

· 回顾与思考 ·

流行病学研究中的偏倚与混杂

任涛 詹思延 沈霞 孟凡亚 胡永华 李立明

对偏倚与混杂的认识和定义虽然在不断深入和改变,但流行病学家们从未忽略控制偏倚和混杂的重要性。历史上,对于偏倚的定义已经包含三个主要意思:①观察者的偏见;②器械错误造成的偏差;③错误的研究设计造成的结果。20世纪70年代出现的偏倚类型学对于定义在流行病学研究设计、实施和分析过程中存在的偏离真实值的情况有很大的帮助^[1,2]。随后,偏倚逐渐形成了更加正式和系统的定义。现代流行病学把偏倚归纳为选择偏倚、信息偏倚和混杂偏倚。混杂可以看作是偏倚的一个特殊类型,在简明牛津英文词典中提到它是一个中世纪的拉丁文词“con-fundere”,意思是混在一起,源自中世纪法语“confondre”。

一、早期的研究

1. 偏倚的历史:关于偏倚和不精确的区分首先在物理学中进行了经典的阐述,这是发生在法国大革命时期关于测量的重要讨论,焦点是关于“米”的定义。最初,米是作为最绝对和“直接的”测量单位(地球经线的 40×10^{-6})。但是,很快人们就发现,在实际工作中“米”很难准确测量,后来还发现包括被收藏在国际档案室“米”的白金标准也有偏差。另外,用这种标准测量的“米”在不同的比较中也很不精确。一般而言,绝对的测量不是很精确但是更加真实,只是得到绝对测量的结果比较困难。与此相反,相对的测量结果会更加精确,但是当它们涉及到与自然标准进行比较时可能会有偏差。

在医学调查中认识到偏倚已经有很长时间了。1843年Guy在研究职业与健康关系的时候,在排字工人中发现肺结核与其他疾病的比是1:3.47,而在印刷工人中为1:5.12,由此提出由于排字工人缺乏体育锻炼而导致肺结核的高风险假设。Guy的假设是,这种风险的差异可以通过工人的自身选择来解释——那些自身体弱或多病的人可能会选择久坐的工作,这就是最初的选择偏倚,但Guy随后走访了503名工人,发现只有11名工人是由于身体状况的影响而对工作进行了选择,于是选择偏倚被排除了。

流行病学中研究设计的概念被提炼出来以后产生了很多关于偏倚的重要讨论。20世纪50年代,随机临床试验(randomized clinical trial, RCT)被用于实践,在医学研究中扮演着“金标准”的角色。同时,也导致了偏倚可以用随机化的方法避免的想法。但是,Cornfield在1954年就指出,RCT只是各种推理工具中的一种。在70年代,对于不同研究设计的优缺点开始了激烈的讨论,尤其是对于病例对照研究,主

要争论集中在病例和对照的选择(特别是医院为基础的研究),如病例的纳入和排除标准,匹配的精确性等。在这一时期,除了一般的关于偏倚来源的讨论外,1976年Murphy和1979年Sackett提出了偏倚的分类学。Murphy提出,偏倚可以出现在以下几种情况:①没有控制的研究;②测量方法的不同;③在处理自发疾病时的对照研究;④不同时间的比较;⑤估计的偏倚;⑥分析假设中的偏倚;⑦假设检验中的偏倚;⑧报告中的偏倚。

Sackett等^[3,4]列举了35种偏倚可以简化为6种:①阅读相关领域资料时的偏倚;②在分类和选择研究样本时的偏倚;③实施试验过程中的偏倚;④在测量暴露和结局时的偏倚;⑤在分析数据时的偏倚;⑥在解释分析结果时的偏倚。在这35种偏倚中,其中有9种被Sackett详细的描述过,这些偏倚至今还为人们所熟悉。包括:现患病例-新发病例偏倚、入院率偏倚、检出征候偏倚、无应答偏倚、选择偏倚、诊断怀疑偏倚、暴露怀疑偏倚、回忆偏倚、家庭信息偏倚。Sackett试图证明这些偏倚的存在与错误估计相对危险度是有联系的。

2. 混杂的历史:混杂的历史在某种意义上可以反映研究设计的历史。可以认为,混杂不是一个统计的或分析的概念,对于混杂的认识,反映了科学推理的逻辑性。在旧约全书中提到,Daniel用一个对照组证实规律饮食的作用来反对巴比伦的国王,这经常被释为“第一个临床试验”。这个例子在1982年被Lilienfeld^[5]引用,作为临床试验研究历史的一页,标题为“其余情况相同”。圣经上的Daniel理解的“混杂”和我们理解的混杂的意义差不多,这确实让我们可能把混杂的历史延展得很远。然而,多数关于混杂的研究还是集中于近几十年。

研究设计的关键在于比较,任何观察性研究的要点也在于比较。比较可以是一个真实的有两组或两组以上之间的比较,也可以是与想象的比较(与我们的预期相比较)。1966年Bernard曾经在医学研究的方法学方面清楚的阐述说,实验性研究和观察性研究有一个共同点,这个共同点就是与对照进行比较。在一个实验性研究中,设计者自己制造对照,但是在观察性研究中,设计者要寻找对照。

Mill关于因果关系的准则中,有关在实验性研究和观察性研究中如何进行比较的内容,在1970年被MacMahon, Pugh^[6]以及1973年Susser引入到流行病学著作中。其中最著名的引文如下:“如果在观察性研究中一个事件发生了,或者没有发生,那在这两种情况下除了一种现象以外,其他现象都相同,那么这种现象就是或可能是这件事例发生的效果或是原因。”这种“求异法”的观点给医学和流行病学很多启

示。当然, Mill 还描述了其他的方法,如“求同法”、“共变法”等。无论是从引用 Mill 的文字,或是从“其余情况都相同”的原则出发,我们让观察性研究尽量等同于实验性研究。一旦“其余情况都相同”的条件不能得到满足,比较就会出现错误,错误的信息导致混乱。这种错误的信息就是混杂,这就是混杂的本质。

Fisher 可能是最早开始将混杂用于研究设计中的人。他在 1937 年《实验设计》一书中用很长的篇幅提到混杂,在他看来,混杂不是什么“洪水猛兽”,是可以避免的,混杂虽然会给研究设计带来麻烦,但是可以将其向好的方面转化。提到混杂,我们还要提到 Kish^[7],他曾经致力于社会科学的方法学理论的研究,他的关于“一些研究设计的统计问题”仍然值得我们阅读,并且是当代参考书的主要的来源之一,当时确实是混杂开始的时期。关于研究设计,他提出四种变量:①说明性变量,或者“实验性”变量,包括独立的和非独立的;②受控的外来性变量(在选择和估计的时候);③未受控的外来性变量(经常与第一种变量混淆);④未受控的外来变量,但是是随机化的或可以按照随机化的方式进行处理。

Kish^[7]对于混杂这个词的理解来源于 Fisher,混淆于测量和选择两种效果内。但是,区别在于 Kish 认为这些不同的变量会影响研究的结局。尽管还不是那么明确,但 Kish 好像已经将混杂同其他类型的偏倚区分开了。他所提出的第二种类型的受控变量就可以进行测量和选择。当今,混杂这个词确实用于两种结果混淆的一种特殊类型,混淆来自于外部原因,即其他因素确实影响了疾病的发病率,比如,年龄、性别、习惯或者居住环境。混杂并不是由于测量或者选择的不同带来的问题,现在我们称其为“信息偏倚”和“选择偏倚”。

二、现代的研究

1. 现代偏倚的正式定义: Miettinen, Cook^[8]认为,偏倚是实验性研究与其他流行病学研究相比真实性的差别。“一般来说,一个非实验性研究设计成功的关键是要仿效实验性研究。”作为金标准的随机临床试验,有三大特点:①安慰剂的使用,即效应的可比性;②随机化的应用,即人群的可比性;③盲法的应用,即信息的可比性。由此,1985 年 Henc 和 Miettinen 提出将偏倚分为比较偏倚、选择偏倚和信息偏倚。这种分类法也出现在其他流行病学的教科书中,比如 Rothman(1986),Henneken 以及 Buring(1987)的课本中。Rothman, Greenland 认为^[9,10]选择偏倚是效应测量的一种失真情况,结果导致选择的人群样本与全人群(或目标人群)的效应估计之间存在差异。在研究健康结局时,选择偏倚是指选择的暴露与非暴露人群中的差异;在研究暴露因素时,是指选择有病患者与健康人群之间的差异。与此类似的,信息偏倚也可以概括为一种分类错误(暴露或结局)无法独立于另一种分类的情况。那么,所有类型的偏倚实际上就是在研究设计方面(人群的选择,信息的收集)无法独立于原先的假设,从而使逻辑上的关联替代了真实的关联。

2. 现代流行病学关于混杂概念的讨论:谈到混杂概念

在历史上的两次飞跃都与 Jerome Cornfield 这个名字有关。第一次是 1959 年发表的一篇重要的论文^[11],在这篇文章中他和 Haenszel, Hammond, Lilienfeld, Shimkin 以及 Wynde 一起讨论是不是吸烟的人比正常人患肺癌的风险高 10 倍,还是由于混入了其他因素所引起的。这篇文章关于吸烟会导致肺癌的基本观念遭到了反对,在反对者中就包括 Fisher。Fisher 认为是某种潜在因素,即能够导致肺癌又使之有吸烟的倾向。因而,吸烟与肺癌之间不是因果联系,仅仅是这个潜在因素导致了两者之间的联系。Cornfield 和他的同事是支持吸烟与肺癌之间存在因果关系的,但是在他们的论文中并没有用混杂这个词,而是用了“病因性物质”和“非病因性物质”,现在我们称之为“混杂因子”和“混淆变量”;问题的关键是,要想否认吸烟是肺癌的一个危险因素,就必须证明一些因素与吸烟有关,比如说,喝咖啡等是肺癌的真正原因。而且,它要比吸烟与肺癌之间的联系更紧密,否则就永远没有充足的理由推翻以前的论证。我们需要找出潜在的混杂,不能让他们成为不能控制的东西而破坏了研究。

Cornfield 的下一个贡献甚至更微妙,那是关于 1971 年的 UGDF(University Group Diabetes study)研究结果的讨论。在此,Cornfield 阐述了在随机化分组以后混杂仍然存在。在 UGDF 的研究中有这样的情况,研究发现口服降糖药的人预后较差,比用胰岛素或者依靠饮食控制的人更多地死于心脏疾病。很快有人指出该研究随机化较差,使口服降糖药组预后较差的原因是该组人群中更多地使用了洋地黄等药物,而且他们年龄偏大,男性较多,肥胖者也较多。

Cornfield 接受了这个挑战,在针对这个问题上,他又没有用混杂这个词,而用了“随机且基线的不均衡性无显著性差异”。他构建了一个多变量预测模型,使用了一个指示变量,并且将其用于不同的治疗组。然后他做了两件事。首先,他演示了在每一个治疗组如何从基线资料估计预后,存在哪些差异,但是这些差异远不及结局中真正的差别;然后,他把所有组按危险因素进行分层,并且再次表明了这样的分层对不同治疗组结局的差异影响甚微。他预示了 Miettinen^[12]提出的多因素混杂因子。Cornfield 的成就在于他说明了即使是在随机化以后,混杂仍需要进行判断。

三、一些历史事例和争论

1. Berkson 偏倚:1946 年 Berkson^[13]对医院为基础的流行病学研究的正确性提出质疑。Berkson 偏倚是指住院患者作为研究对象时,与仅利用医院服务的人相比较,存在的一种固定偏差,产生这种现象的原因主要是具有多种疾病的患者,入院的概率高于独立患有某种单一疾病的人。Berkson 偏倚在流行病学上一直有争论,直到 1978 年他的现实性被 Roberts 等论证。他们重新分析了若干家庭调查研究,获取卫生服务利用信息,信息来自 8 个临床治疗和 6 个药物治疗研究的住院和非住院病例。有联系的所有对子都被分成两组,进行相对危险度显著性差异的统计学检验。通过比较发现,社区为基础的研究与医院为基础的研究相比,关联有所

不同,这种改变很有可能被一些现象所干扰,比如糖尿病、瘫痪、中风、先天性传到阻滞、高血压和胃溃疡等。

2. 检出偏倚:最初关于良性乳腺疾病与乳腺癌的关系的研究表明,良性乳腺疾病患者倾向于患乳腺癌。后来研究表明这可能与检出偏倚有关,即患良性乳腺疾病的人会比一般人做更多的检查,包括连续的胸片,这样就更容易诊断出乳腺癌来。1986 年 Silber 和 Horwitz 在一个病例对照研究中论证了这个偏倚。研究显示,早期乳腺疾病与乳腺癌的粗 OR 值为 2.6,统计上差异有显著性。但是,如果考虑了研究人群接受检查的不均衡性,这种关联就消失了。检出偏倚在病例对照研究中是很常见的问题,相关的讨论都集中在以下两方面:①服用利血平和患乳腺癌的关系;②妇女服用类固醇的危险性。在 1979 年的 Journal of Chronic Disease 上,有对于这两方面病例对照研究方法的彻底讨论,Feinstein 对于检出偏倚进行了方法学上的评价,Sackett 对偏倚的概括也发表在该杂志上^[3]。

3. 健康工人效应:健康工人效应曾被认为是劳动力选择的结果,或者说是雇工自我选择以及雇主选择的结果。基于此种原因,在早期的文章中健康工人效应被认为是选择偏倚。但是,Monson 认为他不是一个选择偏倚,只有当有病的人选择性的进入研究队列才会发生,而健康工人效应是一种特殊类型的混杂,良好的健康状态决定了结局并且与暴露有关(从事相应的工作)。Arrighi, Hertz-Picciotto^[14]阐明了健康工人效应包含了两个过程:①最初的选择过程,身体健康的工人更倾向于选择或者获得特定的工作;②持续性的选择过程,维持工作的人健康状况要比那些放弃工作的人好。

4. 幸存者治疗偏倚:幸存者治疗偏倚是一个严重的问题,曾经在艾滋病的研究中进行过描述。Berkson 偏倚理论出现很多年以后才被经典论证,与之相比,幸存者治疗偏倚直到最近才被发现。Glesby 和 Hoover 在 1996 年的观察性研究中,寿命长的患者将有更多的机会选择治疗,但是那些寿命短的患者就有可能没有接受治疗。这种偏倚的结果就有可能导致错误的结论,即低水平的治疗可以延长寿命。

四、结语

随着流行病学的发展,关于偏倚和混杂的文章不断增多,在 MEDLINE 数据库中以 bias 为主题词进行检索,从 1970~1989 年共 305 篇,1990~1992 年为 2636 篇,而 1993~2003 年达 20904 篇。用 MEDLINE 以 confounding 为主题词进行检索,从 1970~1989 年共 573 篇,其中 70 年代仅 36 篇,80 年代为 537 篇,从 1990 年开始到 2004 年 4 月为 7559 篇。人们对偏倚和混杂有了更加深刻的认识,逐渐明确了偏倚的定义,提出了一些偏倚分类的方法,寻求出偏倚的来源并设法加以控制。在我国,以“偏性或偏倚”为主题词,查找 CMCC 数据库,截止到 2003 年底,共检索到文献 213 篇,其中最早的检索文献为 1980 年全国流行病学学术会议资料汇编,苏德隆教授所著《进入率与样本的偏性》。以混杂为主题词,截止到 2003 年底,共检索到中文文献 89 篇,其中最早的是发表在《中国慢性病预防

与控制杂志》赵宁教授所写“肝癌危险因素的 Meta-analysis:16 个病例对照研究资料的综合分析”。1986 年 5 月,由钱宇平教授主编的《流行病学》(第 2 版)中,对于病例对照研究以及队列研究专门提出了可能在研究中出现的偏倚以及如何进行控制。1991 年 4 月,王天根教授主编的《流行病学研究方法》专门用一个章节介绍流行病学研究中偏倚和交互作用,并将偏倚归类为选择偏倚、信息偏倚和混杂偏倚,同时对于混杂的概念、大小、方向以及控制进行了细致的探讨,其中一些经典的例子至今仍被不断引用。

然而,迄今为止,虽然流行病学家和统计学家曾对混杂进行过很多探讨,也提出过一些混杂判别的方法以及理论,但一直没有一个公认的、确切的混杂的定义。Boivin 和 Wacholder 提出的可压缩准则以及 Miettinen 和 Cook 提出的可比较准则被一些学者认同,同时也被一些学者认为仅仅是混杂的必要条件,不能作为混杂因素的定义。在我国,北京大学、复旦大学和中山大学合作的国家自然科学基金重点项目“流行病学与统计学相结合进行因果探讨的理论方法与研究”着重对混杂进行了探讨,并提出了基于虚拟事实模型的混杂的形式化定义^[15]。该研究在探讨混杂的概念和控制方面和世界处于同步水平。明确研究中的偏倚和混杂并加以控制,将有利于流行病学的研究设计、分析和结论,获得更加真实准确的危险因素与疾病之间的关联信息,从而更好的为人类健康服务。

参 考 文 献

1. Bailer JC III. Mammographic screening: a reappraisal of benefits and risks. Clin Obstet Gynecol, 1978, 21:1-14.
2. Miettinen OS. Matching and design efficiency in retrospective studies. Am J Epidemiol, 1970, 91:111-118.
3. Sackett DL. Bias in analytic research. J Chron Dis, 1979, 32:51-63.
4. Sackett DL, Richardson WS, Rosenberg W, et al. Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone, 1997. 148-149.
5. Lilienfeld AM. Ceteris paribus: the evolution of the clinical trial. Bull Hist Med, 1982, 56:1-18.
6. MacMahon B, Pugh TF. Epidemiology, principles and methods. Boston: Little, Brown and Co, 1970.
7. Kish L. Some statistical problems in research design. Am Sociol Rev, 1959, 26:328-238.
8. Miettinen OS, Cook EF. Confounding, essence and detection. Am J Epidemiol, 1981, 114:593-603.
9. Rothman KJ. Modern epidemiology. Boston, Mass: Little, Brown and Co, 1986.
10. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, 1998.
11. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst, 1959, 22:719-748.
12. Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol, 1976, 104:609-620.
13. Berkson J. Limitations of the application of 4-fold tables to hospital data. Biomet Bull, 1946, 2:47-53.
14. Arrighi M, Hertz-Picciotto I. Definitions, sources, magnitude, effect modifiers and strategies of reduction of the Healthy Worker Effect. JOM, 1993, 5:890-891.
15. 胡永华, 耿直. 关于混杂概念的讨论. 中华流行病学杂志, 2001, 22:459-461.

(收稿日期 2004-08-03)

(本文编辑:尹廉)