504
14 Hemani G, Zheng J, Elsworth B, *et al*. The MR-base platform supports systematic causal inference across the human Phenome. *eLife* 2018;7:e34408.

Alice R Carter is a Senior Scientist in Mendelian randomization at Novo Nordisk Research Centre Oxford, and an honorary Research Fellow at the MRC Integrative Epidemiology Unit, University of Bristol, the UK. Prior to joining Novo Nordisk in 2023, she was a postdoctoral researcher at the Medical Research Council Integrative Epidemiology Unit, University of Bristol, the UK. Dr Carter's applied work has centred around understanding the causes and consequences of socioeconomic inequalities in cardiovascular disease. Additionally, Dr Carter has been involved in the development of advanced Mendelian randomisation methods and pioneered the application of using Mendelian randomisation to conduct mediation analysis. Dr Carter has multiple highly cited publications on this topic along with numerous publications relating to Mendelian randomisation methods and mitigating bias in Mendelian randomisation. She received her PhD from the University of Bristol in 2020, and her MSc from the London School of Hygiene and Tropical Medicine in 2016.

Gen Psych: first published as 10.1136/gpsych-2023-101047 on 10 August 2023. Downloaded from http://gpsych.bmj.com/ on April 27, 2024 by guest. Protected by copyright.

Box 1: Overview of Mendelian randomization (MR) assumptions and methods[2]

Main assumptions of MR

MR has three core assumptions:

1. Relevance assumption: The genetic variant(s) are strongly associated with the exposure of interest.
2. Independence assumption: There are no (unmeasured) confounders between the genetic variants and outcomes of interest.
3. Exclusion restriction criteria (no pleiotropy): There is no pathway between the genetic variant(s) and the outcome other than via the exposure of interest.

Additional, more nuanced assumptions apply to the interpretation of MR analyses. This includes that genetic variants act equally in all individuals (instrument homogeneity), and that genetic variants are truly randomly allocated across the population. This is not always true (e.g., due to ancestral clustering of genetic variants). Consideration should be given to core and advanced assumptions before making strong causal claims.

Common statistical approaches to MR

In the case of summary data MR with multiple genetic variants, typically the "main" MR effect will be estimated with a fixed or random-effects inverse variance weighted method. This combines Wald ratios for individual SNP-outcome effects, giving an overall MR estimate. This method is statistically powerful but can be biased in the presence of pleiotropy.

Pleiotropy robust methods: methods have been developed to account for pleiotropic SNPs, including MR-Egger, the weighted median and weighted mode estimators, MR-PRESSO and MR-RAPs. These make different assumptions about the pleiotropic effects of the genetic variants. A combination of these methods, each giving consistent estimates, allows the greatest confidence that results do not only reflect pleiotropy.

Multivariable MR: this allows the joint effects of two or more exposure to be estimated simultaneously. Multivariable MR can be used, for example, to account for a known pleiotropic pathway or to explore mediated effects.

MR is a rapidly evolving field, with new methodological approaches frequently becoming available. Since all have slightly different assumptions, data requirements, strengths and weaknesses, the most appropriate strategy is the application of multiple methods to test the robustness of findings to these differences.

A more extensive introduction to MR assumptions and methods can be found here.

Supplementary Table 1: Number of SNPs and mean F statistics for each individual analysis for A) blood pressure traits as exposures and B) blood pressure traits as outcomes

Supplementary Table 1a: Number of SNPs and mean F statistics for each individual analysis for blood pressure traits as exposures

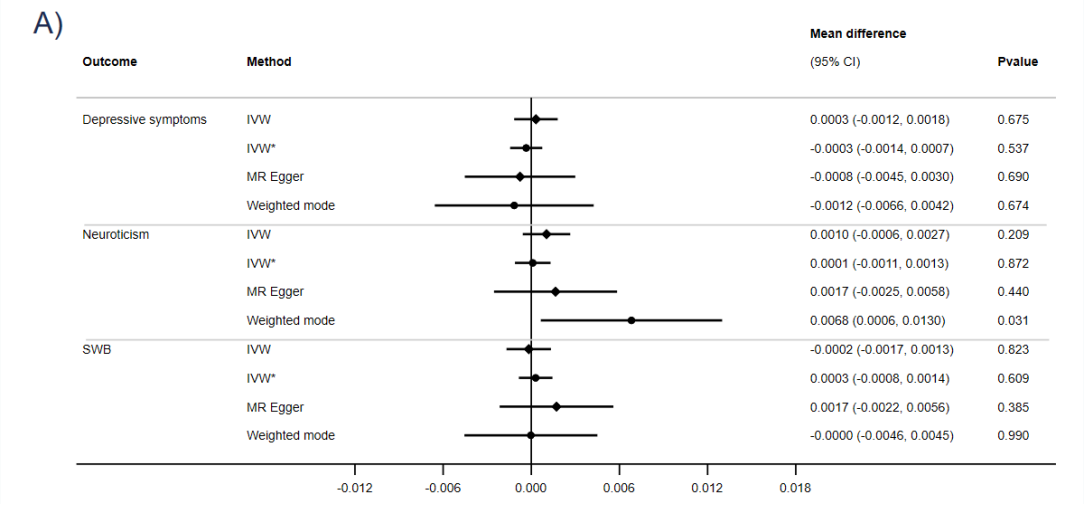
Exposure	**Clumping threshold (r^2)**	**Outcome**							
		Anxiety		**Depressive symptoms**		**Neuroticism**		**Subjective wellbeing**	
		Number of SNPs used in analysis	Mean F statistic	Number of SNPs used in analysis	Mean F statistic	Number of SNPs used in analysis	Mean F statistic	Number of SNPs used in analysis	Mean F statistic
Diastolic blood pressure	<0.05	631	63.61	1041	60.94	1041	60.94	751	61.66
	<0.001	291	80.01	433	80.05	433	80.05	336	81.24
Systolic blood pressure	<0.05	620	60.26	978	59.58	978	59.58	710	60.45
	<0.001	293	74.38	431	75.16	431	75.16	326	76.66
Pulse pressure	<0.05	533	59.82	975	59.70	975	59.67	748	60.41
	<0.001	212	79.94	356	82.28	355	82.38	292	81.55
Hypertension	<0.05	53	48.52	81	46.61	81	46.61	53	48.72
	<0.001	48	49.93	66	48.66	66	48.66	48	50.77

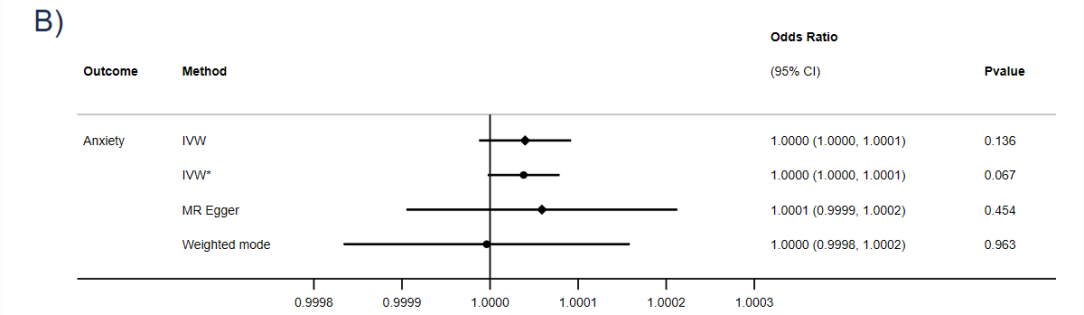
Supplementary Table 1b: Number of SNPs and mean F statistics for each individual analysis for blood pressure traits as outcomes

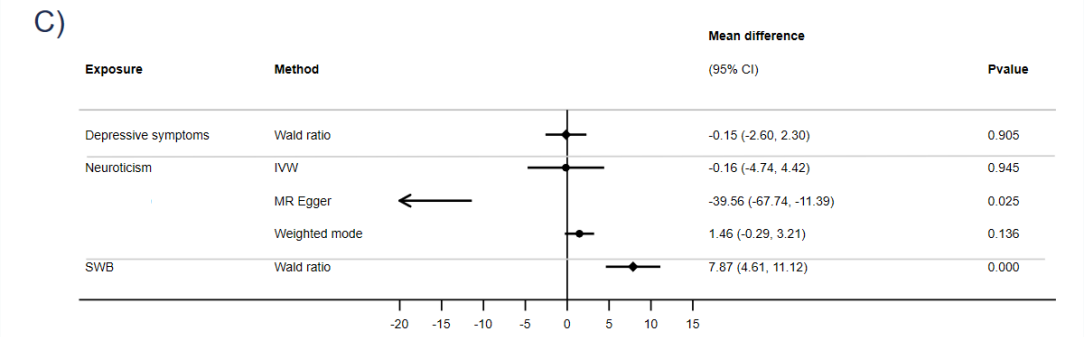
Exposure	Outcome							
	Diastolic blood pressure		Systolic blood pressure		Pulse pressure		Hypertension	
	Number of SNPs used in analysis	Mean F statistic	Number of SNPs used in analysis	Mean F statistic	Number of SNPs used in analysis	Mean F statistic	Number of SNPs used in analysis	Mean F statistic
Anxiety	0	NA	0	NA	0	NA	0	NA
Depressive symptoms	1	38.45	1	38.45	1	38.45	1	38.45
Neuroticism	10	38.45	10	38.45	10	38.45	10	38.45
Subjective wellbeing	1	27.56	1	27.56	1	27.56	1	27.56

Supplementary Figure 1: The bidirectional association between systolic blood pressure (SBP) and anxiety, depressive symptoms, neuroticism and subjective wellbeing. Panel A presents results with SBP as the exposure for continuous outcomes, panel B presents results with SBP as the exposure for dichotomous outcomes, and panel C with SBP as the outcome.

Footnote: IVW = inverse variance weighted method

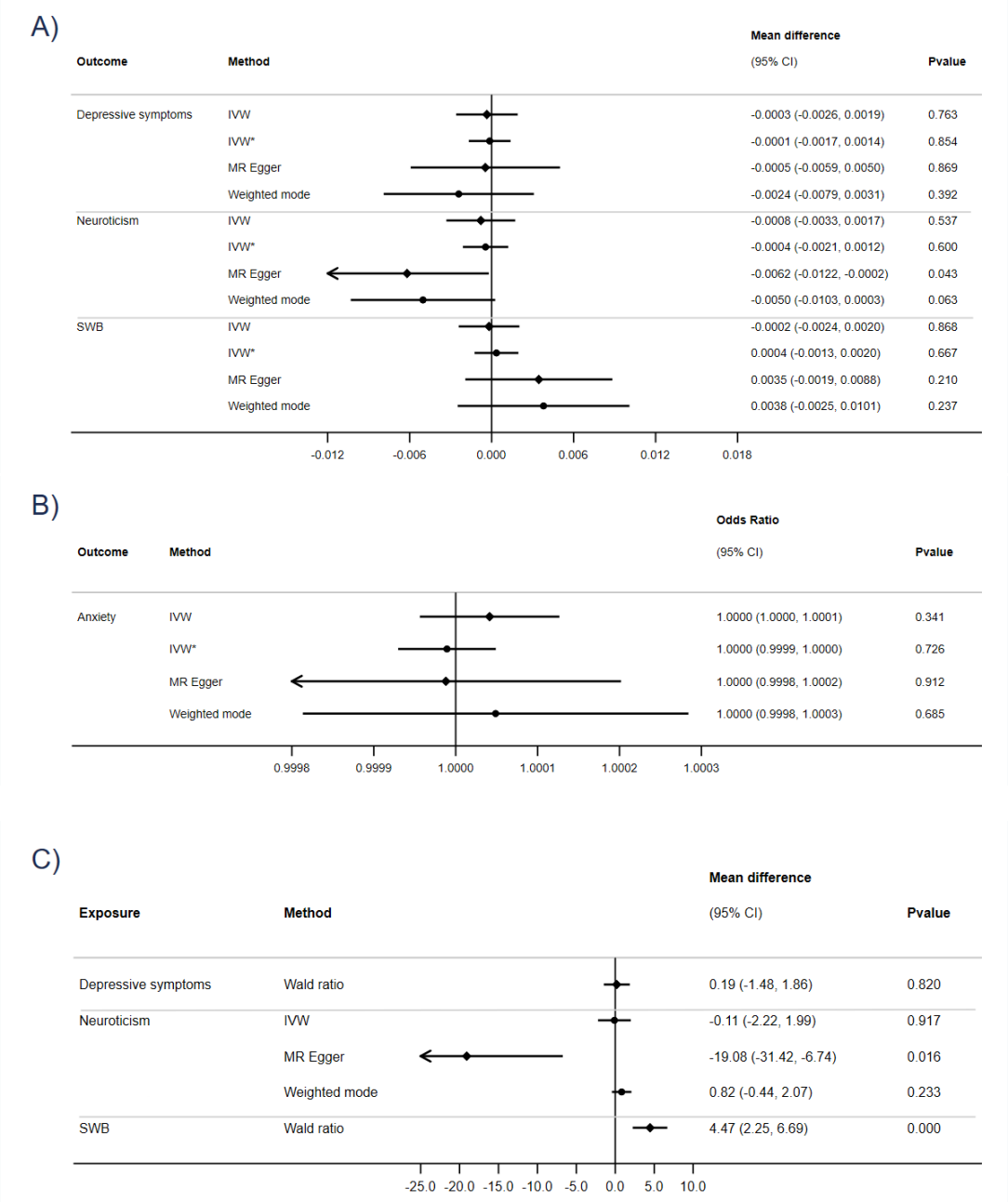


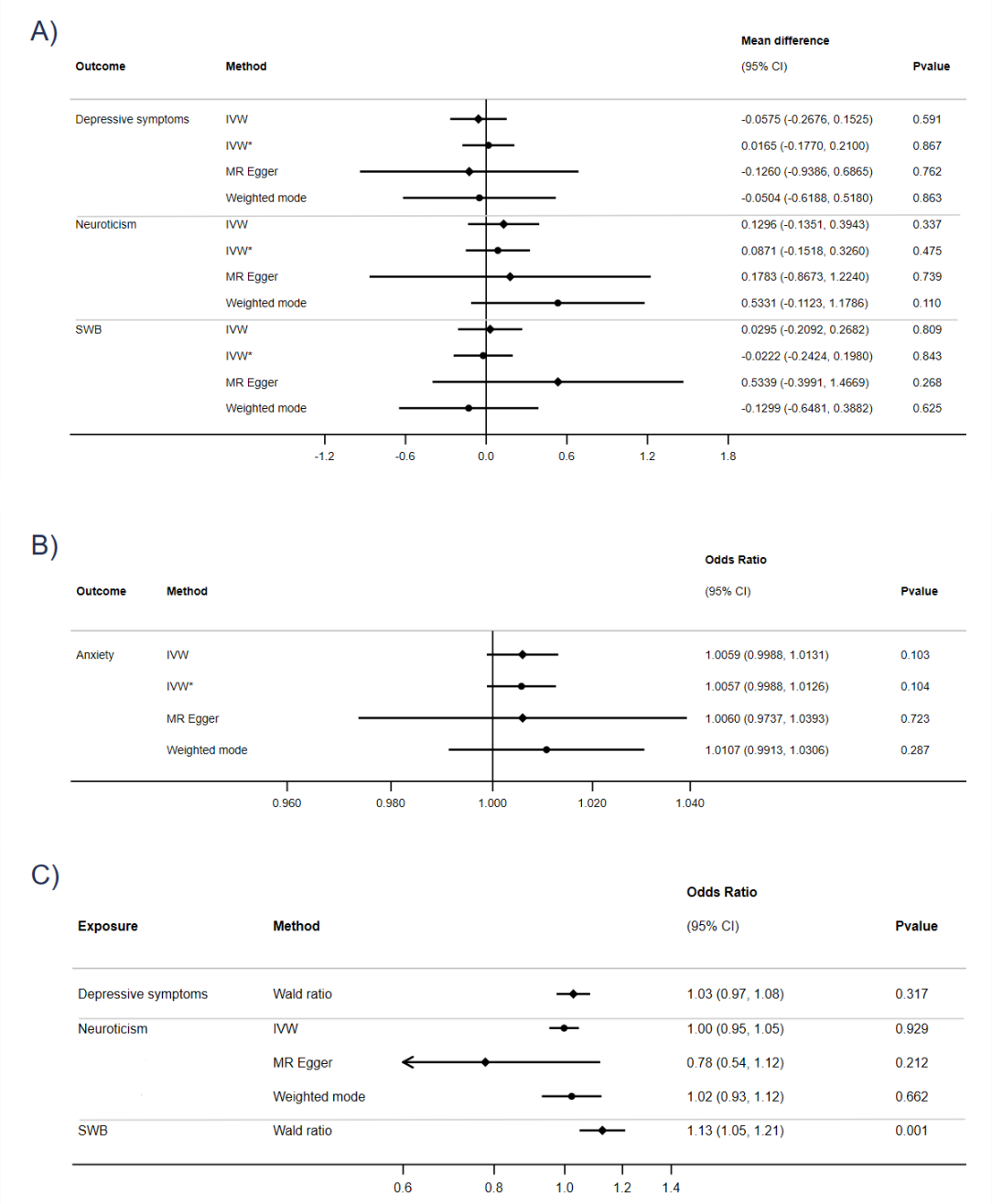




Supplementary Figure 2: The bidirectional association between pulse pressure (PP) and anxiety, depressive symptoms, neuroticism and subjective wellbeing. Panel A presents results with PP as the exposure for continuous outcomes, panel B presents results with PP as the exposure for dichotomous outcomes, and panel C with PP as the outcome.

Footnote: IVW = inverse variance weighted method



Supplementary Figure 3: The bidirectional association between hypertension and anxiety, depressive symptoms, neuroticism and subjective wellbeing. Panel A presents results with hypertension as the exposure for continuous outcomes, panel B presents results with hypertension as the exposure for dichotomous outcomes, and panel C with hypertension as the outcome.

Footnote: IVW = inverse variance weighted method