The Relationship between Education and Mental Health: New Evidence from a Discordant Twin Study

Andrew Halpern-Manners, Landon Schnabel, and Elaine M. Hernandez, Indiana University
Judy L. Silberg and Lindon J. Eaves, Virginia Commonwealth University

Prior research has documented a strong and positive correlation between completed education and adults’ mental health. Researchers often describe this relationship using causal language: higher levels of education are thought to enhance people’s skills, afford important structural advantages, and empower better coping mechanisms, all of which lead to better mental health. An alternative explanation—the social selection hypothesis—suggests that schooling is a proxy for unobserved endowments and/or preexisting conditions that confound the relationship between the two variables. In this article, we seek to adjudicate between these hypotheses using a relatively large, US-based sample of identical adult twins. By relating within-twin-pair differences in education to within-twin-pair differences in mental health, we are able to control for the influence of genetic traits and shared family characteristics that may otherwise bias the estimates associated with educational attainment. Results from our analyses suggest that the observed association between education and mental health is attributable to confounding on unobserved variables. This finding holds across mental health conditions, is robust to several sensitivity checks, and survives a falsification test. Theoretical implications for the study of educational gradients in mental health are discussed.
served endowments (i.e., early onset mental health problems, hard-to-measure family characteristics, genes, and/or biologically based health conditions) that confound the relationship between education and mental health (Kawachi et al. 2013; Kessler et al. 2005; Miech et al. 1999).

In this study, we seek to adjudicate between these possibilities using a large, US-based sample of identical adult twins. By relating within-twin-pair differences in schooling to within-twin-pair differences in mental health, we are able to tease out the influence of genetic traits and shared family characteristics that may otherwise bias the estimates associated with educational attainment. Although this research design—typically referred to as a discordant twin (or co-twin control) study—has been used extensively in past research to examine the relationship between education, overall health, morbidity, and mortality (Amin, Behrman, and Kohler 2015a; Behrman et al. 2011; Kohler, Behrman, and Schnittker 2011; Lundborg 2013; Madsen et al. 2014; Madsen et al. 2011; Webbink, Martin, and Visscher 2010), applications within the mental health literature have been relatively rare (but see Fujiwara and Kawachi 2009).

We believe that this represents a missed opportunity. Carefully parsing the causal structure of the education–mental health relationship is an important prerequisite to understanding the mechanisms through which socioeconomic status and mental health affect each other (Warren 2009), and to addressing long-standing socioeconomic inequalities in psychological well-being (Lundborg 2013). As noted above, convincingly executing such analyses is difficult to accomplish without a strong analytic approach that explicitly controls for a wide array of potential confounding variables (Miech et al. 1999). Conditional on a well-defined set of identifying assumptions (Kohler, Behrman, and Schnittker 2011; McGue, Osler, and Christensen 2010), the within-twin-pair estimator that we employ here is able to resolve this issue, allowing us to obtain unbiased estimates of the causal effect of schooling on various mental health outcomes.

The remainder of this article is divided into four main sections. In the section that follows, we review theories and evidence related to education and mental health, with a particular focus on the causal structure of the hypothesized association. As a part of this discussion, we summarize past attempts to estimate the relative importance of the social selection and causation hypotheses, and give more details about the approach we use to distinguish between these two explanations. In the second and third sections, we describe our data and methods, and then present the results from our empirical analyses (including estimates from a series of supplementary models that are designed to establish the robustness of our estimates). Finally, we conclude by recapitulating our main findings and suggesting avenues of future research for scholars who are interested in studying the relationship between education and mental health.

**Background**

Researchers in sociology, epidemiology, psychology, and other fields have long observed an educational gradient in mental health (Dohrenwend et al. 1992; Fryers, Melzer, and Jenkins 2003; Hollingshead and Redlich 1958; Jürges, Reinhold, and Salm 2011; Kawachi, Adler, and Dow 2010; Kessler et al. 1995;
This finding holds across measures of mental health—including depression, anxiety disorders, happiness, psychopathologies, and stress-related conditions—and has proven to be true using different samples, estimation strategies, and data-collection techniques (Eikemo et al. 2008; Jokela et al. 2010; Jürges, Reinhold, and Salm 2011; Kawachi, Adler, and Dow 2010; Miech et al. 1999; Warren 2009; Yu and Williams 1999). Even studies that challenge some aspect of the relationship between education and mental health still suggest that there is an important association between the two (Adams et al. 2003).

Explanations for these findings generally fall into one of two broad categories: social causation or social selection. Scholars who argue that the relationship between education and mental health is causal assert that education affects health because more education generates additional social and economic resources, leading to fewer stressors, better coping strategies, and more autonomous lifestyles (Link, Lennon, and Dohrenwand 1993; Schieman and Plickert 2008; Søndergaard et al. 2012). Along the way, individuals with higher levels of education develop skills that enhance their ability to solve problems that may be detrimental to their mental health (Ross and Mirowsky 2013), while avoiding (or buffering themselves from) the negative life events, stress exposures, and other risk factors that are thought to precipitate the onset of mental health problems (Lorant et al. 2003). We depict the reduced-form version of this model graphically in the left-hand panel of figure 1, with a causal arrow flowing directly from education to mental health ($S \rightarrow H$).

Other researchers have suggested that the link between education and mental health may derive from “preexisting” differences between individuals (the selection hypothesis). Under this formulation, early life characteristics are seen as the primary drivers of the association between education and mental health—as indicated by the additional set of arrows included in figure 1b. In most cases, prior mental health conditions are identified as the main culprit (Breslau et al. 2008; Eaton et al. 2001; Kessler et al. 2005; Vander Stoep et al. 2003), leading to the early termination of schooling and worse mental health outcomes later in life (Hudson 2005). Although the mechanisms that drive these effects are themselves debated, most scholars agree that functional impairments and/or the stigma and social exclusion that go along with mental health problems are probably to blame (Gove 1982; Link et al. 1989; McLeod, Uemura, and Rohrman 2012; Needham, Crosnoe, and Muller 2004). We refer to these possibilities as “direct selection.”
In other versions of the selection hypothesis, hard-to-observe background characteristics—including contextual environments, genetic endowments, and other biosocial factors—are cited as common causes (Boardman, Domingue, and Daw 2015; Kawachi et al. 2013), confounding the relationship between education and mental health, and ultimately producing spurious effects. Recent meta-analyses suggest that as much as 40 percent of the variance in educational attainment is attributable to genetic differences in the population (Branigan, McCallum, and Freese 2013), and that similar levels of heritability exist for specific mental health conditions (Sullivan, Neale, and Kendler 2000). If both phenotypes are influenced by a common genetic source and/or the same environmental or familial characteristics, we would expect to see an association between the two even in the absence of a causal relationship. Because these processes are distinct from those described above, we refer to them as “indirect selection.”

Although these basic theoretical frameworks (i.e., causation and selection) are not entirely incompatible, long-standing debates exist about their relative importance (see, e.g., Dohrenwend and Dohrenwend 1969; Turner and Gartrell 1978; Wheaton 1978). Early work suggested the presence of strong educational effects on mental health (Kessler et al. 1995), but also revealed important linkages between early onset conditions and completed schooling (Vander Stoep et al. 2003). Subsequent studies reached similar conclusions, often indicating that education shapes mental health through a variety of proximal life experiences and resources, but sometimes demonstrating the reverse (De Ridder et al. 2013; Miech et al. 1999; Ritsher et al. 2001). Some scholars have suggested that these disagreements could stem from variation in the nature of the education–mental health relationship across different measures of mental health (Dohrenwend et al. 1992; Johnson et al. 1999), but even this hypothesis has been difficult to confirm (Lorant et al. 2003).

Part of the problem could be technical. The types of models estimated in prior work have ranged from simple linear regressions using cross-sectional data (see, e.g., Williams et al. 1997), to more complex longitudinal models that allow for direct and indirect effects stemming from early life mental health conditions (i.e., $X \rightarrow S \rightarrow H$ and $X \rightarrow H$ in figure 1b) (see, e.g., Miech et al. 1999). These studies usually include an extensive set of background variables that are thought to covary with education and mental health (Miech and Shanahan 2000), but rarely address concerns about confounding due to heritable traits and genetic endowments (but see Boardman, Domingue, and Daw 2015; Fujiwara and Kawachi 2009), and even less frequently grapple with issues related to measurement error (but see Warren 2009). This has made it difficult to recover unbiased estimates of the true causal relationship between education and mental health ($S \rightarrow H$).

Resolving this problem—while also allowing for selection based on preexisting differences—requires a strong analytic approach. In recent years, researchers who study education and physical health have turned to family-based designs as a potential solution (Fletcher and Frisvold 2009; Krieger et al. 2005; Lundborg 2013; Madsen et al. 2014). The idea is to compare the health outcomes of siblings who are discordant with respect to education, but who were reared in the same social environment and who share all or part of their genetic material. Monozy-
gotic (MZ) twins, who are genetically identical, are of special interest in this research since any within-pair association between education and health cannot be due to genes (Madsen et al. 2011). Studies using this approach have generally found that the association between education and health is either non-existent or substantially attenuated (but see Lundborg 2013).

Despite the widespread use of genetically informed designs in the literature on education and physical health, applications to questions concerning mental health have been much more limited. One of the few exceptions is a recent study by Fujiwara and Kawachi (2009), who found no relationship between years of schooling and depression after differencing out shared genetic and environmental factors using a within-MZ approach. Although intriguing, researchers have questioned whether these results represent the true absence of an effect, or whether they are an artifact of inadequate statistical power and/or an under-specified model (Boardman and Fletcher 2015; Madsen and Osler 2009). Fujiwara and Kawachi (2009) analyzed data from the National Midlife in the United States (MIDUS) study, which includes a limited number of discordant MZ twin pairs \( n = 113 \) and provides no information about early life mental or physical health problems.¹

We share these questions. The utility of the within-twin design—like other estimation strategies—depends on the availability of high-quality data and the validity of various modeling assumptions (as we discuss in more detail in the sections that follow). In this paper, we draw on a relatively large sample of MZ twins that includes information on early life health conditions, and that contains multiple measures of each twin’s completed schooling. These data features allow us to (1) specify a more complete model of the causal pathways that link education to mental health; (2) provide a reasonably well-powered test of the social causation and social selection hypotheses; and (3) correct for issues related to measurement error. In the next section of the paper, we describe our data in more detail and provide further information about the modeling strategy we use to identify the causal relationship between education and mental health.

**Data**

Our analysis makes use of the Virginia 30K (VA30K), a large-scale study of twins (including MZ pairs) ascertained from the population-based Virginia Twin Registry, and supplemented by a nationwide sample of twins recruited through the American Association of Retired Persons (AARP) (for more details on sample ascertainment, see Eaves et al. 1999; Truett et al. 1994). In 1987, a lengthy “Health and Lifestyle” questionnaire was sent to members of the sample and their immediate family members (spouses, siblings, children, and parents); the survey included questions about education, marital history, fertility, labor-market experiences, stressful life events, personality, social attitudes, social support, leisure activities, and mental and physical health. A total of 14,763 twins completed the questionnaire, for a response rate of 70 percent; the response rate among twins’ family members was substantially less.²

Our analytic sample includes complete, MZ twin pairs with non-missing information on their educational attainment and the educational attainment of their
co-twin \( n = 4,038 \) twins or 2,019 pairs). Twins’ zygosity was determined by responses to two questions about perceived physical similarities. For a twin to be counted as MZ, both twins in the pair had to respond that they were “frequently” mistaken for their co-twin while growing up and that they have “such a strong resemblance to each other in stature, coloring, facial features, etc., that people often mistake one for the other, or say they are ‘as alike as two peas in a pod.’” If responses to these items were inconsistent within twin pairs, we coded the twins to an “unknown” category and excluded them from our analysis \( n = 724 \).

The validity of this approach is well established within the twin literature and has been shown to be 95 percent accurate when validated against blood-typing (Eaves, Eysenck, and Martin 1989).

In appendix table A1, we compare the sociodemographic characteristics of our sample to parallel estimates obtained for non-twin siblings, and to contemporaneously collected data from the Current Population Survey (CPS). In general, we find that the twins in our sample are slightly older than members of the CPS, somewhat more educated, disproportionately female, and substantially less diverse. In an effort to evaluate what these differences mean for our analyses, we generated post-stratification weights via an iterative proportional fitting algorithm, and then used these weights to exactly equate our sample to population-control totals obtained from the CPS. The variables that we used for our benchmarks included a subset of those mentioned above: age and gender. In supplementary models, we found that the use of these weights had no effect on the estimated relationship between education and mental health.3

**Measures of Mental Health (H)**

Responses to a self-reported symptom inventory were obtained from each twin using an empirically derived 30-item subset of the 90-item Symptom Checklist (SCL-90) (Derogatis, Lipman, and Covi 1973). In carrying out this task, respondents were asked to consider a variety of mental health problems and then indicate the extent to which that problem had bothered them, or caused discomfort, within the past 30 days. The items included in the symptom inventory were derived from five SCL-90 subscales: depression (10 items), anxiety (7 items), somatization (5 items), phobic anxiety (5 items), and insomnia (3 items). In each case, the response options ranged from (1) not at all to (5) extremely.4 To convert respondents’ answers into distinct measures, we performed a principal component analysis.5 Our preferred solution extracted three readily interpretable factors, which we refer to as depression, panic-phobia, and somatization.6 See figure 2 for a graphical depiction.7

Substantial within-pair differences existed among twins with respect to our three measures of mental health. As shown in the lower panel of table 1, the mean absolute difference within twin pairs was 1.87 for depression \( (SD = 1.95) \), 1.67 for panic-phobia \( (SD = 1.72) \), and 1.28 for somatization \( (SD = 1.92) \). The corresponding intra-pair correlations for these variables were 0.35, 0.38, and 0.34, respectively. The fact that within-twin-pair variation was so great means that the usual reductions to sample size that occur in sibling- and twin-based studies
Figure 2. Distributions for the depression, panic-phobia, and somatization scales ($n = 4,038$). The scales were constructed using a 30-item subset of the 90-item Symptom Checklist (SCL-90) (Derogatis, Lipman, and Covi 1973), and principal component analysis. The full set of loadings—and a short description of each item—can be found in the online appendix. See text for more details.

Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>MZ twins ($n = 4,038$)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean/%</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of schooling</td>
<td>13.40 (2.17)</td>
<td></td>
</tr>
<tr>
<td>Absolute within-pair difference</td>
<td>0.71 (1.20)</td>
<td></td>
</tr>
<tr>
<td>Percent with no difference</td>
<td>70.08</td>
<td></td>
</tr>
<tr>
<td>Intra-pair correlation</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td><strong>Demographic characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>53.61 (17.53)</td>
<td></td>
</tr>
<tr>
<td>Female (%)</td>
<td>70.85</td>
<td></td>
</tr>
<tr>
<td><strong>Mental health scales</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>0.00 (2.38)</td>
<td></td>
</tr>
<tr>
<td>Absolute within-pair difference</td>
<td>1.87 (1.95)</td>
<td></td>
</tr>
<tr>
<td>Intra-pair correlation</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Panic-phobia</td>
<td>0.00 (2.16)</td>
<td></td>
</tr>
<tr>
<td>Absolute within-pair difference</td>
<td>1.67 (1.72)</td>
<td></td>
</tr>
<tr>
<td>Intra-pair correlation</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>Somatization</td>
<td>0.00 (2.00)</td>
<td></td>
</tr>
<tr>
<td>Absolute within-pair difference</td>
<td>1.28 (1.92)</td>
<td></td>
</tr>
<tr>
<td>Intra-pair correlation</td>
<td>0.34</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Educational attainment was coded using respondents’ answers to questions about years of schooling completed, where the response options were 0–7 years of elementary school, 8 years of elementary school, 1–3 years of high school, 4 years of high school, 1–3 years of college, and 4 or more years of college. Mental health scales were obtained via principal component analysis of the 30-item SCL subscale. See text for more details.

(due to a lack of variation on either the independent or dependent variable) are much less severe in our analyses (Boardman and Fletcher 2015; Gilman and Loucks 2014; Kaufman and Glymour 2011; McGue, Osler, and Christensen
This has important implications for our ability to precisely estimate the causal effects associated with progressively higher levels of educational attainment.

**Measures of Educational Attainment (S)**

To characterize each twin’s schooling, we created a continuous variable using respondents’ answers to questions about years of schooling completed (0–7 years of elementary school = 5; 8 years of elementary school = 8; 1–3 years of high school = 10; 4 years of high school = 12; 1–3 years of college = 14; 4 or more years of college = 16); the same strategy was used by Amin, Behrman, and Kohler (2015a) and Boardman et al. (2011) in recent analyses of the VA30K. Basic descriptive statistics for this variable are provided at the top of table 1, along with information about the average size (expressed in absolute terms) of within-twin-pair differences and intra-pair correlations. In supplementary analyses (described below), we employ an alternative, nonlinear version of our educational attainment variable where different levels of education were coded using a series of dichotomous indicators. Descriptive statistics for this variable are available upon request.

**Other “Preexisting” Differences (X)**

Identification of schooling effects ($S \rightarrow H$) requires that we control for variables that vary within twin pairs and that could conceivably influence twins’ educational attainment and mental health. For our purposes, the primary concern is that early life mental health conditions could generate within-twin-pair differences in completed education and within-twin-pair differences in respondents’ post-schooling mental health. To provide a control for this possibility, we created a binary indicator of early life health problems (mental or otherwise) using a lengthy series of health history items (1 = the twin reported experiencing health problems prior to the age of 18; 0 = otherwise). These items included questions about depression and alcoholism, as well as questions that dealt with heart health, asthma, bronchitis, blood pressure, migraines, emphysema, cirrhosis, strokes, and diabetes. In all cases, twins were asked to report the age when the problem was first diagnosed.

**Methodology**

Our goal in this project is to evaluate the independent effect of education on mental health (i.e., $S \rightarrow H$). To achieve this objective, we use a standard fixed effects specification for within-twin pair estimation (Kohler, Behrman, and Schnittker 2011):

$$H_{ij} = \alpha + \beta S_{ij} + \gamma X_{ij} + C_{j} + G_{i} + \varepsilon_{ij},$$

(1)

where $H_{ij}$ is the interval-scaled mental health outcome for individual $i$ ($i = 1, 2$) in twinship $j$ ($j = 1, 2, \ldots, N$); $S_{ij}$ is a similarly annotated measure of completed
schooling; $X_{ij}$ is a binary indicator of early life health problems that could potentially influence both $S$ and $H$; $C_j$ is a measure of unobserved contextual characteristics (e.g., a person’s parental environment, peer group, or neighborhood); $G_j$ is a measure of unobserved genetic endowments; and $\epsilon_{ij}$ is a random individual-level error term that is assumed to be uncorrelated with the other explanatory variables in the model.

In a within-MZ fixed effects estimator, the unobserved components in equation (1) are “controlled away” by taking differences within twin pairs:

$$H_{1j} - H_{2j} = \Delta H_j = \beta \Delta S_j + \gamma \Delta X_j + \Delta \epsilon_j,$$

where the $\Delta$s represent differences between variables for the $j$th pair (i.e., $S_{1j} - S_{2j}$). This approach eliminates the effects of unobserved contextual characteristics ($C_j$)—since the vast majority of twins experience the same family, school, and neighborhood environments while growing up. It also eliminates the effects of unobserved genetic endowments ($G_j$)—since MZ twins are born with identical genetic information. Under this formulation, any remaining differences that exist between twins are assumed to be orthogonal to schooling, conditional on the included covariates ($X$), allowing for unbiased estimates of the true (reduced-form) impact of education on mental health ($\beta$).

In our analysis, we fit the relation shown in equation (2) separately for each of our three measures of mental health, and then compare the resulting estimates to baseline estimates obtained using an unpaired, cross-sectional approach (i.e., ordinary least squares [OLS] regressions with controls for relevant background characteristics). Both estimators will produce consistent and unbiased estimates of the “true” education effect, $\beta$, if all confounds have been taken into account. We argue that this scenario is more tenable in the fixed effects specification because all unobserved endowments common within pairs are differenced out. To ensure that our conclusions with respect to $\beta$ are robust to the presence of measurement error (Griliches 1979), we also fit a series of auxiliary models where each twin’s schooling is instrumented using reports provided by their co-twin (Ashenfelter and Krueger 1994; Kohler, Behrman, and Schnittker 2011). We say more about this approach—and several additional sensitivity analyses—in the Results section below.

Results

Table 2 presents parameter estimates from our unpaired, cross-sectional regressions ($n = 4,038$), with estimates arranged into columns according to model specification and the mental health measure under consideration. To provide a sense for the importance of confounding on observable characteristics (including basic sociodemographic attributes and preexisting mental and/or physical health conditions), we began with a baseline specification that included our continuous measure of education (models 1, 3, and 5); we then entered controls for age and age-squared, a dummy variable reflecting the respondent’s gender ($0 = \text{female}; 1 = \text{male}$), and an indicator for the presence of early life health
issues (1 = yes; 0 = no). In both sets of models, we estimated cluster-robust standard errors in order to account for the interdependence of observations within twin pairs. Estimates that are twice their standard error have been bolded to denote significance.

The results from these models are easy to summarize. Consistent with prior research, we found a negative and statistically significant ($p < .05$) relationship between higher levels of education and poor mental health (columns 1, 3, and 5). This basic pattern held for each mental health condition, was generally true for males and females (see the separate by-gender estimates reported in appendix table A3), and persisted after entering controls for the presence of “preexisting” mental or physical health conditions (which are also highly significant). The effect sizes implied by these models are modest: according to our estimates, a six-year increase in schooling (i.e., the difference between a high school dropout and a college completer) is associated with approximately one-fifth of a standard deviation decrease on the depression scale (–0.40), and one-third of a standard deviation decrease on both the panic-phobia (–0.58) and somatization scales (–0.58).

Do similar findings hold once unobserved genetic and environmental factors are taken into account? In table 3, we present parallel estimates from our within-MZ fixed effects models, with the coefficients again arranged into columns according to model specification and mental health condition. Results from these models suggest that the answer is “no.” The education coefficients are all highly attenuated (e.g., $–0.005/–0.070 = 0.07 \times 100 \text{ percent} = 7 \text{ percent}$ of the estimate we obtained in our fully specified OLS model predicting depression), never significant ($p > 0.10$), and in some cases positively signed—regardless of how we specify the models. These findings run counter to the social causation hypothesis: rather than producing a robust causal relationship, the more stringent within-MZ estimates suggest that the effect of education on mental health is essentially zero. This is consistent with the conceptual diagram shown on the right-hand side of figure 1.

### Table 2. Estimated Effects from Unpaired, Cross-Sectional Regression models ($n = 4,038$)

<table>
<thead>
<tr>
<th></th>
<th>Depression (1)</th>
<th>Depression (2)</th>
<th>Panic-phobia (3)</th>
<th>Panic-phobia (4)</th>
<th>Somatization (5)</th>
<th>Somatization (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of schooling</td>
<td>–0.074 (0.019)</td>
<td>–0.070 (0.020)</td>
<td>–0.124 (0.017)</td>
<td>–0.099 (0.017)</td>
<td>–0.098 (0.016)</td>
<td>–0.087 (0.017)</td>
</tr>
<tr>
<td>Female</td>
<td>0.777 (0.085)</td>
<td>0.612 (0.079)</td>
<td>0.434 (0.072)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>–0.046 (0.017)</td>
<td>–0.029 (0.015)</td>
<td>0.000 (0.015)</td>
<td>0.000 (0.015)</td>
<td>0.000 (0.015)</td>
<td>0.000 (0.015)</td>
</tr>
<tr>
<td>Age-squared</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early life health problem</td>
<td>0.559 (0.170)</td>
<td>0.517 (0.128)</td>
<td>0.269 (0.133)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors, which are shown in parentheses, are clustered by twin pairs. Bolded coefficients are twice their standard error. See text for more details.
The estimates presented in table 3 also present slight complications for the conventional “direct selection” version of the selection hypothesis. After differencing out shared environmental and genetic factors using within-MZ fixed effects, the coefficients associated with early life health problems diminish in magnitude and fail to reach significance by conventional standards ($p > 0.10$). This suggests that the linkage between early life health problems and later-life mental health may originate from a common source: genetic or environmental endowments that precede the completion of respondents’ education and that are present prior to the onset of childhood health conditions. This interpretation would be most consistent with a conceptual model characterized by “indirect selection,” where the relationship between education and mental health is primarily the result of early life genetic and/or social conditions that give rise to both outcomes.

### Within-MZ Estimates for Different Birth Cohorts

The results reported above pertain to twins between the ages of 24 and 93 and, thus, make no distinction between pairs who were born and raised in different time periods. To allow for the possibility that historical factors (e.g., the prevailing educational expectations of the time) help determine the relative importance of unobserved confounds (Heath et al. 1993; Rosenquist et al. 2015), we refit our within-MZ models after partitioning our sample into three regularly spaced birth cohorts. Results from these analyses are given in table 4, with cohorts arranged into rows beginning with those individuals who were born prior to 1917. In general, we find little evidence to suggest that the importance of selection processes varies systematically depending on the historical era: five of the estimated coefficients have negative signs, four have positive signs, none are significant, and a test of the null hypothesis that they are all jointly equal to zero could not be rejected at the $p < .10$ level.

### Sensitivity Analyses

Before drawing further conclusions regarding the merits of the causation hypothesis, it is necessary to first consider a series of plausible alternative explanations. In this section, we summarize the results from several sensitivity analyses in an effort to establish the robustness (and meaning) of the estimates reported above. We begin with a discussion of measurement error.

---

**Table 3. Estimated Effects from Within-MZ Fixed Effects Models ($n = 4,038$)**

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Panic-phobia</th>
<th>Somatization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of schooling</td>
<td>$-0.005$</td>
<td>$-0.005$</td>
<td>$-0.021$</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.043)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Early life health problem</td>
<td>0.029</td>
<td>0.131</td>
<td>$-0.042$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Bolded coefficients are twice their standard error. See text for more details.
Alternative Explanation #1: An Informative Null or Measurement Error?

One of the main critiques of within-MZ models—and fixed effects models more generally—is that they exacerbate problems related to random measurement error (Griliches 1979). In within-MZ analyses, measurement error attenuates parameter estimates (e.g., $S \rightarrow H$) more than in unpaired analyses due to a relatively smaller signal-to-noise ratio (Kohler, Behrman, and Schnittker 2011). This, in turn, leads to an increased likelihood of making a Type II error. Following convention, we sought to mitigate this problem by fitting augmented models where education was instrumented using twin 1’s report about twin 2’s schooling, and vice versa (Ashenfelter and Krueger 1994; Behrman, Rosenzweig, and Taubman 1994; Kohler, Behrman, and Schnittker 2011). Having a co-twin’s report is useful for our purposes because it is likely to be correlated with their twin’s self-report, but uncorrelated with any measurement error contained therein. This allows us to correct for potential attenuation bias.
The results from our augmented models are provided in table 5. To facilitate interpretation, we have arranged the estimates according to the dependent variable (i.e., the mental health condition) and estimation strategy (i.e., standard OLS regression versus within-MZ fixed effects models), and omitted the additional covariates included in our full specification. For the sake of comparison, we have also included the corresponding OLS and within-MZ estimates from tables 2 and 3, respectively. In every case, the measurement-error-corrected coefficients are larger in magnitude than the uncorrected estimates that we presented earlier, suggesting the presence of attenuation bias in our previous analyses. Nowhere in the table, however, do we observe significantly negative coefficients for the within-MZ models—as one would have expected based on the social causation hypothesis. These findings suggest that the patterns we observed in the previous tables are real, and not a function of measurement error on the key independent variable.

**Alternative Explanation #2: Nonlinear Effects?**

To this point, we have assumed a continuous (linear) relationship between years of schooling and adult mental health. This approach—although commonly used in regression analyses examining the relationship between education and mental health conditions—ignores the possibility that educational credentials, and not years of schooling, are what matter most for adults’ psychological well-being.19

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Panic-phobia</th>
<th>Somatization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Within-MZ</td>
<td>OLS</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school (reference group)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>High school</td>
<td>–0.442</td>
<td>–0.222</td>
<td>–0.465</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
<td>(0.234)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>Some college</td>
<td>–0.589</td>
<td>–0.145</td>
<td>–0.549</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(0.263)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>College degree or higher</td>
<td>–0.578</td>
<td>–0.075</td>
<td>–0.757</td>
</tr>
<tr>
<td></td>
<td>(0.159)</td>
<td>(0.299)</td>
<td>(0.140)</td>
</tr>
<tr>
<td>F-test that all coefficients are equal to 0</td>
<td>5.069</td>
<td>0.406</td>
<td>10.732</td>
</tr>
<tr>
<td>p-value</td>
<td>0.002</td>
<td>0.749</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Note:** OLS models include controls for age, age-squared, gender, and the presence of an early life health condition. Standard errors are given in parentheses below the associated parameter estimate; bolded estimates indicate significance at the 0.05 level. The F-statistics reported at the bottom of the table are from tests of the null hypothesis that all of the educational coefficients are equal to 0. See text for more details.
To examine this possibility, we recoded our measure of education so that respondents were grouped into one of four categories: (1) less than high school; (2) high school degree; (3) some college; and (4) a college degree or higher. We then refit our cross-sectional and within-MZ models using less than high school as the reference category. The results are given in table 6. At the bottom of the table, we provide F-statistics from tests of the null hypothesis that all of the schooling coefficients are equal to zero. We were unable to reject this hypothesis in any of our within-MZ models, despite finding significant and substantively large effects using an unpaired approach (the three columns labeled “OLS”). These findings are consistent with our earlier models where we specified a linear functional form.

Alternative Explanation #3: Residual Confounding by Non-Shared Factors?

A third concern with our within-MZ estimates is that they may be subject to non-trivial amounts of endogeneity bias (Bound and Solon 1999). Although twins share the same genes and (generally speaking) the same social environment, that does not mean that they are identical in all possible respects. There could still be residual confounding by non-shared (and non-health-related) factors. This would lead to biased estimates of the true causal effect, where the direction of the bias depends on the nature of the relationship between $S$, $H$, and $\epsilon$ in equation (2). If $\epsilon$ affects $S$ and $H$ in the same direction, estimates of $\beta$ would be upwardly biased, leading to an overemphasis of causal processes in the education–mental health relationship. If $\epsilon$ affects $S$ and $H$ in the opposite direction, then the reverse would be true, leading to underestimates of the true causal effect. Although we do not have the means to address these possibilities conclusively, we can gain some leverage on the question by excluding twin pairs whose experiences growing up were highly dissimilar (making them more susceptible to residual confounding).

In table 7, we provide results from our within-MZ models after excluding twin pairs where at least one twin reported that they never shared the same room while they were children, never had the same playmates, were never dressed alike by their parents, and/or were never in the same classes at school. Corrected models adjust for measurement error; uncorrected models do not. In both cases, standard errors are given in parentheses below the associated parameter estimate, with bolded estimates indicating significance at the $p < 0.05$ level. See text for more details.
dressed alike by their parents, and/or were never in the same classes at school \((n = 3,660)\). The findings from these models remain essentially unchanged. Regardless of whether we make corrections for measurement error, and regardless of which psychological condition we use as the dependent variable, we still see a statistically non-significant relationship between education and mental health in our within-twin-pair models \((p > 0.10\) for all three outcomes). We take this to mean that our inferences regarding the relationship between education and mental health are robust to slight, but theoretically relevant, differences in the social conditions that twins were exposed to prior to the completion of their academic career. While this does not rule out the possibility of residual confounding, it does provide a useful check.

**Alternative Explanation #4: Insufficient Statistical Power?**

Our last set of analyses is designed to address questions related to statistical power. As we mentioned earlier, one of the concerns that is frequently raised when twin studies fail to produce significant results is that they lack the precision required to correctly reject the null hypothesis. Although our sample size is larger than most of the samples used in prior US-based co-twin analyses, we take this concern very seriously. To assess whether our conclusions regarding causation are being driven by inadequate power, we considered an alternative outcome where the evidence in favor of a causal effect is more firmly established: fertility. Prior work suggests that individuals with more schooling tend to have fewer children (Amin and Behrman 2014), are more likely to be childless (Kravdal and Rindfuss 2008), and often delay childrearing until older ages (Black, Devereux, and Salvanes 2008). If we fail to detect similar patterns in our data, it would raise questions about our ability to identify true educational effects in the preceding analyses.21

To construct our dependent variable, we summed respondents’ answers to a pair of questions about the number of sons and daughters they had (respondents were told to count all children, including those not living at home); we then sub-

<table>
<thead>
<tr>
<th>Table 8. Models Predicting Number of Children ((n = 4,038))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within-MZ</strong></td>
</tr>
<tr>
<td><strong>Uncorrected</strong></td>
</tr>
<tr>
<td>Years of schooling</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age-squared</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Note:** In the corrected models, years of schooling was instrumented using twin 1’s report about twin 2’s schooling, and vice versa. Bolded coefficients are twice their standard error. See text for more details.
stituted this measure into equation (2) and fit the resulting within-MZ model, with and without corrections for measurement error. To provide a baseline, we also estimated the corresponding cross-sectional models via OLS, using age, age-squared, and gender as covariates. Results from these models are given in table 8. In agreement with prior research, we find a significant and negative relationship between respondents’ educational attainment and fertility. This finding holds regardless of estimation strategy (i.e., the OLS and within-MZ models produce broadly similar results), and is insensitive to whether or not we make adjustments for measurement error. This is reassuring for our purposes, as it suggests that our within-MZ analyses are sufficiently powered to detect modestly sized education effects.

Discussion

Previous scholarship has described two primary mechanisms that link educational outcomes to mental health: social causation and social selection. In the current study, we subjected these mechanisms to direct empirical testing by exploiting differences in educational attainment within pairs of identical (MZ) twins. Results from these analyses suggest that selection mechanisms dominate the education–mental health relationship, leaving little room for a strong causal interpretation. These findings were consistent across an assortment of mental health conditions (including measures of depression, panic-phobia, and somatization), were robust to several sensitivity checks, and withstood a falsification test that was designed to reveal inherent problems with our data. If there are social factors that systematically drive differences in mental health within pairs of MZ twins, our findings suggest that education is not among them.

What are the broader theoretical implications of these results? Most prior research examining the association between education and mental health has debated whether lower levels of education lead to poorer mental health (social causation), or whether preexisting mental health conditions inhibit individuals’ educational attainment (direct selection). Recent studies using biosocial data suggest a third possibility: that the co-occurrence of low education and poor mental health stems, in part, from genetic factors, heritable traits, and other hard-to-measure early life characteristics that have typically not been included in prior work (indirect selection). Our findings would seem to affirm this latter perspective. Once shared environmental and genetic endowments were taken into account, the empirical association between education and mental health disappeared, calling into question usual attributions about the relationship between the two phenotypes.

Sociologists who are interested in the social determinants of mental health should take this possibility seriously (Freese 2008). If genetic endowments and early life environments are common causes that, alone or in some combination, drive the association between people’s early adult attainments (i.e., schooling) and subsequent well-being (i.e., mental health), then they should not be ignored in our theoretical or empirical work. Doing so risks overstating the true effect of education on mental health, while downplaying the complex underlying processes that give rise to the observed relationship. Although our results do not
preclude the possibility that education—and the various material, social, and personal advantages that it is thought to confer—has a place in the basic causal configuration, they do redirect attention to logically prior variables as the key etiologic factors. We believe that this represents an important theoretical contribution.

The results described above come with some obvious caveats. All statistical estimates must be considered alongside their limitations and judged, in part, by their consistency with other studies using different data sources and alternative estimation techniques. The findings from our analyses are no different. Because participants in our study were not randomly selected, the estimates that we presented are necessarily sample specific, making generalizations to the broader (non-twin) population tentative at best. Although the mechanical adjustments that we applied (i.e., post-stratification weights equating our sample to control totals from the CPS) suggest that this may not be an especially important issue, we cannot know for sure that these adjustments captured all of the relevant differences. Nor can we be certain the same relationships we observed for past birth cohorts of Americans will persist for subsequent generations of adults.

While some may deem this a limitation, we see it as an important opportunity. In order to move forward with this line of work, what we need now are new data—with bigger, more recent, and more demographically representative samples—so that we can generate even stronger statements concerning the linkages between education and mental health in the population at large. Such research need not draw exclusively (or even primarily) on a within-twin-pair design; identification via instrumental variables, natural experiments, and/or some other form of exogenous variation would also be informative. Comparisons between these follow-up studies and our findings would allow us to produce a more general (and methodologically robust) assessment of the education–mental health relationship, and its stability across subpopulations and time. We are optimistic that our study will serve as a useful catalyst in this regard.

Future research should also attempt to further develop the underlying “structural” model that gives rise to the confounding we observed in our analyses. Our primary goal in this project was to evaluate merits of the causation and selection hypotheses (i.e., the “reduced form” models presented in figure 1). In doing so, we left unspecified the relative importance of genetic and environmental factors and did not examine the multiple ways in which their interplay might shape people’s educational and mental health outcomes.25 Pure genetic or environmental confounding, environmental moderation of genetic expression (gene–environment interaction), and selection into environments based on genotype (evocative or active gene–environment correlations) are all potential mechanisms through which the education–mental health association could arise. Each of these possibilities warrants further attention from researchers in the social and health sciences.

We are also hopeful that researchers will extend our analyses to include additional mental health conditions. Although we were able to obtain valid and reliable measures of depression, panic-phobia, and somatization for members of our sample, we were unable to assess other common illnesses that occur in the
population (e.g., attention deficit disorder, obsessive-compulsive disorder, and antisocial behavior). Nor were we able to evaluate the effects of education on more extreme, but less common, pathologies like schizophrenia, psychosis, and bipolar mood disorders. While our results were consistent across the range of mental health conditions that we examined, it remains to be seen whether different (i.e., causal) relationships exist for other forms of mental illness (Dohrenwend et al. 1992; Miech et al. 1999). Future researchers stand to benefit by exploring this issue further using an identification strategy similar to the one we employed here.

We believe that a careful and continued consideration of the education–mental health relationship is crucially important. Educational gradients in mental health have proven to be remarkably robust and persistent over time (Link and Phelan 2006; Murphy et al. 2000; Yu and Williams 1999). A detailed understanding of how these gradients emerge—and whether causal or selection mechanisms are primarily to blame—is necessary before sociologists, psychologists, and others can begin to think about effective strategies for remediation. Our analyses indicate that the main etiologic forces can be located early in the life course, before individuals reach adulthood, and well before they complete their formal education. These findings suggest the need for a refined conceptual model that emphasizes the importance of environmental and genetic endowments as key upstream variables in the production of adult mental health disparities.

Supplementary Material

Supplementary material is available at Social Forces online, http://sf.oxfordjournals.org/.

Notes

1. It is also possible that their point estimates were attenuated due to measurement error on the key independent variable (years of schooling).
2. The reasonably high response rate suggests that first-order selection effects—for example one twin dying prior to the survey due to issues related to their mental health—are unlikely to change our results. For more discussion of selection in twin studies, see Amin et al. (2015b).
3. Post-stratification weights correct for known discrepancies between a sample and a target population, but they do nothing to adjust for unknown or unobserved differences. This is an issue we will return to in the Discussion section.
4. The 30-item subset included standard symptoms such as “feeling blue,” “feeling hopeless about the future,” “nervousness or shakiness inside,” “feeling weak in parts of your body,” “feeling everything is an effort,” and “loss of sexual interest or pleasure.” See appendix table A2 for more information.
5. The Kaiser-Meyer-Oklin measure of sampling adequacy was 0.95, suggesting that a low-dimensional representation of the symptom inventory is quite reasonable for these data. To confirm the appropriate number of components, we (1) inspected a scree plot and (2) ran a parallel analysis using the fapara command in Stata. Both sets of diagnostics agreed that our preferred solution provided the best fit. Finally, to provide further validation, we compared respondents’ depression scores with self-
reports of whether or not they had ever been diagnosed or treated by a physician for depression. Those who answered the self-reported item in the affirmative had a mean of 2.40 on our measure of depression; those who said they had never been diagnosed or treated had a mean score of –0.23. This amounts to just over a standard deviation difference and is significant at better than the $p < .001$ level.

6. Two additional points worth noting. First, we identified but did not analyze a fourth component, insomnia. Second, the panic-phobia component included panic-like symptoms (e.g., terror spells, suddenly scared) and items that reflect phobic behavior (e.g., avoiding frightening things, feeling afraid in open spaces). Out of the six items that loaded most highly on this component, four derived from the SCL phobic-anxiety subscale and two from the anxiety subscale (see appendix table A2 for the full set of loadings). Similar results have been obtained in other analyses using these data (see, e.g., Kender et al. 1995).

7. All three measures of mental health are positively skewed, as shown in figure 2. In auxiliary analyses, we took a log transformation after adding a small constant to each variable (to account for negative values) and then reran the key models in our analysis. The results were substantively identical.

8. Models where we distinguished between early life mental and physical health conditions produced equivalent results.

9. We implemented our identification strategy in Stata using `xtreg`, which can be used to estimate a within-twin pair fixed effects model like the one given in equation (1). We use the language of differences above because it offers a convenient way to think about what our estimation strategy can and cannot accomplish with respect to inference. For more on this class of models, see Allison (2009).

10. Of the twins in our sample, 99 percent reported growing up together, 91 percent said they were in the same classes in school, and 89 percent said they usually or always had the same playmates.

11. Epigenetic modification can lead to genetic differentiation within pairs of MZ twins. Research suggests that differentiation of this sort tends to be most pronounced for older twins and for twins who were reared apart (Fraga et al. 2005).

12. Violations of this assumption—that is, a non-zero correlation between $\epsilon$ and $S$ in equation (2)—would imply that education is not exogenously given, conditional on $X$, resulting in biased estimates of the causal relationship between schooling and mental health. We discuss this possibility in more detail later in the paper.

13. Estimates from the within-MZ models are necessarily less precise because variance between twin pairs is ignored.

14. We find similar patterns when we stratify our within-MZ fixed effects models by gender. See appendix table A4 for more details.

15. More finely grained analyses of cohort-specific patterns were unfortunately not feasible due to sample size considerations, nor were analyses that differentiate between possible age, period, and cohort effects.

16. The concern here is confined to the right-hand side of the equation. As Allison (1990) demonstrates, measurement error on the left-hand side is not a particular problem for fixed effects models or the equivalent difference score approach.

17. The $t$-statistic for education from the first-stage within-MZ instrumental variable model was 11.3, as noted at the bottom of table 5. The $F$-statistic was 64.3.

18. The goal is to locate a variable that is correlated with the respondent’s “true” educational attainment but uncorrelated with any reporting errors that they made, and then use that variable as an instrument. Cross-twin reports provide an ideal source for this type of information and have been used extensively in past research to obtain consistent estimates of education effects. See Ashenfelter and Kruger (1994), Behrman,

19. Significant nonlinearities have been detected in research on the education–mortality relationship (Montez, Hummer, and Hayward 2012), the education–earnings relationship (Jaeger and Page 1996), and the education–health relationship (Lundborg 2013; Mirowsky and Ross 2003). We are not aware of any studies that have specifically focused on this issue as it pertains to the relationship between education and mental health.

20. Most of the 378 twins who were dropped were dropped because they reported that they were never in the same classes at school \((n = 207)\).

21. These analyses are not meant to be comprehensive. A more rigorous approach would need to distinguish, analytically, between births that occurred before the end of a person’s educational career and after. See Amin and Behrman (2014) for an excellent example.

22. Estimates from similarly specified within-MZ Poisson fixed effects models and unpaired Poisson regression models were substantively identical to those shown in table 8.

23. It is worth noting that the effects presented in table 8 are not especially large and, as a result, should not be easier to detect than education effects on mental health. When we standardize all of the variables in our analysis and fit an uncorrected within-MZ model, we find that a one-standard-deviation increase in years of schooling corresponds to a tenth of a standard deviation decrease in number of children.

24. The fact that the within-MZ estimates are larger (although only marginally so in the uncorrected models) suggests that the OLS estimates may underestimate the true education effect. This would occur if there are unobserved confounds that constrain individuals’ education and promote fertility. In a recent study, Kohler, Behrman, and Schnittker (2011) observed similar results using data from the Minnesota Twin Registry.

25. We also left unspecified the mediating variables through which education is thought to influence adults’ mental health outcomes. One possible concern is that these intervening variables could disrupt or otherwise suppress the relationship we seek to estimate in this article. Although a formal causal mediation analysis is beyond the scope of this paper, we have a hard time imagining how such a scenario would arise. Most scholars argue that obtaining higher levels of schooling sets in motion a series of processes and events that widen the gap between those with more schooling and those with less (on average). The more highly educated secure better jobs, enjoy more stable relationships, develop better coping mechanisms and support systems, and are relatively more able to navigate and comprehend complex health information related to their physical and mental well-being. These mechanisms should all facilitate better mental health outcomes during adulthood, producing a stronger causal relationship between the two variables of primary interest. The opposite would have to be true for a suppressor effect to be present.

About the Authors

Andrew Halpern-Manners is an assistant professor in the Department of Sociology at Indiana University. His research focuses on educational gradients in adult attainments, school effects on students’ academic outcomes, and various issues related to longitudinal data collection and analysis. He has recently published
articles in these areas in the *American Journal of Sociology, Social Forces,* and *Demography.*

**Landon Schnabel** is a PhD candidate in the Department of Sociology at Indiana University. His research examines how beliefs, values, and identities influence politics, stratification, and well-being. He has recently published on these issues in *Public Opinion Quarterly, Gender & Society, Journal for the Scientific Study of Religion,* and *Social Science Research.*

**Elaine M. Hernandez** is an assistant professor in the Department of Sociology at Indiana University. Her research addresses the structural forces that contribute to the formation and reproduction of social inequalities in health. Her work has appeared in the *Journal of Health & Social Behavior, Social Science and Medicine,* and *Society & Mental Health.*

**Judy L. Silberg** is an associate professor in the Department of Human and Molecular Genetics at Virginia Commonwealth University, and the Scientific Director of the Mid-Atlantic Twin Registry. Her research interests include the etiology of behavioral and emotional disorders in children, the nature of familial transmission, and the development of genetically informative statistical methods.

**Lindon J. Eaves** is Distinguished Professor Emeritus of Human Genetics and Psychiatry at Virginia Commonwealth University. His research examines the genetics of cognitive function, personality, social attitudes, political behavior and juvenile and adult psychopathology, with a particular focus on modeling the effects of genes and the home environment.

**References**


