

横断面研究——那些“断”不清的相关与因果

作者：李楠 赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/1409.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

横断面研究

又称横断面调查，因为所获得的描述性资料是在某一时点或在一个较短时间区间内收集的，所以它客观地反映了这一时点的疾病分布以及人们的某些特征与疾病之间的关联。由于所收集的资料是调查当时所得到的现况资料，故又称现况研究或现况调查(prevalence survey);又因横断面研究所用的指标主要是患病率，又称患病率调查。

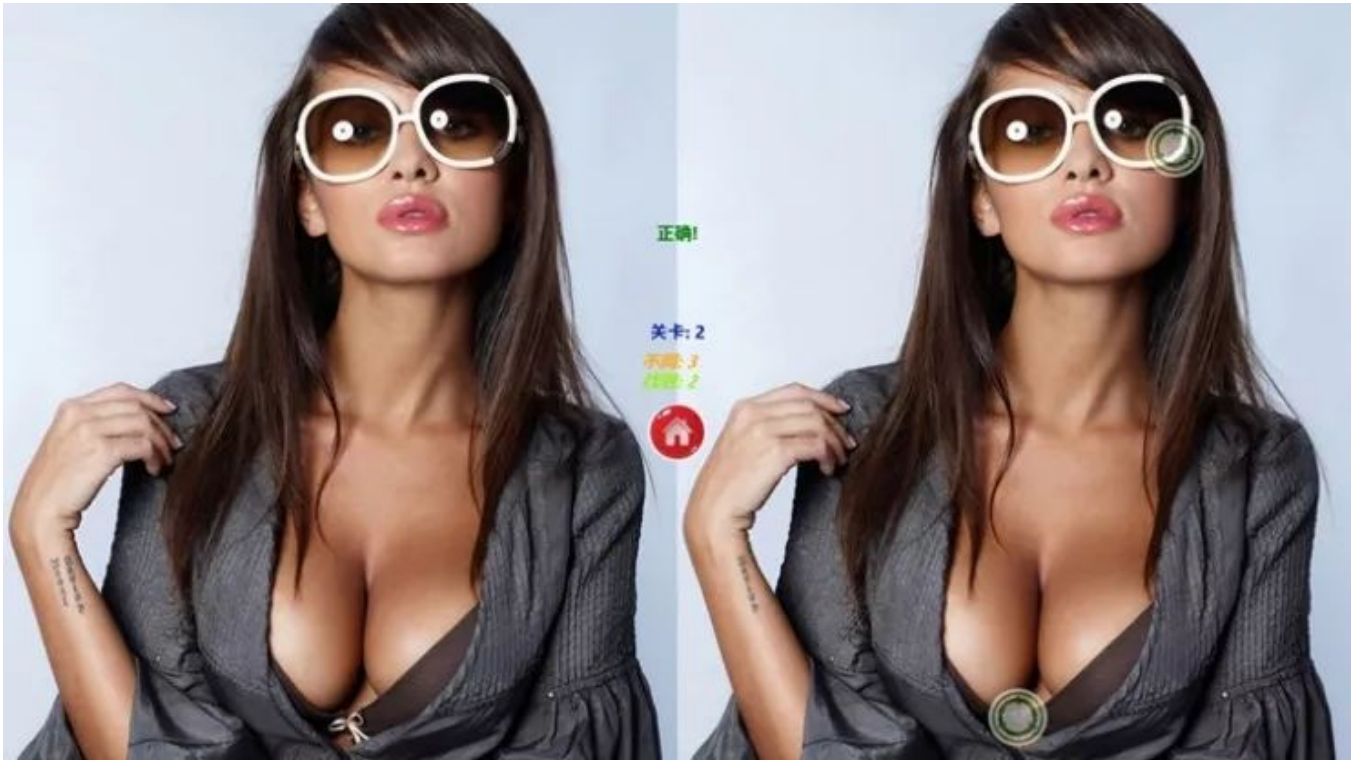
以上是流行病学对于横断面研究的定义。当“横断面”大神被请到临床研究里面时，很多情况下我们对他的期望有不仅仅是得到一个患病率，我们还想用它去探索一些潜在的关联，从而为后续的因果推断找到方向。

比如，我们可以把某类疾病患者(比如首次诊断为肺癌患者)的门诊就诊记录看做是一个点，那么所有首次诊断肺Ca患者的门诊资料就可以看做是一个横断面资料了。如果我们想看看首诊肺Ca患者中小细胞肺癌的百分比，那么像上述横断面研究的定义一样描述一下就好了。最终的结论可能是XXX医院内(如果是区域数据，就是XX地区)门诊首次发现肺癌患者中小细胞肺癌占20%(17%-23%)。

我们都知道小细胞肺癌的恶性度较高，我们如果把肺癌分为恶性度相对较高、较低两大类，那么您一定会想我们既然有了肺癌患者，那么没道理不看看哪些因素和肺癌的恶性度之间有关系吧。问题来了，这真的可行么?我们得到的结论真的可信么?



大家一起来找茬，这么做对不对——管他三七二十一，来个多因素分析搞定。



这是我们看各种论文时，经常见到的套路。很多研究者可能觉得，既然分出来恶性度高低了，那我们先简单比比两类患者在同时收集的其他指标上，哪些指标可能有区别。然后选出一部分，结合工人的潜在病因，一起建立了多因素回归模型(比如Logistic回归模型)，然后等着看结果就行了。

问题来了：作为自变量考虑的这些指标，真的都是影响良恶性的“原因”么？

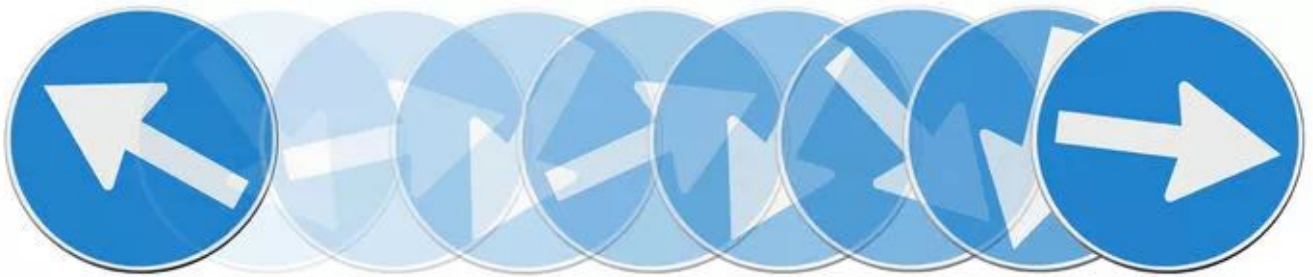
我们在以前很多文章中都介绍过，进行多因素分析的时候，因果关系是否明确是非常重要的。如果把因变量Y(也就是例子里恶性度高)的下游指标(高恶性度带来的激发改变，比如组织坏死引起的指标等)当做原因，把他们作为自变量X，就很可能导致模型本身的错误，当然，任何其他的自变量的结论也都会变得不可靠。(这涉及到流行病学因果推断相关的理论，可以搜索“因果”或“因果图”看对应文章)

让人苦恼的是，在横断面研究中，所有变量都是同事获得的。这样看起来，我们似乎没办法确定谁是Y的“原因”，谁是Y的“结果”。看起来确实没什么建立回归模型的依据，因此一些激进的审稿专家会直接建议：“横断面研究不应建立多因素回归模型，也无法支持病因学研究”



我们当然是不甘心的，如果不是肺癌的良恶性，而是其他更罕见、需要更艰苦的数据收集才能获得的横断面资料，如果我们放弃这个机会重新设计一个可以用于明确病因的研究，可能要在过几十年才能得到探索性的结论了。因此有时难免要硬着头皮做一做分析。这时候需要注意的是：

基于横断面研究的关联分析仅仅是探索!



千万不要说我们发现与结局事件存在关联的因素是“影响因素”、“危险因素”、“潜在病因”甚至是“病因”。我们能说的仅仅是关联，当然在讨论部分，适当的做一些合理推测还行，但结论里面只能用“关联”来表述。

建立模型的时候千万别走“描述→单因素分析筛选变量→建立多因素回归模型”这样的分析套路。



套路玩得深
谁把谁当真

毕竟此时我们找到的任何组间差异都不一定是“上游变量”，因此我们可以进行单因素分析，但是模型的建立本身应该是一个独立过程。至于哪些变量能够进入模型，我想可以考虑两类：

- a. 已有文献报道的病因(结局的危险因素)，同时这些病因指标应该是稳定的，即不会随结局的发生与否而改变。
- b. 从时间上，某一个因素必然是发生在结局之前的，且在相当长时间内一直稳定不变，比如患者的基因、家族史、居住地。当然判断这一指标结果的时候需要排除选择偏倚，比如家里有人得过类似疾病(发生过类似结局)的患者可能更容易来就诊，也会带来这一指标和结局直接存在关联，这一关联有可能仅仅是因为更容易就诊而导致的偏倚。因此当我们纳入这一条里的指标时，要仔细考虑和衡量潜在的偏倚风险。如果不存在上述问题，那么我们可以考虑将它纳入模型。

如果能力允许，建议结合已发表文献和研究者的数据构建因果图，并基于因果图开展后续分析。(具体可搜索文章“因果图”)

小结一下，横断面研究用于探讨可能的因果关系时要注意：

- 1、各种分析仅仅是探索!探索!!探索!!!不能下结论。
- 2、万一要进行多因素分析的探索，已有证据、时间顺序、指标的稳定性，模型中这三位大神一定要同时附体。
- 2、横断面资料是各种偏倚的集散地，任何结论都可能是假的，需要进一步研究。
- 3、后续的临床研究十分必要，但不少情况下进一步临床研究探讨病因的可行性极差，此时只能向基础研究求助了。此时，临床向基础的转化几乎成了唯一选项。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发