JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12399

DOES EDUCATION REALLY IMPROVE HEALTH? A META-ANALYSIS

Xindong Xue
School of Public Administration
Zhongnan University of Economics and Law
Wuhan, 430073, China

Mingmei Cheng
School of Public Finance and Taxation
Zhongnan University of Economics and Law
Wuhan, 430073, China

Wangyongxin Zhang
School of Public Administration
Zhongnan University of Economics and Law
Wuhan, 430073, China

Abstract. While numerous studies assess the relationship between education and health, no consensus has been reached on whether education really improves health. We perform a meta-analysis of 4866 estimates gleaned from 99 published studies that examine the health effects of education. We find that the current literature suffers from moderate publication bias towards the positive effects of education on health. After correcting for publication bias with an array of sophisticated methods, we find that the overall effect size is practically zero, indicating that education generates no discernible benefits to health. The heterogeneity analysis by Bayesian Model Averaging (BMA) and Frequentist Model Averaging (FMA) reveals that the reported estimates can be largely explained by whether the econometric models control for endogeneity of education, the types of data and the differences in health measurements. Our results also suggest that education may not be an effective policy option for promoting population health.

Keywords. Bayesian model averaging; Education; Health; Human capital; Meta-analysis; Publication bias

1. Introduction

The education-health nexus is one of most recognized and documented topics in the field of social sciences. Since the seminal paper of Grossman (1972), the literature on the subject has grown and expanded rapidly (Albarrán *et al.*, 2020; Avendano *et al.*, 2020; Janke *et al.*, 2020). While numerous studies have attempted to investigate this relationship, it remains ambiguous regarding whether education

Corresponding author contact email: xuexindong@zuel.edu.cn. Tel:+86 27 88387901.

really improves health. Understanding this question is of considerable importance. On the one hand, it can directly test the hypothesis of health production theory, which predicts that education improves health (Grossman, 1972). On the other hand, if education does have a large, beneficial effect on health, then education polices might serve as more effective tools for promoting health than merely increasing public health spending (Lleras-Muney, 2005; Clark and Royer, 2013)

The size and sign of the reported estimates in the literature vary greatly, depending the differences in estimation methods, measurements of health and country contexts. While some studies find significant effects of education on health (e.g. Lleras-Muney, 2005; Oreopoulos, 2006, 2007; Van Kipperslius *et al.*, 2011), some report small or no effects (e.g. Albouy and Lequien, 2009; Braakmann, 2011; Clark and Royer, 2013; Meghir *et al.*, 2018). Moreover, some studies yield mixed evidence across different aspects of health or sub-groups (e.g. Webbink *et al.*, 2010; Kemptner *et al.*, 2011; James, 2015). Given the diversity of findings in the current literature, we collect 4866 estimates from 99 published studies and employ meta-analysis to quantify the effects of education on health. Meta-analysis is a reliable and objective way to synthesize research findings and has been extensively used in the field of economics, especially in the case that the empirical literature lacks consensus (Stanley and Doucouliagos, 2012). We seek to answer the following questions: what is overall effect of education on health? Is there any publication bias in the reported estimates? What are the factors responsible for the heterogeneity of the estimates?

To the best of our knowledge, there are only two meta-analytic studies on the effects of education on health.[1] Furnee *et al.* (2008) is the first to conduct a meta-analysis on 88 estimates from 40 studies. They report that education has a significant effect on self-reported health. Hamad *et al.* (2018) provide a meta-analysis of the quasi-experimental studies of compulsory schooling laws. Their findings indicate that education has beneficial effects on most health outcomes. However, the two studies fail to correct for publication bias, which may distort the results of meta-analysis. Our study contributes to the existing literature in several ways. First, we carry out a systematic search and reporting procedure to extract standardized effect size estimates from different studies. Our study represents the most comprehensive analysis on the effects of education on health up to date, both in terms of sample size and number of health variables investigated. Second, we employ an array of methods, especially some recently developed sophisticated techniques to correct for publication bias, Third, as highlighted in Galama *et al.* (2020): '*there is substantial evidence of genuine heterogeneity in the estimated effects of education on health*'. To explore why the reported estimates differ in the current studies, we code 39 explanatory variables concerning the study attributes and employ both Bayesian Model Averaging (BMA) and Frequentist Model Averaging (FMA) to address model uncertainty.

Our main finding is that current literature suffers from moderate publication bias. After correcting for publication bias, the effect of education on health is close to zero, suggesting that education has no positive effects on health. In particular, studies ignoring endogeneity are prone to report larger effects of education on health. The heterogeneity in reported estimates can be largely explained by whether the econometric models control for the endogeneity of education, the types of data and the differences in health measurements. Our results cast doubt on the feasibility of policy options designed to promote population health through education interventions.

The rest of the paper proceeds as follows. Section 2 briefly reviews the literature. Section 3 describes the meta-dataset. Section 4 deals with publication bias. Section 5 explores the heterogeneity. Section 6 summarizes and concludes.

2. Literature Review

As noted above, there is a large literature that examines the health effects of education. In this section, we provide a brief review on the current studies. Our aim is not to be exhaustive because Section 5 provides detailed descriptions of the studies included in the final sample. Broad reviews can be found in Grossman and Kaestner (1997), Grossman (2006), Cutler and Lleras-Muney (2006, 2012) and Galama *et al.* (2020).

Several mechanisms have been proposed on how education might improve health. The first is the productive efficiency hypothesis, which asserts that education increases the productive efficiency from the given quantities of health inputs (Grossman, 1972, 2006). The second is the allocative efficiency hypothesis, which argues that education improves health through optimizing the mix of health inputs. Better-educated people usually have more information on the deleterious effects of smoking and bad habits, so that they are more likely to have healthy lifestyles (Rosenzweig and Schulz, 1989). The third is that education improves health through channels such as better labour market opportunities, higher income, better living conditions, higher quality of care and living environment (Card, 1999; Cutler and Lleras-Muney, 2012).

A pervasive problem in identifying the causal effect of education on health is endogeneity. First, the association between education and health could be spurious due to reverse causality. Healthier people usually have higher education (Behrman and Rosenzweig, 2004). Second, education and health may be simultaneously determined by the unobserved variables such as time preference (Fuchs, 1982), genetic endowments (Behrman *et al*., 2011) or family background (Bijwaard *et al*., 2015).

Several identification strategies have been implemented to address endogeneity and establish the causal relationship between education and health. The most widely used approach is to exploit an exogenous variation in education levels generated by the reform in compulsory schooling laws (CSL). The changes in CLS are often used as instrumental variables (IV), regression discontinuity design (RDD) or reduced form (RF) to estimate the parameter. The evidences so far are inconsistent (Galama *et al*., 2020). While Lleras-Muney (2005), in her influential study, find that education significantly reduces mortality in the USA, some studies report small or no such effects in the USA (Mazumder,2008; Black *et al*., 2015), the France (Albouy and Lequien, 2009), the UK (Clark and Royer, 2013), the Holland (van Kippersluis *et al*., 2011), the Sweden (Megir *et al*., 2018) and, more generally, European countries as a whole (Gathmann *et al*., 2015). Furthermore, the evidences regarding the impacts of education on other aspects of health are also mixed. Although some studies find that education improves self-reported general health (Oreopoulos, 2006; Silles, 2009; Fletcher, 2015) and mental health (Crespo *et al*., 2014; Mazzonna, 2014), and reduces the prevalence of obesity and diabetes (Brunello *et al*., 2013; James, 2015), some recent studies report no or negative effects of education on mental health (Albarrán *et al*., 2020; Avendano *et al*., 2020) and the occurence of a number of chronic conditions (Janke *et al*., 2020).

An alternative approach to account for endogeneity is twin fixed effects estimation. As twins share almost the same characteristics such as genetic inheritance, family background and innate ability, twin fixed effects can eliminate the confounding effects of these omitted variables. Using the Danish twin samples, Behrman *et al*. (2011) find that education has no significant effect on mortality, but Lundborg *et al*. (2016) report that education significantly reduces mortality in Sweden twins. The mixed findings are also reported in Madsen *et al*. (2010) and Lundborg (2013). Some other studies demonstrate that education reduces the overweight for men (Webbink *et al*., 2010) but not for women (Amin *et al*., 2013).

It is obvious that the previous studies report heterogenous results. Thus, it becomes an empirical question to ascertain whether and to what extent education improves health. The meta-analysis can do this job.

3. The Meta-Data Set

3.1 *Search Strategy and Inclusion Criteria*

The first step in meta-analysis is to extract estimates from current studies. Following the reporting guidelines proposed by Havranek *et al*. (2020), we search for the potential studies in the following databases: *EconLit, Web of Knowledge, Google Scholar, JSTOR, EBSCO, RePEc, IDEAS, SSRN, Scopus, NBER, IZA, OECD Library* and *World Bank Publications*. The following key words are combined: ‘education’, ‘schooling’, ‘health’, ‘mortality’, ‘disease’, ‘obesity’, ‘BMI’, ‘morbidity’, ‘depression’,

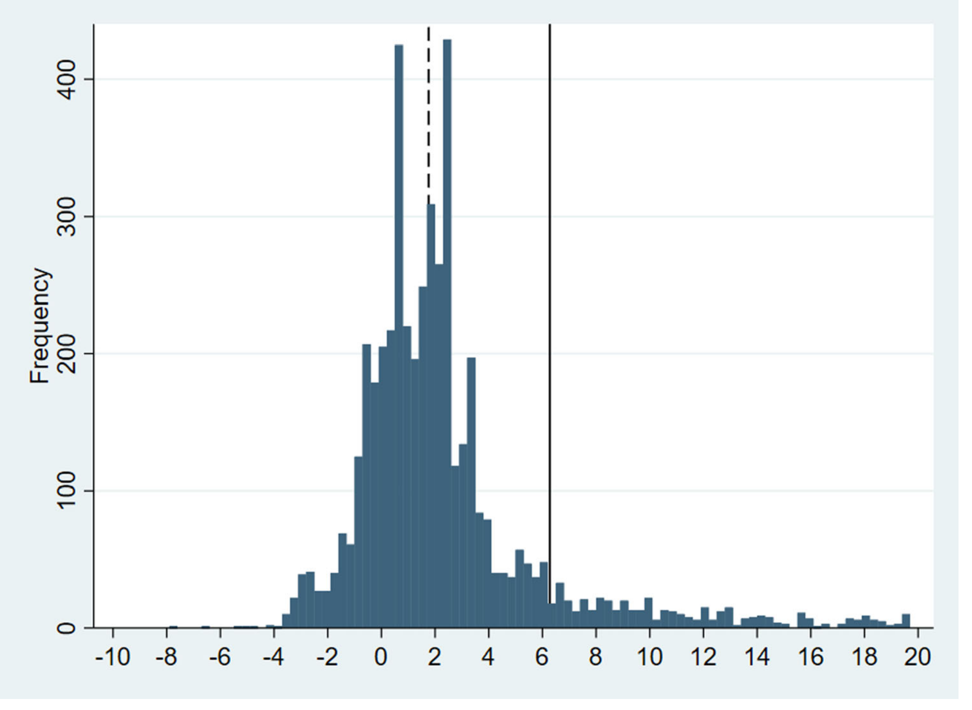

Figure 1. Distribution of t-Statistics. [Colour figure can be viewed at wileyonlinelibrary.com]
Notes: The solid line shows the mean of reported t-statistics; the dashed line denotes the mean of the median estimates of the study.

'cognition', 'life expectancy' and 'survival'. To be as comprehensive as possible, reference snowballing techniques is also leveraged to identify papers through the search engine process. The searching process was completed at the end of December 2018 and updated in July 2020.

Our search produces 508 studies. To make the analysis coherent, we apply the following inclusion criteria. First, the study must investigate the relationship between one's education and his/her own health with health as the dependent variable. This criterion excludes the studies which investigate the intergenerational or spillover effects of education on health (e.g. mother's education on infant health), and the impacts of education on health-related behaviours. Second, the study must report at least one empirical estimate quantifying the impact of education on health, which eliminates the theoretical studies or systematic reviews. Third, the study must contain sufficient information to compute t-statistics (standard errors, t-statistics, p-value, 95% confidence Intervals etc.).[2] Fourth, we exclude studies that include interaction terms and quadratic specifications of the education variable because it is difficult to extract the associated marginal effects (Gunby *et al*., 2017; Xue *et al*., 2020).[3] Finally, we only consider published studies because it is becoming a norm to focus on published studies in recent meta-analysis studies (see Havraneck *et al*., 2017; Cazachevici *et al*., 2020; Havranek and Sokolova, 2020). Compared to unpublished manuscripts, published studies are subject to peer-reviewing process and less likely to suffer from mistakes.[4] As a result, our final dataset consists of 4866 estimates from 99 published papers. The list of included studies is provided in Appendix A.

Figure 1 plots the distribution of t-statistics used in our dataset. It can be seen that the distribution is right-skewed with many studies reporting positive estimates. Of the 4866 t-statistics, 52.49% (2554) lie between –2 and 2, 44.16 % (2149) are greater than 2, and only 3.35% (163) report negative values fewer than –2. The mean and median values of *t*-statistics are 6.273 and 1.762, with minimum and maximum values of 12.81 and 8680. The degrees of freedom also vary greatly, from 49 to 3,781,410. The extreme values raise concerns about outliers, which may distort the validity and robustness of the conclusions from a meta-analysis (Viechtbauer and Cheung, 2010). To alleviate this problem, we choose to winsorize *t*-statistics and degrees of freedom at the top and bottom of 5% level. Winsorization is a method that has been advocated for use in meta-analysis (Lipsey and Wilson, 2001). It replaces the extreme values with the highest values in given percentiles, without loss of observations. Note that we also winsorize our data at 1% level, but our main conclusions remain quantitatively unchanged.

3.2 *Effect Sizes*

Let a typical study *i* employs the following regression specification to estimate the effect of education on health:

$$H_{in} = \delta_i + \beta_i educ_{in} + \sum_{k=1}^{K} \gamma_{ik} X_{ikn} + \varepsilon_i \tag{1}$$

where H_{in} and $educ_{in}$ denote the health and education of individual *n*, and X_{ikn} is a measure of individual characteristics. β_i is the parameter to be estimated, representing the effect of education on health. In meta-analysis, $\hat{\beta}_i$ becomes the dependent variable of interest.[5]

However, the estimated coefficients $\hat{\beta}_i$ cannot be directly used in meta-analysis because they are not comparable across studies. Estimated effects differ because studies use different measures of health and education, or apply different estimation methods. Studies use a variety of health measures, including mortality, disease, self-reported health, activities of daily living (ADL), BMI, depression etc. The same is true for education measures, which can be continuous (actual years of education) or categorical (primary, secondary or tertiary). Appendix B provides detailed description of these measures.

To make the reported estimates comparable, we follow a common approach and use partial correlation coefficients (PCCs) to convert estimates into a unitless, comparable measure. PCC is calculated as follows:

$$\mathrm{PCC_i} = \frac{t_i}{\sqrt{{t_i}^2 + df_i}}$$

where t_i and df_i are the *t*-statistic and degrees of freedom for the estimated effect of education on health in study *i*. The value of PCCs lies between –1 and 1. The corresponding standard error is:

$$\mathrm{SE(PCC_i)} = \sqrt{\frac{1 - {PCC_i}^2}{df_i}} = \frac{PCC_i}{t_i}$$

Both PCC and $SE(PCC_i)$ will be used in the following analysis.

Figure 2 displays box plots of within-study PCC values for the 99 studies in our sample. It appears that PCCs vary widely both within and across studies. Although most studies report positive estimates, about half of these PPC values are less than 0.1, implying that education may have a small impact on health.

To get a sense of the heterogeneous effects of education on health, we partition the sample of all estimates into subsamples according to (i) the different measures of health and education, (ii) whether the study controls for endogeneity of education, (iii) whether the study appears in economics journal or (iv) comes from high-income countries.[6] Table 1 reports the mean PCC values for all estimates and different groups of estimates. Column (1) shows unweighted means and column (2) reports weighted means. The overall mean of the PCC values is approximately 0.03. Studies controlling for endogeneity report smaller PCC values on average, about 0.018. Among the different journal types, economics journals appear to

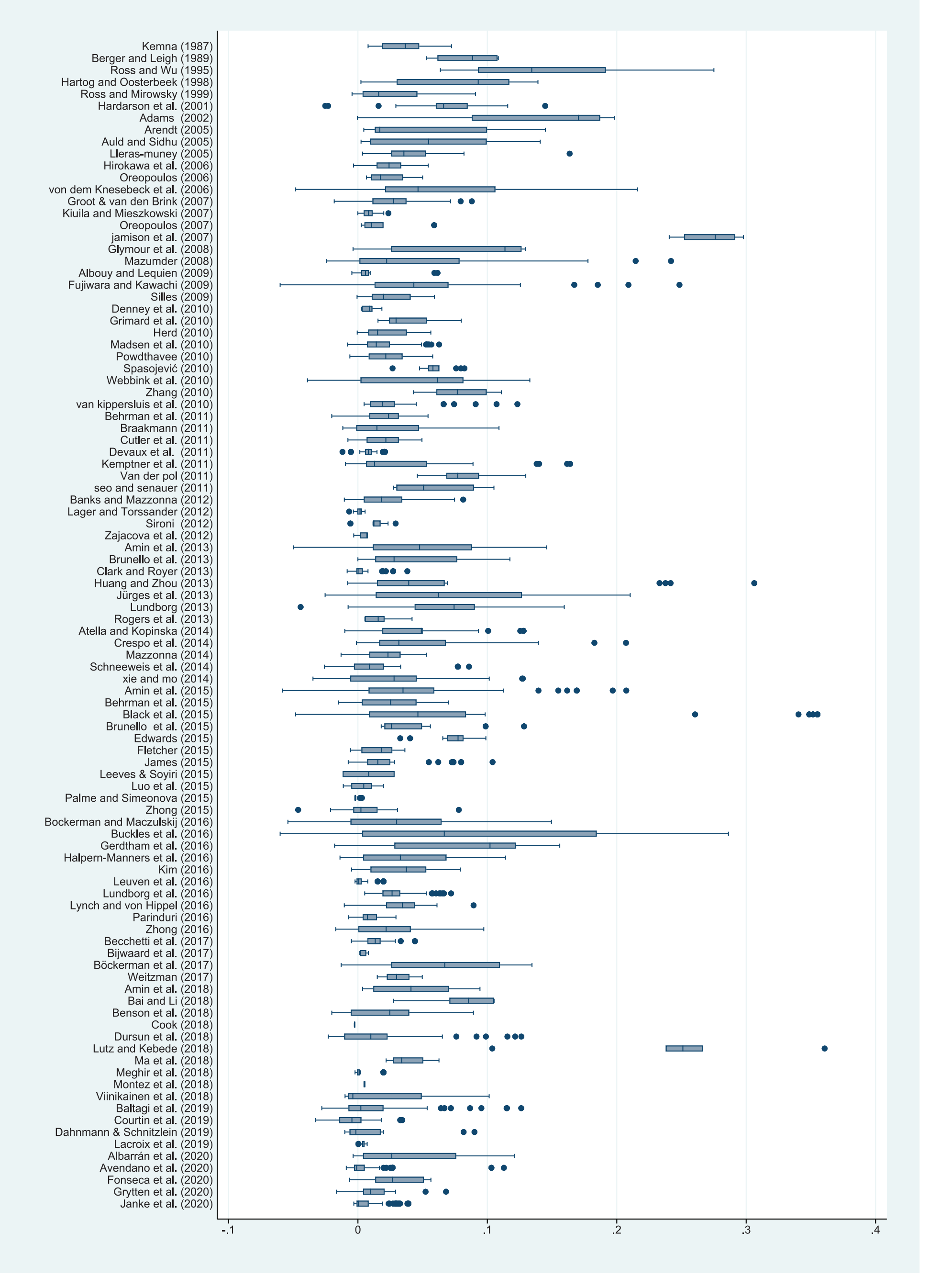

Figure 2. The Box Plot of PCCs within Studies. [Colour figure can be viewed at wileyonlinelibrary.com]
Notes: Dots denote outliers.

Table 1. The Mean Effect Sizes of Education on Health.

	Unweighted	Weighted	Observations
All estimates	0.03	0.039	4866
Subsample estimates			
Physical health	0.024	0.035	3280
Mental health	0.02	0.03	597
Self-reported health	0.056	0.053	989
Primary education	0.04	0.041	358
Secondary education	0.019	0.027	1061
Tertiary education	0.031	0.031	541
Years of education	0.032	0.043	2906
Studies controlling endogeneity	0.015	0.022	2337
Studies not controlling endogeneity	0.043	0.054	2529
Economics journals	0.032	0.041	2920
Non-economics journals	0.027	0.035	1946
High-income countries	0.03	0.038	4484
Middle-low income countries	0.027	0.047	382

Notes: The table presents the mean PCCs of the education effects on health for all estimates and for selected subsamples. In weighted means, PCC values are weighted by the inverse of the number of reported estimates per study.

report slightly larger estimates than non-economics journals. The difference between high-income and middle-low income countries is not significant. Generally, all the means in the subsamples are very close to zero. According to the guidelines proposed by Doucouliagos (2011), PCC values less than 0.07 can be considered ‘small’, with 0.17 the threshold for ‘moderate’, and 0.33 and above ‘large’. None of the mean PCC values even attain the ‘small’ threshold, indicating that education has a very small effect on health. A caveat is that the simple overall mean effects should be interpreted with caution due to the possibility of publication bias. In the next section, we will investigate whether there is publication bias and how it might affect the reported estimates in the current literature.

4. Publication Bias

In meta-analysis, publication bias poses a serious threat to the validity of analytic results. Publication bias arises when journals and editors are more likely to publish significant results, or when authors choose to bury their insignificant results. This latter phenomenon is known as the ‘file drawer problem’ (Rosenthal, 1979). As a result, publication bias leads to distorted estimates of true effect size.

An intuitive way to detect publication bias is the funnel plot, which plots the effect size on the horizontal-axis and the precision of the estimates on the vertical axis (Egger *et al*., 1997). If there is no publication bias, the distribution of the standard error will be symmetrically distributed around the mean line. Publication bias introduces asymmetry into the funnel plot. In the presence of an upward bias, the scatter-dot will cluster on the right of the mean line, or vice versa. Figure 3 shows the funnel plot of the PCCs and their standard errors in the sample. It is obvious that as the standard errors increase, the PCC values are skewed to the right, indicating an upward publication bias towards the positive impacts of education on health.

A formal, widely adopted way to test publication bias is the ‘Funnel Asymmetry Test’ – ‘Precision Effect Test’ (FAT-PET) (Stanley, 2005; Stanley, 2008). FAT-PET is a simple meta-regression of the PCCs

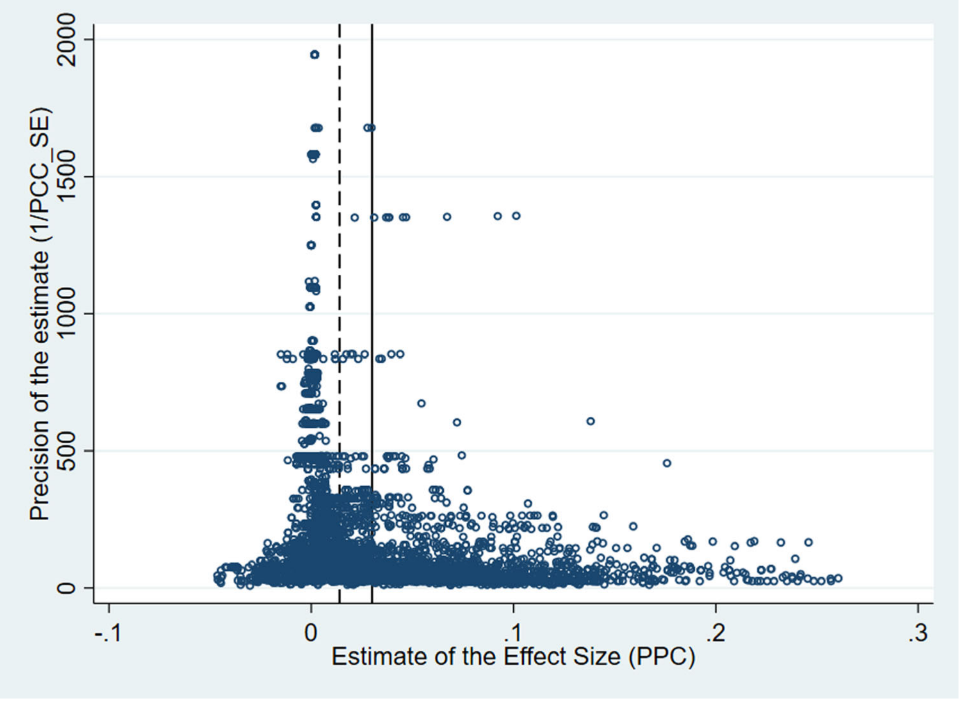

Figure 3. Funnel Plot of PCC. [Colour figure can be viewed at wileyonlinelibrary.com]
Notes: The horizontal axis represents the PCC values. The vertical represents the inverse of the standard errors of PCC. The solid vertical line denotes the mean of the sample. The dashed vertical line displays the median of the sample.

on their standard errors:

$$PCC_i = \alpha_0 + \alpha_1 SE\,(PCC_i) + \epsilon_i \tag{2}$$

where PCC_i is the ith partial correlation coefficient and $SE(PCC_i)$ is the corresponding standard error. α_1 measures the severity of publication bias. α_0 is the true effect after correcting for publication bias. In the presence of publication bias, there will be correlation between *PCC* and *SE*(*PCC*). Accordingly, publication bias is indicated by the statistical significance of the estimate of α_1.

Stanley (2008) argues that α_0 in equation (2) may be biased downward when the null hypothesis is rejected. Stanley and Doucouliagos (2014) propose to replace the standard error with its square. In this case, the meta-regression is called the Precision Effect Estimate with Standard Error (PEESE), which is specified as follows:

$$PCC_i = \alpha_0 + \alpha_1 SE(PCC_i)^2 + \upsilon_i \tag{3}$$

In practice, equations (2) and (3) are not directly estimated because of the apparent heteroskedasticity. Weighted least square (WLS) is routinely employed to address heteroskedasticity. Two estimators are commonly used in the meta-analysis literature: Fixed Effects (FE) and Random Effects (RE), not to be confused with the panel data estimators of the same name. WLS-FE implicitly assumes that there is a single underlying true effect and the reason for different estimates across studies is due to sampling error. The weight for WLS-FE is the inverse of the variance of the estimated effect, $1/\{SE(PCC_i)^2\}$. In contrast, in WLS-RE, true effects are assumed to differ across studies. The variation in estimated effects consists of two parts: sampling error and heterogeneity in the true effect. The weight for WLS-RE is $1/\{SE(PCC_i)^2 + \tau^2\}$, where τ^2 measures the variance of the true effect.

Another potential concern with equations (2) and (3) is the possible endogeneity of standard error. If some unobserved study characteristics (e.g. estimation techniques, study quality) drive the reported estimates and their standard errors in a systematic way, the coefficient of publication bias could be false. In line with previous literature (e.g. Havranek *et al.*, 2018; Cazachevici *et al.*, 2020), we employ an instrumental variables (IV) approach with the inverse of the square root of the degrees of freedom as an instrument for standard errors. The identification condition is that the square root is plausibly correlated with the standard errors but is unlikely to be correlated with the use of a particular estimation method.

Table 2 displays the results of four specifications based on equation (2): simple OLS, WLS-FE, WLS-RE and IV. For each specification, we report cluster-robust standard errors, with clustering by study to accommodate within-study correlation of estimates. Following Cameron *et al.* (2008), we also report wild-bootstrap confidence intervals.[7] Furthermore, we employ two weighting schemes for each specification: equal weights for each estimate (weight 1) and equal weights for each study (weight 2). In three of the four columns, we reject the null hypothesis: $\alpha_0 = 0$, with the estimates of α_0 at least significant at the 10% level. The only exception is the 'WLS-FE' regression.

In general, the precision effect coefficients suggest that education is positively correlated with health. The publication bias-corrected estimates of the mean true effect of education on health range from 0 to 0.008, far below the value that Doucouliagos (2011) identifies as being 'small'. All the four columns also reject the null hypothesis: $\alpha_1 = 0$ at the 5 % significance level, indicating the presence of publication bias. The positive publication bias coefficients suggest upward publication bias, indicating that the current literature favours the publication of positive impacts of education on health. According to Doucouliagos and Stanley (2013), the literature suffers from substantial publication selectivity if the absolute value of publication bias coefficient (α_1) lies between 1 and 2. Our estimated bias coefficients lie between 1.6 and 2.5. Given that the bias-corrected estimates are very close to the simple mean of all estimates, we argue that there is moderate publication bias in the current literature.

Table 3 presents the PEESE results. Compared to Table 2, all the four columns reject the null hypothesis $\alpha_1 = 0$, $\alpha_0 = 0$. The effects of bias become larger and significant at the 1% level, further confirming the publication bias in the literature. The publication bias-adjusted estimate of the mean true effect is more significant and slightly larger, ranging from 0.006 to 0.023. Again, these values are still far below the threshold that Doucouliagos (2011) identifies as being 'small'. In sum, Tables 2 and 3 consistently support the conclusion that the overall effect of education on health is close to zero and the current literature suffers from positive publication bias.

In addition to the FAT-PET-PEESE procedure described above, a number of more sophisticated methods have been developed by researchers. We employ the following four methods to estimate the bias-adjusted true effect: the weighted average of adequately powered (WAAP) estimator by Stanley *et al.* (2017), the endogenous Kink (EK) estimator by Bom and Rachinger (2019), the AK estimator by Andrews and Kasy (2019) and the Stem-based method by Furukawa (2019). Table 4 reports the results of these estimates. First, WAAP computes the unrestricted WLS-weighted average on those estimates that are 'adequately powered', usually defined as their standard error is smaller than the WLS estimates divided by 2.8 (Stanley *et al.*, 2017). Monte-Carlo simulations demonstrate that WAAP reduces bias compared to FE and RE estimators in the case of selective publication. The estimate of WAAP, as shown in Table 4, is 0.004. Second, the EK estimator is a refinement of the PET-PEESE approach in that it attempts to better fit the non-linearity of the relationship between the estimated effect and the SE in the presence of publication bias. There is no selective publication when the standard error is very small. The publication selection usually increases with the standard error. EK estimates this threshold and introduces a non-linearity test of publication bias. Table 4 reports that the EK estimate is almost zero.

Table 2. FAT-PET.

	OLS	WLS-FE	WLS-RE	IV
Weight 1: Equal weight to each estimate				
Publication bias (α_1)	1.656*** (0.205)	2.537*** (0.308)	1.821*** (0.213)	2.077*** (0.235)
	[1.23; 2.15]	[1.89; 3.19]	[1.38; 2.27]	[1.62; 2.54]
Precision effect (α_0)	0.008** (0.003)	–0.00002 (–0.02)	0.006** (0.003)	0.003* (0.001)
	[0.002; 0.015]	[–0.002; 0.003]	[0.001; 0.012]	[–0.0004; 0.006]
Adjusted R^2	0.15	0.00	0.38	0.34
Observations	4866	4866	4866	4866
Number of studies	99	99	99	99
Weight 2: Equal weight to each study				
Publication bias (α_1)	2.117*** (0.473)	2.907*** (0.278)	2.234*** (0.415)	2.495*** (0.33)
	[1.24; 3.13]	[2.34; 3.47]	[1.44; 3.07]	[1.85; 3.19]
True effect (α_0)	0.008* (0.005)	0.0004 (0.001)	0.007* (0.004)	0.003 (0.002)
	[–0.001; 0.017]	[–0.002; 0.003]	[–0.001; 0.014]	[–0.001; 0.007]
Adjusted R^2	0.17	0.0003	0.43	0.39
Observations	4866	4866	4866	4866
Number of studies	99	99	99	99

Notes: The table shows the results of regression in equation (2). Robust standard errors clustered at study level are shown in parentheses. Wild bootstrap confidence intervals are reported in square brackets.
OLS = ordinary least squares, WLS-FE = weighted least square fixed effects; WLS-RE = weighted least square-random effects. IV = instrument variables regression with the inverse of the square root of the degree of freedom used as an instrument.
***Significant at 1%.
**Significant at 5%.
*Significant at 10%.

Table 3. PEESE.

	OLS	WLS-FE	WLS-RE	IV
Weight 1: Equal weight to each estimate				
Publication bias (α_1)	34.76^{***} (5.56)	80.11^{***} (11.53)	40.43^{***} (5.9)	80.65^{***} (11.56)
	[22.96; 49.5]	[58.05; 105]	[28.07; 55.16]	[57.36; 108.8]
True effect (α_0)	0.019^{***} (0.003)	0.006^{***} (0.001)	0.018^{***} (0.003)	0.006^{***} (0.001)
	[0.013; 0.026]	[0.004; 0.01]	[0.012; 0.024]	[0.004; 0.009]
Adjusted R^2	0.11	0.26	0.34	0.27
Observations	4866	4866	4866	4866
Number of studies	99	99	99	99
Weight 2: Equal weight to each study				
Publication bias (α_1)	46.23^{***} (12.75)	95.42^{***} (12.57)	52.17^{***} (12.37)	96.2^{***} (12.66)
	[24.3; 76.93]	[70.17; 121.5]	[29.01; 79.89]	[70.96; 121.9]
True effect (α_0)	0.023^{***} (0.003)	0.007^{***} (0.001)	0.021^{***} (0.003)	0.007^{***} (0.001)
	[0.017; 0.03]	[0.004; 0.01]	[0.016; 0.027]	[0.004; 0.01]
Adjusted R^2	0.14	0.31	0.4	0.32
Observations	4866	4866	4866	4866
Number of studies	99	99	99	99

Notes: The table reports the results of regression in equation (3). Robust standard errors clustered at study level are shown in parentheses. Wild bootstrap confidence intervals are reported in square brackets.

OLS = ordinary least squares, WLS-FE = weighted least square fixed effects; WLS-RE = weighted least square-random effects. IV = instrument variables regression with the inverse of the square root of the degree of freedom used as an instrument.

***Significant at 1%.

Table 4. Bias-Adjusted Mean Effects with Modern Methods.

	Mean effect	Standard error
WAAP	0.004***	0.0003
EK	–0.00002	0.001
AK1 (symmetric)	0.036***	0.001
AK2 (asymmetric)	0.006***	0.002
Stem-based method	0.02***	0.0003

WAAP = weighted average adequately powered; EK = endogenous kink; AK = Andrew & Kasy.
***Significant at 1%.

Third, Andrews and Kasy (2019) propose two approaches to correct for publication bias. AK1 (symmetric) estimator accounts for the selective publication on statistical significance and AK2 (asymmetric) estimator addresses the selective publication caused by both statistical significance and the sign of the estimates. The results of the AK estimators show that the true effect of education on health is also very small. Last, Furukawa's (2019) stem-based method is a non-parametric estimator that focuses on the most precise studies. It is a generally conservative approach to create a bias-corrected estimate that can work under many different publication selection processes. The estimation of stem-based method yields a mean effect of 0.02. In summary, the effect of education on health is still very small in all the four methodological approaches. None of them attains Doucouliagos's cut-off value for 'small'.

To further shed light on the robustness of our results with regard to whether endogeneity matters for the mean effects, we re-estimate equation (2) with the subsamples of observations that either do or do not address endogeneity. This can give us a 'genuine', causal effect of education on health. The results are reported in Appendix C.[8] In the subsample of estimates controlling for endogeneity, estimates of the mean effect are close to zero and statistically insignificant, smaller than the estimated mean effect that do not control for endogeneity.

One of the critics of PCCs is that it can be difficult to interpret the economic significance of the transformed PCCs. In our dataset, there is no study reporting the elasticity of health with respect to education. The only substantial effect that would allow us to assess economic impact is the marginal effects of years of education on mortality, that is, how large an additional year of education reduces mortality. To this end, we construct a sample of 497 observations from 17 studies and repeat equation (2). The results are shown in Appendix D. Despite that the mean value of raw estimates is –0.488, all the effects become insignificant and negligible after publication bias is corrected for. The overall results again confirm our primary conclusion: education generates no sizable benefits to health.

5. Heterogeneity Analysis

5.1 *Variables Description*

To capture the factors underlying the systematic differences among the reported estimates, we code 39 variables according to the following categories: the measure of health, the measure of education, sample characteristics, data characteristics, estimation methods, calculation of t-statistics, publication characteristics, regions and income levels of the countries. Before we proceed, we conduct multicollinearity diagnostic test on all the variables. The values of variance-inflation factors for all the variables are below 7. Table 5 presents the definition, mean, standard deviation, and the weighted mean (with the weight being the inverse of the number of estimates per study) of all variables included for heterogeneity analysis.

Table 5. Description and Summary Statistics of Variables.

Variable	Definition	Mean	SD	WM
Effect size				
PCC	Partial correlation coefficient	0.030	0.047	0.039
SE	Standard error of the PCC	0.013	0.011	0.014
Health measures				
Physical health (Ref.)	=1 if health measure is activities of daily living, health limitations disease, mortality, obesity etc., 0 otherwise	0.674	0.469	0.652
General health	=1 if health measure is general health, 0 otherwise	0.203	0.402	0.231
Mental health	=1 if health measure is mental health, 0 otherwise	0.123	0.328	0.117
Self-reported	=1 if health is self-reported, 0 otherwise	0.587	0.492	0.577
Education measures				
Primary (Ref.)	=1 if education is primary, 0 otherwise	0.0745	0.261	0.046
Secondary	=1 if education is secondary, 0 otherwise	0.218	0.413	0.15
Tertiary	=1 if education is tertiary, 0 otherwise	0.111	0.314	0.145
Years of education	=1 if education is actual years of education, 0 otherwise	0.597	0.491	0.658
No. of educ. variables	Number of education variables in estimation equation	1.740	1.424	0.688
Sample characteristics				
Whole sample (Ref.)	=1 if estimate is from whole population, 0 otherwise	0.51	0.500	0.572
Subsample	=1 if estimate is from subsample population, 0 otherwise	0.49	0.500	0.428
Twin_sample	=1 if estimate is from twin sample, 0 otherwise	0.151	0.358	0.131
Data characteristics				
Cross-section (Ref.)	=1 if estimate is from cross-sectional data, 0 otherwise	0.389	0.487	0.434
Panel	=1 if estimate is from panel data, 0 otherwise	0.611	0.487	0.566
Aggregate data (Ref.)	=1 if estimate is from aggregate-level data, 0 otherwise	0.016	0.124	0.044
Individual data	=1 if estimate is from individual-level data, 0 otherwise	0.984	0.124	0.956
Survey data	=1 if estimate is from survey data, 0 from administrative data	0.767	0.423	0.779

(*Continued*)

Table 5. *Continued.*

Variable	Definition	Mean	SD	WM
Time span	Logarithm of number of years in the sample	1.941	1.235	1.9
Number of variables	Logarithm of number of explanatory variables in estimation	2.383	0.703	2.394
Estimation methods				
OLS (Ref.)	=1 if ordinary least squares are used for estimation, 0 otherwise	0.227	0.419	0.267
Non-linear	=1 if nonlinear models are used for estimation, 0 otherwise	0.292	0.455	0.261
FE	=1 if fixed effects are used for estimation, 0 otherwise	0.081	0.273	0.08
TSLS	=1 if two-stage least squares are used for estimation, 0 otherwise	0.278	0.448	0.286
RD	=1 if regression discontinuity is used for estimation, 0 otherwise	0.060	0.238	0.062
RF	=1 if reduced forms are used for estimation, 0 otherwise	0.061	0.240	0.046
Calculation of t-statistics				
tNormal (Ref.)	=1 if *t*-statistic=coefficient/standard error, 0 otherwise	0.663	0.473	0.763
tCalculatedBypValue	=1 if *t*-statistic is derived from *p*-values, 0 otherwise	0.183	0.387	0.147
tCalculatedByCI	=1 if *t*-statistic is derived from confidence intervals, 0 otherwise	0.153	0.360	0.09
SE spherical (Ref.)	=1 if standard error is spherical, 0 otherwise	0.407	0.491	0.391
SE non-spherical	=1 if standard error is non-spherical, 0 otherwise	0.593	0.491	0.609
Publication characteristics				
Publication year	Logarithm of publication year	7.608	0.002	7.607
Citescore	Scopus Citescore Metrics (2019)	4.656	2.566	4.897
Non-economic journal (Ref.)	=1 if published in non-economic journals (medicine, public health, epidemiology, sociology, demography etc.), 0 otherwise	0.400	0.490	0.414
Economic journal	=1 if published in economics journals, 0 otherwise	0.600	0.490	0.586
Regions				
Europe	=1 if countries in Europe included, 0 otherwise	0.582	0.493	0.506
North America	=1 if countries in North America included, 0 otherwise	0.3	0.458	0.35

(*Continued*)

Table 5. *Continued.*

Variable	Definition	Mean	SD	WM
Asia Pacific	=1 if countries in Asia Pacific included, 0 otherwise	0.111	0.314	0.114
Global	=1 if global countries included, 0 otherwise	0.007	0.083	0.03
Income level				
High income	=1 if countries with high-income country included, 0 otherwise	0.921	0.269	0.879

Notes: Citescores are collected from the 2019 Scopus Journal Metrics.
SD = standard deviation; WM = mean weighted by the inverse of the number of estimates per study; Ref. = reference category.

5.1.1 *Health measures*

World Health Organization (WHO, 1946) defines health as '*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*'. Since health is a multi-dimensional concept, the current studies use a variety of health measures. We categorize these measures into three broad types: physical health, mental health and general health. Physical health includes mortality, obesity (BMI, overweight, body size), onset of a particular illness or disease such as hypertension, heart disease, diabetes, and Activities of Daily Living (ADL). Mental health consists of depression and cognition. For measures of general health, it is a common practice to elicit individual's general health through a five-point, subjective self-assessment: very poor, poor, fair, good and excellent. We also create a dummy variable to indicate whether the health measure is self-reported. In our dataset, around two-thirds (67.4%) of the estimates use physical health as dependent variable. General health and mental health account for 20.3 % and 12.3 % of the total estimates, respectively.

5.1.2 *Education measures*

Education can be defined as continuous and categorical. Continuous measure refers to the actual years of education or schooling. Categorical education is coded into three levels: primary, secondary and tertiary. In our dataset, around 60% of the estimates use years of schooling to measure education. The remaining 40% use categorical education. The mean number of education variables in the dataset is 1.74.

5.1.3 *Sample and data characteristics*

The sample and data characteristics include a set of dummy variables to indicate whether the estimates are samples from the whole population or a subsample of the population, or from a twin sample, and the type of the dataset. About half of the estimates (49%) are from the whole population, 15.1 % are from twin samples and 61% of the estimates are based on panel data. The majority (98.4%) of the estimates are from individual-level data. And 76.7% of estimates are from survey data. The mean time span in the dataset is 13.8 years. The mean number of explanatory variables in the dataset is 2.38.

5.1.4 *Estimation methods*

While many studies use OLS or non-linear models (probit, logit, ordered probit, Cox proportional hazard), some studies control for the endogeneity by employing panel fixed effects (FE), instrumental variables (IV), regression discontinuity (RD) and reduced form (RE). Accordingly, we code the estimation methods into following six categories: OLS, non-linear model, FE, IV, RD and RF. We will examine whether estimation methods make a great difference for the effects of education on health.

5.1.5 *Calculation of* t-*statistics and standard error*

We also code the methods used to derive *t*-statistics and standard errors. There are three corresponding dummy variables: *tNormal*, *tCalculatedBypValue* and *tCalculatedByCI*. 66.3% of the *t*-statistics are calculated by normal ways (coefficient/SE), followed by confidence intervals (18.3%) and *p*-values (15.3%). The types of standard error associated with estimates are coded as *SEspherical* and *SEnonspherical*. About half (49.2%) of the estimates assume non-spherical standard errors.

5.1.6 *Publications characteristics*

We construct three sets of variables: 'publication year' to control for the potential time trend, 'Citescore metrics' to account for study quality and 'economics journal' to evaluate whether studies published in economics journal report systematically different estimates compared to studies published in non-economics journals.

5.1.7 *Regions and income levels*

We create four regional categories: Europe, North America, Asia Pacific and Global countries. 58.2% of the estimates are from Europe. 11.1% are from North American countries. 11.1 % are from Asian-Pacific countries. As education may have different effects on health in rich and poor countries, we categorize a country's income level on the basis of UN classifications.[9] Most estimates (92.1%) use data from high-income counties. Only a small fraction (7.9 %) use data from middle-low income countries.

5.2 *Methods*

We undertake the heterogeneity analysis by adding a set of the variables into equation (2):

$$PCC_i = \alpha_0 + \alpha_1 SE\left(PCC_i\right) + \sum_{k=1}^{K} \alpha_{k+1} X_{ik} + \varepsilon_i \tag{4}$$

where X_{ik} captures various study- and regression-specific characteristics associated with the estimate from study i. Similarly, equation (4) is divided by $SE(PCC_{ij})$ to account for heteroskedasticity:

$$\frac{PCC_i}{SE\left(PCC_i\right)} = \alpha_0 \frac{1}{SE\left(PCC_i\right)} + \alpha_1 + \sum_{k=1}^{K} \alpha_{k+1} \cdot \frac{1}{SE\left(PCC_i\right)} \cdot X_{ik} + \varepsilon_i \frac{1}{SE\left(PCC_i\right)} \tag{5}$$

A fundamental problem is estimating equation (5) is model uncertainty associated with the variables to include. Having the wrong variables in the equation leads to misspecification bias and invalid inference. Model selection and model averaging are two popular strategies employed in the literature to address model uncertainty (Steel, 2020). The most frequently used model selection is stepwise regression.

However, this approach fails to account for the selection process when reporting the final equation's results. In contrast, model averaging takes into account all possible models and, thus, its results are not dependent on a particular model selection. Hence, we follow the recent literature on meta-analysis (see Havranek *et al*., 2017; Havranek *et al*., 2018; Havranek and Sokolova, 2020; Zigraiova and Havranek, 2016) and apply the increasingly used Bayesian Model Averaging (BMA) and Frequentist Model Averaging (FMA) approaches to deal with model uncertainty.

Intuitively, BMA runs regressions with all the possible combinations of explanatory variables and computes the weighted averages of the estimated coefficients. The weight used in BMA is termed as 'posterior model probabilities' (PMP), which measures the 'goodness of fit' of each model with the data. To facilitate the computing process, Markov chain Monte Carlo (MCMC) algorithm is routinely employed to select models with largest PMP. For each variable in the model, BMA reports three parameters: posterior mean, posterior standard deviation and posterior inclusion probability (PIP). PIP aggregates the PMPs of all the models in which the variable included. PIP above 0.5 is usually regarded as the threshold to include variables into the model (Eicher *et al*., 2011). In the process of implementing BMA, we set the uniform model prior which assigns the same prior weight to each model and unit information g-prior which gives the zero-weight to the prior of each coefficient.[10]

Compared to BMA, FMA does not require explicit prior information. FMA focuses on estimators with the optimal properties (usually the minimum squared errors) from the repeated samples. We employ Mallow's model averaging estimator to select the asymptotically optimal weights (Hansen, 2007). Furthermore, to reduce the number of estimated models, we follow Amini and Parmeter (2012) and orthogonalize the covariate space.

5.3 *Results*

Figures 4 shows the graphics of BMA results. The vertical axis lists the explanatory variables sorted by PIP in descending order. The horizontal axis is the PMP of each model sorted in ascending order. The blue (dark) colour indicates the positive sign of the variable in the model, and the red colour (light) denotes the negative sign of the variable. The blank cell suggests that the variable is excluded from the regression model.

Table 6 presents the empirical results of BMA and FMA. We also report the OLS using the variables from BMA with PIP higher 0.5. Overall, FMA and OLS largely confirm the results of BMA. The BMA results shows that there are 18 variables with PIP higher than 0.5, indicating that they are relevant for the differences in the estimated effects of education on health.

The publication bias term is statistically significant and positive in all the specifications. The positive, significant coefficients of publication bias after controlling for a set of moderator variables suggests that the results from Tables 2 and 3 are not a spurious outcome caused by omitted variables. It further proves that the current literature suffers from publication bias

According to our results, the studies that use general health and self-reported health measures are more likely to report a more positive effects of education on health. This finding is consistent with the study by Johnston *et al*. (2009), who find that self-reported/general health measures tend to inflate the real health status because the respondents may be unaware of their chronic conditions. However, studies using mental health as the dependent variable reporter weaker effects of education on health. This finding is in line with Averndano *et al*. (2020), who show that increasing years of education through compulsory schooling laws could incur psychological and emotional burdens on those forced to be in school. Another possible reason for the weaker effects is that more education years are often accompanied by higher psychological stress, which may in turn offset the potential mental health benefits (Dahmann and Schnitzlein, 2019). Regarding education measures, we find that studies controlling for more education variables tend to report smaller effects of education on health.

Table 6. Explaining Heterogeneity in the Estimated Effects of Education on Health.

	BMA			FMA			OLS		
	PostMean	Post SD	PIP	Coef.	SE	p-Value	Coef.	SE	p-Value
Publication bias (α_1)	1.953	N.A.	1.000	1.971	0.115	0.000	1.934	0.241	0.000
Precision (α_0)	6.008	0.866	1.000	5.538	2.667	0.038	5.871	2.712	0.033
Health measures									
Physical health (Ref.)									
General health	0.007	0.001	**1.000**	0.006	0.001	0.000	0.007	0.003	0.018
Mental health	–0.001	0.001	**0.530**	–0.002	0.001	0.003	–0.002	0.001	0.125
Self-reported	0.006	0.001	**1.000**	0.006	0.002	0.000	0.007	0.002	0.009
Education measures									
Primary (Ref.)									
Secondary	0.000	0.000	0.017	0.002	0.001	0.175			
Tertiary	0.000	0.000	0.015	0.002	0.001	0.116			
Years of education	0.002	0.002	0.472	0.005	0.002	0.004			
No. of educ. vars	–0.0001	0.000	**0.612**	–0.001	0.000	0.211	–0.001	0.000	0.077
Sample characteristics									
Subsample (Ref.)									
Whole sample	0.000	0.001	0.183	0.002	0.001	0.061			
Twin_sample	0.002	0.003	0.490	0.006	0.002	0.002			
Data characteristics									
Cross-sectional (Ref.)									
Panel	0.001	0.001	0.284	0.001	0.001	0.169			

(*Continued*)

Table 6. *Continued.*

	BMA			FMA			OLS		
	PostMean	Post SD	PIP	Coef.	SE	*p*-Value	Coef.	SE	*p*-Value
Aggregate data (Ref)									
Individual data	–0.005	0.005	**0.553**	–0.010	0.004	0.009	–0.010	0.011	0.387
Administrative data (Ref.)					0.001	0.001			
Survey data	–0.006	0.001	**1.000**	–0.006	0.001	0.000	–0.005	0.002	0.030
Time span	0.000	0.000	0.138	–0.001	0.000	0.006			
Number of vars	0.000	0.000	0.057	0.001	0.001	0.261			
Estimation methods									
OLS (Ref.)									
Non-linear	–0.008	0.001	**1.000**	–0.011	–0.007	0.001	–0.008	0.003	0.023
FE	–0.020	0.003	**1.000**	–0.021	–0.021	0.004	–0.017	0.006	0.007
TSLS	–0.021	0.001	**1.000**	–0.022	–0.020	0.002	–0.021	0.002	0.000
RD	–0.020	0.001	**1.000**	–0.024	–0.020	0.002	–0.020	0.004	0.000
Reduced form	–0.022	0.001	**1.000**	–0.022	–0.020	0.002	–0.022	0.002	0.000
Calculation of t-statistics									
tNormal (Ref.)									
tCalculatedBypValue	–0.011	0.001	**1.000**	–0.011	0.002	0.000	–0.011	0.002	0.000
tCalculatedByCI	0.000	0.000	0.034	–0.003	0.002	0.070			
SE spherical (Ref.)									
SE non-spherical	–0.001	0.002	0.428	–0.004	0.002	0.063	–0.004	0.001	0.000
Publication characteristics									
Publication year	–0.786	0.114	**1.000**	–0.724	0.350	0.039	–0.768	0.357	0.034
Citescore	0.000	0.000	0.025	0.000	0.000	0.046			
Non-economics journal (Ref.)									
Economics journal	–0.001	0.001	**0.593**	–0.002	0.001	0.096	–0.002	0.002	0.342
Regions									
Europe (Ref.)									
North America	0.000	0.000	0.053	–0.001	0.001	0.222			
Asia Pacific	–0.007	0.001	**0.999**	–0.007	0.004	0.059	–0.007	0.002	0.003
Global	0.025	0.003	**1.000**	0.018	0.008	0.028	0.025	0.003	0.000
Income level									
Middle-low income (Ref.)									
High income	0.000	0.000	0.015	0.000	0.001	1.000			

Notes: The results are from the specifications with the weight being the inverse of SE. OLS includes the variables that are above 0.5 in BMA, with robust standard errors clustering at the study level. All variables are described in Table 5. Variables with PIP above 0.5 are emphasized in bold.
SD = standard deviation. SE = standard error. PIP = posterior inclusion probability. N.A. = not available.

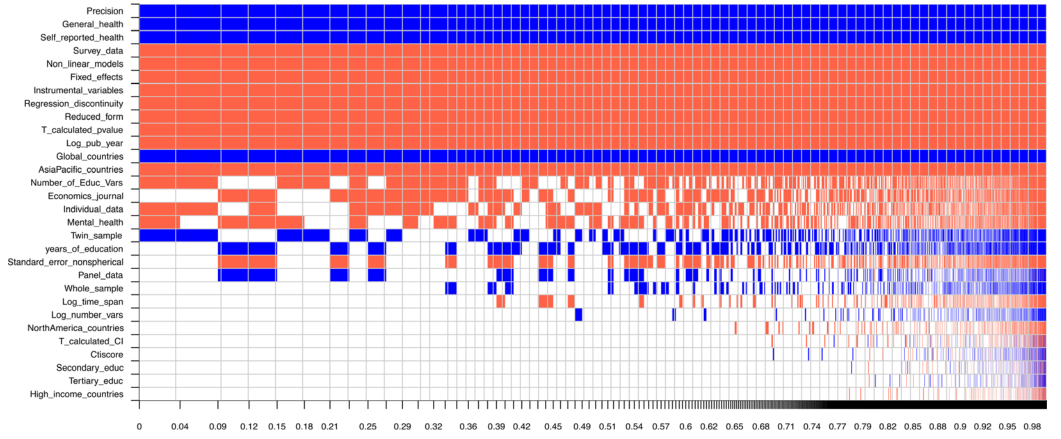

Figure 4. Model Inclusion in Bayesian Model Averaging. [Colour figure can be viewed at wileyonlinelibrary.com]

Notes: The figure depicts the results of Bayesian Model Averaging. The explanatory variables are ranked according to their posterior inclusion probabilities from the highest on the top to the lowest at the bottom. The horizontal axis shows the values of cumulative posterior probability. Blue and red colours denote the positive and negative sign of the estimated parameter of explanatory variable, respectively. No colour means the corresponding explanatory variable is not included in the model. Numerical results are reported in Table 6. All variables are described in Table 5. The results are based on the specifications weighted by the inverse of SE

We find that data characteristics make a difference for the estimated effects. Studies using individual data report smaller effects of education on health than studies using aggregate data. This is perhaps because statistical results using aggregated data do not necessarily reflect the underlying individual behavioural relationships and leads to ‘aggregation bias’ (Kenneday, 2008, p. 506). At the same time, survey data seems to generate larger estimated effects than administrative data. It is plausible that administrative data has fewer problems with attrition, non-response, and measurement error, while survey data is more likely to suffer from reporting bias (Card *et al*., 2010).

The overall results for estimation methods show that non-linear models are associated with reporting lower estimated effects. Whether the studies address endogeneity or not systematically affects estimates. The results consistently show that controlling for endogeneity will significantly reduce the estimated effects by 0.02.

Concerning the calculation of *t*-statistics, the negative coefficient of *tCalculatedbypValue* indicates that the estimated effect of education on health tend to be smaller if the *t*-statistics is calculated by *p*-value. One possible concern is that the mean effect may be downward biased by including estimated effects using the *t*-statistics calculated by *p*-value. Table 6 shows that the coefficient is –0.01. However, after accounting for this effect, the mean effect size is still below the threshold value which Doucouliagos (2011) identifies as ‘small’.

As for publication characteristics, we find that the estimated effects of education on health tend to diminish over time. This coincides with the results of the recent studies uncovering that education has no effects on health. Moreover, compared to non-economics journal, economics journals seem to report smaller effects of education on health.

Our results show that the health effects of education are dependent on countries or regions. We find that education has a smaller health benefits in Asia Pacific countries than in Europe, but have larger effects in cross-country studies. This finding is consistent with several studies in developing countries and in Asia (Lowry and Xie, 2009; Luo *et al*., 2015). The plausible explanation is that in a collectivity-oriented Asian society like China, health-related decisions are usually made by family rather than individual person. Therefore, individual education levels may play a smaller role in determining health behaviours and health care decisions.

To check the robustness of Table 6, we perform BMA analysis of the data with no weights, and with weight being the inverse of number of estimates per study. The results are reported in Table 7. They generally support the main conclusion of Table 6.

Finally, we use the results of BMA to obtain a predicted estimate of the effect of education assuming 'best study' characteristics. We *a priori* decide on a set of characteristics that constitute a 'best study'; that is, a study that is ideally designed to reliably estimate the effect of education on health. We determined these to be: a study that uses physical health for the dependent variable, uses actual years of education as its educational measure, samples from the whole population, is based on individual data that follows subjects over time (panel data), employs a regression discontinuity approach, calculates *t*-statistics in the 'normal' manner of estimate over standard error, allows non-spherical errors, uses data from high income, North American countries and is published in economics journals. Sample mean values were assumed for number of education variables and total number of explanatory variables in the regression, time span, publication year, and Citescore. SE was set equal to zero (thus no publication bias). The associated prediction, which represents the model-weighted average across the models estimated using BMA, was –0.0015, with a standard error of 0.0351. Thus, even using 'best study' characteristics, we find no evidence that education is positively associated with better health.

6. Conclusion

This study provides a meta-analysis on the extensive literature that examines the impact of education on health. Our final sample consists of 4866 estimates from 99 studies. To the best of our knowledge, this paper represents the most comprehensive analysis on the effects of education on health up to date, both in terms of sample size, measures of health, and methods to correct for publication bias. We complement the previous reviews of the literature (Furnee *et al*., 2008; Hamad *et al*., 2018; Galama *et al*., 2020) with a formal treatment of publication bias and heterogeneity in the reported estimates on the effects of education on health. We find that there is moderate publication bias favouring the positive impacts of education on health. In particular, studies ignoring endogeneity are prone to overstate the effects of education on health. The heterogeneity in reported estimates can be largely explained by whether the econometric models control for the endogeneity of education, the types of data, and differences in measures of health.

Our main finding is that the economic effect of education on health is practically zero after correcting for publication bias. Our results echo a number of recent studies by Clark and Royer (2013), Meghir *et al*., (2018), Albarrán *et al*., (2020) and Avenado *et al*. (2020) that report that education plays no role in improving health. Our study also implies that the theory of demand for health capital that assumes a positive role of education deserves to be reassessed.

Education has been proposed as an important health policy initiative in countries including the USA and UK (Clark and Royer, 2013). Similar policy initiatives appear regularly in international organizations such as the OECD and WHO (OECD, 2010; WHO, 2015). However, the findings of this study cast doubt on the feasibility of policies designed to improve health through education interventions.

It should be noted that our study is subject to several limitations. First, we didn't investigate spillover or intergenerational effects of education on health. This doesn't necessarily mean that education has no such effects. Previous studies show that mother's education plays an important role in shaping infant health (Currie and Moretti, 2003) and children's education also affects parents' health (Ma, 2019). Second, we

Table 7. Robustness Checks.

	BMA-unweighted			BMA-weighted by number of estimates per study		
	Post mean	Post SD	PIP	Post mean	Post SD	PIP
Publication bias(α_1)	1.51	0.06	1.00	1.990	0.059	1.000
Precision(α_0)	12.26	N.A.	1.00	0.000	N.A.	1.000
Health measures						
Physical health (Ref.)						
General health	0.02	0.00	**1.00**	0.019	0.001	**1.000**
Mental health	0.01	0.00	**1.00**	0.008	0.002	**0.985**
Self-reported	0.00	0.00	0.12	0.000	0.000	0.015
Education measures						
Primary (Ref.)						
Secondary	–0.01	0.00	**0.94**	–0.015	0.002	**1.000**
Tertiary	–0.01	0.00	**0.97**	–0.024	0.002	**1.000**
Years of education	0.00	0.00	0.07	0.000	0.001	0.040
No. of educ. vars	0.00	0.00	0.01	0.000	0.000	0.044
Sample characteristics						
Subsample (Ref.)						
Whole sample	0.00	0.00	0.23	0.005	0.001	**0.972**
Twin sample	–0.01	0.00	**1.00**	–0.017	0.002	**1.000**
Data characteristics						
Cross-sectional (Ref.)						
Panel	0.00	0.00	0.02	0.000	0.001	0.137
Aggregate data (Ref)						
Individual data	–0.03	0.00	**1.00**	–0.066	0.003	**1.000**
Administrative data (Ref.)						
Survey data	–0.01	0.00	**1.00**	0.000	0.000	0.014
Time span	0.00	0.00	0.13	0.003	0.001	**1.000**

(*Continued*)

Table 7. *Continued.*

	BMA-unweighted			BMA-weighted by number of estimates per study		
	Post mean	Post SD	PIP	Post mean	Post SD	PIP
Number of variables	0.00	0.00	**1.00**	0.000	0.000	0.014
Estimation methods						
OLS (Ref.)						
Non-linear	–0.03	0.00	**1.00**	–0.024	0.002	**1.000**
FE	–0.04	0.00	**1.00**	–0.036	0.003	**1.000**
TSLS	–0.05	0.00	**1.00**	–0.053	0.002	**1.000**
Regression discontinuity	–0.05	0.00	**1.00**	–0.057	0.003	**1.000**
Reduced form	–0.05	0.00	**1.00**	–0.046	0.003	**1.000**
Calculation of t-statistics						
tNormal (Ref.)						
tCalculatedBypValue	–0.01	0.00	**1.00**	–0.015	0.002	**1.000**
tCalculatedByCI	0.00	0.00	0.02	0.000	0.000	0.015
SE spherical (Ref.)						
SE non-spherical	0.00	0.00	0.05	0.000	0.000	0.019
Publication characteristics						
Publication year	–1.60	0.29	**1.00**	0.012	0.001	**1.000**
Citescore	0.00	0.00	0.03	0.000	0.000	0.017
Non-economics journal (Ref.)						
Economic journal	0.00	0.00	0.02	0.000	0.000	0.025
Regions						
Europe (Ref.)						
North America	0.00	0.00	**0.87**	0.011	0.002	**1.000**
Asia Pacific	–0.01	0.00	**0.99**	–0.001	0.003	0.222
Global	0.05	0.01	**1.00**	0.054	0.004	**1.000**
Income level						
Middle-low income (Ref.)						
High income	0.00	0.00	0.17	0.000	0.001	0.046

Notes: The results are based on the specifications with no weight and with weight being the inverse of number of estimates per study. All variables are described in Table 5. Variables with PIP above 0.5 are emphasized in bold
SD = standard deviation. SE = standard error. PIP = posterior inclusion probability.N.A.=not available.

failed to examine the effects of educational quality on health. Pischke and Vonwachter (2008) suggest that compulsory education reforms may increase years of education, but not the quality of education. As quantity and quality of education may face tradeoffs, the investigation on the health effects of education quality may reveal a different picture. Third, although we find that education generates no measurable benefits to health, there is still much we need to learn about which factors are instrumental for health outcomes. It is hoped that that this study will stimulate more research in this regard.

Acknowledgements

We acknowledge financial support from Ministry of Education Humanities and Social Sciences Fund of China (Grant No.20YJC630011), and the Youth Innovation Research Project from Hubei Province (Grant No. T201932). We thank the editor Tomas Havranek and four anonymous referees for their excellent comments. We also thank W. Robert Reed for his helpful suggestions. All errors are our own.

Supplementary file: The data used in the paper is publicly available at Open Science Framework(OSF). It can be accessed at:osf.io/cgyeb/

Notes

1. Another closely related literature is Galama et al. (2020). They review the experimental and quasi-experimental evidence on the effects of education on health and mortality. However, they don't conduct a formal meta-analysis.
2. For p-values, we use the TINV function in Excel to calculate the t-statistics (Stanley and Doucouliagos, 2012). There are some studies only reporting the levels of statistical significance with ***, ** and *. We follow the rule of thumb to assume that p-value is 0.01 for ***, 0.05 for **, 0.1 for * and 0.5 for no star. For 95% confidence intervals, the standard error in OLS is calculated by $SE_i = \frac{upperbound - lowerbound}{2\times 1.96}$. For some studies employing non-linear estimations (e.g. Probit/logit, cox proportional hazard model) and reporting only the odds ratio (OR) with standard error or 95% confidence intervals, we calculate t-statistics by $t_i = \frac{\ln(\hat{\beta}_{1i})\cdot\widehat{\beta}_{1i}}{SE_i}$ or $t_i = \frac{\ln(\hat{\beta}_{1i})\cdot\widehat{\beta}_{1i}}{SE_i}$, where $SE_i = \frac{\ln(upperbound_i) - \ln(lowerbound_i)}{2\times 1.96}$.
3. We exclude 14 papers with interaction or quadratic terms from our data set. Consider the following two specifications: (1) $\mathrm{H} = \beta_0 + \beta_1 Educ + \beta_2(Educ \times Z) + error$; (2) $\mathrm{H} = \beta_0 + \beta_1 Educ + \beta_2 Educ^2 + error$. The associated partial effects are given by $\partial H/\partial Educ = \beta_1 + \beta_2 Z$ in (1) and $\partial H/\partial Educ = \beta_1 + \beta_2 Educ$ in (2). It is obvious that neither produces the real partial effects of education on health.
4. As discussed by Stanley & Doucouliagos (2012. p13): 'in a large, mature and well-established literature, exclusion of unpublished studies is unlikely to affect the results'. Our final dataset contains 99 published paper, which is a large number in meta-analysis.
5. A noteworthy point is that some studies report the estimated effects of higher education on bad health (e.g. mortality, disease, depression). In this case, we transform the coefficients by multiplying –1. The same procedure applies when the studies estimate the effect of lower education on good health.
6. Estimation methods addressing endogeneity of education include fixed effects, instrumental variables, regression discontinuity and reduced form. Section 5 gives further discussion and details.
7. We compute the wild-bootstrap confidence intervals using the *boottest* program developed by Roodman *et al*. (2019).
8. We only report the results of FAT-PET in Appendix C because PEESE and other modern approaches yield similar outcomes. The results are available upon request. We take the same rule in Appendix D.
9. https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf

10. BMA analysis is implemented by *bms* package in R, which is developed by Feldkircher and Zeugner (2009).

References

Albarran, P., Hidalgo-Hidalgo, M. and Iturbe-Ormaetxe, I. (2020) Education and adult health: is there a causal effect? *Social Science & Medicine* 249: 112830.

Amin, S.M. and Parmeter, C.F. (2012) Comparison of model averaging techniques: assessing growth determinants. *Journal of Applied Econometrics* 27: 870–876.

Amin, V., Behrman, J.R. and Spector, T.D. (2013) Does more schooling improve health outcomes and health related behaviors? Evidence from U.K. twins. *Economics of Education Review* 35: 134–148.

Andrews, I. and Kasy, M. (2019) Identification of and correction for publication bias. *American Economic Review* 109(8): 2766–2794.

Avendano, M., de Coulon, A. and Nafilyan, V. (2020) Does longer compulsory schooling affect mental health? Evidence from a British reform. *Journal of Public Economics* 183: 104137.

Albouy, V. and Lequien, L. (2009) Does compulsory education lower mortality? *Journal of Health Economics* 28: 155–168.

Behrman, J.R. and Rosenzweig, M.R. (2004) Returns to birthweight. *Review of Economics and Statistics* 86(2): 586–601.

Behrman, J.R., Kohler, H.-P., Jensen, V.M., Pedersen, D., Petersen, I., Bingley, P. and Christensen, K. (2011) Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography* 48(4): 1347–1375.

Bijwaard, G.E., Kippersluis, H. and Veenman, J. (2015) Education and health: the role of cognitive ability. *Journal of Health Economics* 42:29–43.

Black, D., Hsu, Y.C. and Taylor, L.J. (2015) The effect of early-life education on later-life mortality. *Journal of Health Economics* 44: 1–9.

Bom, P.R. and Rachinger, H. (2019) A kinked meta-regression model for publication bias correction. *Research Synthesis Methods* 10: 497–514.

Braakmann, N. (2011) The causal relationship between education, health and health related behavior: evidence from a natural experiment in England. *Journal of Health Economics* 30: 753–763.

Brunello, G., Fabbri, D. and Fort, M. (2013) The causal effect of education on body mass: evidence from Europe. *Journal of Labor Economics* 31(1): 195–223.

Cameron, A.C., Gelbach, J.B. and Miller, D.L. (2008) Bootstrap-based improvements for inference with clustered errors. *Review of Economics & Statistics* 90(3): 414–427.

Card, D.E. (1999) The causal effect of education on earnings. In O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics*, Vol. 3A* (pp. 1801–1863). Amsterdam: Elsevier.

Card, D., Chetty, R., Feldstein, M. and Saez, E. (2010) Expanding Access to administrative data for research in the United States. http://www.rajchetty.com/chettyfiles/NSFdataaccess.pdf, Accessed 24 August, 2020.

Cazachevici, A., Havranek, T. and Horvath, R. (2020) Remittances and economic growth: a meta-analysis. *World Development* 134: 105021.

Clark, D. and Royer, H. (2013) The effects of education on adult mortality and health: evidence from Britain. *American Economic Review* 103(6): 2087–2120.

Crespo, L., López-Noval, B. and Mira, P. (2014) Compulsory schooling, education, depression and memory: new evidence from SHARELIFE. *Economics of Education Review* 43: 36–46.

Currie, J. and Moretti, E. (2003) Mother's education and the intergenerational transmission of human capital: evidence from college openings. *Quarterly Journal of Economics* 4: 1495–1532.

Cutler, D.M. and Lleras-Muney, A. (2006) Education and health: evaluating theories and evidence. NBER Working Paper 12352.

Cutler, D.M. and Lleras-Muney, A. (2012) Education and health: insights from international comparisons. NBER Working Paper 17738.

Dahmann, S.C. and Schnitzlein, D.D. (2019) No evidence for a protective effect of education on mental health. *Social Science & Medicine* 241: 112584.

Doucouliagos, H. (2011) How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Economics Series 2011/5, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Doucouliagos, H. and Stanlley, T.D. (2013) Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys* 27(2):316–339.

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315: 629–634.

Eicher, T.S., Papageorgiou, C. and Raftery, A.E. (2011) Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics* 26: 30–55.

Feldkircher, M. and Zeugner, S. (2009) Benchmark priors revisited: on adaptive shrinkage and the supermodel effect in Bayesian model averaging. IMF Working Papers/9/202.

Fletcher, J.M. (2015) New evidence of the effects of education on health in the US: compulsory schooling laws revisited. *Social Science & Medicine* 127: 101–107.

Fuchs, V.R. (1982) Time preference and health: an exploratory study. In Fuchs, V. (ed.), *Economic Aspects of Health*. Chicago: The University of Chicago Press.

Furnee, C.A., Groot, W. and van den Brink, H. (2008) The health effects of education: a meta-analysis. *European Journal of Public Health* 18(4): 417–421.

Furukawa, C. (2019) Publication bias under aggregation frictions: theory, evidence, and a new correction method. MIT working paper.

Galama, T.J., Lleras-Muney, A. and van Kippersluis, H. (2020) The effects of education on health and mortality: a review of experimental and quasi-experimental evidence. Andrews M. Jones (ed.), In *The Oxford Encyclopedia of Health Economics*. UK: Oxford University Press.

Gathmann, C., Jürges, H. and Reinhold, S. (2015) Compulsory schooling reforms, education and mortality in Twentieth Century Europe. *Social Science & Medicine* 127: 74–82.

Galama, T.J., Lleras-Muney, A. and van Kippersluis, H. (2020) The effects of education on health and mortality: a review of experimental and quasi-experimental evidence. In Andrews M. Jones (ed.), *The Oxford Encyclopedia of Health Economics: 3-Volume Set*. UK: Oxford University Press.

Grossman, M. (1972) On the concept of health capital and the demand for health. *Journal of Political Economy* 80: 223–255.

Grossman, M. and Kaestner, R. (1997) Effects of education on health. In J. R. Behrman and N. Stacey (eds.), *The Social Benefits of Education*. Ann Arbor: University of Michigan Press.

Grossman, M. (2006) Education and nonmarket outcomes. In E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Vol. 1 (pp. 577–633). Amsterdam: North-Holland.

Gunby, P., Jin, Y. and Reed, W.R. (2017) Did FDI really cause Chinese economic growth? A meta-analysis. *World Development* 90: 242–255.

Hamad, R., Elserb, H., Tranc, D.C., Rehkopfc, D.H. and Goodman, S.N. (2018) How and why studies disagree about the effects of education on health: a systematic review and meta-analysis of studies of compulsory schooling laws. *Social Science & Medicine* 212: 168–178.

Hansen, B. (2007) Least squares model averaging. *Econometrica* 75: 1175–1189.

Havranek, T., Rusnak, M. and Sokolova, A. (2017) Habit formation in consumption: a meta-analysis. *European Economic Review* 95(C): 142–167.

Havranek, T., Irsova, Z. and Zeynalova, O. (2018) Tuition fees and university enrolment: a meta-regression analysis. *Oxford Bulletin of Economics and Statistics* 80: 1145–1184.

Havránek, T., Stanley, T.D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W.R., Rost, K. and van Aert, R.C.M. (2020) Reporting guidelines for meta-analysis in Economics. *Journal of Economic Surveys* 34: 469–475.

Havranek, T. and Sokolova, A. (2020) Do consumers really follow a rule of thumb? Three thousand estimates from 144 studies say 'Probably Not'. *Review of Economic Dynamics* 35: 97–122.

Janke, K., Johnston, D.W., Propper, C. and Shields, M.A. (2020) The causal effect of education on chronic health conditions in the UK. *Journal of Health Economics* 70: 102252.

James, J. (2015) Health and education expansion. *Economics of Education Review* 49: 193–215.

Johnston, D.W., Propper, C. and Shields, M.A. (2009) Comparing subjective and objective measures of health: evidence from hypertension for the income/health gradient. *Journal of Health Economics* 28: 540–552.

Kemptner, D., Jürges, H. and Reinhold, S. (2011) Changes in compulsory schooling and the causal effect of education on health: evidence from Germany. *Journal of Health Economics* 30: 340–354.

Kennedy, P. (2008) *A Guide to Econometrics* (6th edition). Malden, MA: Blackwell Publishing.

Lipsey, M.W. and Wilson, D.B. (2001) *Practical Meta-Analysis*, Vol. 49. Thousand Oaks; London; New Delhi: Sage Publications.

Lleras-Muney, A. (2005) The relationship between education and adult mortality in the United States. *Review of Economic Studies* 72(1): 189–221.

Lowry, D. and Xie, Y. (2009) *Socioeconomic Status and Health Differentials in China: Convergence or Divergence at Older Ages?* Ann Arbor, MI: Population Studies Center, University of Michigan.

Lundborg, P. (2013) The health returns to schooling – what can we learn from twins? *Journal of Population Economics* 26(2): 673–701.

Lundborg, P., Lyttkens, C.H. and Nystedt, P. (2016) The effect of schooling on mortality: new evidence from 50,000 Swedish Twins. *Demography* 53(4): 1135–1168.

Luo, Y., Zhang, Z. and Gu, D. (2015) Education and mortality among older adults in China. *Social Science & Medicine* 127: 134–142.

Ma, M.M. (2019) Does children's education matter for parents' health and cognition? Evidence from China. *Journal of Health Economics* 66: 222–240.

Madsen, M., Andersen, A.M., Christensen, K., Andersen, P.K. and Osler, M. (2010) Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology* 172(2): 225–234.

Mazumder, B. (2008) Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives* 32(2): 2–16.

Mazzonna, F. (2014). The long-lasting effects of education on old age health: evidence of gender differences. *Social Science & Medicine* 101: 129–138.

Meghir, C., Palme, M. and Simeonova, M. (2018) Education and mortality: evidence from a social experiment. *American Economic Journal: Applied Economics* 10(2): 234–256.

OECD.(2010) *Improve Health and Social Cohesion through Education*. OECD: Paris.

Oreopoulos, P. (2006) Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* 96: 152–175.

Oreopoulos, P. (2007) Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics* 91: 2213–2229.

Pischke, J. and VonWachter, T. (2008) Zero returns to compulsory schooling in Germany: evidence and interpretation. *Review Economics & Statistics* 90(3): 592–598.

Rosenthal, R. (1979) The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86(3): 638–641.

Rosenzweig, M.R. and Schultz, T.P. (1989) Schooling, information and nonmarket productivity: contraceptive use and its effectiveness. *International Economic Review* 30: 457–477.

Roodman, D, MacKinnon, J., Nielsen, M. and Webb, M. (2019) Fast and wild: bootstrap inference in Stata using boottest. *Stata Journal* 19(1): 4–60.

Silles, M.A. (2009) The causal effect of education on health: evidence from the United Kingdom. *Economics of Education Review* 28(1): 122–128.

Stanley, T.D. (2005) Beyond publication bias. *Journal of Economic Surveys* 19(3):309–345.

Stanley, T.D. (2008) Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* 70(1): 103–127.

Stanley, T.D. and Doucouliagos, H. (2012) *Meta-Regression Analysis in Economics and Business*. Abingdon: Routledge.

Stanley, T.D. and Doucouliagos, H. (2014) Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods* 5: 60–78.

Stanley, T.D., Doucouliagos, H. and Ioannidis, J.P. (2017) Finding the power to reduce publication bias. *Statistics in Medicine* 36: 1580–1598.

Steel, M.F. (2020) Model averaging and its use in economics. *Journal of Economic Literature* 58(3): 644–719.

Van Kippersluis, H., O'Donnell, O. and Van Doorslaer, E. (2011) Long-run returns to education: does schooling lead to an extended old age? *Journal of Human Resources* 46(4): 695–721.

Viechtbauer, W. and Cheung, M.W.-L. (2010) Outlier and influence diagnostics for meta-analysis. *Research on Synthesis Method* 1: 112–125.

Webbink, D., Martin, N.G. and Visscher, P.M. (2010) Does education reduce the probability of being overweight? *Journal of Health Economics* 29: 29–38.

WHO. (1946). *Preamble to the Constitution of the World Health Organization as adopted by the International Health conference*, New York, 19–22 June, 1946. New York: WHO.

WHO. (2015) *Health 2020: Education and Health through the Life-Course*. WHO: Geneva.

Xue, X.D., Reed, W.R. and Menclova, A. (2020) Social capital and health: a meta-analysis. *Journal of Health Economics* 72: 102317.

Zigraiova, D. and Havranek, T. (2016) Bank competition and financial stability: much ado about nothing? *Journal of Economic Surveys* 30(5): 944–981.

Appendix A: List of Studies Included in the Sample

1. Adams, S. J. (2002). Educational attainment and health: evidence from a sample of older adults. *Education Economics*, 10(1), 97–109.
2. Albarran, P., Hidalgo-Hidalgo, M., Iturbe-Ormaetxe, I. (2020). Education and adult health: is there a causal effect? *Social Science & Medicine*, 249, 112830.
3. Albouy, V., Lequien, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics*, 28(1), 155–168.
4. Amin, V., Behrman, J. R., Kohler, H. P. (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. *Social Science & Medicine*, 127, 181–189.
5. Amin, V., Behrman, J. R., Spector, T. D. (2013). Does more schooling improve health outcomes and health related behaviors? Evidence from U.K. twins. *Economics of Education Review*, 35, 134–148.
6. Amin, V., Dunn, P., Spector, T. (2018). Does education attenuate the genetic risk of obesity? Evidence from U.K. Twins. *Economics & Human Biology*, 31, 200–208.
7. Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review*, 24(2), 149–160.
8. Atella, V., Kopinska, J. (2014). Body weight, eating patterns, and physical activity: the role of education. *Demography*, 51(4), 1225–1249.
9. Auld, M. C., Sidhu, N. (2005). Schooling, cognitive ability and health. *Health Economics*, 14(10), 1019–1034.
10. Avendano, M., de Coulon, A., Nafilyan, V. (2020). Does longer compulsory schooling affect mental health? Evidence from a British reform. *Journal of Public Economics*, 183, 104137.
11. Bai, Y., & Li, Y. (2018). A sibling-pair analysis for causal effect of education on health. *The B.E. Journal of Economic Analysis & Policy*, 18(4).
12. Baltagi, B. H., Flores-Lagunes, A., Karatas, H. M. (2019). The effect of education on health: evidence from the 1997 compulsory schooling reform in Turkey. *Regional Science and Urban Economics*, 77, 205–221.
13. Banks, J., Mazzonna, F. (2012). The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *Economic Journal*, 122(560), 418–448.
14. Becchetti, L., Conzo, P., Pisani, F. (2017). Education and health in Europe. *Applied Economics*, 50(12), 1362–1377.
15. Behrman, J. R., Kohler, H. P., Jensen, V. M., Pedersen, D., Petersen, I., Bingley, P., Christensen, K. (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography*, 48(4), 1347–1375.

16. Behrman, J. R., Xiong, Y., Zhang, J. (2015). Cross-sectional schooling-health associations misrepresented causal schooling effects on adult health and health-related behaviours: evidence from the Chinese Adults Twins Survey. *Social Science & Medicine*, 127, 190–197.
17. Benson, R., von Hippel, P. T., Lynch, J. L. (2018). Does more education cause lower BMI, or do lower-BMI individuals become more educated? Evidence from the National Longitudinal Survey of Youth 1979. *Social Science & Medicine*, 211, 370–377.
18. Berger, M. C., Leigh, J. P. (1989). Schooling, self-selection, and health. *The Journal of Human Resources*, 24(3), 433–455.
19. Bijwaard, G. E., Myrskyla, M., Tynelius, P., Rasmussen, F. (2017). Educational gains in cause-specific mortality: accounting for cognitive ability and family-level confounders using propensity score weighting. *Social Science & Medicine*, 184, 49–56.
20. Black, D. A., Hsu, Y. C., Taylor, L. J. (2015). The effect of early-life education on later-life mortality. *Journal of Health Economics*, 44, 1–9.
21. Böckerman, P., Maczulskij, T. (2016). The education-health nexus: fact and fiction. *Social Science & Medicine*, 150, 112–116.
22. Böckerman, P., Viinikainen, J., Pulkki-Raback, L., Hakulinen, C., Pitkanen, N., Lehtimaki, T., Raitakari, O. T. (2017). Does higher education protect against obesity? Evidence using Mendelian randomization. *Preventive Medicine*, 101, 195–198.
23. Braakmann, N. (2011). The causal relationship between education, health and health related behaviour: evidence from a natural experiment in England. *Journal of Health Economics*, 30(4), 753–763.
24. Brunello, G., Fort, M., Schneeweis, N., Winter-Ebmer, R. (2016). The causal effect of education on health: what is the role of health behaviours? *Health Economics*, 25(3), 314–336.
25. Brunello, G., Fabbri, D., Fort, M. (2013). The causal effect of education on body mass: evidence from Europe. *Journal of Labor Economics*, 31(1), 195-223.
26. Buckles, K., Hagemann, A., Malamud, O., Morrill, M., Wozniak, A. (2016). The effect of college education on mortality. *Journal of Health Economics*, 50, 99–114.
27. Catherine, E. R., Chia-ling, W. (1995). The links between education and health. *American Sociological Review*, 60(5), 719–745.
28. Catherine, E. R., Mirowsky, J. (1999). Refining the association between education and health: the effects of quantity, credential, and selectivity. *Demography*, 36(4), 445–460.
29. Clark, D., Royer, H. (2013). The effect of education on adult mortality and health: evidence from Britain. *American Economic Review*, 103(6), 2087–2120.
30. Cook, A. (2018). The education–suicide mortality gradient. *Applied Economics Letters*, 26(9), 717–721.
31. Courtin, E., Nafilyan, V., Avendano, M., Meneton, P., Berkman, L. F., Goldberg, M., … Dowd, J. B. (2019). Longer schooling but not better off? A quasi-experimental study of the effect of compulsory schooling on biomarkers in France. *Social Science & Medicine*, 220, 379–386.
32. Courtin, E., Nafilyan, V., Glymour, M., Goldberg, M., Berr, C., Berkman, L. F., Avendano, M. (2019). Long-term effects of compulsory schooling on physical, mental and cognitive ageing: a natural experiment. *Journal of Epidemiology & Community Health*, 73(4), 370–376.
33. Crespo, L., López-Noval, B., & Mira, P. (2014). Compulsory schooling, education, depression and memory: new evidence from SHARELIFE. *Economics of Education Review*, 43, 36–46.
34. Cutler, D. M., Lange, F., Meara, E., Richards-Shubik, S., Ruhm, C. J. (2011). Rising educational gradients in mortality: the role of behavioral risk factors. *Journal of Health Economics*, 30(6), 1174–1187.
35. Dahmann, S. C., Schnitzlein, D. D. (2019). No evidence for a protective effect of education on mental health. *Social Science & Medicine*, 241, 112584.

36. Denney, J. T., Rogers, R. G., Hummer, R. A., Pampel, F. C. (2010). Education inequality in mortality: the age and gender specific mediating effects of cigarette smoking. *Social Science Research*, 39(4), 662–673.
37. Devaux, M., Sassi, F., Church, J., Cecchini, M., Borgonovi, F. (2011). Exploring the relationship between education and obesity. *OECD Journal: Economic Studies*, 2011(1), 1–40.
38. Dursun, B., Cesur, R., Mocan, N. (2018). The impact of education on health outcomes and behaviors in a middle-income, low-education country. *Economics & Human Biology*, 31, 94–114.
39. Edwards, R. D. (2015). Health, SES, and the timing of education among military retirees. *Education Economics*, 24(4), 393–410.
40. Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: compulsory schooling laws revisited. *Social Science & Medicine*, 127, 101–107.
41. Fonseca, R., Michaud, P. C., & Zheng, Y. (2020). The effect of education on health: evidence from national compulsory schooling reforms. *SERIEs*, 11(1), 83–103.
42. Fujiwara, T., Kawachi, I. (2009). Is education causally related to better health? A twin fixed-effect study in the USA. *International Journal of Epidemiology*, 38(5), 1310–1322.
43. Gerdtham, U. G., Lundborg, P., Lyttkens, C. H., Nystedt, P. (2016). Do education and income really explain inequalities in health? Applying a twin design. *The Scandinavian Journal of Economics*, 118(1), 25–48.
44. Glymour, M. M., Kawachi, I., Jencks, C. S., Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health*, 62(6), 532–537.
45. Grimard, F., Laszlo, S., Lim, W. (2010). Health, aging and childhood socio-economic conditions in Mexico. *Journal of Health Economics*, 29(5), 630–640.
46. Groot, W., Maassen van den Brink, H. (2007). The health effects of education. *Economics of Education Review*, 26(2), 186–200.
47. Grytten, J., Skau, I., Sorensen, R. (2020). Who dies early? Education, mortality and causes of death in Norway. *Social Science & Medicine*, 245, 112601.
48. Halpern-Manners, A., Schnabel, L., Hernandez, E. M., Silberg, J. L., Eaves, L. J. (2016). The relationship between education and mental health: new evidence from a discordant twin study. *Social Forces*, 95(1), 107–131.
49. Hardarson, T., Gardarsdóttir, M., Gudmundsson, K. TH., Thorgeirsson, G., Sigvaldason, H., Sigfússon, N. (2001). The relationship between educational level and mortality. The Reykjavík Study. *Journal of Internal Medicine*, 249, 495–502.
50. Hartog, J., Oosterbeek, H. (1998). Health, wealth and happiness: why pursue a higher education? *Economics of Education Review*, 17(3), 245–256.
51. Herd, P. (2010). Education and health in late-life among high school graduates: cognitive versus. *Journal of Health and Social Behavior*, 51(4), 478–496.
52. Hirokawa, K. Tsutusmi, A., Kayaba, K. (2006). Impacts of educational level and employment status on mortality for Japanese women and men: The Jichi Medical School Cohort Study. *European Journal of Epidemiology*, 21(9), 641–651.
53. Huang, W., Zhou, Y. (2013). Effects of education on cognition at older ages: evidence from China's Great Famine. *Social Science & Medicine*, 98, 54–62.
54. James, J. (2015). Health and education expansion. *Economics of Education Review*, 49, 193–215.
55. Jamison, E. A., Jamison, D. T., Hanushek, E. A. (2007). The effects of education quality on income growth and mortality decline. *Economics of Education Review*, 26, 772–789.
56. Janke, K., Johnston, D. W., Propper, C., Shields, M. A. (2020). The causal effect of education on chronic health conditions in the UK. *Journal of Health Economics*, 70, 102252.
57. Jürges, H., Kruk, E., Reinhold, S. (2013). The effect of compulsory schooling on health – evidence from biomarkers. *Journal of Population Economics*, 26(2), 645–672.

58. Kemna, H. J. M. I. (1987). Working conditions and the relationship between schooling and health. *Journal of Health Economics*, 6, 189–210.
59. Kemptner, D., Jurges, H., Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: evidence from Germany. *Journal of Health Economics*, 30(2), 340–354.
60. Kim, Y. J. (2016). The long-run effect of education on obesity in the US. *Economics & Human Biology*, 21, 100–109.
61. Kiuila, O., Mieszkowski, P. (2007). The effects of income, education and age on health. *Health Economics*, 16(8), 781–798.
62. Lacroix, G., Laliberte-Auger, F., Michaud, P. C., Parent, D. (2019). The effect of college education on health and mortality: Evidence from Canada. *Health Economics*.
63. Lager, A. C., Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceeding of National Academy of Science (USA)*, 109(22), 8461–8466.
64. Leeves, G., Soyiri, I. (2015). Does more education always lead to better health? Evidence from rural Malaysia. *Biomedical Research International*, 2015, 539212.
65. Leuven, E., Plug, E., Rønning, M. (2016). Education and cancer risk. *Labour Economics*, 43, 106–121.
66. Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies*, 72, 189–221.
67. Lundborg, P. (2013). The health returns to schooling – what can we learn from twins? *Journal of Population Economics*, 26(2), 673–701.
68. Lundborg, P., Lyttkens, C. H., Nystedt, P. (2016). The effect of schooling on mortality: new evidence from 50,000 Swedish twins. *Demography*, 53(4), 1135–1168.
69. Luo, Y., Zhang, Z., Gu, D. (2015). Education and mortality among older adults in China. *Social Science & Medicine*, 127, 134–142.
70. Lutz, W., Kebede, E. B. (2018). Education and health: redrawing the Preston curve. *Population and Development Review*, 00(0), 1–19.
71. Lynch, J. L., von Hippel, P. T. (2016). An education gradient in health, a health gradient in education, or a confounded gradient in both? *Social Science & Medicine*, 154, 18–27.
72. Ma, Y., Nolan, A., Smith, J. P. (2018). The value of education to health: evidence from Ireland. *Economics & Human Biology*, 31, 14–25.
73. Madsen, M., Andersen, A. M., Christensen, K., Andersen, P. K., Osler, M. (2010). Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology*, 172(2), 225–234.
74. Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives*, 32(2), 2–16.
75. Mazzonna, F. (2014). The long-lasting effects of education on old age health: evidence of gender differences. *Social Science & Medicine*, 101, 129–138.
76. Meghir, C., Palme, M., Simeonova, E. (2018). Education and mortality: evidence from a social experiment. *American Economic Journal: Applied Economics*, 10(2), 234–256.
77. Montez, J. K., Zhang, W., Zajacova, A., Hamilton, T. G. (2018). Does college major matter for women's and men's health in midlife? Examining the horizontal dimensions of educational attainment. *Social Science & Medicine*, 198, 130–138.
78. Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1), 152–175.
79. Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, 91(11–12), 2213–2229.

80. Palme, M., Simeonova, E. (2015). Does women's education affect breast cancer risk and survival? Evidence from a population based social experiment in education. *Journal of Health Economics*, 42, 115–124.
81. Parinduri, R. A. (2016). Does education improve health? Evidence from Indonesia. *The Journal of Development Studies*, 53(9), 1358–1375.
82. Powdthavee, N. (2010). Does education reduce the risk of hypertension? Estimating the biomarker effect of compulsory schooling in England. *Journal of Human Capital*, 4(2), 173–202.
83. Rogers, R. G., Hummer, R. A., Everett, B. G. (2013). Educational differentials in US adult mortality: an examination of mediating factors. *Social Science Research*, 42(2), 465–481.
84. Schneeweis, N., Skirbekk, V., Winter-Ebmer, R. (2014). Does education improve cognitive performance four decades after school completion? *Demography*, 51(2), 619–643.
85. Seo, B., Senauer, B. (2011). The effect of education on health among US residents in relation to country of birth. *Health Economics*, 20(1), 45–55.
86. Silles, M. A. (2009). The causal effect of education on health: evidence from the United Kingdom. *Economics of Education Review*, 28(1), 122–128.
87. Sironi, M. (2012). Education and mental health in Europe. *International Journal of Mental Health*, 41(3), 79–105.
88. Spasojević, J. (2010). Effects of education on adult health in Sweden: results from a natural experiment. *Current Issues in Health Economics*, 179–199.
89. Van der Pol, M. (2011). Health, education and time preference. *Health Economics*, 20(8), 917–929.
90. van Kippersluis, H., O'Donnell, O., van Doorslaer, E. (2011). Long-run returns to education does schooling lead to an extended old age? *The Journal of Human Resources*, 46(4), 695–721.
91. Viinikainen, J., Bryson, A., Bockerman, P., Elovainio, M., Pitkanen, N., Pulkki-Raback, L., Pehkonen, J. (2018). Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization. *Preventive Medicine*, 115, 134–139.
92. von dem Knesebeck, O., Verde, P. E., Dragano, N. (2006). Education and health in 22 European countries. *Social Science & Medicine*, 63(5), 1344–1351.
93. Webbink, D., Martin, N. G., Visscher, P. M. (2010). Does education reduce the probability of being overweight? *Journal of Health Economics*, 29(1), 29–38.
94. Weitzman, A. (2017). The effects of women's education on maternal health: evidence from Peru. *Social Science & Medicine*, 180, 1–9.
95. Xie, S., Mo, T. (2014). The impact of education on health in China. *China Economic Review*, 29, 1–18.
96. Zajacova, A., Rogers, R. G., Johnson-Lawrence, V. (2012). Glitch in the gradient: additional education does not uniformly equal better health. *Social Science & Medicine*, 75(11), 2007–2012.
97. Zhang, W., McCubbin, H., McCubbin, L., Chen, Q., Foley, S., Strom, I., Kehl, L. (2010). Education and self-rated health: An individual and neighborhood level analysis of Asian Americans, Hawaiians, and Caucasians in Hawaii. *Social Science & Medicine*, 70(4), 561–569.
98. Zhong, H. (2015). Does a college education cause better health and health behaviors? *Applied Economics*, 47(7), 639–653.
99. Zhong, H. (2016). Effects of quantity of education on health: a regression discontinuity design approach based on the Chinese Cultural Revolution. *China Economic Review*, 41, 62–74.

Appendix B: Measurements of Health and Education

	Estimates	percentage
Health measures		
Physical health		
Mortality	1229	25.26
Obesity	891	18.31
Disease	855	17.57
ADL	303	6.23
Mental health	599	
Depression	369	7.58
Cognition	230	4.73
General health	989	20.32
Total	4866	
Education measures		
Continuous (years of schooling)	2906	59.72
Categorical (education levels)	1960	
Primary	358	7.36
Secondary	1061	21.8
Tertiary	541	11.12
Total	4866	

Note: Author's calculations.

Appendix C: Robustness Checks: Subsample Analysis

Subsample controlling for endogeneity

	OLS	WLS-FE	WLS-RE	IV
Weight 1: Equal weight to each estimate				
Publication bias (α_1)	1.149*** (0.178)	1.222*** (0.174)	1.139*** (0.174)	1.16*** (0.155)
Precision effect (α_0)	–0.0009 (0.002)	–0.0001* (0.0006)	–0.0008 (0.002)	–0.0009 (0.001)
Adjusted R^2	0.14	0.002	0.24	0.21
Observations	2337	2337	2337	2337
Number of studies	71	71	71	71

.

Subsample controlling for endogeneity				
	OLS	WLS-FE	WLS-RE	IV
Weight 2: Equal weight to each study				
Publication bias (α_1)	1.511*** (0.478)	1.553*** (0.29)	1.561*** (0.536)	1.546*** (0.428)
True effect (α_0)	–0.0004 (0.003)	–0.0007 (0.001)	–0.001 (0.004)	–0.0006 (0.002)
Adjusted R^2	0.123	0.003	0.26	0.28
Observations	2337	2337	2337	2337
Number of studies	71	71	71	71
Subsample not controlling for endogeneity				
	OLS	WLS-FE	WLS-RE	IV
Weight 1: Equal weight to each estimate				
Publication bias (α_1)	2.498*** (0.263)	3.474*** (0.546)	2.776*** (0.31)	3.113*** (0.372)
Precision effect (α_0)	0.012** (0.005)	0.003 (0.002)	0.01** (0.004)	0.005* (0.003)
Adjusted R^2	0.25	0.011	0.55	0.51
Observations	2529	2529	2529	2529
Number of studies	89	89	89	89
Weight 2: Equal weight to each study				
Publication bias (α_1)	2.918*** (0.69)	4.053*** (0.463)	3.03*** (0.55)	3.456*** (0.445)
True effect (α_0)	0.013* (0.007)	0.002 (0.003)	0.01* (0.006)	0.006* (0.003)
Adjusted R^2	0.283	0.006	0.6	0.534
Observations	2529	2529	2529	2529
Number of studies	89	89	89	89

Notes: The table reports the FAT-PET results of regression in equation (2). Robust standard errors clustered at study -level are in parentheses. Wild-bootstrap confidence intervals are not reported because they yield similar statistics with robust clustered standard errors.

OLS = ordinary least squares; WLS-FE = weighted least square-fixed effects; WLS-RE = weighted least square-random effects; IV = instrument variables regression with the inverse of the square root of the degree of freedom used as an instrument.

***Significant at 1%.

**Significant at 5%.

*Significant at 10%.

Appendix D: The Marginal Effects of Years of Education on Mortality

.

	OLS	WLS-FE	WLS-RE	IV
Weight 1: Equal weight to each estimate				
Publication bias (α_1)	–1.974*** (0.03)	–4.292*** (1.144)	–1.393** (0.55)	–3.826*** (1.294)
True effect (α_0)	–0.04 (0.096)	–0.0001 (0.0003)	–0.01 (0.018)	–0.0006 (0.002)
Adjusted R^2	0.569	0.006	0.168	0.022
Observations	497	497	497	497
Number of studies	17	17	17	17
Weight 2: Equal weight to each study				
Publication bias (α_1)	–1.938*** (0.09)	–3.794*** (1.003)	–1.55* (0.846)	–3.437*** (0.838)
True effect (α_0)	0.02 (0.1)	0.00005 (0.0002)	–0.016 (0.018)	0.000008 (0.0002)
Adjusted R^2	0.546	0.007	0.035	0.08
Observations	497	497	497	497
Number of studies	17	17	17	17

Notes: The table shows the FAT-PET results of regression in equation (2). Robust standard errors clustered at study level are in parentheses. Wild-bootstrap confidence intervals are not reported because they yield similar statistics with robust clustered standard errors.

OLS = ordinary least squares; WLS-FE = weighted least square-fixed effects; WLS-RE = weighted least square-random effects; IV = instrument variables regression with the inverse of the square root of the degree of freedom used as an instrument.

***Significant at 1%.
**Significant at 5%.
*Significant at 10%.