

倾向指数用于中医药非随机设计临床疗效评价的思考

张颖 李迅 费宇彤 刘建平

(北京中医药大学循证医学中心, 北京, 100029)

摘要 尽管随机对照试验是临床干预研究因果推断的金标准,但是在中医药临床研究的现实中,由于伦理、治疗措施的特征、费用、时间以及外部效度问题,随机对照试验往往无法实施甚至在某些情况下不适用。因果推论的本质是一个反事实问题。随机对照试验通过随机化来平衡可能影响干预分配的特征变量。倾向指数可以应用于非随机设计的观察性研究,如前瞻性队列研究,作为诸多特征变量的一个复合函数,用于均衡暴露组 and 对照组之间特征变量的分布。在满足一定前提条件的情况下,倾向指数分析法可以提供干预效果的无偏估计。因此,在中医药观察性的临床研究中,倾向指数可能具有广阔的应用价值和潜力。

关键词 干预;非随机设计;因果推断;倾向指数

Propensity Score and the Intervention Effects of Chinese Medicine in Non-randomized Clinical Study

Zhang Ying, Li Xun, Fei Yutong, Liu Jianping

(Evidence Based Medicine Center, Beijing University of Chinese Medicine, Beijing 100029, China)

Abstract Randomized controlled trial(RCT) is the golden standard design to make causal inference on intervention effects in clinical trials. While, in the research reality of Chinese medicine, RCTs could not be conducted, even not applicable, due to ethical issue, characters of the intervention, cost, time and external validity. The essence of causal inference is counterfactual framework. All kinds of covariates could be balanced in RCTs via randomization. Propensity score(PS) is a composite score which could balance the distribution of covariates between intervention and control group. PS could provide unbiased estimate of intervention effect when critical assumptions are met. For the observational studies of Chinese medicine, such as prospective cohort study, PS will have wide application and huge potential.

Key Words Intervention; Non-randomized design; Causal inference; Propensity score

中图分类号: R195 文献标识码: A doi:10.3969/j.issn.1673-7202.2014.10.005

1 随机对照试验与反事实

从循证医学的证据分级体系来看,随机对照临床试验(Randomized Controlled Trial, RCT)是科学研究探索因果关系的最理想的金标准设计方案,位于证据体的最高级^[1]。因此,在理想情况下,随机对照临床试验是评价医疗干预效果的最佳方式。统计学的传统观点认为,对采用随机化方法得到的试验数据所进行的因果推论,其本质是一个反事实问题。反事实就是在导致结果的原因(例如,影响疾病预后的治疗措施)不存在的情况下会发生的潜在结果(potential outcome)或事件状态(state of affairs)^[2]。所以,在处于干预状态下的对象而言,反事实就是处于控制状态下的潜在结果;对于处于控制状态的对象而言,反事实就是处在干预状态下的潜在结果。在统计学理论界,通常将此概念框架称为因果关系的Neyman-Rubin反事实框架(Neyman-Rubin counterfactual framework of causality)^[3-4]。

既然反事实是一种潜在结果,那么它在真实数据

(real data)中是无法被观测到的,而随机对照试验是如何去获得反事实的呢?最为关键的环节是随机化。随机化创造一种类似于实验条件的研究环境,使得干预组和对照组的分配符合预先设定的概率(例如,对于1:1设计的试验,每个对象将有50%的概率被分配到干预组或对照组),从而保证2组的一切特征变量(协变量),包括可观察的和不可观察的,具有相同的分布。从统计学的角度,认为经过随机分配的干预组和对照组是同质的,即使存在差异,也不应影响到评价结果,因为这种差异往往是由于机遇造成的。由此,可以认为干预组的干预效果和对照组参与干预的效果相同。借助统计学的手段,通过考察样本中未暴露于干预中的对象的结果来探究干预对样本对象(已接受干预的)的因果效应。

2 中医药非随机设计临床研究与反事实

中医药作为防治疾病的措施,已经有几千年的历史^[5]。虽然其产生和发展过程伴随着长期应用于人体的经验探索和实践,但对其对有效性的论证依然无法

回避因果关系的推论过程。

目前,国内已经有大量的中医药随机对照试验的报告发表,但是其质量往往参差不齐,而且引起了国内外的广泛关注^[6]。要改善这一现象,一方面要提高中药随机对照试验的质量和报告规范,另一方面则要探索适宜中医药特色的设计方法。事实上,在中医药的临床研究实践中,随机对照试验并不适用于任何情况。首先,和西方现代医学相比,中医药的理论、诊断、治疗和语言都自成一体^[7],目前全面借鉴国际统一的化学药品临床研究的方法和标准,不利于中药的评价,也不能体现中医药的特点和优势^[8]。同时,在中医药临床试验中采用随机、盲法和安慰剂也会面临一系列伦理问题^[9]。

随机对照试验无法解决治疗个体化的问题,即在试验过程中不允许按照每个对象“望闻问切”的不同情况,随时调整不同组方的配伍和剂量。而使用了辩证论治过程的随机对照试验,也会面临来自研究者的诊疗偏倚和研究结果的外部真实性较差的挑战。还有,与化学药物的随机对照试验不同,到中医医院就诊的患者,对中医药治疗有很强的选择倾向。在这种情况下,无法通过随机化对患者实施非中医药的干预,哪怕是对照组接受单纯西药,对许多慢性病或肿瘤患者来说也是不符合伦理的。从另一个方面看,在中医药医疗实践的一些特殊情况中,如对中医综合治疗效果的评价,非随机化研究是不可或缺的研究方法。例如,当研究的关注点在于考察中医药在更广泛条件下远期疗效和安全性问题时,就需要采用注册试验(registry study)或真实世界研究(real world study),通常是经过长时间的队列观察才能实现。

非随机设计的观察性研究,可以克服诸多的伦理学因素的限制,而且数据更能代表临床实践中实际存在的疾病谱和患者的构成。然而,观察性研究在进行因果推断时能否像随机对照那样去获得反事实呢?现实中,采用非随机化的方法得到的观察性研究数据,往往会在诸多特征变量上出现组间不平衡的状态,这样对于干预效果的分析和判断就会比随机对照试验更加复杂。用于校正组间不平衡状态的常用方法主要是最小二乘(Ordinary Least Squares, OLS)回归,这种方法将重要的特征变量都纳入到回归方程。如果能满足 Gauss-Markov 经典线性回归模型的全部假设,即误差项的均值为零、同方差和互不相关,最小二乘估计将是最佳线性的无偏估计(Best Linear Unbiased Estimator, BLUE)。但遗憾的是,在真实世界中应用最小二乘回归进行因果推断时,这些假设必将面临一系列的质疑。

3 倾向指数的提出及性质

为了克服传统统计方法的局限,倾向指数(P propensity Score, PS)正获得越来越广阔的应用空间。倾向指数,又称倾向值、倾向分数或倾向评分,它的提出缘起于研究者们对社会科学领域中广泛存在的观察性研究的因果关系的推断方法的探讨。Rosenbaum 和 Rubin 在 1983 年首次提出倾向指数这一术语^[10],它是指在给定已观察到的协变量的条件下暴露于干预的概率。倾向指数从统计方法的角度按照反事实框架获得观察性研究的因果推论^[11]。它将观察对象的多个特征变量表示为一个函数,然后通过倾向指数的匹配来均衡暴露组和对照组间的特征变量的分布。以此过程来实现比较 2 组观察对象在结果变量(因变量)上的差异的目的。

倾向指数的应用就类似对观察性研究数据的特征变量进行了类似随机化的均衡处理。Rosenbaum 和 Rubin 认为,倾向指数具有如下的性质:

1) 倾向指数可以均衡暴露组和对照组之间在诸多特征变量上的差异。倾向指数概括了所有观测到的特征变量的值,是一个复合值(coarsest score),而每一个特征变量都是一个精确值(finest score)。经过推导可以证明,一个复合值可以充分平衡暴露组与对照组之间在精确值上被观察到的差异,也即,暴露组和对照组的对象,只要具有相同的倾向指数,那么在观察到的特征变量上具有相同的分布。暴露组和对照组的对象如果在某个特征变量的取值上存在差异,这种差异也是随机差异(chance differences),或机遇造成的差异,而不是系统差异(systematic differences)。

2) 在既定倾向指数的情况下,干预分配和观测到的特征变量有条件地相互独立,即满足强可忽略治疗分配(strongly ignorable treatment assignment)的假定。这一性质使得在控制了倾向指数的条件下,每一个观察对象具有相同的概率被分配到暴露组,因此和随机对照试验的效果一样。

3) 对于具有相同倾向指数的所有对象来讲,暴露组和对照组之间的结局变量上的均值差是该倾向指数上的平均干预效应(Average Treatment Effect, ATE)的无偏估计。根据反事实框架理论,研究者不能同时观察到同一研究对象上的两种潜在结局(即暴露状态下和非暴露状态下的结果),所以,必须借助于暴露组和对照组之间的平均值来估计反事实。

然而,倾向指数也有自身的局限。它的应用前提是,与治疗分配相关的所有特征变量都被准确的测量,

(下接第 1277 页)

- Bmj, 2005, 331 (7513) : 376 – 382.
- [8] Linde K, Streng A, Jurgens S, Hoppe A, Brinkhaus B, Witt C, Wagenpfeil S, Pfaffenrath V, Hammes MG, Weidenhammer W, et al: Acupuncture for patients with migraine: a randomized controlled trial [J]. *Jama*, 2005, 293 (17) : 2118 – 2125.
- [9] Witt C, Brinkhaus B, Jena S, Linde K, Streng A, Wagenpfeil S, Hummelsberger J, Walther HU, Melchart D, Willich SN: Acupuncture in patients with osteoarthritis of the knee: a randomised trial [J]. *Lancet*, 2005, 366 (9480) : 136 – 143.
- [10] Haake M, Muller HH, Schade – Brittinger C, Basler HD, Schafer H, Maier C, Endres HG, Trampisch HJ, Molsberger A: German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel – group trial with 3 groups [J]. *Arch Intern Med*, 2007, 167 (17) : 1892 – 1898.
- [11] Scharf HP, Mansmann U, Streitberger K, Witte S, Kramer J, Maier C, Trampisch HJ, Victor N: Acupuncture and knee osteoarthritis: a three-armed randomized trial [J]. *Ann Intern Med*, 2006, 145 (1) : 12 – 20.
- [12] Diener HC, Kronfeld K, Boewing G, Lungenhausen M, Maier C, Molsberger A, Tegenthoff M, Trampisch HJ, Zenz M, Meinert R: Efficacy of acupuncture for the prophylaxis of migraine: a multicentre randomised controlled clinical trial [J]. *Lancet Neurol*, 2006, 5 (4) : 310 – 316.
- [13] Endres HG, Bowing G, Diener HC, Lange S, Maier C, Molsberger A, Zenz M, Vickers AJ, Tegenthoff M: Acupuncture for tension – type headache: a multicentre, sham – controlled, patient – and observer – blinded, randomised trial [J]. *J Headache Pain*, 2007, 8 (5) : 306 – 314.
- [14] Witt CM, Jena S, Selim D, Brinkhaus B, Reinhold T, Wruck K, Liecker B, Linde K, Wegscheider K, Willich SN: Pragmatic Randomized Trial Evaluating the Clinical and Economic Effectiveness of Acupuncture for Chronic Low Back Pain [J]. *Am J Epidemiol*, 2006, 164 (5) : 487 – 496.
- [15] Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN: Acupuncture for patients with chronic neck pain [J]. *Pain*, 2006, 125 (1 – 2) : 98 – 106.
- [16] Jena S, Witt CM, Brinkhaus B, Wegscheider K, Willich SN: Acupuncture in patients with headache [J]. *Cephalalgia*, 2008, 28 (9) : 969 – 979.
- [17] Witt CM, Jena S, Brinkhaus B, Liecker B, Wegscheider K, Willich SN: Acupuncture in patients with osteoarthritis of the knee or hip: a randomized, controlled trial with an additional nonrandomized arm [J]. *Arthritis Rheum*, 2006, 54 (11) : 3485 – 3493.
- [18] MacPherson H, Richmond S, Bland M, Brealey S, Gabe R, Hopton A, Keding A, Lansdown H, Perren S, Sculpher M, Spackman E, Torgerson D, Watt I: Acupuncture and counselling for depression in primary care: a randomised controlled trial [J]. *PLoS Med*, 2013, 10 (9) : e1001518. doi:10.1371/journal.pmed.1001518.
- [19] Daniel L. Riddle. et al. The Pragmatic – Explanatory Continuum Indicator Summary (PRECIS) instrument was useful for refining a randomized trial design: Experiences from an investigative team [J]. *Journal of Clinical Epidemiology*, 2010, 63 (11) : 1271 – 5.
- [20] Harold C. Sox. Initial National Priorities for Comparative Effectiveness Research; Report Brief. 2009; <http://www.iom.edu/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx>.
- [21] Witt CM, Michalsen A, Roll S, Morandi A, Gupta S, Rosenberg M, Kroppass L, Stapelfeldt E, Hissar S, Müller M, Kessler C. Comparative effectiveness of a complex Ayurvedic treatment and conventional standard care in osteoarthritis of the knee – study protocol for a randomized controlled trial [J]. *Trials*, 2013, 23 (14) : 149. doi:10.1186/1745 – 6215 – 14 – 149.

(2014 – 09 – 21 收稿 责任编辑:洪志强)

(上接第 1273 页)

在此前提下,从倾向指数获得的干预效果的估计才是无偏的。而且,暴露组和对照组的组间特征变量要有足够的重叠,即有相当大的共同支持域(common support region)。鉴于特征变量的混杂效应广泛的存在于非随机的观察性研究中,倾向指数的应用需要有大样本量作为前提,从而保证暴露组和对照组之间倾向指数分布的均衡性。样本量越大,特征变量之间的均衡性越好。

总之,在中医药治疗的临床评价研究中,非随机的大样本观察性研究数据对于治疗效果的因果推论仍然具有很大的价值和潜力。倾向指数的应用将会更好的控制可观测的特征变量的差异,获得最佳效应的估计值。

参考文献

- [1] 刘建平. 传统医学证据体的构成及证据分级的建议 [J]. *中国中西医结合杂志*, 2007, 27 (12) : 1061 – 1065.
- [2] Shadish, W. R., Cook, T. D., & Campbell, D. T. . Experimental and quasi – experimental designs for generalized causal inference [M]. Boston: Houghton Mifflin, 2002.
- [3] Neyman, J. . Statistical problems in agricultural experiments [J]. *Supplement to the Journal of the Royal Statistical Society*, 1935, 2 (2) : 107 – 180.
- [4] Rubin, D. B. . Estimating causal effects of treatments in randomized and nonrandomized studies [J]. *Journal of Educational Psychology*, 1974, 66 (5) : 688 – 701.
- [5] 陈蔚文. 中药学 [M]. 2 版. 北京: 人民卫生出版社, 2013.
- [6] 费宇彤, 刘建平. 国际草药 CONSORT 声明及中药临床试验报告规范化问题思考 [J]. *中国中药杂志*, 2008, 33 (1) : 89 – 94.
- [7] 刘建平. 中医药临床试验的方法学问题与挑战: 循证医学的观点 [J]. *中西医结合学报*, 2006, 4 (1) : 1 – 6.
- [8] 沈玉红, 彭真, 张珂良. 中药临床试验现状及对策分析 [J]. *中国新药杂志*, 2013, 22 (20) : 2365 – 2368.
- [9] 范昕, 李康, 马葆华, 等. 中药临床试验中若干医学伦理学问题 [J]. *中医药学报*, 2012, 40 (1) : 7 – 10.
- [10] 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.
- [11] 辛涛, 姜宇. 教育科学的因果推论: 困境与超越 [J]. *清华大学教育研究*, 2013, 34 (3) : 1 – 6.

(2014 – 09 – 21 收稿 责任编辑:洪志强)