

临床研究设计要点

易侗学院 陈常中

有了明确的研究假设，开始研究的第一步是做好研究设计。就像盖一座大楼首先要有设计，打一场战役首先要有作战计划。没有计划的行动不会产生需要的结果。本章简略回顾一下流行病学研究设计类型、原理和要点。

第一节 临床研究基本类型

一、观察性研究与试验性研究

流行病学研究设计首先要紧紧围绕一个研究假设，这相当于做一篇文章首先要有个主题。流行病学研究假设是要验证一个危险因素 X 与疾病或结局 Y 有没有联系，根据 X 是不是人为分配的，分为观察性 (Observational) 研究和试验性 (Experimental) 研究。如下图 3-1 所示。

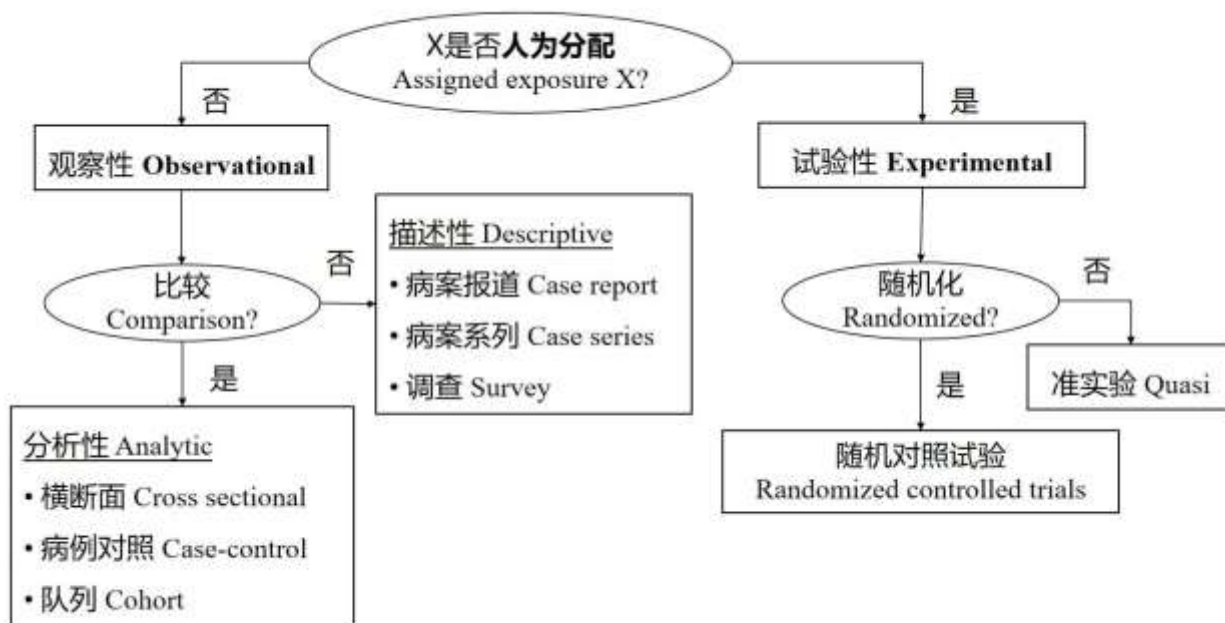


图 3-1 研究设计

试验在英文里是 trial，这个词最早出现于 15 世纪，其用意是 the action or process of trying or putting to the proof。临床试验指的是在人为制造的一种环境或场景下观察事物的发生与发展过程，目的是为了提供证据验证某假设。这里的关键是“人为制造”的，制造什么？当然是制

造“因”了，因在前果在后，我们通过操纵因来观察果的出现，这就是试验。如果不人为操纵因，仅仅是观察因、果的先后出现，那就是观察性研究。

有人说区别一项研究是试验性还是观察性，要看有没有“干预”，然后就有人会陷入到“干预”的字义里拨不出来。临床实践中离不开给病人用药、手术、检查等治疗或干预，临床研究中的“因”本身就是一个人为的干预，也就可以理解为是一个人为制造的环境或场景。说到这里，我们不得不意识到文字的局限性。其实，要拨出来很简单，只要看“干预”的动机，也就是目的是什么。如果“干预”目的只是为了诊断与治疗病人，也就是说，不管你做不做这个研究，这个过程都会发生，那是临床实践的自然过程，就不属于人为制造的，用于研究目的的环境或场景。如果“干预”的目的中有了为了试验，要达到试验目的，那就会人为分配不同的干预，从而人为制造不同的场景，这就是试验性研究。这好比看一个人是否孝顺父母，要看他内心是不是孝顺，而不能只着眼于看他怎么做的，否则贫家无孝子。看一个研究是观察性的还是试验性，不是看有没有干预，而是看干预的目的是什么。

（一）试验性研究比观察性研究更有说服力吗？

有这样一个故事，一首长带着一警卫员在出差的途中，换乘班车时一时买不到票，时间紧迫，急中生智，首长装作警卫员闯进售票室，向售票员要求马上给他们两张票，因为首长有紧急任务不能耽搁。售票员看了看眼前的“警卫员”，再望望站在不远处的“首长”，就卖给了他们两张票。原来警卫员长得高大魁梧，装得派头十足，而首长本人却身材平平，但首长毕竟是首长，说话有水平又有底气，这么调换一下成功的把握大多了。

这个故事说明了什么？事实上首长不一定就比警卫员身材高大魁梧，看起来更有派头，但人们的眼里总会有这个偏见。试验性研究是否比观察性研究更有说服力，看的是其设计与实施过程是否完美。

一座建筑的美不是因为其建筑材料多高级，不是因为其占地面积多大，而在于它的设计、它的架构、它的坚固、它的厚实、它的实用。一栋大楼如果设计不好，浪费材料事小，建成后经不起风吹雨打事大。

一项流行病学研究的美，也是如此。它在设计上的巧妙最终体现在它收集证据的效率与证据的力度。好的研究不仅花费少、时间短，而且偏性控制得严实，混杂因素收集得全面，各变量采集的时间既清晰又完整，各指标测量的方法既明确又统一。而糟糕的研究设计不仅花钱多，费时费力，而且收集的数据漏洞百出，用作证据时经不起推敲。

（二）RCT 研究特点

根据是否做到随机化，实验性研究分为随机对照试验（Randomized controlled trials, RCT）和准实验（Quasi）。随机对照试验，首先从时间上看 X 发生在前 Y 在后，非常明确。其次，RCT 研究一般都要做到盲法，目的是用盲法防止偏性。再有，RCT 研究做到随机化，目的是控制混杂，使除 X 外的影响因素在组间随机分布。

（三）病例对照、队列研究和横断面研究的特点

如图 1 所示，根据是否有比较，观察性研究分为描述性（Descriptive）研究和分析性（Analytic）研究两类。描述性包括病例或病例系列报告和抽样调查。病例或病例系列报告通常是描述一个不寻常或新发生的疾病或事件，对病人的症状、体征、诊断、治疗和随访的详细报告，也包含病人的人口统计资料。人群抽样调查报告主要是描述人群中疾病或健康状况及暴露因素的分布情况，为提出病因假设和进一步研究提供线索。描述性研究是分析性研究的基础。一个专科医院或一个科室，在开展针对其某一特色疾病的新治疗措施研究前，首先要有对某一段时间内收治的所有病人，从临床特征、治疗措施、治疗效果和预后等的总结报告，有了基础水平的评估，才知道是否有提高。

分析性研究又分为横断面研究（Cross-sectional study）、病例对照研究（Case-control study）和队列研究（Cohort study）。分析性观察研究是最常见的研究类型，它们之间的主要区别是收集 X 和 Y 的时间点不同。

如图 3-2 所示，时间线（Time）从左往右表示从过去至现在。①队列研究是先知道暴露（X）然后观察以后结局（Y）发生情况，有比较明确的 X 在前 Y 在后的时相关系。②病例对照研究是先知道结局（Y），然后看过去的暴露（X）情况。一般来说，如果 Y 的发生时间不明确，不容易确定 X 在前 Y 在后的时相关系。③横断面研究：是在一个时点同时检测暴露和结局。一般来说更不容易确定 X 在前 Y 在后的时相关系。

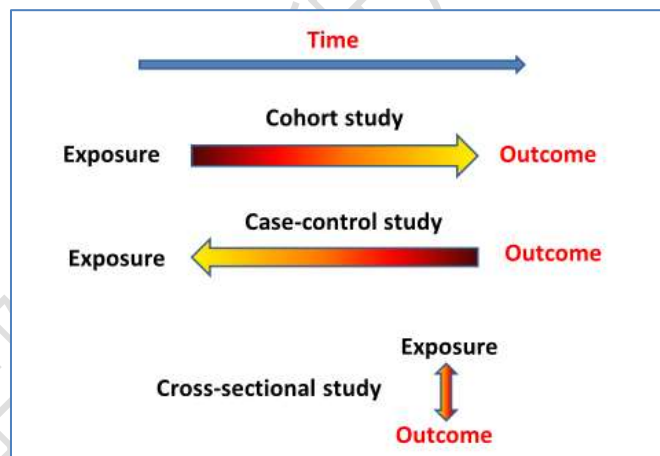


图 3-2 三种观察性研究时间特点

明确因在前果在后的时间关系、防止偏性、控制混杂是研究因果联系三大要素。做好了这三点即使是观察性的研究也是很有说服力的。

二、 队列研究基本原理

流行病学研究 X 与 Y 是否有因果联系。有没有联系是通过观察 X 变化，Y 是否跟着变化来发现的。X 是因，Y 是果，X 变化在前，Y 变化在后。队列研究，观察一个队列，也就是一批人，测定或给予其 X，然后通过随访看他们 Y 的变化，这个过程演示了 X 变化在前 Y 变化在后的过程。因此，队列研究是

必须要理解的流行病学研究类型。其它类型的研究，如横断面研究、病例对照研究、随机双盲临床试验等，都可以看做是队列研究的衍生物。

我们以两分类的 X 与两分类的 Y 为例，分析一下队列研究的原理。假设某种肿瘤有两种不同的手术方案， $X=0$ 表示方案 I， $X=1$ 表示方案 II，关心的结局是术后复发， $Y=0$ 没有复发， $Y=1$ 复发，伴随复发的另一个变量是复发的时间 T ， T 是从手术的那一天开始算起。研究 X 与 Y 是否有联系。队列就是一批患该肿瘤的人，先给他们做手术，一部分人用方案 I ($X=0$)，另一部分用方案 II ($X=1$)， X 要有差异，如果所有的人都是 $X=0$ 不行，都是 $X=1$ 也不行。 X 变化越大，越容易观察出 X 与 Y 是否有联系，如果大部分人是 $X=0$ ，少数人 $X=1$ ，表示 X 变化不大，如果 50% 的人是 $X=0$ ，50% 的人是 $X=1$ ，表示 X 变化最大。 X 在前， Y 变化在后，观察的是术后复发， Y 一定是在后的。随访时不仅记录是否复发，还记录随访时间。

我们可以设想一下，最终的观察结果会是哪几种情况。（1）一部分人复发了，一部分人没有复发。我们就可以比较 X 不同，复发率是否不同。（2）所有人都复发了，但复发时间有很大差异。我们也可以生存分析方法，把复发时间用上，比较 X 不同，复发风险是否不同。（3）所有人没有复发，可以再延长随访时间，直到有复发为止。（1）和（2）两组情况都是 Y 有变化，是否复发的状态有差异或复发的时间有差异。如果 Y 的变化（差异）越大，越容易能观察到 X 与 Y 是否有联系。

队列研究的第一个环节是入选队列人群，入选要求是理论上每个人都有可能发生 Y 。换句话说就是那些肯定不会发生 Y 的不要，肯定会发生 Y 的也不要，因为这两类人在里面只会稀释 X 的作用。

第二个环节是观察 X ，如果 X 是一个干预，如手术、用药，就把 X 发生的时间与类别记录下来，如果 X 是现有的某个特征如体重指数，那就把体检时间与结果记录下来；如果是体内某指标的水平，那就把采样时间记录下来。

第三个环节是观察 Y ，虽然队列里每个入选的研究对象都有可能发生 Y ，但最终不会是每个人都发生，即使每个人都发生了，也有发生得快与慢之别。如果 Y 是死亡，我们知道每个人最终都会死亡，但存活的时间不一样。即使 Y 是 0/1 两分类的状态，也常常会有伴随的量变指标。如是否高血压，伴随的量变指标就是收缩压与舒张压的值。如是否糖尿病，伴随的就是空腹血糖值。如果没有特定的量变指标，也是可以通过数据挖掘找找与之相关的量变指标。这就需要我们多设想、多收集一些观察指标。

在实际人群中实施队列研究时，首先我们不可能让一批人排队等着我们去入选；如果 X 是干预措施，我们不可能让所有的人同时实施 X 干预；每个人进入队列的时间不在同一个时点，开始随访的时间也不在同一个时点，每个人发生 Y 的时间也一定不在同一时点，如果发生 Y 了就停止随访，一般来说也不能等到每个人都发生 Y 了才中止研究。因此，所观察的队列是动态变化的，随时都可能有人进，随时都可能有人出。这就牵涉到很多变数，如让谁进，让谁出。

变数越多，越不牢靠。三角形只有三个变数，最牢靠；四边形、五边形呢，边越多越不牢。要想队列研究最后得出的结论牢靠，不被质疑的关键是减少入选、观察过程中的变数。首先在入选队列时：
①要固定纳排标准；②固定参与单位，即在哪个医院或/和哪个科室入选队列；③固定开始入选时间，

从这个时点开始，所有符合纳排标准的都进入队列；④固定中止入选时间，在这个时点之后，再好的病人也不能纳入队列。如果没有固定的入选开始与入选中止的时点，或者有这前后两个时点，但在这期间有符合纳排标准的人没有入选，人们就会怀疑你有偏性地选择研究对象来达到预想的研究结果，到那时你也无法澄清人们对你的怀疑。其次，⑤要固定随访中止时间。如果阳性事件（如复发、死亡）发生了，随访中止；如没有发生，也只随访到某一预先规定的时间，这个时间可以是整个研究的统一截止时点，例如均随访到 2021 年 1 月 1 日，也可以是一个统一的随访时长，如随访到患者出院即中止。如果随访中止时间没有固定，就有可能被质疑通过操作随访中止时间来操纵最终结局状态。

除此之外，在实施过程中要有书面的操作手册，包括岗位设置，岗位职责和各岗位操作规范，并有培训上岗标准和考核。

在实际人群中进行队列研究要比我们想象的复杂多了，通常需要先做一个预研究，以发现实施过程中可能出现的问题，检验和修订规范化的实施方案。

队列研究的优势是可以计算 Y 的发病率，发病率的分子是发生 Y 的人数。病例对照研究、横断面研究都没有人时，无法计算发病率。

第二节 流行病学研究中的偏倚

一、什么是偏倚？

流行病学教材里一个重要名词是“bias”，翻译成偏倚。测量学中偏倚是指一切测量值对真值的偏离。临床研究中的偏倚（bias）是指从研究设计、到实施、到数据处理和分析的各个环节中产生的系统误差。可分为三大类，选择偏倚、信息偏倚、混杂偏倚。每一大类里又可细分为很多类。如选择偏倚有入院率偏倚、新发病例偏倚、检出信号偏倚、无应答偏倚等；信息偏倚有回忆偏倚、报告偏倚、调查者偏倚等，而且还会有新的分类出现。

对初学者来说，偏倚本身就是一个陌生的名词，再具体到这一类那一类的偏倚，就更陌生了。即使记住了偏倚的定义，并能区别这些偏倚的分类，不等于研究设计就会做得好。相反，这些定义与分类很容易让初学者陷入概念的区分中。

有一则哲理故事：有一艘船载着三个人，其中一个是美国著名的物理学家，另外两个分别是美国著名的生物学家和数学家。不料在海上发生了意外，为了挽救另外两个人的生命，把损失降到最小，必须把一个人扔下去，那么应当把谁扔下去呢？美国一家著名的报社以 1000 美金向全国征求最佳答案。大家各抒己见，陷入对物理、生物、数学哪一个更重要的大规模争论中，谁也无法说服谁。最后，奖金的得主竟是一个年仅十岁的小孩，他的答案很简单：把那个最胖的扔下去！

理解偏倚其实很简单，复杂起来的原因往往是我们自己人为造成的。偏倚就如打靶打偏了，研究结果偏离“真理”就是偏倚。控制偏倚就是要消除研究过程中任何可能导致研究结果偏离“真理”的因素。如果“真理”是 X 与 Y 没有联系，研究过程中从入选研究对象、收集资料到后面的数据分析，任何一个环节都可能出现偏倚，使得研究结果有联系。反之亦然，如果“真理”是 X 与 Y 有联系，偏

倚的结果有两种可能，一是偏向无联系，即偏向无效假设，二是偏向更强的联系（或更强的效应）。最终一项研究的结果是不是可信，是否有科研价值，就要看这个结果是否由偏倚造成的。人们会从各个角度对这个问题提出疑问。我们需要思考的是偏倚各种可能出现的地方、偏倚的方向和如何控制这些偏倚，最终探索到“真理”所在。观察性研究报告指南（STROBE）中明确要求对偏倚要结合研究假设（X和Y的关系）解读偏倚的方向和大小（Discuss both direction and magnitude of any potential bias）。譬如，研究结果是阳性，而偏倚的方向是朝向无效假设（即阴性），那么研究结果仍然是可信的，因为“真理”是X与Y有联系而且X对Y的效应比观察到的更强。

二、“测不准”不等于“区别对待”

偏倚是系统误差，区别于随机误差。随机误差也会导致研究结果偏离真理，所以很容易与偏倚混为一谈。然而，真正理解偏倚并将其与随机误差区别开来，却是真正理解流行病学研究的原理，做好临床研究设计的关键。这里我们无意帮助读者区别各种偏倚，而希望能帮助读者认识到什么才是流行病学研究真正意义上的偏倚。

我们可以简单地将随机误差理解为“测不准”，流行病学研究的目的是回答X与Y有没有联系？如果X与Y有联系，但因为对X或/和Y测不准，没有观察到X与Y有联系，这是统计上所讲的研究效能（power）的问题。样本量不够大也是导致测不准的一种原因。当研究效能不高的时候，没有观察到X与Y有联系，不等于X与Y确实没有联系，这是共识。因此，如果研究结果是阴性的，要计算并报告检验效能，帮助读者判断和解释结果的意义。

这里，我们引进一个通俗的名词“偏性”，偏性即偏心，偏性一定会带来偏倚，但偏倚不一定都是由偏性引起的。偏性换句话说就是“区别对待”，英文为“discrimination”。如果“测不准”没有区别对待，那就没有偏性。希望大家留意文献中在讨论部分常看到的“nondiscrimination”或“not differ by group”，可以帮助我们提高对偏性，乃至流行病学研究的基本原理的认识，这也是为什么要引进“偏性”这个词的原因。什么是“区别对待”？一定要有因人而异。在测量X或Y时，如对X高的人测量Y时偏向高，而对X低的人测量Y时偏向低，反之亦然，这就是因人而异。在选择研究人群时，在X高的人中倾向于选择Y偏高（或低）的人，在X低的人中倾向于选择Y偏低（或高）的人，这也是因人而异，就是偏性。如果对所有的人测Y时都偏高（或低），就没有因人而异；在选择人群时单纯选择X偏高的人或Y偏高的人，也没有因人而异，都不是偏性。

流行病学研究“测不准”的最差结果是“未能发现”，没有发现不等于不存在，这个大家都能接受。而“区别对待”的结果呢？会无中生有，歪曲真理，这是要坚决避免的。

下面以2019年发表在JAMA Pediatrics的一篇题为“Association of In Vitro Fertilization With Childhood Cancer in the United States”的文章为例，说明“测不准”不等于偏性。该研究是一项回顾性队列研究，其X是体外受精（是/否），Y是子代儿童癌症（是/否），作者在讨论的局限性写道：我们无法检测儿童何时离开参与州或在婴儿期死亡，因此我们研究中的人时数（person-time）可能被稍微高估（原文：We also had no way to detect when children moved out of participating states or died after infancy, with the result that person-time in our study may be slightly overestimated）。高估人时会导致低估癌症发病率，但只要分母误差在两组之间没

有差异，就不应该使 HR 有偏倚（原文：Overestimation of person-time would lead to underestimation of absolute cancer rates, but so long as denominator error did not differ by group, it should not have biased the HRs）^[1]。

俗话说，不患贫而患不均。“测不准”没关系但不能有“区别对待”。

偏性是人的本性，是客观存在的自然规律。临床研究是人进行的对人的研究，如果不能客观地理解和正视人的偏性，就不可能真正跨入临床科研之门，做不好临床科研。本章在下面两节将重点讨论观察偏性（信息偏性）和选择偏性。

第三节 研究过程中的观察偏性

一、观察偏性无法度量

疑邻盗斧：从前有个人，丢了一把斧子。他怀疑是邻居家的儿子偷去了，便观察那人，那人走路的样子，像是偷斧子的；看那人的脸色表情，也像是偷斧子的；听他的言谈话语，更像是偷斧子的，那人的一言一行，一举一动，无一不像偷斧子的。不久后，他在翻动谷堆时发现了斧子，第二天又见到邻居家的儿子，就觉得他言行举止没有一处像是偷斧子的人了。

如果你想验证 X 与 Y 有联系，你会不会像丢斧子的这个人一样，处处都觉得 X 与 Y 有联系并有意无意地朝这个方向收集数据呢？会！这是人性，是自然规律，忽视这个偏性就是不遵从自然规律。

也许你会承认你有观察偏性，所以你不亲自参与治疗，不亲自护理，不亲自观察，不亲自收集资料，你认为这样就可以避免观察偏性吗？错！楚王好细腰，宫中多饿死，上有所好，下必甚焉。说不定你课题组的调查员为了满足你的愿望，偏性比你还要大。

有人会想，既然观察偏性一定存在，我们是否可以在数据分析时，在回归方程中引进一个是谁观察的这样一个变量来调整观察偏性呢？答案是否定的。观察过程中的偏性对结果的影响是无法在数据分析时调整的。人心膨胀的幅度与差异是无法用数字来度量的。也就是说主观带来的观察偏性的程度无法度量，也就不可能用变量来调整掉。

由主观观念带来的观察偏性，就像一粒种子，丢到土里就会发芽，长出来可能是一株小草，也可能是一颗参天大树，无法预计。唯一能控制的办法是不把它种进土里，让它无处发芽无处生长。这就是流行病学研究设计中使用盲法的原理，我们无法控制人的主观偏性这一粒种子最后能长多大，但我们可以让它无处发芽。在我们不知道所观察的对象是对照组还是试验组的时候，或者说在观察 Y 的时候不知道 X，在观察 X 的时候不知道 Y，就无法将 X 与 Y 联系起来，就无法“区别对待”，这样主观偏性这粒种子就无处发芽。

观察偏性客观存在并且无法度量。当我们尊重观察偏性这个自然规律，才能真正理解为什么要用盲法，才能真正做好盲法。观察偏性如没有控制好，整个研究就可能报废。

认识到观察偏性的存在，就能看到回顾性研究的优势，就不会总有前瞻性研究比回顾性研究好的偏见。回顾性研究收集的是既往已经发生的数据资料。因为过去在临床工作中，医护们并不知道你今天要研究这个 X 和 Y 的关系，因此对患者是一视同仁的，在治疗方案的选择、指标测量、预后评估和沟通过程中，无法偏向于你今天的预期结果。

二、如何区分回顾与前瞻？

很多流行病学教材都有回顾性研究与前瞻性研究的定义，有整段的文字详细解释这两者的区别。然而，即使是这样，人们在遇到问题时仍有可能混淆不清。某网站上就曾有一个求助贴：回顾性分析某个已结束的大型前瞻队列研究的数据，是前瞻性队列研究还是回顾性队列研究？求助者说他把《流行病学》第三版，关于回顾性研究与前瞻性研究的定义与优缺点比较，搬出来一一对照，仍不得其解。

首先，我们要意识到这不是书的作者水平有问题，这是文字表达的局限，文字表达无法呈现思想的全貌。“书不尽言，言不尽意，自觉圣智，完成人格”不管看了多少，听了多少，都需要结合自己实际情况，形成研究设计。其次，我们需要关注的不是名词，而是真理。回顾也罢，前瞻也罢，就像一个人的名字，是叫张三，还是李四，不重要，重要的是他是什么人。就如六祖慧能所说，“真理是与文字无关的，真理好像天上的明月，而文字只是指月的手指，手指可以指出明月的所在，但手指并不是明月”。

那这里的“明月”是什么？关键是要理解为什么要区别回顾与前瞻？就研究设计而言，之所以要区别前瞻与回顾，体现在“回顾防缺失，前瞻需设盲”这句话上。回顾指的是站在今天看过去已经发生的事，数据已经存在，资料收集的过程或部分过程已经完成。前瞻则相反，站在今天看未来，数据尚未发生。“今天”指的不是此时此刻，而是你收集数据的那个时刻。当你查处方记录看在某一天之前研究对象有没有某药的处方记录的时候，数据已经发生，记录已经存在。过去在观察和记录该数据的时候，记录人并不知道今天要做此研究。而前瞻则相反，从某一时间开始，你随访研究对象看他有没有某 X（暴露因素）有没有某 Y（结局变量）发生，你观察 X 与 Y 的目的是要做此项研究。前瞻如果不设盲，在观察的时候就会有主观意识（种子）导致的观察偏性的问题，有种子也有发芽的地方。回顾呢，你有充足的理由说在观察的时候，种子还没有形成，或者说没有人给观察者这个种子。

区别回顾与前瞻不是目的，目的是鉴别观察过程中有没有主观偏性的种子存在。主观偏性换一种说法就是“区别对待”，这里强调的是“区别”，对 X=0 的人与 X=1 的人观察 Y 时有区别，或对 Y=0 的人与 Y=1 的人观察 X 时有区别，叫区别对待。这样的区别对待会导致 X 与 Y 有联系，偏倚的方向是偏离无效假设。如果对所有的人都一样的“好”或一样的“差”，一样的“严”或一样的“松”，不是区别对待，是观察误差。

三、“观察”的全过程

是不是回顾性研究就没有观察偏性的问题呢？不是的。还是这句话，文字只是帮助你看到月亮的手指。“观察”这个词本身就是很笼统的概念，一个变量的观察，如：是否服用过口服磷酸盐，要到该变量的取值进入数据库后才算完成。询问是一个过程，记录是一个过程，从病历资料中摘录这个变

量也是一个过程。一个指标如血糖的检测，从采样、样品的储存、运输、检测、记录、录入数据库是一个个过程，每个过程中都需要分析是否会有偏性存在。

回顾性研究如果采集数据只是靠询问，让研究对象回忆过去有没有服用某药就会有偏性。因为此时虽然真实的是否服药已经发生，但资料未收集，没有记录，实际能采集到的是否服药受询问者和被问者的主观倾向影响，完全可能脱离真实。如当你问一个高血压（ $Y=1$ ）的病人过去是否吸烟，他回答否，你可能会加问一句“真的没有？”，而如果问的是一个非高血压（ $Y=0$ ）的人呢，你可能就不会多问一句。这就是区别对待。被问者也是一样，有病（病例）的人比无病（对照）的人更倾向于回忆起以前的暴露情况，他们对待是否有暴露这个问题是有区别的。

回顾性研究如果是从现有资料（如病历资料）里摘录数据进行研究，只是询问与记录的过程、生物样品采集的过程已经发生了。在询问和记录的过程中，没有主观的种子，不等于摘录病历资料的时候没有。样品采集的时候没有偏性，不等于储存、运输、实验室检测的时候没有。要看种子什么时候形成，一旦形成了，就需要控制，否则掉到地里就会发芽生长。以“是否服用过口服磷酸盐”为例，如果通过人工查阅处方记录来获得这个变量，查阅处方人在查阅的时候是否有偏性呢？这要看他是否能知道处方上的病人 Y 的信息，如果研究的 Y 是骨折，当他知道病人是否有骨折，在查阅其处方时，自然就有可能对有骨折的人与无骨折的人仔细程度不同，这就是区别对待。

有人会想，开了口服磷酸盐处方的人中一定有一小部分人并没有服用，这是不是偏性？会不会使研究结果产生偏倚？如果 X 为是否口服磷酸盐，虽然有没有偏性要看是否因 Y 不同而有不同程度的“开了处方但不服用”，这要结合专业知识判断，然而 Y 是发生在后的，开了口服磷酸盐处方的患者在决定是否服用时， Y 没有发生，所以一般来说不会因为 Y 不同而有不同，我们可以说这里没有偏性。会不会产生结果偏倚呢？会，开了处方的人我们认为他服用了实际上没有服用，而没有开处方的人可以确定是没有服用，结果会缩小真正服用者与未服用者的差别，从而使研究结果偏向阴性，偏倚的方向是朝向无效假设。如果最终得出来的研究结果是口服磷酸盐与某 Y 有联系，那么真实的 X 对 Y 的效应应该更强。试想，有黑白 A、B 两瓶液体，如果在 A 瓶里滴了几滴 B 液后，你仍能分辨 A、B 两瓶熟黑熟白，那么在这之前 A、B 两瓶一定是更加黑白分明。

当我们把观察的过程扩展到数据进入电子化的数据库后才算观察完成时，就会发现，在这一步步的观察过程中，不仅有主观观念带来的观察偏性，还可能有非主观的无意造成的“区别对待”。例如，一项病例对照研究分析血清中 DDT（双对氯苯基三氯乙烷）水平与自然流产的关系^[2]，病例组是自然流产，对照组是活产。DDT 水平是实验室测出来，用盲法使实验室测量人员测量 DDT 时不知道 Y 是很容易做到的，但是不是这样就没有偏倚了呢？不是的，血样是什么时候采的很关键。最理想的是采集怀孕前一天的血样，然而病例对照研究是在确定了病例与对照后再去测 X ，实际血样是在活产若干月后采的。对照组是活产，活产就会有哺乳，而 DDE（DDT 的代谢产物）可以通过母乳排泄，这就自然造成对照组 DDE 偏低，这是研究设计的本身的缺陷。

样本的采集、储存、运输与检测过程中都有可能带来偏性，有些是意外造成的，如停电导致温度失控，样本破坏。如果意外的结果只影响到试验组或只影响到对照组，或者说对试验组与对照组的影响不同，这个意外就会改变 X 与 Y 的联系，就带来了偏倚。如果这个意外的发生可以追踪，其对哪些样本有影响，或者说其对 X 和 Y 的影响可以度量，我们就可以引进一个变量来调整这种影响。有时候

意外发生了，但我们没有觉察，其带来的偏倚是无形的，不可度量的。意外有时候又是不可避免的，因此，只能依赖于过程控制。在同一批运输样品中，均匀地放入试验组和对照组的样品，就是一种有效的控制方法，能想到这一点是因为对偏性的深刻理解，对偏性可能出现的环节的捕捉能力。而这个能力不是靠熟记了偏倚的定义与分类就能做到的。文献实例：一项巢式病例对照研究^[3]，分析 X 和冠心病的关系。病例组是发生冠心病的人，对照组是未发生冠心病者。核心结果是 X 与冠心病有关联，较高的 X 与较低的冠心病风险有关联 OR=0.80 (95% CI 0.70-0.90)。正文方法部分写到：对所有样品分两批运输，每批中的病例数和对照数相似；所有分析都针对批号进行了调整。原文 All samples were shipped (and consequently assayed) in two batches with a similar number of cases and controls within each batch; all analyses were adjusted for batch number. 为何运输时每批病例数和对照数要相似？因为如果运输过程中出现问题，例如冷藏用的干冰失效导致样本 X 的测量值偏高或偏低，那么病例组和对照组都面临同样的问题，不带来偏向性；然而如果一批样本都是对照组（无冠心病），一旦出现运输问题导致 X 测量值偏低，就会导致结果偏向 X 和 Y 有联系（X 低的偏向于没有冠心病），那么这篇论文的结论就可能是偏性导致的假象了。

第四节 流行病学研究中的选择偏性

流行病学研究因果联系。通过观察 X 变化，Y 是否跟着变化，得出 X 与 Y 有没有联系；观察 X 变化一个单位，Y 变化多大，回答 X 对 Y 的影响多大。最终得出的结果 X 与 Y 是否有联系，X 对 Y 的影响有多大，取决于三个方面，（1）X 与 Y 是否真的有联系，X 对 Y 的作用真值有多大，也就是说真理是什么；（2）我们在什么样人群中观察；（3）怎么观察的。流行病学研究的目的是要证明 X 与 Y 有联系，而是要揭示规律，发现真理。有联系就是有联系，没有联系就是没有联系。在正确的人群中用正确的方法进行观察，才能接近真理。相反，如果人群选择的不对，观察方法不对，研究得出的结果可能会歪曲真理，这是我们需要警醒的地方。

一、选择适当的研究人群以提高研究效率

X 与 Y 有没有联系，要通过观察 X 的变化与 Y 的变化来发现，首先选择的人群里，X 与 Y 要有变化。

X 要有变化。如果所有的人都是男性，你就无法观察性别的差异；如果所有的人都是小于 45 岁的女性，你也就无法观察绝经对某指标 Y 的影响。X 变化越大，越容易观察到 X 对 Y 的影响。如果 X 是连续性变量，看它的取值范围与标准差就知道 X 变化大不大。如果 X 是体重指数，取值范围仅在 21-24 之间研究人群，是很难观察到 X 与什么 Y 有联系。事物的变化存在着一个从量变到质变的过程，从 21 到 24，这个变化量是很小的，我估计饱吃几顿就差不多可以做到。如果 X 是分类变量，X 变化大不大体现在落在各类的比例是否均衡。以 0/1 两分类为例，如果所有的人都是 0，或所有的人都是 1，那就是没有变化；如果大多数人是 0，极少数人是 1，那是变化很小；如果一半人是 0，一半人是 1，那是变化最大。对于三分类的 X 来说，如果每一类的人数都是 1/3 那是变化最大。

Y 也要有变化。如果所有的人最后都活着，你就无法观察某 X 与死亡的关系。如果所有的人在观察期内都死亡了，你就要看死亡时间，也就是生存时间有没有变化，变化大不大。同理，Y 变化越大，越容易观察到 X 与 Y 是否有联系。衡量 Y 的变化是大还是小的原理与 X 是一样的。

虽然，X 变化越大，越容易观察到 X 与 Y 是否有联系；Y 变化越大，也越容易观察到 X 与 Y 是否有联系。然而，X 是因，Y 是果，X 变化在前，Y 变化在后；如 X 与 Y 没有联系，Y 就不会因为 X 变化大而有变化。我们在研究设计时，一般不会同时看到 X 与 Y 的变化。如果是队列研究或试验研究，先有 X 然后观察 Y，我们在选择研究人群时，就希望 X 变化大。如果是随机化对照的试验，最好是各试验组人数相同。如果是病例对照研究，先确定 Y 然后观察 X，希望病例组与对照组人数相同，这样 Y 的变化最大，研究效率最高。

在队列研究制订纳排标准时，不仅要考虑选择 X 变化大的人，同时要将那些已经或肯定要发生 Y 的人排除，一定不会发生 Y 的也要排除，以提高研究的敏感性。从病因通路模型上不难理解，假设 Y 的发生有①X+A+B、②D+C+E 两条通路，如果所有人都没有 A 或 B，不可能通过通路①发生 Y，我们也无法确定通路①是否存在。如我们不可能在男性中研究痛经。如果所有人都能通过通路②发生 Y，我们也无法确定通路①是否存在。如一项队列研究入选新婚妇女，用记日记的方式跟踪每天是否有被动吸烟、是否有月经和痛经，分析被动吸烟是否与痛经有联系，在分析时要将基线有痛经史的妇女排除在外，就是这个道理。

上面说的是为了提高研究敏感性，研究人群如选择不当，不容易观察到 X 与 Y 的联系。这是研究的敏感性问题，不是偏性，这里没有“区别对待”。区别对待是有区别地选择研究人群，具体地说就是会因为其 X 和 Y 的不同而有不同的选择倾向。选择偏性最突出的表现是在病例对照研究对照的选择上。

二、病例对照研究如何选择对照？

病例对照研究是通过病例组与对照的比较发现哪些因素与病例的发生有关。有比较，就存在着跟谁比的问题。俗话说比上不足比下有余，这句话道出了选择偏性的源头与结果。选择跟谁比就能得到什么样的结果。给我一个病例组 and 要研究的暴露 X，我一定能找出一个对照组使得两组 X 的水平显著不同，从而得出阳性结果。这是为什么说病例对照研究最脆弱、最易受到质疑的就是对照的选择。

病例对照研究怎么选择对照？一句话：要独立于暴露 X。什么叫独立？也就是说你不能因为他的 X 高或 X 低你就选他，包括主观和非主观有倾向性地选到他。不要主观地去选 X 高或 X 低的人做对照，这一点我们都可以做到。然而要避免非主观有倾向性地选到 X 偏向性的高或低的人，就比较难了。因为，这受到我们目前认知水平的限制。在医院里开展病例对照研究，病例资源比较丰富，然而选对照比较难，有人想到从另外一个科室选对照，这时问题的关键就在于到这个对照科室看病的病人与 X 是否独立？也就是说这个科室的病人是否 X 倾向性地高或低？如果就我们目前的认知，所选的科室与 X 无关，然而谁能保证，随着我们认知水平的提高，若干年后不会发现这里面有关系呢？例如：一位神经内科的大夫想研究某指标 X 和脑卒中的关系，本科室脑卒中的患者很多，可以找到病例组，然而去哪里找对照呢？眼科的青光眼患者可以吗？如果 X 高的人容易发生青光眼，那么对照组 X 水平会偏高。如果得出的结果是 X 高的人脑卒中风险低，就很可能是对照组的选择偏性导致的，而非真实关系。

病例对照研究该到哪里找对照？一句话：从还原队列里找对照。最理想的是巢式病例对照研究，病例与对照来自同一队列。什么是还原队列？首先我们想象一个人群，在某一时间点先把已经患病（Y=1）的人从这个人群中剔除，剩下的所有 Y=0 的人就是一个队列，跟踪这个队列若干时间后，就会发现有

一部分人发病了，Y 由 0 变成 1，这些人构成了我们所研究的病例组；另一部分人 Y 仍然是 0，这些人就是我们的备选对照，我们要选的对照应该是他们的代表。之所以做病例对照研究，是为了省去当初确定队列和跟踪随访的时间。换句话说，病例对照研究对照的选择要从提供病例的源人群中选。

从提供病例的源人群中选对照，只能在其中，不能在其外。如图 3-3 所示，A 代表一个大人群，B 代表提供病例的源人群，那么对照就应该从 B 里选，不能从 A 里选，也不能从 C 里选。B 是 A 的子集，A 中 B 之外的部分与 B 内没有发病的人可能是不一样的。同理 C 与 B 有交集，C 中 B 之外的部分与 B 内没有发病的也可能不一样。病例对照研究在设计时要仔细鉴别病例源人群。一般来说从另外一个科室选对照甚至从另外一个医院选对照是不可取的。

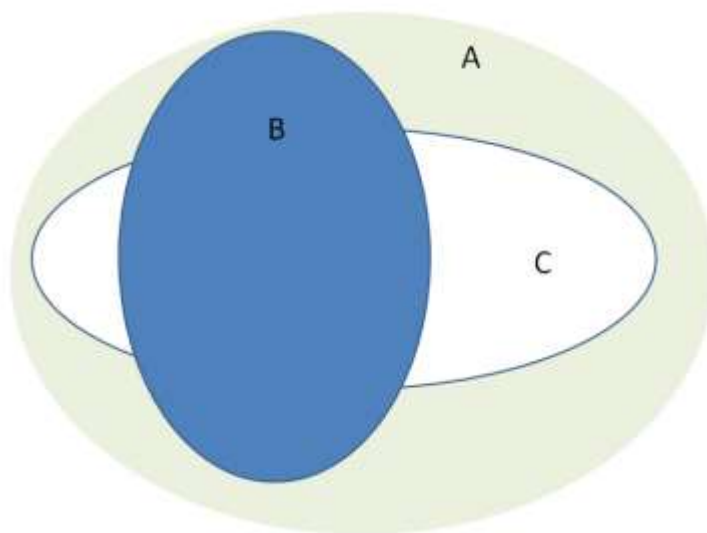


图 3-3 提供病例的源人群

这里要注意的是通过匹配选对照的方法，假如病例限定在年龄 50 岁以上，那么源人群也是 50 岁以上。然而，匹配并不能保证病例的源人群与对照的源人群相同。相反，因为观察性研究的 X 不是随机分配的，必然有某因素 Z 与 X 相关，如果把 Z 匹配了，就观察不到 X 的效应，过度的匹配会适得其反。

三、什么是选择偏性？

流行病学研究的目的是确定 X 与 Y 有没有联系，我们在选择研究人群时，特别是横断面研究和病例对照研究，要保持 X 与 Y 的独立，否则就会专门挑选出一批 X 与 Y 有联系的人。来证明 X 与 Y 有联系。什么叫 X 与 Y 独立呢？如果 X 是连续性变量，单纯挑选 X 偏高的人不是偏性，同理单纯挑选 Y 偏高的人也不是偏性。因为单纯针对 X 或单纯针对 Y 的选择，没有“区别对待”，X 与 Y 仍然是独立的。但如果不是“单纯选 X 偏高”，而是因为其 Y 的高低而有区别，如在 Y 高的人中多选 X 偏高的人，或在 Y 低的人中多选 X 偏低的人，这就不独立了，就有“区别对待”，反之亦然。看下面的这个图 3-4，X 与 Y 本来没有联系，中间一条红色虚线表示 X 对 Y 的回归线，回归系数为零。

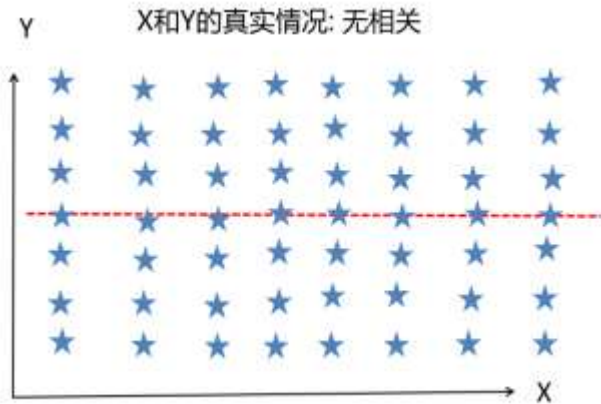


图 3-4 X 和 Y 关系的示意图

如果在图中单纯选择 Y 偏高（框内部分）的人，看 X 与 Y 的关系，还是无相关，回归系数仍为零（图 3-5）。

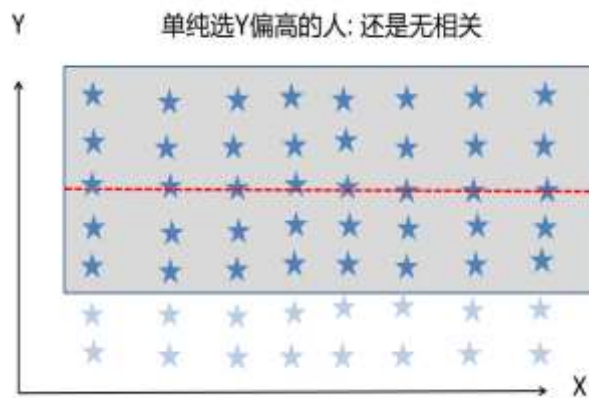


图 3-5 X 和 Y 关系的示意图：选 Y 偏高的

如果单纯选择 X 偏高（框内部分）的人看 X 与 Y 的关系，也是无相关，回归系数仍为零（图 3-6）。

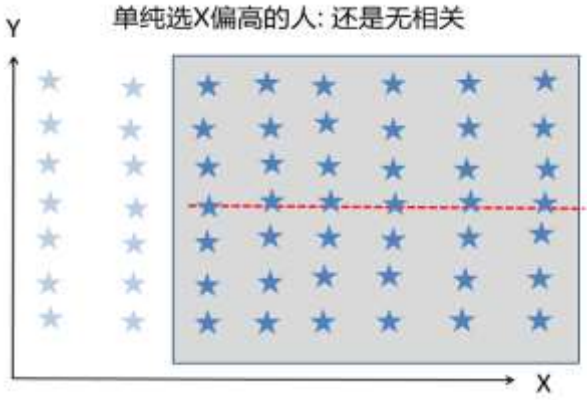


图 3-6 X 和 Y 关系的示意图：选 X 偏高的

但如果在 X 低的人中过多地选择 Y 低的人，在 X 高的人在过多地选择 Y 高的人，此时 X 与 Y 就不独立了，就会导致结果是 X 与 Y 正相关，回归系数大于零（图 3-7）。

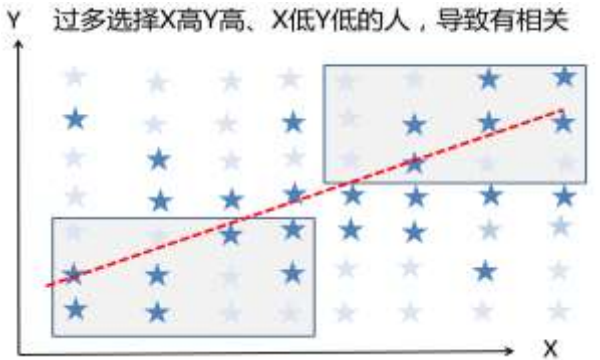


图 3-7 X 和 Y 关系的示意图：选 X 高 Y 高和 X 低 Y 低的

下面再一起讨论两个例子：

例 1：研究口服双磷酸盐（X）与骨折（Y）的关系，只选择老年女性做研究，会不会有偏性？

解答：老年女性发生骨折的比例高，老年女性发生骨质疏松的人也多，口服双磷酸盐是治疗骨质疏松症的，因此老年女性 X、Y 都可能偏高。可以提高研究的敏感性，但不会带来偏性。但如果在老年女性中过多地选择有骨质疏松的人，就是对 X 高 Y 也高的人有偏心，就可能带来偏性。

例 2：某病有多种类型（如脊柱炎），现在要利用医院病历资料研究其中某一大类（如强直性脊柱炎）的发生与某 X 是否有关。研究人群是某段时间内到医院就诊的所有患脊柱炎的病人。如果通过搜索出院诊断来纳入病人，考虑到出院诊断脊柱炎可能多种名称，你在几个主要科室的病历里用可能出现的名称去搜索。这样做会不会出现偏性？如果会，是怎么发生的？

解答：有可能出现偏性。如果我们不能把该段时间内所有到医院就诊的患脊柱炎的病人纳入，那就要思考没有纳入的人是哪些人？他们是否多是 $Y=0$ （非强直性脊柱炎）的人？如果是，他们是否是 X 偏高或偏低的人？因为到医院不同科室就诊的病人其 X 水平很可能不同，我们忽略掉的那些科室很可能多是 $Y=0$ 而 X 偏高（或偏低）的病人，所以很有可能出现偏性。

如何避免选择偏性呢？需要我们静下心来，仔细分析研究人群入选流程与纳排标准，捕捉偏性可能出现的环节，从而加以规避。

四、RCT 招募与分组过程中的偏性

随机对照试验通常通过双盲来控制观察偏性，双盲指的是研究者包括资料分析者和受试者不知道受试者是在对照组还是试验组。如果做到了双盲是不是就完全没有偏性了呢？不是。招募与随机分组过程中还可能出现偏性。有人自然会说，既然是随机分组，由随机数决定病人到哪一组，怎么会有偏性呢？前面说过，偏性是人性决定的，研究者对研究结果有倾向性，通常倾向于两组有差异，非劣效性研究也可能倾向于无差异，有倾向性就会自觉和不自觉地向这个方向努力。理解了这一点，不难想象研究对象是研究人员通过纳排标准入选的，虽有纳排标准，但如果事先知道待入选的对象会分到哪一组，掌握纳排标准的时候是否会不一样？会。

有这样一个“7个和尚分粥”的哲理故事，从前山上的寺庙有七个和尚，他们每天分食一大桶粥，每天的粥都不够吃。一开始，他们拟定由一个小和尚负责分粥。但很快就发现，除了小和尚每天能吃饱，其他人都要饿肚子。于是和尚们轮流分粥，一人一天。这样下来，他们一周只有自己分粥的那一天是饱的。然后大家推选一个公认道德高尚的长者来分粥。开始这位长者还基本公平，但不久就开始为自己和讨好他的人多分，把整个小团体搞得乌烟瘴气。这种状态维持不了多久，和尚们决定分别组成三人的分粥委员会和四人的监督委员会，这样公平的问题基本解决了，但每次分粥方案很难达成一致，不等分完粥早就凉了。

最后，他们想出了一个办法，还是每人轮流值日分粥，但分粥的那个人要等到其他人都挑完后再拿剩下的最后一碗。令人惊奇的是，在这个制度下，7只碗里的粥每次都几乎是一样多，就像用科学仪器量过一样，从此和尚们都能够均等地吃上热粥。

这个故事告诉我们一个简单的道理：（1）仅有标准是不够，谁都知道分粥要均匀，要一视同仁。然而人有私心（偏性），即使公认道德高尚的长者也会被利益拉下水。（2）盲法是控制偏性的基本方法。分粥的人不知道谁会拿到哪一碗粥，就规避了偏性。

随机分组过程中的偏性控制要点是入选时不知道分组，即先入选后分组。只有做到了盲法，入选标准才能被无偏地执行。如何做到入选病人时不知道分组，又要入选后即时分组并尽可能使随机分到各组的人数相同，这有很大的挑战性。随机区组大小的区组随机是一个有效的方法，详见“易倚 DataWeb 系统中央随机”部分。

第五节 混杂控制和时相关系

一、研究设计中的混杂控制

一个疾病或结局事件的发生有多种原因，可由多条通路导致。研究危险因素 X 与结局 Y 是否有联系，实际上是确定是否有“X+A+B”导致 Y 这样一条通路存在，此处 A 和 B 代表其它未知的因素。如果把人群按 X 分成有、无两组，在没有 X 的一组只能通过其它通路导致 Y，而有 X 的则多了“X+A+B”这条通路，如果这条通路不存在，那么两组发生 Y 的几率相同，否则有 X 组发生 Y 的几率就高。这里有一个前提，就是两组通过其它通路导致 Y 的几率相同。

随机化的对照试验，就是通过随机分组的方法，使得两组发生 Y 的基础概率相同，然后再在试验组添加 X，比较试验组与对照组发生 Y 的几率。随机的目的是控制混杂。然而不是所有的 X 都可以随机分配，尤其是临床研究。如 X 是吸烟，我们不能随机分配谁吸烟，谁不吸烟。如果 X 不是随机分配的，是自选的或其它习惯和条件造成的，必然有因素与 X 有关。如男性吸烟的比例大于女性。这就不能保证通过其它通路导致 Y 的发生几率相同。因此，观察性研究所需要面对的最大挑战是如何控制混杂。

某因素 Z 构成混杂的条件其一是与 X 有关。如果 Z 与 X 无关，那就是说明在 X=0 与 X=1 两组通过有 Z 参与的其它通路导致 Y 的几率是相同的。其二是 Z 与 Y 有关，如果 Z 不在任何一条导致 Y 的通路出现，那么 Z 多一点少一点不影响发生 Y 的几率。

观察性研究的混杂控制有赖于在设计与资料收集时把可能构成混杂的因素收集全，然后在数据分析时通过多元回归模型或/和分层分析控制混杂因素的影响。

即使是随机化的对照试验，当样本量比较小的时候，随机也不一定能保证试验组与对照组混杂因素的均衡，因此也有必要收集可能的混杂因素，以便在后续的数据分析时做进一步的调整。

哪些因素可能构成混杂，需要结合临床知识和经验并查文献。首先将与 X 有关的因素、与 Y 有关的因素列出来，两者有交集的部分一定是要收集的。除此之外，反映研究人群特征的一些变量，如年龄、性别、职业、文化程度、地区等，一般来说不可忽略。在确定要收集哪些变量时，需要理解变量的定义及其含义。变量根据含义，一般可以分为两种类型：

(1) 直接变量：直接测量出来的反映某指标的值。如身高、体重指数、收缩压值、舒张压值、血清中某指标的浓度等。

(2) 间接变量(surrogate)：综合代表某一类特征。如教育程度，不只是代表上了多少年的学校，而且间接反映了经济状况、居住环境、工作性质等有关因素。流行病学上一些常见的研究人群特征变量，如文化程度、职业、性别、年份等，是间接变量。

在研究设计时，一般来说，Y 要用直接变量；X 也要用直接变量。试想，如果 X 是间接变量，如职业，结果得出职业与高血压有关，那到底是职业暴露中的哪种因素与血压有关呢？是工作环境中的某种有害物质的暴露，还是职业压力？如果是职业压力，不如直接用工作压力测量指标作为 X。调整变

量 Z 可以用直接变量或间接变量。很多流行病学研究都要调整年龄、职业、文化程度等。校正这些变量，不只是校正了变量本身的含义，而且在一定程度上校正了它们所代表的很多其它因素的综合作用。

二、暴露、混杂和结局之间的时相关系

曾有这样一个电视剧情节，一个检察官从外地调查取证回来，一下车就被公安拷走，原因是在他家搜出毒品。他知道是被贪污集团栽赃陷害了，目的是阻止他第二天出庭。情急之下，他想到借助包装毒品的报纸上的日期，证明这包毒品不是他放的，因为在这个日期之前他已经去外地了。结果，第二天他顺利出庭将犯罪集团一举拿下。这个例子可以启发我们在流行病学研究中，该如何收集证据、提交证据和思辨证据力度。

时相关系顾名思义就是时间的前后关系。因在前果在后，混杂因素也是导致 Y 的因素，当然也应发生在 Y 之前。在研究设计时要充分考虑到何时收集这些变量，在二次课题设计提取数据时要充分考虑提取何时采集的数据，以确保因在前果在后。当然，在数据分析构建模型时要注意放入的变量所代表的时间。譬如一项分析导致院内感染的危险因素研究，要控制是否服用过抗生素这个变量，这里的关键就是抗生素是在院内感染发生前还是发生后服用的。

这里要注意的是变量的采集时间不等于测量时间，如血样采集的时间在前，测量时间在后，甚至是若干年后，然而测量结果代表的是研究对象采样时的水平。有些问卷指标询问的是以前某一时间段的情况，不是研究对象询问时的情况。这些在设计和分析时都要注意到。

如果某指标以前的水平测不到了，能否用现在的水平代替以前的水平呢？这就要结合测量的指标特性和研究对象的状况，具体问题具体分析。如人体内 DDE (1,1-二氯-2,2-雙(對-氯苯基)乙烯)的水平比较稳定，一般来说可以用现在测得的水平代替以前的水平。然而，如果研究对象是有过哺乳的妇女，DDE 可以通过母乳排泄，现在的水平就不能代替哺乳前的水平。

后期干预的作用：不难想象一项队列研究，研究对象入选后测到一些基线指标，要分析某基线指标 X 对最后结局 Y（如死亡）的影响，在入选之后对病人所做的一系列干预措施（A）对结局的影响要不要调整？如何调整？

首先在研究设计时，要不要收集这一系列的干预信息？这取决于研究者预先设定的目标。如果目标比较大，要把病人诊疗过程中的每个环节对最终结局的影响都研究清楚，就需要收集，如果收集就不仅要记录具体的干预措施是什么，还要记录每项措施的实施时间？在具体分析时，需要结合专业知识。如果说后期的干预措施完全是由基线 X 的水平或状态决定的，在分析 X 对 Y 的影响时，就不能控制后面的干预措施。如果说根据 X 的水平或状态，有不同的干预措施，你可以研究在 X 相同的情况下，不同干预措施对 Y 的影响，或分析 X 与干预措施交互作用对 Y 的影响。这又回到要明确研究假设上了，不仅明确 X 是什么 Y 是什么，还要设想以假设为前提的临床场景是什么？各变量之间关系的表现形式是什么样的？如果说根据 X 的水平或状态，有不同的干预措施，你在分析 X 对 Y 的影响时，如果未结合 X 之后的干预措施进行分析，意味着这个研究有缺陷。如果这是个前瞻性的队列研究，人们会怀疑你通过操纵后面的干预措施有目的地使 X 与 Y 有联系。

三、研究设计要点小结

总结研究设计要点，一句话“偏性、混杂与时相，研究设计细思量”。流行病学研究设计的三大要素是：防止偏性、控制混杂、明确时相关系。之所以说“细思量”，是要建立在常识的基础上，静下心来，思考研究过程的各个环节，而不是生搬硬套一些术语和概念。

有这么一个故事，父亲丢了块表，他在房间里边抱怨边翻腾，四处寻找，可半天也找不到。等他出去了，儿子悄悄进屋，不一会儿就找到了表。父亲问：怎么找到的？儿子说：我就安静的坐着，一会就能听到滴答滴答的声音，表就找到了。只要我们静下心来，聆听内心的声音，就能从偏性、混杂和时相关系三方面，捕捉研究过程中各个环节可能出现的问题，从而加以控制。

流行病学研究是一门艺术，要在实践中学习、提高。在实践中本着实事求是的态度，具体问题具体分析，既要按设计方案与操作流程严格执行，又要保持一定的灵活性，过于死板与过于灵活都是不可取的。结合实践，阅读文献，就能捕捉他人在研究设计和实施过程中的闪光之处。譬如，一项随机化对照的临床试验（RCT），对比老年患者远端桡骨骨折外科手术与闭合复位的临床效果，在实施过程中，接受随机分组的只有 166 人，另有 134 人不愿意接受随机分组，如严格按 RCT 执行，这 134 人就要丢弃，而研究者考虑到招募研究对象的难度，对这 134 人按病人自己的选择入组治疗、观察，形成了一个平行的观察性队列研究。原来的一个研究分成了两部分，RCT 部分保证了随机化控制混杂和盲法控制偏性，观察性研究部分贴近临床场景，既可与 RCT 部分的结果相互印证，又提高了结果的普适性。

参考文献

- [1] L-G Spector, Brown M-B, Wantman E, et al. Association of In Vitro Fertilization With Childhood Cancer in the United States[J]. JAMA Pediatr, 2019, 173(6): e190392.
- [2] Susan-A Korrick, Chen Changzhong, Damokosh Andrew-I, et al. Association of DDT with Spontaneous Abortion: A Case-Control Study[J]. Annals of Epidemiology, 2001, 11(7): 491-496.
- [3] D Saleheen, Scott R, Javad S, et al. Association of HDL cholesterol efflux capacity with incident coronary heart disease events: a prospective case-control study[J]. Lancet Diabetes Endocrinol, 2015, 3(7): 507-513.