

临床研究系列讲座

临床试验设计: 随机对照

李静

随机对照临床试验对于评价治疗措施的疗效与安全性是金标准,而疗效和安全性是临床医生最关心的问题。因此,临床研究者应学习如何科学设计随机对照临床试验。此外,在循证医学时代,正确解读临床试验结果应该成为临床医生的一项基本功,就像医生要会看化验单、读 X 线胸片一样。正确解读临床试验也离不开对科学设计原则的了解。

一项好的临床试验必须满足两个条件,一是回答重要的临床问题,能够指导临床实践;二是结论科学可靠,有效控制误差。因此好的临床试验必须是起源于临床实践,依托流行病学的方法。科学的设计目的就是要合理控制误差,保证研究结果的准确可靠。误差分为系统误差和偶然误差,系统误差是指结局指标的差异由其他因素造成,而非由待评价的治疗所致,又称偏倚。系统误差是观察性研究常见的问题,可能高估或者低估疗效,但很难确定其方向和程度。随机误差是机会所致,见于观察性研究和临床试验,但可以量化。只有当真实疗效明显大于误差产生的影响时,研究结论才是可靠的。随机对照与合理的样本量分别能够有效地控制系统误差和随机误差,是临床试验研究设计的重要原则。

1 随机对照

1.1 随机分组

随机对照是消除系统误差的唯一途径,科学的随机分组方法能够确保治疗组和对照组具有真正的可比性。随机是提供一个时间点,从此开始研究各组分开,受试者分到各组的机会相等,而且不能预知分组的情况。既然是通过随机的分组,理论上能够确保所有的因素在各组之间都是均衡分布的。这一点较观察性研究具有无可比拟的优越性。因为观察性研究虽然可以通过分层、匹配等方法使已知对结局指标有影响的因素在各组之间达到平衡,但完全无法平衡未知的影响因素。根据随机的定义,我们可以判断一个研究是否真正做到了随机。有些研究号称“随机”分组,但实际上并不能满足随机的定义,例如,根据患者就诊日期的奇偶数分组、根据生日的奇偶数分组、根据患者参加研究顺序的奇偶数分组,就都不是真正的随机。因为这些情况下,研究者都可以在随机前预知受试者将分到哪个组别,如果研究者带有倾向性,希望受试者分到某个特定的组别,他就有可能采取措施干预受试者的分组,例如如果是根据研究顺序分组,研究者可以让受试者等到下一个受试者之后再随机入组,就会被分到另一个组别。一旦研究者可以干预分组,那就破坏了随机,无法保证各组间的均衡。

一项临床试验的结果发表时,一定要报告各组受试者的基线特征比较,通常是用表来展示。如果各组之间某些基线特征存在明显差异,往往是由于随机的设计存在缺陷或执行过程中出了问题。这会让人质疑结论的可靠性。

随机分组是双盲设计的前提条件。1948 年现代流行病学的鼻祖英国人 Austin Bradford Hill 开展了世界上第一项随机对照临床试验,评价链霉素治疗肺结核的疗效。这项试验大获成功,终结了肺结核作为不治之症的权势。但这项试验更重要影响在于确立了随机对照临床试验的设计原理,他也首次提出双盲的原则,以及评价指标应客观简单。

1.2 意向治疗分析

为了保证随机对照不被破坏,临床试验的分析要遵循意向治疗分析原则(intention to treat, ITT)。ITT 原则是指分析时应包含所有被随机分组的受试者,而且按照随机分配组别进行分析,不考虑受试者是否符合入选标准,是否违背研究方案,也不考虑受试者实际接受了何种治疗。举个极端的例子,如果一名受试者随机后,当天就因意外死亡了,一片研究药物都没吃,按照 ITT 原则,也应该纳入分析,而不能被排除。同样,所有失访的受试者也应纳入分析。正是出于这条原则,目前临床试验结果报告所必须遵循的 CONSORT 原则要求以流程图(图 1)的方式,清楚报告受试者随机及随访的过程,具体到每一步骤、每一组别和每一种情况所对应的受试者人数^[1]。有了这些信息,就能够清楚判断作者是否遵循了 ITT 原则。

举 2 个例子,SPARCLE 研究^[2]评价阿托伐他汀预防脑卒中的疗效,2 365 例受试者分到治疗组,2 366 例分到安慰剂组,治疗组有 78 例撤回知情同意,其中 63 例生存状态不明,此外,还有 15 例失访。对照组与之相似。最后分析时,治疗组的 2 365 例和安慰剂组的 2 366 例受试者全部纳入疗效分析和安全性分析,一个都没少。再看另外一个例子,是评价美托洛尔治疗心衰疗效的 MERIT-HF 研究^[3],随机分组后,治疗组有 1 995 例受试者,安慰剂组有 2 001 例。随访做得非常好,没有失访的。但研究结束时,治疗组和安慰剂组分别有 228 例和 242 例没有接受研究治疗的。最后也是所有受试者都纳入分析。

为什么设立 ITT 原则?分析时为什么不能排除停药的受试者?为什么不能排除失访的受试者?首先从统计学角度而言,ITT 分析是对随机的保证,研究随机分组的时候,分开的各组是平衡的,但随机分组之后各个组之间出现任何

作者单位:100037 北京市,中国医学科学院 北京协和医学院 国家心血管病中心 国家心血管疾病临床医学研究中心 心血管疾病国家重点实验室 中国牛津国际医学研究中心

作者简介:李静 教授 博士研究生导师 主要从事心血管领域大规模随机对照临床试验和医疗结果评价研究 Email:jing.li@fwoxford.org
中图分类号:R54 文献标识码:C 文章编号:1000-3614 (2017) 11-1126-02 doi:10.3969/j.issn.1000-3614.2017.11.022

差异,都有可能是由于各组采取的干预措施不同而导致的,如果排除某些受试者,就会破坏原本随机分组时建立的平衡,导致偏倚。按照随机分配的组别,对所有随机的受试者进行分析,就是为了保证随机不被破坏,是对随机最大程度的尊重。其次,从临床角度而言,停药往往是由于受试者不耐受研究治疗,如果排除停药的受试者,就有可能掩盖研究治疗的不良反应。与之类似,失访有可能是由于死亡,或者由于不耐受研究治疗而不再配合随访,这也是与研究治疗的疗效或安全性直接相关的,如果排除此类患者,势必抹杀了研究治疗与对照之间的差异。

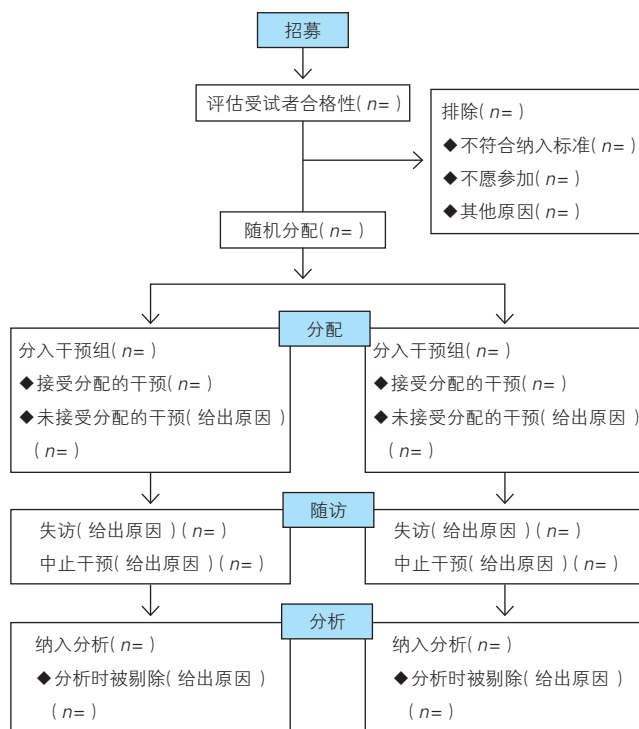


图1 CONSORT 流程图

因此, ITT 分析是随机对照临床试验的首要分析方法,在此基础上,可以考虑对符合方案的受试者进一步分析,补充安全性方面的信息。采用 ITT 分析,是为了避免系统误差,但不依从研究方案和失访者都会导致研究的统计学把握度大大降低,使随机误差增加。因此临床试验实施管理的一个重点就是尽量避免失访和不依从方案的情况发生。如何避免呢?首先是从研究设计的角度考虑,其次是加强随访的管理,与受试者建立良好的沟通和彼此的信任。

2 合理的样本量

合理的样本量是减少随机误差的有效方法。科学的设计包含对样本量的合理估计。样本量不足是临床研究中常见的问题,将导致研究无法给出明确结论,说不清楚到底是研究治疗无效,还是假阴性。相反,样本量也不是“韩信点兵,多多益善”,因为样本量增加意味着人力、物力和时

间成本的增加,以及纳入更多的中心,质量控制的难度增加、随访质量的下降等。不适当的样本量也是不伦理的,过大的样本量会延缓研究结果的产生和发布,延误有效治疗在临床实践中的广泛应用。过小的样本量不足以得出科学结论,会影响受试者的获益。

计算样本量的公式比较复杂,可使用专门的软件,按照提示输入相应的参数,软件就能给出计算结果。通常影响样本量的参数是四个,其中两个是统计学参数,包括统计学显著度和把握度。如果用 α 和 β 分别表示 I 型错误率(假阳性率)和 II 型错误率(假阴性率),则显著度是 α 值,统一规定为 0.05。把握度是 $1-\beta$ 值,由研究者来决定,越大越好,但把握度越大就意味着需要的样本量越大。还有两个参数与研究要回答的问题本身有关。一个是对照组终点事件发生的频率,另一个是预期的疗效,也就是治疗组与对照之间的差异。确定对照组终点事件发生的频率是根据既往研究或预试验的结果。预期的疗效只能是估计值,因为如果已经知晓效果就不需要开展这项临床试验了。这几项参数中,疗效的不确定性最大,其估计的准确性对于样本量计算的影响也最大。估计疗效的根据是既往同类研究结果或待评价治疗的作用机制及流行病学数据^[4]。

根据公式算出来的只是理论上所需的样本量,临床试验实施过程也会影响到最终实际需要的样本量。研究实施中未必全部受试者都能依从研究方案的要求接受研究治疗,依从者在所有受试者中所占的比例是依从率。实际所需的样本量是理论样本量除以依从率的平方。因此,如果依从率是 70%,实际所需样本量就是理论样本量的 2 倍。这一点再次说明临床试验的设计必须考虑研究实施的合理性,而实施质量对于保证科学设计得以实现是至关重要的^[5]。

总而言之,随机对照是临床试验设计的精髓,因为其能有效控制系统误差,从而使得随机对照临床试验成为疗效评价的金标准。为了保证随机的效果不被破坏,临床试验的首要分析是 ITT 分析。预先计算合理的样本量能够减少偶然误差。

参考文献

- [1] CONSORT Group. CONSORT 2010[EB/OL]. <http://www.consort-statement.org/>, 2017-09-22.
- [2] Amarenco P, Bogousslavsky J, Callahan 3rd A, et al. High-dose atorvastatin after stroke or transient ischemic attack. *N Engl J Med*, 2006, 355: 549-559.
- [3] Hjalmarson A, Goldstein S, Fagerberg B, et al. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in congestive heart failure (MERIT-HF). MERIT-HF Study Group. *JAMA*, 2000, 283: 1295-1302.
- [4] Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet (London, England)*, 2005, 365: 1348-1353.
- [5] 陈铮鸣, 邵永孚. 临床随机试验原理和方法. 中华肿瘤杂志, 1994, 16: 234-236.

(收稿日期: 2017-09-20)

(编辑: 宁田海)