



Commentary

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

Jason D. Boardman^{a,*,1}, Jason M. Fletcher^{b,1}^a University of Colorado, Boulder, CO, United States^b University of Wisconsin, Madison, United States

ARTICLE INFO

Article history:

Available online 8 October 2014

Keywords:

Identical twins
Causal estimates
Health and education

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. (2015) 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 2015), 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., 2015). 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. (2015), 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 one more year of schooling than they would otherwise, follow these individuals (and a non-treated sample) over time, and ask whether individuals with experimentally-induced increases in education around age 20 had better health outcomes around age 50.² Turning to twin/sibling models, what assumptions do we need

² 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 (2015) for further discussion and an example of the approach.

* Corresponding author.

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

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 data³ (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 ($p < .0001$) higher mean level of education (mean = 14.53 years) compared to the concordant pairs (mean = 13.51 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. (2015) 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 ($r = -.03$, $p < .65$) but a significant association ($r = -.17$, $p < .03$) 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 genotypically 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 genotypically 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

³ Results, and those that follow, are available from the authors upon request.

(misreporting) in the variables and spillovers (interference) across observations. The first issue is dealt with in Amin et al. in the standard way (see Ashenfelter and Krueger, 1994) by using a twin's report of her co-twin's educational attainment as an instrumental variable to reduce any measurement error in her co-twin's measure. The results in Amin et al., as in most of the literature, suggest minor implications of measurement error for their results. However, absent in the Amin et al. paper is a discussion of interference (peer effects) between twins. One might worry that a twin's educational achievement and attainment may have direct causal effects on her co-twin's achievement, as is sometimes found in the peer effects literature on classmates (Rivkin et al., 2005; Fletcher, 2010) and has been found in the case of spillovers in health outcomes (ADHD) among siblings (Fletcher and Wolfe, 2008). In these cases, the implications of the interference will be driven by the direction of the influence; if twins seek to differentiate/rebel against one another in educational attainments, the twin-differenced results will be biased upward but if twins conform to each other, the results of within-twin difference empirical exercises will be biased downward. This issue also has implications of which twins are used in the data analysis—for example the sample may become comprised of families who promote non-conforming children for within MZ twin analysis.

There are additional sample issues that should be considered. The Amin et al. study leverages samples that may be unrepresentative in several ways. First, twins are a small subsample of the experiences and environments in the US. They also are likely treated differently by families and society than non-twins. This issue is raised in all studies of twins and discussed in Amin et al., but bears repeating. Second, while all datasets of twins may be unrepresentative, the three twin datasets used in the analysis might pose particular concerns—they are all white, mostly female samples of two states (Minnesota and Virginia) and of WWII military veteran brother-pairs. While the authors rightfully conjecture that their estimates may have advantages in estimating average treatment effects (for these specific samples), the external validity of the effects are unclear.

4. Summary

In summary, the Amin et al. results suggesting the lack of causal effects of schooling on later health outcomes are quite provocative and align with the results from many related papers in the literature using twin samples as well as alternative research designs, such as instrumental variables and regression discontinuity (Fletcher, 2015; Clark and Royer, 2010). Indeed, their results are largely in line with the conclusions from a study in *this issue* using very different methods among unrelated persons (Boardman et al., 2015). The purpose of this commentary is not to challenge the results or interpretations of the Amin et al. paper or other papers that have examined comparable questions with twins (Madsen et al., 2014; Lundborg, 2013). Rather, our primary aim is to remind the readers of the limitations of identical twin models for causal inference that have been detailed elsewhere (McGue et al., 2010) and to highlight additional considerations that are more specific to the health literature and to the emerging findings from molecular genetics. Despite these potential limitations, the findings from their study challenge us to think harder about theoretical models and empirical applications that suggest education and health would be strongly and causally linked. Like any single study, the analyses face data and conceptual limitations that do not allow the findings to be conclusive on the topic. Our own thinking on this topic and on these methods would guard against claims that twin studies are useful in producing causal effects—our own preference is that researchers expunge causal language in papers that use these methods. These analyses may indeed *rule out* the likelihood of causal processes, but

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

- Amin, Vikesh, Behrman, Jere R., Kohler, Hans-Peter, 2015. Schooling has smaller or insignificant effects on adult health in the US than suggested by cross-sectional associations: new estimates using relatively large samples of identical twins. *Soc. Sci. Med.* 127, 181–189.
- Ashenfelter, O., Krueger, A., 1994. Estimates of the economic return to schooling from a new sample of twins. *Am. Econ. Rev.* 84 (5), 1157–1173.
- Boardman, Jason D., 2009. State-level moderation of genetic tendencies to smoke. *Am. J. Public Health* 99 (3), 480–486.
- Boardman, Jason D., Blalock, Casey L., Pampel, Fred C., Hatemi, Peter K., Heath, Andrew C., Eaves, Lindon J., 2011. Population composition, public policy, and the genetics of smoking. *Demography* 48 (4), 1517–1533.
- Bound, J., Solon, G., 1999. Double trouble: on the value of twins-based estimation of the returns to schooling. *Econ. Educ. Rev.* 18 (2), 169–182.
- Boardman, Jason D., Domingue, Benjamin W., Daw, Jonathan, 2015. What can genes tell us about the relationship between education and health? *Soc. Sci. Med.* 127, 171–180.
- Brim, Orville G., Balthes, Paul B., Bumpass, Larry L., Cleary, Paul D., Featherman, David L., Hazzard, William R., Kessler, Ronald C., Lachman, Margie E., Markus, Hazel Rose, Marmot, Michael G., Rossi, Alice S., Ryff, Carol D., Shweder, Richard A., 1995–1996. National Survey of Midlife Development in the United States (MIDUS). ICPSR02760-v8. Inter-university Consortium for Political and Social Research [distributor], Ann Arbor, MI, 2011-10-25. <http://doi.org/10.3886/ICPSR02760.v8>.
- Burt, Callie H., Simons, Ronald L., 2014. Pulling back the curtain on heritability studies: biosocial criminology in the post genomic era. *Criminology* 52 (2), 223–262.
- Caspi, Avshalom, Sugden, Karen, Moffitt, Terrie E., Taylor, Alan, Craig, Ian W., Harrington, Hona Lee, McClay, Joseph, Mill, Jonathan, Martin, Judy, Braithwaite, Antony, Poulton, Richie, 2003. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT Gene. *Science* 301, 386–389.
- Clark, D., Royer, H., 2010. The Effect of Education on Adult Health and Mortality: Evidence From Britain (No. w16013). National Bureau of Economic Research.
- Conley, Dalton, Rauscher, Emily, Siegal, Mark L., 2013. Beyond orchids and dandelions: testing the 5HTT “risky” allele for evidence of phenotypic capacitance and frequency dependent selection. *Biodemogr. Soc. Biol.* 59 (1), 37–56. <http://dx.doi.org/10.1080/19485565.2013.774620>.
- Farrell, Phillip, Fuchs, Victor R., 1982. Schooling and health: the cigarette connection. *J. Health Econ.* 217–230.
- Fletcher, Jason, Wolfe, Barbara, 2008. Child mental health and human capital accumulation: the case of ADHD revisited. *J. Health Econ.* 27 (3), 794–800.
- Fletcher, Jason, 2010. Spillover effects of inclusion of classmates with emotional problems on test scores in early elementary school. *J. Policy Anal. Manag.* 29 (1), 69–83.
- Fletcher, J.M., 2013. Adolescent depression and labor market outcomes. *South. Econ. J.* 80 (1), 26–49.
- Fletcher, J.M., 2011. Long term effects of health investments and parental favoritism: the case of breastfeeding. *Health Econ.* 20 (11), 1349–1361.
- Fletcher, J.M., 2014a. The effects of childhood ADHD on adult labor market outcomes. *Health Econ.* 23 (2), 159–181.
- Fletcher, Jason M., 2014b. Friends or family? Revisiting the effects of high school popularity on adult earnings. *Appl. Econ.* 46 (20), 2408–2417.
- Fletcher, J.M., 2015. New evidence of the effects of education on health in the US: compulsory schooling laws revisited. *Soc. Sci. Med.* 127, 101–107.
- Gilman, Stephen E., Loucks, Eric B., 2014. Another casualty of sibling fixed-effects analysis of education and health: an informative null, or null information? *Soc. Sci. Med.* 118, 191–193.
- Lundborg, P., 2013. The health returns to education – what can we learn from twins? *J. Popul. Econ.* 26 (2), 673–701.
- Madsen, Mia, Andersen, Per K., Gerster, Mette, Andersen, Anne-Marie N., Christensen, Kaare, Osler, Merete, 2014. Are the educational differences in incidence of cardiovascular disease explained by underlying familial factors? A twin study. *Soc. Sci. Med.* 118, 182–190.
- McGue, Matt, Osler, Merete, Christensen, Kaare, 2010. Causal inference and observational research the utility of twins. *Perspect. Psychol. Sci.* 5 (5), 546–556.
- Raj, Arjun, Rivkin, Scott A., Andersen, Erik, van Oudenaarden, Alexander, 2010. Variability in gene expression underlies incomplete penetrance. *Nature* 463 (7283), 913–918.
- Risch, Neil, Herrell, Richard, Lehner, Thomas, Liang, Kung-Yee, Eaves, Lindon, Hoh, Josephine, Griem, Andrea, Kovacs, Maria, Ott, Jurg, Ries Merikangas, Kathleen, 2009. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* 301 (23), 2462–2471.
- Rivkin, Steven G., Hanushek, Eric A., Kain, John F., 2005. Teachers, schools, and academic achievement. *Econometrica* 73.2, 417–458.