

再论观察与实验:大数据现实世界研究不能取代随机对照试验

唐金陵

999077 中国香港中文大学公共卫生及基层医疗学院

通信作者:唐金陵, Email: jltang@cuhk.edu.hk

DOI: 10.3760/cma.j.issn.0254-6450.2018.08.021

【摘要】 传统上,流行病学多以干预划分观察和实验,干预研究等于实验研究,还认为干预研究的科学性高于观察性研究。在一般科学实验里,干预指人为施加的改变自然状况的措施。干预并不一定是有益的,也并不一定是研究者当下施加的,研究者、受试者或第三者目前或过去施加的措施都可以形成“有效的”干预。例如,由研究者、受试者和第三者通过某种方法致使视神经损伤,都可以形成有效改变视神经正常功能的干预,研究者可以由此观察到视神经和视力的关系。以此推论,由受试者自己过去施加的不良干预(如吸烟)也属于干预,那么研究吸烟和肺癌的观察性队列研究就等同于实验研究了。由此看来,干预本身并不足以有效地区分观察和实验。如果认为实验的科学性高于观察,那么在干预的基础上,只能从科学性上(即设计特征)区分观察和实验。在评估医学干预效果的临床试验中,随机分组是在传统认为的观察研究基础上引入的最重要的偏倚控制措施,应该是区分观察和实验的核心属性。如果一定要把人群研究分成观察和实验,随机对照试验才是真正的实验研究,非随机分组形成的干预研究属于试验,但不是实验。基于大数据的现实世界研究,如果没有随机分组,不能构成实验,也不能成为对干预效果的最终检验。大数据现实世界研究不能取代随机对照试验,这是本文希望传达的最重要的信息。

【关键词】 观察研究; 实验研究; 研究设计; 随机对照试验; 干预研究; 现实世界研究

Revisit to observation and experiment: real world study cannot replace randomized controlled trial

Tang Jinling

School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong SAR 999077, China

Corresponding author: Tang Jinling, Email: jltang@cuhk.edu.hk

【Abstract】 In epidemiology, intervention is normally used to define what experiment is. Intervention studies are equaled to experimental studies. Experimental studies are also considered scientifically more rigorous than observational ones. Intervention is generally referred to human activities that can interfere or change natural conditions. The intervention by definition may not necessarily be beneficial to the study subjects (although exposing harmful interventions to humans are unethical) and activities by the researcher, by the subject himself, or by any third party and either now or in the past can all form “effective” interventions. For example, interventions that can damage the optic nerve by any of the three parties can all help the researcher establish the relation between the optic nerve and vision. In the same sense, an activity that a study subject initiated in the past, such as smoking, would also constitute a valid intervention. As a result, a cohort study on smoking and lung cancer would also be an experiment. From the above arguments, we can see that intervention alone does not suffice to distinguish between experiment and observation. As we equal experiment to higher scientific rigor than observation, only can study designing features of intervention studies be used to define experiment. In intervention trials, randomization is the defining feature that makes randomized controlled trials differ from, and scientifically more rigorous than, controlled observational studies and has been commonly used to define experiment. If we have to divide clinical research into experiment and observation, randomized controlled trials would be experimental and non-randomized studies of intervention are trials but not experiment. Big data, real-world studies are not experiment and cannot replace randomized trials in confirmation of efficacy if comparison groups are not formed by randomization. Real world studies cannot replace randomized controlled trials. This is the most important message this paper wishes to convey.

【Key words】 Observational study; Experimental study; Study design; Randomized controlled trial; Intervention study; Real world study

近年来医学界对基于大数据和现实世界研究的兴趣,引出了一个十分重要并亟待厘清的流行病学理论议题:基于现实世界的干预研究能否取代随机对照试验用来最终确认医学干预的效果?我国对临床应用型研究的重视程度不断加强,吸引了很多研究者进入临床流行病学研究领域,引起了对流行病学一般理论、概念和方法的广泛兴趣和讨论,但同时也暴露出一些混淆和误解,观察和实验就是其中之一。观察和实验的区分事关整个流行病学研究方法的科学原理、逻辑理性和实践规范,厘清二者的关系和异同十分重要。

观察是研究者对自然现象的直接考察和研究,天文学研究就是典型的观察性研究,因为研究者不能根据自己的意愿改变天文现象而揭示新的规律,而实验则是在人为改变了自然现象的前提下进行考察和推论^[1]。实验是观察的延伸,实验为人类的探索和实践活动提供了新的舞台,是人类通过改变自然世界观察到新规律的方法,是人类用以改造自然世界的探索活动。换言之,如果没有实验性研究,人类就不能找到有效的改变自然的方法。医学的最终目的是通过人为的干预措施,改变疾病的自然病程和转归,使它向着更有利于患者健康的方向发展,药物是医学最常见的干预措施。就此意义上讲,评估药物的作用就是典型的实验性研究。为了便于阐述本文的讨论仅限于防治措施的研究,不包括诊断措施。

医学实验研究可分为在实验室里的实验研究和在人群中的实验研究。二者之间的核心区别是观察单位:实验室研究的观察单位是分子、细胞或动物等,而人群研究的最小观察单位是个体的人。实验室研究关注的是事物的生物学本质和原理,而人群中的实验研究关注的则是效果、效用和效益。实验室研究可以解释现象,也可能引发出崭新的干预方法,但不能直接用于医学实践;人群中实验研究的结果多数是终结性的,多数不产生新理论和新方法,但多数研究结果可直接用于医学实践活动。因此,在人群中进行的针对医学干预措施的实验性研究,对医学实践的进步起着举足轻重的作用。

众所周知,由于伦理学的限制,不能给人施加有害的干预措施,因此在人群中研究病因和危险因素,只能采取观察的方法,即研究者在“自然”条件下进行的比较和考察。医学干预措施是潜在有益的,研究者可以通过给研究对象施加干预并继而观察干预的作用,属于实验性研究。干预是流行病学传统上用来区分观察和实验的标志,探索病因的队列研究

是观察性研究的代表,评估干预效果的随机对照试验是典型的实验研究。

流行病学传统上以干预区分观察和实验,在人群中评估医学干预措施效果的研究属于实验研究(experimental study),而非干预性研究则属于观察性研究。进而把评估预防性干预措施的研究叫做现场试验(field trial),把评估治疗性干预措施的研究叫做临床试验(clinical trial),以区别预防性和治疗性干预研究。公共卫生干预通常不是针对个体而是针对群体的,如社区、工厂、学校等,而且多不能采取随机分组,因此又常把针对群体的干预研究称为社区试验(community trial)。社区试验本质上是以群组为单位的整群试验(cluster trial)。

鉴于实验研究在医学研究中的重要作用,有必要充分厘清实验和观察的区别。鉴于干预在区分实验和观察中的重要作用,推敲干预到底是什么就是一个避不开的问题。什么是干预?假如研究者把受试者视神经切断,发现会引起失明,研究者就此可以推断视神经和视觉有关,该研究属于实验研究(研究一)。在这个研究里,“研究者切断受试者视神经”称为干预,干预是由研究者施加的(虽然现实人群研究中不可能发生),但是干预的本质是造成视神经损伤的行为。那么,如果一个人的视神经因外伤被切断,没有研究者的参与,也形成了“视神经损伤”的干预,研究者同样也可以观察到失明,可靠地推断二者的关系(研究二)。这个由非研究者施加的干预也改变了自然条件(损伤视神经),具有与研究者施加的干预同等科学推理的功效;如果干预等于实验研究,就没有理由不把它也叫做干预,因此也没有理由不把这类干预研究叫做实验研究。同理,一个人自己切断了自己的视神经(研究三),算不算干预?在研究二和研究三里,研究者关注的自然条件(视神经)因非研究者的干预而发生了改变,研究者还是能够可靠地观察到视神经和视觉的关系,似乎没有理由不把第三者或受试者自己施加的干预也叫做干预,因此也没有理由不把这类干预研究也叫做实验研究。

以上分析可以发现,干预的本质是人为地改变(损伤)自然条件(视神经),但是,这个改变完全可以因非研究者的行为而发生。在上述第三项研究里,受试者自身施加的干预也属于干预,那么自己选择吸烟,也就是干预,研究吸烟和肺癌关系的队列研究也就属于干预研究,干预研究就是实验研究。因此,一个典型的观察性队列研究就成了干预性实验研究。如果吸烟和肺癌关系的研究不算实验研究,那

么这个研究和前面3个研究有什么本质区别以至于它不再是实验研究?如果吸烟和肺癌的关系与视神经损伤和失明的关系一样,发生的特异、快速、明确,二者间的因果关系就很容易确立,我们似乎没有理由不把吸烟和肺癌的研究也算做实验。如果算做实验,那么观察性的队列研究和实验性的随机对照试验的边界就开始消失了。另外,如果用干预划分实验与观察,那么评估干预副作用的研究也是关于干预的研究,也应该叫做实验研究,但是很多时候使用的是队列研究和病例对照研究,却把它们叫做观察性研究而不是实验研究。

由此可见,干预本身不能把流行病学观察性研究和实验性研究有效区分开来。而且,流行病学在比较观察与实验时,无论是否有明确表述,都赋予了实验更高的科学价值。这个科学价值至少包括两个层面,一是只有通过干预才能获得的科学推论,二是这个科学推论真实性的高低。在探索潜隐期很短的强因果关系时,如视神经损伤和失明的关系,又如从高处坠落和死亡的关系,干预研究本身就足以同时实现这两个价值,即可靠地证明二者间的因果关系。但是,在探索弱因果关系时,尤其当潜隐期很长的时候,如降压药预防心血管病的效果,有关推论的真实性不是干预研究本身就可以保证的,还需要通过控制施加干预的条件和方法才能实现,例如保证比较组之间始终具有可比性,以便控制混杂揭示因果本质。区分干预这两个层面的重要性,在实验室研究中似乎不是问题,而在人群流行病学研究中就显得十分必要。如果不加以区分,就不足以有效的区分流行病学的观察和实验。

如果把干预进一步分解成干预措施本身以及施加干预措施的方法,那么实验和观察均可以用来评估干预措施的作用,实验区别于观察的核心在于后者,即控制比较的条件和方法,如分配干预的方法,只有使用特殊的能让比较组间所有影响因素可比的分配方法,才能实现实验要求的基本条件。随机对照试验与队列研究(或非随机分组形成的对照试验)的本质区别是随机分组,不是干预;干预使随机分组成为可能,但干预本身还不足以形成实验研究更高的科学性。另外,“randomized controlled trial”里“control”的意思是控制和操纵,而不是对照,如果强调的是对照,完全可以用“compared”或“contrasted”。所谓控制,强调的是对施加干预方法以及试验条件的操纵。因此,随机对照试验就是“用随机分组控制下的(对照)测试”。

因此,目前国际流行病学界普遍认为,在人群研究中测试干预效果的研究是试验(trial),而不是实验(experiment),只有随机对照试验才是真正的实验^[2-6]。国际流行病学协会赞助编写的《流行病学词典》认为,在现代流行病学里,实验就等同于随机对照试验。非随机分组形成的干预研究,不是真正的实验研究,因为存在自然和人为引入的混杂偏倚,其科学性与队列研究没有本质的区别。有些干预研究也可以叫做类实验,如按照患者出生日期或病例档案号码进行分组的试验,其科学性介于队列研究和随机对照试验之间。的确,由于伦理学的限制,人群中的随机对照试验只能用于对干预的研究,但反过来说所有干预研究都是实验研究是不妥当的。

流行病学研究的分类应该有助于选择研究设计类型及判断研究结果真实性。我们的确也是按照科学性的高低来区分病例对照研究、队列研究和随机对照试验所提供的证据的。然而,传统的有关实验和观察研究区别的认识,以及对实验研究的分类并不能有效达到这些目的。而且,对实验研究分类没有采用统一的分类属性,造成了进一步的混乱。例如,对临床试验、现场试验和社区试验的划分,同时使用了干预地点和干预特征两个属性,干预特征又涉及“个体-群体”以及“预防-治疗”两个方面。临床试验是临床上针对个体的治疗性研究,而社区试验则是在临床以外对群体的预防性措施的研究。但是,如果用“地点”、“治疗或预防”以及“个体或群体”3个属性进行分类,共有8种组合,即8种实验研究的类型。目前的3类实验类型显然未包括所有可能的实验研究类型。虽然不存在个体干预的整群研究,但目前的分类里显然排除了实际存在的“临床上的个体预防措施”以及“现场里的个体治疗措施”等研究类型。而且,英文的 trial(试验)不等于 experiment(实验),试验是试一试或测试的意思,而实验则是对应于观察的研究模式。干预重点说的是研究问题,不是研究方法,与其对应的问题应该是探索病因、评估诊断、估计疾病患病率等。研究问题和研究方法隶属不同的属性,不应混淆,更不可等同。

由此看来,在评估治疗干预的问题上,如果一定要把研究划分为观察和实验,就需要借助研究类型的科学性这个分类属性。然而,不同流行病学的研究设计类型的科学性不是二分的,而是逐渐变化的。也许,把研究分为实验和观察的做法太简单化了,采取更细的划分方式才能更好地体现不同研究科学性的高低,临床流行病学和循证医学正是这么

