

临床研究系列讲座

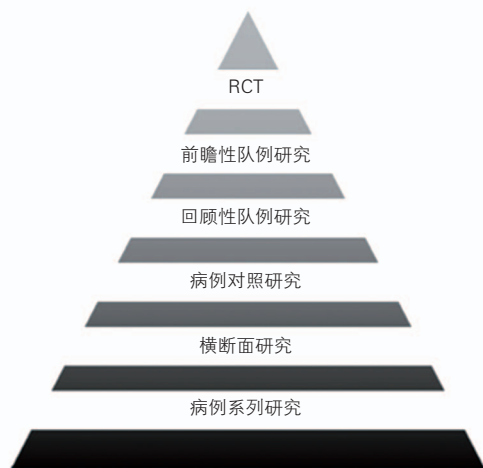
代表性与可比性——决定临床研究价值的核心原则

李希

怎样才能称得上是好的临床研究? 一项好的临床研究应该满足两大条件:(1) 提出了一项重要的问题。(2) 给出了一个可靠的答案。而科学可行的设计是确保上述两条得以实现的基本条件。在临床研究的设计过程中, 需要解决目的遴选、技术路线、对象标准、数据定义、分析计划, 以及质控策略和伦理知情等方方面面的问题。可以说, 相比于大规模多中心临床研究, 小研究在运行过程中所投入的资源和时间肯定要少得多, 但从设计角度来说, 二者的复杂程度并没有明显差别。

1 从临床研究金字塔到随机对照临床试验崇拜误区

谈到研究设计, 相信有一张图是大家都比较熟悉的(图1)——我们称之为研究证据级别的金字塔。其中, 高大上的随机对照临床试验(randomized control trial, RCT)高居塔尖, 备受临床指南等的推崇。这张图在直观展现不同类型研究的证据强弱的同时, 也被很多研究者当做研究设计类型优劣的金字塔, 进而导致了“RCT 崇拜”等一些片面的观念——RCT 成为了解决所有临床研究问题的终极设计和最佳方案。实际上, 只有适合问题的答案才是好答案。要选择最佳的设计方案, 首先要辨别清楚通过研究想解决的临床问题究竟有哪些类型。



注: RCT: 随机对照临床试验

图1 临床研究证据级别的“金字塔”

2 临床研究目的的类型

临床研究要解决的问题, 也就是研究的目的, 可以简单地分为两种类型。第一种称为“情况描述”: 就是要将一群对

象中的某一个临床特征的分布情况描述清楚, 比如说明全国某年急性心肌梗死住院患者的 30 天的平均死亡率是多少, 或者这些患者入院时的血压水平如何。另一种称为“差异比较”或“关联分析”, 比如说明急性心肌梗死的治疗中, 合并糖尿病的患者是否院内事件率更高, 或者长期服用他汀类药物的患者是否预后较好, 就是要通过比较说明不同类型的对象之间的某个临床特征分布有没有差异, 也可以被认为是评价糖尿病或他汀类药物治疗这样的因素与患者预后这一临床特征之间是否存在关联——这是临床研究文章中对同一种情况的两种常见的解释方法。

了解这一分类后, 如果再看一下刚才提到临床研究设计类型的金字塔, 就会只有在“差异比较”或“关联分析”研究, 即针对不同组别对象进行比较分析时(如确定疾病危险因素的强弱, 还是评价治疗手段的优劣), 金字塔中“高端”设计的优势才能显现出来。那么针对不同类型的研究目的来说, 究竟什么样的核心原则会直接影响到结果的可靠性, 进而决定研究的价值呢?

3 情况描述研究中的代表性原则

在试图描述一类患者的分布情况时, 研究所描述的范围越大, 研究本身的信息量也就越大。比如, 一项全国性的调查, 其带给读者的价值要远大于仅覆盖一个地区或单独一家医院的调查。然而, 实际研究中能纳入的对象样本量是有限的。研究样本量越大, 需要投入的资源就越多, 实施的难度也越大。因此, 这类研究的设计中通常需要引入抽样的方法, 以有限的对象(即“样本”)中观察到的结果去反映更大范围群体(即“总体”)的特征。举例来说, 同样是纳入 1 000 例患者的两项研究, 如果其中一个能代表全国急性心肌梗死患者的治疗模式, 另一个只能说明某家医院这类患者的治疗模式。相比之下, 前者当然更容易得到杂志和读者的青睐。所以说, 在情况描述研究中, 研究对象所能代表的群体范围越大, 研究的价值和意义也就越大。因此, 确保“代表性(representativeness)”是研究者在这类研究中所要把握的核心原则。

代表性并不是简单的指研究对象来自于多大的范围, 更重要的是向读者说明研究中样本的结果能够“无偏性”地推论到总体的特征。如果在从总体中纳入样本的过程中, 研究者主观上想选谁选谁, 甚至故意去选择年轻、并发症少的患者(这些患者往往顺应性好, 调查难度小), 所得的样本就不一定能代表总体的实际分布情况——这种“选择性(selectiveness)”可以算得上是代表性最顽固的敌人, 会直接

影响到读者对研究结果的认可程度,也是研究者要尽可能避免的。

要实现研究的代表性,就需要说服读者,研究样本的纳入中没有掺杂主观故意的选择性。从设计角度来说,有三种方法可供选择(图 2)。

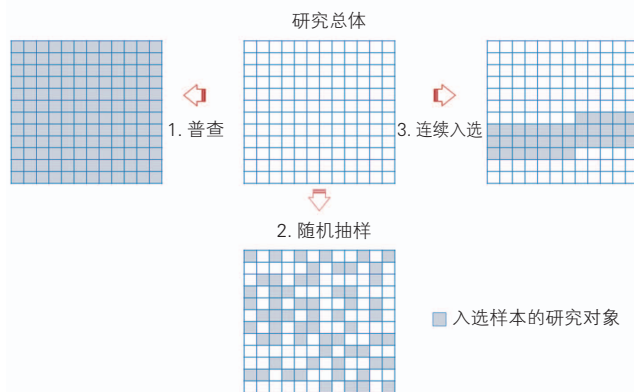


图 2 普查、随机抽样和连续入选：三种实现代表性的设计方法

第一种是普查,就是把想代表的总体中的每一个对象都纳入研究——这个过程中没有任何的选择性,结果无疑最能够代表总体的特征。这样的方法看起来最直接,但是实施起来却最困难。因为如果要做一个覆盖范围较大(如全国)的普查研究,需要投入的人力、物力、财力和时间资源对于单个研究来说都是难以承受的。不仅在临床研究领域,在其他领域也是如此。即使对政府而言,也只有每 10 年才会组织一次全国人口普查,每 5 年才会开展 1 次全国经济普查。

第二种是随机抽样,就是对总体中的每一个对象都根据随机的方法决定其是否纳入研究。“随机”,顾名思义是“根据概率(机会)而定”,而不受研究者意愿的影响。这也就避免了主观“选择性”的干扰。随机抽样的具体设计方式有很多,包括简单抽样、系统抽样等可以在后续的研究设计专题中具体介绍。作为研究者或读者,更重要的是要鉴别一个抽样研究的设计究竟是不是真正的随机抽样或具有代表性。我们经常会看到杂志中一些研究在方法部分自称是基于代表性样本,但实际上并没有真正采用随机抽样的方法,所以只能称为“随意抽样”,也就是“根据研究者主观意愿而定”的样本选择。

第三种是连续入选,就是选择一个有限的时间段,在其中将每一例符合条件的患者都入选到研究样本中,以代表更长时间范围内总体的情况。这也可以避免主观选择性的影响。随着前瞻性临床注册登记研究的流行,这种方法的热度也变得越来越。但这种方法也有其特殊的局限性。比如某项单中心研究受经费和人手所限,选择 6~8 个月连续入选了所有合格对象,希望代表医院全年的急性心肌梗死患者特征和治疗情况——对于再灌注治疗、抗血小板药物的使用率等指标来说,6~8 个月研究样本中的结果有比较大的把握能够反映全年总体的情况,但对于患者入院收缩压水平等明显受季节影响的指标来说,6~8 个月这个时间段的调查结果相比于全年平均水平就可能出现偏差,而不再具有代表性。

4 关联分析研究中的可比性原则

在关联分析研究中,目的不仅仅是比较不同的因素(如治疗)的对象之间临床结局(如预后)差别,而是希望通过这样的差别去推断这种因素是否与结局之间存在独立的关联,进而做出更有把握的因果推论(如某种治疗是否可以改善预后)。这时,就需要在关联分析的比较过程中排除其他混杂因素的影响。比如评价他汀类药物降低心血管事件再发风险的疗效,需要确保服药的患者和不服药的患者之间年龄、性别、合并症,甚至收入水平基本一致,否则,如果服药组的相对要年轻一些,或者合并症相对要少一些,那么即使观察到两组患者之间的预后差异,也很难判断这种差异是由治疗本身造成的,还是年龄等其他因素的影响导致的。所以说,关联分析研究中,研究组间越多的其他因素达到均衡可比,研究推论因果关系的可靠性就越高。因此,确保“可比性(comparability)”是我们在这类研究中所要把握的核心原则。

要实现研究的可比性,从设计角度来说,有两种方法可以选择(图 3)。

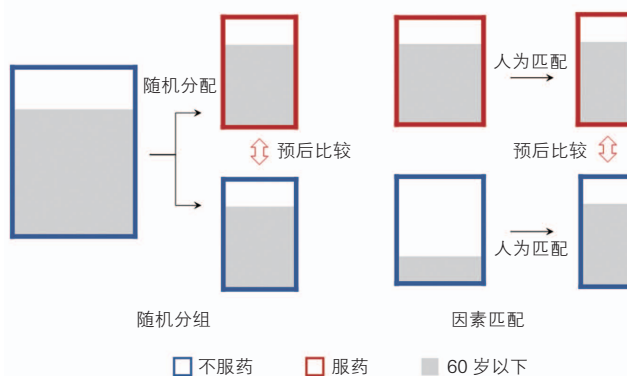


图 3 随机分组和因素匹配：两种实现可比性的设计方法

第一种是随机分组,也是在 RCT 中采用的方法。以药物疗效评价为例,在同一类患者当中,通过随机的方法决定每一例对象是进入治疗组还是对照组,那么在最终得到的两组对象之间,年龄、性别、合并症等因素理论上来说都应该是均衡可比的。这就为两组分别分配治疗方案后的预后比较奠定了基础。

另一种方法是因素匹配,常见于队列研究或病例对照研究这样观察性的临床研究。还是以药物疗效评价为例,在观察性研究中,患者服药或不服药不是由研究决定的,因此在两类患者之间,很可能本身就存在着年龄等因素的差异——例如在临床实践中对高龄患者的用药通常更加保守一些,所以服药组的平均年龄往往比不服药组要低一些。为了排除这种影响,在两组患者入选时,可以从不服药的患者中尽量多入选一些年轻的对象,通过这样的人为措施确保研究纳入的两组患者在年龄上均衡可比。这样当研究观察到两组之间预后存在的差异时,就可以排除年龄这个潜在的混杂因素造成的影响,使得研究对药物疗效的评价更有把握。

当然,相比于 RCT 的随机分组,因素匹配的方法在确保可比性方面的局限性也非常明显。尽管在观察性研究中可以人为努力去匹配一些混杂因素,但能控制的因素总数是有限的——想匹配的因素越多,就给患者入选造成更大的困难。

而且对于一些在研究之前根本没有了解的患者特征,人为匹配也就自然无从谈起。而 RCT 就不受这样的限制,不论存在多少潜在的混杂因素,不论之前是否收集了相关信息,从原理上来说,随机分组都可以实现其均衡可比。这也就是为什么 RCT 在关联分析研究中能超越观察性设计,在金字塔尖傲视同侪。

5 在研究的全过程中确保原则

尽管上面谈了很多在研究设计中确保代表性或可比性的方法,但这对两大原则的把握来说只是一个开始。读者对研究价值的判断不是针对研究的设计方案,而是基于研究运行完成后所得的最终结果。

研究运行过程中,很多环节都可能影响到最终结果的代表性或可比性。比如说,一项问卷调查设计了科学的随机抽样方案,确保了所选患者样本对总体的良好代表性,但在实际运行的对象入选过程当中,却出现了大面积的拒绝应答,最终获得的研究结果就不再能代表设计中目标总体的情况。再比如,在一项随访研究中,尽管前期设计和患者入选都贯彻了代表性原则,但在随访过程中的失访率较高(失访者往往是预后较差、结局事件高发的研究对象),研究最终获得

的预后数据只来自于完成随访的患者,其代表性就大打折扣。除了代表性以外,可比性也面临同样的问题。比如一项临床试验尽管设计了严格的随访分组方案,确保了在研究开始是干预组和对照组对象各方面特征的均衡可比,但在随访过程中,两组失访率存在差异显著(失访者往往是预后较差、年龄较大或知识水平较低的研究对象)。在这种情况下,如果结局的分析需要基于末次随访的调查数据,那么两组之间的比较就不可避免地要受到其他混杂因素的影响,研究对药物疗效的评价结论也就不再可靠了。

因此,要紧扣两大原则提升研究价值,不能仅停留在方案设计阶段,还应该贯穿研究运行的始终。

本文的内容可以总结为三点:(1)临床研究中设计类型的选择要适应研究目的,RCT 并不一定是解决所有问题的最佳方案。(2)针对不同类型的研究目的,要在设计中坚持把代表性或可比性的原则。(3)除了研究设计中把握这些原则,还应当贯穿研究运行的各个环节。这才能从根本上确保临床研究的价值。

(收稿日期:2017-07-19)

(编辑:宁田海)

读者·作者·编者

欢迎关注《中国循环杂志》微信和微博

《中国循环杂志》是一本有着近 30 年历史的,以专业、严谨、务实态度,刊登高质量、高水准学术论文的学术期刊。

《中国循环杂志》是一个标记中国心血管病研究进展的符号,一个让医者捕捉更多心血管专业信息的平台。

与时俱进,也是一本有品质的学术期刊所必备的特征之一。随着互联网在媒介传播领域的渗透,当前已是一个信息大爆炸的时代,微信和微博等各种新颖、快速、高效的信息传播手段席卷而来。为了不让发表在《中国循环杂志》上的优秀论文束之高阁、尘封暗室,让更多的医生了解全球心血管领域的最新进展,杂志的官方新浪微博(中国循环杂志官方微博 V)、官方微信(微信号:zhongguoxunhuanzazhi)也陆续上线。

学术应该是透亮的,像阳光那样,照亮大多数人的生活。希望广大医生和科研工作者关注《中国循环杂志》的微博和微信,共同努力推动中国的学术发展,改善公众健康。



官方微信



官方微博

《中国循环杂志》编辑部