

关于混杂概念的讨论

胡永华 耿直

混杂(confounding)是由混杂因子(confounding factor 或 confounder)所导致的一种偏倚。在流行病学研究中如何避免与控制混杂因子的干扰始终是人们所关注的问题。流行病学家与统计学家曾对混杂现象做过很多探讨 ,提出过一些关于混杂的判别方法与控制理论 ,但至今在国内外流行病学教科书中 ,我们还未见到关于确切的、公认的混杂定义。本文将近年来他人及我们对混杂概念的探讨结果提供给读者 ,期待能有更多学者参与此问题的讨论。

总结既往流行病学家对混杂判定准则 ,大致可以归为两种 :一种是由 Boivin 和 Wacholder 提出的“ 混杂因子的可压缩准则 ”;另一种是由 Miettinen 和 Cook 提出的“ 混杂因子可比较准则 ”。这两种准则曾分别被不同的流行病学家接受或反对 ,也有人认为这两种准则是互为补充的 ,只要遵循这两个准则就可以识别与排除研究中的混杂。

一、可压缩准则(collapsibility-based)^[1-5]

可压缩的定义是 :某外部因素在各水平上的关联测度与边缘上的一致。如不一致则该外部因素为混杂因子。假定 E 为某暴露因素 ,D 为某疾病 ,C 为外部变量。C 可为二分变量也可以为多值变量 ,如 :C 为男、女 ;或为年龄。以 C 分层可有 $C_1, C_2, \dots, C_i, \dots, C_n$ 。在不作分层时 ,我们所得到的危险度测量值为 cRR 。如果进行分层分析则可以得到各层的危险度的测量值 $RR_1, RR_2, \dots, RR_i, \dots, RR_n$ 。可压缩准则是指 :如果 $RR_i = cRR$,表明 C 是可以被压缩的 ,C 不是混杂因子。如果 $RR_i \neq cRR$,表明 C 是不可以被压缩 ,C 是混杂因子。

假定在一项研究中 ,性别为外部变量 ,如果男性、女性与不分层时的危险度都相同 ,那么性别在此研究中不是混杂因子。对于年龄、文化程度等连续或等级变量 ,同样可依据相同的方法判断其是否是混杂因子。事实上这种关于混杂的可压缩准则我们经常在应用。但是 ,这种准则作为混杂的判定标准是否完全真实可靠呢 ? 我们可预先假定有一个外部变量 C ,且 C 不是混杂因子。按照可压缩准则 C 在研究与分析中应该是可以被压缩的。如果 C 不能被压缩 ,那么我们就可以否定可压缩准则的普遍意义。在表 1 的例子中假定无因果作用 ,因此 RR 应为 1 ,但 $RR_i \neq cRR$ 。虽然 C 的分布完全是我们假设的 ,但同样可以说明可压缩准则的缺陷。即 :有些因子虽然不可压缩 ,但仍不是混杂因子。按否定论的观

点 ,一些学者认为可压缩准则是错误的。

表1 C 不能被压缩举例

	C ₁		C ₂		合计(C ₁ + C ₂)	
	暴露	非暴露	暴露	非暴露	暴露	非暴露
有病	15	5	40	50	55	55
无病	85	95	60	50	145	145
合计	100	100	100	100	200	200
RR	RR ₁ = 3.00		RR ₂ = 0.80		cRR = 1.00	

二、可比较准则(comparability-based)^[6]

Miettinen 和 Cook 通过许多例子 ,归纳得到判断外部因子是否是混杂因子的准则 :①该外部因素应是一个独立的危险因素 ,并且 ②在非暴露总体的分布与在暴露总体的分布不同。

到目前为止 ,混杂因子的可比较准则被广泛认同 ,只是在其基础上增加了一条“ 该因子不是病因链的中间环节 ”。某个外部因子是否是“ 中间环节 ”变量可以说主要是依据专业知识 ,在此我们暂且不讨论。我们只需分析一个问题 ,符合可比较准则的因子是否一定是混杂因子 ? 若有例外 ,则只能认为这个准则也不充分。同样我们可预先假定有一个外部变量 C ,无因果作用 , $RR = 1$,且 C 不是混杂因子。按照可比较准则 C 一定是与疾病无关 ,C 在非暴露组与暴露组的分布相同(表 2)。根据我们构建的表 2 , C_1, C_2, C_3 疾病的分布不同 ,说明 C 与疾病有关 , C_1, C_2, C_3 暴露情况也不同 ,说明 C 与暴露也有关。按照可比较准则 C 应该是混杂因子。

还可以构建许多符合比较准则但不是混杂的例子 ,但有一个就够了 ,它足以说明这个准则是有缺陷的 ,即 :符合准则的不都是混杂因子。

表2 C 不可比较举例

	C ₁		C ₂		C ₃		合计(C ₁ + C ₂ + C ₃)	
	暴露	非暴露	暴露	非暴露	暴露	非暴露	暴露	非暴露
有病	95	95	5	5	75	25	175	125
无病	5	5	95	95	75	25	175	125
合计	100	100	100	100	150	50	350	250
RR	RR ₁ = 1.00		RR ₂ = 1.00		RR ₃ = 1.00		cRR = 1.00	

C 与暴露的关系			非暴露组疾病与 C 的关系		
暴露			非暴露		
C ₁	100	100	C ₁		
C ₂	100	100	有病	95	5
C ₃	150	50	无病	5	95

三、可压缩准则与可比较准则的关系^[4,7,8]

我们细致分析一下表 2 ,可以发现 $RR_1 = RR_2 = RR_3 =$

基金项目 :国家自然科学基金资助项目(39930160)和国家杰出青年基金项目(19825103)

作者单位 :100083 北京大学公共卫生学院流行病与卫生统计学系(胡永华) ,北京大学数学科学学院概率统计学系(耿直)

$cRR = 1$ 。按照可压缩准则, C 是可以被压缩的, 所以它不是混杂因子。可压缩准则是对可比较准则的一种补充。我们也可以从表 1 发现, C_1, C_2 均有 100 人暴露, 100 人不暴露, 这表明 C 与暴露因素是无关的。 C 不符合可比较准则, 所以它不应该是混杂。这样看来, 可比较准则也是可压缩准则的一种补充。应该说这两种准则都是正确的, 它们都可以作为判定混杂的必要条件, 二者互为补充。在实际工作中, 常常可以发现这种情况, 流行病学家通常在考虑问题时, 接受“可比较准则”的准则, 而在分析问题时采用“可压缩准则”的准则。但是, 这两种准则又都不是混杂的充分条件, 任意一种准则都不能定义所有混杂因子。那么同时应用这两种准则是否可以定义所有的混杂因子呢? 假定无因果作用, RR 应为 1。见表 3。很明显, 表 3 中 $RR_1 \neq RR_2 \neq cRR$, C 不能被压缩, 应该判为混杂因子。 C_1, C_2 在暴露上的分布不同(与暴露有关), C_1, C_2 疾病的分布也不同(与疾病有关), 按可比较准则也应该判作混杂。此例说明, 即使同时使用这两种准则, 也不能定义所有的混杂因子。

以上所有讨论均是依据给定的实际例子, 只是用了相对风险这个测量指标的实例(没有给出率差的实例), 只是用验证的方法论证的。采用因果的虚拟事实理论, 我们可以用严密的数学形式对以上问题作进一步的证明。

表 3 C 不可压缩也不可比较举例

	C_1		C_2		合计($C_1 + C_2$)	
	暴露	非暴露	暴露	非暴露	暴露	非暴露
有病	60	70	40	180	100	250
无病	40	30	60	220	100	250
合计	100	100	100	400	200	500
RR	$RR_1 = 0.68$		$RR_2 = 0.89$		$cRR = 1.00$	

C 与暴露有关		非暴露组疾病与 C 有关	
暴露	非暴露	C_1	C_2
C_1	100	有病	70
C_2	100	无病	30

四、用虚拟事实模型定义混杂^[9, 10]

Rubin(1974)提出了一种因果作用模型, 此模型与Lewin(1973)的虚拟事实(counterfactuals)的哲学理论相似, 因此称为虚拟事实模型。该模型的最基本概念是引入潜在的虚拟结果。如果能观测到同一个个体接受干预处理和未接受干预处理的两个响应结果的话, 那么, 可以用这两个响应结果的差来评价该干预处理对这个个体的因果作用。但是, 在流行病学和医学研究中, 每一个个体仅处在一个处理状态下, 要么接受干预, 要么未接受干预处理。因此, 我们仅能观测到一个响应结果, 另一个是观测不到的, 这个观测不到的结果是虚拟结果。基于虚拟事实模型进行观察性研究的因果推断时需要一些假定, 而这些假定是不可用观测数据进行检验的。近来, 虚拟事实模型开始被应用于进行因果推断, 应用此模型可以给出关于因果作用的最精确的定义和描述, 同时给混杂予完整的形式化定义^[3, 4, 11, 12]。

1. 因果作用^[5, 12]: 令 E 表示处理(或暴露)状态, 取值 e 表示被处理(或暴露), $\bar{e}$ 表示未处理(或未暴露)。令 D_{eu} 表示个体 u 在被处理状态 $E = e$ 情况下的响应, $D_{\bar{e}u}$ 表示个体 u 在未处理状态 $E = \bar{e}$ 情况下的响应。

定义: 假定在时刻 t_1 , 个体 u 同时被处理 $E = e$ 和未处理 $E = \bar{e}$ 在时刻 t_2 , 两种处理状态下的响应值分别为 D_e 和 $D_{\bar{e}}$, 则在时间段 (t_1, t_2) 内, 处理 e 相对未处理 $\bar{e}$ 对个体 u 的因果作用定义为: $ICE(u) = D_{eu} - D_{\bar{e}u}$, 称为个体 u 的个体因果作用(individual causal effect, ICE)。

设总体 u 有 N 个个体, 则总体的平均因果作用(average causal effect, ACE)定义为总体中所有个体的个体因果作用的平均值:

$$ACE = \frac{1}{N} \sum_u ICE(u) = E(D_e - D_{\bar{e}})$$

平均因果作用是指总体中所有个体都进行处理后的结果的平均值与所有个体都不处理的结果的平均值之差, 因为任何个体只有一种处理状态, 因而 ACE 也是虚拟量。

在有不受处理因素影响的协变量存在的情形下, 设 X 为离散型协变量。协变量 $X = k, k = 1, \dots, K$, 将总体 U 分成若干子总体。

$X = k$ 的子总体的平均因果作用定义为: $ACE_k = E(D_e - D_{\bar{e}} | X = k) = E(D_e | X = k) - E(D_{\bar{e}} | X = k)$, 当 D_e 和 $D_{\bar{e}}$ 是 0-1 值变量时, $ACE_k = P(D_e = 1 | X = k) - P(D_{\bar{e}} = 1 | X = k)$, ACE_k 是指总体中所有 $X = k$ 的个体都进行处理后的结果的平均值与这些个体都不处理的结果的平均值之差。

2. 混杂因素的形式定义和准则^[3, 4, 7]: Kleinbaum, Greenland 和 Robins 曾分别给混杂下了一个定性的定义: 当控制某个因素使混杂偏倚减小时, 那么该因素为混杂因素。Greenland 和 Robins(1986)曾举例说明了可比较性准则仅是混杂因素的必要条件, 不是充分条件, 因此不能作为混杂因素的定义。可压缩性准则也不是充分条件, 它依赖于使用的相关测度和背景因素的水平尺度。

应用虚拟事实模型理论, 我们给出混杂因素的形式化定义。令 C 是一个离散的协变量, 它不受暴露的影响。

设 D_e 和 $D_{\bar{e}}$ 为 0-1 值的变量, 用下面的标准化方法作为调整, 估计虚拟概率, $P(D_{\bar{e}} = 1 | E = e) : P \Delta (D_{\bar{e}} = 1 | E = e) = \sum_{k=1}^K P(D_{\bar{e}} = 1 | E = \bar{e}, C = k) P(C = k | e = E)$

定义 1: 如果: $|P(D_{\bar{e}} = 1 | E = e) - P \Delta (D_{\bar{e}} = 1 | E = e)| < |P(D_{\bar{e}} = 1 | E = e) - P(D_{\bar{e}} = 1 | E = \bar{e})|$, 那么 C 是一个混杂因子。

上面给出的混杂因素的定义表示, 控制 C 得到的测量, 即标准化率比未经标准化调整的粗率更接近待估计的虚拟率。根据 C 进行调整可以减少混杂偏倚, 因此, C 被定义为混杂因素。

定义 2: 如果: $P \Delta (D_{\bar{e}} = 1 | E = e) = P(D_{\bar{e}} = 1 | E = \bar{e})$, 那么 C 是一个无关因素。由定义可以看出, 对无关因素进行调整不能减少混杂偏倚。

定义 3 如果因素 C 存在某个水平的划分 p (即 $p = \{\omega_1 \dots \omega_M\}$ 其中 $\omega_k \neq \emptyset$, $\omega_k \cap \omega_j = \emptyset$, $\bigcup_k \omega_k = \{1 \dots K\}$, $M \leq K$ 使得 $|P(D\bar{e} = 1 | E = e) - P_p(D\bar{e} = 1 | E = e)| < |P(D\bar{e} = 1 | E = e) - P(D\bar{e} = 1 | E = \bar{e})|$ 那么,称 C 为偶然混杂因素。

定义 4:如果 C 的所有可能的水平划分 p 都有 $:P_p(D\bar{e} = 1 | E = e) = P(D\bar{e} = 1 | E = \bar{e})$ 那么,称 C 为一致无关因子。

定义 3 表示如果存在 C 的某种水平划分使得根据该划分调整 C 后能减少混杂偏倚,那么 C 就是一个偶然混杂因素。定义 4 表示如果不存在 C 的任何水平划分使得根据该划分调整 C 后能减少混杂偏倚,那么 C 就是一个一致无关因素,不管如何划分和调整 C 都不能改变混杂偏倚的大小。

根据上面对混杂因素的形式定义,可以证明可压缩性和可比较性两类准则是互补、而不矛盾的准则。两类准则结合使用可以从混杂因素的候选集合中将非混杂因素更多地剔除出去。

外部变量 C 是混杂因素的三个必要条件为:①C 与暴露变量彼此独立,在非暴露组 C 与疾病彼此独立;②风险差关于 C 不是可压缩的;③相对危险度关于 C 不是可压缩的。条件①是可比较准则,条件②和③是可压缩准则。如果 C 是一个由定义 1 定义的混杂因素,那它必须同时满足以上三个条件。

五、如何控制识别与控制混杂^[1,7]

从理论上说,采用随机化设计进行实验研究,如果样本量足够大,可以控制与消除混杂的影响。但在实际研究中,我们采用的方法多为观察性研究,因而在研究设计阶段与资料分析时如何判断识别混杂非常重要。某个或某些因素是不是混杂因子?是否需要调整?调整后所得到的结果是接近真实结果还是更加背离真实结果?回答这些问题可以考虑两种方法:第一种方法是将观察性研究所得到的结果与随机化实验的结果进行比较。例如研究某药物的疗效情况,分别采用观察性研究与实验性研究,通过比较我们自然可以找出观察性研究中的混杂因子。虽然这样的比较没有实际的应用价值,但我们总是可以举一反三,类推其他研究可能存在的混杂。医学伦理学限制我们在很多情况下无法进行实验研究,大量的流行病学研究只能依靠观察法。如吸烟与肺癌关系的研究,雌激素与女性恶性肿瘤关系的研究,职业或环境暴露与疾病关系的研究等。第二种方法是专业知识经验法,即,依据我们已经掌握的科学知识去判断某个因素是否需要控制。从专业知识上判断外部因素与研究的暴露因素是否有关,与疾病的发生是否有关。如在研究许多疾病的病因时,控制年龄、性别所得到的结论通常更加可靠。但具体到某外部因素 C,是否需要控制,目前的统计与流行病学方法还无法直接回答这个问题。因为我们不知道真实值是什么,我们需要向有关专家请教各种情况,控制 C 后得到的测量与专业知识是否符合,与其他类似研究结果的比较,与既往研究结果的比较等,综合判断。

在观察性研究的设计阶段,对选择的研究对象加以各种条件的限制是控制潜在混杂的好方法,但它会造成实施的困难与结论外推的局限。从理论上说,在队列研究中应用匹配方法可以很好地控制混杂,但由于失访存在,往往得不偿失。在病例对照研究中应用匹配方法,看起来似乎可以直接控制混杂,而事实上它常会引入一个可通过分层分析控制的偏倚。

目前的分层分析方法与多因素分析方法很多,可利用的应用软件也很普遍。不少人认为,分层调整后的测量一定比粗测量准确,多元分析方法可解决混杂问题。我们认为,这些观点是不正确的。粗测量应该是第一重要的,其次是分层分析、分析评价各层内的测量效应。如果考虑某因素可能造成混杂,考察其是否符合混杂的“三个必要条件”,首先考虑是分层调整,最后才考虑应用多元分析方法。应用多元分析方法,必须要考虑其适用条件,如变量的独立性问题、分布问题、共线性问题等,盲目使用多元分析方法极为有害。

本文对有关混杂问题作了初步探讨,回顾与评价了混杂的可压缩与可比较的两种判断准则,对其相互关系进行了分析。采用虚拟事实模型理论给出了混杂的形式定义,补充完善了关于混杂的理论。由于篇幅的限制,一些推导过程被简化了,一些较复杂的理论问题也未在文中列出,如强可忽略性假设条件、测量的强可压缩性等问题。基于目前关于混杂的理论,文中简要提出了一些控制混杂的方法。本文的所有讨论均是结合国外近年来关于混杂的理论,依据我们的研究结果作出的,错误之处,敬请同道批评指正。

参 考 文 献

- 1 Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol, 1981, 114:593-603.
- 2 Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Van Nostrand Reinhold, New York. 1982. 244.
- 3 Geng Z. Collapsibility of relative risks in contingency tables with a response variable. J Roy Statist Soc, 1992, B54:585-593.
- 4 Geng Z, Guo J, Lau T, et al. Confounding homogeneity and collapsibility for causal effects in epidemiologic studies. Statistica Sinica 2001, 11:63-75.
- 5 Whittemore AS. Collapsibility of multidimensional contingency tables. J Roy Statist Soc, 1978, 6:328-340.
- 6 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika, 1983, 70:41-55.
- 7 Greenland S, Robins J, Pearl J. Confounding and collapsibility in causal inference. Statist Sci, 1999, 14:29-46.
- 8 Holland PW. Statistics and causal inference. J Am Statist Asso, 1986, 81:945-970.
- 9 Lewis D. Counterfactuals cambridge: Harvard University Press. 1973.
- 10 Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychology, 1974, 68:688-701.
- 11 Greenland S, Robins JM. Identifiability exchangeability and epidemiologic confounding. Int J Epidemiol, 1986, 15:413-419.
- 12 Wickramaratne PJ, Holford TR. Confounding in epidemiologic studies: the adequacy of the control groups as a measure of confounding. Biometrics, 1987, 43:751-765.

(收稿日期 2001-10-15)

(本文编辑:段江娟)