Appendix
A1. Supplementary Figures

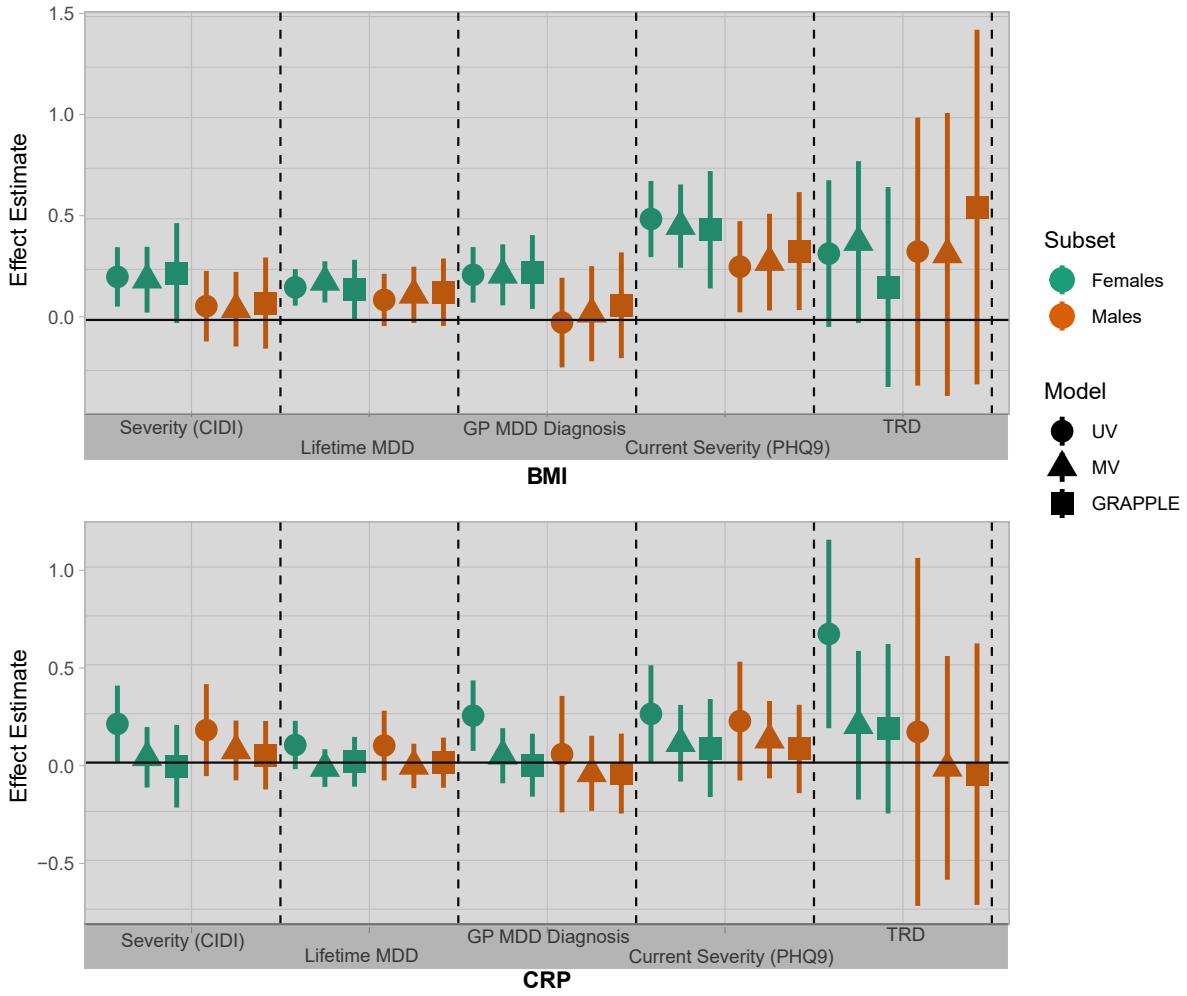

Figure 4: Sex-Stratified Analysis.

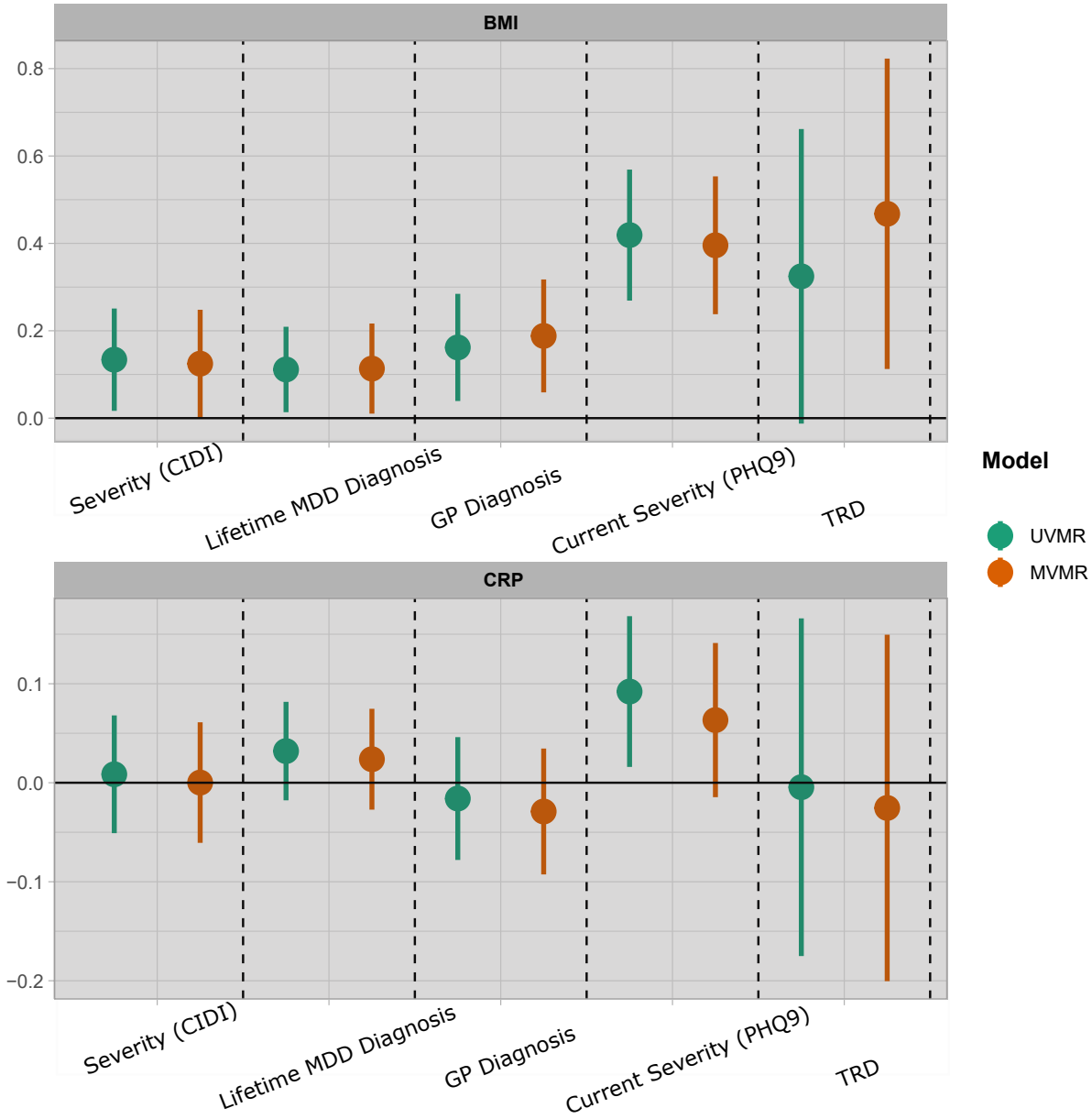

Figure 5: MR Analysis with external weights for BMI [39] and CRP[37]

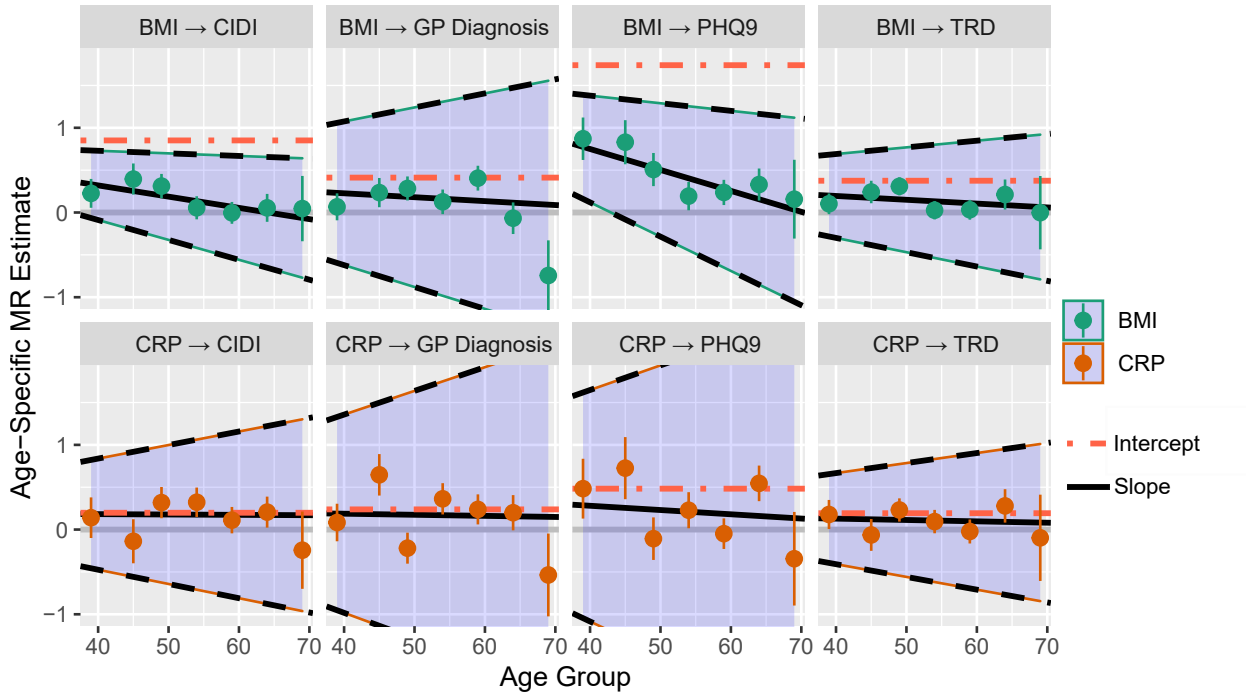

Figure 6: Age as a Moderator of the causal associations of CRP and BMI with mood outcomes. In the visualisation of the meta-regression slope, if the intercept falls within the confidence region of the age slope, then the result is not statistically significant.

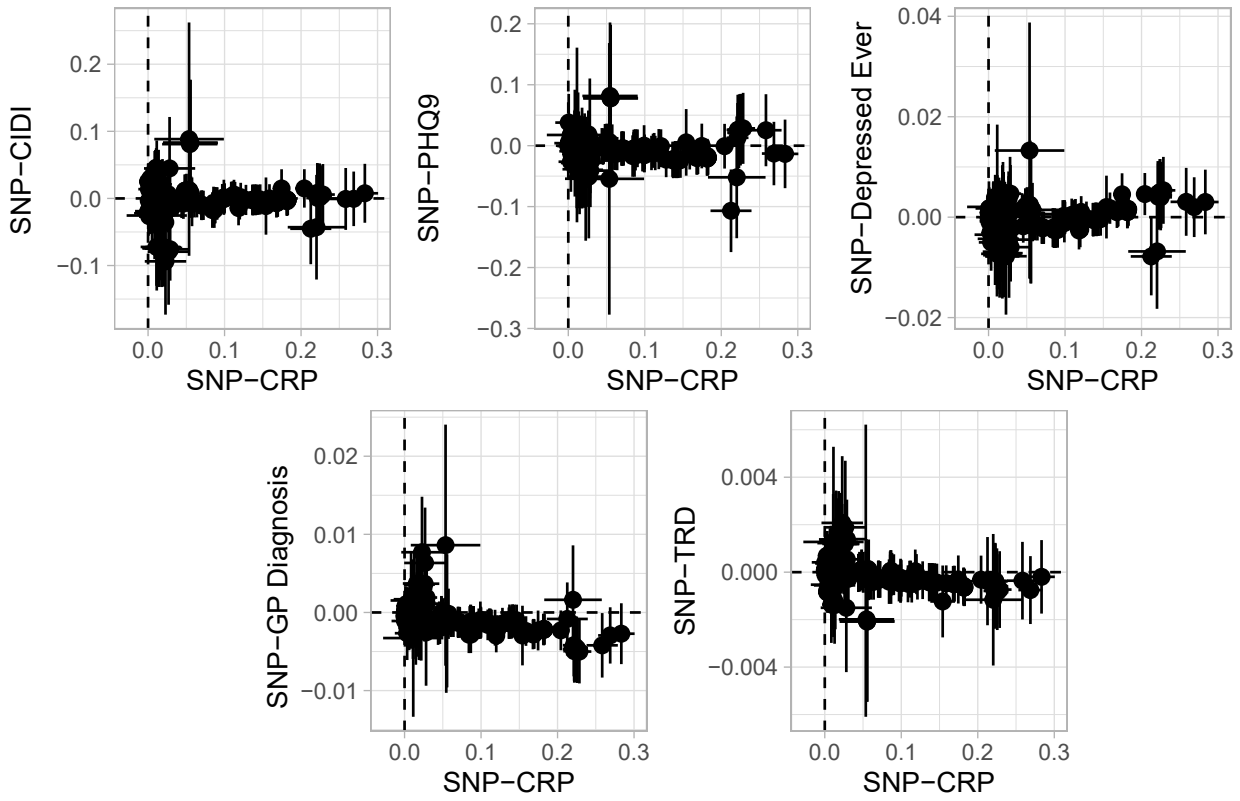

Figure 7: SNP-CRP associations and SNP-depression associations for $n = 194$ SNPs in LD. These genetic associations are then projected to independent genetic components (cis-MR, [42]).

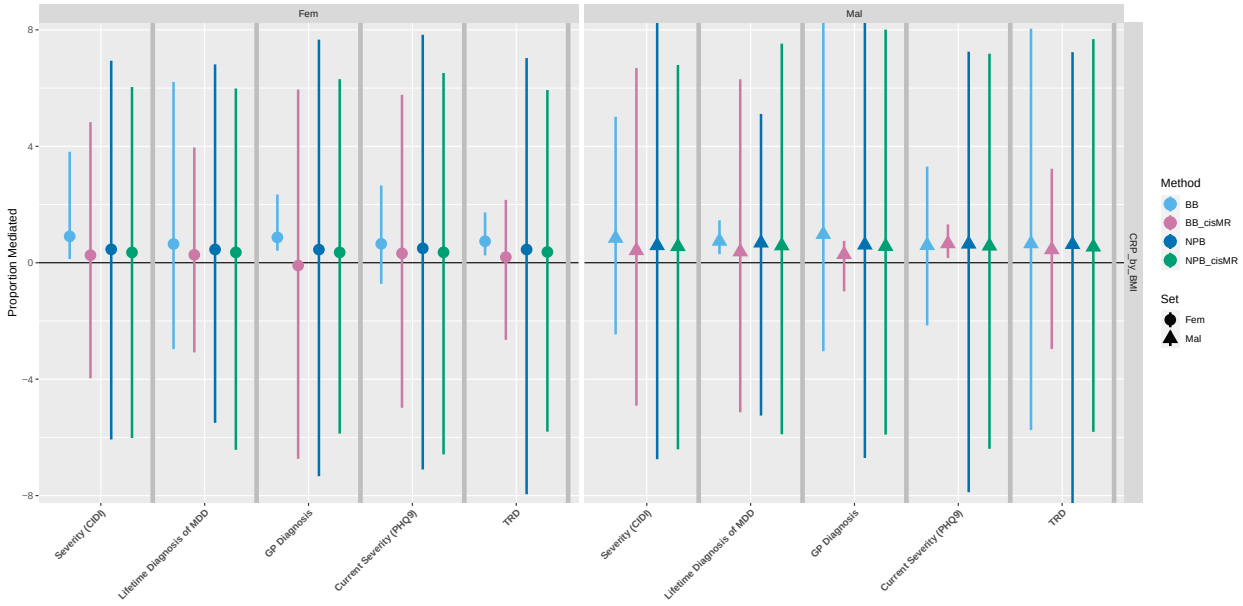

Figure 8: Proportion of CRP effect mediated by BMI in males and females. The CRP effect is defined by two different instruments. cisMR: CRP effect is estimated with one SNP as instrument (rs2794520).

A.2 Supplementary Tables

	Controls	GP-based depression cases	TRD cases
N	173,786	18,330	2,199
Age	57.42 (±8.11)	56.11 (±7.95)	56.43 (±7.82)
% Female	0.485	0.636	0.724
BMI	27.38 (±4.67)	28.26 (±5.37)	29.41 (±5.96)
CRP (mg/L)	2.54 (±4.35)	2.99 (±4.57)	3.64 (±5.42)
CIDI_MDD*	2 (±10.43)	5.45 (±11.55)	6.63 (±1.68)
PHQ9*	2.11 (±2.89)	4.52 (±5.09)	9.06 (±6.83)

Table 1: Individual characteristics. Mean (±SD). *: CIDI and PHQ9 were measured in a different, partly overlapping subset of UKB participants ($n = 146,067$).

		CRP GWAS			**CRP rs2794520**			**BMI**			
Age Group	**N**	***F***	**CFS**	R^2	***F***	**CFS**	R^2	***F***	R^2	**CFS, rs2794520**	**CFS, CRP GWAS**
38-45	43,191	59.51	24.06	0.065	572.17	14.69	0.014	13.05	0.021	10.76	7.609
45-49	43,738	60.77	25.13	0.065	505.94	13.862	0.012	10.75	0.017	9.197	6.48
49-54	64,289	88.31	36.11	0.065	715.76	20.173	0.012	16.38	0.018	13.782	9.696
54-59	76,918	104.54	43.19	0.064	1014.21	26.006	0.014	16.94	0.015	14.355	10.171
59-64	108,124	154.10	61.55	0.067	1453.36	37.431	0.014	20.71	0.013	17.864	12.375
64-69	90,975	143.60	57.14	0.074	1228.65	33.846	0.014	15.63	0.012	13.776	9.533
69-75	16,357	26.58	11.18	0.076	263.45	7.266	0.017	3.72	0.016	3.324	2.592

Table 2: Instrument Strength (Mean $F-$statistic (*F*), conditional *F*-statistic (CFS), R^2) for each age group. N: sample size for each group.

Instrument	**Strength Measure**	**Exposure**	**CRP Instrument Strength**
45 CRP GWAS hits	*F-statistic*	CRP	628.95
45 CRP GWAS hits + 73 BMI SNPs	Conditional F-statistic	CRP	252.71
rs2794520 (C → T)	*F-statistic*	CRP	5732.20
rs2794520 (C → T)	Conditional F-statistic	CRP	147.92
73 BMI SNPs	*F-statistic*	BMI	90.78
45 CRP GWAS hits + 73 BMI SNPs	Conditional F-statistic	BMI	52.75
rs2794520 (C → T) + 73 BMI SNPs	Conditional F-statistic	BMI	78.30

Table 3: Instrument Strength for each MR model presented.

	Estimate	**LCI**	**UCI**
CIDI	0.003	-0.097	0.103
DEP_EVER	0.011	-0.015	0.014
GP Diagnosis	-0.014	-0.01	0.01
PHQ9	-0.046	-0.129	0.13
TRD	-0.003	-0.004	0.004

Table 4: Effect of CRP on depression outcomes, results for the cis-MR approach [42] that includes $n = 194$ SNPs in LD near the *CRP* gene.

	Median	LCI	UCI
GP Diagnosis	1.47%	-253.39%	48.26%
TRD	22.19%	-182.18%	88.19%
DEP	-5.39	-1706.96%	982.97%
CIDI_MDD	33.81%	-374.37%	90.1%
PHQ9	74.87%	-241.37%	1083.2%

Table 5: Proportion of CRP effect mediated by BMI when using one SNP (*rs2794520*) to genetically proxy serum CRP levels.

	CRP	BMI	CRP & BMI MVMR (45 GWAS hits)	CRP & BMI MVMR (rs2794520)
Severity (CIDI)	476.07, 0	1097.50, 0	1155.62, 0	1102.74, 0
Current Severity (PHQ9)	1356.22, 0	2891.75, 0	3007.66, 0	2962.58, 0
GP Diagnosis of Depression	350.86, 0	92.48, 0.0372	703.59, 0	675.73, 0
TRD	162.73, 0	368.44, 0	413.20, 0	386.74, 0
Lifetime MDD Diagnosis	332.43, 0	700.57, 0	765.54, 0	715.02, 0

Table 6: Sargan Test for heterogeneity.

Outcome	Model	BMI Fisher's z	BMI p value	CRP Fisher's z	CRP p value
TRD	UVMR	0.0038	0.979	1.0051	0.332
TRD	MVMR	0.1481	0.882	0.5511	0.534
TRD	GRAPPLE	0.782	0.441	0.5237	0.567
PHQ9	UVMR	1.3176	0.115	0.1597	0.854
PHQ9	MVMR	1.2152	0.269	0.1541	0.899
PHQ9	GRAPPLE	0.5038	0.615	0.0273	0.978
Ever Depressed	UVMR	0.7316	0.441	0.0274	0.978
Ever Depressed	MVMR	0.7348	0.467	0.2168	0.885
Ever Depressed	GRAPPLE	0.0835	0.917	0.0054	0.96
GP Diagnosis	UVMR	1.6839	0.076	1.2134	0.268
GP Diagnosis	MVMR	1.3196	0.178	0.7387	0.459
GP Diagnosis	GRAPPLE	1.0069	0.313	0.3348	0.747
CIDI	UVMR	1.1051	0.213	0.2325	0.838
CIDI	MVMR	1.2101	0.244	0.3163	0.753
CIDI	GRAPPLE	0.7752	0.385	0.4188	0.687

Table 7: Results for Fisher's z statistic to assess sex specificity of CRP & BMI estimates on mood outcomes.

Outcome	Model	BMI Fisher's z	BMI pvalue	CRP Fisher's z	CRP pvalue
Ever Depressed	Observational	0.3735	0.7088	0.2673	0.7892
GP Diagnosis	Observational	0.1584	0.8742	0.2251	0.8219
TRD	Observational	0.0279	0.9777	0.3126	0.7546
PHQ9	Observational	0.0742	0.9409	0.2387	0.8113
Ever Depressed	UVMR	0.5049	0.6136	1.2798	0.2006
GP Diagnosis	UVMR	1.0078	0.3136	0.3615	0.7177
TRD	UVMR	0.7445	0.4566	-0.2475	0.8045
PHQ9	UVMR	0.4049	0.6856	1.2942	0.1956
Ever Depressed	MVMR	0.3463	0.7291	0.7946	0.4269
GP Diagnosis	MVMR	0.7355	0.462	-0.0013	0.9989
TRD	MVMR	0.7037	0.4816	-0.5442	0.5863
PHQ9	MVMR	0.304	0.7612	1.4134	0.1575

Table 8: Fisher's z statistic for sex heterogeneity, smaller CRP instrument

	β_{age^2}	SE	p-value
CRP-PHQ9	-0.0979	0.1825	0.5915
BMI PHQ9	**-0.2240**	**0.1089**	**0.0397**
BMI TRD	-0.0602	0.0920	0.5127
CRP GP_Diag	-0.1147	0.1713	0.5030
BMI GP_Diag	-0.1574	0.0999	0.1152
CRP CIDI	-0.0325	0.1060	0.7592
BMI CIDI	-0.1444	0.0867	0.0953
CRP TRD	-0.0047	0.0992	0.9629

Table 9: Age as a Moderator of the causal associations of CRP and BMI with mood outcomes, results for a quadratic effect (β_{age^2}).

	Q	Q_p	Q'	Q'_p	Q_{Diff}	$Q_{Diff,p}$
BMI → CIDI	5.56	0.474	2.42	0.788	3.139	0.076
BMI → GP Diagnosis	10.365	0.111	9.982	0.076	0.383	0.536
BMI → PHQ9	9.839	0.139	3.345	0.647	**6.494**	**0.011**
BMI → TRD	5.403	0.493	4.905	0.428	0.498	0.48
CRP → CIDI	3.832	0.699	3.83	0.574	0.002	0.965
CRP → GP Diagnosis	11.817	0.066	11.799	0.038	0.018	0.893
CRP → PHQ9	9.815	0.134	9.624	0.087	0.19	0.663
CRP → TRD	3.58	0.733	3.542	0.617	0.038	0.844

Table 10: Heterogeneity Statistics for the age-specific causal effects in the effects of CRP and BMI on mood outcomes.

Sample	**Exposure & Mediator**	**Median**	**Q0.025**	**Q0.05**	**Q0.1**	**Q0.9**	**Q0.95**	**Q0.975**
All	BMI by CIDI	0.287	-2.375	-0.838	0.119	0.918	1.563	3.221
Female	BMI by CIDI	0.182	-1.306	-0.526	0.021	0.683	1.213	2.132
Male	BMI by CIDI	0.105	-1.384	-0.59	-0.263	0.437	0.790	1.464
All	CRP by CIDI	0.230	-3.372	-1.541	-0.803	1.037	2.1492	4.0404
Female	CRP by CIDI	0.096	-2.607	-1.044	-0.437	0.506	0.914	1.903
Male	CRP by CIDI	0.051	-2.050	-0.982	-0.504	0.449	0.936	1.692

Table 11: Proportion of CRP and BMI effects on TRD Mediated by genetically predicted CIDI levels (MDD GRS Instrument with $n = 178$ SNPs [36])

B Assessment of Heterogeneity in Age-Specific Causal Estimates

We implemented the following regression model to see if there was a trend in causal estimates due to age:

$$\hat{\beta}_{XY_i} = \beta_0 + \beta_{age} a\bar{g}e_i + \epsilon_{XYi}.$$

Here $\hat{\beta}_{XY_i}, i = 1, ..., 7$ expresses the causal effect estimate in the ith five-year age stratum, with β_0 the intercept and β_{age} the average age-moderation parameter of the of the mean stratum age $a\bar{g}e_i$). To assess how this model compares against a simple pooled estimate that does not take age into account, we estimated two heterogeneity statistics. Firstly, Cochran's Q statistic:

$$Q = \sum_{i=1}^{7} w_i(\hat{\beta}_{XYi} - \hat{\beta}_{All})^2, \qquad \hat{\beta}_{All} = \sum_{i=1}^{7} w_i \hat{\beta}_{XYi} / \sum_{i=1}^{7} w_i,$$

on 6 degrees of freedom ,and Rucker's Q statistic: [47, 9])

$$Q' = \sum_{i=1}^{7} w_i(\hat{\beta}_{XY_i} - (\hat{\beta}_0 + \hat{\beta}_{age} a\bar{g}e_i))^2,$$

on 5 degrees of freedom. We then calculated their difference Q_{diff}. Under the null hypothesis that age was not a predictor of the causal effect, $Q_{diff} \sim \chi^2_{df=1}$. Results of this analysis are are contained in table 10.

C Applying Collider-Correction to MVMR

In this paper we use the GRAPPLE method [59] to jointly estimate the direct effects of CRP and BMI on the depression-related phenotype of interest in an MVMR analysis,whilst guarding against weak instrument bias and pleiotropy. The method has generally been applied in two-sample MR where the exposures and the outcome are measured in independent sets. This means that the SNP-exposure and SNP-outcome association estimates are independent. When this is not the case, an additional variance component must be added into the software that can be estimated from the sample correlation between the exposure and outcome data sets. Here we describe how we implement the approach in one-sample data (UKB) by combining GRAPPLE with the technique of Collider Correction

[8] which avoids this step. In essence, Collider-Correction is a tool for applying any two-sample method to one sample data by basing it on summary statistics gleaned from the one sample data that are statistically independent. R code for the approach is available at `https://github.com/vaskarageorg/CRP_BMI_Depr_MVMR`.

Consider the following models linking L continuous exposure measurements $\boldsymbol{X_i} = (X_{1i}, \ldots, X_{Li})$ to the outcome Y_i of participant i:

$$\begin{aligned} X_{ai}|G_{i.}, U_i &= \sum_{j=1}^{k} \gamma_{aj} G_{ij} + \beta_{UX,a} U_i + \varepsilon_{X_{ai}}, \quad a = 1, ..., L \\ Y_i|X_{1i} \ldots X_{Li}, G_i, U_i &= \boldsymbol{\beta^T X_i} + \sum_{j=1}^{k} \alpha_j G_{ij} + \beta_{UY} U_i + \varepsilon_{Y_i} \end{aligned}$$

Here, our model for subject i's exposure depends on a set of k genetic variants $G_{i.} = G_{i1}, \ldots, G_{ik}$ and an unmeasured confounder variable U_i. The parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)$ represents the direct causal effect of each exposure on the outcome that we wish to estimate. Consider a regression of the outcome Y on G and $\boldsymbol{X}$ together (but not U). Under our assumed data generating model:

$$E[Y_i|\boldsymbol{X_i}, G_i] = \boldsymbol{\beta^{T*} X_i} + \sum_{j=1}^{k} \alpha_j^* G_{ij} \tag{1}$$

yielding collider biased coefficients $\boldsymbol{\beta^*} = (\beta_1^*, \ldots, \beta_L^*)$ and $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_k^*$. Using similar arguments to those presented in [8], α_j^*, α_j, β^* and β are linked through the following linear relation:

$$\hat{\alpha}_j^* = \alpha_j + (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \boldsymbol{\gamma_j} + \theta_j, \tag{2}$$

where $\boldsymbol{\gamma_j} = (\gamma_{1j}, \ldots, \gamma_{Lj})$ and θ_j is mean zero residual error with $Var(\theta_j) = \sigma^2_{\alpha_j^*} = Var(\hat{\alpha}_j^*)$. This suggests the following general algorithm for estimating the causal effect that is a generalisation of the original algorithm in [8]:

1. Regress Y on $\boldsymbol{X}$ and G to obtain the collider biased parameter estimates $\hat{\beta}_1^*, ..., \hat{\beta}_L^*$ and $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_k^*$.
2. Regress each X_{ai} on G to obtain estimates $\hat{\gamma}_{a1}, \ldots, \hat{\gamma}_{ak}$
3. Fit the MVMR:
$$E[\hat{\alpha}_j^*|\hat{\boldsymbol{\gamma}}_{\boldsymbol{j}}] = \alpha_0 + (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \hat{\boldsymbol{\gamma}}_{\boldsymbol{j}} \tag{3}$$
Under a user specified loss function in order to estimate for the Collider-Correction term $(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T$.
4. Adjust the Collider-biased observational estimate to obtain an estimate for the causal effect β via:
$$\hat{\boldsymbol{\beta}}^{\boldsymbol{T}} = \hat{\boldsymbol{\beta}}^{\boldsymbol{*T}} + \widehat{(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T} \tag{4}$$

We use MR-GRAPPLE to fit step 3 of the Collider correction algorithm for the case of L=2 (BMI and CRP). Specifically this sets the parameter α_0 to 0 under the assumption of balanced pleiotropy (although outlyling SNPs are still penalized), and incorporates an adjustment for weak instrument bias due to uncertainty in $\hat{\boldsymbol{\gamma}}_{\boldsymbol{j}}$.

We performed a simulation study to assess the performance of GRAPPLE with collider correction. The data generating model is represented in Figure 9. The variant-exposure associations are generated in a way that conditional instrument strength was low for both X_1 and X_2 (between 2 and 20) and there is a direct effect of G_{X_1} on the outcome Y (zero mean, balanced pleiotropy).

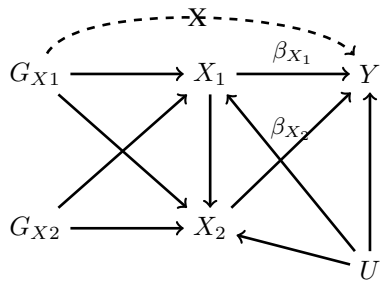

Figure 9: Assumed Data Generated Mechanism for two exposures (X_1, X_2).

Three methods were applied to estimate the causal effects of X_1 and X_2 on Y (which were $\beta_{X_1} = 1$ and $\beta_{X_2} = 0.5$ respectively).

1. MVMR using standard 2SLS, the preferred method in one-sample MR studies;

2. MVMR using GRAPPLE + Collider-Correction (GC);

3. MVMR using GRAPPLE applied directly to SNP-exposure and SNP-outcome summary statistics incorporating sample correlation estimates between X_1,X_2, and Y (GS).

Results for the three implementations are shown in Figure 10. We see that 2SLS estimates for the two causal parameters due to its failure to account for weak instruments. In contrast, both the GC and GS approaches deliver approximately unbiased estimates across the entire range of conditional F-statistic values. The results confirm the validity of our proposed GC algorithm.

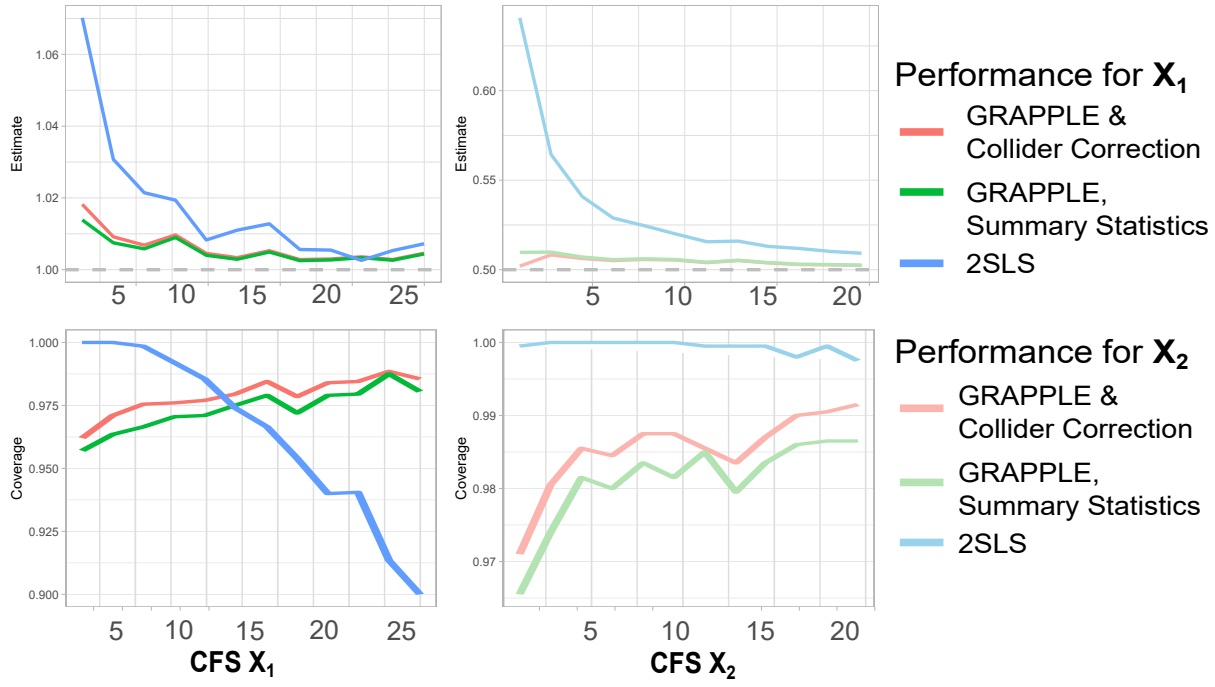

Figure 10: The two causal effects β_{X_1} and β_{X_2} are visualised with grey dashed lines. CFS: Conditional $F-$statistic; CC: Collider Correction; 2SLS: Two-Stage Least Squares; X_1, X_2: Exposures 1 and 2 in Figure 9

.

C.1 Improved performance of the Bayesian Bootstrap: simulation evidence

It was observed that the non-parametric bootstrap gave very large estimates for the uncertainty of $\hat{\pi}_m$, due to the inherent instability of $\hat{\pi}_m$ itself. To understand if (a) this large variability was a true reflection of the uncertainty in $\hat{\pi}_m$ or an over-estimate, and (b) if Rubin's Bayesian bootstrap could provide a more accurate quantification, we performed a simulation study. The data generating model is visualised in Figure 11.

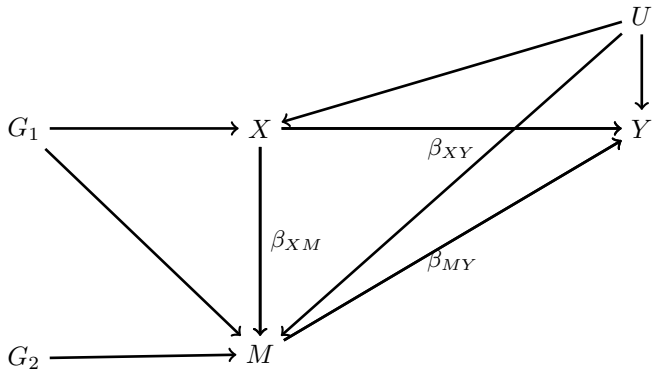

Figure 11: Directed Acyclic Graph

The following data-generating models for the exposure X, mediator M and outcome Y were used, consistent with the DAG in Figure 11:

$$\begin{aligned} X_i &= \sum_{p=1}^{P} \gamma_{Xp} G_{1ip} + U_i + \epsilon_{Xi} \\ M_i &= \beta_{XM} X_i + \sum_{p=1}^{P} \gamma_{1M} G_{1ip} + \sum_{p=1}^{P} \gamma_{2Mp} G_{2ip} + U_i + \epsilon_{Mi} \\ Y_i &= \beta_{XY} X_i + \beta_{MY} M_i + U_i + \epsilon_{Yi} \end{aligned}$$

The UVMR estimate for the causal effect of X on Y, β_{XY},is biased if we exclusively use the set G_1 as instruments for X, since the SNPs exert a pleiotropic effect on Y through M. If, however, we genetically proxy both X and M in an MVMR analysis, this bias can be removed and we can also estimate the extent to which the effect of X on Y is mediated through M [15]. In this setting, the target of the bootstrap methods is to quantify the uncertainty in the mediated proportion estimate:

$$\begin{aligned} \hat{\pi}_m &= \frac{\hat{\beta}_{XM}\hat{\beta}_{MY}}{\hat{\beta}_{XY} + \hat{\beta}_{XM}\hat{\beta}_{MY}} \\ &= 1 - \frac{\hat{\beta}_{XY}}{\hat{\beta}_{XY} + \hat{\beta}_{XM}\hat{\beta}_{MY}}, \end{aligned}$$

where in order to interpret this quantity as a proportion we require that the total effect and mediated effect are of the same sign. We apply three methods to estimate the sampling distribution of $\hat{\pi}_m$.

1. Standard Non-parametric bootstrap: in each iteration, a bootstrapped sample of equal size is constructed by sampling with replacement from the original data:
2. Quasi-Bayesian estimation of confidence intervals [30] implemented with the *mediation* R package;
3. A fully Bayesian bootstrap [46]: in each iteration a vector of Dirichlet weights ($w_i = \frac{\Gamma_i(1,1)}{\sum_{i=1}^{N}\Gamma_i(1,1)}$) is generated and used to fit the the UVMR and MVMR regressions on the original complete data, using regression weights w_i for subject i

Results for our simulation are presented in Figure 12 in which we show the mean (point estimate) and 95% confidence interval for $\hat{\pi}_m$ (top) and its coverage (bottom) based on the three bootstrap procedures. Similar to our previous simulation, data were generated to give a range of conditional F statistic values for X between 2 and 30. We see that the standard non-parametric bootstrap performed poorly in terms of in terms of bias, and large over-coverage due to over-estimating the uncertainty. The Quasi-Bayesian bootstrap performed marginally better than the standard bootstrap in terms of bias and precision, but also exhibited over-coverage. The fully Bayesian bootstrap procedures performed the best, furnishing a sampling distribution for $\hat{\pi}_m$ centred on the true value, from which could be derived confidence intervals with the correct nominal coverage. We therefore recommend the Bayesian bootstrap for use in this context.

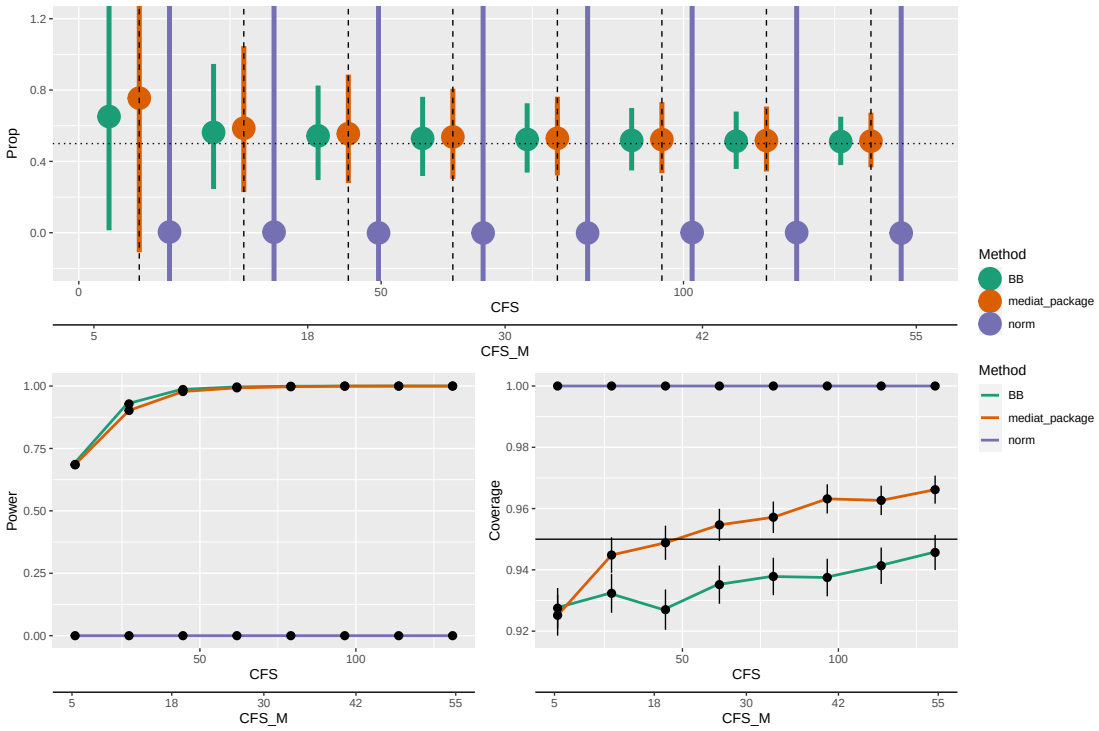

Figure 12: Simulation Results. CFS: Conditional $F-$statistic for the exposure X in Figure 11; CFS: Conditional $F-$statistic for the mediator M. Error bars in the coverage and power plots represent the Monte Carlo error for $s = 6000$ simulations.