
Peer reviewed version

Link to published version (if available):
10.1177/0956797618774532

Link to publication record in Explore Bristol Research
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Sage at DOI: 10.1177/0956797618774532. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
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 sub-groups?

Tom G. Richardson 1,2, George Davey Smith 1,2, Marcus R. Munafò 1,3

1. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
3. UK Centre for Tobacco and Alcohol Studies School of Experimental Psychology, University of Bristol, Bristol, UK

Corresponding author: Tom G. Richardson, MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK. Tel: +44 (0)117 3313370; E-mail: Tom.G.Richardson@bristol.ac.uk
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 sub-groups?

Introduction

Gale and colleagues (Gale et al., 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 those 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”, a finding that generated substantial interest (Time, 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 upon, this can induce collider bias, which can distort observed associations (Greenland, 2003; Munafo, Tilling, Taylor, Evans, & Davey Smith, 2017).

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 favourable self-reporting of health at any objective level of health status). In that case, conditioning upon 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 was 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 upon (Fig 1).

**Figure 1. A directed acyclic graph demonstrating the difference between a confounder and a collider within the context of this study.**

An illustration of two possible scenarios when conditioning upon 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 upon; b) self-reported health is a **collider** which both neuroticism and risk factors for mortality influence, therefore potentially leading to spurious findings when conditioned upon.
We explored this possibility using the same sample drawn from UK Biobank as that 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 upon a collider is one way to condition upon it. Specifically, we firstly analysed all individuals in the sample (i.e., unstratified) and then repeated our analyses within each of the four different subgroups within the sample, based on self-reported health (i.e., stratified).

**Methods**

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 Supplementary Table 1. To verify that our dataset was similar to the one analysed 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 strata as reported by their study.

Linear and logistic regression were used to assess the relationship between neuroticism score and each covariate in turn (as shown in Table 1 in the study by Gale and colleagues) 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 (Supplementary Table 2). Specifically, we observed a hazard ratio > 1 when analysing all individuals in our sample adjusting for age and sex ($P < 1.0 \times 10^{-16}$). In contrast, hazard ratios < 1 in all 4 strata of self-
reported health with p values < 0.001 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 (Supplementary Table 3). In particular, we observed an instance of Simpson’s Paradox (Simpson, 1951) when assessing the relationship between neuroticism and body mass index after stratifying by self-reported health status. This occurs when an association seen in an overall sample attenuates, disappears, or is reversed in each of complete set of subgroups (Hernan, Clayton, & Keiding, 2011).

Fig. 2 illustrates this example of Simpson’s paradox, where there is a negative association between neuroticism and body mass index in every stratum based on self-reported health, but a positive association in the unstratified analysis. This is the 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 analysing forced expiratory volume, cancer and diabetes (see Supplementary Table 3). When the direction of an association between two variables becomes reversed upon conditioning upon a variable, statistical reasoning alone cannot identify the appropriate model. In the present situation we consider our model (b) in Fig. 1 to be more plausible (Hernan et al., 2011).
Figure 2. An illustration of potential collider bias in the UK Biobank study when assessing the relationship between neuroticism and body mass index.

Regression lines from the analysis between neuroticism and body mass index in the UK Biobank study. We observed a positive association in the unstratified analysis (i.e., all individuals, shown in red), but the opposite when we stratified on self-reported health (i.e., a negative association in all strata).
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 influences the relationship between neuroticism and a range of risk factors known to be associated with mortality. Two 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 amongst 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 (Munafo et al., 2017).

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.
Acknowledgements

This work was supported by the Medical Research Council and the University of Bristol (MC_UU_12013/1, MC_UU_12013/6). T.G.R. is a UKRI Innovation Research Fellow (MR/S003886/1). M.R.M. is a member of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research Centre of Excellence. Funding from 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 UK Clinical Research Collaboration, is gratefully acknowledged. UK Biobank data was analysed as part of project 15825.
References


