

## · 综述 ·

## 因果推断在公共卫生领域的研究进展与应用\*

梁胤豪<sup>1,2</sup> 周 磊<sup>3</sup> 方 亚<sup>1,2,Δ</sup>

【提 要】 探究因果关系在公共卫生领域中起的核心作用。回顾潜在结果框架和结构因果模型,介绍这两类框架下常用的因果推断方法,例如倾向得分匹配、工具变量法、断点回归和双重差分、基于机器学习的因果推断方法等,阐述这些因果推断方法的适用条件与优缺点及其在公共卫生领域的应用。

【关键词】 因果推断 公共卫生 因果模型

【中图分类号】 R197.32

【文献标识码】 A

DOI 10.11783/j.issn.1002-3674.2024.06.031

公共卫生学科的历史可以看作是探究疾病因果关系的发展史,公共卫生研究者与政策制定者需要根据处理变量与结局变量的因果关系来决定干预措施,以预防疾病的发生与传染。因果推断作为公共卫生领域中推断疾病病因与制定相关政策的重要手段,已成为其中至关重要的一环。推断因果关系的金标准是随机对照试验(randomized controlled trial, RCT)<sup>[1]</sup>。但在现实工作中, RCT 总是耗时且昂贵的,并需要考虑伦理问题,这在很大程度上限制了它在公共卫生领域的应用。

相比 RCT,在没有任何干预情况下得到的观测数据更容易获得。然而,基于观测数据推断因果关系的核心问题是如何准确得到未观测处理变量的结果。为此研究者提出了潜在结果框架和结构因果模型,以推断处理变量的因果效应。这两类框架被研究者广泛应用于传染病防治、慢性病研究以及公共卫生政策等方面,探明了先前因混杂因素影响而导致的 uncertain 结论。

本文首先回顾了潜在结果框架,并介绍了该框架下常用的因果推断方法及其在公共卫生领域的应用。其次,简要介绍了结构因果模型和基于机器学习的因果推断方法及其在公共卫生领域的相关应用。最后,讨论了目前因果推断在公共卫生领域应用中存在的困难以及未来的研究方向。

## 潜在结果框架

Splawa-Neyman<sup>[2]</sup>在 1923 年提出潜在结果框架,并针对实验性研究给出了因果效应的数学定义, Rubin<sup>[3]</sup>将这一定义推广到观察性研究。潜在结果框架

将个体无法在实验中观测到的结果称为反事实结果,进而将因果效应<sup>[2-4]</sup>定义为相同个体实际结果与反事实结果之差,将平均因果效应定义为个体因果效应的期望。潜在结果框架易于理解,但反事实结果数据缺失,估计较为困难,为此,学者提出倾向得分匹配、工具变量法、断点回归设计、双重差分法等估计方法。

## 1. 倾向得分匹配

倾向得分匹配(propensity score methods, PSM)<sup>[5]</sup>方法由 Rosenbaum 和 Rubin 在 1983 年所提出。PSM 首先通过将样本的多维协变量用一个倾向得分来代替(一般采用 logit 回归或其他机器学习方法计算<sup>[6]</sup>),其次根据倾向得分对样本进行匹配、分层或加权,从而在对照组中找到与处理组相对应的反事实个体,达到两组间协变量分布伪随机化。Wright 等<sup>[7]</sup>探究了医生的临终关怀讨论是否与较少的积极干预有关,通过 PSM 对数据进行调整后,结果显示临终关怀讨论与临终前不太积极的医疗护理和更早的临终关怀有关。Vincent 等<sup>[8]</sup>探究了输血对重症患者死亡率的影响,对于 PSM 分析中的匹配患者,输血患者的 28 天死亡率为 22.7%,未输血患者为 17.1% ( $P=0.02$ ),为流行病学研究提供了输血与器官功能下降、死亡率之间关联的证据。

传统 PSM 并未处理未观测的混杂因素,可能存在模型设定错误和得分估计极端的问题。为了获得更准确的倾向得分,研究者提出了双稳健估计<sup>[9-12]</sup>和协变量平衡倾向得分<sup>[13]</sup>(covariate balancing propensity score, CBPS)方法。双稳健估计是将逆概率加权<sup>[14]</sup>和回归估计相结合,只要其中一个模型正确,双稳健估计就具有相合性。双稳健估计能有效减少模型错误所带来的偏差,被广泛应用于缺失数据分析和因果推断中。在不满足可忽略性假定<sup>[5]</sup>时,双稳健估计的构建比较困难,需要更严格的倾向得分模型假定,并且当两个模型都不正确时,双稳健估计的偏差可能会更大, Kang 等<sup>[15]</sup>对此进行了详细的讨论并且给出相关案例, Rob-

\* 基金项目:国家自然科学基金项目(82073669, 81602941);福建省哲学社会科学规划项目(FJ2020B072)

1. 厦门大学公共卫生学院(361102)

2. 福建省高校卫生技术评估重点实验室

3. 厦门大学公共事务学院

Δ通信作者:方亚, E-mail: fangya@xmu.edu.cn

ins 等<sup>[16-18]</sup>针对这些缺陷提出了去偏差双稳健估计等改进方法。

CBPS 方法在无混杂性假设下通过协变量分布的加权平衡,解决了极端倾向性得分带来的偏倚,对特定 PSM 模型的偏差也更具稳健性。吴浩等<sup>[19]</sup>提出了基于协变量平衡加权法的稳健且有效的估计方法,并通过引入超级学习算法提升了模型在实证应用中的稳健性。

## 2. 工具变量法

工具变量法 (instrumental variable, IV) 是处理内生性问题的重要方法。Angrist 等<sup>[20]</sup>详细介绍了如何使用工具变量识别因果效应。IV 除了关心的处理和结果变量,还需额外观测一个工具变量,并满足以下三个条件:①工具变量对结果无直接作用;②工具变量与未观测的混杂因素独立;③工具变量与处理变量相关<sup>[21]</sup>。IV 的估计通常采用两阶段最小二乘法实现<sup>[22]</sup>。

Chen 等<sup>[23]</sup>使用 IV,以中国新农村养老金计划的推出时间作为工具变量,估计养老金参保情况对心理健康的影响,避免了反向因果与遗漏变量偏差所带来的内生性问题。结果表明,由于参保养老金,抑郁症状程度有所减轻,这种影响在有资格领取养老金、经济拮据以及基线心理健康状况较差的人群中更为明显。Albarran 等<sup>[24]</sup>使用 IV 来探究教育与成人健康之间的因果效应,以义务教育法规定的义务教育年限作为教育的工具变量,结合了来自欧盟多国收入和生活条件统计调查的两个横截面数据,并使用义务教育法导致的各国和队列义务教育年限的外生差异,发现教育与几项健康指标之间都不存在因果关系。

在实际研究中,寻找工具变量并不容易,即使满足工具变量的三个条件,也只能获得因果效应的取值范围,而无法惟一确定因果效应<sup>[25-28]</sup>。为了进一步缩小因果效应的范围,通常还需要额外的信息或模型假定,例如,假设因果效应的同质性<sup>[29-31]</sup>和单调性<sup>[20]</sup>。此外,如何验证工具变量假定也是一个重要研究领域,将基因型作为工具变量的孟德尔随机化<sup>[32]</sup> (Mendelian randomization, MR) 是目前较为流行的方法,由于等位基因遵循随机分配原则,所以基因和疾病之间的关联不会受到传统流行病学研究中的混杂因素和反向因果的影响。Gill 等<sup>[33]</sup>使用英国生物库和 CARDIoGRAMplusC4D 库的遗传相关估计值,进行单变量 MR (年龄 $\leq 55$ 岁)与多变量 MR (年龄 $> 55$ 岁),以探究遗传预测平均动脉压对冠状动脉疾病 (coronary artery disease, CAD) 风险的影响,研究结果表明中年血压对晚年 CAD 风险的影响独立于晚年血压,而且与血压升高对 CAD 风险的累积时间效应一致。

为了解决工具变量寻找困难的问题, Miao<sup>[34-36]</sup>提

出了阴性对照变量,并系统地研究了用阴性对照变量识别因果效应的方法和所需要的条件。首先,作为 IV 方法推广的一般形式,阴性对照变量无需和未观测的混杂变量独立,而是再设定两个辅助变量,即与结果无关的辅助的暴露变量 (阴性对照暴露, NCE) 和不受暴露影响的辅助的结果变量 (阴性对照结果, NCO),使其与混杂变量相关。其次,通常通过检验 NCE 或 NCO 与暴露变量是否关联来判断是否存在混杂因素,并且阴性对照变量还可以用来校正混杂导致的偏差。Gustavson 等<sup>[37]</sup>为探究孕妇产前风险因素与新生儿结局之间的因果效应,构建了包含阴性对照变量的模型,将孕妇妊娠期间饮酒为暴露,孕妇妊娠前饮酒为阴性对照,研究证明当混杂因素的影响程度不同时,采用阴性对照变量也能无偏估计孕妇产前风险因素与新生儿结局之间的因果效应。

## 3. 断点回归设计

断点回归设计 (regression discontinuity design, RDD)<sup>[38]</sup>由美国西北大学心理学家 Campbell 于 1958 年首先提出,具有非常强的实用性。RDD 是一种能够有效利用现实约束条件分析变量之间因果关系的实证方法,可用于评估公共政策的因果效应,近年来得到了公共卫生学界的广泛认可<sup>[39-40]</sup>。在公共卫生研究中的断点可看作是 RCT 中的干预,因此,断点右侧是受到干预处理的处理组,断点左侧是未处理的控制组。在断点附近,控制组观测不到的部分即为处理组观测到的部分的事实,从而推断出因果效应。根据在断点处的处理概率是否从 0 跳跃 1 将 RDD 分为两类,一类是精确断点回归<sup>[41]</sup>,另一类是模糊断点回归<sup>[42-43]</sup>。

Bernal 等<sup>[44]</sup>使用精确断点回归,以居民生活条件为断点,估计了公共保险覆盖面对卫生保健利用和卫生费用支出之间的因果关系。Palmer 等<sup>[45]</sup>使用精确断点回归和模糊断点回归相结合的方法评估了医疗保险对老年人康复情况的影响。

此外,当一项治疗已经成为标准的操作,不可能再开展随机对照试验时, RDD 也可以提供强有力的因果关系证据,即便治疗效果很微小或很难观测,甚至受到来自内部有效性和外部有效性的挑战时, RDD 依然能够捕捉因果关系。

## 4. 双重差分法

双重差分法 (difference in difference, DID)<sup>[46]</sup>常被用来评估公共卫生政策实施的跨期效果,适用于面板数据和纵向数据,并需要满足平行趋势假设。DID 通过比较处理组和对照组之间的差异来评估因果效应。DID 控制了时变特征的异质性,因此不容易受到未测量的混杂因素或测量误差的影响。Carrino 等<sup>[47]</sup>使用 DID 分析了退休政策的变动对健康影响的因果效应,表明提高国家养老金年龄会导致抑郁症状的概

率增加 12%。Bauer 和 Eichenberger 则运用 DID 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,结果发现建筑工人在工作到 60 岁而不是 65 岁时病假增加了 33%,自我报告的健康问题的概率增加了 54%,还发现对 61~65 岁年龄组有微小的负面影响<sup>[48]</sup>。

然而,DID 方法可能会因不完全满足平行趋势假设而使得评估结果产生偏差。Abadie 和 Gardeazabal 在 2003 年提出的合成控制法(synthetic control, SC)<sup>[49]</sup>对 DID 进行了改进,通过对多个控制对象加权来模拟目标对象政策实施前的情况,使之满足平行趋势假设,减少了主观选择的误差,避免了政策内生性问题和过分外推。并且由于 SC 所得权重的稀疏性,使得 SC 有很好的可解释性。Shioda 等<sup>[50]</sup>运用 SC 并开发了分布式滞后模型评估了 2005 年至 2015 年期间,巴西 25 个州接种肺炎球菌结合疫苗对 80 岁及以上全因肺炎住院的影响。

在实际研究中,若处理组样本量大且满足平行趋势假设时,通常使用 DID 方法。若处理组含样本量少,无法满足平行趋势假定时,则可以使用 SC 方法,两种方法的本质都是在寻找一个最优的参照组。Arkhangelsky 等在 2021 年充分将 DID 和 SC 的优点相结合,提出了具有双重稳健性的合成双重差分法(synthetic difference-in-difference, SDID)<sup>[51]</sup>。考虑到公共卫生政策实施的分布在地区与时间上并不是随机出现的,SDID 不仅通过个体权重找到与处理组相近的控制组个体,还通过时间权重找到与政策后处理期相似的政策前处理期,并分别赋予它们更大的个体权重和时间权重,使得最终结果更加精确。

### 结构因果模型

因果推断中另一个有影响力的框架是结构因果模型(structural causal model, SCM)<sup>[52-53]</sup>,由 Pearl 在 1988 年提出。SCM 由因果图和结构方程模型(structural equation model, SEM)两个部分组成,因果图可以帮助确定模型的适当变量和关联关系,承担了为几乎任何因果推断方法指定统计模型的重要步骤<sup>[54]</sup>。因果图中的每个有向边都使用一个确定的函数来表示,这些函数共同组成一个结构方程模型,用来计算各个变量之间的因果效应<sup>[54-55]</sup>。

Hu 等<sup>[56]</sup>以 ROSMAP 项目为基础,运用 SCM 研究了阿兹海默症(Alzheimer's disease, AD)和 2 型糖尿病(type 2 diabetes mellitus, T2DM)之间的共同因果路径,并建立了多组学因果网络,确定了 13 个共享的因果基因、16 个 AD 和 T2DM 之间共享的因果路径,发现了胰岛素信号传导与 AD 病理学之间的相关,还发现了与  $\beta$ -淀粉样蛋白沉积相关的 GSK $\beta$  多态性之

间的相互作用。Cacani 等<sup>[57]</sup>通过使用 SCM 研究昆虫数量、牛巴贝虫感染水平和体重增加之间的潜在因果关系。先采用贝叶斯多性状模型将协方差划分为遗传和残差分量,将 IC(inductive causation)算法用于搜索合理的因果结构,最后使用 SEM 进行分析,得到的有向无环图表明昆虫数量的干预会直接影响从出生到断奶期间的体重增加,进而有助于降低牛巴贝虫感染水平。

### 基于机器学习的因果推断算法

机器学习与因果推断方法的结合一方面极大程度上提升了倾向得分匹配、工具变量法等传统因果推断方法的估计效果。另一方面,在公共卫生领域,机器学习广泛应用于慢性病预测、传染病预防、疾病诊断等领域,机器学习算法能以较高的准确率判断患者患有何种疾病、病因是什么,却很难通过某种指标给出可解释的依据。因此,因果推断对于克服当前的机器学习限制至关重要。Zeevi 等<sup>[58]</sup>将图神经网络和 SCM 相结合,建立了基于图神经网络的因果推断方法。

Cox<sup>[59]</sup>介绍了诸多研究因果推断的机器学习算法,例如双重机器学习<sup>[60-61]</sup>、推理算法<sup>[62]</sup>、介入因果预测算法<sup>[63-64]</sup>、BN 学习与因果发现算法<sup>[65]</sup>、因果效应估计算法<sup>[66]</sup>、解释算法<sup>[67]</sup>、决策优化和概率规划算法<sup>[68]</sup>,解决了数据分析最常见的问题,例如未测量的潜在变量对因果估计的影响<sup>[69-70]</sup>;数据的缺失问题<sup>[71-72]</sup>;因果模型的最优选择<sup>[70,73]</sup>;个体间的异质性<sup>[74]</sup>;因果效应估计的内部与外部有效性<sup>[75-79]</sup>等。

García-Peña 等<sup>[80]</sup>使用墨西哥健康和老龄化研究的数据,进行了衰弱和老龄化的网络分析。以缺陷和死亡为分析节点,通过结构学习获得连接这些节点的边,以 HC(hill-climbing)和 PC(Peter and Clark)算法来推导图结构,并导出与离散概率图形模型相关的无向图。边的数量随着衰弱指数评分的分位数增加而增加。随着衰弱指数评分的提高,相互关联的缺陷组增加,节点的连接方式也发生了变化,最中心的节点是与移动能力有关的缺陷。

### 总结与展望

本文回顾了潜在结果框架和结构因果模型,并介绍了这两种框架下常用的因果推断方法及其在公共卫生领域的应用。此外,简要介绍了基于机器学习的因果推断算法。值得注意的是,由于因果推断的一些必要的基本假定在观察性研究或实验中可能难以满足,所以要求公共卫生研究人员与临床研究人员进行合作,根据专业知识或者先验信息来判断这些假定是否成立,以确定目标实验和因果问题,并且每一步的选择和实施需要与研究目标相一致,特别是其相关的因果估计。

在大数据时代,公共卫生的数据收集也应当更加给予重视,通过少量样本推断出的因果效应在整个群体上未必是成立的,难以满足外部有效性,而根据数据推断因果效应和寻找因果关系将成为推动公共卫生领域发展的重要动力。在对公共卫生大数据收集时,我们首先要在保障居民个人信息安全性的前提下,整合与居民健康以及公共卫生相关的数据,形成可供研究的公共卫生健康大数据集,为科学研究提供数据支持,通过研究影响健康的诸多因素和健康之间的因果效应,为政府出台科学的公共卫生政策提供依据。

其次,尝试将因果推断与深度学习相融合,通过挖掘数据内在的因果结构提升模型的泛化性。变量的连续性、高维变量的复杂性、未观测的混杂因素以及现有机器学习算法的可解释性也是因果推断研究中需要深入研究的问题。最后,研究者们应关注基于探索的因果推断方法,通过因果强化学习等探索策略解决未知的因果问题。

#### 参 考 文 献

- [ 1 ] Fisher RA. Design of experiments [J]. British Medical Journal, 1936, 1(3923): 554.
- [ 2 ] Splawa-Neyman J, Speed TP. On the application of probability theory to agricultural experiments. Essay on principles. Section 9 [J]. Statistical Science, 1990, 5(4): 465-472.
- [ 3 ] Rubin, Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies [J]. Journal of Educational Psychology, 1974, 66(5): 688-701.
- [ 4 ] Rubin DB. Bayesian Inference for Causal Effects [J]. The Annals of Statistics, 1978, 6: 34-58.
- [ 5 ] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects [J]. Biometrika, 1983, 70(1): 41-55.
- [ 6 ] Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning [J]. Statistics in Medicine, 2010, 29(3): 337-346.
- [ 7 ] Wright AA, Zhang B, Ray A, et al. Associations Between End-of-Life Discussions, Patient Mental Health, Medical Care Near Death, and Caregiver Bereavement Adjustment [J]. JAMA, 2008, 300(14): 1665-1673.
- [ 8 ] Vincent JL, Baron JF, Reinhart K, et al. Anemia and Blood Transfusion in Critically Ill Patients [J]. JAMA, 2002, 288(12): 1499-1507.
- [ 9 ] Robins JM, Rotnitzky A, Zhao LP. Estimation of Regression Coefficients When Some Regressors are not Always Observed [J]. Journal of the American Statistical Association, 1994, 89(427): 846-866.
- [ 10 ] Rotnitzky A, Scharfstein R. Semiparametric Regression for Repeated Outcomes with Nonignorable Nonresponse [J]. Journal of the American Statistical Association, 1998, 93(444): 1321-1339.
- [ 11 ] Scharfstein DO, Robins R. Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models [J]. Journal of the American Statistical Association, 1999, 94(448): 1096-1120.
- [ 12 ] Funk MJ, Westreich D, Wiesen C, et al. Doubly Robust Estimation of Causal Effects [J]. American Journal of Epidemiology, 2011, 173(7): 761-767.
- [ 13 ] 范菊逸, 詹铭峰, 蔡宗武, 等. 带有变量选择的协变量平衡倾向得分的估计: 基于 GMM-LASSO 方法 [J]. 系统工程理论与与实践, 2021, 41(10): 9.
- [ 14 ] Ertefaie A, Stephens DA. Comparing approaches to causal inference for longitudinal data: inverse probability weighting versus propensity scores [J]. The International Journal of Biostatistics, 2010, 6(2): 14.
- [ 15 ] Kang J, Schafer JL. Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data [J]. Statistical Science, 2007, 22(4): 574-580.
- [ 16 ] Robins J, Sued M, Lei-Gomez Q, et al. Comment: Performance of Double-Robust Estimators When "Inverse Probability" Weights Are Highly Variable [J]. Statistical Science, 2012, 22(4): 544-559.
- [ 17 ] Tan Z. Bounded, efficient and doubly robust estimation with inverse weighting [J]. Biometrika, 2010, 97(3): 661-682.
- [ 18 ] Vermeulen K, Vansteelandt S. Bias-Reduced Doubly Robust Estimation [J]. Journal of the American Statistical Association, 2015, 110(511): 1024-1036.
- [ 19 ] 吴浩, 彭非. 基于协变量平衡加权的平均处理效应的稳健有效估计 [J]. 统计研究, 2020, 37(4): 114-128.
- [ 20 ] Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables [J]. Social Science Electronic Publishing, 1996, 91(434): 465-468.
- [ 21 ] 苗旺, 刘春辰, 耿直. 因果推断的统计方法 [J]. 中国科学: 数学, 2018, 48(12): 26.
- [ 22 ] Angrist JD, Imbens GW. Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity [J]. Publications of the American Statistical Association, 1995, 90(430): 431-442.
- [ 23 ] Chen X, Wang T, Busch SH. Does money relieve depression? Evidence from social pension expansions in China [J]. Social Science Medicine, 2018, 220: 411-420.
- [ 24 ] Albarran P, Hidalgo-Hidalgo M, Iturbe-Ormaetxe I. Education and adult health: Is there a causal effect? [J]. Social Science & Medicine, 2020, 249: 113094.
- [ 25 ] Heckman JJ. Randomization as an Instrumental Variable [J]. National Bureau of Economic Research, 1996, 78(2): 336-341.
- [ 26 ] Manski CF. Nonparametric Bounds on Treatment Effects [J]. The American Economic Review, 1990, 80(2): 319-323.
- [ 27 ] Bernardo JM, Bayarri MJ, Berger JO, et al. Transparent Parametrizations of Models for Potential Outcomes [J]. Oxford University Press, 2011: 569-610.
- [ 28 ] Pearl BJ. Bounds on Treatment Effects From Studies With Imperfect Compliance [J]. Journal of the American Statistical Association, 1997, 92(439): 1171-1176.
- [ 29 ] Goldberger A. Structural Equation Methods in the Social Sciences [J]. Econometric Society, 1972, 40(6): 979-1001.
- [ 30 ] Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models [J]. Communications in Statistics Theory, 1994, 23(8): 2379-2412.
- [ 31 ] Hernán M, Robins JM. Estimating causal effects from epidemiological data [J]. Journal of Epidemiology Community Health, 2006, 60(7): 578-586.
- [ 32 ] Pierce BL, Stephen B. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators [J]. American Journal of Epidemiology, 2013, (7): 1177-1184.
- [ 33 ] Gill D, Georgakis MK, Zuber V, et al. Genetically Predicted Midlife Blood Pressure and Coronary Artery Disease Risk: Mendelian Randomization Analysis [J]. Journal of the American Heart Association, 2020, 9(14): e016773.
- [ 34 ] Miao W, Tchetgen E. A Confounding Bridge Approach for Double Negative Control Inference on Causal Effects [EB/OL]. (2018) [2021-06-21]. <https://arxiv.org/pdf/1808.04945>.

- [35] Miao W, Tchetgen E. Invited Commentary: Bias Attenuation and Identification of Causal Effects With Multiple Negative Controls [J]. *American Journal of Epidemiology*, 2017, 185(10): 950-953.
- [36] Miao W, Geng Z, Tchetgen E. Identifying causal effects with proxy variables of an unmeasured confounder [J]. *Biometrika*, 2016, 105(4): 987-993.
- [37] Gustavson K, Smith GD, Eilertsen EM. Handling unobserved confounding in the relation between prenatal risk factors and child outcomes: a latent variable strategy [J]. *European Journal of Epidemiology*, 2022, 37(5): 477-494.
- [38] Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: An alternative to the ex post facto experiment [J]. *Journal of Educational Psychology*, 1960, 51(6): 309-317.
- [39] Bor J, Moscoe E, Mutevedzi P, et al. Regression discontinuity designs in epidemiology: causal inference without randomized trials [J]. *Epidemiology*, 2014, 25(5): 729-737.
- [40] 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.
- [41] Stancanelli E. Couples' Retirement Under Individual Pension Design: A Regression Discontinuity Study for France [J]. *Labour Economics*, 2018, 49: 14-26.
- [42] Chen S, Geldsetzer P, Brnighausen T. The causal effect of retirement on stress in older adults in China: A regression discontinuity study [J]. *SSM Population Health*, 2019, 10: 100462.
- [43] Zhang Y, Salm M, van Soest A. The effect of retirement on health-care utilization: Evidence from China [J]. *Journal of Health Economics*, 2018, 62: 165-177.
- [44] Bernal N, Carpio MA, Klein TJ. The effects of access to health insurance: Evidence from a regression discontinuity design in Peru [J]. *Journal of Public Economics*, 2017, 154: 122-136.
- [45] Palmer M, Mitra S, Mont D, et al. The impact of health insurance for children under age 6 in Vietnam: A regression discontinuity approach [J]. *Social Science & Medicine*, 2015, 145: 217-226.
- [46] Wing C, Simon K, Bello-Gomez RA. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research [J]. *Annual Review of Public Health*, 2018, 39(1): 453-469.
- [47] 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.
- [48] Bauer AB, Eichenberger R. Worsening Workers' Health by Lowering Retirement Age: The Malign Consequences of a Benign Reform [J]. *CREMA Working Paper Series*, 2018, 18:100296.
- [49] Abadie A, Gardeazabal J. The Economic Costs of Conflict: A Case Study of the Basque Country [J]. *American Economic Review*, 2003, 93(1): 113-132.
- [50] Shioda K, Cai J, Warren JL, et al. Incorporating Information on Control Diseases Across Space and Time to Improve Estimation of the Population-level Impact of Vaccines [J]. *Epidemiology*, 2021, 32(3): 360-367.
- [51] Arkhangelsky D, Imbens GW, Lei L, et al. Double-Robust Two-Way-Fixed-Effects Regression For Panel Data [J]. *arXiv e-prints*, 2021.
- [52] Andersen S. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference [J]. *Artificial Intelligence*, 1991, 48(1): 117-124.
- [53] Pearl J. *Causality: models, reasoning, and inference* [M]. Cambridge: Cambridge University Press, 2000.
- [54] Pearl J. Causal diagrams for empirical research [J]. *Biometrika*, 1995, 82(4): 702-710.
- [55] Swanson, Norman R. *Causality: Models, Reasoning, and Inference* [J]. *Journal of Economic Literature*, 2002, 40(3): 925-926.
- [56] Hu Z, Jiao R, Wang P, et al. Shared Causal Paths underlying Alzheimer's dementia and Type 2 Diabetes [J]. *Scientific Reports*, 2020, 10(1):4107.
- [57] Cacani L, Lopes FB, Gigliotti R, et al. Inferring phenotypic causal networks for tick infestation, *Babesia bovis* infection, and weight gain in Hereford and Braford cattle using structural equation models [J]. *Livestock Science*, 2020, 238: 104032.
- [58] Zeevi M, Dhimi DS, Velikovi P, et al. Relating Graph Neural Networks to Structural Causal Models [J]. *eprint arXiv:210904173*, 2021.
- [59] Cox LA Jr. Toward practical causal epidemiology [J]. *Global Epidemiology*, 2021, 3: 100065.
- [60] Victor C, Denis C, Mert D, et al. Double/debiased machine learning for treatment and structural parameters [J]. *The Econometrics Journal*, 2018, 21(1): 1.
- [61] Knaus MC. Double Machine Learning based Program Evaluation under Unconfoundedness [J]. *Econometrics Journal*, 2022, 25(3): 602-627.
- [62] Salmerón A, Langseth H, Nielsen TD, et al. A Review of Inference Algorithms for Hybrid Bayesian Networks [J]. *Journal of Artificial Intelligence Research*, 2018, 62: 799-828.
- [63] Hendry DF, Morgan MS. The Statistical Implications of a System of Simultaneous Equations [J]. *The Foundations of Econometric Analysis*, 1995(40): 454-463.
- [64] Lin L, Sperrin M, Jenkins DA, et al. A scoping review of causal methods enabling predictions under hypothetical interventions [J]. *Diagnostic and Prognostic Research*, 2020, 5(1): 3.
- [65] Glymour C, Zhang K, Spirtes P. Review of Causal Discovery Methods Based on Graphical Models [J]. *Frontiers in Genetics*, 2019, 10: 524.
- [66] Textor J, Zander B, Gilthorpe MS, et al. Robust causal inference using directed acyclic graphs: the R package 'dagitty' [J]. *International Journal of Epidemiology*, 2016, 45(6): 1887-1894.
- [67] Yuan C, Lim H, Lu T. Most Relevant Explanation in Bayesian Networks [J]. *Journal of Artificial Intelligence Research*, 2011, 42: 309-352.
- [68] Hansen EA. An integrated approach to solving influence diagrams and finite-horizon partially observable decision processes [J]. *Artificial Intelligence*, 2021, 294: 103431.
- [69] Chobtham K, Constantinou AC. Bayesian network structure learning with causal effects in the presence of latent variables [C]. 2020.
- [70] Li M, Zhang R, Liu K. A New Ensemble Learning Algorithm Combined with Causal Analysis for Bayesian Network Structural Learning [J]. *Symmetry-Basel*, 2021, 12(12): 2054.
- [71] Masegosa AR, Feelders AJ, Gaag L. Learning from incomplete data in Bayesian networks with qualitative influences [J]. *International Journal of Approximate Reasoning*, 2016, 69: 18-34.
- [72] Myers JW, Laskey KB, Levitt TS. Learning Bayesian Networks from Incomplete Data with Stochastic Search Algorithms [C]. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, 1999.
- [73] Cox LA Jr. Implications of nonlinearity, confounding, and interactions for estimating exposure concentration-response functions in quantitative risk analysis [J]. *Environmental Research*, 2020, 187: 109638.
- [74] Athey S, Tibshirani J, Wager S. Generalized Random Forests [EB/OL]. (2018-04-05) [2022-02-08]. <https://arxiv.org/pdf/1610.01271>.
- [75] Bareinboim E, Pearl J. Causal Transportability with Limited Experiments [C]. *Proceedings of the National Conference on Artificial Intelligence*, 2013.