

再谈临床研究中的匹配

作者：曾琳，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/5267.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！

再谈临床研究中的匹配。临床研究最让人头疼或最有意思的地方在于研究对象也就是患者之间的差异，在统计上我们称之为变异

。由于这些变异的存在，我们常常很难回答吃白加黑感冒3天就好的患者和吃泰诺感冒6天才好的患者的疗效差

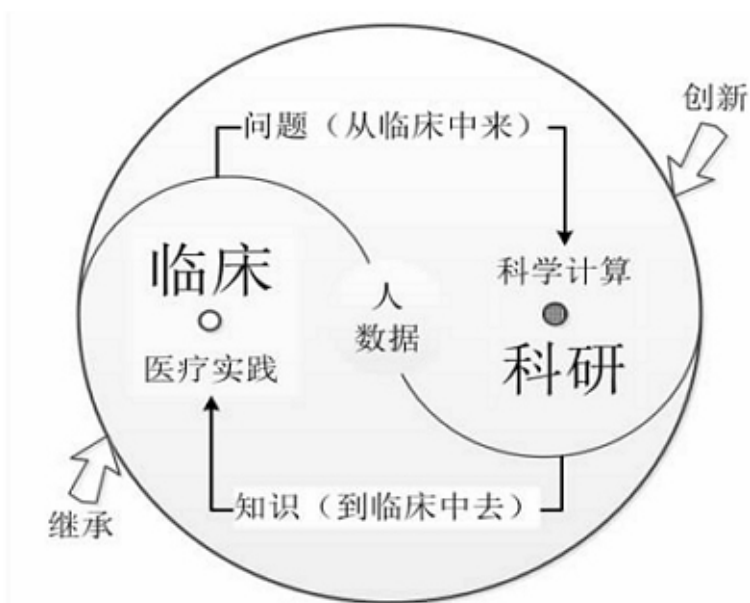
别是药物的作用还是个体差

异导致的。在临床研究设计里，可以通过随机化分配研究对象

来得到两组或多组各方面特征(包括人口学特征、健康或疾病状态、生活习惯等)很相近的研究对象，那么研究结果的组间差异

我们就可以归因到研究因素导致的。随机化

其实是一个简单、粗暴、又行之有效的办法。可现实研究中我们往往难以采用RCT设计，这时，混杂、组间均衡性等等问题又在研究设计时涌上心头，困扰着我们。这个时候，我们想到的尚方宝剑就是两个字——匹配。



在研究设计阶段，通过匹配选择合适的研究对象

，得到和病例组/试验组在某些特征上齐同的对照来提高组间均衡性，或者说来获得可比性好的对照，是最为常见的方法。但是，我们都知道，寻找各方面特征齐同匹配的对照并不是容易的事情，尤其是希望匹配的因素过多的时候。比如我们想比较两种治疗方案的疗效差异时，由于考虑年龄、性别、起病诱因、疾病的阶段、既往病史等等都可能会影响疗效，所以我们希望把这些因

素都一一匹配，这时我们会发现，别说1:4寻找对照了，就连1:1寻找对照都不一定能找到可以和试验组匹配的。

匹配可能会造成研究效率的降低，也可能会低估研究因素的作用大小，我们把这一现象称为匹配过度(over matching)。举个例子，我想研究肥胖和体重正常者以往日常运动量的差异，在研究中我希望控制除了运动量以外的其它可能影响肥胖的因素，比如起始的BMI，饮食摄入，久坐等，于是我匹配了这些因素，可是我竟然发现运动量和肥胖竟然没有关联了，这时为什么呢？哦，经过深入的分析，我发现，爱运动的人，往往不会采用久坐的生活或工作方式，常常会坐下工作/娱乐一段时间后就有意识的起来走走。而我在研究中对久坐进行了匹配，也就是可能把运动-肥胖因果链条中的某个环节——久坐在病例组和对照组中进行了匹配，那么我们就有可能看不到运动和肥胖在统计学上的关联了。

在匹配因素的选择上，我们还要考虑增加匹配因素会对我们研究带来的难度。如果匹配的因素很多，那么对于某些病例/试验组对象，我们可能难以找到合适的对照，那么我们马上要面临的就是要不要舍弃这些病例/试验组对象的问题了。在实际研究中，研究者往往是不愿意随意舍弃研究对象的。而从研究设计的角度，因为找不到合适的对照而舍弃某些病例，虽说是完成了匹配的过程，却可能因为这个舍弃的动作而引入了选择偏倚。这也是得不偿失的。

因此，匹配还是不匹配，这真是值得考虑的问题。一定要具体情况、具体分析。如果想不太清楚，建议即使选择了匹配，匹配因素不要太多。大家想想，即使我们不匹配，我们在分析阶段也可以采用分层分析、多因素分析等方法来调整、控制混杂。但是如果真的匹配过度了，还真是回天乏术了。

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](http://www.iikx.com)转发