

国际临床研究相关报告标准概述

廖 星 谢雁鸣 申 浩

摘要 设计合理并执行正确的临床研究,无论哪种证据级别,都可以为临床实践的干预措施提供有效的证据。但如果是设计和报告质量低的临床研究会得出偏倚的结果,从而误导临床医生的决策。因此,研究者只有准确、充分地报告临床试验的设计、实施和分析,才可经得起研究质量的严格评价,才能得出令人信服的研究结果。本文旨在简析目前国际临床研究范围里针对各种临床研究或系统综述的报告标准,为中医药临床研究实践提供参考标准。

关键词 临床研究;报告标准

An Overview of Reporting Standards for a Series of Systematic Reviews of International Clinical Trials LIAO Xing, XIE Yan-ming, and SHEN Hao *Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing (100700), China*

ABSTRACT Any kind of clinical study can supply clinical decision-making with useful information when they are designed and reported properly. However, results with bias or mistake can misguide doctors' decisions, if trials with poor design quality were reported wrongly. Therefore, researchers should report trials accurately and critically, so that the quality of trials could reach a high level and get trustworthy results. This article made an overview of many kinds of reporting standards about systematic reviews and all kinds of trials. We hope these international published standards could be helpful for Chinese medicine researchers in their practice.

KEYWORDS clinical research; reporting standard

国内外已有许多研究表明目前临床研究报告质量整体水平不高,而中医临床研究的报告质量更是成为中医临床科研中的一块短板。方法学专家深知不恰当、不充分、不准确的报告及试验设计将使评价结果产生偏倚。这种偏差将严重损害研究结果的真实性和可靠性。为此国际上相关科学家和编辑们共同制定了一系列用来规范临床研究报告的统一标准。如针对干预性研究的系统综述/Meta 分析报告、针对观察性研究的系统综述/Meta 分析报告、针对临床随机对照试验(Randomized Controlled Trials, RCT)、针对临床观察性研究、针对非随机对照试验等都有相应的国际报告标准和规范。此外还有英国国家医疗体系中公共卫生医疗办法委员会主导的“严格评价技术研究计划”,开发了包括有:随机对照试验、系统综述、队列研

究、病例对照研究、定性研究、经济学研究、诊断试验、临床预测准则、定性研究等多种研究类型的报告标准。

目前国际上许多优秀期刊已经采用了这一系列报告标准。报告标准的出台,为临床研究者提高临床试验报告质量提供了基本原则,使得评价和解释临床研究结果更为方便和规范。以下将对不同类型的研究报告标准做一概述。

1 系统综述/Meta 分析报告的规范

关于系统综述报告质量的评估标准,目前国际上主要有 4 种可参考的评价工具:英国国家医疗体系(NHS)公共卫生资源部发布的“严格评价技巧研究计划”(Critical Appraisal Skills Program, CASP)中的系统综述方法学评估清单^[1];加拿大多伦多大学循证医学中心发布的系统综述清单^[2];英国伯明翰大学激进情报研究所(Aggressive Research Intelligence Facility, ARIF)开发了评价系统综述的 ARIF 清单^[3];荷兰 EMGO 医疗研究机构开发的多种系统综述评价工具:AMSTAR-37 条(Measurement Tool for the Assessment of Multiple Systematic Reviews, AMSTAR)^[4]。

基金项目:中国中医科学院第六批自主选题(No.Z0215);2012 年度留学人员科技活动项目(No.Z0241)

作者单位:中国中医科学院中医临床基础医学研究所(北京 100700)

通讯作者:谢雁鸣, Tel:010-64014411 转 3302, E-mail: zhi-nanb2012@yahoo.com.cn

为了提高 Meta 分析报告质量,国际上也先后出台了相关的报告规范。如主要针对干预性研究、随机对照试验的 Meta 分析报告标准,就有由英国约克大学开发的标准,即:1999 年 QUOROM 标准(Quality of Reporting Meta-analyses)^[5]和 2009 年的 PRISMA 标准(Preferred Reporting Items for Systematic Reviews and Meta-Analyses)^[6]。QUOROM 制订委员会从 6 个方面规范了 Meta 分析报告标准:检索证据、关于纳入证据的决策、描述原始研究的特征、定量数据的合并、可靠性和与内在真实性(即质量)相关的问题及与外在真实性(即可推广性)相关的临床意义。最终 QUOROM 标准包括一个 21 个项目的清单及 Meta 分析过程中文献检索与筛选的流程图。在 QUOROM 基础上,PRISMA 制订委员会对内容进行了修订并扩充(<http://www.prisma-statement.org/>)。PRISMA 清单包括 4 个阶段的流程图和 7 个方面的规范,由 27 个条目组成:标题,结构式摘要,引言(基本原理、目的),方法(方案与注册、纳入与排除标准、信息来源、检索策略、研究的选择、数据收集过程、数据项、单个研究偏倚的风险、结局指标、结果合成、不同研究之间的偏倚风险、附加分析),结果(研究选择、研究特征、研究中的偏倚风险、单个研究的结果、结果合成、不同研究之间的偏倚风险、附加分析),讨论(证据小结、局限性、结论),资助情况。此外,针对 Cochrane 系统综述(主要为随机对照试验类),还有专门的 Cochrane Handbook 用于评价其规范性^[7]。

针对非干预性的观察性研究的系统综述/Meta 分析报告规范则来自 2000 年颁布的 MOOSE(Meta-analysis of Observational Studies in Epidemiology)规范^[8]。该规范包括 6 个部分,35 个条目:(1)研究的背景:定义研究问题、设定研究假设、确定研究结局、定义好干预暴露/干预措施、研究设计、研究人群的选择等;(2)文献检索策略:规定检索人员的资质、制定检索策略、如纳入研究的时间段和将使用的关键词、尽可能获得所有文献、将要检索的数据库、手工检索、纳入和排除标准、发表语种的处理、交代对仅有摘要文献和未发表文献的处理方法等;(3)研究方法:纳入研究是否符合适用于评估研究假设、文献纳入的合理性、文献处理过程的存档、偏倚的处理、研究质量的评价、统计学方法的描述、异质性的评估、图表的合理使用等;(4)研究结果:用图/表表示单个研究效应和所有研究合成后的总体效应、进行敏感性分析,如亚组分析、对不确定统计学结果的分析;(5)讨论:对偏倚的定量评估(如发表偏倚)、排除理由的陈述(如排除非英语

文献),对纳入研究进行质量评估;(6)结论:进一步考虑对于观察结果的其他解释、总结结论(如结果的可推性分析,可以是基于该文献系统综述范围内)、对未来研究的指导、标明资助来源。

以上这些有关系统综述/Meta 分析报告的规范为提高系统综述/Meta 分析报告的质量提供了保障,增强了报告的清晰性和条理性,使报告不至于遗漏重要信息,能最大程度地提供研究信息。国际上现已有许多著名的杂志包括 Lancet、JAMA 等,要求作者在投送系统综述/Meta 分析时,不但应当在论文中提供各种规范清单中的所有信息,还要注明清单中各条目对应于原稿的页数,稿件寄送时将论文原稿连同该核查单一并发给审稿人。

2 RCT 报告标准: CONSORT 声明与 CONSORT 声明修订版/扩展版

1996 年由杂志编辑、临床试验者和方法学家共同开会制定并发表了《随机对照试验报告规范》(the Consolidated Standards of Reporting Trials, CONSORT statement)即 CONSORT 声明(简称 CONSORT, <http://www.Consort-statement.org>)。CONSORT 由清单和流程图组成,即由 22 个条目清单和一个描述整个试验过程的流程图组成。22 个条目从属于 5 部分内容:文题和摘要、引言、方法、结果、讨论。文题和摘要部分仅 1 个条目:如何随机;引言部分有 1 个条目:科学背景和原理解释;方法部分有:受试者、干预、目的、测量指标、样本量、随机化(序列产生、分配隐藏、实施、盲法、统计方法);结果包括:受试者流程、募集受试者、基线资料、分析的人数、结果和估计值、辅助分析、不良事件;讨论包括:解释、可推广性、全部证据。

CONSORT 最初是用于报告“标准化”的两组平行设计试验。其本身属于 1 个开创性的报告规范,尚存在很多不足之处,其内容并未涵盖临床研究的所有细节,如伦理委员会同意、患者知情同意、资料安全性及监察、资金来源等。对安全性评估的内容仅有 1 条涉及,此外还有诸如成本效益分析和生活质量评估也未涉及。因此,自 1996 年以来,CONSORT 一直在不断更新和修订之中。根据不同特点的随机试验,按照 3 大类别进行了分类:不同设计类别,如整群随机、非劣效和等效试验、实用型随机对照试验;不同干预类别:草药干预、非药物干预、针灸干预;不同数据特征:伤害类报告、摘要式报告。根据这些分类,2003 年诞生 CONSORT 声明扩展版:更好地报告随机试验中的危害(Better Reporting of Harms in Randomized Trials: An Exten-

sion of the CONSORT Statement)^[9]、2004 年发布整群随机对照试验报告规范 (CONSORT statement: extension to cluster randomised trials)、2006 年发布非劣效性和等效性随机对照试验的报告规范 (Reporting of non-inferiority and equivalence randomized trials: An extension of the CONSORT statement)^[10]、2006 年发布草药随机对照试验报告规范 (Reporting Randomized, Controlled Trials of Herbal Interventions: An Elaborated CONSORT Statement)、2008 年发布实用型随机对照试验报告规范 (Improving the reporting of pragmatic trials: an extension of the CONSORT statement)^[11]。2008 年发布非药物干预治疗随机对照试验报告规范^[12] 2010 年发布针灸干预随机对照试验报告规范^[13]。2008 年发布如何在杂志和会议上发表随机对照试验摘要的报告规范^[14]。对于最初版本的 CONSORT, 自 2001 年更新后最新修订版本为 2010 年的 CONSORT^[15]。

3 临床观察性研究报告标准 (strengthening the reporting of observational studies in epidemiology, STROBE)

由流行病学家、方法学家、统计学家、研究者和编辑组成的一个国际性合作小组共同起草, 主要目的是为观察性研究提供报告规范, 从而改进这类研究的质量 (<http://www.strobe-statement.org/>), STROBE 评价内容主要有 5 个方面: 题目和摘要、前言、方法、结果和讨论, 共计 22 个条目。其中题目和摘要 1 个条目, 前言部分包括背景/原理和目标 2 个条目; 方法部分包括研究设计、研究现场、研究对象、研究变量、测量、偏倚、样本大小、统计方法、计量变量和资助情况等 10 个条目; 结果部分包括研究对象、描述性资料、结局资料、主要结果和其他分析等 5 个条目; 讨论部分包括重要结果、局限性、可推广性和解释等 4 个条目。所有条目针对 3 种观察性研究类别: 病例对照研究、队列研究和横断面研究。有通用原则也有各自特性化的要求。

4 非随机对照试验报告规范: TREND 规范

2004 年: 美国疾病预防控制中心 (CDC) HIV/AIDS 综合防治研究 (PRS) 小组研究者及其 TREND 研发小组发布了 TREND (Transparent Reporting of Evaluations with Nonrandomized Designs)^[16] (<http://www.cdc.gov/trendstatement/>), 即用于评估非随机对照试验研究报告的规范。TREND 清单只适用于采用非随机设计的干预评价研究, 而非所有采用非随机设计的研究。该规范包涵 5 部分, 22 条清

单。标题和摘要部分含有 1 条条目; 介绍部分包含 1 条; 方法部分包括 9 条: 研究对象、干预、目标、样本大小、分配方法、盲法、分析单元、统计方法。结果部分包括 8 条: 研究对象的流动情况、研究对象招募情况、基线均衡、数据分析、结局和评价、辅助分析、不良反应事件; 讨论部分包括 3 条: 解释、可推广性、全部证据。

5 队列研究、病例系列研究、病例对照研究等其他类型研究的报告标准

英国国家医疗体系 (NHS) 中公共卫生医疗办法委员会 (Solutions for Public Health, SPH) 主导的严格评价技术研究计划 (Critical Appraisal Skills Program, CASP), 开发了一系列针对不同研究类型的评估工具, 该计划也是国际 CASP 网络中的一部分。其中涉及的研究类型有: 随机对照试验、系统综述、队列研究、病例对照研究、定性研究、经济学研究、诊断试验、临床预测规则研究 (clinical prediction rule, CPR)、定性研究。

举例来说队列研究, 评估以 3 大类问题 12 个小问题的方式进行^[17]。第一大问题: 研究结果是否真实有效? 具体涵盖了 7 个小问题: (1) 该研究是否明确了研究中的关键点? (2) 研究者是否使用了合适的研究方法回答研究问题? (3) 研究中队列的招募方式是否合理? (4) 暴露是否在偏倚最低时进行了精确的测量? (5) 结局是否在偏倚最低时进行了测量? (6) 研究者是否分辨出了所有重要的混杂因素? 研究者是否在设计或分析阶段考虑了这些混杂因素? (7) 所有研究对象的随访时间是否足够? 所有研究对象的随访时间是否足够长? 第二大问题: 研究结果有哪些? 具体涵盖了 3 个小问题: (1) 该研究的结果都有哪些? (2) 这些结果的精确度如何? 风险评估的精确度如何? (3) 作为读者, 你相信这些结果吗? 第 3 大问题: 研究结果会对我有类似作用吗? 具体涵盖了 2 个小问题: (1) 这些结果是否可以应用到一定范围的人群? (2) 该研究结果是否和其他可获得证据类似?

对于病例对照研究, 评估也以 3 大问题和 11 个小问题的方式进行^[18]。第一大问题: 研究结果是否真实有效? 具体涵盖了 6 个小问题: (1) 该研究是否明确了研究中的关键点? (2) 研究者是否使用了合适的研究方法回答研究问题? (3) 研究中病例的纳入方式是否合理? (4) 对照组的选择方式是否合理? (5) 暴露是否在偏倚最低时进行了精确的测量? (6) 研究者都列出了那些混杂因素? 研究者是否在设计或分析阶段考虑了这些混杂因素? 第二大问题: 研究结果有哪些? 具体涵盖了 3 个小问题: (1) 该研究的结果都

有哪些? (2) 这些结果的精确度如何? 风险评估的精确度如何? (3) 作为读者, 你相信这些结果吗? 第三大问题: 研究结果会对我有类似作用吗? 具体涵盖了 2 个小问题: (1) 这些结果是否可以应用到一定范围的人群? (2) 该研究结果是否和其他可获得证据类似?

6 个案报告规范 有关临床不良反应/不良事件 (ADR/AE) 报告规范

牛津大学学者 2003 年在英国医学杂志 (BMJ) 发表文章, 建议在医学期刊上规范化发表西药 ADR 报告^[19,20]。2007 年由国际药物流行病学组织 (The International Society for Pharmacoepidemiology, ISPE) 和国际药物警戒协会 (The International Society of Pharmacovigilance, ISoP) 就如何进行不良事件报告发布了一份指南^[21]。指南包含了 5 部分内容: 标题, 患者, 药物, 不良事件, 讨论。其中患者部分, 又有 5 个方面的内容: 人口学资料、目前健康状态、用药史、体格检查、患者处境。药物部分包括 5 个方面的内容: 具体的药物名称、剂量、医嘱情况、药物-反应发生的时间间隔、合并用药。不良事件: 描述情况。讨论: 因果关系的判断及其相关证据。

我国虽然已建立了较为完整的 ADR 监测体系, 实行可疑即报的政策, 由 ADR 监测中心鉴定其性质, 但 AE 上报法规是一个管理法规, 其报告过程却并没有公开。基于此, 国内有学者做了对《中药不良反应/实践报告规范建议》的研究^[22]。

7 诊断性试验报告规范

目前有关诊断性试验报告的规范有: 发起于 1999 年的 STARD (standards for reporting of diagnostic accuracy) 规范^[23], 该规范包含 25 个条目和 1 个流程图 (<http://www.stard-statement.org/>)。还有 2000 年颁布的医疗检测诊断性精确性报告规范-33 条清单^[24]。

8 卫生经济学研究报告规范

2005 年由国际药物经济学和结局研究协会 (International Society for Pharmacoeconomics and Outcomes Research, ISPOR) 特许良好研究实践特别小组: 随机对照试验——成本效果分析, 进行临床试验中有关成本效果分析报告的规范撰写^[25]。该规范主要包括了 5 个方面的问题: 试验设计, 数据选择, 数据库的设计和管理, 分析, 结果报告。还有 1996 年 BMJ 杂志 BMJ 卫生经济学评估工作组制定的“BMJ 告作者和审评专家有关经济学评价报告指南”^[26], 以及 2008 年发布的有关血友病经济学评估报告建议书^[27], 都在不同方面对临床研究中有关卫生经济学研

究报告规范提出了好的建议。

9 统计分析报告规范

目前国际上主要有如下系列统计分析报告规范可供研究者参考使用: (1) 贝叶斯统计方法在医学研究领域应用中的分析和评价标准^[28,29]; (2) 缺失数据的分析标准^[30]; (3) 亚组分析报告规范^[31]; (4) 异质性分析和评价原则^[32]; (4) 多水平模型分析要素^[33]; (5) 多变量 logistic 回归模型分析评估和报告建议^[34]; (6) 病例对照设计中统计分析结果解释要点^[35]。

此外, 还有许多其他医学研究领域的研究报告规范, 如遗传关联研究报告规范 (STrengthening the REporting of Genetic Association Studies, STREGA)、遗传风险预测研究报告规范 (Strengthening the reporting of Genetic Risk Prediction Studies, GRIPS) 等等。

总之, 准确报告设计方案、实施过程和结果测量方法, 是正确评估中医药临床试验结果真实性和证据强度的基础。这对指导中医临床实践和卫生政策决策, 合理配置和使用医疗资源有重要意义。为了能够规范地报告研究, 也为了能够使研究更加国际化, 上述各种报告规范绝大多数在国内杂志上都有中文版本, 我们倡导中医药研究者在以后的研究中参考以上这些国际规范, 力争让中医药临床研究的结果能够获得更为广泛的认可。

参 考 文 献

- [1] http://www.sph.nhs.uk/sph-files/S_Reviews%20Appraisal%20Tool.pdf [OL].
- [2] <http://ktclearinghouse.ca/cebm/teaching/worksheets/sr> [OL].
- [3] <http://www.arif.bham.ac.uk/critical-appraisal-checklist.shtml> [OL].
- [4] Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews [J]. BMC Med Res Methodol, 2007, 7 (10): 1-7.
- [5] Moher D, Cook DI, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUOROM Statement, Quality of Reporting of Meta-analyses [J]. Lancet, 1999, 354 (9193): 1896-1900.
- [6] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statements [J]. PLoS Med, 2009, 6 (7): e1000097.
- [7] <http://www.cochrane-handbook.org/> [OL].
- [8] Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology:

- a proposal for reporting [J]. *JAMA*, 2000, 283 (15): 2008–2012.
- [9] Ioannidis JP, Evans SJ, Gotsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement [J]. *Ann Intern Med*, 2004, 141(10): 781–788.
 - [10] Piaggio G, Elbourne DR, Altman DG, et al. Reporting of non-inferiority and equivalence randomized trials: an extension of the CONSORT statement [J]. *JAMA*, 2006, 295: 1152–1160.
 - [11] Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement [J]. *BMJ*, 2008, 337: a2390.
 - [12] Boutron I, Moher D, Altman DG, et al. Extending the CONSORT Statement to randomized trials of non-pharmacologic treatment: explanation and elaboration [J]. *Ann Intern Med*, 2008; 295–309.
 - [13] MacPherson H, Altman DG, Hammerschlag R, et al. Revised Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement [J]. *PLoS Med*, 2010, 7(6): e1000261.
 - [14] Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration [J]. *PLoS Med*, 2008, 5(1): e20.
 - [15] Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials [J]. *Ann Int Med*, 2010, 152(11): 1–8.
 - [16] <http://www.cdc.gov/trendstatement/> [OL].
 - [17] <http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/cohort%2012%20questions> [OL].
 - [18] <http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/Case%20Control%2011%20Questions.pdf> [OL].
 - [19] Aronson JK. Anecdotes as evidence [J]. *BMJ*, 2003, 326(7403): 1346.
 - [20] Aronson JK. Publishing histories of adverse reactions to medicaments anecdotally: the PHARMA guidelines for reporting suspected adverse drug reactions [J]. *Drug Saf*, 2008, 31(4): 355–356.
 - [21] Kelly WN, Arellano FM, Barnes J, et al. Guidelines for submitting adverse event reports for publication pharmacoepidemiology and drug safety [J]. 2007, 16: 581–587.
 - [22] 吴泰相, 商洪才, 卞兆祥, 等. 中药不良反应/不良事件报告规范建议 [J]. *中国循证医学杂志*, 2012, 10(2): 215–219.
 - [23] Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for reporting of diagnostic accuracy [J]. *Clin Chem*, 2003, 49: 1–6.
 - [24] Bruns DE, Huth EJ, Magid E, et al. Toward a checklist for reporting of studies of diagnostic accuracy of medical tests [J]. *Clin Chem*, 2000, 46: 893–895.
 - [25] Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report [J]. *Value Health*, 2005, 8: 521–533.
 - [26] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party [J]. *BMJ*, 1996, 313: 275–283.
 - [27] Nicholson A, Berger K, Bohn R, et al. Recommendations for reporting economic evaluations of haemophilia prophylaxis: a nominal group consensus statement on behalf of the Economics Expert Working Group of the International Prophylaxis Study Group [J]. *Haemophilia*, 2008, 14: 127–132.
 - [28] Spiegelhalter DJ, Myles JP, Jones DR, et al. Bayesian methods in health technology assessment: a review [J]. *Health Technol Assess*, 2000, 4(38): 130–131.
 - [29] Sung L, Hayden J, Greenberg ML, et al. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study [J]. *J Clin Epidemiol*, 2005, 58: 261–268.
 - [30] Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls [J]. *BMJ*, 2009, 338: b2393.
 - [31] Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine-reporting of subgroup analyses in clinical trials [J]. *N Engl J Med*, 2007, 357(21): 2189–2194.
 - [32] Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal [J]. *Trials*, 2010, 11: 85.
 - [33] Jackson DL. Reporting results of latent growth modeling and multi-level modeling analyses: some recommendations for rehabilitation psychology [J]. *Rehabil Psychol*, 2010, 55(3): 272–285.
 - [34] Kalil AC, Mattei J, Florescu DF, et al. Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature [J]. *Am J Transplant*, 2010, 10(7): 1686–1694.
 - [35] Crawford JR, Garthwaite PH, Porter S. Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards [J]. *Cogn Neuropsychol*, 2010, 27(3): 245–260.

(收稿:2012-07-09 修回:2013-01-07)