Power of MR in within-family analysis

2018-08-24

Simulation of genotypes

Create n mothers and fathers. They are unrelated. Each mother-father pair has 2 children. All individuals have genomes of 90 SNPs. Calculate IBD at each locus for each sibling pair. The distribution of IBD across the 90 SNPs is $N(\mu = 0.5, \sigma = 0.037)$, as per theory, because there are on average 90 recombination events separating siblings. Hence, we are assuming each SNP has an effect and that all effects are independent.

Dynastic effects on phenotype

See Figure 1. Parents have confounders u, exposure x and outcome y phenotypes. The confounder influences x and y by some effect β_{ux}, β_{uy}, and x influences y with β_{xy}.

The children have the same setup, except the parent's x values influence the children's u values also, with effect β_{xu}. This represents the dynastic effect.

The genetic influence of each of the 90 SNPs on x amounts to explaining V_{gx} of the variance in x. There is no pleiotropy.

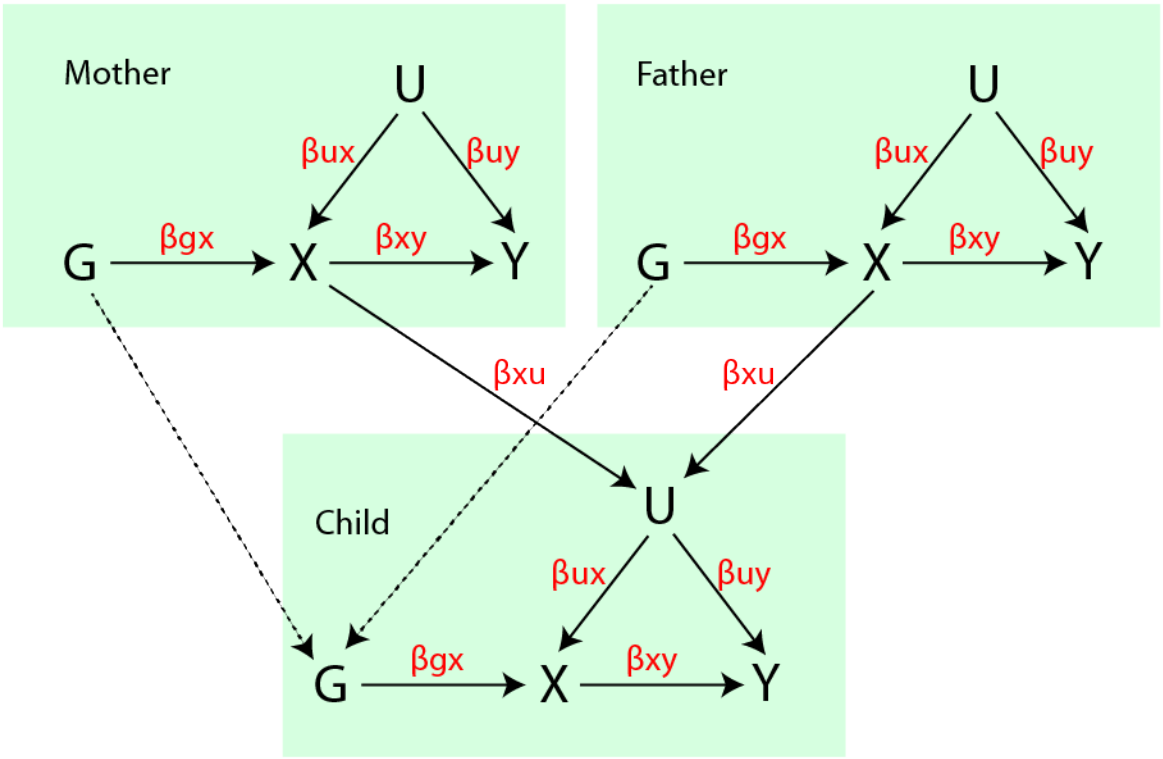

Figure 1: Model of dynastic effects used in these simulations

Statistical models

Standard MR

Use just one sibling from each family and perform MR as usual using IVW. This involves obtaining the SNP-exposure association for each of the 90 SNPs; the SNP-outcome association for each of the 90 SNPs, and estimating the slope underlying the relationship using inverse squared standard errors for weights.

Within-family D^2

The SNP-exposure estimates are performed within family. e.g. For SNP j

$$(x_{i1} - x_{i2})^2 = \beta_{xj}\hat{\pi}_{ij} + e_{ij}$$

and similarly the SNP-outcome estimates are obtained as

$$(y_{i1} - y_{i2})^2 = \beta_{yj}\hat{\pi}_{ij} + e_{ij}$$

The β_{xj} and β_{yj} estimates are then used in a standard IVW framework to obtain the causal effect.

Within-family S^2

Identical to the D^2 except the LHS of the equation is replaced with $(y_{i1} + y_{i2})^2$.

Within-family $D^2 + S^2$

Following from Visscher and Hopper 2001 this approximates the maximum likelihood estimate of the within-family linkage analysis by meta-analysing the D^2 and S^2 results of each SNP-exposure and SNP-outcome estimate. The results are subsequently used to estimate the causal effect in the standard IVW framework.

$$
\begin{aligned}
w &= \frac{\frac{1}{v_D}}{\frac{1}{v_D} + \frac{1}{v_S}} \\
b_{DS} &= \frac{1}{2}((1-w)b_S - wb_D) \\
v_{DS} &= \frac{1}{4}wv_D
\end{aligned}
$$

Hybrid

Here the effect estimates for the SNP-exposure association are obtained externally (e.g. from a GWAS). In the case of the simulations, they are obtained from estimating $x = \beta_{xj}g + e$ using the fathers' simulated data. A weighted genetic difference between siblings is then obtained:

$$\delta_{x_g,j} = \hat{\beta}_x j g_{i_1,j} - \hat{\beta}_x j g_{i_2,j}$$

The SNP's influence on the relationship between sibling genotypic difference and exposure difference is then obtained by regression:

$$(x_{i1} - x_{i2})^2 = \beta_{xj}\delta_{x_g,j} + e_{ij}$$

and similarly for the outcome

$$(y_{i1} - y_{i2})^2 = \beta_{yj}\delta_{x_g,j} + e_{ij}$$

The $\hat{\beta}_{yj}$ and $\hat{\beta}_{xj}$ estimates are then used in a standard IVW framework.

Results

All estimates based on $V_{GX} = 0.1$ and 90 independent causal variants (i.e. somewhat similar to BMI with latest GIANT results).

1. The standard MR estimate has high power for detecting association, but when confounding and dynastic effects are large there is an extremely high false discovery rate
2. The within-family methods have extremely low power
3. The hybrid approach appears to be well powered and with low FDR for the null model with dynastic effects. Also shows little bias in these simulations.
4. Trio method is comparable to hybrid, with slightly better power when sample sizes are smaller
5. The use of IBD, IBS and unweighted IBS are largely identical

Is there a difference in method performance based on whether IBD, IBS or unweighted IBS is used?

test	eff_xy == 0	sig
Standard MR	FALSE	0.0075769
Standard MR	TRUE	0.0026397
D2 ~ IBD	FALSE	0.2821813

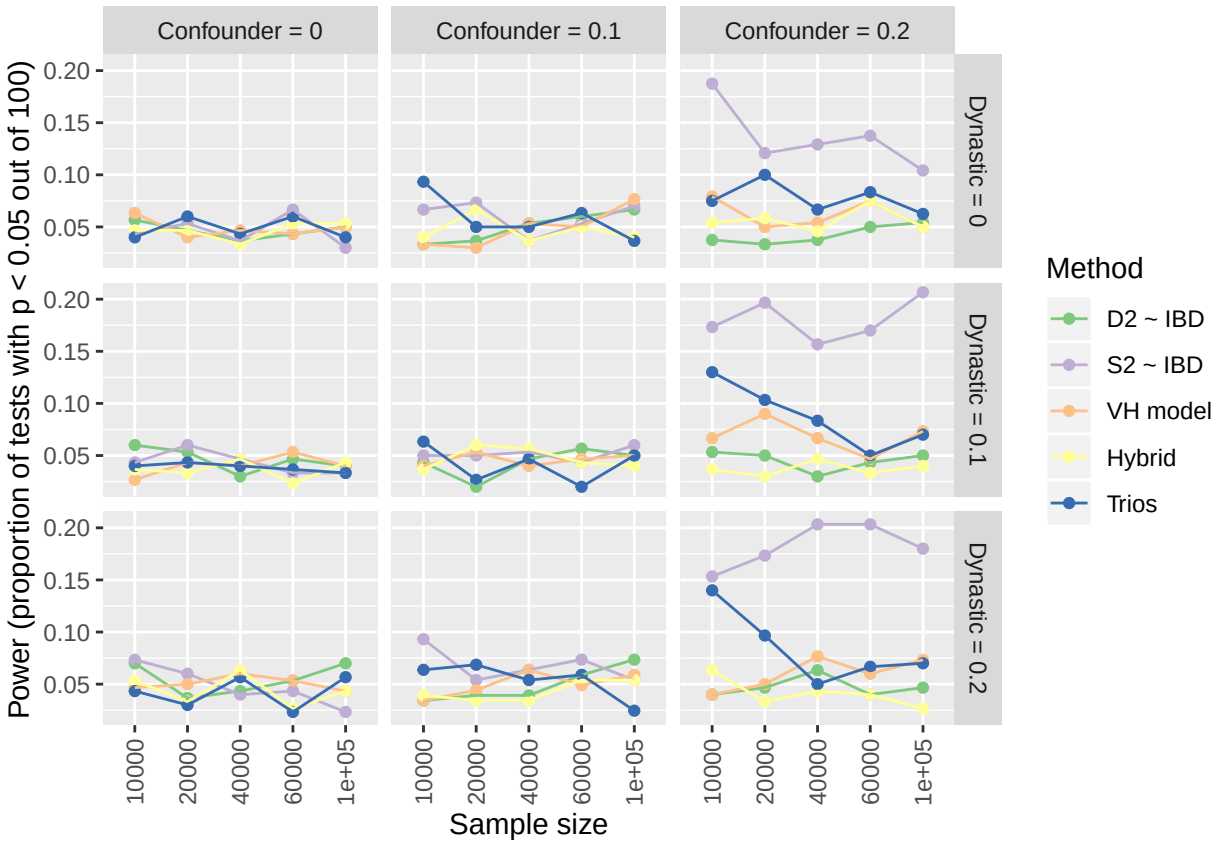

Figure 2: Null model. When confounding and dynastic effects are present then standard MR has very high false discovery rates. Similarly so for the S2 model also.

test	eff_xy == 0	sig
D2 ~ IBD	TRUE	0.2321316
S2 ~ IBD	FALSE	0.5132197
S2 ~ IBD	TRUE	0.3645507
VH model	FALSE	0.0006302
VH model	TRUE	0.1292254
Hybrid	FALSE	0.0000040
Hybrid	TRUE	0.7952225
Trios	FALSE	0.0004679
Trios	TRUE	0.7322850

Plotting again with only hybrid, trio and standard for purpose of cleaner presentations

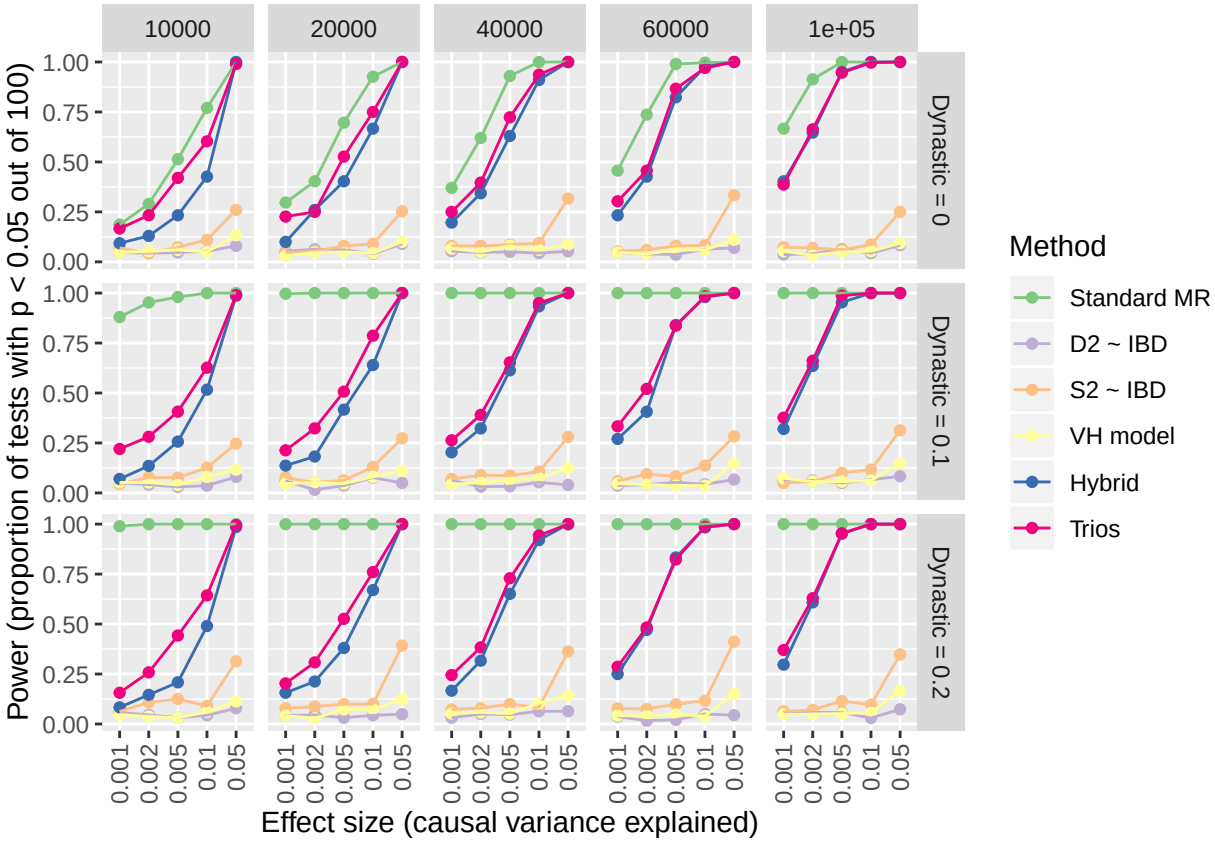

Figure 3: Non-null models where confounders are fixed at 0.1. Standard model has highest power, but hybrid model also performs well. Within-family models are substantially less powerful.

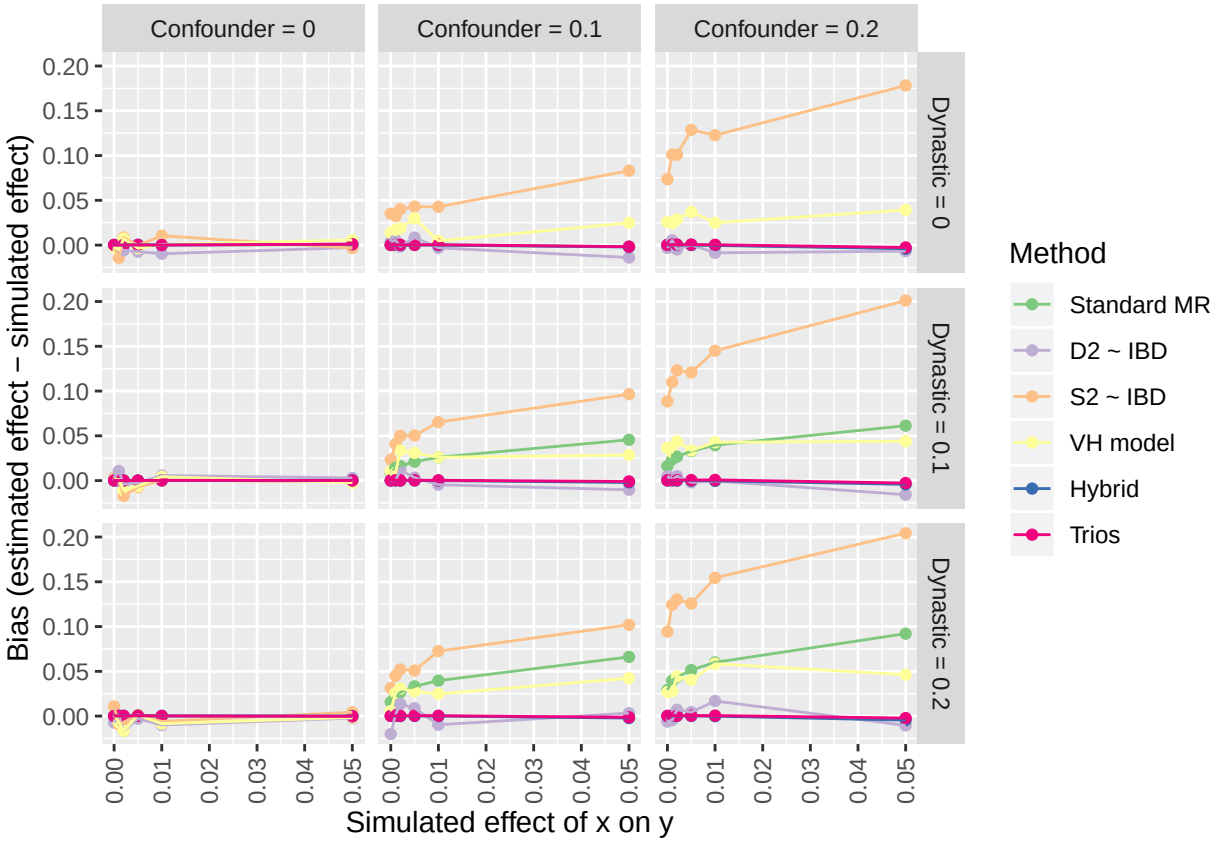

Figure 4: Bias in effect size estimates for different methods. No bias in within family or hybrid methods, but a lot in MR estimates when confounding and dynastic effects are non zero.

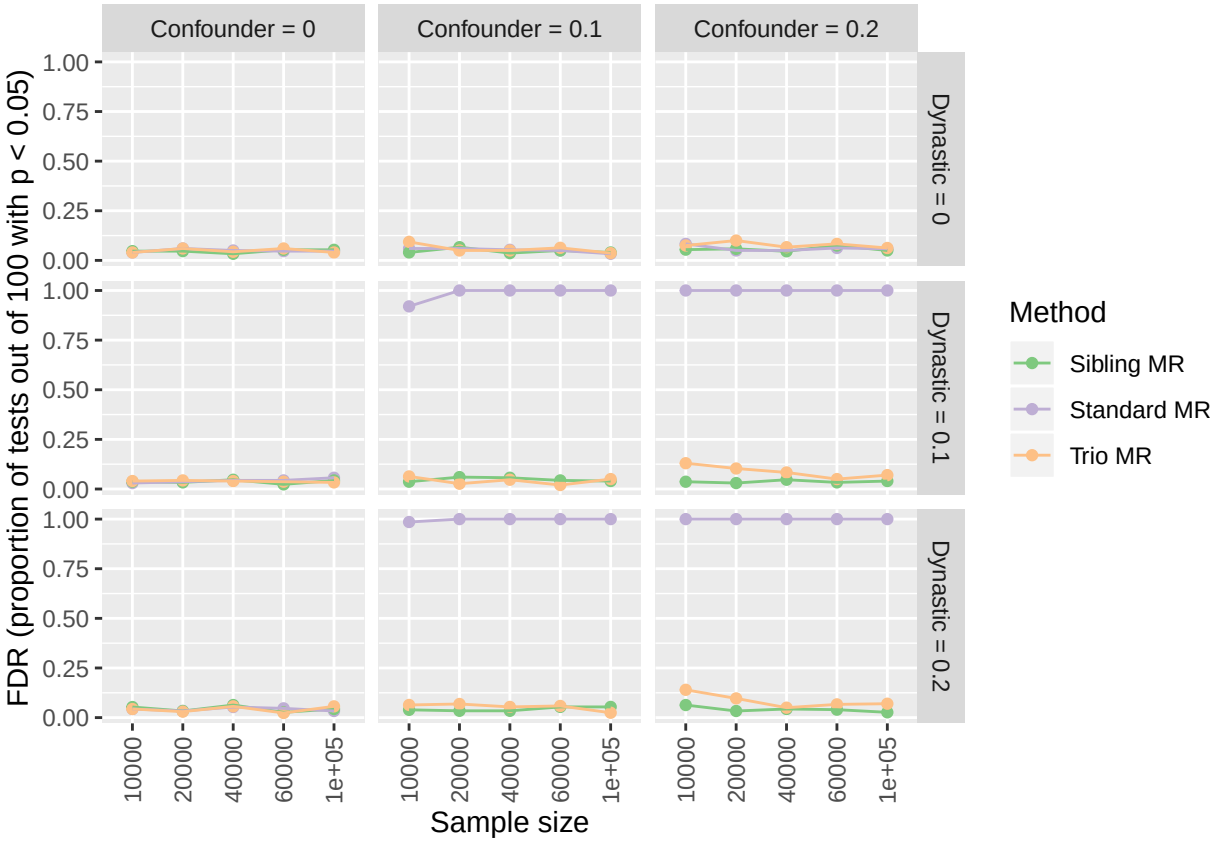

Figure 5: Null model. When confounding and dynastic effects are present then standard MR has very high false discovery rates. Similarly so for the S2 model also.

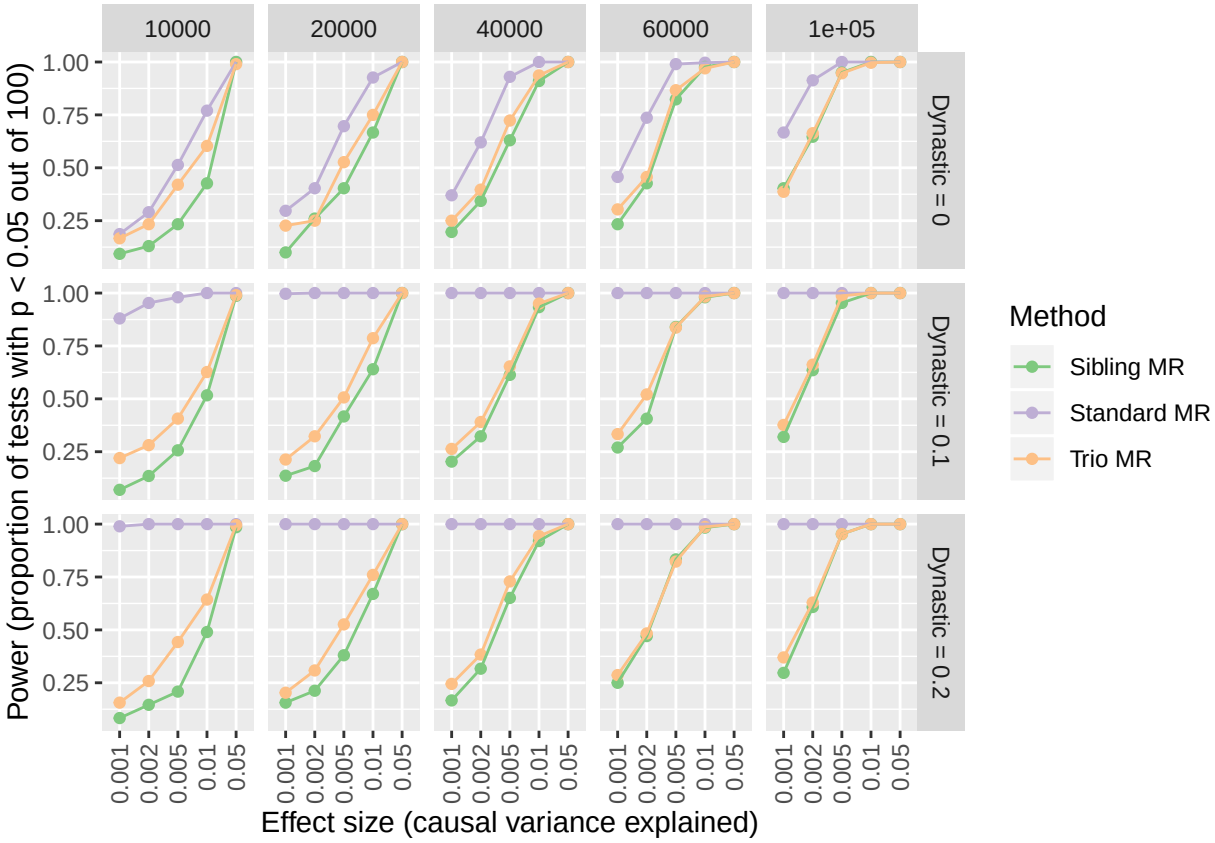

Figure 6: Non-null models where confounders are fixed at 0.1. Standard model has highest power, but hybrid model also performs well. Within-family models are substantially less powerful.

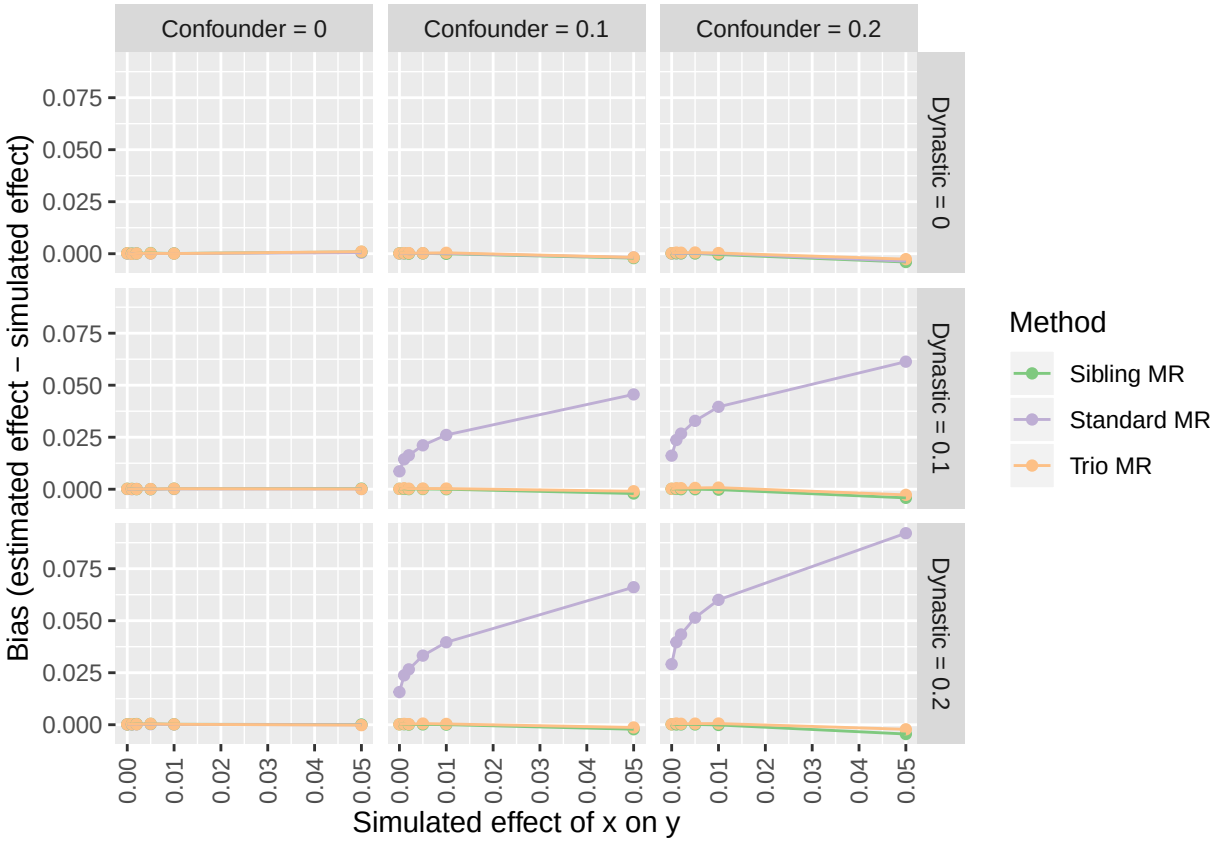

Figure 7: Bias in effect size estimates for different methods. No bias in within family or hybrid methods, but a lot in MR estimates when confounding and dynastic effects are non zero.