

RCT 论文统计学报告自查清单与报告指南的应用与评价*

刘清海¹⁾ 方积乾²⁾ 甘章平³⁾ 朱佩玲⁴⁾ 常湘珍⁵⁾

收稿日期:2008-08-24
修回日期:2009-06-23

- 1) 中山大学学报编辑部//编辑与出版学研究中心, 510080 广州中山二路 74 号, E-mail: liuqh@mail.sysu.edu.cn
2) 中山大学公共卫生学院统计学教研室, 510080 广州中山二路 74 号, E-mail: fangjq@mail.sysu.edu.cn;
3) 中山大学《中国神经精神疾病杂志》编辑部, 510080 广州中山二路 74 号, E-mail: editorgan@yahoo.com.cn
4) 中山大学《新医学》编辑部, 510080 广州中山二路 74 号, E-mail: zhupeiling@163.com
5) 中山大学《中华显微外科杂志》编辑部, 510080 广州中山二路 74 号, E-mail: cxzzch@163.com

摘要 国内未见有对于医学论文的统计学报告进行指导的清单或报告指南。我们在发表随机对照临床试验(RCT)论文统计学报告自查清单后,又为其起草了相应的报告指南。为了了解这份清单和报告指南在实践中应用的实际效果,我们交由 4 种核心期刊的编辑请作者依自查清单进行自查,必要时参考报告指南,并在作者修改前后对论文的统计学报告质量进行评分。结果表明,自查修改后论文统计学报告质量得分平均由 13.4 分提高到 19.9 分,但离满分(27 分)还有一定差距,且有不少项目缺省。这说明自查清单有助于提高相应设计的临床研究的统计学报告质量,同时也反映了自查清单要求作者统计学项目报告详尽和作者实际统计学水平和时间比较有限的矛盾。我们建议,一方面,医学编辑和医学统计学家应进一步修改和完善有关的自查清单和报告指南;另一方面,医学研究界应自我约束和自我提高,努力提升自己的统计学素质,与国际接轨,努力报告好医学论文的统计学项目。

关键词 随机对照临床试验(RCT) 医学论文 统计学报告质量 自查清单 报告指南

自从 1996 年旨在提高随机对照临床试验报告质量的 CONSORT(consolidated standard of reporting trials)声明发表以来^[1],国际主要医学期刊发表的随机对照临床试验(randomized controlled trial, RCT)论文的报告质量开始提高^[2]。本世纪以来,各种旨在提高不同设计类型医学论文报告质量的声明如雨后春笋似的开发出来,仅 CONSORT 修订版^[3]及其长长的解释与说明^[4]就衍生出适合不同目的多种扩展版,如针对群随机对照研究(cluster randomized trial, CRT)的 CONSORT、针对研究不良反应(harm)的 CONSORT、针对非劣效性与等效性(non-inferiority and equivalence)研究的 CONSORT、针对草药(herbal)的 CONSORT,2008 年还研制出了针对非药物处理(non-pharmaceutical)的 CONSORT^[5]、针对会议摘要的 CONSORT^[6]等。其他设计类型的报告指南

包括针对观察性研究的 STROBE 声明^[7]、针对行为与公共卫生干预评价的非随机化研究报告的 TREND 声明^[8]、针对诊断精确性试验的 STARD 声明^[9]等及其解释和说明。然而,国内直到近年才有文献开始介绍国外的各类报告清单^[10,11]。我们依靠国内 50 多位流行病学、卫生统计学、医学编辑和临床研究专家研制出了适合国内应用的随机对照临床试验论文的自查清单^[12],并为其起草了报告指南,之后将其应用于实践当中,对其使用情况进行评价。

1 自查清单与报告指南的起草与应用

依据国内流行病学、卫生统计学、医学编辑及临床研究

* 基金项目:广东省软科学项目(2005B70101121)

等领域 50 余位专家的评分,我们制定了随机对照临床试验论文统计学报告自查清单,共 61 项,发表在《中华医学杂志》上^[12]。该自查清单主要适用于随机对照的临床试验研究,也可以用于非随机但有对照的临床试验参考,不适用于病例总结及实验性研究。为了便于应用,我们将项目改成了列表的形式。清单制定以后,我们参考 CONSORT 研究小组的做法,为自查清单的每一项逐项地为其增加报告的理由说明,也补充了报告的形式,还举了些例子(部分例子取自 CONSORT 声明的解释与说明)^[4],以便作者更好地理解指南,参照写作。为方便描述,摘要部分项目的解释放在了正文项目解释之后。起草完后,将草案交由本研究的专家组讨论修改和定形。

将定形的草案连同自查清单全表呈请中山大学主办的几种医学期刊的编辑应用,并评价其使用效果。应用的期刊包括《中山大学学报(医学科学版)》、《中国神经精神疾病杂志》、《新医学》、《中华显微外科杂志》等。应用时说明本清单和指南适用于随机对照临床研究论文,或至少是有对照的临床研究论文,不适合病例报告、经验总结、实验性研究等。使用时编辑将需要修改的相应设计类型的论文退请作者修改,同时附上《随机对照研究论文统计学报告自查清单》和《随机对照研究论文统计学报告指南》,提请作者对照清单自查,如有不明则参考指南的说明和举例,修改后寄回编辑部。编辑在退修前对论文的统计学报告质量以《随机对照临床研究论文统计学报告质量评价量表》^[13]进行一次评分,作者修改完成寄回后再以同一量表对其进行另一次评分,并对修改前后的评分进行对照分析。

2 应用结果与评价

经使用 3 个月,编辑发现,完全随机对照的研究较少,扩大应用到有对照的临床试验研究后,共收回 21 篇适合论文。作者与编辑皆评价认为自查清单和报告指南比较清楚,应用起来没有问题,但由于项目较多,有些项目难以报告,尤其是非随机对照的临床研究论文。这些难以报告的项目主要包括:确定样本量及确定理由、确定有临床意义的最小差值或比值、抽样的具体方法、实验和对照因素盲法效果的描述、实施者与实验过程可比性的说明、主要指标测量方法与精确度、负性反应或事件的测量范围与方法、数据收集的方法与质量保证措施、个体观察终点与整体研究终点的定义、控制可能偏倚的努力、对主要指标拟行单侧还是双侧检验、对主要指标进行检验的 α 水平、研究或试验的起止时间、征集对象例数、符合研究标准例数、效果分析采取的数据集及各组

样本量、随访数、失访数、偏离计划数及偏离原因、干预前后主要指标的集中与离散趋势描述并明确标记、主要指标干预前后差值或比值的均数与置信区间、计划内多重比较的具体方法、图示是否符合制图原则、 $a \pm b$ 形式中 b 有无明确标记、对设计中可能存在的偏倚的说明、比较利弊得出总的临床性结论、临床结论的适用性/外推性说明。分析发现,较多作者在自查时对含有“主要指标”、“检验的单双侧”、“检验水平”、“分析集”、“偏倚”、“集散趋势”、“负性反应或事件”等术语的项目缺项。

用评价量表对这些应用过自查清单的论文进行应用前后对比研究发现,应用前平均得分为 13.4 分,按评价量表的分级属于二级,应用后平均得分上升一个档次(19.9 分),属于一级,但是离满分 27 分还有一定距离,应用后得分平均升高约 6.4 分,一是说明自查清单和报告指南有助于临床研究论文统计学报告质量的提高,二是说明这几期刊的统计学报告水平也仍有待进一步提高(表 1)。

表 1 几种中文期刊应用自查清单和报告指南后统计学报告质量得分

	论文数	得分均数	得分标准差
修改前	21	13.4	3.0
修改后	21	19.9	4.1
差值	21	6.4	2.7

经配对 t 检验, $t = 10.788, P < 0.001$;

得分为修约值,导致差值与修改前后差值不尽一致。

3 结果分析

我们的评价说明了两个问题:

一是本研究小组制定的随机对照临床研究论文统计学报告自查清单与报告指南基本上适用于国内至少有对照的临床研究论文的作者自查,大部分项目简单明了,易于理解,有助于提高论文的统计学报告质量,但是项目偏多,有些项目不是每篇有对照的临床研究都会涉及到,因此项目还可以进一步精简。

二是国内医学期刊的统计学报告质量偏低,还有待于大幅度提高。虽然我们并非按照随机抽样原则抽取国内医学期刊——这主要是因为我们的研究目的是对自查清单和报告指南的应用评价,而非对国内医学期刊统计学报告水平的评价,但是所抽取期刊皆为医学核心期刊/国家级期刊,在同类刊物中是有一定影响的,可以在一定程度上说明全国的状况。

仔细分析作者自查时经常缺省填写的项目发现,涉及“主要指标”、“检验的单双侧”、“偏倚”、“分析集”、“集散趋势”、“负性反应”等术语的项目比较常见缺省填写。这些术语有些是最基本的统计学或流行病学术语,另有些则可算是一般论文写作时的术语,这些术语在我们的报告指南里都有说明,但是作者仍然缺省填写较多。分析其可能原因是部分作者统计学知识过于贫乏,连基本的统计学术语也不明白或忘了,而写作时一是请统计学专家撰写统计学部分,二是参照已发表文献依葫芦画瓢,在编辑要求作者依自查清单和报告指南自查而遇上困难时,或因畏难而放弃查找报告指南,或因怕费时放弃填写。这充分说明我国医学论文的作者统计学底子较薄,要迎头赶上恐怕要下很大功夫和费很长时间,而提高思想上的认识可能更费周折。

4 进一步的工作和建议

虽然我们的自查清单和报告指南在作者使用后能够提高随机对照研究论文的统计学报告水平,但是我们也发现清单项目偏多,导致报告指南偏长,部分用到统计学术语的项目作者漏报较多,其原因可能与作者的统计学基础薄弱有关。这说明,我们至少要做两方面的工作:一是进一步精简和完善自查清单和报告指南;二是努力提高医学研究者也即作者的统计学意识和水平。参考国外 CONSORT 等清单的研制过程,国内也应该集中一批相关专家座谈数次,清单才可能达到比较好的效果。国内中国循证医学中心开始牵头组织为针对中国传统医药的临床试验报告研制相应的 CONSORT 清单^[14],这是一个很好的开始,希望能有更多的单位和个人加入到这个队伍当中。然而,努力提高医学研究者的统计学意识和水平不是一件简单的事,这涉及到医学统计学的教育、医学研究者的继续教育、医学科研的管理、人事政策、医学期刊的编辑把关等一系列部门的工作,是一项复杂的系统工程,不可能一蹴而就。

此外,我们的研究成果与 CONSORT 的一系列成果类似^[1-6],都是基于随机对照的临床研究,国外的其他类似成果也基本都是针对临床研究的,如 STROBE、TREND、STARD 等声明^[7-9],而针对动物实验和微细(细胞与分子)实验研究的类似成果尚未见出台。可以预见,在未来的数年内,这些研究类型的报告清单与说明也将研制成功。因此,我们的医学统计学专家、医学研究者、医学编辑等应踊跃地加入到这些清单的研制过程当中,贡献自己的一份力量,也为更好地融入世界

医学科研与报告的潮流当中,进一步促进中国的医学研究与报告水平的提高。

(本文所指报告指南较长,为节约版面,本文仅附自查清单,需要报告指南原稿者可向作者索取)

参考文献

- 1 Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: The CONSORT Statement. *JAMA*, 1996,276(8):637-639
- 2 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(15):1992-1995
- 3 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *The Lancet*, 2001,357:1191-1194
- 4 Altman DG, Schulz KF, Moher D, et al. The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration. *Ann Intern Med*, 2001,134:663-694
- 5 Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomised trials in journal and conference abstracts: Explanation and Elaboration. *PLoS Med*, 2008,5(1):e20. DOI:10.1371/journal.pmed.0050020
- 6 Boutron I, Moher D, Altman DG, et al. Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration. *Ann Intern Med*, 2008,148:295-309
- 7 von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Med*, 2007, e296. doi: 10.1371/journal.pmed.0040296
- 8 Des Jarlais DC, Lyles C, Crepaz N, et al. Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement. *Am J Public Health*, 2004,94(3):361-366.
- 9 Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration. *Ann Intern Med*, 2003,138(1):W1-W12
- 10 刘玉秀,姚晨,盛梅. 随机对照试验的统一报告格式——CONSORT 声明. *中国临床药理学杂志*,2004,20(1):76-80
- 11 王波,詹思延. 如何撰写高质量的流行病学研究论文 第一讲 观察性流行病学研究报告规范——STROBE 介绍. *中华流行病学杂志*, 2006,27(6):547-549
- 12 刘清海,方积乾. 应重视医学临床试验论文统计学问题——统计学项目自查清单的研制. *中华医学杂志*, 2007,87(34):2446-2448
- 13 刘清海,方积乾. 医学论文统计学报告水平评价量表的研制及意义. *编辑学报*,2008,20(3):278-280
- 14 吴泰相,李幼平,卞兆祥等. 中医药临床随机对照试验报告规范(征求意见稿). *中国循证医学*,2007,7(8):601-605

· 附件：随机对照临床试验论文统计学项目自查清单 ·

为提高医学论文统计学报告质量,我们特制定本清单。本清单经调查国内 50 余位有关专家而形成。请医学论文作者投稿时对照全文,核实或补充文章各部分应报告的统计学项目,并将其所在的页号和行号填入本表,其中,A 代表摘要,I 代表引言,M 代表材料与方法,R 为结果(R3 ~ R8 的例数可以流程图的方式给出),D 为讨论部分。

项目	页号/行号	项目	页号/行号
A1 分组的具体方法,应说明如何“随机分组”	/	R1 研究或实验的起止时间	/
A2 实验的实施与评价是否实行盲法及谁对什么“盲”	/	R2 随访的起止时间	/
A3 样本总量与分组样本量	/	R3 征集对象例数	/
A4 应说明分析的主要指标	/	R4 符合研究标准数	/
A5 对主要指标使用的统计检验方法	/	R5 实际行分组数	/
A6 主要指标的集中趋势(如均数或比值)与离散趋势(如标准差或置信区间)	/	R6 完成干预例数	/
A7 主要指标比较的精确 P 值	/	R7 偏离计划数及偏离原因	/
A8 关于两组主要指标差异的临床结论	/	R8 随访数、失访数	/
I1 研究类型的定性陈述(“探索”或“确证”)	/	R9 效果分析采取的数据集及各组样本量	/
I2 清楚陈述研究目的及研究假设(优效、非劣效或等效性检验)	/	R10 负性反应或事件的分析集	/
M1 目标人群描述如人口、地理、医院性质、是否转诊、诊断	/	R11 各组人口学及临床特征的基线水平的可比性与不同	/
M2 明确的诊断标准	/	R12 分析主要指标的各组例数与样本数(人/牙/眼……)	/
M3 入选标准与排除标准	/	R13 干预前、后主要指标的集中与离散趋势描述并明确标记	/
M4 确定样本量及确定理由	/	R14 主要指标干预前后差值或比值的均数与置信区间	/
M5 确定有临床意义的最小差值或比值	/	R15 有无进行特殊数据处理(如异常值、数据转换等)	/
M6 抽样的具体方法	/	R16 主要指标统计检验的实际方法	/
M7 分组的具体方法	/	R17 主要指标检验的统计量值	/
M8 是否盲法及对谁“盲”	/	R18 主要指标检验的精确 P 值而不是大于或小于某界值	/
M9 实验和对照因素盲法效果的描述如外观、剂量、用法、疗程等	/	R19 对引言的假设做接受或拒绝的决定	/
M10 实施者和实验过程可比性的说明如术者经验、个体化干预	/	R20 负性反应或事件的各组人数、次数、性质、程度及统计分析	/
M11 研究的单位,如人、肿瘤、眼……	/	R21 计划内多重比较的具体方法	/
M12 效果评价的主要指标	/	R22 图示是否符合制图原则(图形性质、坐标刻度、变异度显示等)	/
M13 主要指标的测量方法与精确度	/	R23 “ $a \pm b$ ”形式中 b 有无明确标记?	/
M14 负性反应或事件的测量范围与方法	/	R24 比率中分母清楚吗?	/
M15 数据收集的方法与质量保证措施	/	D1 与引言对应,说明本研究的性质	/
M16 个体观察终点与整体研究终点的定义	/	D2 对主要指标结果的临床结论或生物医学解释	/
M17 控制可能偏倚的努力如混杂变量	/	D3 对设计中可能存在偏倚的说明	/
M18 统计学方法使用的软件及版本	/	D4 比较利弊,得出总的临床性结论	/
M19 对主要指标拟行比较的统计学方法	/	D5 临床结论的适用性/外推性说明	/
M20 对主要指标拟行单侧还是双侧检验,若单侧检验则其理由	/	D6 结合其它文献加强或平衡本文结论	/
M21 对主要指标进行检验的 α 水平	/		