

科学研究中因果推断的方法、应用与展望 ——以个体健康研究为例

任国强¹, 王于丹¹, 周云波²

(1. 天津理工大学 管理学院, 天津 300384;

2. 南开大学 经济学院, 天津 300071)

摘要: 以个体健康研究为例, 探讨科学研究中因果推断的方法、应用与展望。因果推断方法作为研究个体健康影响机制的重要手段, 有助于促进健康相关政策制定的科学化、合理化, 为个体健康提供更为可靠的社会保障, 对提高居民健康水平具有十分重要的社会意义。从个体健康研究中的常见问题、因果推断理论框架、实证研究中的因果推断三个方面对现有文献进行分析和评价。作为涵盖范围最广泛的研究领域, 个体健康水平受到社会资本、收入、教育、保险、迁移、退休、工作等多类因素的影响, 同时, 实证研究中主要存在测量偏误、遗漏变量、互为因果、共同原因和选择偏差五类问题, 解决这些问题常用的因果推断方法主要包括随机控制实验、倾向得分匹配、工具变量法、双重差分法、断点回归设计以及个体固定效应模型。研究阐述了这些因果推断模型的适用条件及优缺点, 并对各类模型在个体健康实证研究中的应用作了简要的总结和分析。研究还有助于学者在个体健康研究领域中选择合适的因果推断方法, 或进一步综合应用已有的方法。随着大数据技术的发展和因果推断方法的深入了解, 未来应加强机器学习和因果推断方法的结合, 丰富已有的因果推断工具, 保障研究结果的稳健性。

关键词: 个体健康; 内生性; 混杂因素; 因果推断

中图分类号: C92-03 **文献标识码:** A **文章编号:** 1000-4149 (2022) 02-0001-25

DOI: 10.3969/j.issn.1000-4149.2021.00.070

收稿日期: 2021-09-09; 修订日期: 2021-12-01

基金项目: 国家自然科学基金项目“我国企业的收入分配对企业效益影响的定量研究”(71672086); 国家社会科学基金重大项目“基于多维视角的2020年以后我国相对贫困问题研究”(19ZDA052)。

作者简介: 任国强, 经济学博士, 天津理工大学管理学院教授; 王于丹, 天津理工大学管理学院硕士研究生; 周云波, 经济学博士, 南开大学经济学院教授, 博士生导师。

一、引言

健康不仅是人类的普遍愿望和基本需求,而且作为人力资本的重要组成部分,对个体、家庭和社会都会产生重要的影响。从个体角度来看,健康可以促进教育水平的提高和社会资本的形成,从而影响就业状态、就业类型,进而影响就业收入和社会阶层地位^[1];从家庭角度来看,健康可以增加家庭的收入,减少医疗费用的支出,促进家庭幸福^[2];从社会经济角度来看,健康有利于促进劳动力供给,延长就业者的工作年限,促进劳动生产率,从而促进经济增长^[3]。正是由于健康的重要性,对健康的研究一直是学术界研究的热点,以“all in title: health”作为搜索条件,用谷歌学术进行搜索,搜索结果达到142万余条,再加上以各种疾病研究为论文标题的文献就更多。我们在Web of Science数据库中,在标题、摘要、关键词中搜索“causal & health”,然后再分别检索各个因果推断方法和健康(例如:“Propensity Score & health”),共得到了七万余条检索记录,尽管其中可能有重复文献,但即使剔除这些文献仍有大量检索记录。不难看出,因果推断方法在健康领域得到了广泛的应用。该领域的研究早在20世纪90年代就进入大众视野,2010年前后得到了更多的关注,如今更是蓬勃发展,研究范围也逐步从公共环境卫生、临床医学、生物学等领域,开始扩大至经济学、社会学的范畴,而影响个体健康的相关因素除了作为控制异质性的个体人口学特征以外,还包括了教育、工作、收入和社会资本等个人因素,以及迁移、退休和保险等政策性因素。本文就是从经济学、社会学、人口学等领域入手,从多角度考虑各领域因素对个体健康的影响。

尽管健康研究很重要,研究的成果又非常丰富,但是由于其内在复杂性,健康研究仍然存在很多问题,主要表现在以下几个方面:一是被解释变量——健康的测量误差,很多研究采用自评健康作为研究对象,对健康状态的回答主要根据个体主观判断,而个体由于认知能力的差别,往往做不到完全理性,因此回答的结果和自己真实的健康状况通常有一定的差别;二是遗漏变量的存在会对结果造成干扰,使得解释变量的估计结果不准确,例如在教育对健康的影响时,由于无法准确衡量身体禀赋这一不可观测变量,回归时往往遗漏该变量,从而过高估计了教育对健康的影响;三是解释变量和被解释变量之间可能互为因果,比如工作时间过长可能会导致健康状况下降,反之良好的健康状况又会让劳动者有更多的精力投身于工作^[4];四是存在同时影响解释变量和结果的第三方变量,比如教育水平越高相应的收入水平也越高,健康状况可能也越好,而收入对健康也有直接的影响,所以教育就是影响收入和健康的共同原因;五是样本自选择偏差,以迁移对健康的影响为例,迁移的个体往往比不迁移个体更健康,但是这个结果并不一定是迁移政策造成的,可能是迁移个体的健康状况往往更好,因为个人健康状况的好坏决定着其流动机会和流动决策。

为了解决上述健康研究中存在的问题,学者们采用了相应的因果推断方法作为分析工具。例如在研究退休对健康的因果效应时,工具变量的方法可以最大限度地减少对测量误差的担忧^[5];与之类似,库鲁克利斯(Kourouklis)等人在研究收入对健康的影响时,使用工具变量方法更好地处理了因为遗漏变量所带来的潜在偏差^[6];阿塔莱(Atalay)等人为解决双向因果关系和测量误差,使用固定效应模型研究了妻子退休对丈夫心理健康的影响^[7];

龙翠江等人利用倾向得分匹配与双重差分相结合的方法对迁移和健康之间的因果关系展开了深入研究,解决了选择性偏差和互为因果带来的问题^[8]。这些方法已经被广泛应用于个体健康研究,用来解决研究过程中产生的一系列问题。当然对同一类问题的研究根据数据特征和研究背景的不同,可以采用多种因果推断的方法。

在个体健康研究文献中,由于采用的研究方法存在缺陷,许多文献所指出的影响因素与健康之间的关系是相关而非因果关系,而影响因素与健康之间的因果关系对政策的制定和实施至关重要。尽管很多学者采用了因果推断的方法来研究影响因素和健康之间的因果关系,也有学者从公共健康的角度对相关研究进行归纳^[9],但是鲜有研究从健康的社会经济决定方面对个体健康研究中的因果推断问题进行总结。本文的目的在于通过识别个体健康研究中的常见问题,评价常用的因果推断方法及其在个体健康研究中的应用。由于现有个体健康研究的文献过多,因此我们选择近十年的文献进行分析,尽可能涵盖多个方面,但研究文献中未包括临床医学方面的研究。

二、个体健康研究的常见问题

个体健康研究中常见的问题主要来源于被解释变量、解释变量、解释变量和被解释变量关系、混杂因素以及研究样本的选择性五个方面。

在进行健康研究时,作为被解释变量的健康除了可以采用发病率指数^[10-11]、身体质量指数 BMI^[12] 和特定疾病诊断^[13] 等客观指标外,还会采用自评健康 (Self-Reported Health, SRH)^[14] 和心理健康^[15] 等主观指标。当使用像 SRH 这样有序的主观测量作为客观健康的代理时,研究者必须假设 SRH 是对个体健康状态的真实反映。无论我们是根据性别、种族、年龄还是社会经济地位来定义群体,当不同的群体以不同的方式自我报告他们的健康状况时,都有可能造成较大的测度误差^[16]。SRH 的测度误差可能会对研究结果产生不利的影响,一方面是影响结果的可靠性,另一方面对自评健康的过高评价可能导致风险健康行为^[17]。尽管对报告偏差和解释存在担忧,但是由于其易于收集的特点,SRH 仍然是研究人员最常用的健康衡量标准,一些研究结果也表明 SRH 不仅与客观健康指标具有较强的相关性,而且确实可以预测未来的客观健康状况^[16]。虽然学术界对 SRH 是不是可靠的健康指标并没有达成一致,学者们还是希望 SRH 越可靠越好,一个可选择的方案是对自我健康状况进行多次评估。以中国健康与养老追踪调查 (CHARLS) 数据为例,CHARLS 问卷有两次关于自评健康状况的调查,第一次受访者未经任何准备直接被询问健康状况,回答可能过于主观,第二次则是在被询问有关慢性病等客观指标的相关问题之后,一部分受访者的自评健康水平发生了变化,所以受访者第二次的自评答案更加真实可靠。

个体健康研究中解释变量方面存在的问题主要是由遗漏变量所造成的。遗漏变量使得残差项与解释变量相关,导致解释变量的回归系数是有偏的。但是在进行计量分析时,很难做到不遗漏变量,一方面很难判断哪些变量要加入控制变量集,即使知道某个变量应该加入控制变量中,实际上也可能无法测度 (如天赋和能力),另一方面是可能存在未知的控制变量。以保险和健康的关系为例,有学者指出,保险与健康之间存在一些不可观察的变量,主

要包括：①个人的健康意识。个人的健康意识越强，越可能拥有健康的生活方式，也越可能参加医疗保险以预防和治疗疾病。②个人的风险偏好。风险规避者往往倾向于参加医疗保险，同时更注重采取预防保健措施和避免危害健康的行为，这些无法度量的变量使得研究者很难建立起医疗保险与健康状况之间真正的因果关系^[18]。

被解释变量与解释变量之间可能会存在双向因果关系。如果两个变量互为因果，任何一方都可以作为对方的解释变量，那么任何一个单方面的回归都可能带来内生性问题。许多健康的影响因素和健康之间存在双向因果关系，例如收入与健康之间的关系、工作时长与健康之间的关系等。以收入与健康之间的关系为例，很多学者认为较高的收入会带来较高的健康水平，较好的健康也会使得收入增加，但这并不能说明二者之间的因果关系，要分析收入对健康的因果影响，一定要保证收入变量是外生的，但在实证研究中却很难做到：一方面，可以有力预测收入的因素往往也直接影响健康，如教育水平、失业和工作经历；另一方面，存在不可观测的因素，如遗传、风险偏好和社会背景等，可能与收入和健康有关^[19]。为此诸多文献已经转向使用各种类型的收入冲击来估计收入对健康变动的影响，例如有学者利用彩票中奖金额的随机变化来估计家庭收入对自我报告健康状况的因果影响，确实发现了收入和健康之间的因果关系^[20]。与收入对健康的影响类似，健康对收入的影响也可能由于存在诸如教育、能力、工作经历等混杂因素而严重偏离。那么什么样的变量既与健康密切相关，又是外生的呢？有些学者尝试用和遗传有关的变量作为工具变量来分析健康对收入的影响，采用孟德尔随机化研究分析表明健康和收入之间确实存在因果关系^[1]。

个体健康研究中还有一个常见问题是存在混杂因素同时影响解释变量（也叫处理或暴露）和结果，因此也被称为处理和结果的共同原因。混杂因素可能会导致解释变量对健康的有偏估计。例如前面分析收入对健康的因果关系时，教育往往就是一个混杂因素。人力资本理论认为教育会对收入产生显著的正向影响，受教育水平越高的个体相应的收入水平也越高；很多实证分析也表明教育和健康之间有显著的因果关系^[21-22]，所以教育水平会同时影响收入和健康的变动。而减少混杂的策略是侧重于打破混杂因素与结果的关联（如回归调整）；打破混杂因素与处理的关联（例如基于倾向分数的匹配、调整或加权）；或者同时打破与处理和结果的关联（例如双重稳健方法）^[9]。

样本的自选择偏差在个体健康研究中也颇为常见。在自选择问题中，被解释变量在每个子样本中都可观测，但此时个体的选择行为存在非随机性。例如是否选择迁移，是否选择购买医疗保险，或者是否继续接受高等教育等行为都是内生的，因为不可观测的因素（如个人意识偏好、天赋智力等）会影响人们的决定，这是自我选择的结果。格拉夫娃（Grafova）等人重点关注了邻里环境的自我选择对健康影响的估计为何产生向下偏差的问题，具体来说，健康每况愈下的老人更有可能会选择改变他们的生活方式，改善生活环境，以便为自己的健康投资，这就可能会导致邻里环境对健康影响的结果偏差^[23]。

学者们采用了很多方法来解决个体健康研究中存在的上述问题，这些方法有些是因果推断的方法，有些是常用的计量经济学方法^[24-25]。我们归纳了个体健康研究中的诸多文献，包括准自然实验^[26]以及混杂控制^[9]等方面，还有一些虽然研究对象不是个体健康，但是

涉及内生性问题处理的文献^[27]，最终得到个体健康研究中常见问题及解决办法的综合图，详见图 1。

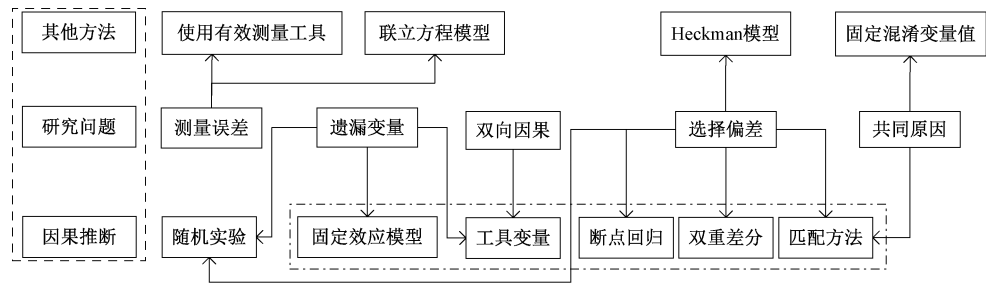


图 1 个体健康研究中的常见问题及解决方法

注：由克雷格（Craig）等人^[26]、马泰（Matthay）等人^[9]、希尔（Hill）等人^[27] 的研究内容综合整理得到。

上述分析表明，个体健康研究中存在的问题往往并不是单一的，可能涉及上述问题中的多个，各类问题综合作用的结果使得探究个体健康研究中的因果关系变得尤为困难。由图 1 可知，对个体健康研究中出现的同一问题，可以采用不同的因果推断方法，那么如何根据研究主题、背景和数据等条件选择合适的因果推断方法，也是研究者亟待解决的问题。因此有必要对个体健康研究中常用的因果推断方法进行归纳、总结，对未来研究提供有益的帮助。

三、因果推断方法

很多学者致力于探讨变量之间的因果关系并估计因果效应，以期根据因果关系给出合理的政策建议。下面主要介绍潜在结果模型和因果效应，以便为后续分析个体健康研究中的因果推断建立更好的理论基础。

1. 潜在结果和因果效应

PO 模型是因果推断重要的理论模型之一，其有三个基本的要素：潜在结果、个体处理稳定性假设（Stable Unit Treatment Value Assumption, SUTVA）以及分配机制（Assignment Mechanism）。考虑一个二值处理（treatment）变量 A ， $a = 1$ 代表处理组，0 代表控制组，那么对于接受处理的研究对象而言，不接受处理时的状态就是一种“反事实”状态，反之亦然，所以 PO 框架又被一些研究者称之为反事实框架（Counter Factual Framework）^[28]。我们用 Y_a 表示假如接受处理 $A = a$ 后的结果，称之为潜在结果；对于每一个 a 值，潜在结果 Y_a 都是一个分布在总体的随机变量，而且每个特定个体的实际结果仅取决于该个体的处理值，这样就排除了个体之间的干扰，这种假设通常被称为个体处理稳定性假设（SUTVA）；在满足 SUTVA 的前提条件下，分配机制就直接决定了个体被分配到处理组或控制组的状态，从而观察到对应的潜在结果。

在实际应用中，因果推断理论是基于因果效应来体现的，总体的平均因果效应（Average Causal Effect, ACE）指的是，假设该总体中所有个体都接受处理平均结果 $E[Y_{a=1}]$ 与假设该总体中所有个体都不接受处理的平均结果 $E[Y_{a=0}]$ 之差，即 $ACE =$

$$E[Y_{a=1} - Y_{a=0}] = E[Y_{a=1}] - E[Y_{a=0}]。$$

但是,苗旺等认为在实际中不可能让所有的个体都接受处理 $a = 1$,再接受对照 $a = 0$;而且对同一个体 i 来说,先接受处理的潜在结果 $Y_{a=1}$ 与后接受处理的潜在结果 $Y'_{a=1}$ 可能也不一致^[29]。也就是说,因果效应是在同一时间(处理前后)对同一个体的对比,它依赖于潜在结果而不仅仅是实际观测结果。然而,这种“潜在结果”在实际生活中普遍不存在,此时随机控制实验就成为识别因果作用最有效的方法。

2. 随机控制实验

统计学家费舍尔(Fisher)提出在所有个体中将处理 A 进行随机分配,使得每个个体只接受处理或者只接受不处理有且仅有在对应处理下的一种结果^①。除了处理的随机化分配外,还需设定随机分配下的处理 A 与潜在结果 Y 以及其他协变量的取值表现无关,即满足 $(Y_{a=1}, Y_{a=0}) \perp A$,这就是著名的“随机控制实验(Randomized Controlled Trials, RCT)”。在 RCT 中,处理的随机性为项目评估提供了最成功和最广泛的科学研究设计基础。在随机化分配下,平均因果作用表示为观测到的结果变量 Y 在处理组 $a = 1$ 与对照组 $a = 0$ 的期望之差,不再含有潜在结果变量 $Y_{a=1}$ 和 $Y_{a=0}$,这就极大地解决了因果推断中的不可识别性。

伊本斯(Imbens)强调了随机实验证据的优势,从某种意义上讲,随机控制实验是最可信的一种政策评估方法,通过随机化,使得已知和未知的混杂因素在处理组和控制组都可以平衡分布,也就是说,随机化所带来的最大好处就在于,它消除了混杂偏倚,使研究者能够量化不确定性^[30]。因此,随机对照实验通常被称为临床实验的黄金标准,在自然科学以及美国、欧洲和其他地方的药物批准过程中发挥着中心作用^[31]。

随机控制实验在健康干预对健康的影响研究方面也有广泛的使用,主要包括工作场所的健康干预、收入转移或现金激励干预以及直接的健康干预。工作场所的健康干预措施包括管理人员的心理健康培训^[32]、工作场所福利项目提供^[33]和心理健康意识培训^[34]等,通常把研究对象随机分配到干预组和对照组,干预实施一定阶段后,再分析健康干预政策的效果,分析结果表明工作场所的健康干预有利于提高劳动者的健康。收入转移或现金激励分为三种情况:一种情况是对处理组给予现金激励,对照组不给^[35];另一种是分为多个组,每组给予的现金激励不同^[36];第三种是总的激励数额相同,但是分发的频率不同^[37]。实验结果表明给予现金激励对健康有积极的影响,给得多比给得少的健康结果要好,激励金额支付频率高比支付频率低对健康的影响要大。直接的健康干预则是通过健康访谈^[38]和基于网络的健康指导^[39]来提升干预组的心理健康。

虽然随机实验的说服力强,但是迪顿(Deaton)和卡特莱特(Cartwright)对随机实验在科学证据层次的突出地位提出了质疑,因为 RCT 有其自身的局限性:①实验的样本容量可能很小,而个体间又可能存在较大差异,所以估计量的精确度不足;②在理想随机实验中,可以随机分配处理组和对照组的成员,但是在实际的研究中,人们往往可以自由选择是否参加,存在自我选择效应;③实验要求控制组完全不受政策的影响,但实际中很难排除政

① 这方面的思想可参考:FISHER R A. Design of experiments [M]. Edinburgh: Oliver and Boyd, 1935

策的溢出效应和替代效应; ④可能存在实验效应, 当人们知道自己处于实验中时, 自我心理和行为就可能发生变化, 从而影响实验结果; ⑤随机实验的系统误差是无法完全避免的, 比如处理组和控制组可能在不同的地点被观察, 在同一天不同时间被观察, 或者由不同的观察员所观察, 这些因素都有可能造成实验结果的偏差。基于以上原因, 虽然随机控制实验在实证研究的各个领域都有广泛应用, 但是由于其过高的实验成本且需花费大量时间, 所以很多学者越来越热衷于使用说服力类似且几乎无成本的准自然实验来替代随机实验进行研究^[40]。

3. 准自然实验

英国医学研究委员会 (Medical Research Council, MRC) 将准自然实验 (Natural Experiment, NE) 定义为: 一项适用于利用处理变化来分析其影响的研究, 其中引起处理的事件、干预措施或政策不受研究者控制。即在准自然实验中, “干预不是为了研究目的而进行的”, 而是“采用一种试图得出因果关系的推断方法来分析处理和结果的变化”^[41]。给定的准自然实验对健康研究的价值取决于一系列因素, 包括受影响人口的规模、可能影响的时间、产生处理差异的过程以及数据收集的实用性等^[26]。如果样本数量足够大, 则可将样本范围内的处理人群和未处理人群 (或处理水平不同的人群) 进行比较, 并可获得关于处理、结果和潜在混杂因素的准确数据, 从而利用准自然实验进行分析研究^[9,42]。但是, 如果缺乏充分的调查数据, 或对处理和结果的评估不准确, 那么就必须根据干预实施的背景谨慎选择准自然实验评估方法, 还应该进行假设检验和公开透明的报告, 提供对于干预和分配过程的清晰描述, 并且最大限度地减少偏差, 提高外部有效性^[43]。

准自然实验的方法已应用于许多领域的经验研究, 如公共卫生、经济学和社会科学等。评估方法包括倾向得分匹配法、工具变量法、双重差分法、断点回归法、固定效应模型等, 如何根据研究主题和数据背景选择合适的评估方法, 也就成为研究人员需要重视的问题。

四、个体健康研究中的准自然实验评估方法与应用

近年来, 准自然实验在社会科学研究中占有越来越重要的地位^[26]。相较于随机实验而言, 其在实验伦理上具有极为重要的优势, 如在研究吸烟对健康影响的过程中, 实验者明知吸烟行为具有成瘾性与危害性, 强行从样本中选取个体要求其接受处理 (即吸烟) 是不道德的; 但可以选取准自然实验的方式, 从社会选取吸烟者群体与不吸烟者群体, 研究其健康差异, 但要考虑到两个群体间是否具有样本自选择问题, 如是否较低健康水平的群体更倾向于不吸烟以维护自身健康。准自然实验的常用评估方法及其在健康领域的经典研究总结如下。

1. 倾向得分匹配

早在 1983 年罗森鲍姆 (Rosenbaum) 和鲁宾 (Rubin) 就提出了倾向得分匹配 (Propensity Score Methods, PSM) 方法, 作为一种近似实验的方法, PSM 常被用于克服横截面样本数据的非随机问题^[44], 尤其是在临床生物医学领域应用更为广泛^[45]。它的基本原理是, 对于接受处理的个体, 找到具有相同可观测特征的未被处理的个体, 通过逆概率加

权^[46-49]、偏差矫正匹配^[50]等方法,比较他们观测结果的差异,达到估计处理效应的目的。使用PSM有一个关键的前提限制,即“强可忽略性处理分配”,该方法假定控制协变量之后,具有相同特征的个体对政策具有相同的反应。换句话说,不可观测因素不影响个体是否接受政策干预的决策,只有在估计倾向得分的回归方程中包含的可观测变量才能影响被暴露于政策的概率。

如果存在样本选择偏差,一般可以采用PSM方法来估计因果效应,该方法广泛应用于研究迁移^[51]、退休^[52]、收入^[53]、低保参与^[25]、工作场所的健康促进措施^[54]等多方面因素对健康的影响。PSM估计因果效应分为三个步骤:首先用所有匹配的协变量估计处理变量的倾向得分,估计方法为二元或多元logit回归;然后根据得分进行匹配;最后根据研究目的计算平均处理效应,或者对剔除未匹配样本后的新数据集进行回归分析。我们以迁移和低保参与对健康的影响为例加以说明。有学者以迁移作为处理变量,采用二元logit回归计算倾向得分,然后利用局部线性回归匹配方法进行倾向得分匹配,同时使用近邻匹配和半径匹配作为稳健性检验,最终的平均处理效应表明迁移者的健康状况要优于未迁移者^[51]。霍萱等人采用PSM方法,研究了低保参与和青少年心理健康之间的因果关系,处理变量是享受低保的状态:进入、离开和保持。研究分为三个阶段,首先使用青少年的个人和家庭等特征以及心理健康得分,采用多元logit模型来预测获得相应福利状态的概率,即倾向得分;然后使用半径匹配方法将每一个在特定处理组的青少年与未接受低保但有相同倾向分数的人进行匹配;最后使用PSM过程中产生的权重回归调整模型来估计各种低保参与状态对心理健康的可能影响。这样的处理过程很好地解决了选择偏差问题^[25]。

但是由于PSM没有处理未观察到的混杂因素,该方法有可能导致有偏差的效果估计。与此同时,如果在使用匹配方法时把样本局限在共同支撑域范围,匹配方法得到的结论仅适用于共同支撑域范围里的样本,而不适用于所有样本。如果共同支撑域样本比较小,那么得到的结论普遍性就相对较差。

2. 工具变量估计

工具变量(Instrumental Variable, IV)也称“辅助变量”,是在模型估计过程中被作为工具使用,以替代模型中与随机误差项相关的随机解释变量,进而估计因果效应。工具变量方法在遗传学领域得到了广泛应用,进一步为孟德尔随机化研究提供了稳健性检验^[55-57]。实证研究中操作IV的方法一般是两阶段最小二乘法2SLS,以回归模型 $Y = \beta_0 + \beta_1 X_1 + \sum_{k=2}^n \beta_k X_k + \mu$ 为例加以说明, $X_k (k = 2, 3, \dots, k, k+1, \dots, n)$ 表示其他控制变量, μ 表示随机干扰项,假定只有第一个解释变量 X_1 为内生变量,找到一个与 X_1 高度相关但是与 μ 无关的工具变量 Z ,即满足 $Cov(X_1, Z) \neq 0$ 且 $Cov(Z, \mu) = 0$,首先用工具变量 Z 对内生解释变量 X_1 进行OLS回归,得到拟合值 $\hat{X}_1$,再用第一阶段回归的拟合值 $\hat{X}_1$ 对被解释变量 Y 进行OLS回归,最终得到两阶段最小二乘法的估计值。关于有效工具变量的选择要满足以下几点重要条件:①相关性,即工具变量与内生解释变量高度相关;②外生性,即工具变量与扰动项不相关;③工具变量要与模型中的其他解释变量无关;④当需要引入多个工具变量

时,这些工具变量之间无关^[42,58]。

在迁移对健康影响的研究中,学者们常采用迁移者的社会网络作为迁移的工具变量,因为它们有助于降低成本,增加移民者的机会,计量分析时常采用一个村的迁移率作为研究指标^[59]。在退休对健康影响研究中学者们通常采用法定退休年龄或养老金领取资格年龄作为退休的工具变量,因为老年人大多集中在这个年龄退休^[60-61];在收入对个体健康影响方面,一些学者以中国新农村养老金计划(NRPS)的推出时间作为工具变量,避免了反向因果与遗漏变量偏差所带来的内生性问题^[24];夸德罗斯-梅纳卡(Cuadros-Meñaca)通过组合移民的迁移率和主要目的地的失业率,构造了失业冲击变量作为替代汇款指标的工具变量^[62]。在教育对健康影响的研究中,学者们大多采用义务教育法规定的义务教育年限作为教育的工具变量^[22,63]。在保险对健康的影响方面,学者们采用了多种工具变量,例如奇尔瓦(Chirwa)等人采用社区的入学率作为保险的工具变量,其理由是社区的入学率越高,保险就越有吸引力,从而增加了个人保险注册的几率^[64];李(Lee)和金(Kim)设计利用26岁截止年龄时保险状态的离散变化作为工具变量,以解决保险状态和精神卫生保健利用之间的内生性问题^[65];而潘杰等采用各城市对参保人群的政府补助比例作为工具变量,结果表明医保制度可以促进参保居民的健康水平^[66]。在工作时间对健康影响的研究方面,万德(Wunder)等人利用德国公共部门法定工作周规定的工作时间作为工具变量来处理工作时间的内生性^[67];沙拉夫(Sharaf)和拉沙德(Rashad)采用建筑、农业和运输工作在11个就业岗位中的普遍程度来衡量工作不稳定性,以克服不同类型工人对不同类型工作有明显的选择性而带来的内生性^[68]。在社会资本对健康影响的研究方面,梁航等人采用结构性社会资本的滞后一期变量为工具变量,原因是结构性社会资本与滞后变量相关,但滞后一期的变量已经发生,可能与本期无关,解决了社会资本和健康之间可能存在的双向因果关系^[69]。

由此我们发现,工具变量法虽然是一个相对简单的估计方法,但是只有当个体对政策反应的异质性不影响决策时,工具变量才是有效的,而且,如何选择合适的工具变量也是研究者们急需解决的一大问题。

3. 双重差分法

双重差分(Difference In Difference, DID)方法是用来估计政策干预和事件处理效应的一个常用方法,被广泛应用于公共卫生政策研究^[70]。这些政策或事件的特点是,它们并不在同一时间影响所有个体,或者对个体的影响并不相同。DID方法评估政策效果的基本思想是通过比较受到影响的群体(处理组)和未受到影响的群体(对照组)的差异来评估政策效果。假设模型为 $Y_{it} = \alpha_0 + \alpha_1 Treat + \alpha_2 Post + \alpha_3 Treat \cdot Post + \varepsilon_{it}$,设定 Y_{it} 表示结果变量,分组虚拟变量 $Treat = 1$ 或 0 分别表示是否对该组样本进行了“处理”;时间虚拟变量 $Post = 1$ 或 0 分别表示“政策实施后”和“政策实施前”; $Treat \cdot Post$ 为分组虚拟变量和时间虚拟变量的交互项,其系数 α_3 反映了政策实施的净效应。使用DID方法的前提条件主要有两点,一是需要使用面板数据,二是需要满足平行趋势假设,即在政策未实施时,处理组和控制组的结果变量随时间变化的路径平行。

在迁移对健康影响的研究中,学者们通常把农村移民设置为处理组,非移民代表对照

组,利用 DID 来估计内部迁移的健康影响,同时控制观察到的特征和未观察到不随时间变化的因素,这两个群体共有的因素可能与内部迁移和健康结果相关^[71]。当退休政策发生改变时,可以用双重差分模型来分析退休政策变动对健康影响的因果效应,退休政策有两种变化:一个是延长退休年龄,一个是降低退休年龄。对于前者,卡里诺(Carrino)等人把因最低养老金计划改变而无法领取国家养老金的妇女作为处理组,把年龄和特征相似的妇女作为对照组,研究了女性领取国家养老金年龄增加6年的健康效应,结果表明提高领取国家养老金年龄会导致抑郁症状的概率增加12个百分点,同时在较低职业等级的女性中,自我报告的医学诊断抑郁症的概率也会增加^[72]。关于降低退休年龄的研究,鲍尔(Bauer)和艾兴伯格(Eichenberger)研究了瑞士建筑业养老金领取资格的政策变化对健康的影响,该变化将退休年龄从65岁降至60岁,实证分析时采用了两种处理方法,一种是将建筑工人作为处理组,其他蓝领工人作为对照组,另一种是令岁数大的建筑工人作为处理组,年轻的建筑工人作为对照组,研究结果表明降低退休年龄并没有改善工人的健康^[73];但是哈尔伯格(Hallberg)等人对于瑞典军人退休年龄由60岁降低到55岁这一政策的研究却得到了相反的结论,该研究以军人为处理组,政府的其他雇员为对照组,研究结果表明提前退休有助于降低死亡率和住院护理比例;结论的差异可能源于行业特点的差异^[74]。养老金收入政策变化对健康的影响也可以采用双重差分的方法进行分析。施纳尔岑贝格(Schnalzenberger)研究了养老金降低对死亡率的影响,处理组为在公共养老金体系中缴费年限超过40年的退休人员,对照组由早期60岁的正常退休人员组成,他们不受这些改革的影响,研究结果表明养老金降低对死亡率没有影响^[53]。工作时间政策的变化对健康的影响也适用双重差分方法。法国1998年通过一个法案,规定从2002年1月起所有私人企业把每周工作时间由39小时降低为35小时,但保持收入不变,伯尼尔(Berniell)和比滕贝克(Bietenbeck)研究了该政策对个体健康的影响,处理变量是2002年雇主是否采用35小时工作制,是取值为1,否则为0,研究结果表明减少工作时间的政策可能会带来重要的健康益处^[4]。除了降低工作时间外,保持收入不变的工作量降低制度,也会对就业者的健康带来有益的影响。布拉特贝格(Bratberg)等人研究了减少55岁以上教师的工作量,但保持相同的工资这一政策对教师健康的影响,处理组为55岁以上的教师,对照组是年龄稍小的教师,不符合减少工作量的条件,研究结果表明工作量的减少会导致男性患病缺勤的减少和心理健康的改善^[75]。税收抵免政策作为收入的外源性变化来源,可以消除或显著减少因与收入相关的冲击而导致的遗漏变量偏差,并可以用来估计接受收入增加对接受处理者的健康处理效果,为收入对健康的因果效应提供了证据。如伦哈特(Lenhart)研究了1996年美国的税收抵免政策对个体健康的影响,该政策使有两个或两个以上子女家庭的最高福利增加了一倍多,而有一个合格子女家庭的福利仅略有增加,以两孩以上家庭为处理组,其他家庭为对照组,采用双重差分方法的分析结果表明收入使受影响的户主报告健康状况极佳或非常好的可能性提高了6.9至8.9个百分点^[19]。

不容忽视的是,DID方法同样有一定的局限性:①DID方法以面板数据模型为基础,对数据要求更加苛刻;②由于存在个体时点效应,所以平行趋势假设很难满足;③DID方法假

定环境因素冲击对处于相同环境中的个体会产生相同的影响,但实际中,处理组和对照组个体可能因为某些不可观测因素的影响,使得其在面临相同的环境因素冲击时做出不同的反应;④还有一点限制因素是溢出效应的风险,即处理组群体可能会影响到对照组群体^[42,70]。

4. 断点回归设计

断点回归设计(Regression Discontinuity Design, RDD)早在1960年就被提出^[76],主要应用于医学领域^[77-78],直到20世纪90年代末之后才被国外大量应用于经济研究。RDD是一种类似于随机实验的方法,李(Lee)认为在随机实验不可得的情况下,断点回归能够避免参数估计的内生性问题,从而真实反映出变量之间的因果关系^[79]。通过RDD来估计因果效应主要分为两部分:首先是图形分析,这一步骤的完成需要满足一些前提条件。①RDD的数据需要包含三个基本变量,分别为驱动变量、断点和观测结果。驱动变量是个体的一个连续特征变量,匹配变量的值是否大于断点将决定个体是否接受处理;断点用于决定个体是否接受处理的阈值;而观测结果则是个体接受处理或未接受处理的观测结果。②驱动变量的值在断点附近无法被准确操作。③断点的选择不受驱动变量的影响。④除了处理状态在断点处发生跳跃式变化外,其他未处理的个体特征变量在断点处没有显著差异。其次是估计因果效应。如果是精确断点回归,可以利用局部线性回归来识别平均处理效应^[80],如果是模糊断点回归,则可以结合工具变量,利用两阶段最小二乘法(2SLS)来估计因果效应^[81-82]。此外,在断点回归分析中进行全局高阶多项式回归往往会受到阶数敏感性、置信区间覆盖率等因素的影响,因此更建议研究人员基于局部多项式或二次多项式进行估计^[83]。作为RDD方法的扩展,回归拐点设计(Regression Kink Design, RKD)同样受到学者的关注,尤其适用于福利政策相关的因果效应研究。在RKD方法中,被干预可能性的斜率在拐点处发生了变化,从而导致赋值函数的一阶导数不连续。该模型使用也具有一定的前提要求,即在拐点处不存在对驱动变量的操纵,而且其他协变量在拐点处的斜率不应该发生变化。

在个体健康研究领域采用断点回归分析方法最多的是退休对健康的影响,通常是以法定退休年龄作为断点,研究内容包括退休对自己健康的影响和伴侣的健康溢出效应。大部分研究表明退休对健康有积极的影响^[84],减轻与工作相关的压力和紧张、增加睡眠时间、更频繁的体育锻炼和户外运动似乎是退休影响健康的关键机制^[5];但有些研究表明退休对健康有不利的影 响,可能是由不良的健康行为所引起的^[12],雷晓燕等则是认为正常年龄退休对男性有显著的负面影响,对女性没有明显影响,可能是因为女性比男性更易于调整心态^[85]。伴侣间溢出效应表明男性配偶的退休改善了女性的心理健康^[86],很可能是通过增加社交和锻炼的频率来实现的^[87],但也有个别文献认为主观健康受到配偶退休的负面影响,原因是随着伴侣退休,饮酒频率和强度显著增加,适度体育活动显著减少^[88]。也有部分学者采用模糊断点回归设计,分析了义务教育法改革带来的额外学校教育年限对心理健康的影响,如阿文达诺(Avendano)等人以改革后的义务教育年限作为断点,研究结果表明改革并没有直接改善心理健康,延长义务教育的持续时间会通过增加教育程度以外的渠道影响心理健康^[89];扬克(Janke)等以最低离校年龄改革前后的出生年月作为驱动变量,以1957年9

月该出生年份为断点, 研究结果表明额外的一年学校教育对大多数慢性健康状况的患病率没有统计上可识别的影响^[90]。还有学者研究了教育质量对健康的影响, 如钟海认为对教育质量产生重大冲击的典型事件是“文化大革命”, 对于1947年之前出生的人, 在1966年“文化大革命”开始时普遍已经高中毕业, 不会受到“文化大革命”的巨大冲击, 1959年之后出生的人在1976年“文化大革命”结束时正处在初高中阶段, 教育制度已经恢复, 同样不会被“文化大革命”所影响, 因此, 钟海以出生年份作为驱动变量, 以1947和1960年作为断点, 分别对应“文化大革命”开始年份和结束年份, 研究结果表明更好的教育会改善自我评估健康状况, 降低残疾的可能性, 但对身体功能不佳或身体不适没有显著影响^[91]。梅内塞斯-菲略 (Menezes-Filho) 和波利蒂 (Politi) 在研究私人健康保险的因果效应时, 以所得税收入上限作为扭结点, 收入取对数作为驱动变量, 因为收入高于扭结点的个人往往会增加私人保险购买金额, 从而降低自己的税收等级, 所以私人保险持有率在扭结点附近有明显的斜率变化, 而这一变化会进一步影响个人的健康行为和健康结果, 最终研究表明私人保险对预防服务的使用、健康结果、体育锻炼和戒烟都有积极影响^[92]。另外有学者研究了健康保险在年轻人心理健康保健中的作用, 由于健康保险状况在26岁左右呈离散变化, 因此作者以26岁作为断点, 26岁左右的月份作为驱动变量, 研究结果表明健康保险增加了年轻人的心理健康保健^[65]。还有学者以贫困线作为断点, 研究了最低生活保障制度对健康的影响, 发现该制度产生了显著的福利污名化效应, 对受助者的心理健康和福祉产生了负面影响^[93]。

虽然RDD是准实验方法中最具有可信性的方法, 但是RDD方法并不能完美地模拟随机实验的普遍性。之所以如此首先是阈值的选择问题, 阈值必须是一个真正的随机因素, 一个本身与结果无关的因素。其次, 断点回归只使用了接近处理临界值的个体数据, 即估计了局部的平均处理效果, 因此与随机实验相比, 由于样本量减少, RDD方法的作用是有限的。对于RKD方法而言, 通常需要比RDD更大的样本量才能保证估计值的准确, 但是与RDD类似, RKD只能估计拐点附近的因果效应, 无法简单地推广至总体。

5. 个体固定效应模型

个体固定效应模型 (Fixed Effects Model) 分析比较了某一时间段内同一个人或群体内的多个观察结果, 认为不可观测且不随时间变化的变量存在, 且与其他可观测的解释变量通常是相关的。因此不能把这一变量当作干扰项的一部分, 而必须将其等同于其他解释变量进行处理。假设固定效应方程为 $y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \delta X_{it} + \varepsilon_{it}$, 其中, α_i 表示不随时间变化的不可观测个体特征, γ_t 为时间效应, D 是处理变量, X 为控制变量, ε_{it} 为残差项, i 表示个体, t 表示时间。以 $t=0$ 和 $t=1$ 对上式做差就可以消除个人特征 α_i , 得到 $y_{i1} - y_{i0} = \gamma_1 - \gamma_0 + \beta(D_{i1} - D_{i0}) + \delta(X_{i1} - X_{i0}) + \varepsilon_{i1} - \varepsilon_{i0}$ 。从这一方程可以看出, 个体固定效应控制住了所有个人层面不随时间变化的异质性, 这样就解决了不随时间而变但随个体而异的遗漏变量问题。

个体固定效应模型在健康研究中得到了广泛的应用。一些学者采用个体固定效应模型, 通过控制未观察到的和不随时间变化的因素, 研究迁移对个体健康的影响, 研究结果揭示了农村人口向城市迁移对中年人而非老年人自评健康的积极影响^[94]。还有学者采用个体固定效用模型研究了退休对健康的影响, 除了把年龄、年龄平方、受教育年限、婚姻状况、家庭

规模和波度作为控制变量外,未观察到的不随时间变化的异质性由个体固定效应控制,研究结果表明退休状态对妇女自我报告的健康、身体和精神健康结果有积极和显著的影响,较长的退休时间会带来明显的额外健康福利^[95]。沃森(Watson)和奥斯伯格(Osberg)在研究收入预期和收入焦虑对心理健康的冲击时,使用个体固定效应和工具变量相结合的方法,以失业率为工具变量,同时控制了教育、关系压力、家庭规模、居住地区和年龄等可能与心理痛苦有关的因素,研究结果表明,经济损失对心理痛苦的影响要大于同等规模的收益,而且随着损失规模的增加对心理健康的边际影响会逐渐递减^[96]。哈洛宁(Halonen)等人使用瑞典纵向职业健康调查研究,在模型设计中加入了性别、遗传特征和个性这些不随时间变化的因素,利用条件 Logistic 回归进行固定效应分析,研究结果表明,如果一个人每周工作时间超过正常时间,那么长时间的通勤会导致身体缺乏活动,增加睡眠问题的风险^[97]。通常我们认为受教育年限越长身体越健康,但是这种关联很可能会被未观察到的因素所混淆,为此,藤原(Fujiwara)和川内(Kawachi)使用双固定效应模型来确定教育对健康和健康行为的因果影响,在这一模型中,因为样本选择全部为有共同生活环境的双胞胎,就可以排除遗传能力、性格或早期家庭环境等可能影响教育和健康的未知共同因素,研究结果表明,从个体层面考虑,教育与健康结果正向相关,但是在固定效应分析中,尽管二者仍然相关,结果却并不显著,这也就意味着,个体层面对教育和健康关系的研究确实存在着某些未知变量的干扰^[98]。

虽然个体固定效应模型优于 PSM,可以控制未观察到的不随时间变化的个体特征,但是仍然有一定的局限性。为了达到控制个体不可观测且不随时间变化的因素,个体固定效应模型使用了个体内变化的信息。如果变量有充分的个体内变化信息,固定效应模型能够有效地处理缺失个体固定效应造成的缺失变量误差;但如果变量个体内变化很少,其系数的估计方差就可能很大且不显著。因此在使用面板数据前,对于每个变量,尤其是所关注的变量的信息来源要有充分的了解,如果变量的系数在加入固定效应后变得不显著,不能简单地推断它没有因果影响,而是要具体情况具体分析。

我们对以上各个典型方法的适用条件、优缺点及应用于健康领域研究的典型文献进行了总结,见表 1。

6. 多方法的综合使用

在前面的内容中,我们总结出了个体健康研究中针对不同问题的具体解决方法,但是,一些研究往往涉及个体健康研究中存在的多个问题,只使用某一种方法可能只能解决某一类问题,不能解决研究中存在的多个问题,存在一定的局限性,并不能给出可靠的因果结论,因此有必要把多种方法结合使用。

常用的结合方式是工具变量与固定效应结合(FE-IV)。如果一个研究问题解释变量是内生的,还存在不随时间变化的遗漏变量,则应该把固定效应与工具变量相结合。FE-IV 方法在个体健康研究中得到了广泛的应用,主要应用于分析退休对健康的因果效应,有的文献具体标明了 FE-IV,有的文献体现的是固定效应与模糊断点回归的综合使用,但本质上是一样的。例如有学者采用 $E_{it} = I(\text{Age}_{it} \geq \text{Age}_t^p)$ 作为工具变量(如果当前年龄 Age_{it} 大于或等于

表1 个体健康研究中常用的因果推断方法

研究方法	适用条件	优点	缺点	涉及的健康影响因素及例证
随机控制实验	随机分配处理组与对照组,对未观察到的混杂因素进行随机对照	从某种意义上讲,随机受控实验是最可信的一种评估方法;处理 X 与潜在结果及协变量的取值无关,可以保证潜在结果与处理分配 X 独立;通过随机化,可以使得已知的和未知的混杂因素在处理组和未处理组都达到平衡	样本容量很小时,估计量的精确度不足;存在自我选择效应;很难排除政策的溢出效应、替代效应;可能存在实验效应;系统误差无法完全避免;实施成本高,时间长	收入 (Undurraga et al, 2016; Bradley & Saunders, 2020) ^[35-36] ; 工作 (Song & Baicker 2019; Tan et al, 2021) ^[33-34]
倾向得分匹配	适用于截面数据和纵向数据。必须满足很强的假设前提,并且要具有相当的数据量	不需要对可观测因素的条件均值函数和不可观测因素的概率分布进行假设;当许多因素与处理相关时,允许进行平衡比较,假定所有因素均已被精确测量,不会有无法测量的混杂因素	极强的前提假设;不能为所有的处理组个体找到控制组个体;数据量要求极大;结果的稳健性受到多种挑战	迁移 (Wang & Hu 2019) ^[51] ; 退休 (Barban et al, 2020) ^[52] ; 收入 (Schnalzenberger, 2016) ^[53] ; 工作 (Huber et al, 2015) ^[54]
工具变量	适用于截面数据、面板数据和纵向数据。工具变量要和内生解释变量相关,但是与扰动项和其他解释变量无关	在实验个体的处理不符合随机分配的情况下最容易被采用;可以很好地解决遗漏变量和双向因果问题	要找出满足条件的工具变量并不容易;只有当个体对政策反应的异质性并不影响参与决策时,工具变量才能被识别,但这是一个很强的假定	收入 (Chen et al, 2019) ^[24] ; 社会资本 (Liang et al, 2020) ^[69] ; 退休 (Gorry & Slavov, 2021) ^[61] ; 教育 (Albar-rán et al, 2020) ^[22] ; 迁移 (Pan & Dong, 2020) ^[59] ; 保险 (Lee & Kim, 2020) ^[65] ; 工作 (Cygan-Rehm & Wunder, 2018) ^[67]
双重差分	适用于面板数据和纵向数据。需要满足平行趋势假设	控制了未观测的和可观测的不随时间变化特征的异质性,因此不太容易因未测量的混杂因素或测量误差而导致偏差	数据要求更加苛刻;个体时点效应未得到控制;未考虑个体所处的环境对个体的不同影响;存在溢出效应的风险	社会资本 (Grafova et al, 2014) ^[23] ; 收入 (Schnalzen-berger, 2016) ^[53] ; 迁移 (Lu et al, 2020) ^[71] ; 退休 (Bauer & Eichenberger, 2021) ^[73] ; 工作 (Berniell & Bietenbeck, 2020) ^[4]
断点回归	适用于截面数据和纵向数据。除处理变量以外,其他影响结果的协变量在分界点两侧不应有明显的断点;分组变量是连续的	RDD 类似于随机实验,因果推断最为清晰,结果最为可信,假设的可检验性也最强	RDD 方法不能完美地模拟随机实验的普遍性;阈值必须是一个真正的随机因素。因此,该方法可适用的范围较小	收入 (Qi & Wu, 2018) ^[93] ; 教育 (Janke et al, 2020) ^[90] ; 保险 (Lee & Kim, 2020) ^[65] ; 退休 (Rose, 2020; Zang, 2020) ^[84,87]
个体固定效应	适用于纵向数据。分析比较同一时间段内同一个人或群体内的多个观察结果,并揭示该策略导致的平均结果变化	由于每个人或组都会随时间与自己进行比较,因此可以消除即使经过测量仍保持不变的人或组之间的差异,并且不会混淆结果	容易受到不可观测的时变混杂因素的影响;剔除了个体之间所有截面数据的变化;随着时间的流逝容易出现测量误差	收入 (Watson & Osberg, 2019) ^[96] ; 工作 (Halonen et al, 2020) ^[97] ; 教育 (Fujiwara & Kawachi, 2009) ^[98] ; 退休 (Zhu, 2016) ^[95]

资料来源:见参考文献 [9]、[26]、[42]、[58] 等。

法定退休年龄 Age_i^p ， $E_{it} = 1$ 取值为 1，其他情况取值为 0），采用具有个体固定效应的两阶段最小二乘模型来分析退休对健康的影响，第一阶段以 E_{it} 为工具变量，以退休为因变量，采用个体固定效应模型得到退休的估计值，第二阶段把退休的估计值作为自变量，采用个体固定效应模型分析该变量对退休的影响^[95]。

另一种常用的结合方式是 PSM 与 DID 相结合，即 PSM-DID，用 DID 进行因果推断有一个基本假设：平行趋势假设，即组间不同的混杂变量不随时间改变，随时间改变的混杂变量在各组之间完全相同。进行实证分析时“平行趋势假设”检验并不一定能通过，如果不能通过则需要先进行 PSM。PSM-DID 模型的实现主要包括两个步骤：第一步，使用 PSM 模型，依据倾向得分为处理组寻找相似度尽可能高的控制组个体，使得控制组与处理组满足平行趋势假设；第二步，使用 DID 模型，通过两次差分处理个体效应和时间效应，从而识别政策冲击带来的净效应。PSM-DID 在个体健康研究中同样应用广泛，主要应用于迁移、医保政策、就业对健康的因果效应研究中。例如一些学者认为迁移组和非迁移组的混杂变量并不满足“平行趋势假设”，因此首先采用倾向得分匹配来减轻选择性偏差，并获得更具可比性的处理组和对照组，然后通过比较迁移组和非迁移组之间的健康变化来估计迁移对健康的平均处理效果^[8]。医保政策对健康的影响主要包括“流动人口基本公共卫生和计划生育服务均等化计划”对迁移者健康的影响^[99]，城乡居民医疗保险一体化对居民医疗服务利用的影响^[100] 和政府健康干预政策对老年人健康影响^[101-102]。上述研究均采用 PSM-DID 方法研究了健康相关政策的影响；斯蒂芬（Stephan）等人则是利用 PSM-DID 对工作和健康之间的因果关系展开了深入研究，首先使用 PSM 来创建可比较的个体经营者和员工群体，然后采用双重差分方法来估计个体就业者和员工群体间的健康差异^[103]。

根据以上总结不难发现，由于健康影响因素的多样化和因果推断方法的普及，基于不同的研究数据和背景，选择合适的研究方法解决问题就显得极为重要。

7. 基于主题的个体健康因果推断方法概括

因果推断在个体健康研究领域被广泛应用，以评估政策干预和其他处理效应，而个体健康水平的变动通常是社会资本、收入、教育、保险、迁移、退休、工作等多重因素作用的结果，所以我们从影响健康的相关因素考虑，对各个不同因素的适用方法进行总结概括。表 2 展示了个体健康研究领域不同影响因素常用的因果推断方法。

表 2 各方法涉及的健康影响因素							
研究方法	社会资本	收入	教育	保险	迁移	退休	工作
随机控制实验		√					√
倾向得分匹配		√		√	√	√	√
工具变量	√	√	√	√	√	√	√
双重差分	√	√		√	√	√	√
断点回归		√	√	√		√	
个体固定效应模型		√	√	√		√	√

注：表中所列仅为常用因果推断方法，以及实证研究中较为常见的健康影响因素，并非全部研究。

收入作为影响健康的主要因素，同时涵盖了所有的因果推断方法。许多学者利用随机控制实验研究现金转移/激励对个体健康的影响^[35-37]；施纳尔岑贝格使用 PSM-DID 方法研究

了养老金收入降低对死亡率的影响^[53]；有学者以中国新农村养老金计划（NRPS）的推出时间作为工具变量，避免了反向因果与遗漏变量偏差所带来的内生性问题^[24]；也有学者使用RDD方法，研究了最低生活保障制度对健康的影响^[93]；沃森和奥斯伯格则是使用个体固定效应和工具变量相结合的方法分析了收入变动和健康之间的因果关系^[96]。

关于工作对健康的影响，主要涉及工作场所的健康干预和工作政策变化两个方面。许多学者利用RCT^[32-34]或者PSM方法^[54]分析健康干预政策的效果；一些学者使用工具变量方法来解决工作时间对健康影响的内生性问题^[67-68]，伯尼尔和比滕贝克利用DID方法研究了法国工作时间政策变化对个体健康的影响^[4]，而哈洛宁等人使用固定效应模型证明了长时间通勤的消极作用^[97]。

研究退休对健康的影响时最常用的是RDD方法，通常是以法定退休年龄作为断点，研究退休对自己健康的影响^[5,12,84,104]和伴侣溢出效应^[86-88]。还可以采用法定退休年龄或养老金领取资格年龄作为退休的工具变量^[60-61,85]或者采用个体固定效用模型^[95]进行分析。当退休政策发生改变时，则可以用双重差分模型来分析退休政策变动对健康影响的因果效应^[72-74]。

此外，关于社会资本、教育、保险、迁移等因素对健康的影响研究也十分广泛。在研究社会资本对健康的影响时，梁航等人采用工具变量解决了社会资本和健康之间可能存在的双向因果关系^[69]，而格拉夫娃等人则是通过使用双重差分方法分析了社区环境对健康的影响^[23]。在教育对健康的影响研究中，通常我们认为受教育年限越长身体越健康，但是藤原和川内使用双固定效应模型研究发现，个体层面对教育和健康关系的研究确实存在着某些未知变量的混淆^[98]。为此，学者们大多采用义务教育法规定的义务教育年限作为教育的工具变量^[22,63]，并结合RDD进行研究^[89-90]。在保险对健康的影响方面，很多学者采用工具变量解决保险状态的内生性问题^[64]，并结合断点回归设计研究保险对健康水平的影响^[65]，而程令国等则是采用固定效应模型控制了不可观测的个人、家庭层面的异质性，并使用PSM-DID进一步减少估计偏差，证明了新农合政策对参保者健康水平的积极效应^[105]。在迁移对健康的影响研究中，学者们常采用迁移者的社会网络作为迁移的工具变量^[59]，还有学者采用双重差分模型^[71]，或者利用PSM方法分析迁移者的健康状况^[51]，还可以采用个体固定效应模型控制未观察到的不随时间变化的因素^[94]，以此来研究迁移对个体健康的影响。

通过对以上基于主题的个体健康因果推断方法的总结概括可以发现，对收入、退休、工作等因素的分析几乎涵盖了所有常用的因果推断方法，而很少有因果推断方法涉及对社会资本的研究。可能的原因主要在于因果推断方法被广泛用于评估政策干预和其他处理效应，有相当多关于收入、退休和工作变动的政策，却很少有社会资本的相关政策，而且，社会资本是一个过于宽泛的概念，表现为个体在一定的社会关系网络中获取信息和资源的能力，包括了邻里关系、社区参与等多个方面，很难判断社会资本这一因素对个体健康的具体因果效应，但是这也为我们以后的研究提供了新的发展方向，或许可以进一步扩展个体健康因果推断的研究领域。

五、未来研究的方向

自 21 世纪以来,以“实验”思想为基础的定量实证分析方法已经成为健康研究领域的前沿热点。不管是建立有针对性设计的“随机控制实验”,还是对基于政策变动或外生事件所产生的“准自然实验”进行的研究,其共同目标都是对健康相关政策的有效性和科学性作出准确的评估,这一类典型的因果推断研究方法也进一步证实了诸多影响因素与个体健康之间的因果关系并非简单的相关关系。从 Web of Science 的检索结果和本文的研究分析中可以看出,因果推断方法在现今及未来较长的时间内,仍然是主流的健康政策评估方法,这一点也得到了众多学者的肯定。

尽管因果推断方法在个体健康研究中得到了广泛的应用,但是该方法在实际应用阶段仍然存在一定的局限性,主要表现在两个方面:一方面是由样本数据的质量和数量带来的可靠性和外部有效性问题。无论是随机控制实验,还是准自然实验,都需要满足一定的前提条件,例如样本的随机分配、平行趋势假设、共同支撑假设等,但是现有的数据未必能满足这些条件,即使变量之间存在因果关系,也未必能得到验证,或者尽管可以在样本范围内得到明确的因果关系,也通过了异质性和稳健性检验,由于样本数量较小,估计结果的外部有效性也会令人质疑。另一方面是分析过程中存在的问题。在实证研究中,通常是假设某一因素和健康之间存在因果关系,然后使用因果推断的方法进行验证,这些因果关系可能源于学者们已有的研究经验,然而可能还存在一些经验之外的因果关系,也就是说,因果推断方法更侧重于验证因果关系。因此,未来研究中首先要着重关注数据的质量和数量,为研究奠定坚实的数据基础,然后采用机器学习与因果推断相结合的方法,解决传统因果推断中存在的问题。

现有的研究一般基于调查数据,样本量往往只有几千或者几万,十几万的数据集都比较少见,但是一个政策可能会涉及几千万,甚至上亿人,通过少量样本推断出的因果效应在整个受众群体上未必是成立的,也就是说很难满足外部有效性。随着信息技术和大数据技术的发展,居民健康相关的信息正以前所未有的速度产生和储存,这些信息可能来自医疗、社会保险、就业和收入、税收、社交数据等,还有政府和机构组织的各种调查和普查数据,每类数据只包含健康及健康影响因素的某一局部方面,例如税务局掌控着每个纳税人的纳税记录,是研究个人所得税改革对健康影响的理想解释变量来源,但是纳税人的健康数据由医疗部门管理,这两者没有交集,因此不能合并为一个可供研究的数据集,只能通过问卷调查进行分析,使研究的可靠性大大降低。由于这些数据分别被不同的机构储存,各机构之间的信息相互独立,没有形成有效的数据共享机制,使得不同的人只能看到健康及其影响因素的某个侧面,难以对健康及健康相关因素进行全局把握。为此,政府应该在保障居民个人信息安全性、私密性的前提下,合理整合居民健康及影响因素的相关数据,形成健康大数据系统,为科学研究提供数据支持,通过研究诸多影响因素和健康之间的因果效应,为政府出台科学的健康政策提供依据。

如果有健康大数据作为支持,那么很多机器学习的分析方法就可以应用于个体健康研究中的因果推断。不同于传统因果推断方法需要基于反事实的潜在结果来定义因果关系,机器

学习以数据为基础,通过精准的预测或干预,将“非随机化”的观测样本尽可能向“随机化”实验靠拢,由此可以从传统数据中发现不易察觉的因果关系^[106]。另外,机器学习还具有优秀的样本和变量选择能力,例如:①在处理组和对照组之间建立平衡,包括三种方法:倾向得分匹配方法,平衡处理组和对照组之间的分布混杂;逆概率加权,即直接优化观察值的权重,从而使处理组和对照组的协变量平均值相同^[107];直接生成处理组和对照组,并使其协变量自动平衡,其算法包括“遗传匹配”算法^[108]、整数规划优化算法^[109-110]。②评估所谓的干扰模型(如倾向得分,或结果的条件期望),用机器学习方法可以分别估计倾向得分^[111]或潜在结果^[112],还可以采用双稳健方法把条件结果期望的估计与倾向得分的估计相结合^[113],以及将条件结果期望的估计与协变量平衡相结合^[114]。③在具有大量协变量的情况下使用机器学习进行变量选择,贝洛尼(Belloni)等人提出了一种提供原则性变量选择的解决方案——“双重套索”,同时考虑了协变量结果和协变量处理分配关联^[115]。除此之外,还可以利用机器学习交叉验证的思想,将数据集分为训练集和测试集,在训练集中使用传统因果推断方法识别因果效应,然后在测试集中进行多次重复测试,进一步验证识别结果的外部有效性。

虽然目前机器学习技术尚未成为经济学研究的主流方法,对整个社会科学因果识别研究范式的冲击有限,但随着未来经济研究领域的发展,可提供样本量的增多,将机器学习方法引入因果关系推断,不仅可以改进因果推断的计量方法,其在分类、预测、验证等方面的优异表现也将会进一步丰富研究者的分析手段和研究范畴,使得机器学习和因果推断互相补充,相互促进。

六、结论

本文以个体健康研究为主题,通过对已有文献的分析和总结,发现个体健康研究中存在着测量偏误、遗漏变量、互为因果、共同原因和选择偏差等诸多问题,随机控制实验、倾向得分匹配、工具变量、双重差分、断点回归设计、固定效应模型等因果推断方法,在社会资本、收入、教育、保险、迁移、退休、工作等因素对个体健康的因果效应研究中均得到了广泛的应用,每一种方法都有其优缺点,在实证研究中应根据研究主题和应用条件进行合理选择,有时需要综合应用多种因果推断方法。本研究为健康政策制定者提供了更加科学合理的决策依据,对于提高居民健康水平具有十分重要的社会意义,还有助于研究者在个体健康研究领域中选择合适的因果推断方法,或综合应用已有的方法。随着大数据技术的发展,在未来的研究中,应加强机器学习和因果推断方法的结合,丰富已有的因果推断工具,以保障研究结果的稳健性。

参考文献:

- [1] HARRISON S, DAVIES A R, DICKSON M. The causal effects of health conditions and risk factors on social and socioeconomic outcomes: mendelian randomization in UK Biobank [J]. International Journal of Epidemiology, 2020, 49 (5): 1661-1681.

- [2] LAMU A N, OLSEN J A. Yes, health is important, but as much for its importance via social life: the direct and indirect effects of health on subjective well-being in chronically ill individuals [J]. *Health Economics*, 2018, 27 (1): 209–222.
- [3] CAI L. The effects of health on the extensive and intensive margins of labour supply [J]. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 2021, 184 (1): 87–117.
- [4] BERNIELL I, BIETENBECK J. The effect of working hours on health [J]. *Economics and Human Biology*, 2020, 39 (4): 1–25.
- [5] BINH TRAN D, ZIKOS V. The causal effect of retirement on health: understanding the mechanisms [J]. *Australian Economic Review*, 2019, 52 (4): 427–446.
- [6] KOUROUKLIS D, VERROPOULOU G, TSIMBOS C. The impact of wealth and income on the depression of older adults across European welfare regimes [J]. *Ageing and Society*, 2020, 40 (11): 2448–2479.
- [7] ATALAY K, ZHU R. The effect of a wife’s retirement on her husband’s mental health [J]. *Applied Economics*, 2018, 50 (43): 4606–4616.
- [8] LONG C, HAN J, LIU Y. Has rural-urban migration promoted the health of Chinese migrant workers? [J]. *International Journal of Environmental Research and Public Health*, 2020, 17 (4): 1–23.
- [9] MATTHAY E C, HAGAN E, GOTTLIEB L M. Alternative causal inference methods in population health research: evaluating tradeoffs and triangulating evidence [J]. *SSM-Population Health*, 2020, 10 (1): 1–9.
- [10] LI Z, FU C, YANG F. Prevalence and risk factors of hypertension for the middle-aged population in China: results from the China Health and Retirement Longitudinal Study (CHARLS) [J]. *Clinical and Experimental Hypertension*, 2019, 41 (1): 80–86.
- [11] YAO S S, CAO G Y, HAN L. Prevalence and patterns of multimorbidity in a nationally representative sample of older Chinese: results from the China Health and Retirement Longitudinal Study [J]. *Journals of Gerontology. Series A: Biological Sciences and Medical Sciences*, 2020, 75 (10): 1974–1980.
- [12] FENG J, LI Q, SMITH J P. Retirement effect on health status and health behaviors in urban China [J]. *World Development*, 2020, 126 (2): 1–10.
- [13] VANAJAN A, BÜLTMANN U, HENKENS K. Why do older workers with chronic health conditions prefer to retire early? [J]. *Age and Ageing*, 2020, 49 (3): 403–410.
- [14] LU W J, ZHANG A W, MOSSIALOS E. Parental migration and self-reported health status of adolescents in China: a cross-sectional study [J]. *EClinicalMedicine*, 2020, 22 (5): 1–7.
- [15] NOH J W, KWON Y D, LEE L J. Gender differences in the impact of retirement on depressive symptoms among middle-aged and older adults: a propensity score matching approach [J]. *PLoS ONE*, 2019, 14 (3): 1–9.
- [16] DOIRON D, FIEBIG D G, JOHAR M. Does self-assessed health measure health? [J]. *Applied Economics*, 2015, 47 (2): 180–194.
- [17] ARNI P, DRAGONE D, GOETTE L. Biased health perceptions and risky health behaviors: theory and evidence [J]. *Journal of Health Economics*, 2021, 76 (2): 1–26.
- [18] DONG Y. How health insurance affects health care demand: a structural analysis of behavioral moral hazard and adverse selection [J]. *Economic Inquiry*, 2013, 51 (2): 1324–1344.
- [19] LENHART O. The effects of income on health: new evidence from the earned income tax credit [J]. *Review of Economics of the Household*, 2019, 17 (2): 377–410.
- [20] KIM S, KOH K. The effects of income on health: evidence from lottery wins in Singapore [J]. *Journal of Health Economics*, 2021, 76 (2): 1–15.
- [21] HECKMAN J J, HUMPHRIES J E, VERAMENDI G. Returns to education: the causal effects of education on earnings, health, and smoking [J]. *Journal of Political Economy*, 2018, 126 (s1): 197–246.

- [22] ALBARRÁN P, HIDALGO-HIDALGO M, ITURBE-ORMAETXE I. Education and adult health; is there a causal effect? [J]. *Social Science and Medicine*, 2020, 249: 2–9.
- [23] GRAFOVA I B, FREEDMAN V A, LURIE N. The difference-in-difference method; assessing the selection bias in the effects of neighborhood environment on health [J]. *Economics and Human Biology*, 2014, 13 (1): 20–33.
- [24] CHEN X, WANG T, BUSCH S H. Does money relieve depression? evidence from social pension expansions in China [J]. *Social Science and Medicine*, 2019, 220: 411–420.
- [25] HUO X, GAO Q, ZHAI F. Effects of welfare entry and exit on adolescent mental health; evidence from panel data in China [J]. *Social Science and Medicine*, 2020, 253: 1–10.
- [26] CRAIG P, KATIKIREDDI S V, LEYLAND A. Natural experiments; an overview of methods, approaches, and contributions to public health intervention research [J]. *Annual Review of Public Health*, 2017, 38 (1): 39–56.
- [27] HILL A D, JOHNSON S G, GRECO L M. Endogeneity; a review and agenda for the methodology; practice divide affecting micro and macro research [J]. *Journal of Management*, 2021, 47 (1): 105–143.
- [28] YAMADA K, KUROKI M. New traffic conflict measure based on a potential outcome model [J]. *Journal of Causal Inference*, 2019, 7 (1): 1–19.
- [29] 苗旺, 刘春辰, 耿直. 因果推断的统计方法 [J]. *中国科学: 数学*, 2018 (12): 1753–1778.
- [30] IMBENS G W. Better LATE than nothing; some comments on Deaton (2009) and Heckman and Urzua (2009) [J]. *Journal of Economic Literature*, 2010, 48 (2): 399–423.
- [31] BOTHWELL L E, GREENE J A, PODOLSKY S H. Assessing the gold standard; lessons from the history of RCTs [J]. *New England Journal of Medicine*, 2016, 374 (22): 2175–2181.
- [32] MILLIGAN-SAVILLE J S, TAN L, GAYED A. Workplace mental health training for managers and its effect on sick leave in employees; a cluster randomised controlled trial [J]. *The Lancet Psychiatry*, 2017, 4 (11): 850–858.
- [33] SONG Z, BAICKER K. Effect of a workplace wellness program on employee health and economic outcomes; a randomized clinical trial [J]. *Journal of the American Medical Association*, 2019, 321 (15): 1491–1501.
- [34] TAN L, HARVEY S B, DEADY M. Workplace mental health awareness training [J]. *Journal of Occupational & Environmental Medicine*, 2021, 63 (4): 311–316.
- [35] UNDURRAGA E A, BEHRMAN J R, LEONARD W R. The effects of community income inequality on health; evidence from a randomized control trial in the Bolivian Amazon [J]. *Social Science and Medicine*, 2016, 149 (2): 66–75.
- [36] BRADLEY C J, SAUNDERS H G. Impact of cash incentives for low-income individuals to seek a primary care visit on mental health outcomes; evidence from a randomized controlled trial [J]. *Social Science and Medicine*, 2020, 265 (22): 1–8.
- [37] AGUILA E, SMITH J P. Supplemental income program design; a cluster-randomized controlled trial to examine the health and wellbeing effects on older adults by gender, duration, and payment frequency [J]. *Social Science and Medicine*, 2020, 259 (16): 1–10.
- [38] GIANNAKOPOULOS G, SOLANTAUS T, TZAVARA C. Mental health promotion and prevention interventions in families with parental depression; a randomized controlled trial [J]. *Journal of Affective Disorders*, 2021, 278 (1): 114–121.
- [39] GAILLARD A, SULTAN-TAÏEB H, SYLVAIN C. Economic evaluations of mental health interventions; a systematic review of interventions with work-focused components [J]. *Safety Science*, 2020, 132 (12): 1–13.
- [40] DEATON A, CARTWRIGHT N. Understanding and misunderstanding randomized controlled trials [J]. *Social Science and Medicine*, 2018, 210 (15): 2–21.
- [41] CRAIG P, COOPER C, GUNNELL D. Using natural experiments to evaluate population health interventions; new medical research council guidance [J]. *Journal of Epidemiology and Community Health*, 2012, 66 (12): 1182–1186.
- [42] HU Y, VAN LENTHE F J, HOFFMANN R. Assessing the impact of natural policy experiments on socioeconomic inequalities in health; how to apply commonly used quantitative analytical methods? [J]. *BMC Medical Research Methodology*, 2017, 17

- (1): 1-17.
- [43] DAWSON A, SIM J. The nature and ethics of natural experiments [J]. *Journal of Medical Ethics*, 2015, 41 (10): 848-853.
 - [44] ROSENBAUM P R, RUBIN D B. The central role of the propensity score in observational studies for causal effects [J]. *Biometrika*, 1983, 70 (1): 41-55.
 - [45] BENCHIMOL E I, SMEETH L, GUTTMANN A. The reporting of studies conducted using observational routinely-collected health data statement for pharmacoepidemiology [J]. *PLoS Medicine*, 2015, 12 (10): 1-22.
 - [46] AUSTIN P C. A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality [J]. *Multivariate Behavioral Research*, 2011, 46 (1): 119-151.
 - [47] AUSTIN P C. An introduction to propensity score methods for reducing the effects of confounding in observational studies [J]. *Multivariate Behavioral Research*, 2011, 46 (3): 399-424.
 - [48] AUSTIN P C, STUART E A. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies [J]. *Statistics in Medicine*, 2015, 34 (28): 3661-3679.
 - [49] SEAMAN S R, WHITE I R. Review of inverse probability weighting for dealing with missing data [J]. *Statistical Methods in Medical Research*, 2013, 22 (3): 278-295.
 - [50] ABADIE A, IMBENS G W. Bias-corrected matching estimators for average treatment effects [J]. *Journal of Business and Economic Statistics*, 2011, 29 (1): 1-11.
 - [51] WANG S, HU Y. Migration and health in China: linking sending and host societies [J]. *Population, Space and Place*, 2019, 25 (6): 1-11.
 - [52] BARBAN N, DE LUNA X, LUNDHOLM E. Causal effects of the timing of life-course events: age at retirement and subsequent health [J]. *Sociological Methods and Research*, 2020, 49 (1): 216-249.
 - [53] SCHNALZENBERGER M. Causal effect of income on health: investigating two closely related policy reforms in Austria [J]. *Journal of the Economics of Ageing*, 2016, 7 (1): 6-16.
 - [54] HUBER M, LECHNER M, WUNSCH C. Workplace health promotion and labour market performance of employees [J]. *Journal of Health Economics*, 2015, 43 (5): 170-189.
 - [55] BOWDEN J, SMITH G D, BURGESS S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression [J]. *International Journal of Epidemiology*, 2015, 44 (2): 512-525.
 - [56] BOWDEN J, DAVEY SMITH G, HAYCOCK P C. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator [J]. *Genetic Epidemiology*, 2016, 40 (4): 304-314.
 - [57] BURGESS S, BUTTERWORTH A, THOMPSON S G. Mendelian randomization analysis with multiple genetic variants using summarized data [J]. *Genetic Epidemiology*, 2013, 37 (7): 658-665.
 - [58] BASU S, MEGHANI A, SIDDIQI A. Evaluating the health impact of large-scale public policy changes: classical and novel approaches [J]. *Annual Review of Public Health*, 2017, 38 (1): 351-370.
 - [59] PAN Z, DONG W. Can money substitute adult Children's absence? measuring remittances' compensation effect on the health of rural migrants' left-behind elderly parents [J]. *Journal of Rural Studies*, 2020, 79 (7): 216-225.
 - [60] ATALAY K, BARRETT G F. The causal effect of retirement on health: new evidence from Australian pension reform [J]. *Economics Letters*, 2014, 125 (3): 392-395.
 - [61] GORRY D, SLAVOV S N. The effect of retirement on health biomarkers [J]. *Economics and Human Biology*, 2021, 40 (1): 1-10.
 - [62] CUADROS-MEÑACA A. Remittances, health insurance, and pension contributions: evidence from Colombia [J]. *World Development*, 2020, 127 (3): 1-12.

- [63] GLYMOUR M M, MANLY J J. Compulsory schooling laws as quasi-experiments for the health effects of education: reconsidering mechanisms to understand inconsistent results [J]. *Social Science and Medicine*, 2018, 214 (19): 67-69.
- [64] CHIRWA G C, SUHRCKE M, MORENO-SERRA R. The impact of Ghana's national health insurance on psychological distress [J]. *Applied Health Economics and Health Policy*, 2020, 18 (2): 249-259.
- [65] LEE J, KIM J. The role of health insurance in mental health care for young adults [J]. *Applied Economics*, 2020, 52 (42): 4577-4593.
- [66] 潘杰, 雷晓燕, 刘国恩. 医疗保险促进健康吗? ——基于中国城镇居民基本医疗保险的实证分析 [J]. *经济研究*, 2013 (4): 130-142, 156.
- [67] CYGAN-REHM K, WUNDER C. Do working hours affect health? evidence from statutory workweek regulations in Germany [J]. *Labour Economics*, 2018, 53 (4): 162-171.
- [68] SHARAF M F, RASHAD A S. Does precarious employment ruin youth health and marriage? evidence from Egypt using longitudinal data [J]. *International Journal of Development Issues*, 2020, 19 (3): 391-406.
- [69] LIANG H, YUE Z, LIU E. How does social capital affect individual health among the elderly in rural China?: mediating effect analysis of physical exercise and positive attitude [J]. *PLoS ONE*, 2020, 15 (7): 1-18.
- [70] WING C, SIMON K, BELLO-GOMEZ R A. Designing difference in difference studies: best practices for public health policy research [J]. *Annual Review of Public Health*, 2018, 39 (1): 453-469.
- [71] LU H, KANDILOV I T, ZHU R. The impact of internal migration on the health of rural migrants: evidence from longitudinal data in China [J]. *Journal of Development Studies*, 2020, 56 (4): 840-855.
- [72] CARRINO L, GLASER K, AVENDANO M. Later retirement, job strain, and health: evidence from the new state pension age in the United Kingdom [J]. *Health Economics*, 2020, 29 (8): 891-912.
- [73] BAUER A B, EICHENBERGER R. Worsening workers' health by lowering retirement age: the malign consequences of a benign reform [J]. *Journal of the Economics of Ageing*, 2021, 18 (1): 1-17.
- [74] HALLBERG D, JOHANSSON P, JOSEPHSON M. Is an early retirement offer good for your health? quasi-experimental evidence from the army [J]. *Journal of Health Economics*, 2015, 44 (6): 274-285.
- [75] BRATBERG E, HOLMÅS T H, MONSTAD K. Health effects of reduced workload for older employees [J]. *Health Economics*, 2020, 29 (5): 554-566.
- [76] THISTLETHWAITE D L, CAMPBELL D T. Regression-discontinuity analysis: an alternative to the ex-post facto experiment [J]. *Journal of Educational Psychology*, 1960, 51 (6): 309-317.
- [77] BOR J, MOSCOE E, MUTEVEDZI P. Regression discontinuity designs in epidemiology: causal inference without randomized trials [J]. *Epidemiology*, 2014, 25 (5): 729-737.
- [78] MOSCOE E, BOR J, BÄRNIGHAUSEN T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice [J]. *Journal of Clinical Epidemiology*, 2015, 68 (2): 132-143.
- [79] LEE D S, LEMIEUX T. Regression discontinuity designs in economics [J]. *Journal of Economic Literature*, 2010, 48 (2): 281-355.
- [80] STANCANELLI E. Couples' retirement under individual pension design: a regression discontinuity study for France [J]. *Labour Economics*, 2017, 49 (6): 14-26.
- [81] CHEN S, GELDSETZER P, BÄRNIGHAUSEN T. The causal effect of retirement on stress in older adults in China: a regression discontinuity study [J]. *SSM-Population Health*, 2020, 10 (1): 1-11.
- [82] ZHANG Y, SALM M, VAN SOEST A. The effect of retirement on healthcare utilization: evidence from China [J]. *Journal of Health Economics*, 2018, 62 (6): 165-177.
- [83] GELMAN A, IMBENS G. Why high-order polynomials should not be used in regression discontinuity designs [J]. *Journal of Business and Economic Statistics*, 2019, 37 (3): 447-456.

- [84] ROSE L. Retirement and health; evidence from England [J]. *Journal of Health Economics*, 2020, 73 (5): 1-26.
- [85] 雷晓燕, 谭力, 赵耀辉. 退休会影响健康吗? [J]. *经济学 (季刊)*, 2010 (4): 1539-1558.
- [86] PICCHIO M, OURS J V. Mental health effects of retirement [J]. *De Economist*, 2020, 168 (3): 419-452.
- [87] ZANG E. Spillover effects of a husband's retirement on a woman's health; evidence from urban China [J]. *Social Science and Medicine*, 2020, 245 (2): 1-10.
- [88] MÜLLER T, SHAIKH M. Your retirement and my health behavior; evidence on retirement externalities from a fuzzy regression discontinuity design [J]. *Journal of Health Economics*, 2018, 57 (1): 45-59.
- [89] AVENDANO M, DE COULON A, NAFILYAN V. Does longer compulsory schooling affect mental health? evidence from a British reform [J]. *Journal of Public Economics*, 2020, 183 (3): 1-34.
- [90] JANKE K, JOHNSTON D W, PROPPER C. The causal effect of education on chronic health conditions in the UK [J]. *Journal of Health Economics*, 2020, 70 (2): 1-15.
- [91] ZHONG H. Effects of quantity of education on health; a regression discontinuity design approach based on the Chinese Cultural Revolution [J]. *China Economic Review*, 2016, 41 (5): 62-74.
- [92] MENEZES-FILHO N, POLITI R. Estimating the causal effects of private health insurance in Brazil; evidence from a regression kink design [J]. *Social Science and Medicine*, 2020, 264 (21): 1-9.
- [93] QI D, WU Y. Does welfare stigma exist in China? policy evaluation of the minimum living security system on recipients' psychological health and wellbeing [J]. *Social Science and Medicine*, 2018, 205 (10): 26-36.
- [94] GAO L, PENNING M J, WU Z. Internal migration and the health of middle-aged and older persons in China; the healthy migrant effect reconsidered [J]. *Research on Aging*, 2021, 43 (9-10): 1-13.
- [95] ZHU R. Retirement and its consequences for women's health in Australia [J]. *Social Science and Medicine*, 2016, 163 (15): 117-125.
- [96] WATSON B, OSBERG L. Can positive income anticipations reverse the mental health impacts of negative income anxieties? [J]. *Economics and Human Biology*, 2019, 35 (4): 107-122.
- [97] HALONEN J I, PULAKKA A, VAHTERA J. Commuting time to work and behaviour-related health; a fixed-effect analysis [J]. *Occupational and Environmental Medicine*, 2020, 77 (2): 77-83.
- [98] FUJIWARA T, KAWACHI I. Is education causally related to better health? a twin fixed-effect study in the USA [J]. *International Journal of Epidemiology*, 2009, 38 (5): 1310-1322.
- [99] FU M, LIU C, YANG M. Effects of public health policies on the health status and medical service utilization of Chinese internal migrants [J]. *China Economic Review*, 2020, 62 (4): 1-13.
- [100] SU D, CHEN Y C, GAO H X. Effect of integrated urban and rural residents medical insurance on the utilisation of medical services by residents in China; a propensity score matching with difference-in-differences regression approach [J]. *BMJ Open*, 2019, 9 (2): 1-10.
- [101] HAO X, YANG Y, GAO X. Evaluating the effectiveness of the health management program for the elderly on health-related quality of life among elderly people in China; findings from the China health and retirement longitudinal study [J]. *International Journal of Environmental Research and Public Health*, 2019, 16 (1): 1-9.
- [102] 宗庆庆, 张熠, 陈玉宇. 老年健康与照料需求: 理论和来自随机实验的证据 [J]. *经济研究*, 2020 (2): 36-51.
- [103] STEPHAN U, LI J, QU J. A fresh look at self-employment, stress and health; accounting for self-selection, time and gender [J]. *International Journal of Entrepreneurial Behaviour and Research*, 2020, 26 (5): 1133-1177.
- [104] EIBICH P. Understanding the effect of retirement on health; mechanisms and heterogeneity [J]. *Journal of Health Economics*, 2015, 43 (5): 1-12.
- [105] 程令国, 张晔. “新农合”: 经济绩效还是健康绩效? [J]. *经济研究*, 2012 (1): 120-133.
- [106] BRADY N. Introduction to causal inference from a machine learning perspective [EB/OL]. (2020-12-17) [2021-12-

- 01]. <https://www.bradyn Neal.com/Introduction to Causal Inference-Dec17 2020-Neal.pdf>.
- [107] ZUBIZARRETA J R. Stable weights that balance covariates for estimation with incomplete outcome data [J]. *Journal of the American Statistical Association*, 2015, 110 (511): 910-922.
- [108] DIAMOND A, SEKHON J S. Genetic matching for estimating causal effects: balance in observational studies [J]. *Review of Economics and Statistics*, 2013, 95 (3): 932-945.
- [109] HAINMUELLER J. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies [J]. *Political Analysis*, 2012, 20 (1): 25-46.
- [110] ZUBIZARRETA J R. Using mixed integer programming for matching in an observational study of kidney failure after surgery [J]. *Journal of the American Statistical Association*, 2012, 107 (500): 1360-1371.
- [111] GOLLE D, LECHNER M, MOCZALL A. Does the estimation of the propensity score by machine learning improve matching estimation? the case of Germany's programmes for long term unemployed [J]. *Labour Economics*, 2020, 65 (4): 1-21.
- [112] HAHN P R, MURRAY J S, CARVALHO C M. Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects [J]. *Bayesian Analysis*, 2020, 15 (3): 965-1056.
- [113] CHERNOZHUKOV V, CHETVERIKOV D, DEMIRER M. Double/debiased machine learning for treatment and structural parameters [J]. *Econometrics Journal*, 2018, 21 (1): 1-68.
- [114] ATHEY S, IMBENS G W, WAGER S. Approximate residual balancing: debiased inference of average treatment effects in high dimensions [J]. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 2018, 80 (4): 597-623.
- [115] BELLONI A, CHERNOZHUKOV V, HANSEN C. Inference on treatment effects after selection among high-dimensional controls [J]. *Review of Economic Studies*, 2013, 81 (2): 608-650.

Methods, Applications and Prospects of Causal Inference in Scientific Research: A Study of Individual Health Research

REN Guoqiang¹, WANG Yudan¹, ZHOU Yunbo²

(1. School of Management, Tianjin University of Technology, Tianjin 300384, China;

2. School of Economics, Nankai University, Tianjin 300071, China)

Abstract: Causal inference is an important method to study the mechanism of individual health effects, which helps to promote the scientific and rationalization of health-related policies, provides more reliable social security for individual health, and has very important social significance for improving the health of residents. This article analyzes and evaluates the existing literature from three aspects: common problems in individual health research, theoretical framework of causal inference, and causal inference in empirical research. As the most extensive research field, individual health level is affected by social capital, income, education, insurance, migration, retirement, work and other factors. Meanwhile, there are mainly five types of problems such as measurement error, omitted variable, reverse causation, common causes, and selection bias. The common causal inference methods for these problems include randomized controlled trials, propensity score methods, instrumental variables, double difference methods, regression discontinuity design, and individual fixed

effects model. In this article, the applicable conditions, advantages and disadvantages of these causal inference models are expounded; meanwhile, the application of various models in empirical research on individual health is briefly summarized and analyzed. The study also helps scholars to choose appropriate causal inference methods in the field of individual health research, or further comprehensively apply existing methods. With the development of big data technology and an in-depth understanding of causal inference methods, in future research, the combination of machine learning and causal inference methods should be strengthened to enrich the existing causal inference tools and ensure the robustness of research results.

Keywords: individual health; endogeneity; confounding factors; causal inference

[责任编辑 武 玉]