

中国教育健康效应

——基于 Meta 回归的研究

彭树宏

[摘要] 源于教育的健康效应研究中存在相互冲突的研究结论的现象, 本文对中英文期刊发表的中国教育的健康效应的实证研究文献进行了 Meta 分析。FAT-PET 检验和 PEESE 检验表明, 已发表的研究文献中存在着显著的发表偏误, 在剔除发表偏误后, 中国教育对健康的真实影响效应接近于零; 多元 Meta 回归模型的 Post Lasso 估计量和 Double Post Lasso 估计量表明, 健康指标选取、教育指标选取以及是否解决内生性是导致研究结论差异的三个重要来源。

[关键词] 教育; 健康; 中国; Meta 分析; 发表偏误

一、引言

教育和健康是人力资本的两种主要表现形式。自 Grossman(1972)建立健康需求理论模型, 将教育作为重要参数纳入其中后, 教育对健康的影响就一直是人力资本领域的一个重要研究话题(Grossman, 2000)。研究教育的健康效应, 不仅能直接检验健康资本需求(Grossman, 1972)、童年发展对成年影响(Heckman, 2007; Conti et al., 2010)以及宏观经济增长(Acemoglu and Johnson, 2007; Cervellati and Sunde, 2005)理论模型的预测假设, 还能为相关的教育和健康政策制定提供决策参考。特别是, 随着现代社会的人们对健康问题的日益关注, 教育作为健康的一种重要的社会性决定因素更加受到人们的关注和重视。

但遗憾的是, 教育的健康效应研究尚未取得一致结论。一些研究发现,

[收稿日期] 2021-08-19

[基金项目] 江西省教育科学规划课题“大学教育对健康的影响效应及其机制研究”(19YB051)。

[作者简介] 彭树宏, 江西财经大学财税与公共管理学院, 电子邮箱地址: shuhong.peng@gmail.com。

教育对健康有显著的正向效应；另一些研究则发现，教育对健康的影响很小，甚至没有影响(Grossman, 2015; Galama, 2018)。即使用相同的健康指标，不同国家的研究结论也不一样。例如，Lleras-Muney(2005)使用美国义务教育法在各州的不同变化做工具变量，研究发现教育对死亡率有很大的显著影响效应。Clark and Royer(2013)利用英国义务教育法改革的自然实验进行断点回归则发现，教育对死亡率没有显著影响。美英之外，教育对死亡率的影响在新西兰(Kippersluis et al., 2011)和德国(Kempton et al., 2011)都得到了证实，但在法国(Albouy and Lequien, 2009)和瑞士(Lager and Torssander, 2012)却没有得到证实。即使是针对中国的研究，相关文献也给出了不同的结论。例如，胡安宁(2010)、程令国等(2014)研究发现，教育对居民健康有显著的正向影响。但这两项研究并没有解决反向因果和遗漏不可观测变量导致的内生性问题。李军和刘生龙(2019)借助中国义务教育法实施的自然实验，采用两阶段最小二乘法检验中国教育与国民健康之间的因果关系，研究发现教育与中国成年男性自报健康、女性保持正常体型之间存在显著因果关系。但郭四维等(2019)同样基于中国义务教育法实施的自然实验，采用模糊断点回归方法考察教育对居民健康水平的影响，结果却表明教育对自评健康和体型均无影响。

虽然研究结论的不一致可能表明教育的健康效应存在异质性，但在模型参数设置和相关政策制定时却需要有一个统一的明确结论。在这种情况下，传统的定性文献综述方法无能为力，因为它存在着多方面的局限性：(1)无法对估计效应给出一个综合性结论；(2)无法系统性考察差异化研究结论的产生原因；(3)文献的选择没有明确的标准，存在主观性的选择偏差；(4)无法识别已有文献中是否存在发表偏误。Meta 分析是一种将多个具有相同研究主题的研究结果进行定量合成并对研究结果间差异的来源进行检验的统计分析方法(Porta, 2016)，可以有效地避免上述传统定性文献综述方法的局限性，已在国外经济学定量综述领域得到广泛的应用(Havranek et al., 2018; Xue et al., 2021)。

针对中国教育的健康效应的研究结论存在异质性的问题，本文运用 Stanley 和 Doucouliagos(2012)等提出的 Meta 回归分析方法来进行定量文献研究。研究结论表明，现已发表的研究文献存在显著的发表偏误；在剔除发表偏误后，中国教育对健康的真实影响效应接近于零；健康指标选取、教育指标选取以及是否解决内生性是导致研究结论差异的三个重要来源。本文的主要贡献在于：(1)首次针对中国教育的健康效应问题采用 Meta 分析方法得出定量合成结论并检验了发表偏误的存在性；(2)首次采用 Lasso 方法解决运

用 Meta 回归方法考察研究异质性来源时的模型选择的不确定性问题,并得到研究结论差异的三个主要来源。

本文的第二部分是研究设计,包括文献检索与筛选标准、数据整理与编码、效应值确定与模型设定;第三部分是 Meta 分析结果,包括合成效应值大小、发表偏误、异质性来源分析;第四部分是研究结论。

二、研究设计

(一)文献检索与筛选

Meta 分析首要的工作是文献搜集。本文将研究对象限定在中国教育的健康效应的实证研究。因此,我们在中国知网中搜索标题或关键词中同时含有“教育”和“健康”(或“死亡率”“肥胖”等)的中文论文;在 Google Scholar、Elsevier Science Direct、Web of Science、Wiley Online Library 中搜索标题或关键词中同时含有“education”(或“schooling”)和“health”(或“mortality”“obesity”等)的英文论文。为了避免遗漏任何重要文献,我们进一步通过“滚雪球”方式查询所搜集文献的参考文献和被引文献,并对照了作者之前在研究此问题上所积累的重要文献。文献搜索截至时间为 2021 年 2 月。

初步检索后,我们通过以下准则筛选所搜集到的文献:(1)只纳入已发表论文,不纳入工作论文;(2)中文为 CSSCI 论文,英文为 SSCI 论文;(3)报告了教育对健康影响的实证结果,排除理论研究、文献综述和仅进行描述性统计研究的论文;(4)以中国全国数据为样本,排除以外国数据或以中国某个地方的数据为样本的研究;(5)排除同一作者已发表中文论文的英文翻译版;(6)考察个体教育与其自身健康的关系,排除教育的代际影响,比如父母教育对小孩健康的影响,小孩教育对父母健康的影响等;(7)报告了估计系数、 t 统计量(或能计算出 t 统计量的标准误、 p 值、星号、置信区间等)和样本量;^①(8)排除设置了教育变量的二次项和交互项的模型;(9)排除考察教育对健康影响机制的回归,如教育对吸烟、酗酒、均衡饮食、锻炼的影响等。

^① 本文所纳入的文献没有只报告置信区间和 p 值的,但有只报告星号的。在兼顾样本量和文献可信性程度的情形下,我们纳入了只报告星号的英文 SSCI 论文中的估计,而没有纳入只报告星号的中文 CSSCI 论文中的估计。后文的 Lasso 估计表明, t 值是否由星号转化计算而得对效应值大小没有影响;进一步,我们也做了剔除全部只报告星号的估计后的 Meta 分析,发现研究结论没有变化。

通过上述筛选准则后纳入本文 Meta 分析的文献共计 22 篇(其中中文文献 15 篇, 英文文献 7 篇), 330 组估计值。纳入文献的列表见本文附录。

(二)数据整理与编码

我们从所纳入的文献中提取了两类研究所需的数据。第一类是每个估计中的 t 统计量、估计系数、标准误、表示显著性水平的星号数量、样本量、变量个数等, 用于计算效应值和检验发表偏误; 第二类是描述研究特征的变量, 包括健康和教育的度量指标选取、出版特征、模型设定和样本特征的变量, 用于考察研究结果异质性的来源。

1. t 统计量和自由度

对于原始回归结果中报告了 t 统计量的, 我们直接提取该 t 统计量的值; 对于原始回归结果中没有报告 t 统计量, 但报告了估计系数和标准误的, 我们将估计系数除以标准误得到 t 统计量; 对于原始回归结果中没有报告 t 统计量和标准误, 而只报告代表显著性水平的 $***$ 、 $**$ 、 $*$ 时, 我们遵照现有 Meta 分析文献的通用规则, 将其分别与 0.01、0.05、0.1 的 p 值对应, 没有 $*$ 的对应 0.5 的 p 值, 然后通过 R 函数 $qt(p/2, n)$ 来计算对应的 t 值, t 值的符号由回归系数的符号确定(Stanley and Doucouliagos, 2012)。

本文所纳入的文献中的教育变量都是正向设定的, 即数值越大代表教育程度越高(以虚拟变量设定时, 1 代表更高的学历水平)。但是, 健康指标的设定既有正向设定(数值越大代表健康水平越高), 也有负向设定(数值越大代表健康水平越低)。如果健康指标是正向设定的, t 统计量就是上述得到的 t 统计量; 如果健康指标是负向设定的, 那么我们取上述 t 统计量的相反数作为最终分析所用的 t 统计量。

从图 1 可以看到, t 统计量的分布右偏, 大多数的 t 值都大于零。有 43.3% 的 t 值大于 2, 51.2% 的 t 值在 $[-2, 2]$ 内, 只有 5.5% 的 t 值小于 -2。这意味着, 大多数研究报告了教育对健康有正向影响, 其中约一半有着显著的正向影响。

我们用 $n - k - 1$ 来计算自由度 d , 其中 n 为样本量, k 为自变量个数。为了减少极端值的影响, 我们将 t 统计量和自由度 d 在其分布上下 5% 分位处进行缩尾, 以缩尾后的 t 统计量和自由度来计算效应值。

2. 研究特征

我们从健康度量指标选取、教育度量指标选取、出版特征、模型设定、样本特征四个方面提取影响研究结果的研究特征变量。相关变量的描述性统计如表 1 所示。

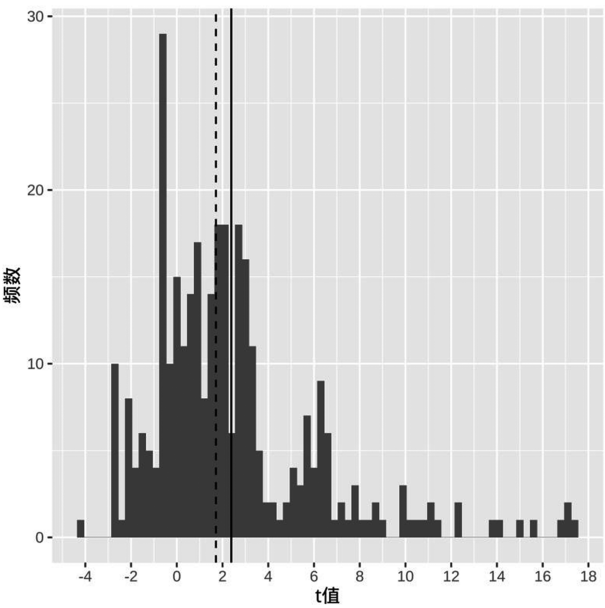


图 1 t 统计量的分布图

注：实的竖直线和虚的竖直线分别代表 t 统计量的均值和中位数。

表 1 研究特征变量的定义和描述性统计

变量	定义	均值	标准差	加权平均值
健康指标：				
生理健康	健康指标以 ADL 或是否患某种疾病来度量 =1；否则=0	0.206	0.405	0.244
认知功能	健康指标以记忆能力、计算能力等认知功能 来度量=1，否则=0	0.103	0.304	0.098
心理健康	健康指标以抑郁量表等心理测量表或相关 问题度量=1，否则=0	0.042	0.202	0.068
体重	健康指标以是否肥胖或体重过低来度量=1， 否则=0	0.191	0.394	0.082
死亡率	健康指标以死亡率来度量=1，否则=0	0.121	0.327	0.061
总体自评健康	健康指标以总体自评健康来度量=1，否则 =0	0.336	0.473	0.446
教育指标：				
小学	教育变量以是否上过小学来度量=1，否则 =0	0.194	0.396	0.159
中学	教育变量以是否上过中学来度量=1，否则 =0	0.161	0.368	0.174
大学	教育变量以是否上过大学来度量=1，否则 =0	0.045	0.209	0.061

续表				
变量	定义	均值	标准差	加权平均值
接受教育	教育变量以是否受过教育来度量=1, 否则=0	0.036	0.187	0.045
教育年限	教育变量以受教育年限来度量=1, 否则=0	0.564	0.497	0.561
出版特征:				
出版年份的对数	论文发表的年份取对数	7.609	0.001	7.609
英文期刊	在英文期刊发表=1, 否则=0	0.636	0.482	0.318
由星号计算得到 t 值	t 统计量由表示显著性水平的星号转化而得=1, 否则=0	0.139	0.347	0.068
模型设定:				
最小二乘法	模型采用 OLS 估计=1, 否则=0	0.361	0.481	0.347
非线性模型	采用 probit、logit 等非线性模型进行估计=1, 否则=0	0.188	0.391	0.288
固定效应	模型采用固定效应估计=1, 否则=0	0.073	0.260	0.068
工具变量	模型采用 IV 估计=1, 否则=0	0.279	0.449	0.235
断点回归	模型采用 RD 估计=1, 否则=0	0.142	0.350	0.108
变量数目的对数	对模型中的变量数目取对数	2.505	0.801	2.421
样本特征:				
横截面	数据为单一年份横截面数据=1, 否则=0	0.355	0.479	0.530
老年人	数据样本为老年人=1, 否则=0	0.176	0.381	0.227
男性	数据样本为男性=1, 否则=0	0.176	0.381	0.076
女性	数据样本为女性=1, 否则=0	0.176	0.381	0.076
城镇	数据样本为城镇个体=1, 否则=0	0.027	0.163	0.068
农村	数据样本为农村个体=1, 否则=0	0.021	0.144	0.091

注: 加权平均值的权重为每篇文献中纳入本文分析的回归模型数量的倒数。

从健康指标来看, 现有研究使用最多的健康指标是总体自评健康, 占比超过 1/3; 其次是生理健康和体重指标, 各自占比达 1/5。有关教育对心理健康的影响的研究严重匮乏, 只占全部研究的 4.2%。从教育指标来看, 以教育年限为核心自变量的研究最多, 达到 56.4%; 考察基础教育(含小学和中学)和高等教育对健康影响的回归的占比分别为 35.5%和 4.5%, 后者的研究严重不足。

论文的出版年份在 2010 年到 2020 年之间, 表明有关中国教育对健康影响的研究在近十年才受到关注。虽然英文期刊中的回归模型数量占到了全部样本的 63.6%, 但其加权平均却只有 31.8%, 这是因为一篇英文期刊论文中的回归模型数量要大于中文期刊论文。本文样本中, 一篇英文期刊论文和一篇中文期刊论文中平均的回归模型数量分别为 30 和 8。

从模型估计方法来看,采用 OLS 估计的回归模型最多,其次为 IV 估计,两者分别占比 36.1%和 27.9%;采用固定效应估计的回归模型最少,主要应用在采用双胞胎数据的研究中。考虑内生性问题的模型(FE、IV、RD)和不考虑内生性问题的模型(OLS、Probit、Logit)各占一半。在考虑内生性问题的模型中,用于识别的自然实验(占比)为义务教育法(31.1%)、大饥荒(24.3%)、文化大革命(18.2%)、童工法(9.1%)和高校扩招(6.1%),其中义务教育法、大饥荒和文化大革命是使用最多的三种自然实验。此外,还有运用配偶教育(4.5%)和组平均(0.8%)做 IV,以及运用双胞胎样本(6.1%)做固定效应分析的研究。

从数据样本来看,所用数据全部为微观调查数据,其中有中国健康与营养调查(CHNS, 30%)、中国家庭追踪调查(CFPS, 17.9%)、中国老龄健康追踪调查(CLHLS, 16.4%)、中国健康与养老追踪调查(CHARLS, 10.3%)、中国综合社会调查(CGSS, 9.4%)、中国城市劳动力市场调查(CULS, 7.6%)、人口普查(Census, 5.5%)、中国成年双胞胎调查(CATS, 2.4%)和中国劳动力动态调查(CLDS, 0.6%)。本文样本中的估计,只有在因变量为死亡率时,才使用了面板数据,且使用的均是 CLHLS 数据;其他情形采用的都是横截面数据,包括单一年份横截面数据和混合年份横截面数据。使用面板数据和单一年份横截面数据的估计在全部估计中的占比分别为 12.1%和 35.5%。

(三)效应值

本文样本中的文献都采用以下的回归模型来估计教育对健康的影响效应:

$$Health_{ij} = \alpha_j + \beta_j Education_{ij} + \gamma_j' X_{ij} + \epsilon_{ij} \quad (1)$$

其中, $Health_{ij}$ 和 $Education_{ij}$ 分别代表第 j 个回归中的个体 i 的健康和教育状况, X_{ij} 是一组控制变量,表示除教育外,其他影响健康的一系列因素。回归系数 β_j 反映了教育对健康的影响程度和方向,各个研究中估计得到的 β_j 是 Meta 分析中感兴趣的因变量。但如前所述,不同研究中的健康和教育采用了不同的度量指标,因而不同研究中的 β_j 不能直接比较。我们遵照文献中的通常处理方式,将 $\hat{\beta}_j$ 转化为无量纲的偏相关系数后纳入 Meta 分析(Stanley and Doucouliagos, 2012)。偏相关系数的定义如下:

$$PCC_j = \frac{t_j}{\sqrt{t_j^2 + df_j}} \quad (2)$$

其中, t_j 和 df_j 分别是第 j 个估计的 t 统计量和自由度。相应的标准误差为:

$$SE(PCC_j)=\frac{PCC_j}{t_j}$$

(3)

三、Meta 分析结果

(一)整合分析

我们计算了每项研究中所有估计的偏相关系数及标准误的平均值，并以森林图的形式展示于图 2。从图中可以看到，不同研究之间的估计结果存在很大的差异，虽然绝大部分的研究都表明教育对健康有显著的正向影响，但是影响的效应却很小。除一项研究的平均偏相关系数为 0.12 外，其他所有研究中的平均偏相关系数都没有超过 0.01，远低于 Doucouliagos (2011)提出的 0.07 的门槛值。^①

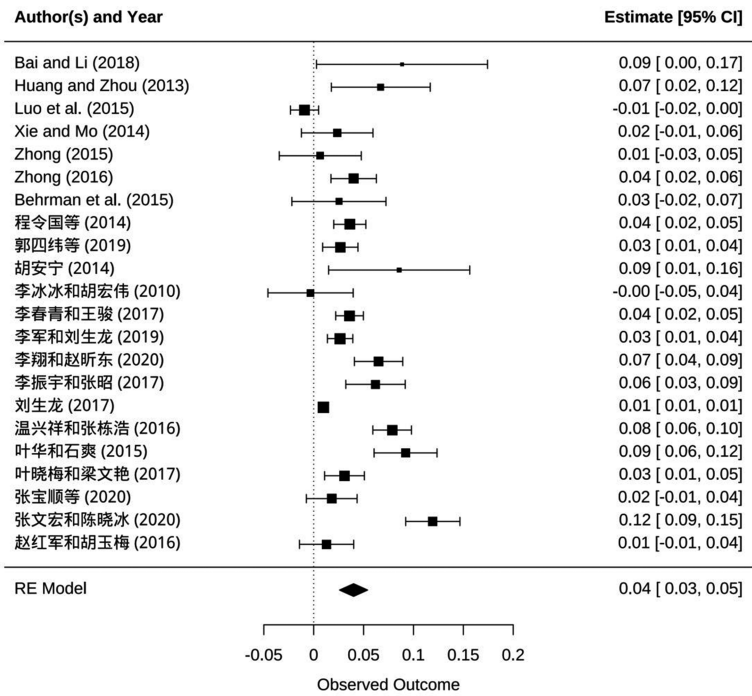


图 2 森林图

① Cohen(1988)提出用 0.1、0.3 和 0.5 作为相关系数表明的相关性小、中和大的区分门槛。相关系数是零阶相关，偏相关系数是控制其他变量情形下两个变量之间的相关，为一阶相关。Doucouliagos (2011) 针对偏相关系数，提出了对于经济学领域 0.070、0.173、0.327 的小、中、大的门槛值。

虽然现有文献的研究结论不一,但是在理论模型参数设置和相关政策制定时却需要有一个统一的明确结论。Meta分析采用固定效应模型和随机效应模型来合成现有的效应值。^①固定效应模型假设所有研究的样本都来自于一个相同的总体,真实的效应值只有一个,不同研究估计得到的不同效应值来源于抽样误差(Borenstein et al., 2011)。固定效应模型通过将所有估计的效应值进行加权平均而得到一个总的平均效应值。权重为估计方差的倒数,即给予更精确的估计(方差更小)更大的权重。随机效应模型则认为不同的研究不是来自于一个单一的总体,而是来自于多个总体,因而真实的效应值有多个,不同研究估计得到的不同效应值既源于抽样误差,也源于抽样总体的不同(Schwarzer et al., 2015)。随机效应模型假设真实的效应值服从一个分布,该分布的方差 τ^2 度量了真实效应值的异质性。随机效应模型先对 τ^2 进行估计,然后将其与样本中各估计的方差(代表抽样方差)相加,以其和的倒数为权重,对所有估计的效应值进行加权平均而得到一个总的平均效应值。

我们分别运用固定效应模型和随机效应模型针对全部样本和各子样本计算了平均效应值,如表2所示。^②表2的结果印证了图2的直观结论。固定效应模型和随机效应模型的估计结果都表明,教育对健康有显著的正向影响,但是影响很小,远没有达到Doucouliagos(2011)提出的0.07的门槛值。子样本的估计结果与全部样本的估计结果类似,几乎全部子样本的估计结果都没有达到0.07的门槛值。

不同健康指标和教育指标下的估计结果差异较大。以生理健康、体重、死亡率和自评健康为健康指标的研究得到的估计效应较小,以认知能力和精神健康为健康指标的研究得到的估计效应较大,其中以认知能力为健康指标的研究得到的估计效应达到了0.07的门槛值,是子样本中唯一一类达到此门槛值的估计。^③以受教育年限为核心解释变量的研究比以学历虚拟变量为核心解释变量的研究得到的估计效应大,以大学学历的虚拟变量为核心解释变量的研究得到的估计效应值非常小且不显著,这表明高等教育对健康并没有显著的促进作用,教育的健康效应主要来自于基础教育。

① Meta分析中的固定效应模型和随机效应模型有自身的独特含义,不同于面板数据分析中的固定效应模型和随机效应模型。

② 当Meta分析的样本是总体或总体中的一个随机样本时,加权平均的效应值是总体的一个无偏估计(Hunter and Schmidt, 2004)。当存在较大的异质性时,随机效应模型更合适;但是当存在发表偏误时,模拟结果显示固定效应模型的偏误更小(Stanley, 2008)。

③ 死亡率是一个负向的健康指标,其估计效应为负值表明教育对健康有正向影响。

在研究教育的健康效应时,会面临健康对教育的反向因果和遗漏时间偏好等引发的内生性问题。如果没有对内生性问题进行很好的处理,那么估计结果将会有偏。表 2 的结果表明,考虑内生性的研究比不考虑内生性的研究得到的估计效应要小很多。这意味着,在研究教育对健康的影响效应时,采用 OLS、Probit、Logit 等估计方法能得到显著的结果,但采用 IV、RD 等却未必能得到显著的结果。在研究教育的健康效应时,是否考虑内生性会影响到研究结论的可信性,内生性问题是此研究主题必须解决的问题。从表 2 中还可以看到,中文期刊论文比英文期刊论文得到的估计效应要大,这主要是因为中文期刊论文中解决内生问题的估计比英文期刊要少。

表 2 平均效应

	固定效应模型		随机效应模型		观测数
	估计值	标准误	估计值	标准误	
全部样本	0.014 ***	0.000	0.030 ***	0.002	330
子样本					
生理健康	0.014 ***	0.001	0.033 ***	0.004	68
认知能力	0.074 ***	0.002	0.069 ***	0.012	34
精神健康	0.040 ***	0.003	0.047 ***	0.011	14
体重	0.012 ***	0.002	0.023 ***	0.005	63
死亡率	-0.010 ***	0.001	-0.010 ***	0.001	40
总体自评健康	0.016 ***	0.001	0.032 ***	0.003	111
小学	0.009 ***	0.0004	0.024 ***	0.007	64
中学	0.009 ***	0.001	0.017 ***	0.005	53
大学	0.002	0.005	-0.001	0.007	15
受教育	0.038 ***	0.002	0.036 ***	0.012	12
教育年限	0.035 ***	0.001	0.038 ***	0.003	186
老年人	0.003 ***	0.001	0.005	0.004	58
男性	0.009 ***	0.001	0.003	0.003	58
女性	0.008 ***	0.001	0.021 ***	0.005	58
城镇	0.064 ***	0.004	0.066 ***	0.015	9
农村	0.055 ***	0.004	0.088 ***	0.024	7
考虑内生性	0.006 ***	0.0005	0.015 ***	0.002	149
未考虑内生性	0.024 ***	0.001	0.036 ***	0.004	181
英文期刊	0.009 ***	0.001	0.026 ***	0.003	210
中文期刊	0.015 ***	0.0004	0.035 ***	0.004	120

(二)发表偏误分析

期刊编辑和审稿人更偏好发表那些能确认某种事前信念或意料之外的统计显著的结果的论文。这使得研究者有很强的激励根据最终发表的可能性来挑选研究结果撰写论文并投稿。这种现象被称之为发表偏误(Andrews and Kasy, 2019)。发表偏误是在学术发表过程中普遍存在的现象,它使得很多不具有统计显著性结果的研究被埋没,从而也被称之为抽屉文件问题。在研究教育对健康的影响时,人们倾向于先验地认为教育会对健康产生正向影响。那些不具统计显著性,得出教育不会对健康产生影响的研究被发表的可能性将很低。

若存在发表偏误,则我们进行 Meta 分析的样本将不是所有潜在研究的一个随机样本,前面整合分析得到的平均效应值将是有偏的。当估计值较大时, t 统计量会更大,从而更可能得到统计显著的结果。所以,存在发表偏误时,已发表的研究中的估计值普遍偏大,Meta 整合得到的平均效应值会比真实的效应值偏大。

即使估计值较小,大样本的研究也容易得到统计显著的结果。而且,大样本数据的搜集和整理会花费更多的时间和精力,所以使用大样本数据的研究存在发表偏误的风险较小。相反,在小样本中,只有估计值较大的研究才会得到统计显著的结果,发表偏误更容易出现在小样本的研究中。可以通过漏斗图来直观展示发表偏误的存在性(Egger et al., 1997)。漏斗图的横轴是估计效应值,纵轴是相应标准误的倒数,用于刻画估计精度。如果不存在发表偏误,则漏斗图中的点将围绕估计效应的均值而对称分布。如果存在向上的发表偏误,则漏斗图中的点将更多地聚集在均值线的右侧。

如图 3 所示,在漏斗图的上方,估计精度较高,较小的估计值也能得到统计显著的结果,研究者挑选研究结果的概率较低,效应值对称分布在均值线两侧;在漏斗图的下方,估计精度较低,只有较大的估计值才能得到统计显著的结果,研究者挑选研究结果的概率较高,效应值更多地聚集在均值线的右侧。这表明,现有关于中国教育健康效应的研究中存在发表偏误。

可以通过下面的回归模型对发表偏误进行检验:

对漏斗图的对称性进行检验(funnel-asymmetry test, FAT):

$$PCC_i = \beta_0 + \beta_1 SE(PCC_i) + \epsilon_i \quad (4)$$

PCC_i 和 $SE(PCC_i)$ 为第 i 个估计的偏相关系数及其标准误。若 $\beta_1 = 0$, 则表明漏斗图是均值对称的,既有研究不存在发表偏误;反之则表明漏斗图不对称,既有研究存在发表偏误。统计显著的 β_1 是发表偏误存在的证据,根据模型(4)对 β_1 进行检验被称之为漏斗图的对称性检验(funnel-asymmetry

test, FAT)。 β_0 是纠正发表偏误后的真实效应, 对 β_0 进行的检验被称之为真实效应检验(precision-effect test, PET)。根据模型(4)来检验发表偏误和真实效应值被统称为 FAT-PET 检验。

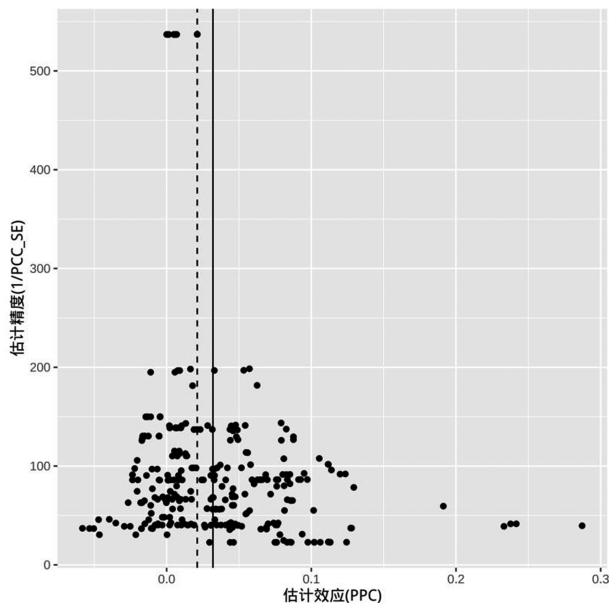


图3 漏斗图

注: 实的竖直线和虚的竖直线分别代表估计效应的均值和中位数。

模拟研究表明, 当真实效应 β_0 不为零时, FAT-PET 检验存在向下的偏误, 低估真实的 β_0 (Stanley, 2008; Stanley and Doucouliagos, 2012)。将模型(4)中偏相关系数的标准误 $SE(PCC_i)$ 以其方差 $SE(PCC_i)^2$ 替代, 得到:

$$PCC_i = \beta_0 + \beta_1 SE(PCC_i)^2 + \epsilon_i \quad (5)$$

模拟研究表明, 当真实效应不为零时, 模型(5)对 β_0 提供了更好的估计 (Stanley and Doucouliagos, 2014)。模型(5)中的 β_1 度量了发表偏误, β_0 是纠正发表偏误后的真实效应。根据模型(5)来检验发表偏误和真实效应值被称为 PEESE(precision-effect estimate with standard error)。^①

鉴于模型(4)和模型(5)中存在明显的异方差, 我们对其进行加权最小二乘估计, 权重为 $1/SE(PCC_i)^2$ 。FAT-PET 和 PEESE 的估计结果如表 3 的第(1)、(2)列所示。根据 Stanley and Doucouliagos(2013)的标准, 表 3 的估

^① PEESE 的名称来自于将模型(5)两边同时除以 $SE(PCC_i)$ 得到的加权最小二乘形式的估计模型: $t_i = \beta_1 SE(PCC_i) + \beta_0 [1/SE(PCC_i)] + v_i$ 。

计结果表明有关中国教育对健康的影响的研究中存在着较为严重的发表偏误。^① 纠正发表偏误后, FAT-PET 和 PEESE 分别得出了 0.007 和 0.012 的真实效应估计值。这个结果比前述整合分析中采用固定效应模型和随机效应模型得到的 0.014 和 0.03 的效应值都要低, 更低于 Doucouliagos(2011)提出的 0.07 的门槛值。

Stanley 等(2017)提出了 WAAP (weighted average of adequately powered)估计量来计算偏差调整的真实效应。与固定效应模型计算全部样本不同, WAAP 只计算“功效”较强的估计样本。Stanley 等(2017)将 $SE < |FE|/2.8$ 的估计界定为“功效”强的估计, 其中 SE 为标准误, FE 为固定效应估计值。蒙特卡洛模拟结果表明, 在存在发表偏误时, WAAP 能减少固定效应模型中的偏差。本文样本中符合 Stanley 等(2017)标准的“功效”强的估计只有 20 个, 只占全部样本量的 6.1%。本文样本中的弱功效估计占主导意味着存在发表偏误的概率极大。如表 3 第(3)列所示, WAAP 的计算得到的真实效应为 0.01, 介于 FAT-PET 和 PEESE 之间。

表 3 发表偏误的三种检验

	FAT-PET		PEESE		WAAP
	FE	RE	FE	RE	
	(1)	(2)	(3)	(4)	(5)
真实效应	0.007*** (0.002)	0.010 (0.009)	0.012*** (0.003)	0.020*** (0.007)	0.010*** (0.002)
发表偏误	1.674*** (0.498)	1.387*** (0.522)	56.812*** (13.382)	32.529*** (9.313)	
观测数	330	330	330	330	20
调整 R ²	0.042	0.352	0.320	0.351	

注: 表中数字为估计系数, 括号内是在每项研究上的聚类标准误;***、**、* 分别代表 1%、5%、10% 的统计显著性水平。

(三) 异质性来源分析

现有对中国教育的健康效应的研究结论不一。本文样本中的偏相关系数分布在[-0.058, 0.287]内, 有 26.4% 的偏相关系数小于 0, 73.6% 的偏相关系数大于 0。我们可以通过在模型(4)中加入对估计效应有影响的其他变

① Stanley and Doucouliagos(2013)提出, 如果 β_1 不显著或 $|\beta_1| < 1$, 则发表偏误不严重; 如果 β_1 显著且 $1 \leq |\beta_1| \leq 2$, 则发表偏误比较严重; 如果 β_1 显著且 $|\beta_1| > 2$, 则发表偏误非常严重。

量, 进行多元 Meta 回归来考察估计结果差异的来源。

$$PCC_i = \beta_0 + \beta_1 SE(PCC_i) + \sum_{j=1}^K \beta_j + 1X_{ki} + \epsilon_i \quad (6)$$

为了克服异方差, 我们将模型(6)两边同时除以 $SE(PCC_i)$, 得到:

$$\frac{PCC_i}{SE(PCC_i)} = \beta_0 \cdot \frac{1}{SE(PCC_i)} + \beta_1 + \sum_{j=1}^K \beta_j + 1 \cdot \frac{X_{ki}}{SE(PCC_i)} + \epsilon_i \cdot \frac{1}{SE(PCC_i)} \quad (7)$$

模型(7)中的常数项 β_1 代表发表偏误, 系数 β_0 为真实效应。对模型(7)进行有效估计的一个难点在于变量 X_{ki} 的选择。不同于一般的计量研究项目, Meta 分析的样本量通常不大, 而影响研究结果的因素又特别多。所以, 模型设定对 Meta 分析而言更显重要。

为了避免主观选择 X_{ki} 导致的“设定操纵”问题, 我们首次在 Meta 分析中引入 Lasso 估计来解决变量选择难题。我们借鉴已有的 Meta 分析文献, 从健康指标、教育指标、出版特征、模型设定、样本特征四个方面全面搜集整理了共计 26 个可能对估计结果有影响的研究特征变量(如前述表 1 所示)。相对于本文的样本量来讲, 26 个研究特征变量数目较多, 且这些变量之间存在着多重共线。如果将所有变量放入模型后进行 OLS, 则难以得到精确估计结果。如果主观挑选部分变量进行 OLS, 则又难以避免“设定操纵”之嫌。适用于高维数据的 Lasso 估计通过在 OLS 目标函数中加入参数向量的 1-范数的惩罚项来达到从大量特征变量中, 筛选出真正对因变量有影响的少数变量的目的。

我们将全部 26 个研究特征变量作为 X_{ki} 放入回归模型(7)中, 再对其进行 Lasso 估计。^① 表 4 的第(1)列给出了 Lasso 估计结果, 非零系数对应的变量即是通过 Lasso 选择得到的对因变量有真正影响的变量。但通常的 Lasso 估计只有估计系数, 而没有标准误等表明统计显著性的估计量。一种解决办法是针对 Lasso 筛选出的变量进行 OLS 估计得到 Post Lasso 估计量(Belloni and Chernozhukov, 2013), 结果如表 4 第(2)列所示。

Lasso 虽然能筛选出对因变量有影响的变量, 但如果有多多个对因变量有较强影响但同时相关性又较强的变量, 那么 Lasso 在筛选变量的过程中会去

① 本文采用 R 包 hmd 中的 `rlasso()` 和 `rlassoEffects()` 函数进行 Lasso 估计和 Double Selection Regression, 其主要原因是: (1) 本文的 Meta 回归模型中存在异方差且因变量非正态分布, 而 hmd 包正好可以处理此类问题; (2) hmd 包采用有理论基础的数据驱动方法选择 Lasso 估计中的重要参数——惩罚项 λ , 因而也被称之为严格的(rigorous) Lasso, `rlasso()` 和 `rlassoEffects()` 的前缀 r 即由此而来(Chernozhukov et al., 2016)。

掉其中一些变量,此时 Post Lasso 估计量就会因遗漏变量问题而有偏。因此我们进一步基于 Belloni et al. (2014) 提出的双重选择回归得到更为稳健的 Post Double Lasso 估计量。我们将 Lasso 估计筛选出的变量依次作为关键变量,将因变量和关键变量分别对所有控制变量进行 Lasso 回归,然后对这两个 Lasso 回归所选中的控制变量取并集后代入原方程进行 OLS 回归,以此得到关键变量的 Post Double Lasso 估计。我们对每个 Lasso 估计筛选出的变量都进行一次双重选择回归,以得到每个变量的 Post Double Lasso 估计,结果如表 4 第(3)列所示。^①

表 4 的 Lasso 估计结果表明,在研究中国教育对健康的影响的研究文献中,发表偏误存在,而真实效应却不存在。这进一步印证了前文整合分析和发表偏误检验的研究结论。即使我们在 Lasso 估计中不对真实效应变量进行惩罚,而在 Post Lasso 估计和 Post Double Lasso 估计中加入真实效应变量,估计结果也仍然不显著。

从研究特征变量来看,健康指标、教育指标、出版特征和模型估计方法都会对研究结果有影响,但样本特征并不会影响研究结论。Post Lasso 估计和 Post Double Lasso 估计的结果总体上相近,其中认知功能、大学、英文期刊、最小二乘法四个变量在两个估计中都在 5% 的统计显著性水平上显著。

以认知功能度量健康来研究教育的健康效应时,更容易得到大的效应值,从而更可能得到统计显著的研究结果。这表明,教育对认知功能有显著提升作用。这个结论很容易理解。教育的一个主要目的就是培养认知技能,虽然教育主要在低龄阶段完成,但习得的认知技能会受益终身,甚至对老年认知都有显著的正面促进作用。体重变量和死亡率变量分别在 Post Lasso 估计和 Post Double Lasso 估计中显著,且系数均为负值。这意味着,教育对以客观指标度量的健康的影响效应很小,更难得到统计显著的研究结果。其中的原因可能在于,体重和死亡率更可能跟先天的基因相关,后天的教育对其影响有限。

以大学学历度量教育来研究教育的健康效应时,得到的估计效应值会更小,从而更难得到统计显著的研究结果。这说明,教育的健康效应在教育阶段上存在异质性。初等教育是通识教育、素质教育,目的是促进人的全面发展,其回报不仅体现在将来的劳动力市场回报上,还会体现在健康等非劳动力市场回报上;高等教育是专业教育,目的是让学生掌握某一专业的高级知

^① 模型(7)中,代表发表偏误的是常数项,故表 4 的第(3)列 Post Double Lasso 估计中没有发表偏误的估计结果。

识和技能，其回报更多地体现在个体的就业、工资等劳动力市场回报上。

发表在英文期刊的研究估计得到的教育对健康的影响效应值更小，其主要原因在于英文期刊更注重研究的严谨性，因而会更多地采用工具变量、断点回归等方法处理内生性问题。本文样本中，英文期刊和中文期刊处理了内生性的估计所占的比重分别为 49%和 38%。表(4)的结果表明，采用 OLS 方法估计教育的健康效应时，所得效应值会更大，因而更容易出现统计显著的研究结果。与此对照，那些采用了固定效应、工具变量、断点回归等方法估计教育的健康效应的研究，更难得到统计显著的研究结果。

我们进一步分子样本进行 FAT-PET 检验和 PEESE 检验，结果如表 5 和表 6 所示。表 5 根据是否以认知功能度量健康将样本分为两类。可以看到，在以认知功能度量健康的样本中，不存在发表偏误，其真实效应值为 0.08，达到了 Doucouliagos(2011)提出的 0.07 的门槛值；但在以其他指标度量健康的样本中，则存在十分显著的发表偏误，剔除发表偏误后的真实效应值不超过 0.02。表 6 根据是否控制内生性将样本分为两类。可以看到，在控制了内生性的样本中，无论是 FAT-PET 检验还是 PEESE 检验，也无论是用固定效应加权还是用随机效应加权，都存在着显著的发表偏误。但是，在不控制内生性的样本中，总体上并不存在十分显著的发表偏误。此外，在每类估计中，控制了内生性的样本的真实效应值都比不控制内生性的样本要小。

表 4 异质性来源的 lasso 估计

	Lasso	Post Lasso		Post Double Lasso	
	(1)	(2)		(3)	
	系数	系数	p 值	系数	p 值
真实效应	0	0.001	0.567	−11.189	0.714
发表偏误	2.223	2.087	0.000		
生理健康	0				
认知功能	0.026	0.037	0.000	0.026	0.005
精神健康	0				
体重	−0.011	−0.022	0.014	−0.016	1
死亡率	−0.010	−0.009	0.248	−0.050	0
总体自评健康	0				
小学	0				
中学	0				
大学	−0.017	−0.023	0.107	−0.022	0.017
接受教育	0				

续表					
	Lasso	Post Lasso		Post Double Lasso	
	(1)	(2)		(3)	
	系数	系数	p 值	系数	p 值
教育年限	0.006	0.013	0.002	-0.023	1
出版年份的对数	0				
英文期刊	-0.014	-0.018	0.013	-0.024	0.007
由星号计算得到 t 值	0				
最小二乘法	0.019	0.020	0.000	0.042	0.031
非线性模型	0				
固定效应	0				
工具变量	-0.0003	-0.004	0.288	0.002	0.401
断点回归	0				
变量数目的对数	0				
横截面	0				
老年人	0				
男性	0				
女性	0				
城镇	0				
农村	0				

注：表中数字为估计系数，括号内是在每项研究上的聚类标准误；***、**、* 分别代表 1%、5%、10% 的统计显著性水平。

表 5 子样本分析 1：是否以认知功能度量健康

	Yes				No			
	FAT-PET		PEESE		FAT-PET		PEESE	
	FE	RE	FE	RE	FE	RE	FE	RE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
真实效应	0.082*** (0.009)	0.080*** (0.011)	0.078*** (0.007)	0.076*** (0.009)	0.007*** (0.001)	0.008 (0.009)	0.011*** (0.003)	0.017*** (0.006)
发表偏误	-0.659** (0.327)	-0.523 (0.439)	-17.940* (10.333)	-15.111 (13.230)	1.379*** (0.484)	1.276** (0.538)	48.784*** (10.747)	30.487*** (9.125)
观测数	34	34	34	34	296	296	296	296
调整 R ²	0.407	0.471	0.704	0.471	0.055	0.351	0.317	0.355

表 6 子样本分析 2：是否控制内生性

	Yes				No			
	FAT-PET		PEESE		FAT-PET		PEESE	
	FE	RE	FE	RE	FE	RE	FE	RE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
真实效应	0.002*** (0.001)	-0.006 (0.007)	0.005*** (0.001)	0.004 (0.003)	0.015** (0.006)	0.009 (0.023)	0.019*** (0.007)	0.026* (0.013)
发表偏误	1.057*** (0.323)	1.565*** (0.474)	37.977*** (5.721)	39.041*** (5.446)	1.691* (0.900)	2.352 (1.810)	91.233** (39.622)	59.849 (48.863)
观测数	149	149	149	149	181	181	181	181
调整 R ²	0.033	0.497	0.447	0.550	0.081	0.376	0.421	0.364

四、结论

源于教育的健康效应研究中存在相互冲突的研究结论的现象，本文对中英文期刊发表的中国教育的健康效应的实证研究文献进行了 Meta 分析。我们主要做了三项工作：(1)以通用的偏相关系数作为效应值的度量，分别运用固定效应模型和随机效应模型针对全部样本和各子样本计算了平均效应值；(2)基于 FAT-PET 检验和 PEESE 检验考察了中国教育的健康效应研究中的发表偏误的存在性，并计算了剔除发表偏误后的真实效应值；(3)运用多元 Meta 回归模型考察了研究异质性的来源。

我们得到以下研究结论：(1)无论是总样本还是子样本，固定效应模型和随机效应模型的估计结果都表明，教育对健康虽有显著的正向影响，但是影响效应却很小，远没有达到 Doucouliagos(2011)提出的 0.07 的门槛值；(2)现已发表的研究文献存在显著的发表偏误，在剔除发表偏误后，中国教育对健康的真实影响效应进一步降低至接近于零的水平；(3)健康指标选取、教育指标选取以及是否解决内生性是导致研究结论差异的三个重要来源，以认知功能度量健康的研究和采用 OLS 估计的研究得到的估计效应值更大、更容易得到统计显著的研究结果、存在发表偏误的可能性更低，教育的健康效应在教育阶段上存在异质性，高等教育相比初等教育更难有健康溢价。

本文的主要贡献是：(1)在研究对象上，首次针对中国教育健康效应问题采用 Meta 分析方法得出定量合成结论并检验了发表偏误的存在性；(2)在研究方法上，首次引入 Lasso 估计解决运用 Meta 回归方法考察研究异质性来

源时的模型选择的不确定性问题；(3)在研究结论上，揭示了中国教育的健康效应研究中存在的发表偏误及其接近于零的真实效应值，挖掘得出健康指标选取、教育指标选取以及是否解决内生性是导致研究结论差异的三个重要来源的结论。虽然我们尽可能多地采集了 26 个描述研究特征的变量，但仍可能存在一些无法观测的遗漏变量因素。如何解决这些可能的遗漏变量给 Meta 回归带来的估计偏差问题是未来需要进一步研究的问题。

[参考文献]

- 程令国、张晔、沈可，2015：《教育如何影响了人们的健康？——来自中国老年人的证据》，《经济学(季刊)》第 1 期。
- 郭四维、张明昂、曹静，2019：《教育真的可以影响健康吗？——来自中国 1986 年义务教育法实施的证据》，《经济学报》第 3 期。
- 胡安宁，2014：《教育能否让我们更健康——基于 2010 年中国综合社会调查的城乡比较分析》，《中国社会科学》第 5 期。
- 李军、刘生龙，2019：《教育对健康的影响——基于中国 1986 年义务教育法的实证分析》，《数量经济技术经济研究》第 6 期。
- Acemoglu, D. and S. Johnson, 2007, "Disease and Development: The Effect of Life Expectancy on Economic Growth", *Journal of Political Economy*, 115(6): 925—85.
- Albouy, V. and L. Lequien, 2009, "Does Compulsory Education Lower Mortality?", *Journal of Health Economics*, 28(1): 155—168.
- Andrews, I. and M. Kasy, 2019, "Identification of and Correction for Publication Bias", *American Economic Review*, 109(8): 2766—2794.
- Belloni, A. and V. Chernozhukov, 2013, "Least Squares after Model Selection in High-dimensional Parse Models", *Bernoulli*, 19(2): 521—547.
- Belloni, A., V. Chernozhukov and C. Hansen, 2014, "Inference on Treatment Effects after Selection among High-Dimensional Controls", *Review of Economic Studies*, 81(2): 608—650.
- Borenstein, M., V. Hedges, J. Higgins and H. Rothstein, 2011, *Introduction to Meta-Analysis*, John Wiley & Sons.
- Cervellati, M., and U. Sunde, 2005, "Human Capital Formation, Life Expectancy and the Process of Development", *American Economic Review*, 95(5): 1653—1672.
- Chernozhukov, V., C. Hansen and M. Spindler, 2016, "hdm: High-Dimensional Metrics", *The R Journal*, 8(2): 185—199.
- Clark, D. and H. Royer, 2013, "The Effect of Education on Adult Health and Mortality: Evidence from Britain", *American Economic Review*, 103(6): 2087—2120.
- Cohen, J., 1988, *Statistical Power Analysis in the Behavioral Sciences*, 2nd ed, Hillsdale: Erlbaum.

- Conti, G. , J. Heckman and S. Urzua, 2010, “The Education-Health Gradient”, *American Economic Review*, 100(2): 234—238.
- Doucouliaos, H. , 2011, “How Large is Large? Preliminary and Relative Guidelines for Interpreting Partial Correlations in Economics”, *Working Paper*, Deakin University.
- Egger, M. , G. Smith, M. Schneider and C. Minder, 1997, “Bias in Meta-Analysis Detected by a Simple, Graphical Test”, *British Medical Journal*, 315: 629—634.
- Galama, T. J. , A. A. Lleras-Muney and H. van. Kippersluis, 2018, “The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence”, *Oxford Research Encyclopedia of Economics and Finance*, 1—99.
- Grossman, M. , 1972, “On the Concept of Health Capital and the Demand for Health”, *Journal of Political Economy*, 80(2): 223—255.
- Grossman, M. , 2000, “The Human Capital Model”, In *Handbook of Health Economics* (Vol. 1, pp. 347—408), Elsevier.
- Grossman, M. , 2015, “The Relationship between Health and Schooling: What’s New?”, *Nordic Journal of Health Economics*, 3(1): 7—17.
- Havranek, T. , Z. Irsova and O. Zeynalova, 2018, “Tuition Fees and University Enrolment: A Meta-Regression Analysis”, *Oxford Bulletin of Economics and Statistics*, 80(6): 1145—1184.
- Heckman, J. , 2007, “The Economics, Technology and Neuroscience of Human Capacity Formation”, *Proceedings of the National Academy of Sciences*, 104(33): 13250—13255.
- Hunter, J. E. and F. L. Schmidt, 2004, *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, New York: Sage.
- Kemptoner, D. , H. Jürges and S. Reinhold, 2011, “Changes in Compulsory Schooling and the Causal Effect of Education on Health: Evidence from Germany”, *Journal of Health Economics*, 30(2): 340—354.
- Kippersluis, H. , O. O’Donnell and E. Doorslaer, 2011, “Long-run Returns to Education: Does Schooling Lead to an Extended Old Age?”, *Journal of Human Resources*, 46(4): 695—721.
- Lager, A. , C. Jonas and J. Torssander, 2012, “Causal Effect of Education on Mortality in a Quasi-experiment on 1.2 Million Swedes”, *Proceedings of the National Academy of Sciences*, 109(22): 8461—8466.
- Lleras-Muney, A. , 2005, “The Relationship between Education and Adult Mortality in the United States”, *The Review of Economic Studies*, 72(1): 189—221.
- Porta, M. , 2016, *A Dictionary of Epidemiology*, Oxford University Press.
- Schwarzer, G. , J. R. Carpenter and G. Rücker, 2015, *Meta-Analysis with R*, Springer.
- 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 H. Doucouliagos, 2012, *Meta-regression Analysis in Economics and*

Business, Routledge.

- Stanley, T. D., H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, G. Pugh, R. S. Rosenberger and K. Rost, 2013, “Meta-Analysis of Economics Research Reporting Guidelines”, *Journal of Economic Surveys*, 27(2): 390—394.
- Stanley, T. D. and H. Doucouliagos, 2014, “Meta-regression Approximations to Reduce Publication Selection Bias”, *Research Synthesis Methods*, 5(1): 60—78.
- Stanley, T. D., H. Doucouliagos and J. P. A. Ioannidis, 2017, “Finding the Power to Reduce Publication Bias”, *Statistics in Medicine*, 36(10): 1580—1598.
- Xue, X., M. Cheng and W. Zhang, 2021, “Does Education Really Improve Health? A Meta-Analysis”, *Journal of Economic Surveys*, 35(1): 71—105.

附录：纳入本文 Meta 分析的文献列表

- 程令国、张晔、沈可，2015：《教育如何影响了人们的健康？——来自中国老年人的证据》，《经济学（季刊）》第 1 期。
- 郭四维、张明昂、曹静，2019：《教育真的可以影响健康吗？——来自中国 1986 年义务教育法实施的证据》，《经济学报》第 3 期。
- 胡安宁，2014：《教育能否让我们更健康——基于 2010 年中国综合社会调查的城乡比较分析》，《中国社会科学》第 5 期。
- 李冰水、胡宏伟，2010：《教育与医疗保险对老年人健康状况的影响》，《南方人口》第 6 期。
- 李军、刘生龙，2019：《教育对健康的影响——基于中国 1986 年义务教育法的实证分析》，《数量经济技术经济研究》第 6 期。
- 李春青、王骏，2017：《教育对健康的影响——基于中国家庭追踪调查数据的分析》，《北京社会科学》第 11 期。
- 李振宇、张昭，2017：《教育对个体健康人力资本的影响——以义务教育法实施为工具变量》，《教育与经济》第 3 期。
- 李翔、赵昕东，2020：《教育如何影响我国老年人健康水平？》，《财经研究》第 3 期。
- 刘生龙，2017：《教育对老年健康的影响：来自断点回归模型的实证证据》，《学术研究》第 11 期。
- 温兴祥、张栋浩，2016：《教育能延缓认知老化吗？——来自“教育革命”的证据》，《世界经济文汇》第 1 期。
- 叶晓梅、梁文艳，2017：《教育对中国老年人健康的影响机制研究——来自 2011 年 CLHLS 的证据》，《教育与经济》第 3 期。
- 叶华、石爽，2015：《健康的教育梯度、城乡差异与影响机制》，《学术研究》第 9 期。
- 张文宏、陈晓冰，2020：《教育对个体健康水平的影响》，《山东社会科学》第 7 期。
- 张宝顺、于林月、陈龙，2020：《教育对老年人健康的影响机制研究》，《西北人口》第 1 期。
- 赵红军、胡玉梅，2016：《教育程度一定会提高健康水平吗？——基于中国家庭追踪调查

(CFPS)的实证分析》，《世界经济文汇》第6期。

- Bai, Y. and Y. Li, 2018, “A Sibling-Pair Analysis for Causal Effect of Education on Health”, *The B.E. Journal of Economic Analysis & Policy*, 18(4): 1—10.
- Behrman, J. R., Y. Xiong and J. Zhang, 2015, “Cross-sectional Schooling-health Associations Misrepresented Causal Schooling Effects on Adult Health and Health-related Behaviors: Evidence from the Chinese Adults Twins Survey”, *Social Science & Medicine*, 127: 190—197.
- Huang, W. and Y. Zhou, 2013, “Effects of Education on Cognition at Older Ages: Evidence from China’s Great Famine”, *Social Science & Medicine*, 98: 54—62.
- Luo, Y., Zhang, Z., and Gu, D. (2015). Education and Mortality among Older Adults in China. *Social Science & Medicine*, 127, 134—142.
- Xie, S. and T. Mo, 2014, “The Impact of Education on Health in China”, *China Economic Review*, 29: 1—18.
- Zhong, H., 2015, “Does a College Education Cause Better Health and Health Behaviours?”, *Applied Economics*, 47(7): 639—653.
- 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.

The Health Effects of Education in China: A Meta-Regression Analysis

PENG Shu-hong

(School of Public Finance and Public Administration, Jiangxi University of Finance and Economics)

Abstract: Originating from the phenomenon of conflicting research conclusions in the research on the health effects of education, this paper conducts a meta—analysis of empirical studies on the health effects of education in China. The FAT-PET test and the PEESE test show that there are significant publication errors in the published research literature. After excluding publication errors, the true effect of education on health in China is close to zero. The Post Lasso estimators and Double Post Lasso estimators of multiple meta—regression models show that the selection of health indicators, the selection of education indicators and whether to solve endogeneity are three important sources leading to the differences in research conclusions.

Key words: education; health; China; meta-analysis; publication bias

(责任编辑: 刘泽云 责任校对: 刘泽云 胡咏梅)