

如何撰写高质量的流行病学研究论文

第五讲 整群随机对照试验的报告规范——CONSORT 扩展声明

王波 詹思延

【导读】 整群随机对照试验在公共卫生与临床服务领域的应用日益广泛,并为医疗卫生干预提供了重要的证据。对这类研究结果的正确评价有赖于充分、准确的报告。临床试验报告统一标准(CONSORT)扩展声明是当前权威的整群随机对照试验报告规范。本文介绍了整群随机对照试验的特殊方法学问题、实施与报告现状以及 CONSORT 扩展声明的清单和流程图,并对清单条目进行了解释与说明。

【关键词】 临床试验报告统一标准;整群随机对照试验

近年来,准确、充分地报告随机对照试验已经受到了越来越多的重视。以改进平行对照随机试验(研究对象以个体为单位被随机分配到各干预组)报告为目的的临床试验报告统一标准(CONSORT)声明已经发表(或参见本刊 2006 年第 12 期第 1086-1088 页)^[1]。整群随机对照试验(cluster randomized trials)是将研究对象以群组为单位进行随机分配的一种试验设计。在很多情况下,医疗卫生干预是在群体水平上实施的,如针对社区人群的健康教育、针对医生实施指南的干预等,这种情况下很难以个体为单位进行随机分组;同时,整群随机试验设计可以更好地避免不同于干预之间的污染(contamination),因此在公共卫生与医疗服务领域有着广泛而重要的用途^[2]。

1. 整群随机试验的方法学问题和实施、报告中存在的问题^[3,4]:整群随机试验在设计、实施和分析上比个体化随机试验更为复杂。

在设计方面,由于同一群组的个体往往较不同群组的个体在干预效果上具有更为相似的结果(非独立性),因此在相同的样本量下,整群随机对照试验提供的信息总是少于个体化随机试验,这是确定整群随机对照试验样本量的一个重要考虑^[5]。这种有效样本量的降低程度取决于平均群组大小和群内相关程度的大小,即群内相关系数(ρ)。通常为了补偿整群随机对照试验把握度的降低,样本量应当在个体化随机对照试验基础上扩大 $1 + (m - 1)\rho$ 倍,其中 m 为平均群组大小。

在实施方面,整群随机对照试验也与个体化随机试验有所不同,潜在的偏倚可能出现在群体和个体两个水平:在群体水平,为了防止有偏的分配,应当进行严格、正确的随机化,一旦分配完成后就必须确保各群组接受所分配的干预,并避免退出,以防止退出偏倚;在个体水平,整群随机分配完成后,偏倚可能会在个体对象进入研究时出现,如负责征集研究对象的人员如果知道分组情况,就可能如同个体化随机

试验未能做到分配隐藏一样产生偏倚。

如果在随机化完成之后才征求研究对象是否同意参加试验,可能引起随机化之后的选择偏倚问题,如果没有知情同意还可能引起伦理学问题^[6]。在随机化之前进行知情同意,以及由对分组情况未知的人员去募集和纳入研究对象,可以减少上述偏倚。

在整群随机对照试验的分析方面^[7],由于同一群内个体往往具有一定的相关性(非独立性),违背了个体化随机试验假设检验和统计推断的基本假定。在这种情况下,必须使用可以处理整群设计效应的分析方法,反之会导致研究夸大干预效应,出现假阳性结果。Cornfield^[8]认为“用个体随机化的方法去分析整群随机试验是一种自欺欺人的做法”。然而,实际的研究中对整群随机对照试验的不正确分析相当常见。

对 1990 - 1993 年发表在 2 个主流公共卫生期刊上的 21 个一级预防整群随机试验研究进行回顾发现^[9],在计算样本量或把握度时考虑到整群设计效应的研究不到 20%,在未能考虑到整群设计效应的 17 个研究里,有 3 个为阴性结果,而这可能是由于把握度不够造成的。Divine 等^[10]调查了 54 个关于医生行为的整群随机对照试验研究,发现 70% 的研究使用了错误的分析单位。Donner 等^[11]对 1979 - 1989 年间发表的 16 个整群随机对照试验进行调查后发现,在样本量计算和分析方面分别只有 3 个和 8 个研究考虑了整群设计效应。一项针对指南实施研究的调查发现^[12],1966 - 2000 年发表的 149 篇整群随机对照试验中,报告样本量计算的 37 篇(25%)中只有 19 篇考虑了整群设计效应,有 62 篇(42%)存在着分析单位错误,仅 6 篇报告了群内相关系数。对临床决策支持系统研究的分析发现^[13],2000 年以前发表的 24 篇整群随机对照试验里,没有任何一篇报告考虑到整群设计效应的样本量计算,10 篇(42%)在分析时未考虑到整群设计效应。过去 30 年来在撒哈拉以南非洲开展的 51 个整群随机对照试验中^[14],在样本量计算和分析方面考虑到整群设计效应的研究分别只有 10 个和 19 个,报告了群内相关系数和设计效应的研究分别只有 1 个和 3 个。

2. CONSORT 声明向整群随机对照试验的扩展:由于整

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

群随机对照试验在设计、实施和分析上的复杂性, Elbourne^[15] 在 1997 年提出整群随机对照试验的报告需要专门的指南。2001 年 Elbourne, Campbell^[16] 对 CONSORT 声明进行了修改, 提出了针对两水平、完全随机化的整群随机对照试验的报告规范讨论版。2004 年, 针对整群随机对照试验的 CONSORT 扩展声明正式发表^[3], 包括一个清单(表 1)和一个流程图(图 1)。针对整群随机对照试验的特殊性, 修改后的 CONSORT 声明增加了以下内容: 采取整群设计的原理; 如何考虑整群设计效应进行样本量的计算; 如何考虑到整群设计效应进行分析; 从随机分配到分析过程中整群和个体的流动情况。

3. 清单条目解释与说明: 为了帮助研究者更好地理解和应用这一扩展的 CONSORT 声明, 制定者对有变动的条目进行了解释和说明, 以下逐一进行解读:

条目 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”), 并详细说明分配是以群组为单位进行的。在标题或摘要里注明设计类型的主要目的是确保 Medline 将其正确标记为“整群随机对照试验”, 从而有利于更容易地识

别这类研究。研究者应当考虑在摘要中报告群组的数量。

条目 2 应有科学背景与原理的解释(包括使用整群设计的原理)。根据赫尔辛基宣言的原则^[17], 将人群暴露于不必要的研究危险之下是不道德的。整群随机设计增加了研究的复杂性, 需要更多的研究对象, 因此对采用整群随机设计的原理进行介绍是十分必要的^[11]。

条目 3 说明研究对象和群组的人选标准及数据收集的机构、地点。由于整群随机对照试验有两个统计推断水平: 群体水平和个体水平^[18], 因此群组 and 个体对象的人选标准均需要报告。在一个整群随机试验里, 主要的人选标准通常是所有群组在一个规定的地理区域内。

条目 4 包括各组干预的详细描述、干预针对个体水平还是群体水平(或者两者都有)以及何时、如何实施。如果干预针对的是群体水平, 则应当描述干预实施的细节。

条目 5 设定目标和假说, 及其针对个体水平还是群体水平(或者两者都有)。应清楚说明设定的目标和假说针对的是个体水平, 还是群体水平, 或者两者都有。对统计推断水平的了解将有助于统计方法的解释。

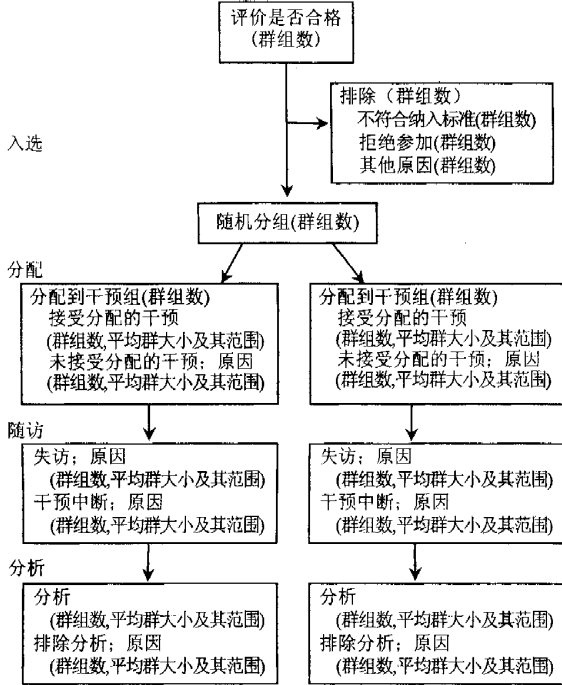
表 1 报告整群随机对照试验时需要包括的条目清单

(斜体部分代表为适用于整群随机对照试验, 在原 CONSROT 声明清单上修改的内容)

内容与主题	条目	描 述
标题与摘要		
设计	1 ^a	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”), 并详细说明分配是以整群为单位进行的
介绍		
背景	2 ^a	科学背景与原理的解释, 包括使用整群设计的原理
方法		
研究对象	3 ^a	研究对象和群组的人选标准, 数据收集的机构和地点
干预	4 ^a	各组干预的详细描述, 干预针对个体水平还是群体水平(或者两者都有), 以及何时、如何实施
目标	5 ^a	设定的目标和假说, 以及它们针对个体水平还是群体水平(或者两者都有)
结局	6 ^a	明确定义主要和次要结局指标, 它们针对个体水平还是群体水平(或者两者都有), 如果可能, 描述改进结局测量质量的方法(如多次观察, 对测量者进行培训)
样本大小	7 ^a	总样本量大小如何确定(包括计算方法, 群组的数量, 群组大小, 群内相关系数及其不确定性的指标), 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 ^a	产生随机分配序列的方法, 包括任何限定情况(如分组、分层、匹配)
分配隐藏	9 ^a	按照产生的序列进行随机分配的方法, 详细说明分配基于整群而不是个体, 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、评价结局者是否对分组未知? 如果是, 盲法是否成功
统计方法	12 ^a	比较各组主要结局的统计学方法, 指出如何处理整群设计效应; 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 ^a	各个阶段群组和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、和进入分析的群组和研究对象数量。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 ^a	如果可能, 报告各组个体水平和群体水平的基线人口学特征和临床特征
分析的数量	16 ^a	纳入每个分析的各干预组的群组和研究对象的数量(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 ^a	对每个主要和次要结局, 如果可能, 应报告个体或群体水平上每个组的综合结果, 估计效应大小和精确度(如 95% 可信区间), 报告各主要结局的群内相关系数
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、多重结局有关的危险, 对结果进行解释
可推广性	21 ^a	试验结果向个体和/或整群(如果相关)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

^a 与原 CONSORT 声明清单相比, 在修改后有变化的条目

条目 6 明确定义主要和次要结局指标,及其针对个体水平还是群体水平(或者两者都有),如果可能,描述改进结局测量质量的方法(如多次观察,对测量者进行培训)。一项研究究竟是在群体水平还是个体水平上被评价,对选择正确的结局资料分析方法十分关键。因此,明确报告测量个体还是群体的结局很重要。



(只适用于结局在群体水平上的情况,如果是多水平分析,则图中还应包括个体对象的流动情况)

图1 整群随机对照试验各个阶段群组与研究对象流程图

条目 7 应说明总样本量大小如何确定(包括计算方法、群组的数量、群组大小、群内相关系数及其不确定性的指标),如果可能,对中期分析和终止试验的条件进行解释。如前所述,要获得与个体化随机试验相同的把握度,整群随机试验的样本量必须扩大,要扩大的比例取决于群内相关系数和群组大小。报告整群随机对照试验时应阐明计算群组数量和群组大小所使用的假定。

条目 8 列出产生随机分配序列的方法,包括任何限定情况(如分组、分层、匹配)。整群随机对照试验可能使用简单、完全随机化设计,也可能是整群匹配设计,或分层设计。在个体化随机试验里,随机分配通常可以保证各比较组间基线特征上的差异仅仅是由于随机误差(机会)造成的,而不是系统化偏倚的结果^[19]。但在整群随机化对照试验里,通常群组的数量相对较少,因此并不能保证比较组间基线特征差异完全由机会造成。因此整群随机对照试验设计中常常施加某种形式的“限制”(匹配或分层),以最大程度地减少比较组间的不平衡。这种“限制”会影响样本量的计算和结果分析,因此在报告中应当明确报告。

条目 9 按照产生的序列进行随机分配的方法,详细说

明分配基于群体而不是个体,清楚阐明在分派干预之前序列是否被隐藏。在个体随机化试验中,如果纳入的研究对象事先知晓分配序列,就可能引起偏倚^[20]。在整群随机对照试验中,由于预先确定了群内各个成员所分配到干预,由此导致在群内产生选择偏倚(选择性纳入研究对象)的可能性是很高的^[2,4]。因此,对那些为了减小选择偏倚而采取的策略(如是否群内所有个体均纳入了研究,或者是否负责纳入研究对象的人对群组分配情况未知)进行描述就十分重要。

条目 12 比较各组主要结局的统计学方法(指出如何处理整群设计效应)及其他分析方法,如亚组分析和调整分析。对统计推断水平的识别有助于读者评价分析方法。例如,如果干预针对的是群体水平,并且结局也是在整群水平上汇总,这时就不需要复杂的整群调整分析。如果结局是在个体病例水平上测量的,分析就需要调整群效应。

条目 13 介绍各个阶段群组和研究对象的流动情况(强烈推荐使用流程图)。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群组以及研究对象数量。描述实际研究偏离研究方案的程度及原因。了解所有群组和个体中未接受分配干预的比例、退出的比例、未获得随访资料的比例,对准确解释研究结果十分重要。例如,在随机分配之后被排除的研究个体可能无法代表研究对象,不同的退出率可能与所接受的干预直接相关^[21]。整群随机对照试验里更容易出现比较组间在按照研究方案接受干预和随访等方面的差别,这是因为群体水平和个体水平都可能出现退出的情况。因此,在报告整群随机对照试验时,应描述群组和个体的流动情况。流程图通常是表达这类信息的最好途径(图 1)。

条目 15 如果有可能应报告各个个体水平和群体水平的基线人口学特征、临床特征。如前所述,个体化随机分配常常可以保证比较组间仅存在机会造成的差异,而不存在系统偏倚。然而这一假定在整群随机对照试验并不一定成立,因此有必要同时报告群体和个体水平的基线特征。

条目 16 纳入每个分析的各干预组群组和研究对象的数量(分母),以及是否进行了意向性(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%)。在个体化随机试验中,研究对象的数量对结果的解释很重要。有时候,并非所有研究对象都参加各个结局的分析。在整群随机对照试验里,这一问题更加复杂,因为除了个体水平外,还有可能并非所有群组都参加分析。由于样本量和把握度是基于假定所有研究对象和群组都参加分析而计算出的,因此应当报告参加分析的研究对象和群组的数量,以便评估统计学把握度是否降低。如果只有一个主要结局,那么流程图可以反映这一信息;如果有不止一个主要结局,则应分别报告参加各个结局分析的研究对象和群组的数量。

条目 17 对每个主要和次要结局,如果可能,应报告个体或群体水平上每个组的综合结果,估计效应大小和精确度(如 95% 可信区间),报告各主要结局的群内相关系数。当报

告整群随机对照试验结果的时候,应当报告各主要结局的点估计和可信区间。考虑到群内相关对研究把握度的影响,应当报告各个结局的群内相关系数。这一信息有助于读者评价最初的样本大小计算是否合适,以及各结局整群效应的大小。同时报告调整和未调整的估计值也有助于判断整群设计效应的大小。一些研究者支持报告群相关系数^[22],认为这有利于将来在相似领域开展整群随机试验。

条目 21 试验结果应具有向个体和/或整群(如果相关)的可推广性(外部有效性)。在任何一个试验报告的讨论部分,都应当考虑结果的外部有效性。整群随机对照试验的外部有效性更为复杂,因为结果既可能被推广到群体,也可能被推广到群内的个体,或者两者均可。因此,应当清楚报告的外部有效性具有的水平。

4. 结论:整群随机对照试验的报告应当包括设计、分析的重要信息,以便读者准确解释其结果。实证研究已经发现^[23],不充分的方法学报告与估计研究效应时的偏倚有关。通过使用针对平行对照随机试验的 CONSORT 声明已经获得了更好的报告质量^[24]。我们相信 CONSORT 扩展声明同样会促进和改善整群随机对照试验的报告质量。

参 考 文 献

- [1] Moher D, Schulz KF, Altman DG, et al. The CONSORT statement; revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 2001, 357: 1191-1194.
- [2] Fayers PM, Jordhoy MS, Kaasa S. Cluster-randomized trials. *Palliat Med*, 2002, 26: 69-70.
- [3] Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomized trials. *BMJ*, 2004, 328: 702-708.
- [4] Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ*, 2003, 327: 785-789.
- [5] Hayes RJ, Bennett S. Simple sample size calculation for clusterrandomized trials. *Int J Epidemiol*, 1999, 28: 319-326.
- [6] Edwards S, Braunholtz DA, Lilford RJ, et al. Ethical issues in the design and conduct of cluster randomised trials. *BMJ*, 1999, 318: 1407-1409.
- [7] Wears RL. Advanced statistics: statistical methods for analyzing cluster and cluster-randomized data. *Acad Emerg Med*, 2002, 9: 330-341.
- [8] Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol*, 1978, 108: 100-102.
- [9] Simpson JM, Klar N, Donner A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Public Health*, 1995, 85: 1378-1383.
- [10] Divine GW, Brown JT, Frazier LM. Unit of analysis error in studies about physicians' patient care behavior. *J Gen Intern Med*, 1992, 7: 623-629.
- [11] Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979 - 1989. *Int J Epidemiol*, 1990, 19: 795-800.
- [12] MacLennan GS, Ramsay CR, Mollison J, et al. Room for improvement in the reporting of cluster randomised trials in behaviour change research. *Control Clin Trials*, 2003, 24: S69-70.
- [13] Chuang JH, Hripcsak G, Jenders RA. Considering clustering: a methodological review of clinical decision support system studies. *Proc AMIA Symp*, 2000: 146-150.
- [14] Isaakidis P, Ioannidis JPA. Evaluation of cluster randomized controlled trials in sub-Saharan Africa. *Am J Epidemiol*, 2003, 158: 921-926.
- [15] Elbourne D. Guidelines are needed for evaluations that use cluster approach. *BMJ*, 1997, 315: 1620-1621.
- [16] Elbourne DR, Campbell MK. Extending the CONSORT statement to cluster randomised trials: for discussion. *Stat Med*, 2001, 20: 489-496.
- [17] World Medical Association. Recommendations guiding physicians in biomedical research involving human subjects. World Medical Association Declaration of Helsinki. *JAMA*, 1997, 277: 925-926.
- [18] Atienza AA, King AC. Community-based health intervention trials: an overview of methodological issues. *Epidemiol Rev*, 2002, 24: 72-79.
- [19] Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet*, 1990, 335: 149-153.
- [20] Schulz K, Chalmers I, Hayes R, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 1995, 273: 408-412.
- [21] Sackett DL, Gent M. Controversies in counting and attributing events in clinical trials. *N Engl J Med*, 1979, 301: 1410-1412.
- [22] Ukoumunne OC, Gulliford MC, Chinn S, et al. Methods for evaluating area-wide and organisation based interventions in health and health care: a systematic review. *Health Technol Assess*, 1999, 3(5): iii-92.
- [23] Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ*, 2001, 323: 42-46.
- [24] Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomised trials: a comparative before-and-after evaluation. *JAMA*, 2001, 285: 1992-1995.

(收稿日期:2007-01-11)

(本文编辑:张林东)