

如何撰写高质量的流行病学研究论文

第四讲 随机平行对照试验报告规范——CONSORT 介绍

王波 詹思延

【导读】 随机对照试验为临床治疗和医疗卫生干预提供重要证据。对随机对照试验的严格评价有赖于充分、准确的报告。临床试验报告统一标准(CONSORT)清单和流程图是当前公认的、最权威的随机平行对照试验的报告规范。本文介绍该清单和流程图的内容,并对一些重要条目进行了详细说明。

【关键词】 临床试验报告统一标准;随机对照试验;清单;流程图

随机对照试验是评估干预效果的最佳研究,但低质量的临床试验会错误地估计疗效。为了使读者能够判断试验的内外部真实性(或叫有效性),随机对照试验报告应当准确、完整地描述关于试验设计、实施、分析和可推广性的信息。

1. 当前随机对照试验报告中存在的问题:既往许多研究表明,随机对照试验报告存在缺陷^[1,2]。1979 和 1980 年发表在 4 个主流杂志上的 67 篇随机对照试验报告里^[3],只有 30% 报告了结局评价是否采用盲法的信息。发表在 1985 年的 45 篇随机对照试验报告里^[4],只有 27% 对主要结局进行了定义。发表在 1990 年的 37 篇结果阴性的随机对照试验报告里^[5],只有 43% 提到了样本大小的计算。在 119 篇声称进行了意向性分析(intention-to-treat, ITT)的报告里^[6],有 15 篇(13%)排除了病例,或者并没有分析所有被分配到各组的病例。在妇产科期刊上的 206 篇随机对照试验报告中^[7],至少 5% 并不是真正随机化的研究。这个估计是保守的,因为大多数报告并没有提供关于分配方法的足够信息。在评价选择性 5-羟色胺再摄取抑制剂(SSRI)作为抗抑郁一线治疗策略的有效性方面,1 篇对 122 个随机对照试验的综述发现只有一项研究(0.8%)充分报告了随机化^[8]。不充分、不准确的报告可能使得随机对照试验结果的解释变得困难。而且,这种不充分、不准确的报告近乎一种不道德的行为,尤其是在存在偏倚的结果受到错误信任的情况下。

2. CONSORT 声明的制定与内容:1995 年,为了改进随机对照试验报告质量,一个由临床试验学者、统计学家、流行病学家和生物医学编辑组成的国际小组制定了 CONSORT 声明,即临床试验报告的统一标准(consolidated standards of reporting trials)^[1,2]。1999 年,CONSORT 制定组织依据最新的关于偏倚产生的证据,对最初的 CONSORT 清单和流程图进行了修订。修订后的 CONSORT 声明包括一个由 22 个条目组成的清单(表 1)和一个流程图(图 1)。同时,为了对清单各个条目的含义及原理进行详细的解释,CONSORT 组织决定制定一份解释和说明性文件。这一文件包括了清单条目所基于的证据和原理。

3. CONSORT 清单中部分条目的解释与说明:为了帮助读者更好地理解和应用 CONSORT 清单,CONSORT 对清单中的 22 条进行了逐一的解释和说明^[9],以下选取其中一些条目进行解读。

条目 1 作者应当在报告的摘要里明确陈述研究对象被随机分配到比较组。同时强烈鼓励在报告的标题里使用“随机化”一词。当需要对报告进行总结时,鼓励使用结构化摘要。结构化摘要向读者提供了一系列与试验设计、实施和分析有关的标题;在各个标题下是标准化的信息。研究发现结构化摘要质量比传统的描述性摘要更高,它们可以使读者更容易找到信息。

条目 3 任何一个随机化对照试验都关注一个与一些人群相关的问题,研究者通过使用入选标准来限定这个感兴趣的人群。通常纳入标准与年龄、性别、临床诊断和其他伴随疾病有关;排除标准则用于保证患者的安全。应当明确定义入选标准。入选标准并不影响一个试验的内部真实性,却影响其外部真实性。

条目 8 作者应当提供足够的信息以便读者可以评价产生随机分配序列的方法和分组过程中产生偏倚的可能性。因此,仅仅报告“随机分组”、“随机化”而没有进一步的说明是不够的。作者应当指出序列产生的方法,如随机数字表或计算机产生的随机数字。一些研究尽管使用了“随机”一词,但却使用了非随机的方法,如轮流交替分组、按医院数字或出生日期分组等。如果研究者使用了这些方法,应当对此进行详细描述,并且不适合使用“随机”一词,即使是“类随机”(quasi-random)都未必恰当。

条目 9 当研究对象进入试验的时候,随机序列如何实施非常重要,这种情况下最理想的方法是使用分配隐藏^[7]。分配隐藏是一个防止分配结局被那些负责纳入研究对象的人预先知晓的严格过程。如果没有充分的分配隐藏,即使随机的分配序列也会被破坏。分配隐藏与盲法是不同的概念。分配隐藏的目的是防止选择偏倚^[10],它负责分配之前随机序列的保护,常常是可以成功实现的。而盲法的目的是防止实施和结局判定中的偏倚^[7],它负责保护的是分配发生后的随机序列,盲法并非总是可以实现的。许多良好的分配隐藏方法都整合了外部的参与。例如,使用药房或中心电话系统是两种常见的技术。如果外部的参与不可行,一个很好的方

作者单位:100083 北京大学医学部公共卫生学院流行病学与卫生统计学系

通讯作者:詹思延, Email: siyan-zhan@bjmu.edu.cn

法是使用编号的容器。干预措施被密封在按顺序编号的同样大小的容器里。使用连续编号的、不透明的、密封的信封也是一个很好的方法。研究者必须确保信封按照顺序打开,并且只有在研究对象的姓名和其他细节都被写在正确的信封上之后才能打开^[11]。

表1 报告平行组设计的随机对照试验时需要包括条目的清单

内容与主题	条目	描述
标题与摘要	1	研究对象是如何分配到各个干预组的(如,“随机分配”或“随机化”)
介绍		
背景	2	科学背景与原理的解释
方法		
研究对象	3	研究对象的人选标准,数据收集的机构和地点
干预	4	各组干预的详细内容以及何时、如何实施的
目标	5	设定的目标和假说
结局	6	明确定义主要和次要结局指标,如果可能,描述改进测量质量的方法(如多次测量,对测量者进行培训)
样本大小	7	样本量大小如何确定,如果可能,对中期分析和终止试验的条件进行解释
随机化		
序列的产生	8	产生随机分配序列的方法,包括任何限定情况(如,分组,分层)
分配隐藏	9	按照产生的序列进行随机分配的方法(如,编号的容器或中心电话),清楚阐明在分派干预之前序列是否被隐藏
实施	10	谁产生分配序列,谁登记研究对象,谁指派研究对象到相应的组
盲法	11	研究对象、实施干预者、评价结局者是否不知道分组情况?如果是,盲法是否成功要评价
统计学方法	12	比较各组主要结局的统计学方法;其他分析方法,如亚组分析和调整分析
结果		
研究对象的流动	13	各个阶段研究对象的流动情况(强烈推荐流程图);特别是报告各组接受随机分配、接受干预、完成试验、和进入分析的研究对象数量;描述实际研究偏离研究方案的程度及原因
研究对象的征集	14	征集研究对象和随访的日期范围
基线数据	15	各组的基线人口学特征和临床特征
分析的数量	16	纳入每个分析的各组研究对象的数量(分母),以及是否进行了ITT分析。如果可行,用绝对数的形式来表达结果(如10/20,而不是50%)
结局和估计	17	对每个主要和次要结局,报告每个组的综合结果,估计效应大小和精确度(如,95%可信区间)
辅助分析	18	报告进行的其他所有分析,包括亚组分析和调整分析,阐明哪些分析是预先设定的,哪些是探索性的,从而关注多重分析问题
不良反应事件	19	各个干预组所有重要的不良反应事件或副作用事件
讨论		
解释	20	结合研究假设、潜在偏倚或不精确的来源以及与分析、结局多重性有关的危险,对结果进行解释
可推广性	21	试验结果的可推广性(外部有效性)
证据总体	22	结合现有的证据,对结果进行全面的解释

条目 11 盲法是指患者、医生(有时也包括收集和分析数据的人)对所分配的干预措施未知,从而预防偏倚的产生。对患者的盲法是必要的,因为那些得知自己正在接受新的治疗措施的患者可能会产生正向的预期或增加焦虑感,而那些接受标准治疗的人则会感到被歧视或很安心。安慰剂与盲

法一起使用目的是为了减少与接受干预有关的非特异效应(安慰剂效应)所造成的偏倚。对患者和医生的盲法可以防止实施偏倚(performance bias)。当医生提供了其他的治疗措施或患者有偏好地去寻求其他治疗措施时,这种类型的偏倚就可能发生。对患者、医生和其他参与结局判定人员(如放射医生)的盲法可以减少检测偏倚(detection bias)的发生。对数据分析人员的盲法也可以预防偏倚,知晓干预分组状况可能影响分析策略和方法的选择^[12]。当结局为主观指标时,盲法尤其重要,例如评价疼痛程度。在报告里作者应当阐明^[13,14],谁处于盲态(如患者、医生、结局评价人员或数据分析人员),盲法的机制(如胶囊或药片)以及治疗特征的相似性(如外观,味道和服用方法)。

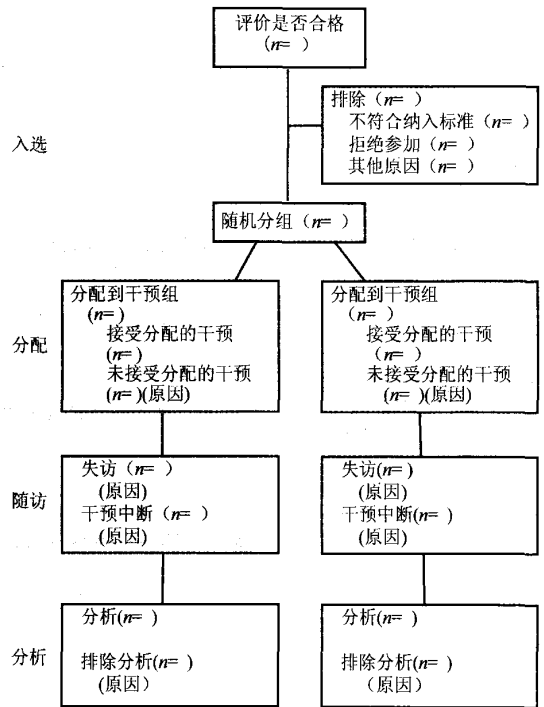


图1 随机对照试验各个阶段

(入选、干预分配、随访和分析)流程图

条目 12 在交互作用的检验方面,一个常用但并不合理的方法是通过比较不同亚组的 P 值来推断是否存在交互作用。如一个亚组 P 值显著而另一亚组不显著,由此推断存在亚组效应(交互作用)是不正确的;这种推断常常有很高的假阳性;此时应当进行交互作用的检验^[15]。由于多重比较的问题增大了 I 类错误的概率,因此通常并不鼓励进行亚组分析;事后的亚组分析就更不容易被进一步的研究证实了;这类分析往往不具有很大的可信性。在调整分析方面,是否进行调整分析以及对哪些因素进行调整分析不应当取决于基线差异是否具有统计学意义^[16]。这是因为即使某些因素在基线差异上有统计学意义,如果这些因素对结局无影响或对结局影响较弱,那么调整分析是无意义的;而即使某些因素在基线差异上无统计学意义,如果这些因素对结局有比较强的影

响,调整分析仍然是必要的。例如,在一个比较溶栓药物疗效的研究里^[17],基线时既往脑出血病史存在 1% 的差异可能在统计学上无差异,却可以影响药物治疗后出血性脑卒中的发生率(研究结局),因此这种基线差异在临床上被认为是很重要的。如果基线状态时存在这种重要特征的差异,就应当进行调整分析。作者应当阐明选择调整变量的依据、调整的方法以及这种调整是方案中计划的还是由数据提示的。

条目 15 通过对基线时各组的人口学特征和临床特征进行比较,可以评价各比较组的可比性。对连续性变量(如血压、年龄、胆固醇水平等)组间相似性可以通过比较均数和标准差(正态分布)或中位数与范围(偏态分布)来判断;对分类变量(如性别、疾病阶段等)组间相似性可以通过比较各类数量和比例来判断。比较组间越相似,研究结果反映真实治疗效应的可信性就越大,特别是在未进行调整分析的情况下。

条目 16 用分数的形式来表达结果,有助于判断是否所有接受随机分配的研究对象均纳入分析,如果不是,有多少被排除在分析之外。因此,结果不应当只表达为一个效应指标(如 RR),还应包括各组的结局发生情况。许多试验都可能出现“偏离研究方案”的情况,如一些患者未接受全程的干预或正确的干预,或者一些不合格的患者被错误地进行了随机分配。在这种情况下,处理这类问题一个广为推荐的方法是根据所有研究对象最初分组的情况进行分析,而不考虑分组后发生的情况(ITT 分析)。如果仅仅分析那些完全依从研究方案完成干预和结局评价的研究对象,则这样的分析为“按实际处理分析”(on-treatment analysis)或“符合方案分析”(per protocol analysis)。分析中未纳入所有的研究对象会导致偏倚的产生。ITT 分析可以避免由于研究对象不随机退出而造成的偏倚^[18]。

条目 17 对每个结局,研究结果中应报告各组结局发生情况(如,发生或未发生事件的比例,或测量指标的均数和标准差),以及组间比较的情况,即效应大小。对二分类结局而言,效应指标可能是相对危险度、比值比或者率差;对生存资料而言,效应指标可能是危险比或中位生存时间差;对连续资料而言,效应指标通常是均差。对所有结局指标而言,应当提供可信区间以提示效应估计的精度(不确定性)。对统计学上不显著的差异而言,可信区间具有特殊的价值,即可能提示存在一个临床上重要的差异。所有计划的主要结局和次要结局结果均应报告,而不仅是那些统计学上具有显著差异的分析结果。

条目 20 从方法学的角度,对研究存在的缺陷进行评价是十分必要的。应当注意区分统计学上的差异和临床上的重要性。作者应避免将一个统计上不存在差异的结果解释为干预效果是一样的。不论 P 值的大小,可信区间提供了重要的信息来判断研究结果是否可能具有临床上的重要性。当存在数个干预或结局以及亚组分析的时候,就会出现多重分析问题。作者应谨慎地去解释多重分析的研究结果。在这种情况下,一些有统计学意义的结果可能仅仅是由于机会造成的。

4. 结论:随机对照试验的报告常常是不充分、不完整的,这种情况妨碍了对研究的严格评价和结果的推广。而 CONSORT 声明有助于研究者和其他人写作或评价临床试验的报告。柳叶刀杂志(The Lancet),英国医学杂志(BMJ),美国医学会杂志(JAMA)和内科年鉴(Annals of Internal Medicine),以及越来越多的生物医学期刊编辑组织,包括国际医学杂志编辑委员会、科学编辑委员会都对 CONSORT 声明提供了官方的支持。初步的研究提示使用 CONSORT 的确有助于改进随机对照试验报告的质量^[19]。我们相信,CONSORT 声明及其解释说明文件的广泛推广,将不断促进和提高随机对照试验报告的质量,最终有益于临床和医疗卫生中干预的评价与实施。

参 考 文 献

- Begg C, Cho D, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA, 1996, 276: 637-639.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Medical Research Methodology, 2001, 1-2.
- DerSimonian R, Charette LJ, McPeck B, et al. Reporting on methods in clinical trials. N Engl J Med, 1982, 306: 1332-1337.
- Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. N Engl J Med, 1987, 317: 426-432.
- Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA, 1994, 272: 122-124.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ, 1999, 319: 670-674.
- Schulz KF, Chalmers I, Grimes DA, et al. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA, 1994, 272: 125-128.
- Hotopf M, Lewis G, Normand C. Putting trials on trial— the costs and consequences of small trials in depression: a systematic review of methodology. J Epidemiol Community Health, 1997, 51: 354-358.
- Altman DG, Schulz KF, Moher D. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med, 2001, 134: 663-694.
- Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA, 1995, 273: 408-412.
- Schulz KF. Subverting randomization in controlled trials. JAMA, 1995, 274: 1456-1458.
- Götzsche PC. Blinding during data analysis and writing of manuscripts. Control Clin Trials, 1996, 17: 285-293.
- A proposal for structured reporting of randomized controlled trials. The standards of reporting trials group. JAMA, 1994, 272: 1926-1931.
- Schulz KF, Grimes DA, Altman DG, et al. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. BMJ, 1996, 312: 742-744.
- Matthews JN, Altman DG. Statistics notes. interaction 2: compare effect sizes not P values. BMJ, 1996, 313: 808.
- Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet, 2000, 355: 1064-1069.
- Wallen L, Swahn E, Kontny F, et al. For FRISC II investigators. Invasive compared with non-invasive treatment in unstable coronary-artery disease: FRISC II prospective randomised multicentre study. Lancet, 1999, 354: 708-715.
- Lachin JL. Statistical considerations in the intent-to-treat principle. Control Clin Trials, 2000, 21: 526.
- Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before and after evaluation? JAMA, 2001, 285: 1992-1995.

(收稿日期:2006-11-08)

(本文编辑:尹廉)