

反事实和归因疾病负担研究

陶庄 杨功焕

【关键词】 反事实; 归因疾病负担; 因果效应; 可比较的风险评估; 原因对照

Counterfactual and burden of disease attributable to risk factors TAO Zhuang, YANG Gong-huan. National Center for Public Health Surveillance and Information Services, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Corresponding author: YANG Gong-huan, Email: yangghuan@vip.sina.com

【Key words】 Counterfactual; Burden of disease attributable to risk factors; Causal effect; Comparative risk assessment; Causal contrast

反事实(counterfactual)的本质意思是指在实际生活中,某些情况并未发生,即与“事实”相反。在哲学和统计学界,有很多学者就是借用这个问题来探索事物发生的原因。笔者曾将“counterfactual”译为“虚拟事实”,这是沿袭耿直的译法^[1],但是究其在逻辑学上的起源,会发现一般直译为“反事实”^[2],由于无论统计学还是现在的社会学都奉此起源为正朔,故本文也遵从该译法。

1. 反事实分析方法在流行病学中的应用:流行病学是研究人群中疾病和健康状况的分布及其影响因素,并研究如何预防控制疾病及促进健康的策略和措施的科学。作为探讨致病原因的一门学科,当然应该遵循哲学和统计学上的基本思维方法。将反事实理论和传统流行病学方法相结合的重要学者是美国统计学家 Greenland^[3-5],特别是2002年该学者与 Maldonado 合著的“对因果效应的估计”(estimating the causal effects)一文^[6],成为反事实理论应用在流行病学中最重要的文献之一。在这篇文献中,作者使用了一个图示模型来说明问题(图1)。

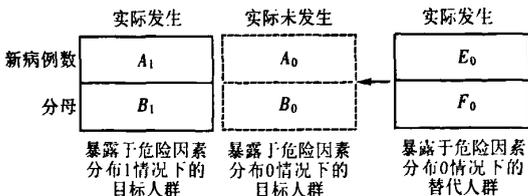


图1 目标人群与替代人群

图1中A和E表示某疾病(如肺癌)的新发例数,B和F是目标人群可能的该病罹患数,而危险因素分布1和0分别表示暴露的两种不同水平,如人群吸烟率为40%定义为分布1,

以吸烟率20%定义为分布0。显然 $R_1 = \frac{A_1}{B_1}$ 、 $R_0 = \frac{A_0}{B_0}$ 和

$R'_0 = \frac{E_0}{F_0}$ 分别为不同人群(或不同暴露状况下)的发病率,

此时可以定义一个 R_1 和 R_0 的原因对照(causal contrast)

$$RR_{原因} = \frac{R_1}{R_0} = \frac{A_1/B_1}{A_0/B_0} \quad (1)$$

如果希望公式(1)有意义,或是说这两个率是可比的,那么最理想的状态当然就是一个人群同时暴露于两个暴露分布水平,换言之, R_1 表示的是某人群吸烟率为40%时的肺癌发病率,而 R_0 是该人群吸烟率为20%时的肺癌发病率,且还应该是在同一时间段上发生的。

很显然,这种理想状态在实际中是不可能的,或者 R_1 不是实际存在,或者 R_0 不是实际存在,甚至于 R_1 和 R_0 两者都不是实际存在。也就是说在实践中,原因对照无法直接测量获得,反事实的情形就出现了,这里必需使用反事实分析的理论。而使用反事实方法的第一步,就是要为不存在的情形找一个替代(substitute),即E和F。

为此,公式(1)相应变为

$$RR_{关联} = \frac{R_1}{\text{对 } R_0 \text{ 的替代}} = \frac{R_1}{E_0/F_0} = \frac{A_1/B_1}{E_0/F_0} \quad (2)$$

注意公式(1)与(2)中RR的下标是不同的,后者使用“关联(association)”替代了前者中的“原因(causal)”。这反映了作者在此处的深刻用意,并暗示着此时的关系——“关联”较“因果”来得弱。

反事实理论模型给病因推断提供了一个通用的框架,不仅可以用于随机试验,同样可以用于观察性研究。但在应用中,使用反事实,必须注意混杂因素(confounder)造成的影响,这主要表现在,所选择的替代人群与目标人群在某些影响因果关系的方面不相同,此时就会发生混杂,如图1中就有

$$\frac{RR_{关联}}{RR_{原因}} = \frac{A_1/B_1}{E_0/F_0} \cdot \frac{A_0/B_0}{A_1/B_1} = \frac{A_0/B_0}{E_0/F_0} \quad (3)$$

显然,只有两个人群完全一样,两个率比才相等。因此,如何选择一个合适的对照组,就成为应用反事实理论的最重要问题。在目前的流行病学实践中,除了经常采用的传统方法(分层、匹配)外,近些年来发展起来了诸如倾向分(propensity score)方法、边际结构模型等,可以部分解决这个问题。其他的学者在近些年也从理论和实践两方面进行了一些研究^[7,8],这些研究促进了反事实分析理论在流行病学领域的发展。

2. 反事实分析方法在疾病负担研究中的应用:作为研究

因果效应的方法论,反事实理论自然可以应用于与因果推断和因果效应评价有关的各个方面,那么归因疾病负担研究当然也不例外。

疾病负担(burden of disease)是指疾病造成的健康、经济、资源的损失,及其产生的生物、心理和对社会的危害,以及对疾病结局如死亡、失能和康复等所带来的后果和影响。疾病负担不仅包括疾病不同转归所带来的负担,也包括疾病对社会带来的负担及资源的消耗^[9]。它一般使用涵盖死亡与患病两方面的综合指标(如 DALY 等)。而归因疾病负担研究是指将总的疾病负担按照不同的部分(不同病因、危险因素)进行分解,以探讨各部分对健康影响的方法。

为了研究归因疾病负担, Murray 和 Lopez^[10]提出了可比较的风险评估(comparative risk assessment),在此框架下,一般使用两种归因方法:分类归因(categorical attribution)和反事实分析(counterfactual analysis)。前者一般是为了研究某个事件如死亡被归因于某个简单的病因(某个疾病或某个危险因素),或者一组病因,这样所有事件都可被归因于不同部分的归因问题,如 ICD 和归因于酗酒的伤害,或职业因素就是其应用的例子。在 1990 年的全球疾病负担研究中使用的就是该方法^[11]。

显然,这里应需要一个较完善的分类体系,这个体系最低需满足所有类的合集是全集及任意类相互间的交集是空集。所以,在对分病因的疾病负担进行研究时,尚可使用,但在研究各种危险因素的归因疾病负担时,就只能使用反事实理论了。

在这个领域,使用反事实理论同样面临着选择什么样的替代(alternative)人群,或称反事实暴露场景(scenarios)的问题。长期以来,占主导地位的反事实暴露场景是整个人群 0 暴露或一个非 0 的定值(如血压,不可能为 0,只能定义为非高血压的定值)。在这种估计基础上的统计量是人群归因分值(PAF),它被定义为:如果暴露水平被减少到 0(或其他定值),疾病或死亡的减少比例为

$$PAF = \frac{P(RR - 1)}{P(RR - 1) + 1} \quad (4)$$

当暴露水平可以被分为许多水平(如 n 个水平)时,并且每一水平都有自己的 RR 时,公式(4)变为公式(5)^[12]

$$PAF = \frac{\sum_{i=1}^n P_i(RR_i - 1)}{\sum_{i=1}^n P_i(RR_i - 1) + 1} \quad (5)$$

在某些情形下使用 0 暴露是一个有用的方法,但对其他一些情形就显得过于苛刻。而危险因素对于疾病或死亡的作用还可应用其他的估计方法,可观察的暴露水平与其他分布(而不是一个定值)的比,这被定义为潜在影响分值(PIF)^[12]

$$PIF = \frac{\int_{x=0}^m RR(x)P(x)dx - \int_{x=0}^m RR(x)P'(x)dx}{\int_{x=0}^m RR(x)P(x)dx} \quad (6)$$

式中, $P(x)$ 为实际人群暴露率,而 $P'(x)$ 就是反事实暴露分布中的暴露率, m 是暴露的最高水平。第一项和第二项分别表

示在现有情况下和反事实情况下的暴露水平。当因子是分类而不是连续的,则公式(6)可为^[12]

$$PIF = \frac{\sum_{i=1}^n P_i RR_i - \sum_{i=1}^n P'_i RR_i}{\sum_{i=1}^n P_i RR_i} \quad (7)$$

进一步,如果把非暴露组作为对照,分析基于一系列分布的多重反事实的多重比较,公式(7)可考虑反事实暴露的 RR ,也就是说这个 RR 还可以依赖于其他因素,如新技术、医疗服务设施等。同样,从事同一种职业,由于具有不同的安全措施,会面临不同的职业危害。因此公式(7)一般的形式^[12]

$$PIF = \frac{\int_{x=0}^m RR(x)P(x)dx - \int_{x=0}^m RR'(x)P'(x)dx}{\int_{x=0}^m RR(x)P(x)dx} \quad (8)$$

可供选择反事实暴露分布的有许多类型, Murray 和 Lopez^[10]介绍了一种对反事实分布的分类方法,为鉴别暴露剂量及如何应用提供了依据。该方法将反事实分布归为 4 类:理论最小风险(theoretical minimum risk)是指来自人群理论上的最低风险,而无论该情况是否真实存在,如前述的 0 暴露;可能真实最小风险(plausible minimum risk)基本上是根据某种理论的设想,但未经统计学的验证;可能合理最小风险(feasible minimum risk)一般是从其他人群或者其他研究获得;费效最小风险(cost-effective minimum risk)考虑的是暴露去除所花费的费用(通过一套已知的费效干预方法)来作为暴露场景的选择方法。

由于实际工作中,各种因素的相互关联性,研究者们还给出了估算多个因素联合的 PIF(或 PAF)^[13]

$$PIF = 1 - \prod (1 - PIF_i) \quad (9)$$

近些年有关疾病负担的研究中,针对归因疾病负担研究所使用的方法也多为反事实理论^[14-16]。

3. 反事实理论在国内的研究现状:在我国只有少数学者对该理论进行了探讨,包括哲学、教育心理学及社会经济学,但这些领域通常使用的是反事实思维(counterfactual thinking)。而在统计学方面,只在研究因果推断时进行了介绍^[17,18]。在公共卫生领域,特别是在我国的归因疾病负担研究中,还只限于总体上的负担研究,归因研究也仅限在分病因等领域,即使用分类归因的方法,还未见使用反事实理论对危险因素进行研究^[19-21]。

参 考 文 献

- [1] Geng Z, Jin H. Counterfactual causal model // Fang JQ, Lu Y. eds. Modern Medical Statistics. Beijing: People's Health Press, 2002:516-518. (in Chinese)
耿直, 金华. 虚拟事实因果模型 // 方积乾, 陆盈, 主编. 现代医学统计学. 北京: 人民卫生出版社, 2002:516-518.
- [2] Goble L. Philosophical logic // Zhang QY, Chen MZ. Beijing: China Renmin University Press, 2008:438-475. (in Chinese)
Goble L. 哲学逻辑. 张清宇, 陈慕泽, 译. 北京: 中国人民大学出版社, 2008:438-475.

[3] Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*, 1986, 15: 413-419.

[4] Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol*, 1987, 125: 761-768.

[5] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci*, 1999, 14: 29-46.

[6] Maldonado G, Greenland S. Estimating the causal effects. *Int J Epidemiol*, 2002, 31: 422-429.

[7] Malcolm M. The case-specular study design and counterfactual. *Epidemiology*, 1998, 9: 6-7.

[8] Blakely T, Woodward A. Counterfactual challenges to social epidemiology. *Australas Epidemiol*, 2000, 7: 3.

[9] Wang BY. The burden of disease research and application. *Chin Endemic J*, 2004, 23(5): 512-513. (in Chinese)
王滨有. 第五讲 疾病负担的研究与应用. *中国地方病学杂志*, 2004, 23(5): 512-513.

[10] Murray CJ, Lopez AD. The global burden of disease. Harvard University and World Health Organization. Boston: Harvard University Press, 1996.

[11] Christopher JL. Summary measures of population health. Geneva, World Health Organization, 2002: 273-313.

[12] Ezzati M. Comparative quantification of health risks. Geneva, World Health Organization, 2004: 1-38.

[13] Lopez AD, Mathers C, Ezzati M, et al. Global burden of disease and risk factors. New York: The World Bank and Oxford University Press, 2006.

[14] Prüss-Ustün A. Introduction and methods: assessing the environmental burden of disease at national and local levels. Geneva, World Health Organization (WHO Environmental Burden of Disease Series, No. 1). 2003.

[15] Norman R, Bradshaw D, Schneider M, et al. Revised burden of disease estimates for the comparative risk factor assessment, South Africa 2000. Methodological Note. Cape Town: South African Medical Research Council, 2006.

[16] Begg S, Vos T, Barker B, et al. The burden of disease and injury in Australia 2003. PHE 82. Canberra: AIHW, 2007.

[17] Hu YH, Geng Z. Discussion of concept of confounding. *Chin J Epidemiol*, 2001, 22(6): 459-461. (in Chinese)
胡永华, 耿直. 关于混杂概念的讨论. *中华流行病学杂志*, 2001, 22(6): 459-461.

[18] Geng Z, Hu YH. Interaction of statistical inference. *Chin J Epidemiol*, 2002, 23(3): 221-224. (in Chinese)
耿直, 胡永华. 交互作用的统计推断. *中华流行病学杂志*, 2002, 23(3): 221-224.

[19] Zheng Y, Li DL, Yu SZ, et al. The burden of disease study of malignant tumors in Shanghai. *Chin Tumor*, 2001, 10(4): 196-198. (in Chinese)
郑莹, 李德录, 俞顺章, 等. 上海市恶性肿瘤疾病负担的研究. *中国肿瘤*, 2001, 10(4): 196-198.

[20] Li D, Xu K. The epidemic characteristics and burden of disease of major chronic diseases in China. *Chin Primary Health Care*, 2003, 17(1): 32-34. (in Chinese)
李栋, 许琨. 我国主要慢性病的流行特点及疾病负担. *中国初级卫生保健*, 2003, 17(1): 32-34.

[21] Zhang AY. Chronic diseases and risk factors for burden of disease analysis-disability-adjusted life years and its significance. *Chin Chronic Dis Prevent Control*, 2006, 14(5): 305-307. (in Chinese)
张安玉. 慢性病与危险因素的疾病负担分析——残疾调整生命年及其意义. *中国慢性病预防与控制*, 2006, 14(5): 305-307.

(收稿日期: 2009-08-20)

(本文编辑: 张林东)

中华流行病学杂志第六届编辑委员会通讯编委名单

- | | | |
|---------------------|-------------------|-----------------------|
| 陈 曦(湖南省疾病预防控制中心) | 秦丰满(成都市疾病预防控制中心) | 高 婷(北京市疾病预防控制中心) |
| 姜宝法(山东大学公共卫生学院) | 李 杰(北京大学医学部) | 李十月(武汉大学公共卫生学院) |
| 李秀央(浙江大学医学院公共卫生学院) | 廖苏苏(中国医学科学院基础医学院) | 林 玫(广西壮族自治区疾病预防控制中心) |
| 林 鹏(广东省疾病预防控制中心) | 刘爱忠(中南大学公共卫生学院) | 刘 刚(四川省疾病预防控制中心) |
| 刘 静(北京安贞医院) | 刘 莉(四川省疾病预防控制中心) | 刘 玮(军事医学科学院微生物流行病研究所) |
| 鲁凤民(北京大学医学部) | 欧剑鸣(福建省疾病预防控制中心) | 彭晓昱(北京市疾病预防控制中心) |
| 邱洪斌(佳木斯大学) | 赛晓勇(解放军总医院) | 苏 虹(安徽医科大学公共卫生学院) |
| 汤 哲(首都医科大学附属宣武医院) | 田庆宝(河北医科大学公共卫生学院) | 王 蓓(东南大学公共卫生学院) |
| 王素萍(山西医科大学公共卫生学院) | 王志萍(山东大学公共卫生学院) | 谢 娟(天津医科大学公共卫生学院) |
| 徐爱强(山东省疾病预防控制中心) | 徐慧芳(广州市疾病预防控制中心) | 严卫丽(新疆医科大学公共卫生学院) |
| 阎丽静(中国乔治中心) | 杨春霞(四川大学华西公共卫生学院) | 余运贤(浙江大学医学院公共卫生学院) |
| 曾哲淳(北京安贞医院) | 张 波(宁夏回族自治区卫生厅) | 张宏伟(第二军医大学) |
| 张茂俊(中国疾病预防控制中心传染病所) | 张卫东(郑州大学公共卫生学院) | 赵亚双(哈尔滨医科大学公共卫生学院) |
| 朱 谦(河南省疾病预防控制中心) | 祖荣强(江苏省疾病预防控制中心) | |