

CONSORT 2010 说明与详述:报告平行对照随机临床试验指南的更新

David Moher¹, Sally Hopewell², Kenneth F Schulz³, Victor Montori⁴, Peter C Gøtzsche⁵, P J Devereaux⁶, Diana Elbourne⁷, Matthias Egger⁸, Douglas G Altman²

1. Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa Hospital, Ottawa, Ontario, Canada, K1H 8L6
2. Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford, UK
3. Family Health International, Research Triangle Park, NC 27709, USA
4. UK Knowledge and Encounter Research Unit, Mayo Clinic, Rochester, MN, USA
5. The Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, Copenhagen, Denmark
6. McMaster University Health Sciences Centre, Hamilton, Canada
7. Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK
8. Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

摘要:

大量证据显示随机对照临床试验(randomised controlled trial, RCT)的报告质量不理想。报告不透明,则读者既不能评判试验结果是否真实可靠,也不能从中提取可用于系统综述的信息。最近的方法学分析表明,报告不充分和设计不合理与对治疗效果产生评价偏倚有关。这种系统误差对 RCT 损害严重,而 RCT 正是以其能减少或避免偏倚而被视为评价干预措施的金标准。

为了提高 RCT 的报告质量,一个由专家和编辑组成的工作组制定了临床试验报告的统一标准(Consolidated Standards of Reporting Trials, CONSORT)声明。CONSORT 声明于 1996 年首次发表,并于 2001 年更新。声明由对照检查清单和流程图组成,供作者在报告 RCT 时使用。许多核心医学期刊和主要国际性编辑组织都已认可 CONSORT 声明。该声明促进了对 RCT 的严格评价和解释。

2001 年,在对 CONSORT 进行修订时,人们就已经清楚地认识到,解释和说明制定 CONSORT 声明的原理,有助于研究人员等撰写或评价临床试验报告。一篇 CONSORT 说明与详述文章于 2001 年同 2001 版 CONSORT 声明一起发表。

2007 年 1 月的专家会议之后,对 CONSORT 声明作了进一步修订并已发表,即“CONSORT 2010 声明”。这次更新对原版对照检查清单作了文字上的修改,使其更为明晰,并收入了与一些新近才认识到的主题相关的建议,如选择性报告结局产生的偏倚。

说明与详述文件旨在加强人们对 CONSORT 声明的理解、应用和传播,这次也作了大量修订,对每一项新增或更新的清单条目的含义和增改理由进行了解释,提供了优秀的报告实例,还尽可能地提供了相关的经验性研究的参考文献。文中收入了若干流程图实例。

“CONSORT 2010 声明”、其说明与详述文件,以及相关网站(www.consort-statement.org),对于改进随机临床试验报告必将有所裨益。

关键词: 临床试验; 随机对照试验; 指南; CONSORT

Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. *J Chin Integr Med.* 2010; 8(8): 701-741.

Received July 21, 2010; accepted July 29, 2010; published online August 15, 2010.

Indexed/abstracted in and full-text link-out at PubMed. Journal title in PubMed; *Zhong Xi Yi Jie He Xue Bao.*

Free full text (HTML and PDF) is available at <http://www.jcimjournal.com>.

Forward linking and reference linking via CrossRef.

DOI: 10.3736/jcim20100801

Open Access

CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials (Chinese version)

David Moher¹, Sally Hopewell², Kenneth F Schulz³, Victor Montori⁴, Peter C Gøtzsche⁵, P J Devereaux⁶, Diana Elbourne⁷, Matthias Egger⁸, Douglas G Altman²

1. Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa Hospital, Ottawa, Ontario, Canada, K1H 8L6
2. Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford, UK
3. Family Health International, Research Triangle Park, NC 27709, USA
4. UK Knowledge and Encounter Research Unit, Mayo Clinic, Rochester, MN, USA
5. The Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, Copenhagen, Denmark
6. McMaster University Health Sciences Centre, Hamilton, Canada
7. Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK
8. Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

Abstract:

Overwhelming evidence shows the quality of reporting of randomised controlled trials (RCTs) is not optimal. Without transparent reporting, readers cannot judge the reliability and validity of trial findings nor extract information for systematic reviews. Recent methodological analyses indicate that inadequate reporting and design are associated with biased estimates of treatment effects. Such systematic error is seriously damaging to RCTs, which are considered the gold standard for evaluating interventions because of their ability to minimise or avoid bias.

A group of scientists and editors developed the CONSORT (Consolidated Standards of Reporting Trials) statement to improve the quality of reporting of RCTs. It was first published in 1996 and updated in 2001. The statement consists of a checklist and flow diagram that authors can use for reporting an RCT. Many leading medical journals and major international editorial groups have endorsed the CONSORT statement. The statement facilitates critical appraisal and interpretation of RCTs.

During the 2001 CONSORT revision, it became clear that explanation and elaboration of the principles underlying the CONSORT statement would help investigators and others to write or appraise trial reports. A CONSORT explanation and elaboration article was published in 2001 alongside the 2001 version of the CONSORT statement.

After an expert meeting in January 2007, the CONSORT statement has been further revised and is published as the CONSORT 2010 Statement. This update improves the wording and clarity of the previous checklist and incorporates recommendations related to topics that have only recently received recognition, such as selective outcome reporting bias.

This explanatory and elaboration document — intended to enhance the use, understanding, and dissemination of the CONSORT statement — has also been extensively revised. It presents the meaning and rationale for each new and updated checklist item providing examples of good reporting and, where possible, references to relevant empirical studies. Several examples of flow diagrams are included.

The CONSORT 2010 Statement, this revised explanatory and elaboration document, and the associated website (www.consort-statement.org) should be helpful resources to improve reporting of randomised trials.

Keywords: clinical trials; randomized controlled trials; guidelines; CONSORT

“整个医学都有赖于临床试验报告的透明化。”^[1]

设计良好又实施得当的随机对照临床试验 (randomised controlled trial, RCT) 能为医疗干预措施的有效性提供最可靠的证据,但如果临床试验存在方法学的不足,就会产生偏倚,尤其会夸大治疗效果^[2-5]。设计和报告质量低劣的临床试验所得出的有偏倚的结果会误导从单个患者的治疗到国家公共卫生政策的制定等各级医疗卫生决策。

只有当作者在 RCT 报告中对试验的设计、实施和分析作出全面而准确的描述时,才有可能对试

验的质量进行严格评价。除了远未达到透明化之外,RCT 报告还常常缺乏完整性^[6-9],更使因方法学质量低劣而产生的诸多问题复杂化^[10-15]。

1 不完整和不精确的报告

诸多综述都分析过临床试验报告的不足之处。例如,2000 年 PubMed 收录的 519 篇临床试验报告中,只有 21% 报告了在试验中使用何种方法将受试者分配入各对照组^[16],而 2006 年收录的 616 篇中也只有 34% 对此进行了报告^[17]。同样,2000 年 PubMed 收录的临床试验报告只有 45%^[16],2006 年

只有 53%^[17]对主要终点指标作出了定义; 2000 年只有 27%, 2006 年只有 45% 报告了样本量的计算方法。试验报告不仅经常缺乏完整性, 有时还不够准确。在 119 篇声称将所有最初分配入组的受试者都纳入了分析[意向性治疗分析(intention-to-treat analysis)]的报告中, 实际上有 15 篇(13%)剔除了部分病例, 或者未将最初分配的全部病例纳入分析^[18]。很多其他综述发现, 报告不充分的情况普遍存在于专业性期刊^[16, 19]和非英语期刊^[20, 21]。

恰当的随机方法可以减少试验入组时的选择性偏倚, 也是高质量 RCT 的关键要素^[22]。成功的随机方法取决于两个步骤: 产生不可预测的分配顺序, 对负责招募受试者的研究人员隐藏该顺序^[2, 23](见框 1)。

遗憾的是, 尽管分配受试者干预措施的方法如此重要, 但对其报告通常也是不充分的。例如, 妇产科期刊发表的 206 篇据称是 RCT 的报告中, 5% 所描述的研究没有真正做到随机^[23]。这仅仅是保守的估计, 因为目前大多数报告都未提供足够的有关分配方法的信息^[20, 23, 30-33]。

2 改进 RCT 的报告: CONSORT 声明

DerSimonian 及其同事提出“编辑向作者提供一份他们希望作者报告时严格遵守的项目列表可以极大地改进临床试验的报告”^[34]。早在 20 世纪 90 年代, 两个由期刊编辑、临床试验专家和方法学专家组成的研究小组各自发表了报告临床试验的建议^[35, 36]。Rennie 在随后发表的一篇评论中敦促这两个小组一起开会拟定一套统一的建议^[37], 结果便产生了临床试验报告的统一标准(Consolidated Standards of Reporting Trials, CONSORT)声明^[38]。

CONSORT 声明(或简称 CONSORT)由报告 RCT 必备的基本项目清单和描述整个试验过程中受试者流程的流程图组成, 主要针对的是两组平行设计的 RCT 报告。CONSORT 中的大多数条目也与很多其他设计类型的临床试验相关, 如非劣效性试验、等效性试验、析因设计试验、群组试验, 以及交叉设计试验等。报告上述部分设计类型试验的 CONSORT 扩展版已经发表^[39-41]。另外, 已经发表的还有报告危害^[42]、非药物治疗^[43]、草药干预^[44], 以及摘要^[45]等的 CONSORT

框 1 治疗方法分配。随机方法有何特别之处?

分配干预措施给受试者的方法是临床试验设计的关键。随机分配是首选的方法, 该方法 50 多年来经常被成功地应用于临床试验中^[24]。随机方法有三大优点^[25]。第一, 如果使用恰当, 可以消除选择性偏倚, 在分配治疗时均衡已知和未知的预后因素。如果不用随机方法, 可能会因为有意或无意地选择特定类型的受试者接受某种特定的治疗而使对治疗结果的比较发生偏差。第二, 随机分配允许使用概率理论来表达仅由于机遇(chance)造成接受不同于干预因素的两组之间结局差异的可能性大小^[26]。第三, 在有些情况下, 随机分配有助于将治疗方法对研究人员、受试者和评估者设盲, 如使用安慰剂, 以减少分配治疗后产生的偏倚^[27]。在这三大优点中, 减少试验入组时的选择性偏倚往往是最重要的^[28]。

实际应用时, 随机方法的成功有赖于两个互为关联的方面: 确保产生不可预测的分配顺序和隐藏分配序列直至完成分配^[2, 23]。关键之处在于, 参与将受试者分配入各组的人是否知道或可以预测这个分配顺序^[29]。因此, 应该建立治疗分配系统以使负责招募受试者的人不能预知下一个人将接受何种治疗, 这个过程称为分配隐藏^[2, 23]。正确的分配隐藏可以防止有人知道下一个分配, 而恰当的随机序列则能防止根据之前的分配情况准确推测将来的分配。

扩展版。

CONSORT 的目的是指导作者如何提高其临床试验报告的质量。临床试验的报告需要清晰、完整和透明。读者、审稿人和编辑还可以利用 CONSORT 来帮助评估和解释 RCT 报告。但 CONSORT 不是用于质量评价的工具, 其内容更多地着眼于那些与临床试验的内部和外部真实性相关的条目。很多 CONSORT 没有明确提到的项目在报告中也应该得以体现, 如有关伦理委员会的批准, 如何获取受试者的知情同意, 甚或是否有数据安全性和监察委员会等方面的信息。此外, 试验的其他方面只要文中提到了就必须准确恰当地报告, 例如与成本效益分析相关的信息等^[46-48]。

自 1996 年发表以来, CONSORT 已经得到 400 多种期刊(见 www.consort-statement.org)和若干编辑组织如国际医学期刊编辑委员会的拥护^[49]。RCT 报告质量的提高与期刊是否采用 CONSORT 有关^[17, 50, 51]。不过, CONSORT 是一项持续性的工作, CONSORT 声明也在定期修订^[3], 最近一次修订是在 9 年前的 2001 年^[52-54]。自那以后, 支撑 CONSORT 的证据基础的积累极为可观, 经验性的

资料提示,与报告 RCT 相关的一些新问题已经突显出来,如选择性报告结局^[55-57]。于是,2007 年 1 月在加拿大召开了一次 CONSORT 工作组会议,

目的在于修订 2001 年版 CONSORT 声明和与其配套的说明与详述文件。修订后的清单见表 1,对流程图(见图 1)未作修订^[52-54]。

表 1(待续) 随机临床试验应报告的信息 CONSORT 2010 对照检查清单*

论文章节/主题	条目号	对照检查的条目	报告页码
文题和摘要			
	1a	文题能识别是随机临床试验	_____
	1b	结构式摘要,包括试验设计、方法、结果、结论几个部分(具体的指导建议参见“CONSORT for abstracts” ^[46, 65])	_____
引言			
背景和目的	2a	科学背景和对试验理由的解释	_____
	2b	具体目的或假设	_____
方法			
试验设计	3a	描述试验设计(诸如平行设计、析因设计),包括受试者分配入各组的比例	_____
	3b	试验开始后对试验方法所作的重要改变(如合格受试者的挑选标准),并说明原因	_____
受试者	4a	受试者合格标准	_____
	4b	资料收集的场所和地点	_____
干预措施	5	详细描述各组干预措施的细节以使他人能够重复,包括它们实际上是在何时、如何实施的	_____
结局指标	6a	完整而确切地说明预先设定的主要和次要结局指标,包括它们是在何时、如何测评的	_____
	6b	试验开始后对结局指标是否有任何更改,并说明原因	_____
样本量	7a	如何确定样本量	_____
	7b	必要时,解释中期分析和试验中止原则	_____
随机方法:			
序列的产生	8a	产生随机分配序列的方法	_____
	8b	随机方法的类型,任何限定的细节(如怎样分区组和各区组样本多少)	_____
分配隐藏机制	9	用于执行随机分配序列的机制(例如按序编码的封藏法),描述干预措施分配之前为隐藏序列号所采取的步骤	_____
实施	10	谁产生随机分配序列,谁招募受试者,谁给受试者分配干预措施	_____
盲法	11a	如果实施了盲法,分配干预措施之后对谁设盲(例如受试者、医护提供者、结局评估者),以及盲法是如何实施的	_____
	11b	如有必要,描述干预措施的相似之处	_____
统计学方法	12a	用于比较各组主要和次要结局指标的统计学方法	_____
	12b	附加分析的方法,诸如亚组分析和校正分析	_____
结果			
受试者流程(极力推荐使用流程图)	13a	随机分配到各组的受试者例数,接受已分配治疗的例数,以及纳入主要结局分析的例数	_____
	13b	随机分组后,各组脱落和被剔除的例数,并说明原因	_____
招募受试者	14a	招募期和随访时间的长短,并说明具体日期	_____
	14b	为什么试验中断或停止	_____
基线资料	15	用一张表格列出每一组受试者的基线数据,包括人口学资料和临床特征	_____
纳入分析的例数	16	各组纳入每一种分析的受试者数目(分母),以及是否按最初的分组分析	_____
结局和估计值	17a	各组每一项主要和次要结局指标的结果,效应估计值及其精确性(如 95% 可信区间)	_____
	17b	对于二分类结局,建议同时提供相对效应值和绝对效应值	_____
辅助分析	18	所做的其他分析的结果,包括亚组分析和校正分析,指出哪些是预先设定的分析,哪些是新尝试的分析	_____
危害	19	各组出现的所有严重危害或意外效应(具体的指导建议参见“CONSORT for harms” ^[42])	_____

表 1(续) 随机临床试验应报告的信息 CONSORT 2010 对照检查清单*

论文章节/主题	条目号	对照检查的条目	报告页码
讨论			
局限性	20	试验的局限性,报告潜在偏倚和不精确的原因,以及出现多种分析结果的原因(如果有这种情况的话)	_____
可推广性	21	试验结果被推广的可能性(外部可靠性,实用性)	_____
解释	22	与结果相对应的解释,权衡试验结果的利弊,并且考虑其他相关证据	_____
其他信息			
试验注册	23	临床试验注册号和注册机构名称	_____
试验方案	24	如果有的话,在哪里可以获取完整的试验方案	_____
资助	25	资助和其他支持(如提供药品)的来源,提供资助者所起的作用	_____

* 我们极力推荐结合“CONSORT 2010 说明与详述”阅读本声明,该文件对全部条目作了详细阐述。我们还推荐必要时阅读关于群组随机试验^[40]、非劣效性和等效性试验^[39]、非药物治疗^[43]、草药干预^[44]以及实效性试验^[41]的各种 CONSORT 扩展版。其他扩展版即将面世。与本清单有关的各种扩展版及最新参考资料,请见 www.consort-statement.org。

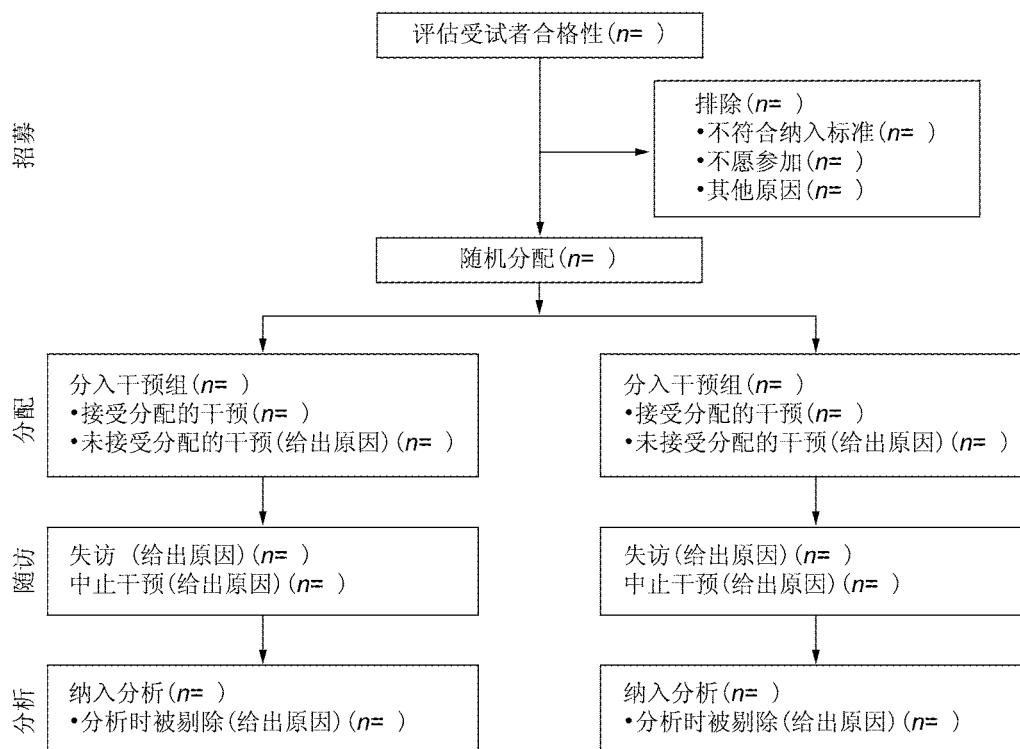


图 1 两组平行对照随机临床试验各阶段(招募受试者、分配干预措施、随访和数据分析)进程的流程图^[52-54]

3 CONSORT 2010 声明:说明与详述

对指导制定 CONSORT 声明的原理的详细说明与阐释有助于研究人员等撰写或评价临床试验报告,这一点在 2001 年修订 CONSORT 时就已经很清楚了。CONSORT 说明与详述^[58]于 2001 年同 2001 版 CONSORT 声明一起发表,讨论了每一项条目的报告理由和科学背景,并提供了已发表的报告范例。修订 CONSORT 说明与详述的理由与上文所述的修订 CONSORT 声明的理由类似。下面我们简要介绍一下新版说明与详述对原版所作的主要增删。

4 CONSORT 2010 说明与详述:改动之处

我们对原版 CONSORT 说明与详述作了几处实质性的改动和一些字面上的修改(详细内容在 2010 版 CONSORT 声明中都已特别指出^[59])。其中有些改动是因为 CONSORT 清单已作了改动,CONSORT 2010 清单增加了 3 项新条目,如条目 24 要求作者报告在何处可以获取临床试验设计方案。我们还更新了一些原有的解释,如补充了方法学证据的新近参考文献,选用了一些更为合适的实例。我们删除了原有的词汇表,该词汇表现在可以从 CONSORT 网站(www.consort-statement.org)下载。

我们尽可能地介绍了相关经验性研究的成果,很多有关临床试验的优秀书籍对方法学问题有更为全面的讨论^[60-62]。最后,尽管我们知道 RCT 所评价的干预措施不全是治疗方法,受试者也不全是患者,但为了方便起见,我们有时还是使用“治疗方法”和“患者”这两个词。

5 清单条目

5.1 文题和摘要

条目 1a. 文题能识别是随机临床试验

举例——“用尼古丁口腔吸入器减少吸烟:有效性和安全性的双盲随机临床试验”^[63]。

解释——在电子数据库中能否识别出某篇报告是随机临床试验很大程度上取决于该文献是如何被标引的。如果某篇报告的作者没有明确地报告是否系随机临床试验的信息,索引编制者可能不会将其归类为随机临床试验^[64]。为了确保某项研究被恰当地标引,并且易于识别,作者应该在文章标题中使用“随机”一词表明受试者是被随机分配到所比较的各组中。

条目 1b. 结构式摘要,包括试验设计、方法、结果、结论几个部分(具体的指导建议参见“CONSORT for abstracts”^[45, 65])

解释——清晰、透明且足够详细的摘要极为重要,因为读者经常根据这些信息作出他们对某项临床试验的评价。有些读者将摘要用作筛选工具来决定是否阅读全文。更要紧的是,由于不能免费阅读所有的临床试验报告,而且有些医疗工作者也无法获取临床试验报告的全文,所以医疗卫生决策有时候是根据随机临床试验的摘要作出的^[66]。

一篇期刊论文的摘要应该包含临床试验足够的信息,以便做到准确记录临床试验的实施过程和结果,并通过使用期刊规定的格式在有限的篇幅内提供临床试验的最佳信息。一篇结构清晰、撰写得当的摘要可以帮助人们快速评价相关的试验结果,帮助人们从电子数据库中检索出相关的试验报告^[67]。摘要应该准确反映全文的内容,不应当有正文中未提到过的信息。有研究对期刊论文摘要中报告的信息和正文报告的信息进行了比较,发现摘要中有与正文不一致或正文根本没有提到的表述^[68-71]。反过来,不在摘要中报告重要危害会严重误导他人对试验结果的解释^[42, 72]。

近来有一篇 CONSORT 声明的扩展版提供了作者在期刊或会议摘要中报告随机临床试验主要结果时应该包括的基本项目清单(见表 2)^[45]。我们极力推荐使用结构式摘要报告随机临床试验,针对试验设

计、实施、分析和解释分列几个小标题为读者提供临床试验的信息^[73]。一些研究发现,结构式摘要与较为传统的描述性摘要相比质量更高^[74, 75],更容易让读者找到所需信息^[76]。我们知道很多期刊有自己制定的摘要写作格式和字数限制,我们并不是想让它们改变格式,只是提出针对应该报告哪些信息的建议。

表 2 期刊报告随机临床试验时摘要应具备的项目

项目	描述
作者	通讯作者的详细联系信息
试验设计	试验设计的描述(例如平行试验、群组试验、非劣效试验等)
方法:	
受试者	受试者的合格标准和数据采集的场所
干预措施	各组拟实施的干预措施
目的	具体目的或假设
结局指标	清楚地说明该项研究的主要结局指标
随机方法	如何分配受试者进入各干预组
盲法(遮蔽)	是否将分组情况对受试者、医护提供者以及结局测评者设盲
结果:	
随机分组的例数	随机分入各组的受试者人数
招募受试者	临床试验状态
纳入分析的例数	每组纳入数据分析的受试者数目
结局	各组主要结局指标的结果及其效应估计值和精确性
危害	重要的不良事件或副作用
结论	对结果的概括性解释
试验注册	临床试验注册号和注册机构名称
资助	资助来源

5.2 引言

条目 2a. 科学背景和对试验理由的解释

举例——“外科手术是 I 期和 II 期非小细胞肺癌(non-small cell lung cancer, NSCLC)患者的治疗选择……一篇有关 NSCLC 的 meta 分析对比较手术和手术加以顺铂为基础的辅助化疗的 8 项随机临床试验结果的合并分析显示,5 年绝对生存率增加仅约 5%(从 50%提高到 55%),没有统计学意义($P=0.08$)。在设计该临床试验时(20 世纪 90 年代中期),辅助化疗还没有成为标准治疗在临床上推行……新辅助化疗在临床上应用的理由有三:能使原发肿瘤缩小,以便于后续的手术治疗,使手术变得简单,甚或可以不手术;检测不到的微转移可以在治疗之初就得到处理;有可能抑制因手术及术后伤口愈合而释放的生长因子刺激残存肿瘤的生长……本试验旨在比较可手术切除的 NSCLC 患者接受单纯手术治疗和术前加 3 个周期以铂类药物为基础的化疗在总生存率、生活质量、病理分期、肿瘤切除率、手术范围,以及复发的时间与部位等方面的差异。”^[77]

解释——典型的引言是,作者用流畅的文字解

释试验的科学背景和试验理由,并介绍试验的整体概况。引言中也可以包括试验目的(见条目 2b)。试验理由可以是解释性试验(如评价某种药物对肾功能的可能影响)或实效性试验(如比较两种治疗方法的益处和危害以便为其应用提供指导)。作者应该报告与试验中的阳性干预措施的利弊有关的证据,并且当干预措施如何起效不够清晰的时候,作者还应该作出恰当的解释^[78]。

“赫尔辛基宣言(Declaration of Helsinki)”指出,涉及人的生物医学研究必须以充分了解科学文献为基础^[79]。也就是说,将人置于不必要的研究危险之中是不道德的。有些临床试验后被发现本来没有必要做,因为对现有文献的系统综述已经或者本可以回答它们要研究的问题^[80, 81]。因此,应该在引言中交待开展新试验的必要性。最理想的是引用一篇以往类似试验的系统综述作为参考文献,或者说明确尚无此类试验^[82]。

条目 2b. 具体目的或假设

举例——“我们在本研究中检验了以下假设,即对初产分娩的积极处理会:(1)减少剖腹产率;(2)减少产程延长的发生率;(3)不影响产妇的满意度。”^[83]

解释——目的指的是试验拟回答的问题,常常与某种特殊的治疗性或预防性干预措施的效果有关。假设是为达到目的而事先设定以待检验的问题。假设比目的更具体,适合用严谨的统计学方法检验。实际上,目的和假设并非总是易于区分。大多数 RCT 报告都能提供关于试验目的和假设的充足信息^[84]。

5.3 方法

条目 3a. 描述试验设计(诸如平行设计、析因设计),包括受试者分配入各组的比例

举例——“本试验是在美国(41 个试验地点)进行的一项多中心、分层(6~11 岁和 12~17 岁,按 2:1 的比例非均衡随机分配)、双盲、安慰剂平行对照研究。”^[85]

解释——“设计(design)”一词常用来指试验各方面是如何安排的,但也有更狭义的解释。广义的试验设计包括很多具体的方面,如随机方法和盲法的细节,这些在 CONSORT 清单的其他条目中也有提到。我们在这里用来指试验类型如平行对照或析因设计,也用来指概念性的试验框架如优效设计或非劣效设计,还用来指清单的其他条目中未提到的其他相关问题。

CONSORT 声明主要针对将受试者逐个随机分配到两个“平行对照(parallel)”组的临床试验。

事实上,在所有已发表的临床试验中,这种设计仅仅刚过半数^[16]。其他主要设计类型有多组平行设计、交叉设计、群组设计^[40],以及析因设计。另外,大多数试验是为明确一项新的干预措施的优效性(假如确实有)而设计的,而其他则是为评价非劣效性或等效性而设计的^[39]。研究人员清晰地描述试验的这些方面是非常重要的,包括随机分配的单元(如患者、全科医生的诊疗、损伤等)。摘要中最好也要有这些细节(见条目 1b)。

如果用的是不常见的试验设计,作者最好说明为什么选用这种设计,因为这类设计往往可能意味着需要更大的样本量或更复杂的分析和解释。

尽管大多数试验采用均衡随机方法(如按 1:1 比例分为两组),还是应该尽可能说明分配的比例。对于药品的临床试验,还应该具体说明是哪一期试验(I~IV 期)。

条目 3b. 试验开始后对试验方法所作的重要改变(如合格受试者的挑选标准),并说明原因

举例——“最初按 1:1:1:1:1:1 的比例将患者随机分配到平行的 6 组中,分别给予 5 种奥米沙班(otamixaban)……治疗方案,或阳性对照药肝素……为了患者的安全,有一个独立的数据监察委员会审查未设盲数据;对有效还是无效未作中期分析。试验期间,该委员会建议停止接受最小剂量奥米沙班[0.035 mg/(kg·h)]的一组试验,因为临床证据表明其抗凝作用不够。试验方案随即根据该建议作了修改,之后的受试者按 2:2:2:2:1 的比例随机分配到各组,分别接受保留下来的几种奥米沙班方案,或阳性对照药。”^[86]

解释——有些试验开始时可能没有固定的计划(也就是说,完全是探索性的),但大多数都会有试验方案详细说明试验将如何实施。偏离最初的方案是有可能的,因为无法预计在整个试验过程中可能出现的每一种环境变化。因此,有些试验在开始后会对试验方法作重要改变。

改变试验计划可能因于从其他研究获得的外部信息,也可能是因为内部经费困难,或招募的受试者数量不足。像这样改变试验方案时,不应该破坏对正在积累的受试者结局指标方面的数据所设的盲法。在有些试验中,独立的数据监察委员会可能会根据他们所看到的未设盲数据而建议更改试验方案,这是其职责之一。此类更改可能影响到研究方法(如治疗方案、入选标准、随机分配比例,或随访时间长短等的更改)或试验实施(如让数据质量低劣的中心退出试验)^[87]。

有些试验一开始就采用“适应型(adaptive)”设计。对这类设计尚没有一种被大家普遍接受的定义,有一种定义可能适用,即“一种利用积累的数据来决定如何调整研究的某些方面而又不损害试验的可靠性和完整性的多阶段研究设计。”^[88]调整通常是针对样本量和治疗方法的种类,以便更快地做出决策,更有效地利用资源。不过在选用这种设计时还要考虑很重要的伦理、统计和可行性等问题^[89, 90]。

无论这些调整本身就是试验设计的一部分,还是因为环境改变而作出,都必须充分地予以报告,这对于读者解释试验结果是有帮助的。试验方案的改变应该报告,这一点目前做得不好。有一篇综述将试验报告与试验方案作了比较,结果显示大约半数报告 RCT 的期刊论文存在主要结局指标与试验方案不一致且没有说明原因的情况^[57]。常见的这类不一致还有随机方法、盲法^[91]以及统计学分析^[92]的细节。

条目 4a. 受试者合格标准

举例——“合格的受试者均为 18 岁及以上的成年艾滋病病毒感染者,根据马拉维国家抗艾滋病病毒治疗指南,符合抗逆转录病毒疗法的人选标准(WHO 临床 III 期或 IV 期,或 WHO 任何一期而 CD4 计数小于 250/mm³),并且在 BMI 小于 18.5 时开始治疗。排除标准为怀孕和哺乳,或正在参加其他补充喂养计划。”^[93]

解释——详尽地描述使用怎样的合格标准筛选受试者是必要的,它能帮助读者正确解释研究结果。尤其是,清楚地理解这些标准是判断试验的结果适用于哪些人(即试验的可推广性或实用性及与临床和公共卫生实践的相关性,见条目 21)必备的几个要素之一^[94]。从这一点看,描述招募方法,如转诊或自选(例如借助广告),也是很重要的。由于受试者合格标准用在随机分配之前,所以不会影响试验的内部真实性(internal validity),但它们是确保试验外部真实性(external validity)的中心环节。

典型而公认的选择标准涉及所研究的疾病的性质和病期,涉及如何排除那些可能特别容易受到所研究的干预措施伤害的人,还涉及如何确保研究工作遵守法律规定和伦理准则的问题。例如,在干预措施的研究中,特别要求做到受试者知情同意。常见的将纳入标准和排除标准区分开来的做法是没有必要的;同样一条标准其实既可以是纳入受试者的标准,也可以是排除受试者的标准^[95]。

尽管受试者合格标准很重要,却经常未被充分报告。例如,有 8 项已发表的曾被国立卫生研究院

(National Institutes of Health) 编成《临床警示(Clinical Alert)》的临床试验,在其原试验方案中列举了平均 31 条合格标准,但在期刊上发表时只提到其中的 63%,而在《临床警示》中只提到 19%^[96]。在有关艾滋病病毒的临床试验中也发现类似的欠缺^[97]。在 364 篇外科的 RCT 报告中,25%根本就没有列出受试者合格标准^[98]。

条目 4b. 资料收集的场所和地点

举例——“该研究于 2006 年 1 月至 2007 年 4 月在位于马拉维布兰太尔(Blantyre)的伊丽莎白皇后中心医院(Queen Elizabeth Central Hospital)的抗逆转录病毒疗法门诊部进行。布兰太尔是马拉维的主要经济城市,人口 100 万,据估计 2004 年有 27%的成年人感染艾滋病病毒。”^[93]

解释——和受试者合格标准(见条目 4a)以及干预措施的描述(见条目 5)一样,场所和地点也是判断试验实用性和可推广性的关键信息。受试者是从一级、二级或三级医疗机构招募,还是从社区招募?不同的医疗卫生机构在组织、经验和资源方面有很大差异,在不同医疗机构开展某种疾病研究的基线危险性也相差悬殊。研究环境的其他方面(包括社会、经济和文化环境以及气候)也会影响研究结果的外部真实性。

作者应该报告研究场所的数目和类型,描述参与的医护提供者,应该报告开展研究的地点,包括国家,最好还有城市,以及周围环境(例如社区、诊所、医院门诊部、病房等)。尤其要清楚地报告试验是在一个还是几个中心(“多中心临床试验”)开展的。这些描述应提供足够的信息以便读者可以判断试验的结果是否适用于他们自己所处的环境。试验实施时的环境,与之后试验结果被用来指导实践或制定政策的环境可能相差甚远^[94, 99]。作者还应该报告任何可能影响观察结果的有关试验场所和地点的其他信息,如交通问题可能影响患者参与或推迟干预措施的使用。

条目 5. 详细描述各组干预措施的细节以使他人能够重复,包括它们实际上是在何时、如何实施的

举例——“在 POISE 试验中,患者于术前 2~4 h 首次服用试验药物(即口服美托洛尔缓释剂 100 mg 或与其匹配的安慰剂)。要求心率大于或等于 50 次/分,收缩压大于或等于 100 mm Hg 才能给予试验药物;这些血液动力学指标在每次给药前检测。在术后 6 h 内的任何时候,只要心率大于或等于 80 次/分,且收缩压大于或等于 100 mm Hg,则

患者服下术后的第一次给药(美托洛尔缓释剂 100 mg 或与其匹配的安慰剂)。如果术后 6 h 内未给试验药,则患者在术后 6 h 时服下术后第一次给药。术后第一次给药后 12 h,患者开始每天服用美托洛尔缓释剂 200 mg 或与其匹配的安慰剂,连续 30 d。如果患者的心率一直低于 45 次/分,或收缩压降至 100 mm Hg 以下,则停用试验药,直至其心率和收缩压恢复,再开始每天给予试验药 100 mg。如果患者的心率维持在 45~49 次/分,且收缩压超过 100 mm Hg,则推迟 12 h 服用试验药。”^[100]

“患者被随机分配,晚上佩戴定制的氯丁橡胶夹板或接受常规治疗。夹板是一种制动矫形器具,推荐只在晚上使用,置于拇指根部和大鱼际隆起处,但不超过腕部(图 1)。夹板由 3 名职业治疗师制作,他们给每个患者调整夹板,使得虎口可以张开,且拇指与食指正好处在相对的位置。患者觉得需要调整夹板时,或者如果戴上夹板时疼痛加剧,或者产生了不良反应(如磨破了皮肤),则应随时联系职业治疗师。由于对这种情况没有称得上金标准的治疗,因此,对照组和干预组的患者均接受常规治疗(由全科医生或风湿科医生酌情处理)。我们决定不使用安慰对照,因为就我们所知,还没有人能按要求成功地对患者设盲而实施夹板的安慰对照。”^[101]

解释——作者应该详尽地描述每一种干预措施,包括对照的干预措施,足以让想使用试验所评价干预措施的临床医生确切地知道如何使用^[102]。对于药物治疗,应该包括药名、剂量、给药方式(如口服、静脉注射等)、给药时间和疗程、什么情况下撤药等信息,如果必要,还应该有什么逐渐增加剂量的方案。如果对照组接受“常规治疗”,详尽描述“常规治疗”都包括哪些治疗方法也很重要。如果对照组或治疗组联合使用几种干预措施,作者应该详尽描述每一种干预措施,解释联合方案中每一种干预措施的施加顺序或撤减顺序,如有必要,说明在什么情况下才开始施加这些干预措施。

CONSORT 声明的几个扩展版专门论述非药物和草药干预措施的报告及其特殊要求(如专业技能,如何对干预措施实行标准化的细节等)^[43, 44]。我们建议读者最好参阅关于非药物和草药干预措施的 CONSORT 声明。

条目 6a. 完整而确切地说明预先设定的主要和次要结局指标,包括它们是在何时、如何测评的

举例——“治疗银屑病效果的主要终点指标是 12 周时用银屑病面积和严重程度指数(psoriasis area and severity index, PASI)测评的银屑病活动

度与基线相比改善 75% 的患者比例。另外,还对 PASI 评分变化的百分率以及银屑病皮损改善程度进行分析。”^[103]

解释——所有 RCT 都通过组间比较来评价因变量或结局指标(终点指标)。大多数试验都有好几个结局指标,其中有些指标比其他指标更有意义。主要结局指标是事先预设的被认为是试验相关的各方(如患者、决策者、医生、资助者等)最重要的结局指标,常用来计算样本量(见条目 7)。有些试验不止一项主要结局指标。不过,使用多项主要结局指标会带来如何解释多种分析结果的问题(见条目 18 和 20),因此不推荐这样做。其他有意义的结局指标是次要结局指标(辅助性的结局指标)。可以有好几个次要结局指标,常常包括未预料到的或不希望出现的干预措施效应(条目 19),但危害始终应该被视为重要的结局指标,不管是把它们列为主要结局指标还是次要结局指标。

所有结局指标,无论是主要还是次要指标,都应该列出并完整地定义。其原则是提供的信息要足以使他人也能够使用这些指标^[102]。如果结局指标在随机分配后的几个时间点测评,作者还应该指明事先确定的最主要的时间点。对许多非药物干预措施来说,需要说明由谁测评结局指标(如是否需要特殊的技能才能做测评)和有多少测评者^[43]。

若有而且适合,就应该使用已有的有效量表或公认的指南^[104, 105],既可以提高测评质量又有助于与同类研究比较^[106]。例如,使用经过验证的测评工具有可能会提高生活质量评估的质量^[107]。作者应该说明量表的来源和特征。

在 196 项非甾体抗炎药治疗类风湿关节炎的 RCT 中使用了 70 多种结局指标^[108];在 2 000 项精神分裂症的临床试验中用过 640 种不同的测评工具,其中 369 种只用过一次^[33],对这 2 000 项试验中的 149 项的调查显示,使用未发表过的量表是产生偏倚的原因之一。在结果认为有治疗优势的非药物临床试验中,三分之一是基于未发表过的量表,若使用已发表的量表来评估则不能得出同样结论^[109]。类似的数据还见于其他报道^[110, 111]。对 2000 年发表的 519 篇 RCT 报告的研究显示,只有 45% 明确指出了主要结局指标^[16],而对 2006 年发表的 614 篇 RCT 报告的类似研究显示,明确指出了主要结局指标者也仅仅占 53%^[17]。

条目 6b. 试验开始后对结局指标是否有任何更改,并说明原因

举例——“最初的主要终点指标是全因死亡率,

但是在做设盲分析(masked analysis)时,数据与安全性监察委员会注意到,总死亡率比预计的要低,试验不可能按照最初计划的样本量和把握度(power)完成。于是指导委员会决定采纳将全因死亡率(最初的主要终点指标)作为共同主要终点指标的方案,即将全因死亡率作为主要终点指标,也可以将因心血管问题入院治疗(原先设计为第一次主要终点指标)作为主要终点指标。”^[112]

解释——导致偏离最初试验方案的原因很多(见条目 24)。作者应该报告对试验方案所作的的所有重大变动,包括临时改变受试者合格标准、干预措施、检查项目、数据收集方法、分析方法,以及结局指标。这些信息经常没有报告。

正如上文所说(见条目 6a),大多数临床试验记录多种结局指标,于是就存在这样的风险,即只选择部分结果予以报告(见条目 17)。预先设定好主要和次要结局指标并依此报告(见条目 6a)就能消除这种危险。然而,在某些临床试验中,环境因素要求改变某种结局指标的评估方法,甚至如上面的实例,要求改用一种完全不同的结局指标。比如,来自其他临床试验或系统综述的外部证据提示终点指标或许不合适,或者招募的受试者人数或总体事件发生率低于预期^[112]。基于未设盲的数据改变某项终点指标尤其成问题,尽管对适应型试验设计来说可以在文中对此特别予以说明^[88]。有任何这种改变作者都应该指出并说明原因。同样,试验开始后如果对将哪些结局指标定为主要或次要指标有所更改,也必须报告并说明原因。

将已发表的 102 篇随机对照临床试验报告与其试验方案比较发现,62%的报告至少有一项主要结局指标在原试验方案基础上作过更改,或是新增的,或者被省略^[55]。一项对加拿大健康研究所资助的 48 项临床试验所作的研究也发现,40%的试验在发表时其主要结局指标与原试验方案有异^[113]。后来对 150 篇临床试验报告的研究显示,没有一篇提到对试验方案的改动,更不用说对此有所解释了。最近发表的一篇有关经验性研究的系统综述检查了报告结局时存在的偏倚,提示其他研究也存在类似问题^[57]。

条目 7a. 如何确定样本量

举例——“为了做到与 Lobo 等^[17]的研究一致,术后住院时间减少 3 天(标准差为 5 天),达到双侧 5%的显著性水平和 80%的把握度,预计脱落率为 10%,则每组必须有 50 例的样本量。预计需要 12 个月的招募期才能招募到该数量的患者。”^[114]

“按照安慰剂组主要综合性终点事件的发生率

在 2.25 年时预计达到 11%,我们计算出,为了检测出伊伐布雷定(ivabradine)组和安慰剂组之间有显著性差异,且达到 90%的把握度,即相当于相对危险度(relative risk)减少 19%(双侧检验, I 类误差为 5%),需要 950 例发生主要终点事件的病例,样本量要达到 9 650 例。我们最初设计的是事件驱动的临床试验,原计划在发生主要终点事件的病例达到 950 例时终止试验。然而,主要终点事件的发生率高于预期,可能归因于所招募患者的基线特征,他们具有的风险高于预期(即 NYHA I 级患者的比例较低,而糖尿病和高血压患者的比例较高)。我们计算,当发生终点事件的病例达到 950 例时,最晚纳入的患者可能只治疗了 3 个月。因此,在 2007 年 1 月,执行委员会决定将本研究从事件驱动改为时间驱动,并继续研究直到最后纳入的患者完成 12 个月的随访。这一改动并未改变原定为 3 年的研究周期。”^[115]

解释——出于科学和伦理的原因,必须仔细计算临床试验的样本量,并在临床意义与统计学要求之间取得平衡。理想的情况是,研究的样本量要足够大,从而具有较大的可能性(把握度)以检测出在该样本量的情况下临床结果的重要差异具有统计学意义,如果这种差异确实存在。具有重要意义的效应值的大小与检测出该效应所需的样本量大小成反比,即差异越小所需样本量越大。计算样本量的要素包括:(1)每组结局指标的估计值(指试验组之间具有重要临床意义的目标差异);(2) α (I类)误差的大小;(3)统计学把握度或 β (II类)误差的大小;(4)若结局指标为连续变量,要有测量值的标准差^[116]。这些要素间的相互影响以及如何报告对群组临床试验^[40]和非劣效性与等效性临床试验^[39]而言会有所不同。

作者应该说明如何确定样本量。如果使用了正规的把握度计算方法,作者应该指出用来计算的主要结局指标(见条目 6a)和计算时用到的所有量值,以及每组算出的目标样本量。更为可取的是采用对照组预期的结果以及作者认为不能忽视的组间差异来计算样本量。还有一种选择是,作者在描述计算方法时可以列出各组事件发生的百分率或均数。应该详细描述是否允许对研究期间损耗或不依从的病例进行补充。

一些方法学家曾经写道,有些所谓低把握度的临床试验或许是可以接受的,因为它们最终可能被纳入某篇系统综述和 meta 分析^[117-119],况且有一些信息总比没有强。在这里要特别提醒,只有满足以下条件才可供 meta 分析使用^[118],即如果不考虑试验结果的话,该试验不应该有偏倚,而且是恰当地

报告和发表。另一方面,很多医学研究人员担心低把握度的结果不确定的临床试验得不到发表,因而坚持认为所有的临床试验就个别而言都应该是有“充分的把握度(sufficient power)”的。这种争论将会继续下去,CONSORT 工作组成员的观点也不尽相同。不过,严格地说,这种争论和那些观点的存在对如何报告临床试验来说是无关紧要的。不论试验的把握度如何,作者都必须恰当地报告其设想的效应大小,同时恰当地报告其研究方法和假设^[118]。这样才能向读者透明地展示试验的把握度,使他们能够藉此评估试验是否达到了原来设想的效应。

有些临床试验利用中期分析来帮助决定是否要提前终止试验,有时候中期分析用来帮助决定是否在原计划试验该结束时还继续招募受试者(见条目 7b)。如果由于某些其他原因(例如招募困难或修改了目标样本量)而使实际样本量与最初设计的样本量不同,则应该予以说明。

小样本的研究在报告时经常会得出各干预组之间没有差异的错误结论,事实上是因为纳入研究的病例太少才得出这样的结论^[120]。对已发表的临床试验的综述一致发现,很大部分试验因为把握度低下不足以检测出具有临床意义的治疗效果^[121-123]。其实,小但真正具有临床意义的差异可能远多于大的差异,但它们需要大规模的临床试验才能检测出来^[124]。

总的来说,已见报道的临床试验样本量偏小。196 项关节炎临床试验的中位样本量是 54 例^[108],73 项皮肤病临床试验的中位样本量为 46 例^[8],而 2 000 项精神分裂症临床试验的中位样本量为 65 例^[33]。以上这些小样本量的情况与一篇对 2000 年 12 月收录于 PubMed 的 519 篇临床试验报告的研究^[16]以及另一篇对 2006 年收录于 PubMed 的 616 篇临床试验报告的类似研究^[17]的结果一致,根据这两篇研究,平行对照试验招募的中位患者数为 80 人。而且,很多综述发现,很少有作者报告如何确定样本量^[8, 14, 32, 33, 123]。

利用试验的结果事后计算统计学把握度几乎没有什么价值,这时试验的把握度适合用可信区间来表述(见条目 17)^[125]。

条目 7b. 必要时,解释中期分析和试验中止原则

举例——“试验期间做过两次中期分析。显著性水平设为总体 P 值 0.05,并且根据 O'Brien-Fleming 终止界值计算得出。本次终期分析用 Z 值 1.985,相应的 P 值为 0.0471。”^[126]

“一个独立的数据和安全性监察委员会定期审核疗效和安全性方面的数据。终止原则以修正的

Haybittle-Peto 界值为准,即对于疗效数据,研究的前半阶段为 4 个标准差,后半阶段为 3 个标准差,对于安全性数据,前半阶段为 3 个标准差,后半阶段为 2 个标准差。当发生主要事件的病例数累计达到预期数的 50%和 75%时,分别对疗效做了两次正规的中期分析;对中期检验所报告的 P 值未作校正。”^[127]

解释——很多临床试验需要很长一段时间招募受试者。如果干预措施效果极好或极差,均应出于伦理原因提前终止研究。这个问题可以通过一边积累数据一边检查结果的做法来解决,最好由独立的数据监察委员会执行。不过,对积累的数据进行多次统计学检验而不经恰当的校正则可能得到错误的结果和导致错误的解释^[128]。如果临床试验的累积数据经过 5 次使用 P 值 0.05 的中期分析,则总体假阳性率接近 19%,而不是通常人们所习惯的仅仅 5%。

有几种成组序贯统计方法可用来校正多次分析^[129-131],其应用应该在临床试验方案中预先设定。使用这些方法时,每次中期分析都对数据进行比较,如果 P 值小于成组序贯方法规定的临界值则提示有统计学意义。一些临床试验研究人员使用成组序贯方法来辅助决策^[132],也有人把它们当作一种正规的终止试验的标准(其用意是如果观察到的 P 值小于临界值就终止试验)。

作者应该报告,他们或数据监察委员会是否多次“查看”数据,如果是,则应该报告“查看”的次数,“查看”的原因,使用的统计学方法(包括任何正规的终止试验的标准),以及是在试验开始之前就已计划好的,还是在数据监察委员会受委派查看中期数据之前,或试验开始后的某个时间才安排的。这些信息常常没有包括在已发表的临床试验报告中^[133],甚至一些比原计划提前终止的临床试验也没有报告这些信息^[134]。

条目 8a. 产生随机分配序列的方法

举例——“独立的药剂师根据计算机生成的随机分配表发放尼古丁吸入器或安慰剂吸入器。”^[63]

“分配受试者时使用计算机生成的随机数字表。”^[135]

解释——应该根据机遇(随机)过程将受试者分配到试验所要比较的各组中,该过程的特点是具有不可预测性(见框 1)。作者应该提供充足的信息让读者可以评价生成随机分配序列的方法,以及分组过程中产生偏倚的可能性。有一点非常重要,与随机分配过程相关的信息要放在正文中,而不要另以补充文件的形式出现,从而导致读者可能错过阅读。

“随机(random)”一词有精确的技术含义。随机分配时,每名受试者在分配前接受每种干预措施的概率是已知的,但实际分配的干预措施却是由机遇过程决定的,无法预测。然而,在文献中“随机”经常被不恰当地用来描述一些使用了非随机的、具有确定性的分配方法的试验,如交替分配,按住院号分配,或按出生日期分配等。若研究人员使用这类非随机方法,就必须准确地对其进行描述,并且不得使用“随机”一词,或该词的任何变化形式。甚至用“半随机(quasi-random)”一词来描述这类试验也是不可接受的。基于非随机方法的临床试验一般都会得到有偏倚的结果^[2-4, 136],偏倚很可能源于未能充分隐藏其分配系统(见条目 9)。

很多序列生成方法都可以满足要求。但如果缺乏进一步的说明,读者仅从“随机分配(random allocation)”、“随机化(randomisation)”、“随机(random)”等词语中是无法做出准确判断的。作者应具体说明序列生成的方法,如随机数字表或计算机随机数字生成程序等。也可按照一种普遍接受的非随机方法,即最小化(minimisation)程序生成序列(见框 2)。

在有些临床试验中,受试者被故意以不均等的数目分入各组,如为了获得某种新方法更多的经验,或为了限制试验的成本。如果是这种情况,作者应该报告随机分配的比例(例如 2 : 1 或对照组每分配 1 名受试者治疗组分配 2 名)(见条目 3a)。

在 2000 年 PubMed 收录的一组具有代表性的临床试验报告中,只有 21% 详细报告了随机序列生成的方法^[16],在 2006 年 PubMed 收录的临床试验报告的研究中,这一比例上升到 34%^[17]。在超过 90% 的这些试验中,研究人员使用了计算机随机数字生成程序或随机数字表。

条目 8b. 随机方法的类型,任何限定的细节(如怎样分区组和各区组样本多少)

举例——“使用 Stata 9.0 (StataCorp, College Station, TX) 统计软件生成随机序列,并按中心分层,采用的随机区组大小分别为 2、4 和 6,按 1 : 1 的比例分配。”^[137]

“按照简单随机程序(计算机生成的随机数字)将受试者随机分配到两个治疗组。”^[138]

解释——有数百名以上受试者的临床试验经常依靠简单随机法使两个试验组的受试者数相近^[139],并使两组在已知和未知的预后因素上大致可比^[140]。对样本量较小的临床试验(见条目 7a),甚至原本设计的样本量并不小只是在达到目标样本量之前就终止的试验,可使用限制性随机法,以达到

组间样本量和基线特征的均衡(见框 2)。

重要的是,要明确指出有没有使用限制方法,可以直接表明没有使用限制方法或直接说明使用了“简单随机法”。不然,就应该在指出随机选择方法的同时,具体说明用于限制随机的方法。对于区组随机法,作者应详细描述如何产生区组(例如,借助计算机随机数字生成程序使用序列区组设计),区组大小,以及区组大小是否固定还是随机改变。如果试验人员知道了区组大小,也应该报告,因为可能会导致密码破译。作者应该具体说明是否应用了分层,如果是,则说明分层因素(如招募点、性别、疾病的阶段等),分层的分类截断值,以及限制的方法。分层是一项有用的技术,尤其是对小样本试验来说更是如此,但如果使用的分层因素过多,则实施就会变得复杂,甚至可能会行不通。若使用了最小化法(见框 2),那么连同其所涉及的因素均应描述清楚;若使用了随机元素,也应该具体说明。

在专业性期刊发表的 206 篇临床试验报告以及在综合性医学期刊发表的 80 篇临床试验报告中分别只有 9%^[23]和 39%^[32] 报告使用了分层。两者各只有约一半的报告提到使用了限制性随机法。但以上研究以及 Adetugbo 和 Williams^[8] 的研究发现,在很多临床试验中,虽然各治疗组样本量相同或相似,却并未提及区组随机或分层随机的使用。对这种样本数高度均衡的一种可能解释是使用了限制性随机法而没有报告。

条目 9. 用于执行随机分配序列的机制(例如按序编码的封藏法),描述干预措施分配之前为隐藏序列号所采取的步骤

举例——“强力霉素和安慰剂均为胶囊剂型,外观相同。分瓶包装并连续编号,根据随机分配表发给每名妇女。每名妇女都分配到一个顺序号并领到编有相应号码的事先装好胶囊的药瓶。”^[146]

“对负责招募和评估受试者的研究人员(Joel A Radford)隐藏分配顺序,分配序列装于按顺序编号的密封的不透明的信封中。信封内衬铝箔纸阻挡强光透过信封。为了防止分配序列被打乱,把受试者的姓名和出生日期写在信封上,并将密封信封和受试者的详细信息记录在录像带中。信封内的复写纸将这些信息转写到信封内的分配卡上,然后再由第二位研究人员(Catherine Cook)观看录像带,确认写上受试者姓名后的信封仍然密封。只有在招募的受试者完成全部基线评估后,才打开相应的信封,这时候再分配干预措施。”^[147]

框 2 随机化和最小化

简单随机法 (simple randomisation)——基于单一分配比例的纯随机称为简单随机法。按 1 : 1 比例分配的简单随机法类似于掷硬币, 尽管我们并不提倡在 RCT 中通过掷硬币进行随机。“简单”一词有点名不副实。尽管其他随机方法貌似更复杂、更高级, 但实际上, 简单随机法更深奥、更不可预测, 在预防偏倚方面更是优于其他所有随机方法。

限制性随机法 (restricted randomisation)——指除简单随机法以外的其他随机方法, 最为常见的是区组随机法。其他限制性随机方法还包括替代 (replacement) 法、偏币 (biased coin) 法和瓮 (urn) 法等, 尽管这些方法用得很少^[141]。

区组随机法 (blocked randomisation)——区组设计用于保证所比较的各组根据预先确定的比例分配, 通常是 1 : 1 或各组样本量基本相同。区组设计的方法可以用来保证试验期间无论何时各组的例数都接近均衡。例如, 若每个区组有 8 名受试者, 则试验的每个干预组会分配 4 名受试者^[142]。提高均衡性的代价是降低了分配序列的不可预测性。尽管每个区组内干预措施的顺序会随机变化, 但如果实施试验的人知道区组大小, 就可能推测出分配给下一例的治疗^[143]。对于干预措施设盲, 使用较大的区组, 以及随机变换区组大小可以改善这个问题。

分层随机法 (stratified randomisation)——分层用于保证各组受试者特征的良好均衡性。特别是在小型的试验中, 偶尔会出现所研究的各组在年龄、疾病分期等基线特征上不完全匹配的情况, 这就会削弱试验的可靠性^[144]。可以避免这种不均衡, 同时又不牺牲随机的优势。分层可以保证接受每种干预的受试者数在每一层内接近均衡。分层随机法是通过将受试者分成 2 个或多个亚组, 然后分别实施随机分配来实现的 (例如, 按研究中心、年龄或疾病严重程度分层)。按中心分层在多中心临床试验中很常用。分层设计要求某些限制性条件 (如在各层内分区组)。不分区组的分层是无效的。

最小化法 (minimisation)——最小化法保证各干预组之间在若干既定的患者因素 (如年龄等) 方面具有均衡性^[22, 60]。第一名患者是真正随机分配的, 而随后每名患者分配的治疗在这时就已明确, 这种分配使各组之间在既定因素上的不均衡最小化。此后可以使用这样分配的治疗, 或者任意选择一种干预措施, 条件是有较高的权重支持选择这种干预措施会使不均衡最小化 (如 0.8 的概率)。随机成分的应用一般来说是可取的。最小化法有一个优点, 就是在整个试验过程中使样本数较小的各组在受试者特征方面接近相同。最小化法是唯一可以接受的随机分配法的替代方法, 甚至有人认为它优于随机分配法^[145]。不过, 最小化法缺乏消除所有已知和未知因素方面偏倚的理论基础。但总的来说, 即使并不包含随机元素 (random element), 使用了最小化法的临床试验在方法学上可视为等同于随机临床试验。

解释——条目 8a 讨论了如何生成不可预测的分配序列。当受试者进入试验后, 如何应用这一分配序列尤为重要 (见框 1)。生成的分配顺序应该通过分配隐藏来执行^[23], 这是一种严格的机制, 可防止负责招募受试者的人因预先知道分配的治疗而影响分配过程。研究人员要在不知道序列中下一个分配的情况下决定是否纳入或排除某个受试者, 并且得到受试者的知情同意^[148]。

分配隐藏不能与盲法混淆 (见条目 11)。分配隐藏意在防止选择性偏倚, 在分配前保密分配序列, 多能成功地执行^[2]; 而盲法旨在防止实施偏倚和测量偏倚, 在分配后保密分配序列, 并非总能得到执行^[23]。若分配隐藏不充分, 即使进行了随机, 还是有可能使不可预测的分配顺序遭到破坏^[2, 149]。

集中分配或“第三方”分配尤为可取。很多好的

分配隐藏机制往往有外部介入。两种常用的方法是药房或中心电话随机系统, 自动分配系统可能会变得更常用^[150]。当外部介入行不通时, 使用编号容器是分配隐藏的极好方法。根据分配序列, 将干预措施 (通常是药) 密封在标有连续数字、外形相同的容器里^[151]。如果组织和监管严密, 将分配序列放在依次编号的不透光的密封信封内不失为一种好的分配隐藏机制。但这种方法可能会失败, 尤其是执行不当时。研究人员要保证信封置于光照下确实不透光, 而且只能在受试者的姓名和其他信息写在相应的信封上后, 才依次打开^[143]。

一些方法学研究提供了支持这些预防措施的经验性证据^[152, 153]。分配序列隐藏不充分或不彻底的试验与作者报告了充分的分配隐藏的试验相比, 前者对治疗效果的估计高于后者。这些结果提供了强有力的经验性证据, 即分配隐藏不充分会在评估治

疗效果时导致偏倚。

尽管分配隐藏机制如此重要，已发表的试验报告却往往忽略这些细节。89%的关于类风湿关节炎的试验^[108]、48%的妇产科期刊报道的试验^[23]，以及44%的综合性医学期刊报道的试验^[32]未报告干预措施的分配机制。在有更广泛代表性的 PubMed 收录的全部随机临床试验报告中，只有 18%对分配隐藏机制进行了报告，但所报告的隐藏机制有些是不充分的^[16]。

条目 10. 谁产生随机分配序列，谁招募受试者，谁给受试者分配干预措施

举例——“参考由 Bradford Hill 教授按试验中心和性别分层制定的根据随机数字生成的统计序列，决定患者是接受链霉素治疗加卧床休息(S 组病例)还是仅仅卧床休息(C 组病例)；所有研究人员和统筹者都不知道序列的详情……试验小组接受患者后，在患者进入链霉素中心前，在中心办公室打开与其相应的标有号码的信封；信封里的卡片会提示患者分配到 S 组还是 C 组，然后再将这些信息告知中心的医务人员。”^[24]

“分组的详情写在彩色卡片上，置于按顺序编号的不透光的密封信封中。这些都在国立围产期流行病学研究室(National Perinatal Epidemiology Unit, NPEU)准备好，并保存在每个病区的确定的地方。第二产程结束时，若助产士认为阴道分娩即将开始，便进行随机分配。助产士打开下一个连续编号的信封，将该妇女纳入研究。”^[154]

“利用计算机生成的随机数字表进行分层随机，随机数字表由一名在试验中不参与临床工作的研究人员准备好。我们按因肿瘤入院接受的治疗进行分层。参与研究的护士得到患者的同意后，打电话给一个不参与招募过程的联系人，委托其进行分配。”^[155]

解释——正如条目 9 提到的，纳入受试者时干预措施的分配隐藏特别重要。因此，除了要知道隐藏分配序列的方法外，了解随机序列是如何执行的，尤其是谁生成分配序列，谁登记受试者，谁分配受试者进入各试验组，也是非常重要的。

试验中随机分配受试者的过程包括 3 个步骤：分配序列生成、分配隐藏和分配实施(见框 3)。尽管同一组人员可以完成每个项目下的多项工作，但研究者应尽量使参与生成分配序列和隐藏分配的人与参与执行分配的人完全分开。所以，如果有人参与了分配序列生成和分配隐藏步骤，那他们最好不要参与分配实施步骤。

框 3 典型的随机分配过程的步骤

序列生成

- 由某些随机程序生成分配序列

分配隐藏

- 建立分配隐藏机制(如编号的完全一样的瓶子或按顺序编号的密封的不透光的信封)
- 从序列生成步骤开始利用分配序列准备分配隐藏机制

实施

- 招募受试者：
 - 评估合格性
 - 讨论试验相关事宜
 - 获取知情同意
 - 纳入受试者
- 确认干预措施的分配(如打开下一个信封)
- 实施干预措施

即使序列生成和分配隐藏完美无瑕，不将分配序列的生成和隐藏与执行试验分配完全分开仍可能导致偏倚。如生成分配序列的人可能保留备份，在与候选受试者面谈时参考。这样的话，不管分配序列是否可以预测，此人都可能使招募和分配过程发生偏倚。因此，研究人员必须保证分配计划不可预测，并将计划锁起来，连序列生成者也看不到(如放在招募地点的人不能接触到的某幢大楼内的保险柜里)。试验报告中应说明研究人员将分配清单藏于何处。

条目 11a. 如果实施了盲法，分配干预措施之后对谁设盲(例如受试者、医护提供者、结局评估者)，以及盲法是如何实施的

举例——“分配到干预组的患者和医生知道他们分配在哪个组，但分配结果对结局评估者和数据分析者设盲。”^[156]

“向施行干预措施的工作人员和受试者强调每种饮食都符合健康原则，都是专家们提倡的有利于长期减肥的，以此来严格保证设盲和均衡。除了施行干预措施的人(营养师和行为心理治疗师)之外，受试者的饮食分配也要对研究人员和工作人员设盲。试验遵守已建立的程序，始终将进行结局指标测量的工作人员与发放饮食的工作人员分开。得到结局指标测量结果的工作人员不知道各组的饮食分配。发放饮食的工作人员、营养师和行为心理治疗师未参与结局指标的测量。向所有的研究人员、工作人员和受试者遮蔽结局指标的测量结果和试验结果。”^[157]

解释——“盲法(blinding)”或“遮蔽(masking)”是指隐瞒干预措施分配信息以免参与试验的人因为知道这些信息而可能受影响。盲法是防止偏倚的重要保护措施,尤其是在评价主观性结局指标时^[153]。

Benjamin Franklin 被公认是第一个在科学实验中使用盲法的人^[158]。他在施行催眠术(一种 18 世纪流行的“医疗剂”)时蒙住受试者的眼睛使他们不知道,其实验结果显示催眠术是一种无效的安慰疗法。科学界从这一实验认识到盲法减少偏倚的效力,从此盲法成为科学实验常用的方法,并沿用至今。

通过解释盲法的术语,框 4 对可能因为知道治疗分配情况而给试验带来偏倚的各类人群(即受试者、医护提供者、数据收集者、结局评判者和数据分析者)进行了定义。受试者如果知道治疗的分配,其疗效可能会不同(比如当他们接受的是新疗法时,就会表现出更好的疗效)^[153]。未施行盲法还可能影响对干预措施的依从性,导致辅助干预措施的使用,以及受试者退出试验的危险。

不对医护提供者设盲会导致类似的偏倚;未设盲的数据收集者可能会有区别地评价结局指标(如评价的频次和时间安排),重复测量异常结果,或在测试中提供帮助;未设盲的结局评判者可能会有区别地评价主观性结局指标;未设盲的数据分析者可能会有选择地使用分析策略从而导致偏倚,如选择有利的时间点或结局指标,以及将受试者从分析中剔除。关于这些偏倚已有很多报道^[71, 153, 159-162]。

不像分配隐藏(见条目 10),盲法并非总是适用或可行,如比较从耳朵或拇指上采血的疼痛程度的试验^[163]。当结局指标涉及主观因素时,使用盲法

尤其重要,如评价疼痛程度。就客观结局指标如任何原因引起的死亡而言,对数据收集者和结局评价者设盲似乎无关紧要。即使这样,如果不对受试者或医护提供者设盲也会导致其他问题,如病例耗损差别^[164]。在某些试验尤其是外科试验中,对受试者和外科医生设盲常常十分困难,甚至不可行,但对数据收集者和结局评判者设盲往往可以成功施行,如可以拍摄治疗前后伤口的照片,让未参与试验的人评估^[165]。不论盲法是否可行,作者都可以而且应该说明设盲的对象(即受试者、医护提供者、数据收集者和结局评判者)。

遗憾的是,作者往往不报告是否采用盲法^[166]。如 506 项囊肿性纤维化试验的报告中有 51%^[167], 196 项类风湿关节炎试验中有 33%^[108], 68 项皮肤病学试验中有 38%^[8] 未报告是否使用盲法。只有当作者改进了试验中盲法应用的报告,读者判断试验的可靠性才不会有困难,他们才有望利用试验结果指导其临床实践。

人们有时更倾向于用“遮蔽”一词来代替“盲法”,以避免与视力丧失的疾病混淆。不过,“盲法”在方法学上的含义似乎已被人们普遍理解,完全可用于报告临床试验^[165, 168]。

条目 11b. 如有必要,描述干预措施的相似之处

举例——“健美生实验室有限公司(Jamieson Laboratories Inc)提供速释烟酸制剂,一种规格为 500 mg 的白色椭圆形分劈药片。我们独立地利用高效液相色谱法证实了其中的成分……安慰剂在味道、颜色、大小,以及微晶纤维素、二氧化硅、磷酸氢钙、硬脂酸镁和硬脂酸含量方面与试验药完全相同。”^[172]

框 4 盲法术语

为了便于应用,一个技术术语的用法和解释必须一致。临床试验报告的作者常使用“双盲(double blind)”一词,而较少使用“单盲(single blind)”和“三盲(triple blind)”。使用这类词汇有一个问题,即临床医生的解释与流行病学教材的定义之间有很大差异^[169]。更有甚者,一项对 200 篇报告使用了双盲的 RCT 的研究,在对这些试验报告的作者调查后发现,设盲的对象实际上有 18 种不同类别人员的组合,这些报告使用了双盲的试验大约每五篇中就有一篇没有对受试者、医护提供者或数据收集者设盲^[170]。

该研究显示存在术语含糊不清的问题,因此,作者和编辑应该放弃使用这些术语。作者应该明晰地报告对相关人员进行设盲的状况,因为对他们是否设盲可能影响试验的可靠性。

医护提供者包括试验期间照护受试者的所有人员(如医生、按摩师、理疗师、护士等)。数据收集者是收集试验结局指标数据的人员。结局评判者是确认受试者最终是否获益的人员。

有些研究人员还提倡对数据监察委员会和文稿撰写者设盲并报告盲法的实施状况^[160]。对这类人员施行盲法并不常见,其价值仍有争议^[171]。

有时候某类人员(如医护提供者)同时又在试验中担当另一角色(如数据收集者)。如果真有这种情况,作者应该明晰地报告对这些人员施行盲法的状况,从而让读者判断试验的可靠性。

解释——正如我们寻求分配隐藏的证据以确认分配是真正随机一样,我们同样寻求盲法实施方法的证据。对受试者和医护提供者施行了盲法的试验,作者应该说明干预措施特征的相似之处(如外观、味道、气味、施予方法等)^[35, 173]。

有些人提倡通过在试验结束时询问受试者或医护提供者,他们认为受试者接受了试验干预措施还是对照干预措施,以检测盲法是否成功^[174]。由于受试者和医护提供者常常会知道受试者是否在主要结局指标方面获益,这使得要确定他们的回答是反映了盲法不成功还是他们对干预措施效果的猜测很准确变得困难^[175]。鉴于这样得到的信息具有不确定性,我们已经将建议报告这类评价盲法是否成功的检测从“CONSORT 2010 声明”中删除。但我们还是提倡作者报告在实施盲法时做过的任何折中处理。例如,作者应该报告在试验实施过程中的任何时间点是否有必要对任何受试者揭盲。

条目 12a. 用于比较各组主要和次要结局的统计学方法

举例——“主要终点指标为意向性治疗人群 20 周研究期间体重的变化……疗效的次要终点指标包括腰围、收缩压和舒张压、代谢综合征发生率……的变化。我们对主要终点指标以及腰围、血压和患者报告的结局指标评分等次要终点指标采用协方差分析(ANCOVA),辅以重复测量分析。协方差分析模型以治疗方法、国家和性别为固定效应,以随机分配时的体重为协变量。我们旨在评价所得数据是否能为不同剂量的利拉鲁肽(liraglutide)优于安慰剂(主要目的)和奥利司他(orlistat)(次要目的)提供证据。”^[176]

解释——数据分析方法有很多,其中某些方法在特定情况下可能并不很合适。有必要说明每种分析所用的统计学过程,并在报告的结果部分进一步阐明。需要遵循的原则为“详细描述统计学方法使具备相关知识而又能获得原始数据的读者能够核实报告的结果”(www.icmje.org)。详细描述统计学分析如意向性治疗分析也很重要(见框 6)。

几乎所有分析方法都会得出一个比较组间结局指标的疗效估计值。作者应同时报告疗效估计值的可信区间,即真实治疗效果的不确定范围。可信区间也可解释为治疗效果与观察数据吻合的数值范围。通常采用 95% 的可信区间,即 100 个类似研究中有 95 个研究的真实值所在的范围。

也可用统计学差异评价研究结果。 P 值表示干预措施的作用确无差异时由机遇造成观察值差异(或更极端的结果)的可能性。极力推荐报告确切的 P 值(如 $P=0.003$),而不是不精确的阈值,如 $P<0.05$ ^[48, 177]。

标准的分析方法都假设数据是“独立”的。就对照试验而言,这通常意味着每例受试者都有一次观察。将一例受试者的多次观察结果视为独立数据是一种严重的错误,当可以在不同身体部位测量结局指标时就会得出这样的数据,比如在口腔和风湿病学研究中。数据分析应建立在每例受试者只计数一次的基础上^[178, 179],或采用更复杂的统计学方法^[180]。196 项类风湿关节炎试验中有 123 项(63%)存在对每个病例多次观察结果进行分析的错误^[108]。

条目 12b. 附加分析的方法,诸如亚组分析和校正分析

举例——“不同治疗组间患者的有效率比较用 Mantel-Haenszel χ^2 检验,对分层变量(是否使用甲氨喋呤)进行校正。”^[103]

“根据是否分配了抗氧化剂治疗或分配的抗氧化剂治疗的不同,是否原有心血管疾病,饮食中叶酸摄入量的大小,是否吸烟,是否患糖尿病,是否使用阿司匹林,是否接受激素疗法,以及是否服用多种维生素,应用分层 Cox 比例风险模型进行预先设定的亚组分析。这些分析使用了基线暴露评估,而且仅限于无亚组基线数据缺失情况的受试者。”^[181]

解释——和主要分析一样,亚组分析也应明确指出所采用的方法。检验效能最强的分析是用辅助的亚组分析(如老年和年轻受试者)来寻找疗效差异的证据,这种比较称为交互检验^[182, 183]。一种常用但带有误导性的方法是对独立分析每一组疗效所得的 P 值进行比较。从一个有统计学意义和一个无统计学意义的 P 值推断出亚组的效果(交互作用)是错误的。这样的推断有很高的假阳性率。

由于亚组分析得出虚假结果的风险很高,因此通常不鼓励使用^[14, 185]。尤其是事后的亚组比较(看到数据后再做分析)得出的结果往往不能被进一步的研究证实。这种分析可信度不高。

有些研究用某种形式的多元回归分析来校正受试者特征的不均衡性。RCT 中需要作校正分析者比流行病学研究少得多,尤其是有一个或多个预后变量时,校正分析可能是明智的选择^[186]。校正分

析最好在研究计划中说明(见条目 24)。例如,对分层变量(见条目 8b)通常推荐进行校正,其原则是分析策略必须符合试验设计类型。在 RCT 中是否使用校正分析不应根据基线资料差异是否具有统计学意义而定(见条目 16)^[183, 187]。采用任何校正分析和统计学方法都应说明理由。

作者应说明选择哪些校正变量,如何处理连续变量,指出其分析是预先计划的还是视数据而选用^[188]。对已发表研究的综述显示,校正分析的以上几个方面报告均不充分^[188-191]。

5.4 结果

条目 13. 受试者流程(极力推荐使用流程图)

条目 13a. 随机分配到各组的受试者例数,接受已分配治疗的例数,以及纳入主要结局分析的例数举例——见图 2 和图 3。

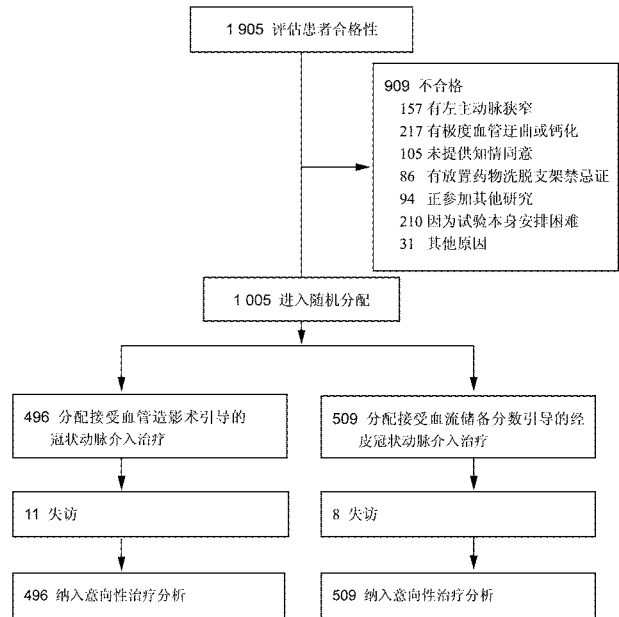


图 2 血流储备分数与血管造影术比较引导经皮冠状动脉介入治疗的多中心临床试验流程图(改编自 Tonino 等^[313])。流程图包括排除受试者的详细信息

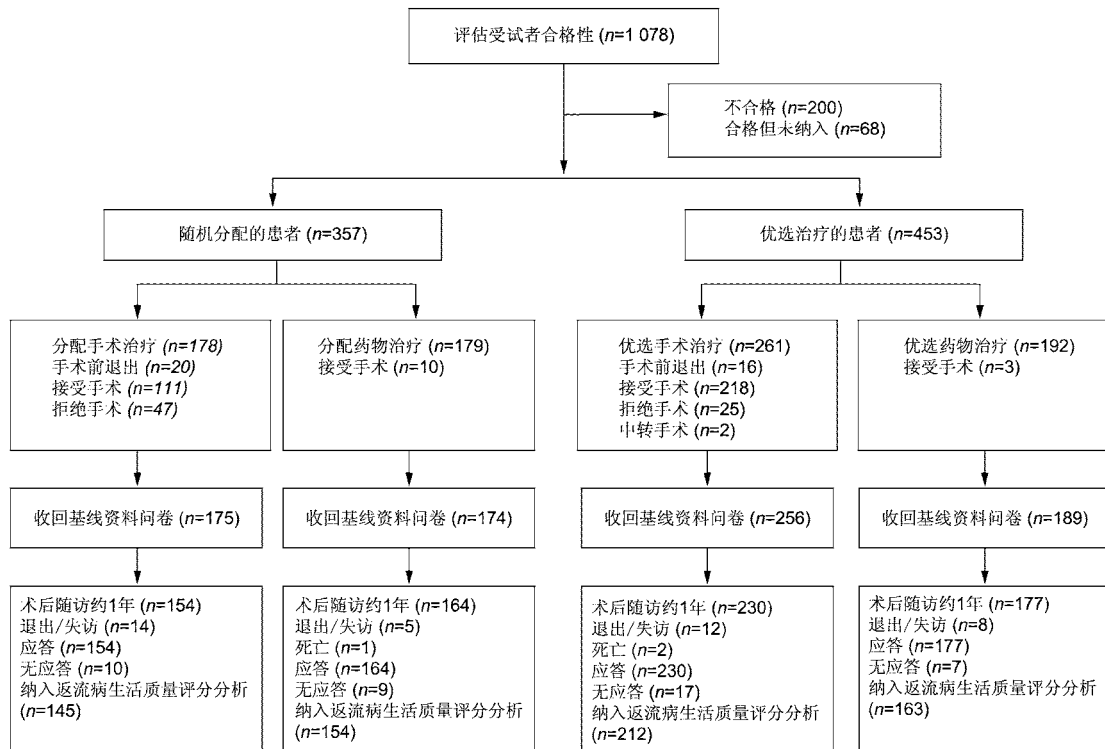


图 3 比较微创手术与药物治疗慢性胃食管返流病的流程图(改编自 Grant 等^[196])。流程图示包含一个优选治疗的非随机平行对照组的临床试验

解释——有些 RCT 的设计和实施较简单,用几句话就可将受试者在试验各阶段的流程描述清楚,特别是没有失访和排除病例时。在较复杂的研究中,读者可能很难辨别受试者是否接受了分配的治疗,为何没有接受分配的治疗,是否以及为何失访,是否以及为何在分析时被排除^[51]。这些信息之

所以很关键有几个原因。分组后被排除者不大可能代表研究中的所有受试者,例如,当其病情恶化或因治疗而受到伤害时,可能无法做随访评价^[22, 192]。

失访一般难以避免。要将失访造成的损耗与研究学者决定排除的病例区分开来,后者的原因有受试者不合格、退出治疗和对试验方案依从性差等。若

有受试者在分析时被排除,则可能得出错误结论,各组之间排除受试者的不均衡更意味着可能存在偏倚^[192-194]。因此,研究人员是否将最初随机分配到各组的所有受试者都纳入分析(意向性治疗分析,见条目 16 和框 6)就特别重要。知道未按分配接受干预措施或未完成治疗的受试者人数才能让读者确定,与理想情况相比,估计的疗效被低估的程度。

可能的话,还应报告所有被评定过是否符合纳入条件的总人数。尽管这一数字仅与外部真实性相关,而且可以说不如其他数字重要^[195],但仍是反映受试者能否代表所有符合纳入标准者的一个有用指标。

一篇对 1998 年发表在 5 种主要综合性医学期刊和内科学期刊上的 RCT 的综述发现,对受试者流向的报告常常不完整,尤其是接受所分配干预措施的受试者人数和失访人数^[51]。多达 20% 的文章未提供随机分配的受试者人数和分析时排除的人数这些基本信息^[51]。在采用了 CONSORT 推荐的受

试者试验流程图的文章中,这些信息的报告要全面得多。这一研究为在 CONSORT 声明修订版中新版流程图的设计提供了依据^[52-54]。建议使用的模版见图 1,要求报告数字的各项在表 3 中有详细描述。

有些信息并非总能知道,如完成受试者合格性评估的总人数^[14]。有些数字可能比其他数字更有用,依试验特性而定。有时候需要针对某项具体试验调整流程图的结构。在某些情况下,增加一些其他信息可能会很有用。例如,一项比较微创手术与药物治疗胃食管返流病的平行对照临床试验的流程图中还包括一个优选治疗的非随机平行对照组(见图 3)^[196]。

流程图的确切形式和内容可依试验的具体特征而变化。例如,许多关于手术或疫苗接种的试验不包括可能中止试验这一项。尽管 CONSORT 竭力推荐使用流程图来表示受试者在研究中的流向,但没有规定一个特定格式。

表 3 描述随机临床试验各阶段受试者流向时必须包括的信息

阶段	纳入人数	未纳入或排除的人数	理由
招募受试者	已评价的招募对象人数	不符合纳入标准的人数或符合纳入标准但拒绝参加试验的人数	这些数字反映参加试验的受试者对全体就诊患者的代表性;它们仅与评价外部真实性相关,它们常常未予报告。
随机分配	随机分配的受试者人数		用于定义试验大小和评价试验是否经意向性治疗分析的关键数字
分配治疗	按分配完成治疗的受试者人数,按组分别列出	未按分配完成治疗的受试者人数,按组分别列出	用于评价内部真实性和解释结果的重要数字;应该给出未接受分配的治疗的原因。
随访	按分配完成治疗的受试者人数,按组分别列出	未按分配完成治疗的受试者人数,按组分别列出	用于评价内部真实性和解释结果的重要数字;应该给出未完成治疗或随访的原因。
	按计划完成随访的受试者人数,按组分别列出	未按计划完成随访的受试者人数,按组分别列出	
分析	纳入主要分析的受试者人数,按组分别列出	未纳入主要分析的受试者人数,按组分别列出	用于评价试验是否经意向性治疗分析的关键数字;应该给出排除受试者的原因。

条目 13b. 随机分组后,各组脱落和被剔除的例数,并说明原因

解释——“仅有 1 名试验组妇女试验方案执行有偏差。她由于骨盆测量结果异常,被安排行择期剖腹产术,但产科主治医师认为可以尝试自然分娩。当第一期产程无进展时,对其施行了剖腹产术。”^[197]

“通过监测研究过程,撤消了 9 个中心,在这些中心,有些患者无法证明确实存在,或有其他严重违反临床试验管理规范的情况。”^[198]

解释——流程图中可以报告一些偏离计划的情况(见条目 13a),如未接受预定干预措施的受试者人数。如果受试者因为被发现不满足合格标准(见条目 16)而在随机分组后被排除(和意向性治疗分析的原则相反),则应该将其包括在流程图中。在发表的文章中仅用“偏离研究计划”一词并不足以证明

排除已经随机分组的受试者是合理的,而应该报告偏离研究计划的性质和排除已随机分组的受试者的确切原因。

条目 14a. 招募期和随访时间的长短,并说明具体日期

举例——“从 1993 年 2 月至 1994 年 9 月……招募年龄符合的受试者……受试者从随机分组时(基线)开始每 6 个月看一次门诊,时限 3 年。”^[199]

解释——知道试验何时开始和在什么时期内招募受试者有助于了解试验所处的历史背景。药物、手术或联合治疗手段日新月异,可能会影响试验期间受试者的常规处理。了解招募受试者的频次同样有用,尤其是对其他研究者而言。

随机分组后,随访期常不固定。在很多 RCT

中,结局指标是出现某事件的时间,每位受试者的随访均结束于某一特定日期,应报告此结束日期,报告最短、最长和中位随访期也是很有用的^[200, 201]。

一篇对肿瘤学期刊中采用了生存分析报告(多数不是 RCT)的综述^[201]发现,约 80%(132 篇报告中的 104 篇)报告了增补病例的起始和结束时间,但仅 24%(132 篇报告中的 32 篇)同时报告了随访结束时间。

条目 14b. 为什么试验中断或停止

举例——“中期分析时,总随访数包括估计在研究结束时可以收集到的患者-年总数的 63%,由 Lan-DeMets α 损耗函数法计算得出阈值为 0.009 5……中期分析时,与对照组比较,干预组的相对危险度为 0.37, P 值为 0.00073,低于阈值。数据和安全性监察委员会建议研究人员中止试验并为对照组施行环切术,之后把对照组患者叫到研究中心来建议他们行“医疗割礼”……因为试验中断,有些受试者在试验中止日期尚未完成全部随访过程,他们没有完成的随访在本文中仍记为‘按计划’。”^[202]

“2000 年 1 月,疫苗供应出现问题,因此必须在全国范围内临时用无细胞百日咳疫苗代替全细胞百白破/Hib 联合疫苗。由于这种疫苗的局部反应原性不同,我们决定提前终止试验。”^[203]

解释——可以认为,临床试验人员在只有很少的事件累积时不依据统计学原则任意进行非计划的中期分析有很高的极端随意“捕捉”数据的危险,这很可能大大地高估治疗的获益。^[204]

以数据驱动方式截取试验结果的试验与达到试验结果以外的目标之后才报告最终研究结果的试验相比,读者将只能从中得出价值很有限的推论。因此,RCT 报告应该说明试验为什么中止(见框 5)。报告还要透露影响做出中止试验决定的试验以外的因素,以及谁做出中止试验的决定,包括报告资助机构在商议和做出中止试验决定时发挥的作用^[134]。

一篇对因为已获益而比计划提前终止的 143 项 RCT 的报告所作的系统综述发现,这些试验在事件累积达到中位数 66 时终止,中位相对危险度估计为 0.47,累积事件数与效应值之间具有高度相关性,事件较少的小型试验得出的治疗效应最大(比值为 31,95% 可信区间为 12 至 82)^[134]。虽然越来越多发表在高影响力医学期刊上的临床试验报告了提前终止的信息,但仅有 0.1% 的试验报告了因为获益而提前终止,这与从模拟研究得出的估计^[205]以及对数据安全性和监察委员会调查的结果对比有很大差别^[206]。因而,很多临床试验以较少的病例数而报告较大的治疗效应,可能较原计划方案提前终止试验,但却未对此予以报告。

框 5 提前终止试验

RCT 可以在样本量达到目标时终止,或在事件数达到目标时终止,或在随访时间长度达到目标时终止,或因为到了计划结束的日期而终止。在这些情况下,试验将以一种不受试验结果影响的方式结束,终止试验不大可能引起结果偏倚。也有另一种情况,RCT 可以是因为中期分析的结果显示获益大于预期或试验观察的干预措施带来危害而比计划提前终止。RCT 还可以是在研究人员发现试验观察的干预措施与对照干预措施之间没有重要差异的证据时比计划提前终止(即因为无效而终止)。另外,由于资助结束和研究人员无法获得合格患者或要研究的干预措施使试验变得不可行,或者由于其他研究的结果使得试验要研究的问题显得无关紧要,试验可能因而提前终止。

全面报告试验终止的原因对循证决策十分重要(见条目 14b)。有研究者检查了因为获益而提前终止的 143 项临床试验的具体原因后发现,很多试验没有报告有关如何做出终止决定的关键的方法学信息,如预先计划的样本量($n=28$),中期分析后终止试验($n=45$),或是否根据某条中止原则作出终止决定($n=48$)^[134]。清单中的条目 7b 要求报告中期分析的时间安排、原因和次数,是预先计划的还是特设的,以及是否根据统计学原则和中止标准才得出终止试验的决定。此外,了解是否有独立的数据监察委员会参与分析(包括委员会由哪些人组成,特别要注意资助者的作用)以及由谁做出终止试验的决定也是有益的。常常是数据安全性和监察委员会提出建议,而资助者(主持试验者)或研究人员做出终止试验的决定。

明显地是由于试验结果以外的原因而提前终止的临床试验,以及因为到了计划结束的时间而终止的临床试验,不大可能因为终止试验而产生偏倚^[207]。在这些情况下,作者应该报告是否做过中期分析以及资助者是否得到过中期分析的结果。

迫于中期分析的结果而改变临床试验的干预措施,使得有可能因为试验进度加快和出现致命的严重情况而对有利的干预措施作出快速评价,这就要求更加仔细地报告提前终止试验的程序和决策过程^[208]。

条目 15. 用一张表格列出每一组的基线数据,包括人口学资料和临床特征

举例——见表 4。

解释——尽管受试者合格标准(见条目 4a)指出谁可以入选试验,但了解实际纳入的受试者的特征也很重要。这些信息可以让读者特别是临床医生判断试验结果与某个特定患者的相关程度。

随机临床试验旨在对仅在干预措施(治疗)方面不同的各组受试者进行比较。恰当的随机分配虽可防止选择性偏倚,却不能保证各组基线一致。然而,基线特征的任何差异都是机遇而非偏倚所致^[32]。应该对各试验组的重要人口学和临床特征等基线资料进行比较,

以便读者能够评估组间的相似性。对在试验之初也能测量的结局指标(如血压等),基线数据尤为重要。

最好用一张表格描述基线信息(见表 4)。对于连续变量,如体重或血压,在报告均数的同时,还要报告数据的变异度。各组的连续变量可用均数和标准差表示。当连续变量的数据呈偏态分布时,最好的办法是用中位数和百分位数范围(如第 25 和第 75 百分位数)来描述^[177]。标准误和可信区间不适合用来描述变异度——它们是推论性统计量而非描述性统计量。不应该把小数目的有序分类的变量(如疾病分期的 I 至 IV 期)当作连续变量处理,而应该报告每一类的例数和构成比^[48, 177]。

表 4 报告人口学和临床特征基线资料举例* (改编自 Yusuf 等^[209]的表 1)

	替米沙坦(N=2 954)	安慰剂(N=2 972)
年龄(岁)	66.9 (7.3)	66.9 (7.4)
性别(女性)	1 280 (43.3%)	1 267 (42.6%)
吸烟情况:		
还在吸烟	293 (9.9%)	289 (9.7%)
过去吸烟	1 273 (43.1%)	1 283 (43.2%)
人种:		
亚洲	637 (21.6%)	624 (21.0%)
阿拉伯	37 (1.3%)	40 (1.3%)
非洲	51 (1.7%)	55 (1.9%)
欧洲	1 801 (61.0%)	1 820 (61.2%)
土著	390 (13.2%)	393 (13.2%)
其他	38 (1.3%)	40 (1.3%)
血压(mm Hg)	140.7 (16.8/81.8) (10.1)	141.3 (16.4/82.0) (10.2)
心率(次/分)	68.8 (11.5)	68.8 (12.1)
胆固醇(mmol/L):		
总胆固醇	5.09 (1.18)	5.08 (1.15)
低密度脂蛋白胆固醇	3.02 (1.01)	3.03 (1.02)
高密度脂蛋白胆固醇	1.27 (0.37)	1.28 (0.41)
冠状动脉疾病	2 211 (74.8%)	2 207 (74.3%)
心肌梗死	1 381 (46.8%)	1 360 (45.8%)
心绞痛	1 412 (47.8%)	1 412 (47.5%)
外周动脉疾病	349 (11.8%)	323 (10.9%)
高血压	2 259 (76.5%)	2 269 (76.3%)
糖尿病	1 059 (35.8%)	1 059 (35.6%)

* 数据为均数(标准差)或计数(%)。

遗憾的是,基线差异的显著性检验仍很常见^[23, 32, 210];1997 年发表于核心综合性医学期刊的 50 项 RCT 报告中有一半这样做^[183]。这种显著性检验评价的是,观察到的基线差异系因机遇而发生的概率;然而我们都知道,任何差异均因机遇所致。检验基线差异并非一定是错误之举,只是不合逻辑^[211]。这种假设检验是多余的,会误导研究人员和读者。基线比较更应考虑所测量的变量的预测作用强度以及因机遇而发生的不均衡的程度^[211]。

条目 16. 各组纳入每一种分析的受试者数目(分母),以及是否按最初的分组分析

举例——“主要分析为意向性治疗分析,包括全部随机分配的病例。”^[212]

“阿仑膦酸钠组 1 例失访,故有 31 例的数据可用于意向性治疗分析。5 例被视为违背方案者……最终保留 26 例作符合方案分析(per-protocol analysis)。”^[213]

解释——各组受试者例数是结果分析的基本要素。尽管可以在流程图(见条目 13a)中列出纳入分

析的受试者例数,但对不同的结局指标而言,这些数目常常会不同。所有的分析都应给出每组受试者例数。对结局指标为二分类变量者,如危险比(risk ratio)和危险差(risk difference),还应报告分母或事件发生率。用分数的形式表达结果还有助于读者评估是否有随机分配的受试者在分析时被剔除。所以,结果不应仅用综合测量指标如相对危险度来表达。

有时候受试者可能并未接受整个干预措施,或者随机分配时错误地纳入了不合格的受试者。被广为推荐用于处理此类问题的方法是,不论分组后发

生了什么情况均按最初的分组对全部受试者进行分析(见框 6)。这种“意向性治疗”分析策略并非总是可以直截了当地执行。常常会有一些患者没有完成试验——他们可能脱落或退出已开始的治疗——因而最终未作评估。如果结局指标是死亡率,根据入组时登记的信息可以将这些患者纳入分析,但如果还有其他结局指标的数据缺失,则可能需要使用估算技术。“意向性治疗分析”一词常被不恰当地使用——例如,用于有受试者因未接受第一次试验用药而在结果分析时被排除的情况^[18]。

框 6 意向性治疗分析

RCT 的独特优点是在向受试者分配干预措施时避免偏倚(见框 1)。这一优点使其能得出强有力的使用其他研究设计无法求证的因果推论。为了充分保留随机分配的巨大好处,我们应该在分析时纳入全部随机分配的受试者,而且这些受试者要全部保留在所分配的试验组中。符合这两个条件的就是“意向性治疗”分析,这一分析策略广受推崇^[18, 223]。意向性治疗分析也即完全按照随机分配的情况对各组进行分析。严格的意向性治疗分析由于以下两个主要原因常常难以做到——部分受试者结局指标缺失和对试验方案的不依从。

结局指标缺失

很多临床试验人员将未观察结局指标的患者排除。这样做通常是合理的,但是,一旦有任何已经随机分配的受试者在分析时被排除,严格来说就不是意向性治疗分析。当然,大多数随机临床试验都会有一些观察值的缺失。试验人员总得作出选择,是排除没有最终结局指标数据的受试者,还是人为估计其缺失的结局指标数据^[225]。“完整病例”(或“可用病例”)分析只纳入那些已知结局指标的病例。当然,少量结局指标数据的缺失不会引起什么问题,半数临床试验有超过 10% 的已随机分配的患者可能存在结局指标数据缺失的情况^[226]。这种通常的做法会因为减少了样本量而使试验的把握度受到损失,如果失访影响到对患者治疗反应的评估,则很可能导致偏倚。当不同干预组之间脱落的例次或原因不同时,尤其要引起重视。

对于有结局指标数据缺失的受试者,只有对其结局指标数据进行估算后才能纳入分析(即他们的结局指标数据系从其他收集的信息估算得出)。估算缺失数据使得分析符合意向性治疗分析,但要求具有可能极难求证的强假设^[227]。简单估算法看似不错,但却不宜使用。具体来说,有一种应用很广的方法叫“末次观察值结转(last observation carried forward)”,即用受试者失访前最后一次测得的结局变量的值代替缺失的最终值。这一方法因为简单,故看似很好,但却可能会导致偏倚^[228],且无把握的估算也得不到承认^[229]。很多作者已对末次观察值结转提出非议^[229-231]。

试验方案不依从

另一个问题是,有些受试者可能没有完全遵从试验方案。常见的例子如:受试者不符合纳入标准(如诊断错误,年龄太小等),受试者接受了本该禁止的联合干预措施,受试者没有全部完成所分配的治疗,或接受了另一种治疗,甚至根本没有接受治疗。处理偏离试验方案的情况,简单的方法是置之不理,可以在分析时纳入全部受试者而不考虑其是否遵从了试验方案,此方法即所谓意向性治疗分析。因此,由于这类原因而排除任何受试者就不是意向性治疗分析了。

“修正的意向性治疗(modified intention-to-treat)”这个词被相当广泛地用于描述排除未完全遵从试验方案尤其是未接受规定的最少次数治疗的受试者的分析方法^[232]。有一个替代的词叫“符合方案(per protocol)”。尽管符合方案分析可能适用于某些场合,但应该注明这是一种非随机的观察性的比较方法。任何在分析时排除患者的做法都有悖于随机分配的本意,而且可能导致结果偏倚。

就像“意向性治疗分析”一样,通过诸如此类的说明来澄清到底纳入了哪些患者是不可靠的。因此,我们在 CONSORT 清单中删去了对意向性治疗分析的具体要求,而赞成清晰准确地描述每种分析纳入了哪些人。

相反,可将分析对象仅限于符合入选标准且完全执行研究方案规定的干预措施和结局测量的受试者,这种分析被称作“符合方案分析”(on-treatment 或 per protocol 分析)。分析时排除部分受试者会导致错误的结论。例如,某试验比较药物与手术治疗颈动脉狭窄,仅对有随访资料的受试者的分析显示,手术可降低一过性缺血发作、中风和死亡风险。但按最初分组纳入全部受试者进行意向性治疗分析发现,手术治疗并不比药物治疗更好^[214]。

意向性治疗分析可避免因受试者由于非随机地丢失所造成的偏倚,故总体来说较受青睐^[215-217]。不论作者是否使用“意向性治疗分析”一词,他们都应明确报告在每一种分析中纳入了哪些受试者及其数目(见条目 13)。未依从分配的治疗可能意味着意向性治疗分析将低估治疗措施的可能益处,此时可以考虑附加分析,如符合方案分析^[218, 219]。然而,必须注意的是,这类分析常有明显缺陷^[220]。

根据一篇对 2002 年发表于 10 种核心医学期刊的 403 篇 RCT 报告的综述,249 篇(62%)报告了在其主要分析中使用了意向性治疗分析。这个比例在遵守 CONSORT 声明的期刊中更高(70%对 48%)。在报告了使用意向性治疗分析的文章中,实际上仅 39%按最初的随机分组对全部受试者进行了分析,超过 60%的文章在其主要分析中存在数据缺失的情况^[221]。其他研究显示类似的结果^[18, 222, 223]。未报告排除受试者的试验在其他方法学质量方面比报告有受试者被排除的试验差^[173],这强烈提示有部分研究者排除了部分受试者但却未报告。另一项研究发现,报告意向性治疗分析的试验在试验设计和报告的其他方面也做得较好,例如描述样本量的计算^[224]。

条目 17a. 各组每一项主要和次要结局指标的结果,效应估计值及其精确性(如 95%可信区间)

举例——见表 5 和表 6。

表 5 报告每组的总结性结果举例(二分类变量结局指标)* (改编自 Mease 等^[103]的表 2)

终点指标	例数 (%)		危险差(95% CI)
	依那西普(n=30)	安慰剂 (n=30)	
主要终点指标			
12 周时达到牛皮癣关节炎缓解标准	26 (87)	7 (23)	63% (44-83)
次要终点指标			
符合 ACR 标准的患者比例:			
ACR20	22 (73)	4 (13)	60% (40-80)
ACR50	15 (50)	1 (3)	47% (28-66)
ACR70	4 (13)	0 (0)	13% (1-26)

* 参见条目 6a 的举例。ACR: American College of Rheumatology(美国风湿病学院)。

表 6 报告每组的总结性结果举例(连续变量结局指标)* (改编自 van Linschoten^[234]的表 3)

	锻炼疗法 (n=65)		对照 (n=66)		12 个月时调整后的差值* (95% CI)
	基线[均数(标准差)]	12 个月时[均数(标准差)]	基线[均数(标准差)]	12 个月时[均数(标准差)]	
功能评分(0-100)	64.4 (13.9)	83.2 (14.8)	65.9 (15.2)	79.8 (17.5)	4.52 (-0.73-9.76)
休息时疼痛评分(0-100)	4.14 (2.3)	1.43 (2.2)	4.03 (2.3)	2.61 (2.9)	-1.29 (-2.16--0.42)
活动时疼痛评分(0-100)	6.32 (2.2)	2.57 (2.9)	5.97 (2.3)	3.54 (3.38)	-1.19 (-2.22--0.16)

* 根据基线水平、年龄和症状持续时间调整后的功能评分

解释——对于每组每个结局指标,都应总结性地报告研究结果(例如,是否发生某事件的受试者人数及其分母,或测量指标的均数和标准差),同时给出组间差异,即所谓效应值。对二分类变量结局指标,效应值可以是危险比或相对危险度、比值比(odds ratio),或者危险差;对于生存时间数据,可以是危害比(hazard ratio)或中位生存时间的差值;而对于连续变量数据,通常是均值的差异。组间比较时,应给出差异的可信区间。常见的错误是报告各组结局指标独立的可信区间而非整个疗效的可信区

间^[233]。试验结果采用表格描述常较使用文字陈述更清晰,如表 5 和表 6。

对所有的结局指标,作者应给出可信区间以表示估计值的精确性(或不确定性)^[48, 235]。通常采用 95%可信区间,但有时也可用其他值。许多杂志要求或强烈建议使用可信区间^[236]。在差异无统计学意义时可信区间尤其有价值,因为在这种情况下常可从可信区间看出,统计结果并不能排除重要临床差异存在的可能。近年,尽管可信区间还未被所有医学专业采用,但其使用已明显增加^[233]。除可信

区间外还可再给出 P 值,但报告结果时不能仅有 P 值^[237, 238]。应报告研究设计中所有预先设定好的主要和次要终点结局指标的结果,而不是仅仅报告差异有统计学意义或“有用”的分析结果。选择性地报告研究工作是一个很普遍很严重的问题^[55, 57]。进行了中期分析的试验,应着重解释试验结束时的结果,而非中期结果^[239]。

若结果为二分类数据和生存时间数据,同时报告为获益需治病例数或出现危害需治病例数(number needed to treat for benefit or harm)也很

有用(见条目 21)^[240, 241]。

条目 17b. 对于二分类结局,建议同时提供相对效应值和绝对效应值

举例——“氧气依赖或死亡风险减少了 16% (95% CI 为 25% 至 7%)。绝对差为 -6.3% (95% CI 为 -9.9% 至 -2.7%),也即每对大约 16 名婴儿早期给药可望防止 1 名婴儿发生死亡或长期依赖氧气。”(也见表 7)^[242]

表 7 同时报告绝对效应值和相对效应值举例(改编自 OSIRIS 协作组^[242]的表 3)

主要结局指标	百分比(例数)		危险比(95% CI)	危险差(95% CI)
	早期给药 ($n=1\ 344$)	选择延迟给药($n=1\ 346$)		
“预产期”时发生死亡或氧气依赖	31.9 (429)	38.2 (514)	0.84 (0.75—0.93)	-6.3 (-9.9—-2.7)

解释——当主要结局指标是二分类变量时,应该同时报告(采用可信区间)相对效应(危险比或相对危险度,或者比值比)和绝对效应(危险差),因为仅用相对效应或绝对效应都不能反映试验效应的全貌及其意义。不同读者可能或偏爱相对危险度,或偏爱绝对危险度(absolute risk),但若治疗效应是用相对危险度描述,则更容易使医生和业外人对其作出过高的估计^[243-245]。危险差比相对危险度更加不易推广至其他人群,因为危险差的大小取决于非暴露组基线危险度的大小,而不同人群的基线危险度多不相同。如果疾病的结局指标是常用的指标,即便相对危险度接近一致,从公共卫生的角度看也可能在临床上有重要差异。相反,如果结局指标是很少用的指标,即使相对危险度较大,对公共卫生而言也不一定具有重要意义(尽管它对处于某种高危险的个体来说可能很重要)。

条目 18. 所做的其他分析的结果,包括亚组分析和校正分析,指出哪些是预先设定的分析,哪些是新尝试的分析

举例——“有一项研究认为,围手术期应用 β 受体阻滞剂的效果可能因基线危险度(baseline risk)而异,据此我们根据修订的心脏病危险指数评分系统预先设定了主要亚组分析。我们还根据性别、手术方式以及使用硬膜外麻醉还是脊髓麻醉作了预先设定的次要亚组分析。全部亚组分析我们都使用了 Cox 比例风险模型进行交互作用检验,以 $P < 0.05$ 为有统计学意义……表 3 示我们预先设定的亚组分析结果,表明疗效一致……如果确实存在亚组疗效,人们可能希望能将其检测出来,但我们的亚组分析

尚不足以检测这种亚组疗效可能存在的微小差异。”^[100]

解释——对同样的数据作多种分析有得出假阳性结果的危险^[246]。作者应该抵住诱惑不作过多的亚组分析^[183, 185, 247]。试验方案预先设定的分析(见条目 24)要比那些视数据情况而采取的分析可靠得多,故作者应该报告哪些分析是预先设定的。如果作了亚组分析,作者应报告分析了哪些亚组,若这些分析是预先设定的,则要报告为什么这样做,还要报告有多少种亚组分析是预先设定的。选择性地报告亚组分析可能导致偏倚^[248]。评价某个亚组时,问题不在于是否该亚组得到有统计学意义的结果,而在于各亚组的疗效是否有显著的组间差异。检验交互作用有助于对此问题作出判断,尽管这类检验的把握度特别低。如果正式作了交互作用检验(见条目 12b),则应予以报告,即要报告各亚组间干预措施效果的估计差值(及其可信区间),而不是只报告 P 值。

有一项调查显示,50 篇临床试验报告中 35 篇报告了亚组分析,而这其中仅 42% 采用了交互作用检验^[183]。通常很难判断试验方案中是否设定过亚组分析。另一项对发表于高影响力期刊上的外科临床试验报告的调查显示,72 篇中有 27 篇共报告了 54 种亚组分析,这其中 91% 系事后分析,仅 6% 的亚组分析采用了交互作用检验来评估是否存在亚组疗效^[249]。

类似的建议也适用于对基线变量作过校正后的分析。若有校正,则校正前和校正后的分析均应报告。作者应该说明,校正分析包括选择校正哪些变量是否是预先计划的。理想的情况是,试验方案要

写明对基线变量的校正是否使用协方差分析^[187]。因为变量的基线值有显著差异而对其进行校正有可能导致对疗效的估计产生偏倚^[187]。一项调查发现,在全部 25 篇报告了亚组分析的试验报告和 28 篇报告了校正分析的试验报告的 23 篇中,试验方案与发表的论文之间有不为人注意的出入^[92]。

条目 19. 各组出现的所有严重危害或意外效应(具体的指导建议参见“CONSORT for harms”^[42])

举例——“rBPI21 组(重组杀菌性/通透性增强蛋白)和安慰剂组出现任何不良事件的患者的比例相近,分别为 88.4% (168/190) 和 88.7% (180/203),但在 12 个出现了不良事件的人体系统中,有 11 个系统发生不良事件的患者比例 rBPI21 治疗组低于安慰剂治疗组……由研究人员判断发生严重不良事件的患者比例 rBPI21 组低于安慰剂组,分别为 27.9% (53/190) 和 36.5% (74/203)。仅报告了 3 例与药物相关的严重不良事件,且均发生在安慰剂组。”^[250]

解释——读者需要有关干预措施益处和危害的信息,以便权衡后作出明智的决定。不良反应的存在及其性质对判断某一项干预措施是否可以接受和是否会产生很大影响。在试验过程中观察到的不良事件不一定是干预措施所致,有些可能是正在被治疗的疾病所致。尽管不能检测出罕见的危害,随机临床试验是获取安全性和有效性数据的最佳方法。

很多 RCT 报告提供的不良事件信息不充分。对 1967 至 1999 年发表的 192 篇药物临床试验报告的调查显示,仅有 39% 充分报告了临床不良事件,29% 充分报告了药物毒性的实验室检测结果^[72]。最近,一项对提交给试验负责机构国立癌症研究所(National Cancer Institute)的不良事件数据与期刊论文中报告的信息所作的比较发现,低等级的不良事件在期刊论文中报告不足。期刊论文报告的高等级事件[“常见毒性标准(Common Toxicity Criteria)”第 3 至第 5 级]与实际不符,对于为什么将不良事件归因于所研究的药物,报告也不完整^[251]。而且,一篇对 2006 和 2007 年发表于 6 种综合性医学期刊中的临床试验报告的综述发现,尽管 133 篇报告中 89% 提及不良事件,但分别有 27% 和 48% 的论文未给出严重不良事件信息和患者因不良事件而退出的信息^[252]。

已有一份 CONSORT 声明扩展版为如何报告随机临床试验中的危害提供了详细建议^[42],正确报告的建议和举例可以在 CONSORT 网站(www.

consort-statement.org) 免费获取。这些建议是对“CONSORT 2010 声明”的补充,应该参考这些建议,尤其是在以危害的研究作为试验的关键目的的时候。简单地说,只要收集了不良事件的数据,就应该列出不良事件并加以定义,必要时参考有关标准。应描述用于收集不良事件数据和判定其原因的方法。对于每一个研究分组,都应报告每一项不良事件的绝对危险度(对反复发生的事件要采用合适的计量方法)以及因为危害而退出的受试者数目。最后,作者应权衡获益和危害并在文中予以讨论^[42]。

5.5 讨论

条目 20. 试验的局限性,报告潜在偏倚和不精确的原因,以及出现多种分析结果的原因(如果有这种情况的话)

举例——“男性患者居多(85%)是本研究的局限……我们使用了裸金属支架,因为直到病例收集的后期,药物洗脱支架才有供应。尽管后者也可视为一种局限,但已发表的资料表明,稳定型冠心病患者接受药物洗脱支架者与接受裸金属支架者比较,在死亡和心肌梗死方面并无获益(包括短期和长期)”^[253]

解释——在科学报告的讨论部分常常满支持作者研究发现的溢美之词^[254],而很少就其研究本身和研究结果提出正反两方面的经过斟酌的论点。一些杂志已尝试通过提倡作者在讨论试验结果时使用更加结构化的写作来纠正这个问题^[255, 256]。例如,《内科学年鉴(*Annals of Internal Medicine*)》建议作者对讨论部分使用结构式写作,包括以下几部分:(1)概述主要结果;(2)提出可能的机制并解释试验结果;(3)与其他已发表研究的相关结果进行比较(尽可能包含将现在所做研究的结果与以往所有相关研究的结果结合起来所作的系统综述);(4)这次研究的局限性(以及为减小和弥补这些局限而采取的措施);(5)酌情用一段简短的话总结该研究对临床和科研的意义^[255]。我们建议作者遵循这些合理的建议,也可以在讨论部分加上合适的小标题。

尽管研究报告往往未讨论局限性^[257],但指出并讨论一项研究的弱点特别重要^[258]。例如,一个外科治疗组报告,腹腔镜胆囊切除术(一种技术上比较难的手术操作)治疗急性胆囊炎,并发症发生率显著低于传统开腹胆囊切除术^[259]。但是,作者未讨论其结果存在的明显偏倚,即所有的腹腔镜胆囊切除术均由研究者亲自操刀,而 80% 的开腹胆囊切除术由实习医生完成。

作者还应讨论研究结果的不准确性。不准确性

的产生可能涉及研究的好几个方面,包括主要结局指标的测量(见条目 6a)或诊断(见条目 4a)。或许所使用的量表是一个在成年人中得到验证的量表,但却用于儿童,或者测量者未受过如何使用测量工具的培训。

要始终牢记统计学意义与临床重要性的区别。作者尤应避免一个常犯的的错误,即将结果无统计学意义解释为干预措施等效。无论 P 值如何,可信区间(见条目 17a)为判断试验结果是否与有临床意义的效果一致提供了有价值的信息^[120]。

使用多种比较的方法来评价试验结果时,作者应特别小心。这种多重性来自多种干预措施、结局测量方法、时间点、亚组分析以及其他因素。在这种情况下,某些有统计学意义的结果很可能仅由机遇产生。

条目 21. 试验结果被推广的可能性(外部可靠性,实用性)

举例——“干预措施在不同性别、各种年龄、各种运动类型,以及不同运动水平的运动员中实施,结果表明,各种年龄的运动员,从年轻新秀到以运动为休闲的年长运动员,以及介于两者之间者,都可因使用本文所述的防止踝部扭伤复发的训练方案而获益。通过纳入不必作医疗处理和要作医疗处理的运动员,我们涵盖的损伤程度范围很广。这提示本训练方案可用于治疗所有运动员。而且,因为将非体育运动所致的踝部扭伤视作类同于体育运动所致的踝部扭伤是合理的,故本训练方案也有益于普通人群。”^[260]

“本研究重复并扩展了 Clarke 等的工作,表明我们的认知行为预防方案可以在不同场所由最初制定该干预方案的研究小组以外的临床医生可靠而有效地实施。其效应值与之前报道的单中心抑郁症预防研究的结果一致,并且减轻抑郁情绪和身体症状两方面的效应值在不同研究地点较为稳健……在本次研究可推广性的临床试验中,我们选择一种与公共卫生相关的干预措施——常规处理进行比较研究……病例主要选自从劳动阶级到中产阶级的不同阶层,他们都有医疗保险。因为有证据显示,认知行为疗法对来自高收入家庭的青少年更有效,因此,检测本预防方案对具有更多样化的经济和种族背景的病例的效果将十分重要。”^[261]

解释——外部真实性,也称可推广性或适用性,是指研究结果能推广到其他情况的程度^[262]。内部真实性,指试验设计和实施消除偏移的可能性的程度,它是外部真实性的前提条件:一个有缺陷的试验

其结果是不可靠的,其外部真实性则无从谈起。不存在绝对的外部真实性,惟有对试验中未直接观察的一些明确限定的疾病来说,外部真实性才有意义。试验结果能推广到年龄、性别、疾病严重程度和并存疾病与试验纳入的受试者不同的个体或群体吗?试验结果适用于同类其他药物吗?适用于不同剂量、不同给药次数、不同给药途径以及有不同辅助疗法时的情况吗?有望在一级、二级和三级保健体系中获得类似结果吗?对试验中未予评价的相关结局指标有何影响?尤其关乎危害时,随访时间和疗程长短的重要性如何?^[263]

外部真实性是一种大体判断,因纳入试验中的受试者特征、试验环境、所观察的治疗方案以及所评价的结局指标而不同^[5, 136]。因此,充分报告受试者合格标准以及试验环境和地点(见条目 4b)、干预措施及其实施过程(见条目 5)、结局指标的定义(见条目 6),以及招募期和随访时间(见条目 14)都很重要。对照组受试者发生不同结局的比例(对照组危险度)也很重要。如流程图(见条目 13)所示,拒绝参加试验的合格受试者比例关系到试验的可推广性,因为它可能提示患者对干预措施的偏好或接受程度。同样也要考虑到临床医生可能存在的偏好^[264, 265]。

当试验结果应用于个体患者时,有几个问题很重要^[266-268]。尽管个体患者与试验中的患者或系统综述中的患者在治疗效果上可能会有差异,但多是量而不是质的差异。

尽管有一些很重要的例外^[268],试验中对较小范围患者有益的疗法(尤其是药物^[269])在实际医疗中还是常常得到较广泛的应用。已有一些评价外部真实性的框架,包括像使用积分法作“过程评估”^[270]这种定性研究,以及“对照检查清单”^[271]。结合基线危险度计算疗效的方法,如每增加一例获得有利结局的病例需要治疗的患者数和每发生一例副作用需要治疗的患者数,有助于评估那些与典型受试者特征不同的个体或群体患者的获益和危险之间的比重^[268, 272, 273]。最后,除了以患者为中心估计干预措施可能的利弊,临床医生还必须将其估计与患者的价值观和意愿治疗结合起来考虑。在评估试验结果对不同环境和不同干预措施的推广性时应考虑类似的问题。

条目 22. 与结果相对应的解释,权衡试验结果的利弊,并且考虑其他相关证据

举例——“1990 年以前发表的研究提示,预防性免疫疗法还可减少极低出生体重儿的院内感染。

然而,这些研究的病例数很少,采用的设计、剂型和剂量不同,纳入研究的人群各异。在这次大规模多中心随机对照临床试验中,对出生体重在 501 ~ 1 500 g 的早产儿预防性重复静脉注射免疫球蛋白,未能显著降低院内感染率。”^[274]

解释——读者可能需要了解当前的试验结果与其他 RCT 结果的关系。最好的实现办法是在试验报告的结果或讨论部分包含正式的系统综述^[83, 275-277]。这种同时作系统综述的办法对临床试验的作者来说可能不切实际,但通常可以引用一篇有关类似临床试验的系统综述。系统综述可能有助于读者评估该 RCT 的结果是否与同一主题的其他试验的结果相似。RCT 报告常未充分注意到这一点^[277]。可使用贝叶斯法对试验数据和以前的证据进行统计学合并处理^[278]。

我们建议,至少讨论应尽量系统化,而且是在全面检索的基础上,而非仅限于支持当前试验结果的研究^[279]。

5.6 其他信息

条目 23. 临床试验注册号 and 注册机构名称

举例——“本临床试验在 ClinicalTrials.gov 注册,注册号为 NCT00244842 [ClinicalTrials.gov]。”^[280]

解释——不完整地报告试验^[281, 282],选择性地报告试验结局,以及该作意向性治疗分析却作了符合方案分析等所导致的后果屡见报道^[55, 56, 283]。变相地重复发表临床试验也会引起很多问题,尤其是对系统综述的作者,他们可能无意中多次纳入同一项临床试验的结果^[284]。

为了减少和避免这些问题,在过去的 25 年中,要求临床试验在开始之前就注册,为其分配唯一的临床试验识别号,并记录试验的其他基本信息的呼声不断,这样做便于公开试验的详细基本情况^[285-288]。受最近发生的隐瞒数据这类严重问题的触动^[299],为实现随机临床试验注册又作了新的努力。世界卫生组织明确指出“对所有干预性临床试验进行注册是一种科学、伦理和道德责任”(www.who.int/ictrp/en)。为一项随机临床试验注册时,作者通常要按最基本的要求报告试验的一些信息,并获得一个唯一的临床试验注册号。

2004 年 9 月国际医学期刊编辑委员会(International Committee of Medical Journal Editors, ICMJE)修改了他们的规定,声明他们将只考虑发表在招募第一名受试者之前就已经注册的临床试验^[290]。这导致临床试验注册数增长惊人^[291]。

ICMJE 对哪些是合格的注册机构提出了指导意见(www.icmje.org/faq.pdf)。

根据最近一项对 165 种高影响因子医学期刊作者须知的调查,44 种期刊特别声明所有新近的临床试验都必须注册,并以此作为向他们的期刊投稿的基本要求^[292]。

作者应该提供注册机构的名称和试验的惟一注册号。如果其临床试验还没有注册,作者应写明没有注册并给出原因。

条目 24. 如果有的话,在哪里可以获取完整的试验方案

举例——“完整而详细的试验方案可以在附录中找到,和本文的全文一起见于 www.nejm.org。”^[293]

解释——完整的试验方案(而不是试验过程中某一特殊步骤的方案)十分重要,因为它预先设定了随机临床试验的方法,如主要结局指标(见条目 6a)。提供试验方案有助于限制试验开始后改变试验方法而不报告以及选择性地报告结局的做法(见条目 6b)。对随机临床试验方案中应包括的要素的描述见于他处^[294]。

为了确保感兴趣的读者获得试验方案,有几种选择作者可以考虑。正如以上举例所述,报告试验主要结果的期刊可以在其网站上提供试验方案。如果是开放获取期刊,试验结果和试验方案更易获取。有些期刊(如 *Trials*)发表试验方案,报告试验的主要结果时可以参考这一做法。临床试验注册(见条目 23)也可以确保试验方案的很多详情可供获取,因为合法的临床试验注册数据库里有最基本的临床试验特征资料,至少包括好几条试验方案条目和试验结果(www.who.int/ictrp/en)。临床试验的研究人员也可以通过项目负责人把他们的试验方案放在某个网站上。不管采用哪种机制,我们鼓励所有临床试验的研究人员都做到,让感兴趣的读者易于获取他们的试验方案。

条目 25. 资助和其他支持(如提供药品)的来源,提供资助者所起的作用

举例——“从国际计划(Plan International)获得基金资助干预措施,从维康基金会(Wellcome Trust)和联合国 HIV/AIDS 联合项目获得基金资助研究。提供资助者在研究设计、数据收集和分析、决定发表或文稿准备中均未承担任何任务。”^[295]

“本研究受葛兰素史克制药公司(GlaxoSmithKline Pharmaceuticals)资助。葛兰素史克参与研究设计

和实施,在试验期间提供后勤保障。其员工与研究人员一起制定了统计学分析方案,但由犹他大学(University of Utah)完成分析。Shaddy 博士和指导委员会成员准备了文稿。葛兰素史克得到允许审阅了文稿并提出修改意见,但只有作者保留对文稿内容的最终决定权。”^[296]

解释——作者应该报告临床试验的资助来源,因为这是读者用于评价一项临床试验的重要信息。研究显示,制药企业资助的研究与其他经费来源资助的研究相比更有可能得到有利于资助研究的公司生产的结果^[297-300]。一篇对 30 项关于资助问题的研究的系统综述发现,由制药企业资助的研究得到有利于资助者的结局的可能性是其他经费来源资助的研究的 4 倍(比值为 4.05,95% 可信区间为 2.98~5.51)^[297]。现在仍有很大一部分试验发表时不报告资助来源。漏报的情况有多严重很难量化。一项对 370 篇药物临床试验报告的调查发现,29% 未报告资助来源^[301]。另一项调查显示,519 篇于 2000 年 12 月发表并被 PubMed 收录的随机临床试验报告中仅 66% 报告了资助来源^[16]。

资助者多大程度参与试验及其对试验设计、实施、分析和报告的影响不一。因此作者详细描述资助者的作用十分重要。若资助者未参与试验,作者也应该报告资助者未参与试验。同样,作者还应报告任何其他提供支持者,如提供和准备药物或仪器设备,或分析数据和撰写文稿等^[302]。

6 非两组对照平行设计 RCT 的报告

CONSORT 的建议主要针对平行设计的有两个治疗组的 RCT。大多数 RCT 采用这种设计,但有相当一部分采用的不是这种设计:2000 年 12 月发表的 RCT 有 45% (233/519)^[16],2006 年 12 月为 39% (242/616)^[17]。

CONSORT 声明的大部分内容同样适用于各种设计类型的临床试验,但针对每一种设计都有一些问题需要补充。在 2001 年 CONSORT 声明修订版发表之前,CONSORT 工作组决定制定与特殊的临床试验设计相应的 CONSORT 声明母版的扩展版。报告群组随机临床试验^[40]和报告非劣效性与等效性临床试验^[39]的扩展版已经发表。由于缺乏资料,其他计划中的扩展版还没有完成;它们将涵盖以下设计类型的临床试验:多组平行设计、析因设计、交叉设计、个体内设计。

在报告群组设计临床试验,或者报告非劣效性或等效性临床试验时,作者除了参考本文之外,还应

参考 CONSORT 的相应建议。在此,我们暂对其他几种设计作一些评论。对每一种设计来说,都要同时在正文中和论文摘要中明晰地报告临床试验的设计。

多组(2 组以上)平行对照临床试验需要对标准 CONSORT 指南所作的调整最小。可对流程图作简单扩充。它与两组对照临床试验的主要区别在于,要阐明研究假设与多组之间有什么关联以及数据分析的相应方法,还有如何解释试验结果。对于析因设计临床试验,一般需要考虑干预措施之间有无交互作用。除了对是否接受试验所观察的每种干预措施的受试者进行全面比较,研究人员还要考虑报告每种治疗组合的结果^[303]。

在交叉设计临床试验中,每一名受试者按随机顺序接受两种(或多种)治疗;要另外处理的主要问题涉及数据的配对性,它会影响到试验设计和分析^[304]。类似问题也影响到个体内比较,在这种情况下,受试者同时接受两种治疗(常常针对配对的器官)。另外,由于分别有存在时间遗留效应和系统遗留效应的危险,因此在以上两种情况中,试验设计的选择都必须有充足的理由。

CONSORT 工作组打算发表 CONSORT 的扩展版以涵盖所有这些试验设计。此外,考虑到 CONSORT 通用版指南的这次重大更新,我们还将发表已有的群组随机临床试验和非劣效性与等效性临床试验指南的更新版。

7 讨论

除非研究人员能确保所作的比较无偏倚,否则对卫生保健干预措施的评价就可能产生误导。试验组的随机分配仍是惟一能消除选择性偏倚和混杂偏倚的方法。与随机临床试验相比,非随机临床试验更可能导致对治疗效果高估^[305, 306]。

然而,如果研究人员实施这类试验不当,即使 RCT 也会因偏倚而受损^[307]。最近一篇合并分析了几项方法学研究结果的系统综述发现,就主观性结局指标而言,使用了不恰当或不明确的分配隐藏方法的试验对疗效的估计比那些使用了恰当的隐藏方法的试验高 31%,而未使用盲法的试验对疗效的估计比使用盲法的试验高 25%^[153]。正如人们说担心的,两者之间有很强的相关性。

RCT 的设计和 implementation 需要方法学和临床专业知识,需要一丝不苟尽心尽力^[143, 308],还需要对意想不到的困难的高度警觉。撰写 RCT 报告同样要时时注意减少偏倚。不应让读者去推测,对试验方法的报告应完整和透明,以便读者容易区分无偏倚的研

究结果和有问题的研究结果。好的科研包含恰当地报告研究结果,而符合伦理道德的临床试验的实施需以良好的科研为基础^[309]。

我们希望这次对 CONSORT 说明文章的更新将帮助作者使用 2010 版 CONSORT,所以我们将以通用的术语解释充分报告临床试验的重要性。CONSORT 声明可以帮助研究人员设计将来的临床试验^[310],也可以指导审稿人和编辑评阅稿件。确实,我们鼓励审稿人和编辑使用 CONSORT 对照检查清单评估作者是否报告了这些条目。这种评估将会提高发表的临床试验的清晰性和透明性。CONSORT 是一份需要不断改进的文件,需要不断地动态评价和凝练,如有必要,还要修改。正因为如此,我们才有了这次对清单和说明文章的更新。随着新证据和批评意见的积累,我们会评估是否有必要再作下一次更新。

初版 CONSORT 声明自 1996 年发表以来,似乎已使采纳 CONSORT 的期刊发表的 RCT 报告的质量有所提高^[50-54]。还有人使用 CONSORT 的模板来提高其他研究设计的报告质量,如诊断性试验^[311]和观察性研究^[312]。

CONSORT 网站 (www.consort-statement.org) 已经开通,提供学习资料和与报告 RCT 相关的资料的存档数据库。该网站收入了很多试验实例,包括本文所举的全部例子。我们将继续向该数据库添加好的和差的报告实例,我们诚邀读者通过该网站与我们联系提交更多的建议。CONSORT 工作组将继续文献调查工作,查找涉及 RCT 报告问题的相关文章。我们诚邀这类文章的作者将这些信息告知我们。所有这些信息都将在 CONSORT 网站上获取,并定期更新。

400 多种核心的综合性和专业性期刊,以及生物医学编辑组织,包括国际医学期刊编辑委员会(ICMJE)、世界医学编辑协会(World Association of Medical Editors)和科学编辑理事会(Council of Science Editors),均对 CONSORT 给予了正式支持。我们诚邀其他关注临床试验报告质量的期刊使用 CONSORT 声明,并通过我们的网站与我们联系,以便我们知道他们支持 CONSORT。为这些集体努力作出最大奉献的应该是那些不论出于何种原因向卫生保健团体寻求治疗的人。

我们感谢 Frank Davidoff 和 Tom Lang,他们参与了 2001 版 CONSORT 说明与详述文本的起草。特别感谢渥太华 CONSORT 协调员 Mary Ocampo 帮助发放本说明与详述文稿和 CONSORT 声明。

为 CONSORT 2010 作出过贡献的 CONSORT 工作组成员: Douglas G Altman, Centre for Statistics in Medicine, University of Oxford, UK; Virginia Barbour, *PLoS Medicine*, UK; Jesse A Berlin, Johnson & Johnson Pharmaceutical Research and Development, USA; Isabelle Boutron, University Paris 7 Denis Diderot, Assistance Publique des Hôpitaux de Paris, INSERM, France; PJ Devereaux, McMaster University, Canada; Kay Dickersin, Johns Hopkins Bloomberg School of Public Health, USA; Diana Elbourne, London School of Hygiene & Tropical Medicine, UK; Susan Ellenberg, University of Pennsylvania School of Medicine, USA; Val Gebski, University of Sydney, Australia; Steven Goodman, *Journal of the Society for Clinical Trials*, USA; Peter C Gotzsche, Nordic Cochrane Centre, Denmark; Trish Groves, *BMJ*, UK; Steven Grunberg, American Society of Clinical Oncology, USA; Brian Haynes, McMaster University, Canada; Sally Hopewell, Centre for Statistics in Medicine, University of Oxford, UK; Astrid James, *Lancet*; Peter Juhn, Johnson & Johnson, USA; Philippa Middleton, University of Adelaide, Australia; Don Minckler, University of California Irvine, USA; David Moher, Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Canada; Victor M Montori, Knowledge and Encounter Research Unit, Mayo Clinic College of Medicine, USA; Cynthia Mulrow, *Annals of Internal Medicine*, USA; Stuart Pocock, London School of Hygiene & Tropical Medicine, UK; Drummond Rennie, *JAMA*, USA; David L Schriger, *Annals of Emergency Medicine*, USA; Kenneth F Schulz, Family Health International, USA; Iveta Simera, EQUATOR Network, UK; Elizabeth Wager, Sideview, UK.

资助:我们衷心感谢以下提供经济资助者。它们是: United Kingdom National Institute for Health Research; Canadian Institutes of Health Research; Presidents Fund, Canadian Institutes of Health Research; Johnson & Johnson; *BMJ*; the American Society for Clinical Oncology.

为了鼓励更多的人使用“CONSORT 2010 声明”,本文可以在 bmj.com 免费获取,也将在 *Journal of Clinical Epidemiology* 发表。本文作者共同拥有本文的版权。有关如何进一步使用本文的详细信息见 CONSORT 网站 (www.consort-statement.org)。

REFERENCES

- 1 Rennie D. CONSORT revised — improving the reporting of randomized trials. *JAMA*. 2001; 285 (15): 2006-2007.
- 2 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological

- quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995; 273(5): 408-412.
- 3 Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. *Consolidated Standards of Reporting Trials*. *JAMA*. 1998; 279(18): 1489-1491.
 - 4 Kjaergard LL, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy. 7th. *Cochrane Colloquium*. Rome. Italy. 1999.
 - 5 Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001; 323(7303): 42-46.
 - 6 Veldhuyzen van Zanten SJ, Cleary C, Talley NJ, Peterson TC, Nyrén O, Bradley LA, Verlinden M, Tytgat GN. Drug treatment of functional dyspepsia: a systematic analysis of trial methodology with recommendations for design of future trials. *Am J Gastroenterol*. 1996; 91(4): 660-673.
 - 7 Talley NJ, Owen BK, Boyce P, Paterson K. Psychological treatments for irritable bowel syndrome: a critique of controlled treatment trials. *Am J Gastroenterol*. 1996; 91(2): 277-283.
 - 8 Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol*. 2000; 136(3): 381-385.
 - 9 Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology*. 1999; 30(5): 1134-1138.
 - 10 Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA*. 1966; 195(13): 1123-1128.
 - 11 Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J*. 1977; 1(6053): 85-87.
 - 12 Hall JC, Hill D, Watts JM. Misuse of statistical methods in the Australasian surgical literature. *Aust N Z J Surg*. 1982; 52(5): 541-543.
 - 13 Altman DG. Statistics in medical journals. *Stat Med*. 1982; 1(1): 59-71.
 - 14 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(7): 426-432.
 - 15 Altman DG. The scandal of poor medical research. *BMJ*. 1994; 308(6924): 283-284.
 - 16 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*. 2005; 365(9465): 1159-1162.
 - 17 Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010; 340: c723.
 - 18 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999; 319(7211): 670-674.
 - 19 Lai TY, Wong VW, Lam RF, Cheng AC, Lam DS, Leung GM. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. *Ophthalmic Epidemiol*. 2007; 14(6): 390-398.
 - 20 Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet*. 1996; 347(8998): 363-366.
 - 21 Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA*. 1998; 280(3): 247-249.
 - 22 Altman DG. Randomisation. *BMJ*. 1991; 302(6791): 1481-1482.
 - 23 Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA*. 1994; 272(2): 125-128.
 - 24 Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ*. 1948; 2: 769-782.
 - 25 Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol*. 1998; 41(2): 245-256.
 - 26 Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990; 1(6): 421-429.
 - 27 Armitage P. The role of randomization in clinical trials. *Stat Med*. 1982; 1(4): 345-352.
 - 28 Kleijnen J, Gøtzsche PC, Kunz R, Oxman AD, Chalmers I. So what's so special about randomisation. In: Maynard A, Chalmers I, eds. *Non-random reflections on health services research*. *BMJ Books*. 1997: 93-106.
 - 29 Chalmers I. Assembling comparison groups to assess the effects of health care. *J R Soc Med*. 1997; 90(7): 379-386.
 - 30 Nicolucci A, Grilli R, Alexanian AA, Apolone G, Torri V, Liberati A. Quality, evolution, and clinical implications of randomized, controlled trials on the treatment of lung cancer. A lost opportunity for meta-analysis. *JAMA*. 1989; 262(15): 2101-2107.
 - 31 Ah-See KW, Molony NC. A qualitative assessment of randomized controlled trials in otolaryngology. *J*

- Laryngol Otol. 1998; 112(5): 460-463.
- 32 Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet*. 1990; 335(8682): 149-153.
- 33 Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*. 1998; 317(7167): 1181-1184.
- 34 DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med*. 1982; 306(22): 1332-1337.
- 35 A proposal for structured reporting of randomized controlled trials. The Standards of Reporting Trials Group. *JAMA*. 1994; 272(24): 1926-1931.
- 36 Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med*. 1994; 121(11): 894-895.
- 37 Rennie D. Reporting randomized controlled trials. An experiment and a call for responses from readers. *JAMA*. 1995; 273(13): 1054-1055.
- 38 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996; 276(8): 637-639.
- 39 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006; 295(10): 1152-1160.
- 40 Campbell MK, Elbourne DR, Altman DG; CONSORT group. CONSORT statement: extension to cluster randomised trials. *BMJ*. 2004; 328(7441): 702-708.
- 41 Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Oxman AD, Moher D; CONSORT group; Pragmatic Trials in Healthcare (PractiHC) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008; 337: a2390.
- 42 Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D; CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004; 141(10): 781-788.
- 43 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P; CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*. 2008; 148(4): 295-309.
- 44 Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C; CONSORT Group. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med*. 2006; 144(5): 364-367.
- 45 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF; CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med*. 2008; 5(1): e20.
- 46 Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996; 276(16): 1339-1341.
- 47 Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *BMJ*. The *BMJ* Economic Evaluation Working Party. *BMJ*. 1996; 313(7052): 275-283.
- 48 Lang TA, Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers. ACP. 1997.
- 49 Davidoff F. News from the International Committee of Medical Journal Editors. *Ann Intern Med*. 2000; 133(3): 229-231.
- 50 Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 2006; 185(5): 263-267.
- 51 Egger M, Jüni P, Bartlett C; CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. *JAMA*. 2001; 285(15): 1996-1999.
- 52 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med*. 2001; 134(8): 657-662.
- 53 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001; 285(15): 1987-1991.
- 54 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001; 357(9263): 1191-1194.
- 55 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004; 291(20): 2457-2465.
- 56 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial

- protocols accepted by The Lancet. *Lancet*. 2008; 372 (9634): 201.
- 57 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008; 3(8): e3081.
- 58 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T; CONSORT GROUP (Consolidated Standards of Reporting Trials). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001; 134(8): 663-694.
- 59 Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials*. 2010; 11: 32.
- 60 Pocock SJ. *Clinical trials: a practical approach*. John Wiley. 1983.
- 61 Meinert CL. *Clinical trials: design, conduct and analysis*. Oxford University Press. 1986.
- 62 Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. 3rd ed. Springer. 1998.
- 63 Bolliger CT, Zellweger JP, Danielsson T, van Biljon X, Robidou A, Westin A, Perruchoud AP, Säwe U. Smoking reduction with oral nicotine inhalers: double blind, randomised clinical trial of efficacy and safety. *BMJ*. 2000; 321(7257): 329-333.
- 64 Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof*. 2002; 25 (1): 38-64.
- 65 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF; CONSORT Group. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet*. 2008; 371(9609): 281-283.
- 66 The impact of open access upon public health. *PLoS Med*. 2006; 3(5): e252.
- 67 Harbourt AM, Knecht LS, Humphreys BL. Structured abstracts in MEDLINE, 1989-1991. *Bull Med Libr Assoc*. 1995; 83(2): 190-195.
- 68 Harris AH, Standard S, Brunning JL, Casey SL, Goldberg JH, Oliver L, Ito K, Marshall JM. The accuracy of abstracts in psychology journals. *J Psychol*. 2002; 136(2): 141-148.
- 69 Pitkin RM, Branagan MA, Burmeister LF. Accuracy of data in abstracts of published research articles. *JAMA*. 1999; 281(12): 1110-1111.
- 70 Ward LG, Kendrach MG, Price SO. Accuracy of abstracts for original research articles in pharmacy journals. *Ann Pharmacother*. 2004; 38(7-8): 1173-1177.
- 71 Gøtzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ*. 2006; 333(7561): 231-234.
- 72 Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA*. 2001; 285(4): 437-443.
- 73 Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med*. 1990; 113(1): 69-76.
- 74 Taddio A, Pain T, Fassos FF, Boon H, Ilersich AL, Einarson TR. Quality of nonstructured and structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association. *CMAJ*. 1994; 150(10): 1611-1615.
- 75 Wager E, Middleton P. Technical editing of research reports in biomedical journals. *Cochrane Database Syst Rev*. 2008. MR000002.
- 76 Hartley J, Sydes M, Blurton A. Obtaining information accurately and quickly: Are structured abstracts more efficient? *J Inform Sci*. 1996; 22(5): 349-356.
- 77 Gilligan D, Nicolson M, Smith I, Groen H, Dalesio O, Goldstraw P, Hatton M, Hopwood P, Manegold C, Schramel F, Smit H, van Meerbeeck J, Nankivell M, Parmar M, Pugh C, Stephens R. Preoperative chemotherapy in patients with resectable non-small cell lung cancer: results of the MRC LU22/NVALT 2/EORTC 08012 multicentre randomised trial and update of systematic review. *Lancet*. 2007; 369 (9577): 1929-1937.
- 78 Sandler AD, Sutton KA, DeWeese J, Girardi MA, Sheppard V, Bodfish JW. Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive developmental disorder. *N Engl J Med*. 1999; 341(24): 1801-1806.
- 79 World Medical Association. Declaration of Helsinki: ethical principle for medical research involving human subjects. 59th WMA General Assembly, Seoul 2008; (accessed 2 June 2009).
- 80 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*. 1992; 327(4): 248-254.
- 81 Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clin Trials*. 2005; 2(3): 218-229.

- 82 Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ*. 1996; 313(7069): 1390-1393.
- 83 Sadler LC, Davison T, McCowan LM. A randomised controlled trial and meta-analysis of active management of labour. *BJOG*. 2000; 107(7): 909-915.
- 84 Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke*. 1998; 29(10): 2203-2210.
- 85 Blumer JL, Findling RL, Shih WJ, Soubrane C, Reed MD. Controlled clinical trial of zolpidem for the treatment of insomnia associated with attention-deficit/hyperactivity disorder in children 6 to 17 years of age. *Pediatrics*. 2009; 123(5): e770-e776.
- 86 Sabatine MS, Antman EM, Widimsky P, Ebrahim IO, Kiss RG, Saaiman A, Polasek R, Contant CF, McCabe CH, Braunwald E. Otimixaban for the treatment of patients with non-ST-elevation acute coronary syndromes (SEPIA-ACS1 TIMI 42): a randomised, double-blind, active-controlled, phase 2 trial. *Lancet*. 2009; 374(9692): 787-795.
- 87 Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, Elbourne DR, McLeer SK, Parmar MK, Pocock SJ, Spiegelhalter DJ, Sydes MR, Walker AE, Wallace SA; DAMOCLES study group. Issues in data monitoring and interim analysis of trials. *Health Technol Assess*. 2005; 9(7):1-238, iii-iv.
- 88 Gallo P, Krams M. PhRMA Working Group on adaptive designs, "White Paper." *Drug Inf J*. 2006; 40(4): 421-482.
- 89 Brown CH, Ten Have TR, Jo B, Dagne G, Wyman PA, Muthen B, Gibbons RD. Adaptive designs for randomized trials in public health. *Annu Rev Public Health*. 2009; 30: 1-25.
- 90 Kelly PJ, Sooriyachchi MR, Stallard N, Todd S. A practical comparison of group-sequential and adaptive designs. *J Biopharm Stat*. 2005; 15(4): 719-738.
- 91 Pildal J, Chan AW, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ*. 2005; 330(7499): 1049.
- 92 Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ*. 2008; 337: a2299.
- 93 Ndekha MJ, van Oosterhout JJ, Zijlstra EE, Manary M, Saloojee H, Manary MJ. Supplementary feeding with either ready-to-use fortified spread or corn-soy blend in wasted adults starting antiretroviral therapy in Malawi: randomised, investigator blinded, controlled trial. *BMJ*. 2009; 338: b1867.
- 94 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet*. 2005. 365(9453): 82-93.
- 95 Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. National Surgical Adjuvant Breast and Bowel Program. Pediatric Oncology Group. *J Clin Epidemiol*. 1998; 51(2): 69-79.
- 96 Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol*. 2000; 53(10): 973-979.
- 97 Gandhi M, Ameli N, Bacchetti P, Sharp GB, French AL, Young M, Gange SJ, Anastos K, Holman S, Levine A, Greenblatt RM. Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS*. 2005; 19(16): 1885-1896.
- 98 Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery*. 1996; 119(4): 466-472.
- 99 Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. *Arch Intern Med*. 2008; 168(2): 133-135.
- 100 POISE Study Group, Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC, Xavier D, Chrolavicius S, Greenspan L, Pogue J, Pais P, Liu L, Xu S, Málaga G, Avezum A, Chan M, Montori VM, Jacka M, Choi P. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet*. 2008; 371(9627): 1839-1847.
- 101 Rannou F, Dimet J, Boutron I, Baron G, Fayad F, Macé Y, Beaudreuil J, Richette P, Ravaud P, Revel M, Poiraudau S. Splint for base-of-thumb osteoarthritis: a randomized trial. *Ann Intern Med*. 2009; 150(10): 661-669.
- 102 Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ*. 2008; 336(7659): 1472-1474.
- 103 Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. *Lancet*. 2000; 356(9227): 385-390.
- 104 McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. 3rd ed. New York:

- Oxford University Press. 2006.
- 105 Streiner D, Norman C. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press. 2003.
- 106 Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials*. 2007; 8: 39.
- 107 Sanders C, Egger M, Donovan J, Tallon D, Frankel S. Reporting on quality of life in randomised controlled trials: bibliographic study. *BMJ*. 1998; 317(7167): 1191-1194.
- 108 Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials*. 1989; 10(1): 31-56.
- 109 Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry*. 2000; 176: 249-252.
- 110 Jadad AR, Boyle M, Cunnigham C, Kim M, Schachar R. Treatment of Attention-Deficit/Hyperactivity Disorder. Evidence Report/Technology Assessment No. 11. Rockville, MD: U. S. Department of Health and Human Services, Public Health Service, Agency for Healthcare Research and Quality. AHQR publication No. 00-E005. 1999.
- 111 Schachter HM, Pham B, King J, Langford S, Moher D. The efficacy and safety of methylphenidate in attention deficit disorder: A systematic review and meta-analysis. Prepared for the Therapeutics Initiative, Vancouver, B.C., and the British Columbia Ministry for Children and Families. 2000.
- 112 Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet*. 2001; 357(9266): 1385-1390.
- 113 Chan AW, Krleza-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ*. 2004; 171(7): 735-740.
- 114 Vermeulen H, Hofland J, Legemate DA, Ubbink DT. Intravenous fluid restriction after major abdominal surgery: a randomized blinded clinical trial. *Trials*. 2009; 10: 50.
- 115 Fox K, Ford I, Steg PG, Tendera M, Ferrari R; BEAUTIFUL Investigators. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2008; 372(9641): 807-816.
- 116 Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ*. 1995; 311(7013): 1145-1148.
- 117 Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter. *PLoS Med*. 2008; 5(1): e4.
- 118 Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005; 365(9467): 1348-1353.
- 119 Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA*. 2002; 288(3): 358-362.
- 120 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995; 311(7003): 485.
- 121 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994; 272(2): 122-124.
- 122 Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med*. 1978; 299(13): 690-694.
- 123 Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ*. 2009; 338: b1732.
- 124 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med*. 1984; 3(4): 409-422.
- 125 Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994; 121(3): 200-206.
- 126 Galgiani JN, Catanzaro A, Cloud GA, Johnson RH, Williams PL, Mirels LF, Nassar F, Lutz JE, Stevens DA, Sharkey PK, Singh VR, Larsen RA, Delgado KL, Flanigan C, Rinaldi MG. Comparison of oral fluconazole and itraconazole for progressive, nonmeningeal coccidioidomycosis. A randomized, double-blind trial. *Mycoses Study Group. Ann Intern Med*. 2000; 133(9): 676-686.
- 127 Connolly SJ, Pogue J, Hart RG, Hohnloser SH, Pfeffer M, Chrolavicius S. Effect of clopidogrel added to aspirin in patients with atrial fibrillation. *N Engl J Med*. 2009; 360(20): 2066-2078.
- 128 Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics*. 1987; 43(1): 213-223.
- 129 Berry DA. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat Med*. 1985; 4(4): 521-526.

- 130 Pocock SJ. When to stop a clinical trial. *BMJ*. 1992; 305(6847): 235-240.
- 131 DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. *Lancet*. 1999; 354(9194): 1983-1988.
- 132 Buyse M. Interim analyses, stopping rules and data monitoring in clinical trials in Europe. *Stat Med*. 1993; 12(5-6): 509-520.
- 133 Sydes MR, Altman DG, Babiker AB, Parmar MK, Spiegelhalter DJ; DAMOCLES Group. Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study. *Clin Trials*. 2004; 1(1): 48-59.
- 134 Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005; 294(17): 2203-2209.
- 135 Coutinho IC, Ramos de Amorim MM, Katz L, Bandeira de Ferraz AA. Uterine exteriorization compared with in situ repair at cesarean delivery: a randomized controlled trial. *Obstet Gynecol*. 2008; 111(3): 639-647.
- 136 Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. BMJ Books. 2001.
- 137 Creinin MD, Meyn LA, Borgatta L, Barnhart K, Jensen J, Burke AE, Westhoff C, Gilliam M, Dutton C, Ballagh SA. Multicenter comparison of the contraceptive ring and patch: a randomized controlled trial. *Obstet Gynecol*. 2008; 111(2 pt 1): 267-277.
- 138 Tate DF, Jackvony EH, Wing RR. Effects of internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. *JAMA*. 2003; 289(14): 1833-1836.
- 139 Lachin JM. Properties of simple randomization in clinical trials. *Control Clin Trials*. 1988; 9(4): 312-326.
- 140 Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976; 34(6): 585-612.
- 141 Schulz KF, Grimes DA. *The Lancet handbook of essential concepts in clinical research*. Elsevier. 2006.
- 142 Altman DG, Bland JM. How to randomise. *BMJ*. 1999; 319(7211): 703-704.
- 143 Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995; 274(18): 1456-1458.
- 144 Enas GG, Enas NH, Spradlin CT, Wilson MG, Wiltse CG. Baseline comparability in clinical trials: prevention of poststudy anxiety. *Drug Inf J*. 1990; 24: 541-548.
- 145 Treasure T, MacRae KD. Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ*. 1998; 317(7155): 362-363.
- 146 Sinei SK, Schulz KF, Lamptey PR, Grimes DA, Mati JK, Rosenthal SM, Rosenberg MJ, Riara G, Njage PN, Bhullar VB, Ogembo HV. Preventing IUCD-related pelvic infection: the efficacy of prophylactic doxycycline at insertion. *Br J Obstet Gynaecol*. 1990; 97(5): 412-419.
- 147 Radford JA, Landorf KB, Buchbinder R, Cook C. Effectiveness of low-Dye taping for the short-term treatment of plantar heel pain: a randomised trial. *BMC Musculoskelet Disord*. 2006; 7: 64.
- 148 Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med*. 1987; 6(3): 315-328.
- 149 Pocock SJ. Statistical aspects of clinical trial design. *Statistician*. 1982; 31(1): 1-18.
- 150 Haag U. Technologies for automating randomized treatment assignment in clinical trials. *Drug Inf J*. 1998; 32: 11.
- 151 Piaggio G, Elbourne D, Schulz KF, Villar J, Pinol AP, Gülmezoglu AM; WHO Research Group to evaluate Misoprostol in the Management of the Third Stage of Labour. The reporting of methods for reducing and detecting bias: an example from the WHO Misoprostol Third Stage of Labour equivalence randomised controlled trial. *BMC Med Res Methodol*. 2003; 3: 19.
- 152 Pildal J, Hróbjartsson A, Jørgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol*. 2007; 36(4): 847-857.
- 153 Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008; 336(7644): 601-605.
- 154 McCandlish R, Bowler U, van Asten H, Berridge G, Winter C, Sames L, Garcia J, Renfrew M, Elbourne D. A randomised controlled trial of care of the perineum during second stage of normal labour. *Br J Obstet Gynaecol*. 1998; 105(12): 1262-1272.

- 155 Webster J, Clarke S, Paterson D, Hutton A, van Dyk S, Gale C, Hopkins T. Routine care of peripheral intravenous catheters versus clinically indicated replacement: randomised controlled trial. *BMJ*. 2008; 337: a339.
- 156 Smith SA, Shah ND, Bryant SC, Christianson TJ, Bjornsen SS, Giesler PD, Krause K, Erwin PJ, Montori VM; Evidens Research Group. Chronic care model and shared care in diabetes: randomized trial of an electronic decision support system. *Mayo Clin Proc*. 2008; 83(7): 747-757.
- 157 Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, McManus K, Champagne CM, Bishop LM, Laranjo N, Leboff MS, Rood JC, de Jonge L, Greenway FL, Loria CM, Obarzanek E, Williamson DA. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med*. 2009; 360(9): 859-873.
- 158 Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med*. 1998; 72(3): 389-433.
- 159 Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL, Fallen EL, Taylor DW. Effect of encouragement on walking test performance. *Thorax*. 1984; 39(11): 818-822.
- 160 Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials*. 1996; 17(4): 285-290.
- 161 Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. *JAMA*. 1975; 231(10): 1038-1042.
- 162 Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*. 1994; 44(1): 16-20.
- 163 Carley SD, Libetta C, Flavin B, Butler J, Tong N, Sammy I. An open prospective randomised trial to reduce the pain of blood glucose testing: ear versus thumb. *BMJ*. 2000; 321(7252): 20.
- 164 Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med*. 2002; 136(3): 254-259.
- 165 Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ*. 2000; 321(7259): 504.
- 166 Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol*. 2002; 55(8): 787-790.
- 167 Cheng K, Smyth RL, Motley J, O'Hea U, Ashby D. Randomized controlled trials in cystic fibrosis (1966-1997) categorized by time, design, and intervention. *Pediatr Pulmonol*. 2000; 29(1): 1-7.
- 168 Lang T. Masking or blinding? An unscientific survey of mostly medical journal editors on the great debate. *Med Gen Med*. 2000; 2(1): E25.
- 169 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, Bhandari M, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001; 285(15): 2000-2003.
- 170 Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. *Clin Trials*. 2006; 3(4): 360-365.
- 171 Meinert CL. Masked monitoring in clinical trials — blind stupidity? *N Engl J Med*. 1998; 338(19): 1381-1382.
- 172 Mills E, Prousky J, Raskin G, Gagnier J, Rachlis B, Montori VM, Juurlink D. The safety of over-the-counter niacin. A randomized placebo-controlled trial [ISRCTN18054903]. *BMC Clin Pharmacol*. 2003; 3: 4.
- 173 Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ*. 1996; 312(7033): 742-744.
- 174 Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ*. 2004; 328(7437): 432.
- 175 Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ*. 2004; 328(7448): 1136.
- 176 Astrup A, Rössner S, Van Gaal L, Rissanen A, Niskanen L, Al Hakim M, Madsen J, Rasmussen MF, Lean ME; NN8022-1807 Study Group. Effects of liraglutide in the treatment of obesity: a randomised, double-blind, placebo-controlled study. *Lancet*. 2009; 374(9701): 1606-1616.
- 177 Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with confidence: confidence intervals and statistical guidelines*. 2nd ed. *BMJ Books*. 2000: 171-190.
- 178 Altman DG, Bland JM. Statistics notes. Units of analysis. *BMJ*. 1997; 314(7098): 1874.
- 179 Bolton S. Independence and statistical inference in clinical trial designs: a tutorial review. *J Clin Pharmacol*. 1998; 38(5): 408-412.
- 180 Greenland S. Principles of multilevel modelling. *Int J Epidemiol*. 2000; 29(1): 158-167.
- 181 Albert CM, Cook NR, Gaziano JM, Zaharris E,

- MacFadyen J, Danielson E, Buring JE, Manson JE. Effect of folic acid and B vitamins on risk of cardiovascular events and total mortality among women at high risk for cardiovascular disease: a randomized trial. *JAMA*. 2008; 299(17): 2027-2036.
- 182 Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ*. 1996; 313(7061): 862.
- 183 Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000; 355(9209): 1064-1069.
- 184 Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *BMJ*. 1996; 313(7060): 808.
- 185 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992; 116(1): 78-84.
- 186 Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *Am Heart J*. 2000; 139(5): 745-751.
- 187 Altman DG. Adjustment for covariate imbalance. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics*. John Wiley. 1998: 1000-1005.
- 188 Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med*. 2002; 136(2): 122-126.
- 189 Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med*. 1993; 118(3): 201-210.
- 190 Bender R, Grouven U. Logistic regression models used in medical research are poorly presented. *BMJ*. 1996; 313(7057): 628.
- 191 Khan KS, Chien PF, Dwarakanath LS. Logistic regression models in obstetrics and gynecology literature. *Obstet Gynecol*. 1999; 93(6): 1014-1020.
- 192 Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med*. 1979; 301(26): 1410-1412.
- 193 May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation*. 1981; 64(4): 669-673.
- 194 Altman DG, Cuzick J, Peto J. More on zidovudine in asymptomatic HIV infection. *N Engl J Med*. 1994; 330(24): 1758-1759.
- 195 Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. *Consolidated Standards of Reporting Trials*. *JAMA*. 1998; 279(18): 1487-1489.
- 196 Grant AM, Wileman SM, Ramsay CR, Mowat NA, Krukowski ZH, Heading RC, Thursz MR, Campbell MK; REFLUX Trial Group. Minimal access surgery compared with medical management for chronic gastroesophageal reflux disease: UK collaborative randomised trial. *BMJ*. 2008; 337: a2664.
- 197 van Loon AJ, Mantingh A, Serlier EK, Kroon G, Mooyaart EL, Huisjes HJ. Randomised controlled trial of magnetic-resonance pelvimetry in breech presentation at term. *Lancet*. 1997; 350(9094): 1799-1804.
- 198 Brown MJ, Palmer CR, Castaigne A, de Leeuw PW, Mancia G, Rosenthal T, Ruilope LM. Morbidity and mortality in patients randomised to double-blind treatment with a long-acting calcium-channel blocker or diuretic in the International Nifedipine GITS study: Intervention as a Goal in Hypertension Treatment (INSIGHT). *Lancet*. 2000; 356(9227): 366-372.
- 199 LaCroix AZ, Ott SM, Ichikawa L, Scholes D, Barlow WE. Low-dose hydrochlorothiazide and preservation of bone mineral density in older adults. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2000; 133(7): 516-526.
- 200 Shuster JJ. Median follow-up in clinical trials. *J Clin Oncol*. 1991; 9(1): 191-192.
- 201 Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995; 72(2): 511-518.
- 202 Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med*. 2005; 2(11): e298.
- 203 Diggle L, Deeks J. Effect of needle length on incidence of local reactions to routine immunisation in infants aged 4 months: randomised controlled trial. *BMJ*. 2000; 321(7266): 931-933.
- 204 Pocock S, White I. Trials stopped early: too good to be true? *Lancet*. 1999; 353(9157): 943-944.
- 205 Hughes MD, Pocock SJ. Stopping rules and estimation problems in clinical trials. *Stat Med*. 1988; 7(12): 1231-1242.
- 206 Kiri A, Tonascia S, Meinert CL. Treatment effects monitoring committees and early stopping in large clinical trials. *Clin Trials*. 2004; 1(1): 40-47.
- 207 Psaty BM, Rennie D. Stopping medical research to save money: a broken pact with researchers and patients. *JAMA*. 2003; 289(16): 2128-2131.
- 208 Temple R. FDA perspective on trials with interim efficacy evaluations. *Stat Med*. 2006; 25(19): 3245-3249.
- 209 Telmisartan Randomised Assessment Study in ACE

- intolerant subjects with cardiovascular Disease (TRANSCEND) Investigators, Yusuf S, Teo K, Anderson C, Pogue J, Dyal L, Copland I, Schumacher H, Dagenais G, Sleight P. Effects of the angiotensin-receptor blocker telmisartan on cardiovascular events in high-risk patients intolerant to angiotensin-converting enzyme inhibitors: a randomised controlled trial. *Lancet*. 2008; 372(9644): 1174-1183.
- 210 Senn S. Base logic: tests of baseline balance in randomized clinical trials. *Clin Res Regul Aff*. 1995; 12(3): 171-182.
- 211 Altman DG. Comparability of randomised groups. *Statistician*. 1985; 34(1): 125-136.
- 212 Heit JA, Elliott CG, Trowbridge AA, Morrey BF, Gent M, Hirsh J. Ardeparin sodium for extended out-of-hospital prophylaxis against venous thromboembolism after total hip or knee replacement. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2000; 132(11): 853-861.
- 213 Haderslev KV, Tjellesen L, Sorensen HA, Staun M. Alendronate increases lumbar spine bone mineral density in patients with Crohn's disease. *Gastroenterology*. 2000; 119(3): 639-646.
- 214 Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD, Macdonald M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA*. 1970; 211(12): 1993-2003.
- 215 Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med*. 1991; 10(10): 1595-1605.
- 216 Lewis JA, Machin D. Intention to treat — who should use ITT? *Br J Cancer*. 1993; 68(4): 647-650.
- 217 Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000; 21(5): 526.
- 218 Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther*. 1995; 57(1): 6-15.
- 219 Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med*. 2000; 19(14): 1849-1864.
- 220 Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine — selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003; 326(7400): 1171-1173.
- 221 Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials*. 2007; 4(4): 350-356.
- 222 Kruse RL, Alper BS, Reust C, Stevermer JJ, Shannon S, Williams RH. Intention-to-treat analysis: who is in? Who is out? *J Fam Pract*. 2002; 51(11): 969-971.
- 223 Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. *J Bone Joint Surg Am*. 2009; 91(9): 2137-2143.
- 224 Ruiz-Canela M, Martínez-González MA, de Irala-Estévez J. Intention to treat analysis is related to methodological quality. *BMJ*. 2000; 320(7240): 1007-1008.
- 225 Altman DG. Missing outcomes in randomized trials: addressing the dilemma. *Open Med*. 2009; 3(2): e51-53.
- 226 Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials*. 2004; 1(4): 368-376.
- 227 Streiner DL. Missing data and the trouble with LOCF. *Evid Based Ment Health*. 2008; 11(1): 3-5.
- 228 Molnar FJ, Hutton B, Fergusson D. Does analysis using “last observation carried forward” introduce bias in dementia research? *CMAJ*. 2008; 179(8): 751-753.
- 229 Ware JH. Interpreting incomplete data in studies of diet and weight loss. *N Engl J Med*. 2003; 348(21): 2136-2137.
- 230 Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry*. 2002; 47(1): 68-75.
- 231 Lane P. Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. *Pharm Stat*. 2008; 7(2): 93-106.
- 232 Abraha I, Montedori A, Romagnoli C. Modified intention to treat: frequency, definition and implication for clinical trials [abstract]. Sao Paulo, Brazil: XV Cochrane Colloquium. 2007: 86-87.
- 233 Altman DG. Confidence intervals in practice. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with confidence*. 2nd ed. BMJ Books. 2000: 6-14.
- 234 van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, Koes BW, Bierma-Zeinstra SM. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. *BMJ*. 2009; 339: b4074.
- 235 Altman DG. Clinical trials and meta-analyses. In: Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with confidence*. 2nd ed. BMJ Books. 2000: 120-138.

- 236 Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *Ann Intern Med.* 1997; 126(1): 36-47.
- 237 Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ.* 1986; 292(6522): 746-750.
- 238 Bailar JC 3rd, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Ann Intern Med.* 1988; 108(2): 266-273.
- 239 Bland JM. Quoting intermediate analyses can only mislead. *BMJ.* 1997; 314(7098): 1907-1908.
- 240 Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ.* 1995; 310(6977): 452-454.
- 241 Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ.* 1999; 319(7223): 1492-1495.
- 242 The OSIRIS Collaborative Group. Early versus delayed neonatal administration of a synthetic surfactant — the judgment of OSIRIS (open study of infants at high risk of or with respiratory insufficiency — the role of surfactant). *Lancet.* 1992; 340(8832): 1363-1369.
- 243 Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexøe J, Nielsen JB. Laypersons' understanding of relative risk reductions: randomised cross-sectional study. *BMC Med Inform Decis Mak.* 2008; 8: 31.
- 244 Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet.* 1994; 343(8907): 1209-1211.
- 245 Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med.* 1992; 117(11): 916-921.
- 246 Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science.* 1977; 198(4318): 679-684.
- 247 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA.* 1991; 266(1): 93-98.
- 248 Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med.* 2000; 19(24): 3325-3336.
- 249 Bhandari M, Devereaux PJ, Li P, Mah D, Lim K, Sch nemann HJ, Tornetta P 3rd. Misuse of baseline comparison tests and subgroup analyses in surgical trials. *Clin Orthop Relat Res.* 2006; 447: 247-251.
- 250 Levin M, Quint PA, Goldstein B, Barton P, Bradley JS, Shemie SD, Yeh T, Kim SS, Cafaro DP, Scannon PJ, Giroir BP. Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: a randomised trial. rBPI21 Meningococcal Sepsis Study Group. *Lancet.* 2000; 356(9234): 961-967.
- 251 Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J Clin Oncol.* 2006; 24(24): 3933-3938.
- 252 Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med.* 2009; 169(19): 1756-1761.
- 253 Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, Knudtson M, Dada M, Casperson P, Harris CL, Chaitman BR, Shaw L, Gosselin G, Nawaz S, Title LM, Gau G, Blaustein AS, Booth DC, Bates ER, Spertus JA, Berman DS, Mancini GB, Weintraub WS; COURAGE Trial Research Group. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med.* 2007; 356(15): 1503-1516.
- 254 Horton R. The rhetoric of research. *BMJ.* 1995; 310(6985): 985-987.
- 255 *Annals of Internal Medicine.* Information for authors. Available at www.annals.org (accessed 15 Jan 2008).
- 256 Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ.* 1999; 318(7193): 1224-1225.
- 257 Purcell GP, Donovan SL, Davidoff F. Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *JAMA.* 1998; 280(3): 227-228.
- 258 Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. *J Clin Epidemiol.* 2007; 60(4): 324-329.
- 259 Kiviluoto T, Sirén J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet.* 1998; 351(9099): 321-325.
- 260 Hupperets MD, Verhagen EA, van Mechelen W. Effect of unsupervised home based proprioceptive training on recurrences of ankle sprain: randomised controlled trial. *BMJ.* 2009; 339: b2684.
- 261 Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TR, DeBar LL, Lynch FL, D'Angelo E, Hollon SD, Shamseddeen W, Iyengar S. Prevention of depression in at-risk adolescents: a randomized controlled trial. *JAMA.* 2009; 301(21):

- 2215-2224.
- 262 CAMPBELL DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull.* 1957; 54(4): 297-312.
- 263 Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials.* 2006; 1(1): e9.
- 264 King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. *Health Technol Assess.* 2005; 9(35): 1-186, iii-iv.
- 265 Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, Kuderer NM, Lyman GH. The uncertainty principle and industry-sponsored research. *Lancet.* 2000; 356(9230): 635-638.
- 266 Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA.* 1998; 279(7): 545-549.
- 267 Smith GD, Egger M. Who benefits from medical interventions? *BMJ.* 1994; 308(6921): 72-74.
- 268 McAlister FA. Applying the results of systematic reviews at the bedside. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context.* BMJ Books. 2001.
- 269 Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, Dieppe P. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess.* 2005; 9(38): iii-iv, ix-x, 1-152.
- 270 Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ.* 2006; 333(7563): 346-349.
- 271 Bornhöft G, Masion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, Gilbertson S, Matthiessen PF. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. *BMC Med Res Methodol.* 2006; 6: 56.
- 272 Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med.* 1988; 318(26): 1728-1733.
- 273 Altman DG. Confidence intervals for the number needed to treat. *BMJ.* 1998; 317(7168): 1309-1312.
- 274 Fanaroff AA, Korones SB, Wright LL, Wright EC, Poland RL, Bauer CB, Tyson JE, Philips JB 3rd, Edwards W, Lucey JF, Catz CS, Shankaran S, Oh W; National Institute of Child Health and Human Development Neonatal Research Network. A controlled trial of intravenous immune globulin to reduce nosocomial infections in very-low-birth-weight infants. *N Engl J Med.* 1994; 330(16): 1107-1113.
- 275 Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. First International Study of Infarct Survival Collaborative Group. *Lancet.* 1986; 2(8498): 57-66.
- 276 Gøtzsche PC, Gjørup I, Bonnén H, Brahe NE, Becker U, Burcharth F. Somatostatin v placebo in bleeding oesophageal varices: randomised trial and meta-analysis. *BMJ.* 1995; 310(6993): 1495-1498.
- 277 Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. *J R Soc Med.* 2007; 100(4): 187-190.
- 278 Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med.* 1999; 130(12): 995-1004.
- 279 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ.* 1987; 295(6599): 654-656.
- 280 Papp K, Bissonnette R, Rosoph L, Wasel N, Lynde CW, Searles G, Shear NH, Huiizinga RB, Maksymowych WP. Efficacy of ISA247 in plaque psoriasis: a randomised, multicentre, double-blind, placebo-controlled phase III study. *Lancet.* 2008; 371(9621): 1337-1342.
- 281 Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev.* 1997; 9(1 Suppl): 15-21.
- 282 Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess.* 2000; 4(10): 1-115.
- 283 Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med.* 2005; 24(10): 1547-1561.
- 284 Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ.* 1997; 315(7109): 635-640.
- 285 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol.* 1986; 4(10): 1529-1541.
- 286 Chalmers I. From optimism to disillusion about commitment to transparency in the medico-industrial complex. *J R Soc Med.* 2006; 99(7): 337-341.
- 287 Tonks A. A clinical trials register for Europe. *BMJ.* 2002; 325(7376): 1314-1315.
- 288 Dickersin K, Rennie D. Registering clinical trials. *JAMA.* 2003; 290(4): 516-523.
- 289 Whittington CJ, Kendall T, Fonagy P, Cottrell D,

- Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet*. 2004; 363(9418): 1341-1345.
- 290 De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJ, Schroeder TV, Sox HC, Van Der Weyden MB; International Committee of Medical Journal Editors. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Lancet*. 2005; 365 (9474): 1827-1829.
- 291 Zarin DA, Ide NC, Tse T, Harlan WR, West JC, Lindberg DA. Issues in the registration of clinical trials. *JAMA*. 2007; 297(19): 2112-2120.
- 292 Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials*. 2008; 9: 20.
- 293 Russell JA, Walley KR, Singer J, Gordon AC, Hébert PC, Cooper DJ, Holmes CL, Mehta S, Granton JT, Storms MM, Cook DJ, Presneill JJ, Ayers D; VASST Investigators. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med*. 2008; 358(9): 877-887.
- 294 Chan AW, Tetzlaff J, Altman D, Gøtzsche PC, Hróbjartsson A, Krleža-Jeric K, et al. The SPIRIT initiative: defining standard protocol items for randomised trials. Oral presentation at the 16th Cochrane Colloquium: Evidence in the era of globalisation; 2008 Oct 3-7; Freiburg, Germany [abstract]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 2008; 102: 27.
- 295 Gregson S, Adamson S, Papaya S, Mundondo J, Nyamukapa CA, Mason PR, Garnett GP, Chandiwana SK, Foster G, Anderson RM. Impact and process evaluation of integrated community and clinic-based HIV-1 control: a cluster-randomised trial in eastern Zimbabwe. *PLoS Med*. 2007; 4(3): e102.
- 296 Shaddy RE, Boucek MM, Hsu DT, Boucek RJ, Canter CE, Mahony L, Ross RD, Pahl E, Blume ED, Dodd DA, Rosenthal DN, Burr J, LaSalle B, Holubkov R, Lukas MA, Tani LY; Pediatric Carvedilol Study Group. Carvedilol for children and adolescents with heart failure: a randomized controlled trial. *JAMA*. 2007; 298(10): 1171-1179.
- 297 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003; 326 (7400): 1167-1170.
- 298 Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the *BMJ*. *BMJ*. 2002; 325(7358): 249.
- 299 Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Med*. 2007; 4 (6): e184.
- 300 Sismondo S. Pharmaceutical company funding and its consequences: a qualitative systematic review. *Contemp Clin Trials*. 2008; 29(2): 109-113.
- 301 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. 2003; 290(7): 921-928.
- 302 Ross JS, Hill KP, Egilman DS, Krumholz HM. Guest authorship and ghostwriting in publications related to rofecoxib: a case study of industry documents from rofecoxib litigation. *JAMA*. 2008; 299 (15): 1800-1812.
- 303 McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA*. 2003; 289(19): 2545-2553.
- 304 Senn S. *Crossover trials in clinical research*. 2nd ed. Wiley. 2002.
- 305 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003; 7(27): iii-x, 1-173.
- 306 Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007. MR000012.
- 307 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet*. 2001; 357(9253): 373-380.
- 308 Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet*. 1996; 348(9027): 596-598.
- 309 Murray GD. Promoting good research practice. *Stat Methods Med Res*. 2000; 9(1): 17-24.
- 310 Narahari SR, Ryan TJ, Aggithaya MG, Bose KS, Prasanna KS. Evidence-based approaches for the Ayurvedic traditional herbal formulations: toward an Ayurvedic CONSORT model. *J Altern Complement Med*. 2008; 14(6): 769-776.
- 311 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW. Towards complete and accurate

reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med.* 2003; 138(1): 40-44.

312 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007; 147(8): 573-577.

313 Tonino PA, De Bruyne B, Pijls NH, Siebert U, Ikeno F, van't Veer M, Klauss V, Manoharan G, Engström

T, Oldroyd KG, Ver Lee PN, MacCarthy PA, Fearon WF; FAME Study Investigators. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N Engl J Med.* 2009; 360(3): 213-224.

(周庆辉翻译,卞兆祥、刘建平审校。本文系根据发表于 *BMJ* 2010; 340: c869 的全文翻译而成。本文的翻译符合 CONSORT 的规定,感谢 David Moher 为此提供的帮助。)