Commentary

To cause or not to cause? That is the question, but identical twins might not have all of the answers

Jason D. Boardmana,⁎, 1, Jason M. Fletcherb, 1

⁎ University of Colorado, Boulder, CO, United States
b University of Wisconsin, Madison, United States

We are pleased to respond to a request for comments on the utility of identical twins to infer causality with respect to the education and health association. Specifically, we were asked to outline some thoughts on the provocative, though not unprecedented, findings from Amin et al. (2014) who show that cross-sectional correlations between years of schooling and multiple measures of health are not robust to specifications comparing educational and health differences among identical twin pairs. As practitioners of twin and sibling differencing models on a variety of topics (Boardman et al., 2011; Boardman, 2009; Fletcher, 2011, 2013, 2014a,b,c), in general we are sympathetic with this approach to assessing potentially confounded relationships. Importantly, the thought experiment of “fixing” genetic variation with identical twin pairs, though without with the advantages and disadvantages of twin samples, has been extended to unrelated persons to gain insights in the exogenous influence of education on health (Boardman et al., 2014). Our comments will first remind readers of the empirical benefits and consequences of using within family analyses and then turn to specific thoughts of utilizing identical twins for this research and we will focus our attention on the Amin et al. findings.

The question of interest—the potential magnitude of the causal linkages between completed schooling and later health outcomes—is of fundamental importance to many branches of social science, but providing valid and reliable empirical evidence on these purported causal effects has proven to be challenging. A key concern is the role of confounding influences such as genetic factors and/or family and environmental factors that are correlated with both schooling and health. Farrell and Fuchs (1982) was among the first to suggest that the ability to delay gratification may be linked to both more education and better health, therefore questioning the causal nature of the observed association. In a recent paper published in this journal, Madsen et al. (2014) use a sample of nearly 12,000 identical twin pairs to address a comparable question as Amin et al. (2014), and they reach similar conclusions: the protective effects of education on health (measured by cardiovascular disease in Madsen et al.) are often significantly reduced or eliminated when examining within twin pair associations. It is therefore important to consider that there is consistent evidence that the links between education and health are either significantly reduced, or completely eliminated, using the within identical twin approach. Our comments are not meant to challenge these results. Rather our goal is to discuss limitations that should be considered when interpreting the meaning of these results. Our comments should be considered in light of these ongoing debates (Gilman and Loucks, 2014) and we hope to provide additional points of clarification.

To briefly repeat what has been outlined in the literature (especially see Bound and Solon, 1999), while sibling/twin models have many advantages in adjusting for potential genetic and environmental confounding, they do present several important statistical as well as conceptual limitations; to be sure, some of these are discussed in Amin et al. An overarching issue is describing the thought experiment that would produce causal estimates that the twin method might attempt to mimic. One such thought experiment, mirroring an actual RCT, is to consider a case where we could randomly assign individuals (twins or otherwise) to attain education around age 20 had better health outcomes around age 50. Turning to twin/sibling models, what assumptions do we need

⁎ Corresponding author.
E-mail address: boardman@colorado.edu (J.D. Boardman).
1 Author names appear alphabetically denoting equal contribution to the paper.

An alternative implementation of this thought experiment has been the use of compulsory schooling law changes in the mid 1900s in the US and around the world, where researchers use an instrumental variables strategy to ask whether law-induced changes in education seem to affect later health status. See Fletcher (this issue) for further discussion and an example of the approach.

http://dx.doi.org/10.1016/j.socscimed.2014.10.013
0277-9536/© 2014 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Boardman, J.D., Fletcher, J.M., To cause or not to cause? That is the question, but identical twins might not have all of the answers, Social Science & Medicine (2014), http://dx.doi.org/10.1016/j.socscimed.2014.10.013
to make in order to mirror our thought experiment? We would need to believe that any within-twin differences in schooling are as-good-as (conditionally) randomly determined; they are uncorrelated with any unobserved (to the researcher) within-twin factors (e.g. motivation, time preference, etc.) that might be related to both health and educational outcomes. Only if we believe this “as-if” randomization is at play, should we think of the results presented in Amin et al. (and noted, perhaps prematurely, as “causal effects” in the abstract) as actual causal estimates. Indeed, to the extent that using twin differences removes exogenous (“experimental”) within-twin variation from the data, results using this approach could be increasing the bias of the “naïve” results due to a reliance mainly/only on endogenous (e.g. motivation) variation (Bound and Solon, 1999) and therefore, moving us away from causal effects instead of toward them. Another way of thinking about this issue is to ask the questions: once controls for genetic and family factors are included, what variation is “left over” to estimate effects on health? And is this variation (wherever it comes from) policy relevant variation? While there are issues with alternative sources of variation (e.g. compulsory schooling laws that shift educational attainments), one might reasonably suggest that these sources of variation are both (1) clear and measurable and (2) policy relevant.

With this backdrop, we believe that there are three important concerns regarding the use of identical twins to infer causality in the social sciences.

1. The meaning of discordance among identical twins

The bulk of the criticisms regarding the use of twins for social scientific and health research has focused primarily on the estimation of heritability (e.g., the proportion of phenotypic variance that is due to genetic variance in the population) because these models assume that monozygotic (MZ) twin pairs share an environment that is, on average, comparable to the shared environment of dizygotic (DZ) twins (Burt and Simons, 2014). While it is important to note that this assumption is not relevant to the use of MZ twins only, it is nevertheless important to consider the constrained variance for phenotypic discordance (e.g., educational differences) among MZ twins compared to the general population. Because the fixed effects model requires that MZ twins to be discordant for both education and health, it forces us to make generalizations about the larger population from a notably select group of identical twin pairs. To illustrate this point we examined the adult MZ pairs from the MIDUS data1 (Brim et al., 1995). Of the 339 pairs with complete information on education and self-rated health the mean level of educational difference (years) among pairs was comparable in our analysis and the Amin et al. analysis (absolute within MZ pair diff = 1.11, sd = 1.45) and only 91 (~27%) pairs were discordant for both education and health. This selection is also true in the Virginia twin data in which 71% of twin pairs had no difference in their reported schooling alone. Importantly, the discordant pairs in the MIDUS data had a significantly (p < .0043) higher mean level of education (mean = 14.53 years) compared to the concordant pairs (mean = 13.79 years). This was not the case for self-rated health. Lundborg (2013) uses these same data to evaluate the causal nature of the education—health association and only shows an association between high school degree vs. less than high school but no educational returns after that point.

Similarly, as described in detail elsewhere (Gilman and Loucks, 2014), the reduction in sample size described above reduces the precision of the estimate. That is, a comparably sized coefficient may be rendered statistically indistinguishable from zero simply because the reduction in sample size inflates the standard error. Thus, the risk of type II error may be enhanced simply because of the reduction in sample size. Amin et al. (2014) have an adequately powered sample for these purposes, however, it is important to consider this limitation for other within MZ pair analyses.

Finally, it is important to consider the meaning of discordance with respect to the age of the twins. Put simply, these methods require both twins to be alive, which means that older twins are increasingly select with respect to health. To illustrate, the mean health of MZ twins in the MIDUS study for those born between 1920 and 1939 (ages 75–56) is identical to those born between 1940 and 1949 (ages 46–55) suggesting that more healthy older pairs are selected. Similarly, there is virtually no association between age and self-rated health among the MZ pairs born prior to 1950 ($r = -.17$, $p < .03$) among those born after 1950. With respect to the Amin et al. paper, the twin registry for the WWII sample is nearly 16,000, of whom only 1902 pairs can be analyzed. It is difficult to know the extent to which these observations affect the results of within MZ analyses, but it is important to consider these limitations when generalizing the results to the broad population.

2. Fixing genotype does not mean that we are eliminating genetic influences

There is increasing evidence that genotypic differences among individuals make them particularly sensitive to the environment. Caspi et al. (2003) were some of the first to show that carriers of the short allele in a gene responsible for serotonin levels were particularly sensitive to environmental stressors as related to depression. This work has been confirmed by some but has been questioned by others (Risch et al., 2009). Some believe that the reason for this weak replication history is that sensitivity alleles such as the 5HTT’S allele or the 7R allele in DRD4 will predict a response in general but not necessarily the same response. That is, identical twins who both carry two copies of the 7R allele and who are exposed to the same environments may be very sensitive to their environments for the same genetic reason but may respond very differently to the same environmental stimuli (Conley et al., 2013). The environmental source of the response may be random, but the probability of response has a foundation in genotype. Thus, genotype can still exert an important influence on discordance among genetically identical twins.

Equally important, there is increasing evidence that fixed genotype does not imply fixed genetic associations. The role of the epigenome and variation in gene expression may be the reason for incomplete gene penetrance observed in many studies. Raj et al. (2010) demonstrate this critical issue with a study of genetically identical worms that reside in identical environments who have marked differences in measurable phenotypes. They attribute this phenotypic variation to stochastic processes outside of the two “fixed” genetic and environmental components. They show epigenetic differences in response to otherwise random variation that may have important implications for health, longevity, and other measures of fitness. Thus, the attraction of the identical twin study design is that genotype is fixed and thus cannot be an explanation for observed differences, but it appears as though genetic processes may still underlie some observed phenotypic differences among genetically identical organisms.

3. The use of twins may exacerbate general concerns with traditional statistical inference

In addition to this issue of the validity of the research design to attain causal estimates, as discussed in Amin et al., within twin estimates also exacerbate two statistical issues: measurement error

---

1 Results, and those that follow, are available from the authors upon request.
it is unclear that the methods themselves can be assumed to approximate an experiment, and therefore rule in causal effects. Our view is that the residual error term in these models is unlikely to be uncorrelated with educational attainment.

References


