4. Results

Contents

Results **1**

Simulation Study . . . 1

Data Simulation . . . 1

Analysis of Simulated Data . . . 2

No Causal Effect . . . 2

Positive Causal Effect . . . 2

Simulation Tables . . . 4

Citations Search . . . 6

Search Results . . . 6

Re-analysis Results . . . 8

Re-analysis vs Reported WME Causal Estimates . . . 8

Re-analysis WME vs MR-Hevo Causal Estimates . . . 8

Word count: 877

Results

Simulation Study

Data Simulation

Data were successfully simulated as intended. A selection of representative visualisations are presented in Figure 1. Full details of testing used to validate model outputs from parameter inputs are given in Appendix **??**. The F-statistic calculated from simulated instruments was >10, indicating that they were sufficiently strongly associated with exposure to meet the relevance assumption of instrumental variable (IV) analysis (Tables 1 and 2).

Analysis of Simulated Data

No Causal Effect

Across all cases where no causal effect was present (Table 1), the mean rate of reporting a causal effect (i.e. false positive rate) for MR-Hevo was 0.64% versus 2% for weighted median estimator (WME). Of the 8 combinations of scenarios and parameters, MR-Hevo exhibited a favourable false positive rate versus WME in 8 (100%). For both MR-Hevo and WME methods, causal report rate generally increased with an increasing proportion of invalid instruments, and also as assumption violations progressively presented greater data variability and bias across scenarios 1 to 3.

The mean causal effect estimate (mean reported 95% confidence interval (CI)s) across all cases was 0.044 (-0.13 to 0.23) for MR-Hevo and 0.025 (-0.15 to 0.2) for WME. The mean (range) standard error (SE) of causal effect estimates across all cases was 0.0015 (0.001 to 0.002) for MR-Hevo and 0.088 (0.078 to 0.094). For both MR-Hevo and WME methods, causal effect estimates, width of CIs and SE all tended to increase slightly both with an increasing proportion of invalid instruments, and as assumption violations progressively presented greater data variability and bias across scenarios 1 to 3.

Positive Causal Effect

Across all cases where positive causal effect was present (Table 2), the mean rate of reporting a causal effect (i.e. sensitivity) for MR-Hevo was 19% versus 19% for WME. Of the 12 combinations of scenarios and parameters, MR-Hevo exhibited a favourable sensitivity versus WME in 5 (42%). For both MR-Hevo and WME methods, causal report rate generally increased with an increasing proportion of invalid instruments, and also as assumption violations progressively presented greater data variability and bias across scenarios 1 to 3.

The mean causal effect estimate (mean reported 95% CIs) across all cases was 0.14 (-0.04 to 0.34) for MR-Hevo and 0.11 (-0.063 to 0.29) for WME. The mean (range) SE of causal effect estimates across all cases was 0.0015 (0.001 to 0.002) for MR-Hevo and 0.09 (0.079 to 0.1). For both MR-Hevo and WME methods, causal effect estimates, width of CIs and SE all tended to increase slightly both with an increasing proportion of invalid instruments, and as assumption violations progressively presented greater data variability and bias across scenarios 1 to 3.

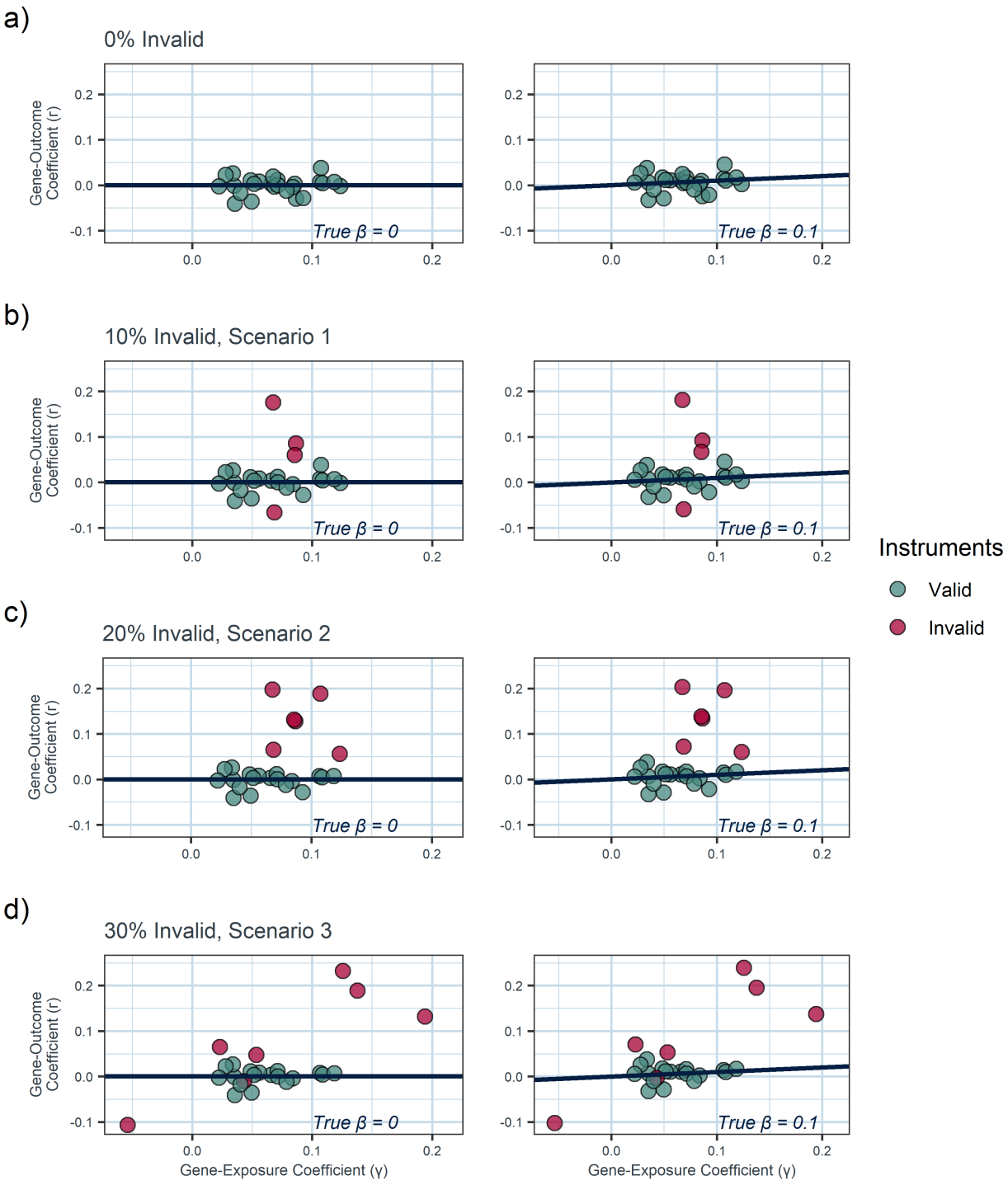

Figure 1: Plots of a representative group of simulated datasets; all simulate genetic instruments from the same index from the same random seed. Left and right columns demonstrate null and positive true causal effects, respectively. The scenario and the proportion of invalid (i.e. pleiotropic) genetic instruments changes with each row. a) 10% of instruments invalid, Scenario 1: balanced pleiotropy introduces noise around the causal effect. b) 20% of instruments invalid, Scenario 2: directional pleiotropy biases in the direction of the invalid instruments. c) 30% of instruments invalid, Scenario 3: directional pleiotropy and InSIDE assumption violation strongly biases towards a positive effect estimate.

Simulation Tables

Table 1: Summary of 1,000 simulated Mendelian randomisation studies per combination of scenario and parameters, all with null causal effect

				Weighted Median			MR Hevo		
N	Invalid IVs	*F*	R^2	Mean Estimate (Mean SE)	Mean 95% CI	Causal Report Rate	Mean Estimate (Mean SE)	Mean 95% CI	Causal Report Rate
Scenario 1: Balanced pleiotropy, InSIDE assumption satisfied									
10,000	0%	11.7	2.8%	0.001 (0.078)	-0.15 to 0.15	0.2%	0.000 (0.001)	-0.12 to 0.12	0%
10,000	10%	11.7	2.8%	0.026 (0.086)	-0.14 to 0.19	1.5%	0.032 (0.001)	-0.13 to 0.2	0%
10,000	20%	11.7	2.8%	0.022 (0.092)	-0.16 to 0.2	2%	0.037 (0.002)	-0.17 to 0.25	0%
10,000	30%	11.7	2.8%	0.014 (0.093)	-0.17 to 0.2	1.6%	0.022 (0.002)	-0.2 to 0.25	0%
20,000	0%	26.2	3.2%	0.003 (0.056)	-0.11 to 0.11	0.3%	0.001 (0)	-0.09 to 0.09	0%
20,000	10%	24.5	3%	0.022 (0.062)	-0.1 to 0.14	0.5%	0.019 (0.001)	-0.1 to 0.14	0.1%
20,000	20%	24.5	3%	0.020 (0.067)	-0.11 to 0.15	1.3%	0.022 (0.001)	-0.13 to 0.18	0%
20,000	30%	24.5	3%	0.012 (0.067)	-0.12 to 0.14	0.8%	0.014 (0.001)	-0.15 to 0.18	0%
Scenario 2: Directional pleiotropy, InSIDE assumption satisfied									
10,000	0%	11.7	2.8%	0.001 (0.078)	-0.15 to 0.15	0.3%	0.000 (0.001)	-0.12 to 0.12	0%
10,000	10%	11.7	2.8%	0.020 (0.087)	-0.15 to 0.19	0.8%	0.039 (0.001)	-0.13 to 0.22	0%
10,000	20%	11.7	2.8%	0.050 (0.093)	-0.13 to 0.23	4.1%	0.098 (0.002)	-0.11 to 0.33	1.5%
10,000	30%	11.7	2.8%	0.066 (0.094)	-0.12 to 0.25	5.8%	0.126 (0.002)	-0.09 to 0.38	3.6%
20,000	0%	24.5	3%	0.004 (0.056)	-0.11 to 0.11	0.2%	0.001 (0)	-0.08 to 0.09	0%
20,000	10%	24.5	3%	0.016 (0.062)	-0.11 to 0.14	0.7%	0.021 (0.001)	-0.1 to 0.15	0.1%
20,000	20%	24.5	3%	0.038 (0.067)	-0.09 to 0.17	2.2%	0.054 (0.001)	-0.1 to 0.22	0.5%
20,000	30%	24.5	3%	0.050 (0.068)	-0.08 to 0.18	4.9%	0.076 (0.002)	-0.08 to 0.25	1.2%
Scenario 3: Directional pleiotropy, InSIDE assumption not satisfied									
10,000	0%	11.7	2.8%	0.001 (0.078)	-0.15 to 0.15	0.2%	0.000 (0.001)	-0.12 to 0.12	0%
10,000	10%	13.7	3.3%	0.077 (0.087)	-0.09 to 0.25	8%	0.044 (0.001)	-0.12 to 0.21	0.1%
10,000	20%	14.9	3.6%	0.144 (0.099)	-0.05 to 0.34	24.7%	0.107 (0.002)	-0.1 to 0.35	1.3%
10,000	30%	12.8	3.1%	0.103 (0.097)	-0.09 to 0.29	11.9%	0.102 (0.002)	-0.11 to 0.36	0.6%

CI: Confidence Interval, IV: Instumental Variable, SE: Standard Error. Null Causal Effect (β = 0)

Table 2: Summary of 1,000 simulated Mendelian randomisation studies per combination of scenario and parameters, all with positive causal effect

				Weighted Median			MR Hevo		
N	Invalid IVs	*F*	R^2	Mean Estimate (Mean SE)	Mean 95% CI	Causal Report Rate	Mean Estimate (Mean SE)	Mean 95% CI	Causal Report Rate
Scenario 1: Balanced pleiotropy, InSIDE assumption satisfied									
10,000	0%	11.7	2.8%	0.070 (0.079)	-0.08 to 0.22	4.9%	0.085 (0.001)	-0.04 to 0.21	6.2%
10,000	10%	11.7	2.8%	0.094 (0.087)	-0.08 to 0.26	11%	0.118 (0.001)	-0.05 to 0.29	12.6%
10,000	20%	11.7	2.8%	0.089 (0.093)	-0.09 to 0.27	10.3%	0.124 (0.002)	-0.08 to 0.34	5.6%
10,000	30%	11.7	2.8%	0.081 (0.094)	-0.1 to 0.27	8.7%	0.108 (0.002)	-0.12 to 0.34	1.6%
20,000	0%	24.5	3%	0.080 (0.056)	-0.03 to 0.19	21.3%	0.089 (0.001)	0 to 0.18	62.2%
20,000	10%	24.5	3%	0.098 (0.063)	-0.03 to 0.22	27.8%	0.108 (0.001)	-0.01 to 0.23	29.9%
20,000	20%	24.5	3%	0.095 (0.067)	-0.04 to 0.23	22.6%	0.113 (0.001)	-0.04 to 0.27	15%
20,000	30%	24.5	3%	0.088 (0.068)	-0.05 to 0.22	17.7%	0.104 (0.001)	-0.06 to 0.27	5.4%
Scenario 2: Directional pleiotropy, InSIDE assumption satisfied									
10,000	0%	11.7	2.8%	0.070 (0.079)	-0.08 to 0.22	5.3%	0.085 (0.001)	-0.04 to 0.21	5.9%
10,000	10%	11.7	2.8%	0.089 (0.088)	-0.08 to 0.26	9%	0.124 (0.001)	-0.05 to 0.31	11.9%
10,000	20%	11.7	2.8%	0.119 (0.094)	-0.07 to 0.3	17.7%	0.187 (0.002)	-0.02 to 0.43	32.3%
10,000	30%	11.7	2.8%	0.133 (0.095)	-0.05 to 0.32	23.3%	0.216 (0.002)	0 to 0.47	46.1%
20,000	0%	24.5	3%	0.080 (0.057)	-0.03 to 0.19	21.1%	0.089 (0.001)	0 to 0.18	62.7%
20,000	10%	24.5	3%	0.093 (0.063)	-0.03 to 0.22	24%	0.109 (0.001)	-0.01 to 0.24	29.1%
20,000	20%	24.5	3%	0.116 (0.068)	-0.02 to 0.25	35.3%	0.146 (0.001)	-0.01 to 0.31	41.2%
20,000	30%	24.5	3%	0.127 (0.069)	-0.01 to 0.26	40.7%	0.168 (0.002)	0.01 to 0.35	56.2%
Scenario 3: Directional pleiotropy, InSIDE assumption not satisfied									
10,000	0%	11.7	2.8%	0.070 (0.079)	-0.08 to 0.22	5.2%	0.085 (0.001)	-0.04 to 0.21	5.7%
10,000	10%	13.7	3.3%	0.150 (0.089)	-0.02 to 0.32	35%	0.137 (0.001)	-0.03 to 0.31	25.1%
10,000	20%	14.9	3.6%	0.213 (0.1)	0.02 to 0.41	55.8%	0.202 (0.002)	-0.01 to 0.46	45.2%
10,000	30%	12.8	3.1%	0.169 (0.099)	-0.03 to 0.36	37.1%	0.191 (0.002)	-0.03 to 0.46	29.1%
20,000	0%	24.5	3%	0.080 (0.057)	-0.03 to 0.19	21.5%	0.089 (0.001)	0 to 0.18	62.8%
20,000	10%	30.4	3.7%	0.144 (0.064)	0.02 to 0.27	66%	0.125 (0.001)	0.01 to 0.25	63%
20,000	20%	32.4	3.9%	0.189 (0.073)	0.05 to 0.33	81.5%	0.154 (0.001)	0.01 to 0.32	58.6%
20,000	30%	31.1	3.8%	0.153 (0.071)	0.01 to 0.29	60.3%	0.146 (0.001)	-0.01 to 0.32	41%

CI: Confidence Interval, IV: Instumental Variable, SE: Standard Error. Null Causal Effect (β = 0)

Citations Search

Search Results

A total of 110 abstracts and 54 full texts were screened to identify the 10 studies[1–10] listed in Table 3; the screening flow diagram is presented in Figure 2.

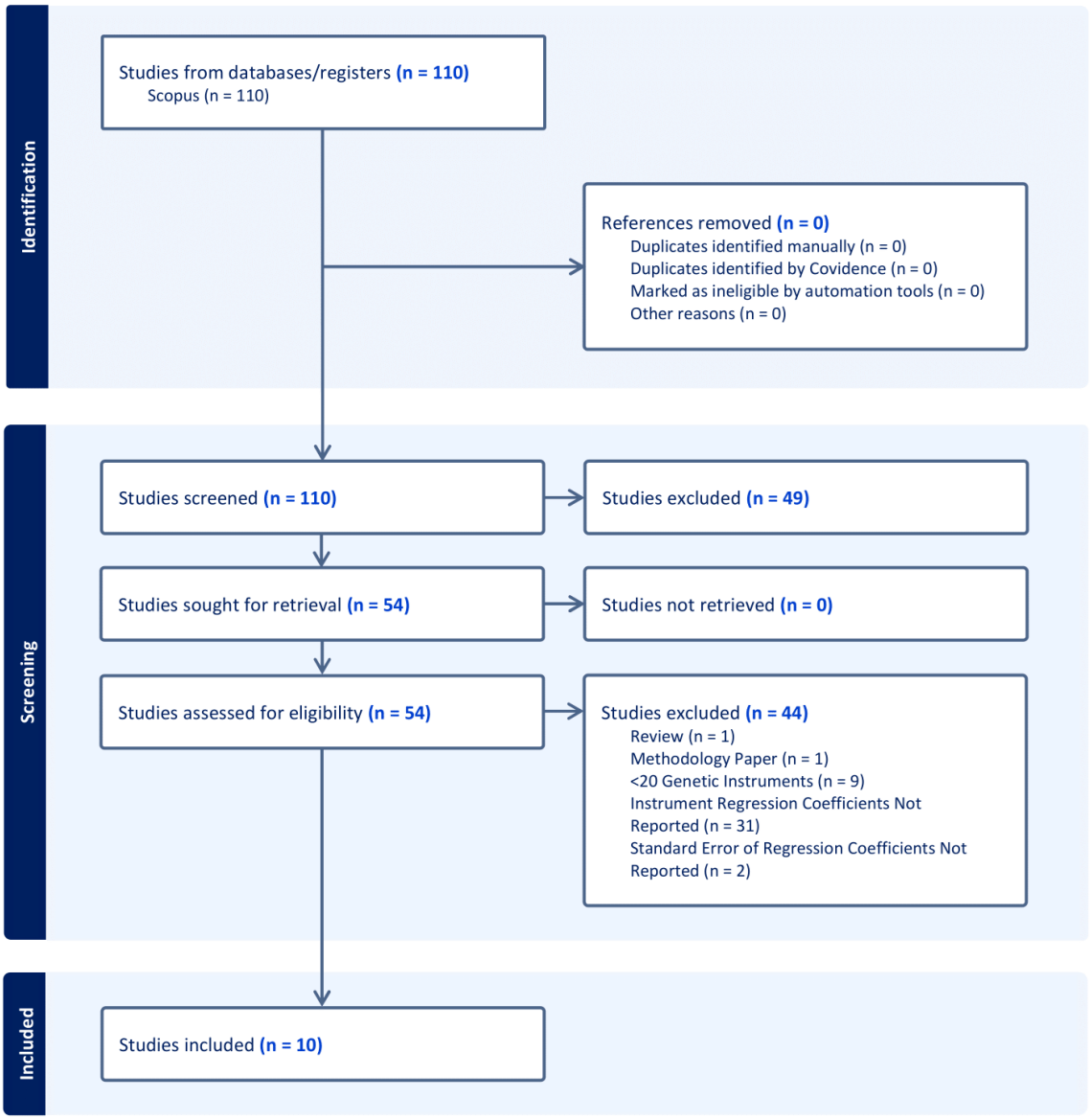

Figure 2: Flow diagram of selection for sample of ten highly-cited two-sample Mendelian randomisation articles reporting a weighted median estimate of casual effect

Table 3: Summary of ten highly-cited two-sample Mendelian randomisation articles reporting a weighted median estimate of casual effect

Author	Citations	Association	*N* Instruments	Participants			Causal Effect		Causality Reported	*p*-value
				N		Maximum Estimated Overlap				
				Exposure	Outcome		Measure	Estimate		
Budu-Aggrey et al, 2019	182	BMI vs Psoriasis	97	339,224	12,559	0%	OR	1.06 (1 to 1.12)	No	-
Carreras-Torres et al, 2017	200	Height vs Pancreatic Cancer	558	253,288	15,002	19%	OR	1.14 (1 to 1.29)	No	0.05
Carter et al, 2019	199	Education vs Coronary Disease	1,267	766,345	184,305	0%	OR	0.62 (0.57 to 0.67)	Yes	<0.001
Choi et al, 2019	492	Activity vs Depression	24	377,234	143,265	0%	OR	1.49 (0.94 to 2.36)	No	0.08
Clift et al, 2022	129	Smoking Initiation vs COVID-19 Infection	378	1,232,091	281,105	36%	OR	1.53 (1.02 to 2.28)	Yes	0.04
Ligthart et al, 2018	298	CRP vs Schizophrenia	52	204,402	82,315	0%	OR	0.89 (0.81 to 0.96)	Yes	0.004
Mokry et al, 2016	199	BMI vs Multiple Sclerosis	70	322,105	38,589	2.5%	OR	1.26 (0.98 to 1.62)	No	0.08
Pasman et al, 2018	328	Schizophrenia vs Cannabis Use	102	150,064	184,765	0%	Beta	0.163 (0.067 to 0.259)	Yes	0.001
Xie et al, 2023	138	T2DM vs NAFLD	449	441,016	218,792	0%	OR	1.61 (1.09 to 2.38)	Yes	<0.001
Xu et al, 2022	183	Coeliac vs Gut Bifidobacterium	105	15,283	24,269	63%	OR	0.998 (0.99 to 1.005)	No	0.56

Re-analysis Results

There were missing gene-outcome coefficients for three instruments from Xie et al[9], and one instrument in Clift et al[10] was reported as having an implausibly large gene-outcome coefficient/standard error (-1243.03 and 19161.64, respectively); these were imputed as the respective mean value per study. Data were otherwise complete as expected per the description in each study manuscript. A summary of the re-analysis results is presented in Table 4; estimates are presented both as β regression coefficients and odds ratio (OR)s to aid comparison across studies.

Re-analysis vs Reported WME Causal Estimates

Four WME estimates generated through re-analysis matched the originally reported estimates poorly. There was a >0.1 difference in re-analysis estimates of OR or β versus the values originally reported by Carreras-Torres et al[6], Ligthart et al[3], Mokry et al[5] and Xie et al[9]. Details of instruments used in re-analysis were re-checked against the relevant manuscripts to ensure no transcription errors had occurred.

There was a general trend of re-analysed WME estimates having far narrower CIs than reported values; exceptions were Carreras-Torres et al[6], Choi et al[1], and Ligthart et al[3].

Conclusions regarding presence of a causal effect were also inconsistent: three re-analysed studies detected a causal effect by WME not reported previously (Choi et al[1], Mokry et al[5] and Xu et al[7]); one study detected no causal effect in re-analysis despite one being reported in the original publication (Ligthart et al[3]).

Compared with reported values, the mean difference for effect estimates (SE, 95% CI) from the was 0(0)

Re-analysis WME vs MR-Hevo Causal Estimates

Across all ten studies re-analysed (Table 4), a causal effect was reported by MR-Hevo in 5 and by WME in 5; causal reporting was discordant in 0

Of the ten studies, MR-Hevo exhibited a favourable sensitivity versus WME in 5 (42%). For both MR-Hevo and WME methods, causal report rate generally increased with an increasing proportion of invalid instruments, and also as assumption violations progressively presented greater data variability and bias across scenarios 1 to 3.

Table 4: Re-analysis of ten highly-cited two-sample Mendelian randomisation articles reporting a weighted median estimate of casual effect, comparing results of both WME and MR-Hevo causal effect estimation methods

Author	Exposure	Outcome	SNPs	Weighted Median				MR-Hevo			
				β	SE	OR	Causality Reported	β	SE	OR	Causality Reported
Budu-Aggrey et al	BMI	Psoriasis	97	0 (-0.28-0.28)	0.144	1 (0.75-1.33)	No	0.08 (-0.16-0.34)	0.002	1.08 (0.86-1.4)	No
Carreras-Torres et al	Height	Pancreatic Cancer	558	0 (-0.13-0.13)	0.064	1 (0.88-1.14)	No	0.29 (-1.18-1.55)	0.688	1.34 (0.31-4.72)	No
Carter et al	Years of Education	Coronary Artery Disease	1,266	-0.46 (-0.55–0.38)	0.043	0.63 (0.58-0.68)	Yes	-0.48 (-0.54–0.42)	0.000	0.62 (0.58-0.66)	Yes
Choi et al	Self-Reported Physical Activity	Major Depressive Disorder	25	0.39 (-0.05-0.83)	0.224	1.47 (0.95-2.28)	No	0.22 (-0.24-0.64)	0.004	1.25 (0.78-1.89)	No
Clift et al	Genetically Determined Smoking Initiation	COVID-19 Infection	378	0.43 (0.03-0.82)	0.201	1.53 (1.03-2.27)	Yes	0.37 (0.1-0.64)	0.001	1.45 (1.11-1.89)	Yes
Ligthart et al	Genetically Determined CRP	Schizophrenia	29	-0.41 (-0.87-0.06)	0.235	0.67 (0.42-1.06)	No	-0.36 (-1.24-0.59)	0.009	0.69 (0.29-1.8)	No
Mokry et al	BMI	Multiple Sclerosis	70	0.34 (0.08-0.6)	0.131	1.41 (1.09-1.82)	Yes	0.34 (0.17-0.51)	0.001	1.4 (1.18-1.66)	Yes
Pasman et al	Liability to Schizophrenia	Cannabis Use	109	0.16 (0.07-0.26)	0.048	1.18 (1.07-1.29)	Yes	0.17 (0.07-0.26)	0.001	1.18 (1.08-1.29)	Yes
Xie et al	T2DM	NAFLD	526	0.48 (0.1-0.86)	0.192	1.61 (1.11-2.35)	Yes	0.51 (0.27-0.75)	0.002	1.66 (1.31-2.12)	Yes
Xu et al	Coeliac Disease	Gut Bifidobacterium	105	0 (-0.01-0.01)	0.004	1 (0.99-1.01)	No	0 (-0.01-0)	0.000	1 (0.99-1)	No

β and OR presented as: estimate (95% CI).

β: causal effect estimate, CI: Confidence Interval, OR: Odds Ratio, SE: Standard Error.

BMI: body mass index, CRP: C-reactive protein, NAFLD: non-alcoholic fatty liver disease, T2DM: type 2 diabetes mellitus

Table 4 reference[2,11]

BMI body mass index
CI confidence interval
CLT central limit theorem
CRP C-reactive protein
IV instrumental variable
InSIDE Instrument Strength Independent of Direct Effect
MR Mendelian randomisation
NAFLD non-alcoholic fatty liver disease
OR odds ratio
RCT randomised-controlled trial
SD standard deviation
SE standard error
SNP single nucleotide polymorphism
T2DM type 2 diabetes mellitus
UMREG Usher Masters Research Ethics Group
WME weighted median estimator

1. Choi KW, Chen CY, Stein MB, Klimentidis YC, Wang MJ, Koenen KC, et al. Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults: A 2-Sample Mendelian Randomization Study. JAMA Psychiatry [Internet]. 2019 Apr [cited 2025 Apr 27];76(4):399–408. Available from: https://doi.org/10.1001/jamapsychiatry.2018.4175

2. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal effect of schizophrenia liability. Nature Neuroscience [Internet]. 2018 Sep [cited 2025 May 11];21(9):1161–70. Available from: https://www.nature.com/articles/s41593-018-0206-1

3. Ligthart S, Vaez A, Võsa U, Stathopoulou MG, Vries PS de, Prins BP, et al. Genome Analyses of >200,000 Individuals Identify 58 Loci for Chronic Inflammation and Highlight Pathways that Link Inflammation and Complex Disorders. The American Journal of Human Genetics [Internet]. 2018 Nov [cited 2025 May 27];103(5):691–706. Available from: https://www.cell.com/ajhg/abstract/S0002-9297(18)30320-3

4. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. BMJ [Internet]. 2019 May [cited 2025 May 27];365:l1855. Available from: https://www.bmj.com/content/365/bmj.l1855

5. Mokry LE, Ross S, Timpson NJ, Sawcer S, Davey Smith G, Richards JB. Obesity and Multiple Sclerosis: A Mendelian Randomization Study. Muraro PA, editor. PLOS Medicine [Internet]. 2016 Jun [cited 2025 May 27];13(6):e1002053. Available from: https://dx.plos.org/10.1371/journal.pmed.1002053

6. Carreras-Torres R, Johansson M, Gaborieau V, Haycock PC, Wade KH, Relton CL, et al. The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study. JNCI: Journal of the National Cancer Institute [Internet]. 2017 Sep [cited 2025 May 27];109(9):djx012. Available from: https://doi.org/10.1093/jnci/djx012

7. Xu Q, Ni JJ, Han BX, Yan SS, Wei XT, Feng GJ, et al. Causal Relationship Between Gut Microbiota and Autoimmune Diseases: A Two-Sample Mendelian Randomization Study. Frontiers in Immunology [Internet]. 2022 Jan [cited 2025 May 28];12. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2021.746998/full

8. Budu-Aggrey A, Brumpton B, Tyrrell J, Watkins S, Modalsli EH, Celis-Morales C, et al. Evidence of a causal relationship between body mass index and psoriasis: A mendelian randomization study. PLOS Medicine [Internet]. 2019 Jan [cited 2025 Jun 2];16(1):e1002739. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002739

9. Xie J, Huang H, Liu Z, Li Y, Yu C, Xu L, et al. The associations between modifiable risk factors and nonalcoholic fatty liver disease: A comprehensive Mendelian randomization study. Hepatology [Internet]. 2023 Mar [cited 2025 Jun 3];77(3):949. Available from: https://journals.lww.com/hep/pages/articleviewer.aspx?year=2023&issue=03000&article=00022&type=Fulltext#T1

10. Clift AK, Ende A von, Tan PS, Sallis HM, Lindson N, Coupland CAC, et al. Smoking and COVID-19 outcomes: An observational and Mendelian randomisation study using the UK Biobank cohort. Thorax [Internet]. 2022 Jan [cited 2025 Jun 3];77(1):65–73. Available from: https://thorax.bmj.com/content/77/1/65

11. Bowden J, Smith GD, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology [Internet]. 2016 Apr [cited 2024 Oct 22];40(4):304. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC4849733/