Simulations to assess the contribution of survival bias to protective associations of body mass index and Parkinson's disease

Gibran Hemani

2016-07-28

Summary

- Performed simulations where BMI is related to mortality to test if survival bias could induce an apparent causal association between BMI and PD
- The simulations suggest that a survival bias effect is induced, which makes it appear that higher BMI is protective of PD even when there is no real biological link between the two
- However, the estimated effects in real data are substantially larger than these induced survival bias effects, suggesting that a survival bias alone is insufficient to explain the result

Background

Two sample Mendelian randomisation (MR) analysis has shown empirically that there is an effect of increasing body mass index (BMI) on reduced risk of Parkinson's disease (PD). One mechanism that this association could manifest without there being any underlying biological link is if individuals with high BMI have higher mortality rates, and therefore those diagnosed with PD are more likely to harbour alleles that are associated with lower BMI.

This analysis seeks to simulate a population in which mortality is related to BMI, and PD is unrelated to BMI. MR is then performed on the sample to estimate the extent to which survival bias (or frailty) can induce an association between BMI and PD. This frailty effect is then compared against the empirical association obtained from MR.

Simulation strategy

The basic model looks like this:

```
exposure ~ snp(s)
mortality ~ age + exposure
outcome ~ age
```

Simulate a large population ($n = 500000$) where each individual has 78 BMI associated SNPs (Locke et al. 2015) (using only European, LD independent SNPs), PD status, age values, BMI values, and alive/dead status. Data looks like this (first 6 rows):

age	cc	bmi	alive	grs
43.08225	0	24.71988	1	1.832
72.05372	0	22.84763	1	1.881
82.48915	0	24.30631	0	1.734
50.57028	0	28.03450	1	1.646
46.99551	0	23.15997	1	2.019
41.55666	0	23.85864	1	1.883

Age

The age values are generated to match the reported age distributions in (M. A. Nalls et al. 2014):

Sample size	Mean age	SD	Case/control
13708	60.60886	12.71536	1
95282	53.14392	17.52808	0
108990	54.08281	17.17714	NA

BMI SNPs

BMI SNPs are generated as a function of their allele frequencies, such that for individual i at SNP j their genotype value is g_{ij} $Binom(2, p_j)$ where p_j is the allele frequency of SNP j.

BMI values

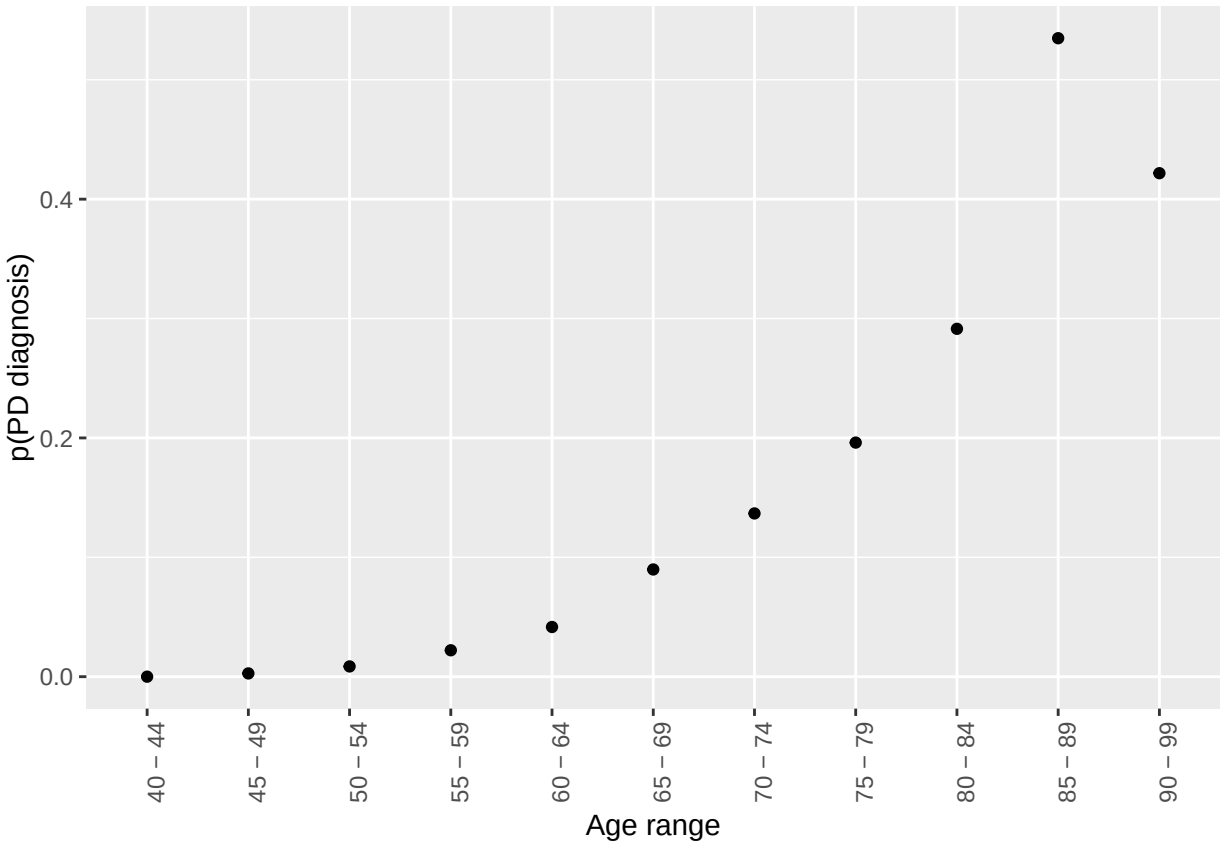
The BMI values are a function of the BMI SNPs, such that

$$x_i = \sum_j g_{ij}\beta_j + e_j$$

where $e_j \sim N(0, \frac{V_E}{V_E+V_G})$, where the genetic variance $V_G = \sum_j 2p_j(1-p_j)\beta_j^2$ and residual variance $V_E = V_P - V_G$. The phenotypic variance, V_P, is the variance of the trait that was used to obtain the effect sizes.

PD status

PD status was simulated as a function of age, based on age related incidence obtained from Driver et al.



The distribution of simulated age stratified by PD looks like this:

Alive/dead status

Alive/dead status was a function of age and BMI values. The baseline survival function was generated from the Gompertz-Makeham law of mortality, with age related hazard function

$$h(t) = a\exp(bt) + \lambda$$

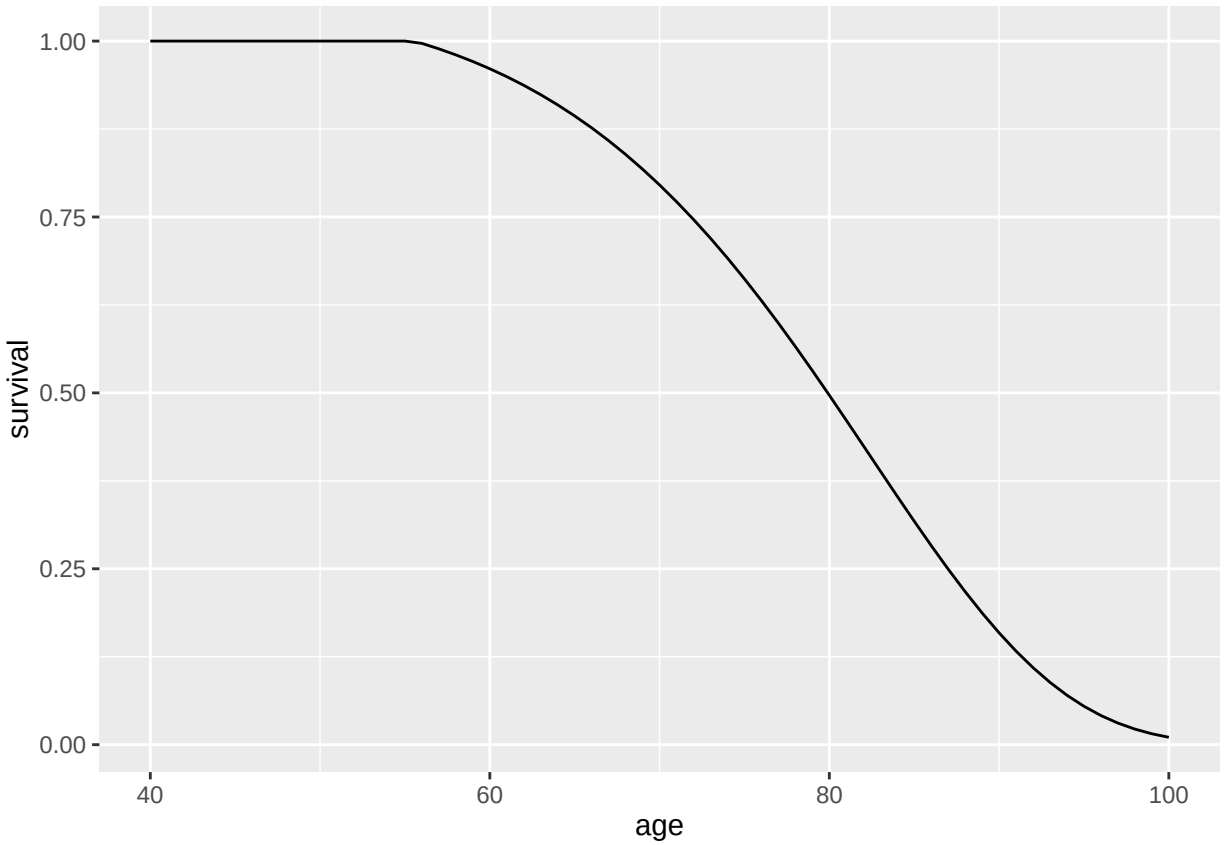
which has CDF

$$F(t) = 1 - \exp(-\lambda t - \frac{a}{b}(e^{bt} - 1))$$

giving the baseline survival function:

$$S_b(t) = 1 - F(t) = -\exp(-\lambda t - \frac{a}{b}(e^{bt} - 1))$$

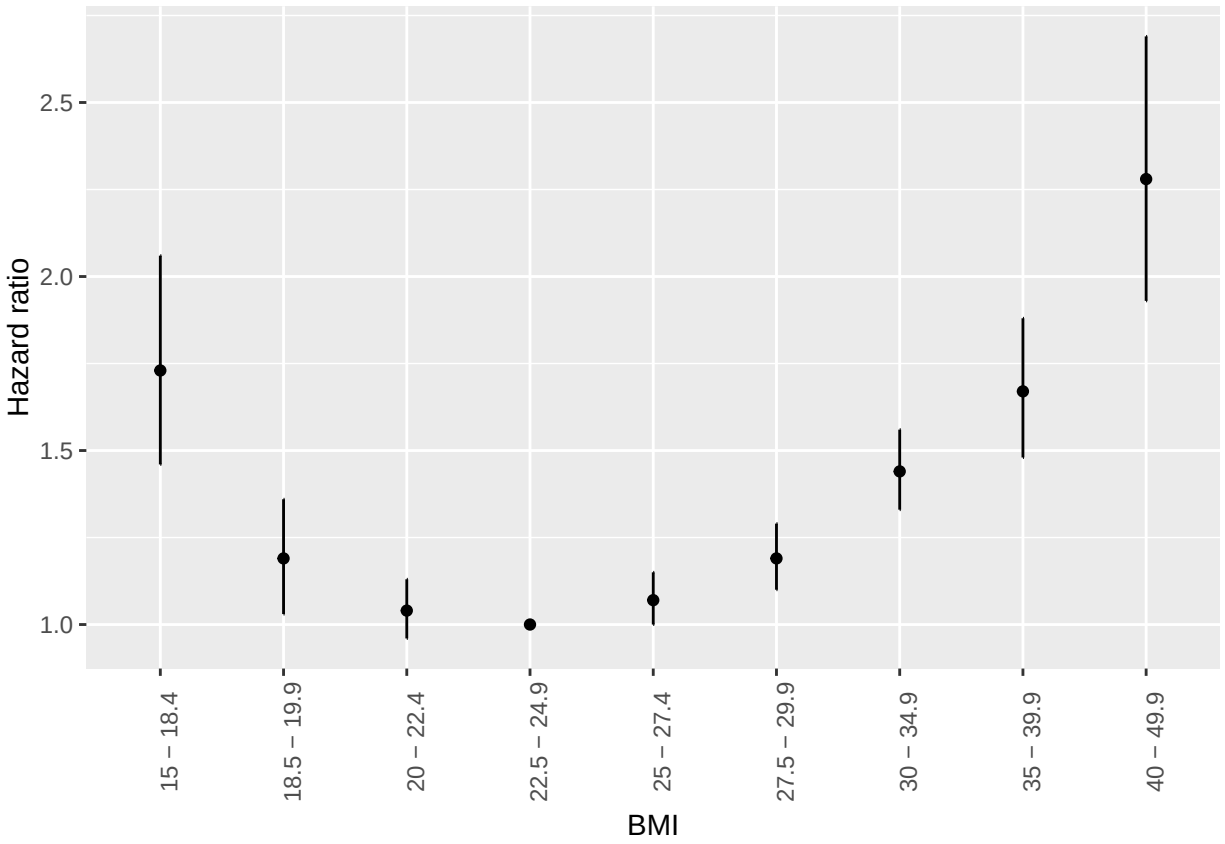
which looks like this:



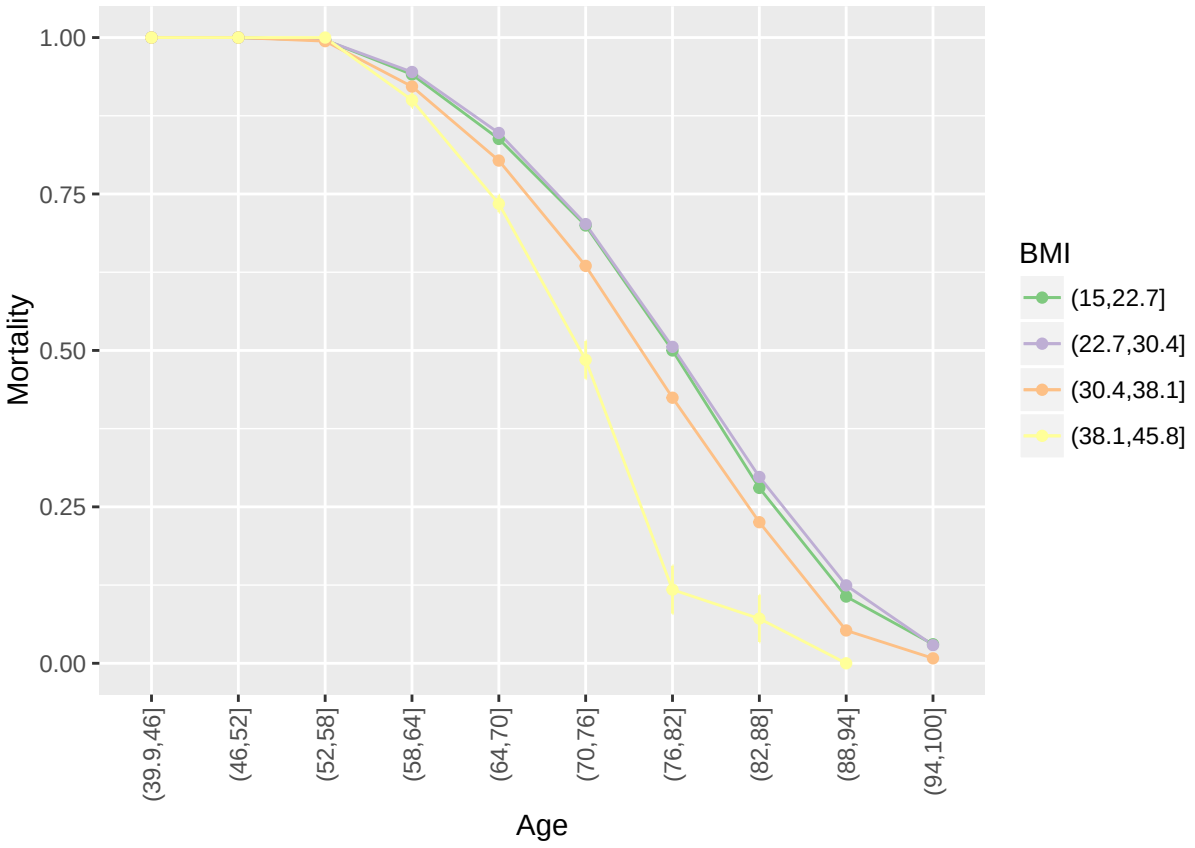
The influence of BMI on survival is then incorporated into the full survival model as

$$S(t) = S_b(t)^{w(x)}$$

where x is the BMI value and $w(x)$ is a function that uses external data to relate BMI with mortality. A J-shaped relationship between BMI and hazard ratios (Berrington de Gonzalez et al. 2010) was simulated:



The influence of BMI on the survival function is shown here, where the curves are the survival functions of quartiles of BMI values:

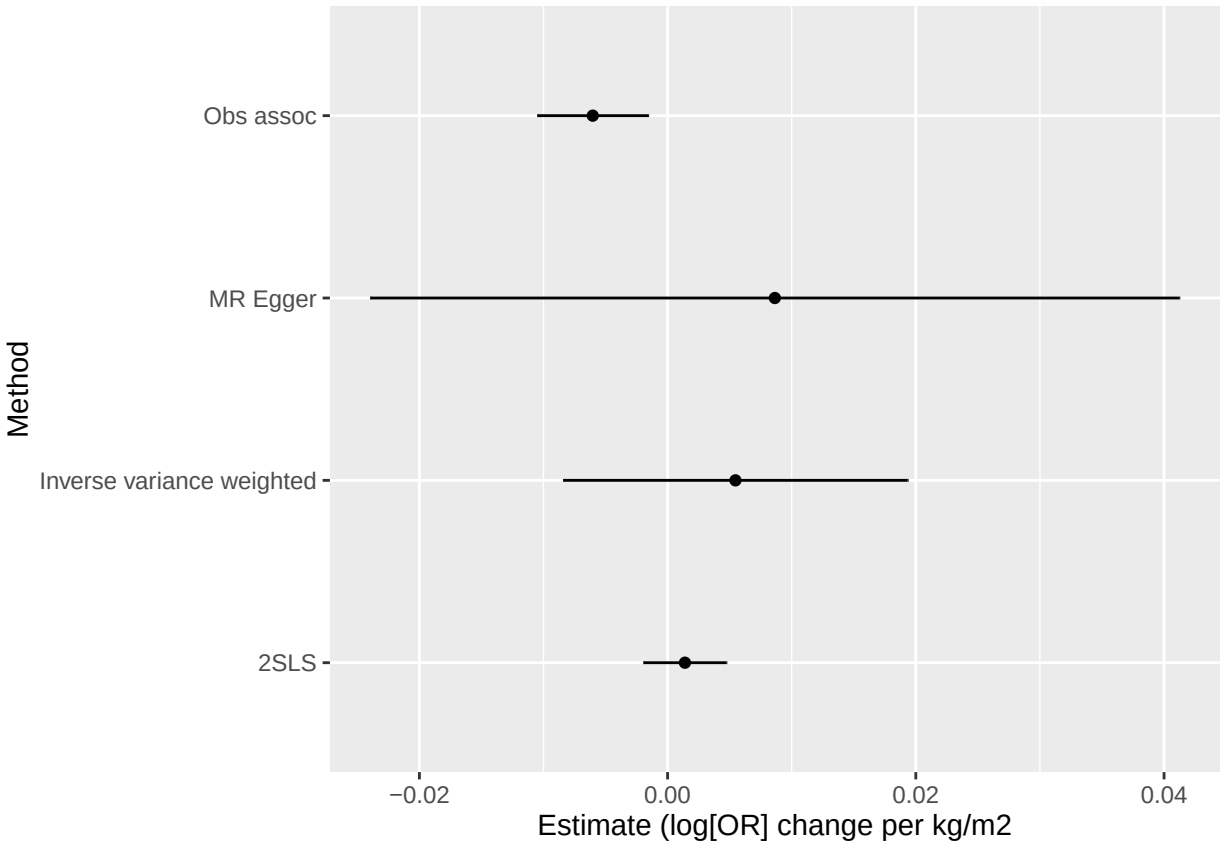


MR analysis of simulated data

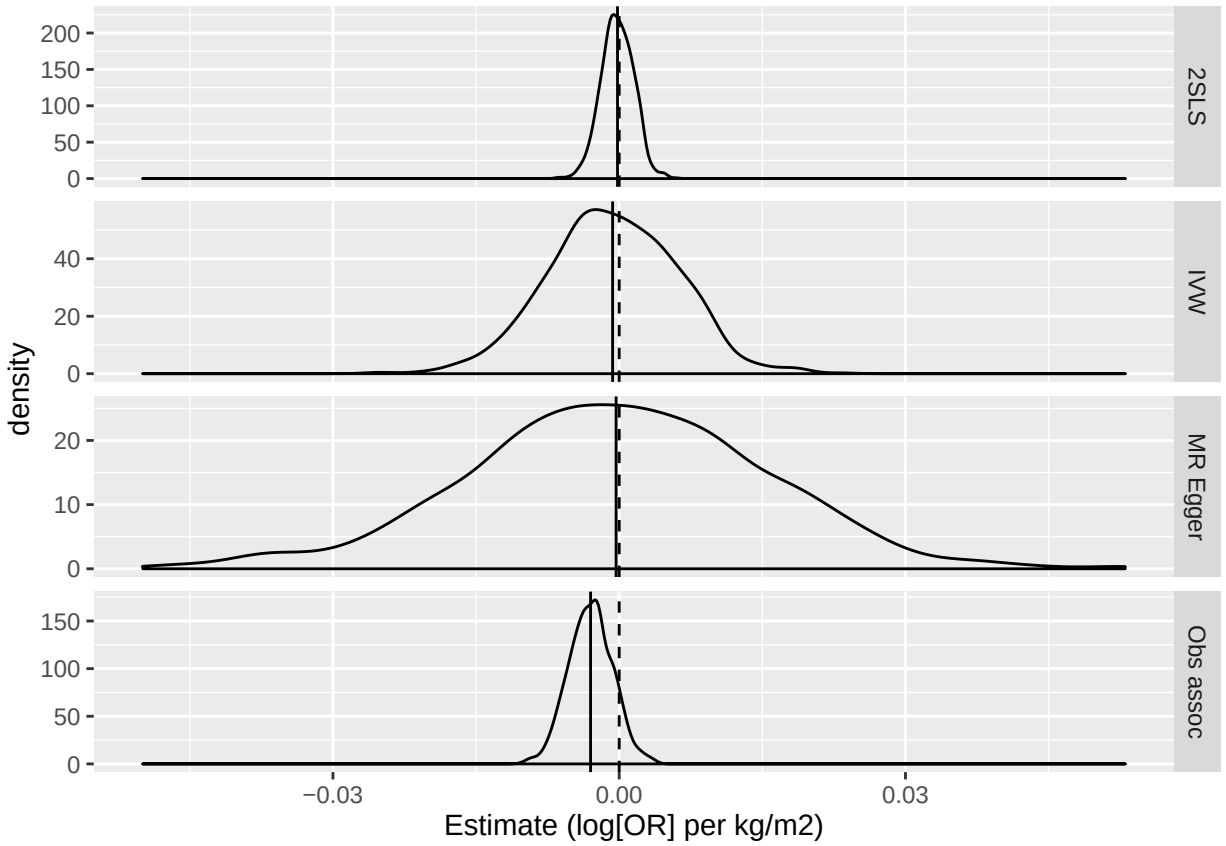
Perform an MR analysis of BMI on PD in the simulated data. A sample of 108990 are randomly sampled from the individuals who have `alive` status to match the sample size and relative numbers of cases and controls in the PD GWAS (M. A. Nalls et al. 2014). With these simulated data the following tests can be performed:

1. Observational association between BMI and PD
2. Two stage least squares estimate of BMI on PD
3. Two sample MR using the BMI effect sizes (Locke et al. 2015) and the estimated association between these 66 simulated SNPs and the simulated PD status. The inverse variance weighting and the MR Egger methods are shown here.

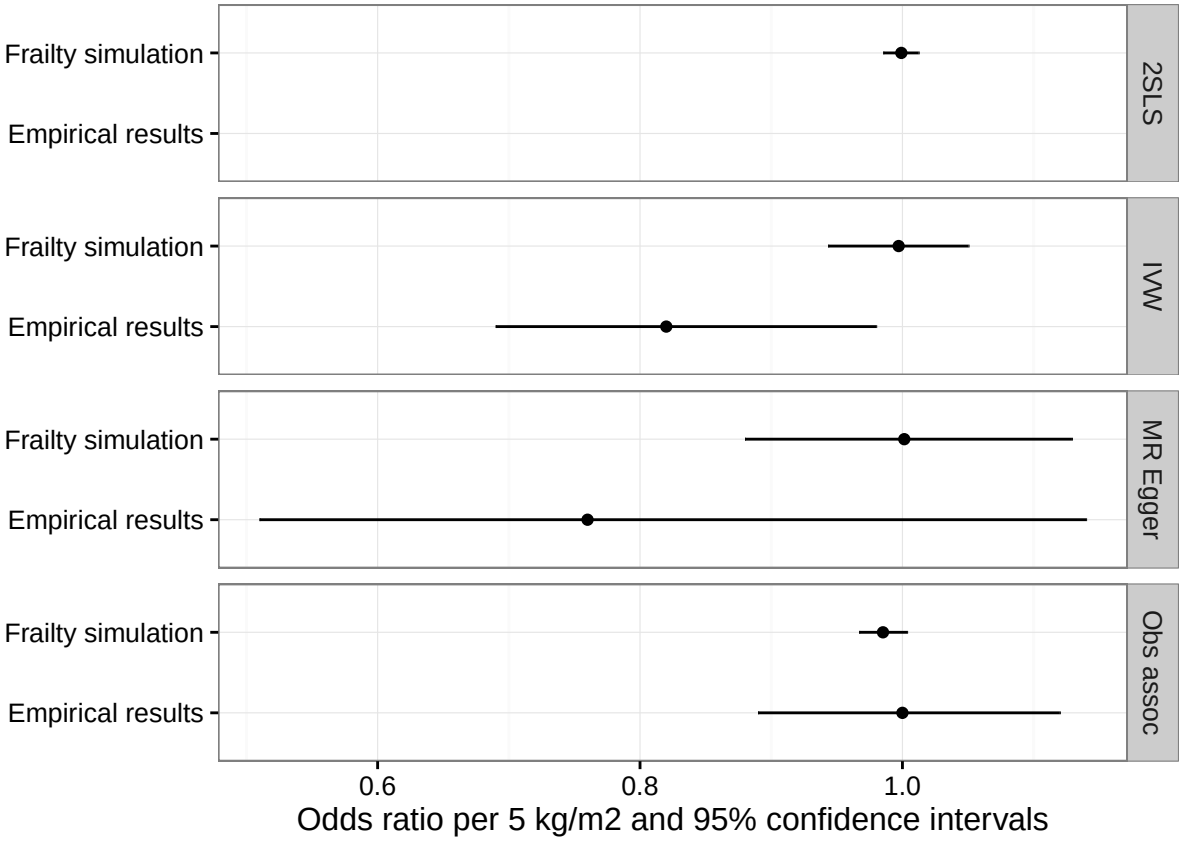
An example of what the result looks like for one simulation is shown here:



Performing the simulation 1000 times allows us to obtain an empirical distribution of these estimates under the null model that BMI is not biologically linked with PD. The results from these simulations look like this:



So for each of the methods (observational association, 2SLS and 2 sample MR) there is evidence that survival bias induces an apparent protective association of BMI on PD. How does this effect relate to those effects that were estimated in the real data?

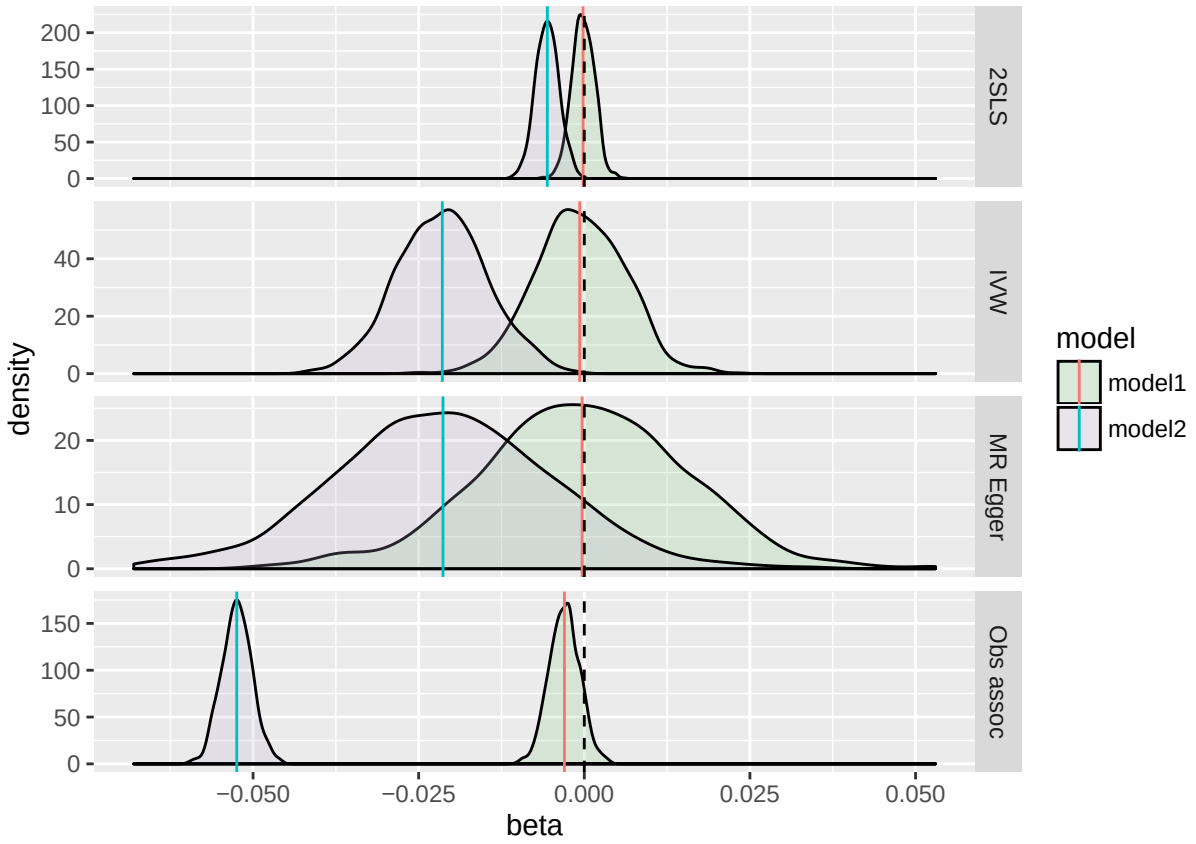


It is clear from this figure that the estimates obtained from the real data are substantially larger than an effect that would be induced spuriously from a frailty effect alone.

Limitations

The simulations rely on external data to provide hazard ratios for BMI and incidence rates for PD. The results are unlikely to be grossly effected by small fluctuations in these estimates.

The frailty simulations allow us to compare the induced associations using three different methods - observational associations, 2SLS, and 2 sample MR. The distribution of estimates from these three methods are not consistent. As an example to illustrate this more clearly, a second simulation was performed using a different hazard model for BMI. Here, a simple and extreme model was used whereby the HR for BMI on mortality was 1 for BMI values less than 27, and 5 for BMI values greater than or equal to 27. The comparison of simulation results from model 1 (J-shaped model described above) and model 2 (extreme) are shown below:



As expected model 2 induces a much larger apparent protective effect, but it's not clear why the different methods give quite drastically different results.

References

Berrington de Gonzalez, Amy, Patricia Hartge, James R. Cerhan, Alan J. Flint, Lindsay Hannan, Robert J. MacInnis, Steven C. Moore, et al. 2010. "Body-Mass Index and Mortality among 1.46 Million White Adults." *New England Journal of Medicine* 363 (23): 2211–19. doi:10.1056/NEJMoa1000367.

Locke, Adam E., Bratati Kahali, Sonja I. Berndt, Anne E. Justice, Tune H. Pers, Felix R. Day, Corey Powell, et al. 2015. "Genetic studies of body mass index yield new insights for obesity biology." *Nature* 518 (7538): 197–206. doi:10.1038/nature14177.

Nalls, Mike A, Nathan Pankratz, Christina M Lill, Chuong B Do, Dena G Hernandez, Mohamad Saad, Anita L DeStefano, et al. 2014. "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease." *Nature Genetics* 46 (9). Nature Publishing Group: 989–93. doi:10.1038/ng.3043.