机器学习方法在因果推断中混杂因素控制的应用

兰雨姗 郑思 李姣[1]
（中国医学科学院/北京协和医学院 医学信息研究所 北京 100020）

[摘要] 本文介绍了医学研究中混杂因素对因果推断的影响及常见的混杂因素识别方法，梳理了机器学习方法在因果推断中控制混杂因素的应用，最后讨论了机器学习方法在混杂因素控制中面临的机遇和挑战。

[关键词] 机器学习；因果推断；混杂因素控制
[中图分类号] R-056 [文献标识码] A

Machine Learning Methods for Confounding Control in Causal Inference

LAN Yushan , ZHENG Si ,LI Jiao*,
Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100020, China

[Abstract] In this study, we introduced the effects of confounders in causal inference in medical studies and the state-of-art confounder identification methods. Furthermore, we summarized the applications of machine learning methods for controlling confounders in causal inference. Lastly, we discussed the opportunities and challenges face by application of machine learning methods in controlling confounders.

[Keywords] Machine learning ; Causal inference; Confounding control

1 引言

因果关系（Causality）是指某因素是结局发生的原因，因果推断（Causal Inference）反映了一种在试验设计和分析过程中对混杂、偏移等的慎重考虑，从而在得出因果关系结论时排除各种干扰的影响，做出正确的结论[1]。混杂因素（Confounder）是指某个既与暴露有关又与结局相关的因素，该因素可能会使暴露和结局之间的因果关系产生偏移。例如，在一项关于口服避孕药和心肌梗死之间关联的研究中，在未考虑混杂因素的情况下得到口服避孕药诱发心肌梗死的 OR（Odds Ratio,OR）值为 2.20，而在控制年龄因素（将研究对象按照年龄是否小于 40 岁进行分层）的影响后，得到口服避孕药和心肌梗死之间的 OR 值为 2.79[2]。年龄这一混杂因素的存在，减弱了口服避孕药和心肌梗死之间的关联。因此，混杂因素的控制和识别是因果推断中的关键[3]。传统流行病研究中常采用限制、配对、随机化、工具变量等方法控制混杂因素，但随着医学大数据的不断积累，混杂因素的维度不断增加，传统方法难以较好地处理高维特征。因此，越来越多的研究将机器学习算法引入混杂因素控制领域，希望借助机器学习算法良好的分类和预测能力提升估计因果效应的能力本文介绍了在观察性研究数据中识别混杂因素的方法及如何利用机器学习方法对识别到的或潜在的混杂因素进行控制（如图 1 所示）。

1[作者简介] 兰雨姗，硕士研究生；郑思，副研究员；通讯作者：李姣，研究员，博士生导师。[基金项目]中国医学科学院医学与健康创新工程“医学知识管理与智能化知识服务关键技术研究”（项目编号：2021-I2M-1-056）；“医学人工智能算法评价标准库构建”（项目编号： 2018-I2M-AI-016）

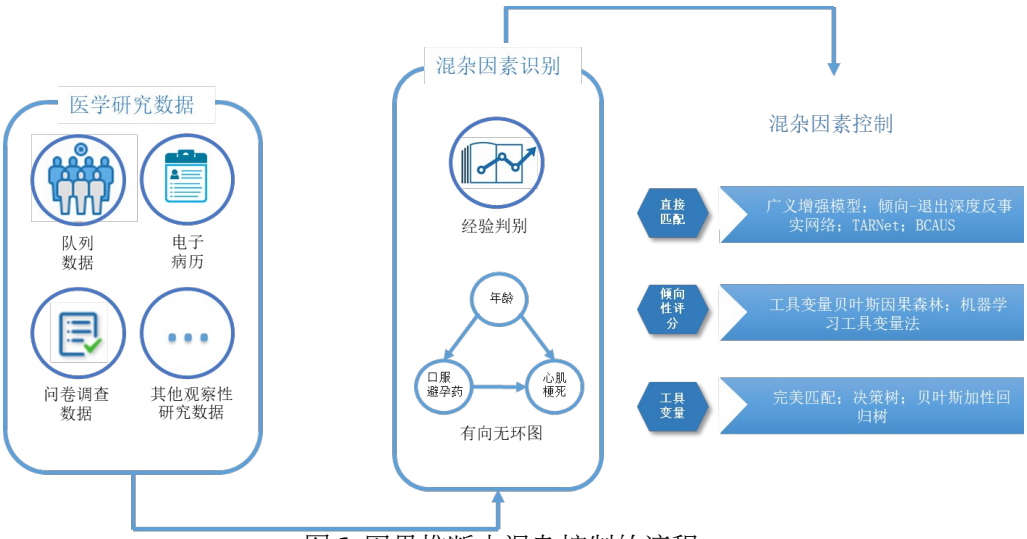

图 1.因果推断中混杂控制的流程

2 因果推断的定义及其方法

因果推断是研究变量间因果关系的学科，流行病学领域提出了一些因果推断理论用于推断出暴露和结局间的因果关系。例如，1856 年 Mill 提出了著名的密尔氏法则，即求同法、求异法、共变法和排除法。1965 年 Hill 提出流行病学病因研究中因果推断的九条准则，即时间顺序、关联强度、剂量反应关系、可重复性、合理性、考虑可替代的解释、实验证据关联的特异性、关联的一致性，该标准被简称为希尔准则（Hill's Criteria），仍广泛地用于人群研究中判断因果关系。

同时，随着相关研究积累，也产生了许多因果推断框架，其中结构性因果模型和潜在结果框架得到了广泛的应用。结构性因果模型由图灵奖获得者 Judea Pearl 教授于 1995 年提出[4]，包括因果图和结构方程，可以用来描述一个系统的因果机制。因果图即有向无环图（Directed Acyclic Graph, DAG），一个有向无环图能唯一确定一个联合分布。结构方程[5]是一种建立、估计和检验因果关系模型的多元回归方法。从结构性因果模型中，Pearl 教授又提取出了 Pearl 因果层次结构（Pearl Causal Hierarchy，PCH）[6]，该结构将因果推断分为了三个层次，包括关联、干预和反事实推论。反事实推论是因果推断中最高层次的问题，用于估计治疗的潜在效果，反事实推论用于回答一个经典问题“如果我采取不同的行动会怎么样？”当混杂因素存在时，会对反事实问题的推导产生影响，例如，上述研究中由于年龄的存在导致估计未服用避孕药的患者是否会患心肌梗死时，低估了口服避孕药对于患心肌梗死的影响。潜在结果框架由哈佛大学 Donald Rubin 教授提出，目的是估计不能被观察到的潜在结果，从而估计实际的干预效果，潜在结果框架又被称为鲁宾因果模型。干预的效果可以被定义为：$ATE = E(x_i|z_i=1) - E(x_i|z_i=0)$ [7]，其中，Z 代表研究所施加的干预措施，X 是研究的结局。干预的效果就是干预措施对结局的因果效应。

3 混杂因素的识别

混杂因素是指与暴露对结局的影响相混淆的另一个风险或保护因素[8]。观察性研究中存在大量已知和未知的混杂因素，这些混杂因素存在时会歪曲暴露和结局之间真实的因果关联。例如，在上述关于口服避孕药与心肌梗死的关联研究中，避孕药就是我们要研究的暴露因素，心肌梗死是本研究的结局，而年龄就是一个会与口服避孕药对心肌梗死的影响相混淆的风险因素，即一个混杂因素。因此，如何识别和控制混杂因素是进行因果推断时需要考虑的重要问题之一。

3.1 利用经验识别混杂

经验识别[9]是指根据已掌握的专业经验知识来判断某个因素是否为混杂因素，当某个因素满足以下条件时即可被判定为混杂因素，即（1）该因素与暴露因素相关。（2）该因素与结局相关。（3）该因素不是暴露与结局之间的中间变量。在许多疾病的病因因果推断中，通常会考虑控制年龄、性别等常见的混杂因素来得到合理的因果关系，但对于是否要控制某个外部因素就需要慎重的考虑。需要参考专家意见，考虑控制该外部因素后得到的结果与专业知识是否符合，与同类研究和既往研究进行比较等，才能得出最终的结论。

3.2 利用有向无环图识别混杂

有向无环图是关于暴露、结局以及其他相关变量之间假设关系的图形表示，当存在多个混杂因素时，有向无环图可以将各变量间的关系更加直观地表示出来，帮助研究人员更加全面地识别混杂因素。有向无环图是由节点和箭头组成，每个箭头的起点被称为父节点而该箭头指向的节点被称为子节点。当一个有向无环图中存在“后门”路径时，提示研究中存在混杂因素，“后门”路径是指存在一个混杂因素既指向研究因素又指向结局，从而导致研究因素和结局之间表现出相关性。如图 1 所示，在有向无环图中存在年龄这一因素既指向口服避孕药，又指向心肌梗死，该路径就是一个“后门”路径。“后门”路径表明年龄既与研究因素有关又与结局有关，是研究中的一个混杂因素。

控制混杂因素的过程实际上就是切断“后门”路径，从而排除混杂因素的干扰。控制混杂因素的过程可以看作是固定混杂因素的值，当混杂因素的值给定之后，暴露和结局间的相关性就与混杂因素无关，相关性就能够反映出因果性。

4 利用机器学习控制混杂因素的方法

混杂因素的存在有时会导致错误的因果关联，根据 Yule - Simpson 悖论，当忽略了第三个变量时，两个变量间的相关性可能会从正相关变为负相关。因此，Yule - Simpson 悖论表明，良好的混杂因素控制方法有助于我们推断出正确的因果关系。Setoguchi 等人利用机器学习算法估计倾向性得分，从而达到控制混杂因素的目的[10]。该研究利用模拟的队列数据比较了 Logistic 回归和机器学习建立的倾向性评分模型，结果表明，由机器学习方法构建的倾向性评分模型具有更小的估计误差，能够更好地控制混杂因素。 传统流行病学中控制混杂因素的方法包括限制、匹配、随机化、分层分析、多元分析等，本研究介绍了机器学习与匹配、倾向性评分及工具变量法相结合，提升控制混杂因素的能力。

4.1 基于机器学习的样本匹配方法

匹配是流行病学中常用的控制混杂因素的方法，可以确保某些变量的分布在暴露组和对照组之间相同或尽可能相同，提高估计因果效应的效率[11]。直接匹配的思想是通过计算处理样本和对照样本间的距离，距离越小的样本间差异越小，将距离暴露组样本最近的对照组样本进行匹配。根据数据集的不同特点可以选择不同的距离函数，例如马氏距离、欧式距离等。随着医学数据的不断积累，协变量特征不断增加，基于机器学习算法的直接匹配将会带来更好的匹配效果。

树模型的原理是根据变量特征对样本进行分组，利用回归树可以将样本协变量特征与阈值依次进行比较，从而将不同的样本纳入不同的分组中，确保暴露组和对照组间协变量分布相似或相同。2009 年，Su 等人[12]提出了交互树的概念，交互树是利用随机森林直接识别影响因果效应异质性的变量的重要程度。之后，Chipman[13]等将贝叶斯算法加入到随机森林模型上，提出了贝叶斯加性回归树（Bayesian Additive Regression Trees, BART）。BART 模型可以自动识别变量之间的非线性关系，并且能够估计异质性因果效应[[14]，该方法在估计因果效应方面得到了广泛的应用[15]。BART 模型对超参数规范有显著的鲁棒性，并且适用

于高维度情况[13]。研究者陆续提出了因果树[16]、因果森林[17]及广义随机森林[18]等方法，这些方法与传统的近邻匹配相比能够更好地匹配样本，并且能够更加准确地估计异质性因果效应。此外，深度学习算法也被用于样本匹配。例如，Schwab 等人[19]在最近邻匹配的思想上结合了神经网络方法提出了完美匹配方法（Perfect Match, PM），这是一种通过训练神经网络来对样本进行直接匹配的方法。PM 方法易于实现，可以应对不同场景和数据集，并且不会增加任何超参数或计算复杂度。Diamond 等人[20]利用搜索算法比较不同样本间的马氏距离，从而提出了一种多元匹配方法。该方法可以用于改善样本匹配后协变量平衡的问题。其中，协变量平衡是指暴露组和对照组的协变量具有相同的联合分布。搜索算法的引入改善了传统需要手工迭代检查不同样本间距离，能够更加高效地进行匹配。Louizos[21]还提出了一种用于估计个体因果效应的有限混合模型，该方法能够用于分析潜在变量对于结局的因果效应，并且可以对样本进行分类。

4.2 基于机器学习的倾向性评分方法

倾向性评分（Propensity Score,PS）是一种仅使用协变量评分来衡量一个人接受治疗的可能性的方法。倾向性评分的主要目标是实现协变量的平衡，从而控制评估治疗或暴露的平均效果时的混杂偏差，对标量倾向得分的调整足以消除由于所有观测协变量造成的偏差[22]。该方法由 Rosenbaum 和 Rubin 在 1983 年提出。倾向性评分技术可以将高维度特征压缩为某一个复合特征，可以直接评价暴露组和对照组在背景特征方面是否相似[23]。将倾向性评分用于因果推断过程可以达到控制混杂效应的作用。合理地利用倾向性评分进行匹配、回归、分层，可以在估计因果效应时减小选择偏倚的影响，实现“事后随机化”。当一个暴露个体和一个未暴露个体具有相同或相近的倾向性评分时，可以认为这两个个体的治疗分配不受任何混杂因素的影响，该暴露个体和未暴露个体之间的差异可以用于回答反事实问题，从而推断暴露因素与结局之间的因果关系。

4.2.1 基于树模型的倾向性评分方法

倾向性评分匹配是最常见的利用倾向性评分的方法。利用决策树可以对样本进行倾向性评分匹配，决策树是机器学习中用于分类和回归的一种非监督学习方式，可以帮助直接进行倾向性评分匹配。利用决策树对一组个体进行分类时，决策树可以将个体划分为多个叶子节点，在每个叶子节点内分类的所有数据点都具有相似的分类概率，因此，决策树可以直接将研究对象分为暴露组和对照组[24]。

此外，通用梯度回归模型（Generalized Boosted Regression Model, GBM）也已应用于估计倾向性评分。在考虑大量预测因素的情况下，GBM 也可以产生平衡不同组间协变量分布的模型[25]。GBM 通过其迭代程序，找到使处理组和对照组之间达到最佳平衡的倾向性评分模型[26]，从而实现研究对象的“随机化”。随机森林、增强型分类和回归树等方法基于协变量对数据进行递归分区来分配治疗组和对照组，从而合理评估变量间的因果效应。其中随机森林的方法还可以有效处理协变量中的缺失数据[27]。在流行病学研究中，随机森林算法已经被广泛用于构建疾病预测模型[28,29]。

4.2.2 基于深度学习的倾向性评分方法

将深度学习算法引入倾向性评分匹配也是当前因果推断中控制混杂的趋势之一。Ghosh[30]等人考虑利用生成式对抗网络（Generative Adversarial Networks, GAN）生成合成数据作为未被匹配的实验个体的对照，从而推断出两组间的差异，判断因果关系。倾向-退出的深度反事实网络(Deep Counterfactual Network with Propensity-Dropout，DCN-PD)是一种多任务学习方法，可以用来减少暴露组和对照组之间的选择偏移。有研究者使用一个深度多任务网络来模拟被试的潜在结果，该网络包含一组事实和反事实结果之间的共享层，以及一组特定结果层。通过倾向性评分计算每个样本被排除的概率，并使用该概率对网络进行正则化，最后利用正则化后的网络输出暴露组和对照组，从而达到控制混杂的目的[31]。

4.2.3 基于神经网络倾向性评分方法

神经网络算法也被用于倾向性评分匹配，Shalit[23]等人开发出了一种可以用于估计个体治疗效果的神经网络——TARNet，研究者利用一个双头多任务模型来估计二元治疗的效果，其中每个样本被分配一个特定的权重，从而平衡了治疗组和对照组之间的混杂因素。

Claudia 等人[32]在 TARNet 的基础上，提出了 Dragonnet，该方法是对 TARNet 进行了修改，利用了倾向性评分来对模型进行正则化修饰，从而避免了模型过拟合。还有研究者提出了 BCAUS[33]方法，这是一种自动因果推理方法，该模型通过指定治疗的错误预测以及逆概率加权变量之间的不平衡程度进行惩罚，之后再与传统的基于倾向得分的方法结合使用，估计变量间的因果效应。其中，逆概率加权（Inverse Probability Weighting ,IPW）是根据倾向性评分计算得到的，为研究对象接受其实际接受的暴露的条件概率的倒数，即 $\frac{1}{1-PS}$，其中 PS 是该研究对象的倾向性评分。逆概率加权与倾向性评分一样，用于评估患者分配到暴露因素的概率，从而平衡暴露组和对照组间的混杂因素。

尽管利用倾向性评分可以较好地控制研究中的混杂因素，但在利用倾向性评分对研究对象进行匹配时，容易将所有因素都当作混杂因素，但并不是所有的因素都是混杂。因此利用倾向性评分控制混杂因素容易导致数据间信息量的降低。为解决这一问题，况琨[34]等提出了一种数据驱动的变量分解（Data-Driven Variable Decomposition ,D2VD）算法，该算法可以自动分离混杂因素和调整变量，同时估计治疗效果。

4.3 基于机器学习的工具变量法

工具变量（Instrumental Variable，IV）法是利用一个额外的工具变量识别变量间的因果关系。工具变量法由 PG. Wright 在 1928 年提出，早期主要应用于经济学和社会学领域，后来用于流行病学研究中混杂因素的控制。工具变量法的基本原理是利用一个只与暴露因素相关，而与其他混杂因素无关的变量来评估暴露和结局之间的因果关系。工具变量的质量是工具变量分析的核心问题[35]，因此，越来越多的研究利用机器学习方法来构建工具变量，从而提高工具变量的质量。

Bargagli 等人[36]将贝叶斯方法与工具变量法相结合，提出工具变量贝叶斯因果森林（Bayesian Causal Forest with Instrumental Variable ,BCF-IV）方法，用于发现和估计异质性因果效应。BCF-IV 是建立在贝叶斯加性回归树上的一个半参数贝叶斯回归模型。Hartford 等人[37]提出了一种利用深度学习网络估计工具变量的方法，将工具变量法分为两个阶段，通过调整损失函数来估计因果效应。Singh 等人[38]提出了一种“Machine Learning Instrument Variables(MLIV)”算法，利用了机器学习算法的非线性建模方式和正则化方法来优化工具变量的构造过程，能够在构造工具变量的同时进行因果推断。

此外，虽然工具变量法可以帮助我们识别变量间的因果关系，但当工具变量构建缺乏精确性时，会导致因果推断结果也缺乏一定的精确性。常见的提高工具变量精确性的方法包括使用多种工具或利用近似最优工具。Belloni[39]等人提出了一种基于“多工具”的渐进方案，利用正则化的机器学习方法构建工具变量，通过将 Lasso 回归的方法与工具变量的估计相结合可以提高构建工具变量的收益。

机器学习方法能够提升混杂因素的控制效果，但可能会存在过拟合等问题。表 1 对各种机器学习方法在控制混杂因素中应用的领域、实例及其优缺点进行了总结。

表 1.机器学习方法及其应用

方法	应用领域	应用	特点（优点/缺点）
决策树	倾向性得分匹配	抗生素治疗和儿科患者死亡率及肾毒性的因果关系。[40]	**优点：** 可直接划分试验组和对照组；可以用来处理分类、有序、连续和缺失数据 **缺点：** 对于异常值和变量的单调变换不敏感；在建模平滑函数和主效应方面存在困难；可能存在过拟合
广义增强模型	倾向性得分匹配	青少年缓刑对于药物滥用的疗效[41]	**优点：** 计算速度快；可以分析非线性效应和交互作用项；可用于拟合平滑模型 **缺点：** 仅能分析有限个因素间的交互作用；
深度学习	倾向性得分匹配 直接匹配 工具变量	低出生体重早产儿中婴儿认知测试得分的影响因素[31] 专家就诊对儿童认知发展的影响[19] 机票价格与客户是否购买的关系[37]	**优点：** 能够处理高维、非线性/非平行治疗分配；可以与传统方法结合使用；可通过调整损失函数来估计因果效应，避免了模型过拟合 **缺点：** 只适用于不存在潜在混杂因素的情况
神经网络	倾向性得分匹配	探究抗糖尿病药物对高水平糖化血红蛋白的影响（HbAlc）[33]	**优点：** 能够处理协变量较多的情况；高度自动化；拟合速度较快； **缺点：** 在少数情况下，难以保证所有协变量平衡
Lasso 回归	工具变量	各种社会经济因素对新冠肺炎病毒传播的影响[43]	**优点：** 避免了模型过拟合的情况 **缺点：** 难以分离出每一项干预的影响
搜索算法	直接匹配	探索加入某职业培训项目的人参与工作的类型的影响因素[20]	**优点：** 较传统方法更加高效；可解决协变量平衡问题 **缺点：** 当治疗分配的估计概率接近 0 或 1 时模型效果不佳
生成式对抗网络	倾向性得分匹配	探索与抗生素耐药性相关基因[30]	**优点：** 可解决样本量不足和权重不稳定的问题；适用于高维数据 **缺点：** 无法克服由于数据不平衡而产生的偏差

5 结语

机器学习方法应用于因果推断领域，在控制混杂因素进而得到合理的因果效应方面发挥着重要作用。本文介绍了因果推断、混杂因素的定义、混杂因素对因果推断的影响及常见的混杂因素识别方法。重点介绍了三种控制混杂因素的方法，即基于机器学习的样本匹配方法、基于机器学习的倾向性评分方法、基于机器学习的工具变量法，以及这些方法如何提升估计因果效应的能力。在基于机器学习的倾向性评分方法中，重点介绍了基于树模型的倾向性评分方法、基于深度学习的倾向性评分方法和基于神经网络的倾向性评分方法。

在因果推断混杂因素控制方面，机器学习较传统的简单线性模型更具有灵活性，能够用来分析暴露和结局间的非线性关系，从而更加全面地评估暴露和结局间的因果关系，提升预测的准确性。在分析高维数据时，机器学习方法能够更好地控制和识别混杂因素，提升暴露组和对照组间的匹配效果，实现“随机化”。此外，许多机器学习算法利用了模型集成的方法来降低单一方法可能存在的预测偏差，使得预测结果更加精准。但是机器学习方法在因果推断中控制混杂因素目前仍然存在不足。包括机器学习的过拟合、训练样本集构建、结果可解释性等问题。由于机器学习模型通过纳入大量参数和复杂的非线性关系提升预测能力，这可能会导致模型出现过拟合的问题，使得机器学习模型进行外部验证时效果不佳。若加入惩罚项来消除模型中的过拟合问题容易造成因果效应估计量的一致性被破坏，当前机器学习领域提出了双重机器学习、样本分割等方法来尝试解决该问题。此外，机器学习模型是一个“黑箱”模型，研究者通常难以了解模型构建过程，导致预测结果缺乏可解释性。因此，在利用机器学习方法定义因果关系时应更加谨慎，需要考虑多种证据后才能得出确切的结论。机器学习算法作为一种统计推断的方法，需要在正确的因果推断框架内使用才能达到最佳的效果。在因果推断领域，良好的试验设计方案是正确估计因果效应的关键所在。

参考文献:

[1] 黄丽红,赵杨,王陵等.获得现实世界证据的因果推断统计学思考[J].中国临床医学,2021,28(5):738–743
[2] 詹思延.流行病学(第八版)[M].北京:人民卫生出版社,2017:87-93.
[3] 陶秋山,李立明.基于虚拟事实理论的病因效应模型[J].中华流行病学杂志,2014,23(1):60–62.
[4] PEARL J.Causal diagrams for empirical research[J].Biometrika,1995,82(4):669–688.
[5] Rex B. Kline. Principles and Practice of Structural Equation Modeling: Fourth Edition [M]. New York City: Guilford Press,2016: 36-38.
[6] PEARL J. Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution [A]. 11th ACM International Conference on Web Search and Data Mining. 2018,3-3
[7] RUBIN D B. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.[J]. Journal of Educational Psychology, 1974, 66(5): 688–701.
[8] HOWARDS P . An Overview of Confounding. Part 1: The Concept and How to Address It[J]. Acta Obstetricia et Gynecologica Scandinavica, 2018, 97(4): 394–399.
[9] 胡永华, 耿直. 关于混杂概念的讨论[J]. 中华流行病学杂志, 2001, 22(6): 459–461.
[10] SETOGUCHI S, SCHNEEWEISS S, BROOKHART M A, et al. Evaluating Uses of Data Mining Techniques in Propensity Score Estimation: A Simulation Study[J]. Pharmacoepidemiology and Drug Safety, 2008, 17(6): 546–555.
[11] MANSOURNIA M A, JEWELL N P, GREENLAND S. Case–Control Matching: Effects, Misconceptions, and Recommendations[J]. European Journal of Epidemiology, 2018, 33(1): 5–14.
[12] SU X, TSAI C-L, WANG H, et al. Subgroup Analysis via Recursive Partitioning[J]. JOURNAL OF MACHINE LEARNING RESEARCH, 2009, 10:141-158.
[13] CHIPMAN H A, GEORGE E I, MCCULLOCH R E. BART: Bayesian additive regression trees[J]. The Annals of Applied Statistics, 2010, 4(1): 266-298.
[14] GREEN D P, KERN H L. Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees[J]. Public Opinion Quarterly, 2012, 76(3): 491–511.
[15] LEONTI M, CABRAS S, WECKERLE C S, et al. The Causal Dependence of Present Plant Knowledge on Herbals--Contemporary Medicinal Plant Use in Campania (Italy) Compared to Matthioli (1568)[J]. Journal of Ethnopharmacology, 2010, 130(2): 379–391.
[16] ATHEY S, IMBENS G. Recursive partitioning for heterogeneous causal effects[J]. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(27): 7353–7360.
[17] WAGER S, ATHEY S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests[J]. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 2017,113(523): 1228-1242.
[18] ATHEY S, TIBSHIRANI J, WAGER S. Generalized Random Forests[J]. arXiv preprint arXiv:1610.01271.
[19] SCHWAB P, LINHARDT L, KARLEN W. Perfect Match: A Simple Method for Learning Representations For Counterfactual Inference With Neural Networks[J]. arXiv preprint arXiv:1810.00656.
[20] DIAMOND A, SEKHON J S. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies[J]. Review of Economics and Statistics, 2013, 95(3): 932–945.
[21] LOUIZOS C, SHALIT U, MOOIJ J, et al. Causal Effect Inference with Deep Latent-Variable Models[A]. 31st Annual Conference on Neural Information Processing Systems (NIPS), 2017,30: 6449–6459.
[22] 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.
[23] SHALIT U, JOHANSSON F D, SONTAG D. Estimating Individual Treatment Effect: Generalization Bounds and Algorithms[A]. 34th International Conference on Machine Learning, 2017,70: 3076–3085.
[24] COOK E F, GOLDMAN L. Asymmetric Stratification. An Outline for an Efficient Method for Controlling Confounding in Cohort Studies[J]. American Journal of Epidemiology, 1988, 127(3): 626–639.
[25] MCCAFFREY D F, RIDGEWAY G, MORRAL A R. Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies[J]. Psychological Methods, 2004, 9(4): 403–425.
[26] MCCAFFREY D F, GRIFFIN B A, ALMIRALL D, et al. A Tutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models[J]. Statistics in Medicine, 2013, 32(19): 3388–3414.
[27] ZHAO P, SU X, GE T, et al. Propensity Score and Proximity Matching Using Random Forest[J]. Contemporary clinical trials, 2016, 47: 85–92.
[28] YANG L, WU H, JIN X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China[J]. Scientific Reports, 2020, 10: 5245.
[29] DI CASTELNUOVO A, BONACCIO M, COSTANZO S, et al. Common Cardiovascular Risk Factors and In-Hospital Mortality in 3,894 Patients with COVID-19: Survival Analysis and Machine Learning-Based Findings from the Multicentre Italian CORIST Study[J]. Nutrition, Metabolism, and Cardiovascular Diseases: NMCD, 2020, 30(11): 1899–1913.
[30] GHOSH S, BOUCHER C, BIAN J, et al. Propensity score synthetic augmentation matching using generative adversarial networks (PSSAM-GAN)[J]. Computer methods and programs in biomedicine update, 2021, 1: 100020.
[31] ALAA A M, WEISZ M, VAN DER SCHAAR M. Deep Counterfactual Networks with Propensity-Dropout[J]. arXiv preprint arXiv:1706.05966.
[32] SHI C, BLEI D M, VEITCH V. Adapting Neural Networks for the Estimation of Treatment Effects[A]. 33rd Conference on Neural Information Processing Systems (NeurIPS), 2019,32: 2503-2513.
[33] BELTHANGADY C, STEDDEN W, NORGEOT B. Minimizing bias in massive multi-arm observational studies with BCAUS: balancing covariates automatically using supervision[J]. BMC Medical Research Methodology, 2021, 21: 190.
[34] KUANG K, CUI P, ZOU H, et al. Data-Driven Variable Decomposition for Treatment Effect Estimation[J]. IEEE Transactions on Knowledge and Data Engineering, 2020: 1–1.
[35] BROOKHART M A, WANG P, SOLOMON D H, et al. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable[J]. Epidemiology (Cambridge, Mass.), 2006, 17(3): 268–275.
[36] BARGAGLI-STOFFI F J, DE-WITTE K, GNECCO G. Heterogeneous Causal Effects with Imperfect Compliance: A Bayesian Machine Learning Approach[J].arXiv preprint ArXiv:1905.12707
[37] HARTFORD J, LEWIS G, LEYTON-BROWN K, et al. Deep IV: A Flexible Approach for Counterfactual Prediction[A]. 34th International Conference on Machine Learning,2017,70: 1414-1423.
[38]Singh, Amandeep and Hosanagar, Kartik and Gandhi, Amit, Machine Learning Instrument Variables for Causal Inference[EB/OL]. [2021–12–20]. https://ssrn.com/abstract=3352957
[39] BELLONI A, CHEN D, CHERNOZHUKOV V, et al. Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain[J]. 2010.80(6): 2369-2429.
[40] TAMMA P D, TURNBULL A E, HARRIS A D, et al. Less Is More: Combination Antibiotic Therapy for the Treatment of Gram-Negative Bacteremia in Pediatric Patients[J]. JAMA pediatrics, 2013, 167(10): 903–910.
[41] MCCAFFREY D F, RIDGEWAY G, MORRAL A R. Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies[J]. Psychological Methods, 2004, 9(4): 403–425.

[42] QIU Y, CHEN X, SHI W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China[J]. Journal of Population Economics, 2020: 1–46.