

断点回归设计在临床治疗性研究中的应用



张华, 王晓晓, 曾琳, 李楠, 陶立元, 石岩岩, 褚红玲, 倪凯文, 赵一鸣

北京大学第三医院临床流行病学研究中心(北京 100191)

【摘要】 随机对照试验的证据等级较高, 但因其实实施成本较高、外部真实性较低和伦理要求等原因而限制了其在临床开展和应用, 而传统的观察性研究由于存在各种混杂因素导致内部真实性降低, 从而降低了证据等级。断点回归设计 (regression discontinuity design, RDD) 是在自然条件下观察比较阈值附近的人群, 其控制混杂的能力仅次于随机对照试验, 可提供较高等级证据。它适用于干预 (或暴露) 与某连续变量的数值直接相关的情况, 如 HIV 患者是否需要抗逆转录病毒治疗主要取决于其 CD4 细胞计数是否低于 $200/\mu\text{L}$ 。因为连续变量的测量存在随机误差, 在阈值附近是否给予干预是接近随机的, 阈值附近干预和非干预组患者基线应该是均衡可比的。根据这一假设, 比较阈值附近人群的结局, 即可估计干预 (或暴露) 与结局的因果效应。RDD 在医学中主要适用于分类结局的研究, 其中两阶段最小二乘法、基于似然比的估计方法、贝叶斯方法是较常用的模型估计方法。然而, RDD 的适用条件和对样本量的要求限制了其在医学中的广泛应用, 随着数据可及性的提高和真实世界研究的发展, RDD 将更多地应用于临床研究中。

【关键词】 断点回归设计; 阈值; 观察性研究; 混杂因素; 平均因果效应

The application of regression discontinuity design in clinical therapeutic research

ZHANG Hua, WANG Xiaoxiao, ZENG Lin, LI Nan, TAO Liyuan, SHI Yanyan, CHU Hongling, NI Kaiwen, ZHAO Yiming

Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, 100191, P.R.China

Corresponding author: ZHAO Yiming, Email: yimingzhao115@163.com

【Abstract】 The level of evidence in randomized controlled studies is high, but it cannot be widely applied due to its high cost, external authenticity, ethics and other reasons. The traditional observational studies reduce the internal authenticity due to various confounding factors, and the level of evidence is low. Regression discontinuity design (RDD) is a design that observes and compares outcome of object around the threshold under practical clinical conditions. Its ability to adjust confounding factors is second only to that of randomized control studies. It can be used in cases where the intervention (or exposure) is directly related to the value of a continuous variable, for example, whether an HIV patient needs antiretroviral treatment mainly depends on whether the CD4 cell count is lower than $200/\mu\text{L}$. Because the measurement of continuous variables has random error, whether intervention is given near the threshold is close to random, and the baseline of patients in the intervention group and non-intervention group near the threshold should be balanced and comparable. Based on this assumption, the causal effect of intervention (or exposure) and outcome can be estimated by comparing the outcomes of populations near the threshold. RDD is mainly applicable to the study of classification outcomes in medicine, among which two-stage least square method, likelihood ratio based estimation method and Bayesian method are more commonly used model estimation methods. However, the application conditions of RDD and the requirement of sample size limit its extensive application in medicine. With the improvement of data accessibility and the development of real world research, RDD will be more widely used in clinical research.

【Key words】 Regression discontinuity design; Threshold; Observational study; Confounding factor; Average causal effect

DOI: 10.7507/1672-2531.201807146

基金项目: 国家科技支撑计划项目 (编号: 2013BAI09B14)

通信作者: 赵一鸣, Email: yimingzhao115@163.com

1 背景

随机对照试验 (randomized controlled trial, RCT) 是临床研究因果推断的最佳研究设计类型, 但其实施受到多方面限制。首先, RCT 的时间成本和经济成本非常高, 实施过程比较困难; 其次, 其外部真实性受到质疑, 原因是研究的人选和排除标准比较严格, 多数 RCT 结果是在理想情况下估计干预措施与结局的因果关系, 而非在真实环境中; 再者, 某些 RCT 方案因不符合伦理原则而不能组织实施。当临床治疗性研究在条件受限时需要调整技术路线, 其中采用观察性研究方案进行因果关系推断是学术界探讨的一个方向。在观察性研究中, 准确因果推断需要一个假设: 没有混杂因素干扰。违反这条假设会导致因果效应的估计出现偏倚, 从而降低了观察性研究的内部真实性。

断点回归设计 (regression discontinuity design, RDD) 是在约束条件下分析变量之间因果关系的实证方法之一^[1,2], 其已在经济学等领域广泛应用。近年来 RDD 逐渐应用于医学领域, 如临床护理、健康行为方式、社会决定因素、环境暴露等, 在临床治疗性研究中也出现了成功运用的案例, 成为我们学习理解 RDD 的窗口。本文将介绍 RDD 的原理, 并举例其在治疗性研究中如何实现因果关系推论。

2 RDD 的定义

当是否实施某种暴露或干预措施是由测量的连续分布变量的阈值决定时, 可考虑采用 RDD 来进行因果关系推断。阈值决定原则在临床医学中很常见, 如胆固醇、血糖或出生体重等, 当患者的测量值高于或低于某个阈值时, 就会给予该患者相应的干预措施。而决定是否给予干预措施的连续变量会因测量误差、抽样变异性及环境温度等偶然因素而受到影响。这种测量值的误差是随机产生的, 这意味着在阈值之上和低于阈值的患者在所有观察到和未观察到的基线资料应该是均衡的, 这与 RCT 相似^[3]。而通过比较这些患者结果, 可进行因果关系推断, 即 RDD。RDD 阈值是指南或政策规定的界值, 如正常值范围的上下限, 或某项治疗的应用指征等。RDD 可分为两类^[4], 第一类, 临界值是确定的 (sharp), 即在临界值一侧的所有观测点都接受了干预; 反之, 在临界值另一侧的所有观测点都没有接受干预, 称为确定型 RDD (sharp regression discontinuity, SRD)。此时, 接受干预的概率从临界值一侧的 0 跳转到另一侧的 1。以艾滋

病抗逆转录病毒治疗为例, 如果 CD4 细胞计数低于阈值 200/ μL , 那么就会对患者实施抗逆转录病毒治疗, 否则不给予治疗^[5]。第二类, 在多数情况下, 干预方案不是由连续变量的阈值完全决定的, 而是有概率的, 即临界点是模糊的 (fussy), 即在临界值附近, 接受处置的概率是单调变化的, 称为模糊型 RDD (fussy regression discontinuity, FRD)。如临床医生仅在一定程度上根据阈值决定原则对患者进行治疗, 而这种判断主要根据医生自身的临床判断, 即是 FRD。以艾滋病抗逆转录病毒治疗为例, 如果 CD4 细胞计数低于阈值, 或患者出现较重的临床症状, 那么他们就会获得抗逆转录病毒治疗, 这种类型即为 FRD^[6]。Hahn 等^[7]证明了在一定假设下, 无论是哪一种类型的断点回归, 都可使用临界值附近的样本情况来推断暴露和结局之间的因果关系。

3 RDD 因果效应估计方法的理论及应用

3.1 因果关系推断的理论基础

本文简要介绍如何使用 RDD 进行因果推论的方法。为方便论述, 我们假设存在两种干预方案, 当然 RDD 也可推广到多种干预方案甚至连续暴露变量。我们以一个 RDD 在临床研究应用中较经典的真实案例^[5,8]来进行说明: 对于 HIV 患者来说, 患者病情进展到哪一阶段开始进行抗逆转录病毒治疗 (anti-retroviral therapy, ART) 一直是存在争议的问题^[9,10]。有研究认为在 CD4 细胞计数在 200 以下时使用 ART, 也在研究者认为在 250 ~ 350 之间就应该进行治疗, 甚至有研究认为低于 550 就需要进行治疗。虽然有相关的 RCT 研究, 但检验效能不足, 未能解决该问题。在非洲国家的实际医疗常规中, 如果 CD4 细胞计数低于 200, 或有 AIDS 四期相关的并发疾病, 或符合某个国家的标准时, 患者将接受 ART。我们将影响干预方案分配的连续变量 (如 CD4 细胞计数) 记为变量 Z , 其阈值 (开始治疗的临界值如 200/ μL) 记为 c ; 结局 (如是否死亡) 记为 Y , 进行治疗记为 1, 不进行治疗记为 0; 治疗与未治疗的结果记为 $Y(1)$ 和 $Y(0)$, 治疗与未治疗的期望结果记为 $E(Y(1))$ 和 $E(Y(0))$, 单个患者的参数用 i 表示。

在上述案例中, 我们想观察 ART 是否可降低 HIV 患者的病死率, 其最佳研究设计类型为 RCT, 但实施难度较大。如果采用观察性研究, 影响患者死亡的因素有很多, 如未能识别和处理设计与分析过程中的混杂因素, 那么 ART 与死亡的因果效应

估计将存在较大偏倚。而一般观察性研究设计不能识别和处理未知的影响因素,这也是观察性研究设计所面临的挑战。RDD通过对阈值 200/ μ L 周围的数据进行分析,可避免和降低各种未知混杂因素的影响。在阈值 200/ μ L 附近的一个小范围内(如 195 ~ 205/ μ L),由于测量误差的存在,使低于阈值 200/ μ L 或高于 200/ μ L 的测量结果是随机出现的,此时对 CD4 细胞计数在 200/ μ L 以下的患者进行 ART 治疗,那这个范围内的患者是否进行治疗是接近随机分配的。再进行一个极端的假设,当这个范围趋于 0 时,即所有患者 CD4 计数约等于 200/ μ L 时,其治疗的分配是独立于其它因素的,类似于随机分配的结果,此时进行因果关系推断就可避免已知和未知混杂因素的影响。

采用 RDD 进行因果推断有两个假设条件^[4,5]:
 ① 连续变量 Z 在阈值 c 附近是连续的,即病例的 CD4 计数在 200/ μ L 处是连续的,呈现均匀分布,如果出现不连续,即分布不均匀,则存在人为操纵 CD4 计数的可能,这会导致治疗方案不是随机分配的,因此不能使用 RDD。
 ② 基线变量与结果 Y 的关系在阈值 c 处是连续的。如果基线资料在 200/ μ L 处不连续,如年龄不连续,在进行因果推断时不能区分到底是年龄对死亡有影响还是是治疗对死亡有影响,因果推断无法确定。如满足这些假设条件,则所有潜在的混杂因素都在阈值附近的一个小范围内保持均衡,治疗组在阈值附近的预期结果可估计,对照组在阈值处的预期结果也可估计,此时两组在阈值处预期结果的差异与治疗存在因果关系。尽管在阈值 c 处的连续性看起来像是一个非常强的假设,但事实上连续变量 Z 存在测量误差,另外多数情况下无法精确地按是否低于阈值对患者进行干预方案的分配,故上述假设条件在临床实际工作中较容易满足。

大多数研究应用断点回归估计两组在阈值附近的预期结果差值,称为平均因果效应(average causal effect, ACE)^[4]。在 SRD 中,患者因为在阈值上(或下)而被分到不同的组。图 1 显示了结局的连续条件期望函数。纵轴 Y 为预期病死率,横轴 Z 为连续变量 CD4 计数。上方曲线为未治疗组病死率与 CD4 计数的函数关系,下方曲线为治疗组病死率与 CD4 计数的函数关系,实线表示观察到的数据,虚线表示预测的条件期望函数,而预测的条件期望函数并没有被实际观察到。可通过已观察到的数据,在阈值处求极限来识别两组阈值处的效应差异(ACE)大小。

$$ACE_{SRD} = \lim_{\{z \uparrow c\}} E[Y_i Z_i = z] - \lim_{\{z \downarrow c\}} E[Y_i Z_i = z] \quad (1)$$

在 FRD 中,公式(1)就是意向性治疗(intention to treat, ITT)分析效果,即低于阈值 c 进行干预、高于阈值 c 不进行干预,进行分析。ITT_{FRD} 衡量的是由阈值规则决定的 ITT 的效果。此外,治疗本身的依从性会对结果产生重要影响,考虑这种影响时,其结果称为依从者平均因果效应(complier average causal effect, CACE_{FRD}),即将 ITT_{FRD} 除以阈值处的治疗概率差异^[4]。

$$CACE_{RDD} = \frac{\lim_{\{z \uparrow c\}} E[Y_i Z_i = z] - \lim_{\{z \downarrow c\}} E[Y_i Z_i = z]}{\lim_{\{z \uparrow c\}} P[T_i = 1 | Z_i = z] - \lim_{\{z \downarrow c\}} P[T_i = 1 | Z_i = z]} \quad (2)$$

当公式(2)的分母等于 1 时,这时就是确定型断点回归,此时 ITT_{FRD} = CACE_{FRD} = ACE_{SRD}; 当分母为 0 时,没有发现不连续点,因果效应差异也没有被识别。

在 SRD 和 FRD 中,都在阈值处计算因果效应大小。如果可证明,无论 CD4 计数是多少,ART 治疗可降低的病死率是恒定的,此时阈值处计算得到的 ACE 可认为是在整个连续变量 Z 范围内治疗效应的平均值。而此时 RCT 可被认为是一个特殊的 RDD,在这个设计中,随机数字是连续变量 Z。如果治疗效应在连续变量各处是不均匀的,那么 RDD 应该被解释为在阈值 c 处局部的治疗效应。由于对 Z 的测量存在随机误差,因此,在连续变量 Z 的单个值中确定的治疗效果可被认为是一个更大范围内(Zi-se, Zi+se)(注:se 为允许误差)真值的加权平均值。此外,即使在整个 Z 范围内的预期效应是不同的,它也可近似于在阈值周围的某个适当范围内的平均值。可通过评估在 c 处的曲线斜率是否变化来检测接近阈值的效应异质性。需要小心的是,局部效应可能无法推广到远离阈值的人群

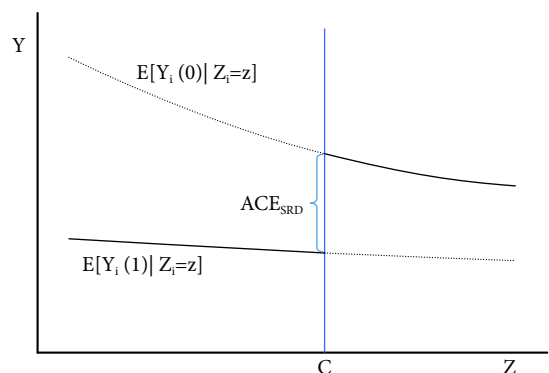


图 1 结局的连续条件期望函数图

中。在阈值处, 替代局部识别的另一种方法是通过拟合整个 Z 范围内的条件期望函数来估计一个整体平均因果效应, 这需要知道两组中 Y 与 Z 的函数关系。

3.2 因果效应估计方法

在 RDD 中进行效应估计需要估计公式 (1) 和 (2) 的极限。然而, 如果曲线斜率在阈值的任意一端是非零的, 那么这些平均值对真实平均值的估计存在偏差。估计局部线性 (或立方) 回归模型可很好地解决这个问题。在实际应用时, ACE_{SRD} 和 ITT_{FRD} 估计通常是使用在阈值范围的数据拟合结局与连续变量 Z 的函数, 并计算阈值 c 处的预测值差异。一般情况下要符合如下模型:

$$E[Y_i|Z_i] = \beta_0 + \beta_1(Z_i - c) + \beta_2 \cdot 1[Z_i < c] + \beta_3(Z_i - c) * 1[Z_i < c] \quad (3)$$

β_1 为阈值以上直线的斜率, $\beta_1 + \beta_3$ 为阈值以下直线的斜率, β_2 为阈值处的差异。交互项 $\beta_3(Z_i - c) \times 1[Z_i < c]$ 表示治疗效果是可异质的。除非效应的函数关系已知, 否则有限样本的效应估计总是存在偏差, 但可通过使用更小的范围来估计模型, 并通过评估 z 的高阶多项式来提高估计结果的稳健性, 从而解决这个问题。

在 RDD 中, 对数据进行无偏差的图形呈现是必要的。断点回归必须满足以下两个条件: ① 数据的密度应该在阈值附近连续。如果患者 (或研究者) 能够精确地操纵连续变量 Z , 则不能使用断点回归估计。② 基线的协变量应该在阈值上保持平衡 (即连续)。正如在 RCTs 中, 基线观测值的均衡性意味着被分配到治疗组和对照组的患者具有可交换性^[11]。

通常 RDD 使用两阶段最小二乘法进行因果效应的估计^[12]。其优点是可得 ACE 的无偏差的估计, 一般的统计软件都可实现, 但对标准误差的估计准确性较差。基于似然比的估计方法也可对因果进行无偏差的估计, 且对效应误差的估计也较准确。贝叶斯方法在 RDD 中有较大优势^[13], 首先, 贝叶斯方法可基于之前 RCT 的结果, 设置先验概率使贝叶斯估计更加准确; 其次贝叶斯使用 MCMC 方法使模型结构更加灵活, 可提供与前两种方法一样的准确性, 特别是可设置条件概率和先验分布进行估计, 因此可用于阈值宽度较大的研究。

3.3 非线性模型的 RDD

RDD 最常应用于经济学领域, 经济学结局指标多为连续性, 所以 RDD 通常使用线性回归模型, 很少有二分类、计数、生存模型的 RDD 应用。

而二分类、计数、生存模型在临床研究中更加常用。对于 ACE_{SRD} 和 ITT_{FRD} 来说, 非线性模型的 RDD 的扩展是非常简单的。只需要将连续性 RDD 中条件期望值换成经 log 或 logit 转换的条件期望值即可^[14]。在上述案例中, 可对病死率进行 logit 转换, 采用 logit (p) 与 CD4 计数进行模型拟合, 从而进行因果关系推断。另外, RDD 可应用到其他的不依靠均值的参数, 如乘法模型和量化回归估计的边际效应等。

3.4 应用案例

如前所述, 在 HIV 疾病进展过程中, 何时进行抗逆转录病毒治疗是一个非常具有争议的问题。在开始 ART 治疗的时机选择上, 没有较多的 RCT 证据。几个较大的队列研究都存在一定的偏倚, 如治疗决策的选择原因及排除了所有未进行 ART 治疗患者等。Bor 等^[6]使用南非的一个大型队列研究数据 (2007 ~ 2011 年), 估计了在阈值 (CD4 计数在 $200/\mu\text{L}$) 处 ART 治疗对患者死亡率的影响。该研究获取了患者的基本信息、治疗信息和随访信息, HIV 患者 CD4 计数在 $200/\mu\text{L}$ 以下或者有艾滋病 4 期相关的并发症即开始 ART 治疗。在注册的 4 391 人中 2 874 人开始 ART 治疗, 共随访 13 139 人/年, 以 10 个细胞为宽度分组, 分别计算每组的病死率, 使用病死率与 CD4 计数的分组拟合模型, 分别得到干预组和非干预组病死率与 CD4 细胞计数的关系, 进而得到在 $200/\mu\text{L}$ 处的患者死亡率差异。相对于在阈值附近且在 $200/\mu\text{L}$ 以上的病例, 在 $200/\mu\text{L}$ 以下的死亡率下降 35% ($ITT_{FRD}=0.65$), 即认为在阈值处 ART 治疗可减少 35% 的死亡率。

4 优点和局限性

作为观察性研究的一种, RDD 有很大的优势。首先, 该方法因果推断的证据等级较高, 有研究者建议将 RDD 与 RCT 并列作为高证据等级的研究设计类型^[1]。Lee 等^[15]认为在不适合开展 RCT 的情况下, RDD 能够有效避免各种偏倚, 特别是降低甚至消除未知的混杂因素影响, 从而反映出变量之间真实的因果关系^[9]。其次, RDD 主要关注阈值附近的因果效应差异, 这也正是临床决策的难点。例如 CD4 计数低于 $200/\mu\text{L}$ 时需要使用 ART, 高于 $600/\mu\text{L}$ 时不需要使用 ART, 但在 $200 \sim 600/\mu\text{L}$ 之间时是否需要服药存在较大的争议, 而 RDD 关注的正是这部分人群的干预效应, 能够给临床决策提供数据支持。

值得注意的是, RDD 并不是完美的方法: ①

相对于传统 RCT, RDD 的统计效能较低, 因为 RDD 只用到邻近阈值处的数据, 要达到与 RCT 相同的统计效能时, RDD 的样本量需要达到传统 RCT 的 3 ~ 4 倍以上^[11]。因此, RDD 更适用于样本量较大的研究及真实世界研究。② 需要预先评价阈值附近协变量的均衡性和连续性, 如果不均衡, 可能不适用 RDD。③ 因为是在阈值处进行因果效应估计, 故在向阈值远处进行推广时, 应进行敏感性分析, 并慎重进行结论^[9]。

我们希望在医学领域更广泛地使用 RDD, 特别是在越来越多地使用计算机观测数据和医保数据的情况下, 海量数据可避免局部回归模型样本量较小导致的统计效能不足的缺点。同时在使用这一方法时, 应注意确保对参数估计是准确的。

参考文献

- 1 Tang Y, Cook TD. Statistical power for the comparative regression discontinuity design with a pretest no-treatment control function: theory and evidence from the national head start impact study. *Eval Rev*, 2018, 42(1): 71-110.
- 2 Maas IL, Nolte S, Walter OB, *et al*. The regression discontinuity design showed to be a valid alternative to a randomized controlled trial for estimating treatment effects. *J Clin Epidemiol*, 2017, 82: 94-102.
- 3 Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ*, 2016, 352: i1216.
- 4 Oldenburg CE, Moscoe E, Bärnighausen T. Regression discontinuity for causal effect estimation in epidemiology. *Curr Epidemiol Rep*, 2016, 3: 233-241.
- 5 Smith LM, Lévesque LE, Kaufman JS, *et al*. Strategies for evaluating the assumptions of the regression discontinuity design: a case study using a human papillomavirus vaccination programme. *Int J Epidemiol*, 2017, 46(3): 939-949.
- 6 Bor J, Ahmed S, Fox MP, *et al*. Effect of eliminating CD4-count thresholds on HIV treatment initiation in South Africa: An empirical modeling study. *PLoS One*, 2017, 12(6): e0178249.
- 7 Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 2001, 69(1): 201-209.
- 8 Bor J, Moscoe E, Mutevedzi P, *et al*. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology*, 2014, 25(5): 729-737.
- 9 Severe P, Juste MA, Ambrose A, *et al*. Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *N Engl J Med*, 2010, 363(3): 257-265.
- 10 Cohen MS, Chen YQ, McCauley M, *et al*. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*, 2011, 365(6): 493-505.
- 11 Walkey AJ, Drainoni ML, Cordella N, *et al*. Advancing quality improvement with regression discontinuity designs. *Ann Am Thorac Soc*, 2018, 15(5): 523-529.
- 12 Bor J, Moscoe E, Bärnighausen T. Three approaches to causal inference in regression discontinuity designs. *Epidemiology*, 2015, 26(2): e28-e30.
- 13 Geneletti S, O’Keeffe AG, Sharples LD, *et al*. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Stat Med*, 2015, 34(15): 2334-2352.
- 14 van Leeuwen N, Lingsma HF, Mooijaart SP, *et al*. Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials. *J Clin Epidemiol*, 2018, 98: 70-79.
- 15 Lee DS, Card D. Regression discontinuity inference with specification error. *J Econom*, 2006, 142(2): 655-674.

收稿日期: 2018-08-02 修回日期: 2018-10-08

本文编辑: 熊鹰