


# Conditioning on a Collider May Induce Spurious Associations: Do the Results of Gale et al. (2017) Support a Health-Protective Effect of Neuroticism in Population Subgroups?

Tom G. Richardson<sup>1,2</sup> , George Davey Smith<sup>1,2</sup>, and Marcus R. Munafò<sup>1,3</sup>

<sup>1</sup>MRC Integrative Epidemiology Unit, University of Bristol; <sup>2</sup>Population Health Sciences, Bristol Medical School, University of Bristol; and <sup>3</sup>UK Centre for Tobacco and Alcohol Studies, School of Psychological Science, University of Bristol

Received 12/9/17; Revision accepted 4/11/18

Gale and colleagues (2017) examined the association between neuroticism and mortality in a large sample ( $N > 300,000$ ) drawn from the UK Biobank study (Sudlow et al., 2015). They observed that neuroticism was associated with higher all-cause mortality but that following adjustment for self-rated health, neuroticism was associated with lower all-cause mortality. Further analyses stratified on self-rated health suggested that higher neuroticism was associated with reduced mortality only among individuals with fair or poor self-rated health. The authors concluded that “neuroticism becomes protective against mortality from all causes and cancer in people with fair or poor self-rated health” (p. 1355), a finding that generated substantial interest (Macmillan, 2017), reflected in an Altmetric score (at the time of writing) of 416.

The availability of very large cohort studies such as UK Biobank in principle allows researchers to identify associations where the absolute effect size may be small but population-level impact considerable (as is the case of the results reported by Gale and colleagues). This is of increasing relevance as cohort studies continue to grow in scale, given that the introduction of even modest bias could lead to robust, but spurious, findings. For instance, when two variables independently influence a third variable and that third variable is conditioned on, this can induce collider bias, which can distort observed associations (Greenland, 2003; Munafò, Tilling, Taylor, Evans, & Davey Smith, 2018).

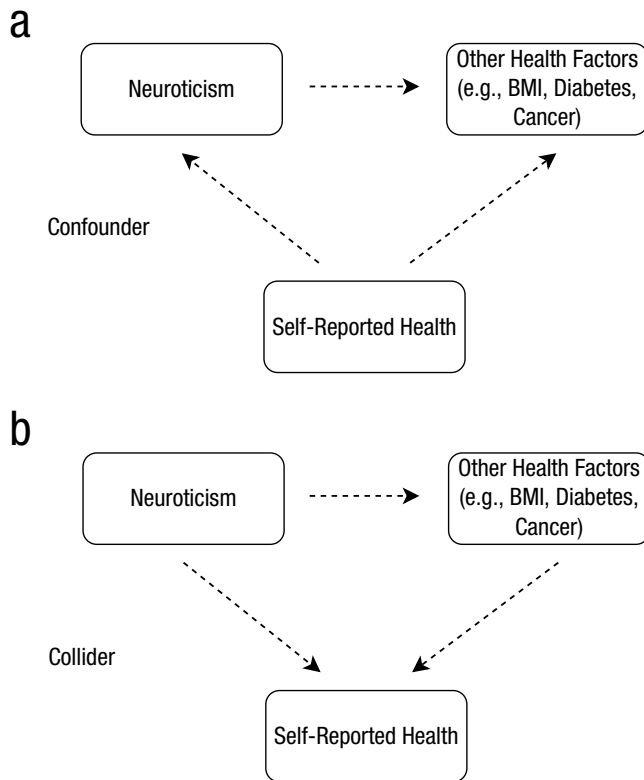
In the case of neuroticism, self-reported health, and mortality, it is plausible that both neuroticism and risk

factors for all-cause mortality might influence self-reported health (note that neuroticism could do this by generating less favorable self-reporting of health at any objective level of health status). In that case, conditioning on self-reported health might induce collider bias and generate spurious or distorted associations between neuroticism and both risk factors associated with all-cause mortality and all-cause mortality itself. However, if self-reported health were known to influence neuroticism and risk factors for all-cause mortality, then it would be a confounder and should not lead to distorted findings when conditioned on (Fig. 1).

We explored this possibility using the same sample drawn from UK Biobank as used by Gale and colleagues. This was done by examining the association between neuroticism and a range of risk factors known to be associated with all-cause mortality, both unstratified and stratified by self-reported health, as stratifying on a collider is one way to condition on it. Specifically, we first analyzed all individuals in the sample (i.e., unstratified) and then repeated our analyses within each of the four different subgroups within the sample, on the basis of self-reported health (i.e., stratified).

## Corresponding Author:

Tom G. Richardson, MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom  
E-mail: Tom.G.Richardson@bristol.ac.uk



**Fig. 1.** A directed acyclic graph demonstrating the difference between a confounder and a collider within the context of this study. The figure illustrates two possible scenarios when conditioning on self-reported health in the analysis between neuroticism and risk factors for mortality: (a) Self-reported health is a confounder that influences both neuroticism and risk factors for mortality and is therefore appropriate to condition on; (b) self-reported health is a collider that both neuroticism and risk factors for mortality influence, thereby potentially leading to spurious findings when conditioned on. BMI = body mass index.

## Method

We reproduced the analyses reported by Gale and colleagues as closely as possible using data derived from the UK Biobank study. A full list of the variables we used can be found in Table S1 in the Supplemental Material available online. To verify that our data set was similar to the one analyzed by Gale and colleagues, we used Cox proportional-hazards regression to reproduce the hazard ratios for all-cause mortality in all individuals and within each self-rated health stratum, as reported in their article.

Linear and logistic regression were used to assess the relationship between neuroticism score and each covariate in turn (the covariates are listed by Gale and colleagues in Table 1 of their study) for continuous and binary traits, respectively, with adjustment for age and sex. Analyses for each covariate were then repeated after stratifying individuals according to their self-rated

health status. Our interest was in the comparison between the unstratified and stratified analyses.

## Results

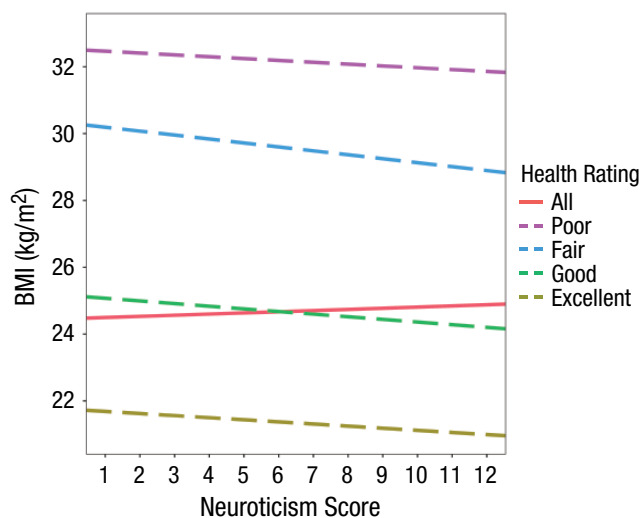
We were able to reproduce the observations reported by Gale and colleagues when evaluating the relationship between neuroticism and mortality (Table S2 in the Supplemental Material). Specifically, we observed a hazard ratio greater than 1 when analyzing all individuals in our sample, adjusting for age and sex ( $p < 1.0 \times 10^{-16}$ ). In contrast, hazard ratios less than 1 in all four strata of self-reported health with  $p$  values less than .001 were observed in the “fair” and “poor” self-reported health strata.

However, we also observed evidence suggesting that conditioning on self-reported health status may strongly influence the relationship between neuroticism and other risk factors in this study (Table S3 in the Supplemental Material). In particular, we observed an instance of Simpson’s (1951) paradox when assessing the relationship between neuroticism and body mass index (BMI) after stratifying by self-reported health status. This occurs when an association seen in an overall sample attenuates, disappears, or is reversed in each complete set of subgroups (Hernan, Clayton, & Keiding, 2011).

Figure 2 illustrates this example of Simpson’s paradox, where there is a negative association between neuroticism and BMI in every stratum based on self-reported health but a positive association in the unstratified analysis. This is similar to the effect observed between neuroticism and mortality by Gale and colleagues in the age- and sex-adjusted analyses. Further examples of collider bias were also observed in analyses with other risk factors (with the exception of reaction time and Townsend index), with particularly marked effects observed when analyzing forced expiratory volume, cancer, and diabetes (see Table S3). When the direction of an association between two variables becomes reversed when conditioning on a variable, statistical reasoning alone cannot identify the appropriate model. In the present situation, we consider the collider model shown in Figure 1b to be more plausible (Hernan et al., 2011).

## Discussion

Our results suggest that the findings reported by Gale and colleagues should be interpreted in the context of the potential for collider bias. Specifically, conditioning on self-reported health status strongly influenced the relationship between neuroticism and a range of risk factors known to be associated with mortality. Two



**Fig. 2.** The relationship between body mass index (BMI) and neuroticism after conditioning on a collider. Results are shown separately for four strata of self-reported health and also from the unstratified analysis. Regression lines were obtained from the analysis undertaken in the UK Biobank study.

factors lead us to believe that these associations are spurious. First, for many risk factors (e.g., BMI), the associations are clearly negative in every stratum but positive in the unstratified analysis, indicating that a form of Simpson's paradox is operating. Second, we do not consider it likely that neuroticism could have the kind of protective effect suggested by Gale and colleagues across all of the risk factors we observed (particularly given evidence of Simpson's paradox). Although the statistical considerations of this Commentary suggest this, it would also be worthwhile using triangulation (i.e., investigating results derived from various approaches that rest on different and, ideally, orthogonal assumptions) to confirm this (Lawlor, Tilling, & Davey Smith, 2016).

The results we observed could be due to a confounding effect of self-reported health status on neuroticism and other risk factors or due to collider bias if neuroticism causes self-reported health status. Differentiating between these possibilities would require stronger evidence that neuroticism causes self-reported health status, for example, using Mendelian randomization (Davey Smith & Ebrahim, 2003; Davey Smith & Hemani, 2014). This method can be used to infer causal relationships among correlated traits in epidemiology by using genetic variants as instrumental variables. Investigating the genetic contribution to distinct facets of neuroticism should also prove worthwhile in terms of investigating causal relationships in this paradigm (Hill, Weiss, McIntosh, Gale, & Deary, 2017). However, it is worth noting that collider bias can still influence the analysis of genetic factors, which are protected from some

biases in observational studies but not from this form of bias (Munafò et al., 2018).

It is unclear in the study by Gale and colleagues whether risk factors and self-reported health were considered potential mediators of the effect of neuroticism on mortality when adjusted for. Although this may not be the case in the study by Gale et al., it should be noted that adjusting for mediators without considering the implications of doing so may also lead to biased results (Rohrer, 2018).

Overall, our results serve as a cautionary note that while large cohort studies provide unparalleled power to elucidate associations between risk factors and disease outcomes, the ability to detect ever smaller effect sizes increases the risk that relatively weak biases may distort our findings. In other words, with great (statistical) power comes great responsibility.


### Action Editor

D. Stephen Lindsay served as action editor for this article.

### Author Contributions

G. Davey Smith and M. R. Munafò designed the study. T. G. Richardson performed the statistical analyses. All of the authors contributed to the initial draft of the manuscript and approved the final version for submission.

### ORCID iD

Tom G. Richardson  <https://orcid.org/0000-0002-7918-2040>

### Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

### Funding

This work was supported by the Medical Research Council and the University of Bristol (MC\_UU\_12013/1, MC\_UU\_12013/6, MC\_UU\_00011/1, and MC\_UU\_00011/7). T. G. Richardson is a UK Research and Innovation Fellow (MR/S003886/1). M. R. Munafò is a member of the UK Centre for Tobacco and Alcohol Studies, a UK Clinical Research Collaboration (UKCRC) Public Health Research Centre of Excellence. Funding from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UKCRC, is gratefully acknowledged. UK Biobank data were analyzed as part of Project 15825.

### Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618774532>

## References

- Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*, 1–22.
- Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–R98.
- Gale, C. R., Čukić, I., Batty, G. D., McIntosh, A. M., Weiss, A., & Deary, I. J. (2017). When is higher neuroticism protective against death? Findings from UK Biobank. *Psychological Science*, *28*, 1345–1357. doi:10.1177/0956797617709813
- Greenland, S. (2003). Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology*, *14*, 300–306.
- Hernan, M. A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology*, *40*, 780–785.
- Hill, W. D., Weiss, A., McIntosh, A. M., Gale, C. R., & Deary, I. J. (2017). *Genetic contribution to two factors of neuroticism is associated with affluence, better health, and longer life*. Retrieved from bioRxiv: <https://www.biorxiv.org/content/10.1101/146787v1>
- Lawlor, D. A., Tilling, K., & Davey Smith, G. (2016). Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, *45*, 1866–1886.
- Macmillan, A. (2017, July 25). Being neurotic may help you live longer. *Time*. <http://time.com/4872545/neurotic-longer-life/>
- Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Davey Smith, G. (2018). Collider scope: When selection bias can substantially influence observed associations. *International Journal of Epidemiology*, *47*, 226–235. doi:10.1093/ije/dyx206
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*, 27–42. doi:10.1177/2515245917745629
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society B: Methodological*, *13*, 238–241.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Collins, R. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, *12*(3), Article e1001779. doi:10.1371/journal.pmed.1001779