

·循证医学与中医·

临床试验中的结局指标及效应测量

张宏伟 刘建平

(北京中医药大学基础医学院循证医学中心, 北京市朝阳区北三环东路11号, 100029)

[摘要] 临床干预措施对人体施加的影响是多层次、多方位的, 临床试验中结局指标的选取直接决定对干预措施的认识。临床试验中的结局指标主要可分为3类, 即有效性、安全性和卫生经济学评价指标。对试验结局效应的测量主要是通过比较治疗组和对照组的相对和绝对效应来表达的。对临床试验中常用的结局指标类型, 以及不同的效应测量方法和临床意义进行了分析和介绍。

[关键词] 临床试验; 结局指标; 效应测量

临床试验中结局指标的选取直接决定了我们对研究目标措施的认识, 仅凭借某单一指标或某方面的指标, 无法全面了解干预措施对人体的作用以及影响, 如果用于指导临床决策, 就有可能导致无用甚至有害、或是无害却耗费医疗资源较多的干预措施在临床使用。正确的临床决策必须基于对目标措施正确而全面的认识。因此, 了解临床试验结局指标的主要类型可以帮助我们更全面地认识临床干预措施, 根据试验目的和研究对象选取适当的结局指标。对试验结局效应的测量主要通过比较治疗组和对照组的相对和绝对效应来表达, 有效的效应表达方式可以更明确地传达干预措施效应的大小, 更有利于指导临床决策。

1 临床结局指标的主要类型

临床干预措施对人体施加的影响是多层次、多方位的, 涉及人体生理、心理、社会功能以及周围环境等多个方面。一般情况下, 与干预措施最直接相关的评价就是有效性和安全性。一个具有临床应用价值的干预措施首先必须是有效的, 同时也是安全的。几乎所有医疗干预措施都会对人体存在或多或少的不良影响, 因此评价某项干预措施, 不但要了解其作用是否有效, 还要知道其不良作用, 才能综合权衡其益害比, 只有益处大于害处的干预措施才有应用的价值^[1]。另一方面, 在现有医疗卫生资源条件下, 如何有效而公平地使用有限的卫生资源也是需要重视的问题。因此, 对临床干预措施的评估主要着重于以下3个方面, 即有效性、安全性和卫生经济学评价。据此, 临床试验中的结局指标主要可分为这3种类型。

1.1 有效性指标

1.1.1 终点指标 终点指标是临床试验主要评估指

标的最佳选择。它一般是指对患者影响最大、患者最为关心的、与患者的切身利益最为相关的事件, 主要包括对患者生存或死亡、残障水平或其他一些重要临床事件, 如疾病复发等的测量^[2]。终点结局指标由于与患者最为相关, 因此对临床决策最具参考价值。终点结局指标往往可以用率来表示, 例如病死率、治愈率、缓解率、复发率、副反应率、生存率等。通常需要进行长期随访的试验研究来测量这些指标。

1.1.2 替代指标 在终点指标的测量不可行(如需要很长时间)的情况下, 就需要采用替代指标来评估干预措施的效果。替代指标一般易于测量, 如常用的单纯生物学指标, 包括实验室理化检测和体征发现, 如血脂、血糖、血压、血清胆固醇含量、实体肿瘤体积的缩小等。采用替代指标必须有足够证据支持其与临床终点结局的关系, 并可预测疾病结局。其应用的前提是替代指标的改善也将会相应改善疾病的终点结局。替代指标选择不当有可能导致错误估计干预措施对临床最终结局的作用。例如, 为了减少心肌梗死患者发生心律失常而应用抗心律失常药物的结果反而是导致病死率增加^[3]。

1.1.3 症状与体征 患者的主观症状、主诉以及医生在诊疗过程中探查发现的体征, 例如患者疼痛的性质和程度、心率、肝肿大程度、脉象和舌象等, 都可以用来评估干预措施对人体造成的影响。中医辨证论治通常就是依据患者的综合症状和体征来评价干预的效果。

有关主观症状和功能方面的指标常被称为软指标, 临床可靠性差, 重复性难于保证, 不易量化, 可以通过经科学设计和检验的问卷与量表进行调查。此外, 对一些认为是连续变化的特征或症状, 如疼痛的程度, 可以采用视觉模拟标尺(visual analogue scale, VAS)来进行测量^[4]。

1.1.4 生存质量 患者报告的结局指标(patient reported outcomes, PRO)是临床评价的重要内容,在对临床疗效的整体评估中日益受到重视,它包括患者描述的功能状况、症状以及与健康相关的生存质量(HRQOL)^[5]。生存质量一般通过专门设计并经科学检验的量表来进行评估,生存质量量表通常分为适用于一般人群的普适性量表和针对特定疾病人群的特异性量表。目前国内采用的量表主要有引用国外现成的量表以及自行设计的量表。选择量表时,除了考虑量表信度、效度和可信度等方面的指标,还要对量表的内容有全面的了解,针对研究目的和研究对象来选择适合的量表。

1.1.5 患者的满意度 循证医学的临床决策除了要基于目前最可靠的研究证据,考虑可利用的资源外,还要兼顾患者的意愿和价值取向。毕竟,改善患者的生存质量,令患者感到满意是医学干预措施的重要目的之一。特别是在很多慢性病没有十分有效的治疗方法的情况下,通过治疗让患者感到满意和舒心,也是很有价值的。以上指标的设立多是从医生的立场或角度出发,作为医务工作者,还需要从患者的角度评估疗效,比如,患者对治疗过程的满意程度、对治疗的接受程度,也是通过经科学评估和测试的问卷和量表来调查。

1.2 安全性指标

干预措施在对病情发生影响的同时,有可能会对患者造成不适甚至伤害。临床上主要通过观察、记录或及时报告不良反应、不良事件(包括严重不良事件)和副作用,来对治疗的安全性进行评估。应向受试者告知潜在的风险,提醒受试者报告试验期间发生的所有不良反应,有助于及时地获得有关不良事件的信息。

对不良事件的记录应参照《药品不良反应监测管理办法(试行)》的要求,对每一个不良事件尽量予以确诊或记录症状、体征。在出现无任何临床症状或体征的实验室检查异常时,应首先排除实验室误差(通过复查或检查仪器和试剂等),以及其他合并用药,或正常活动等的影响,然后参照WHO的实验室检查的毒性分级标准进行记录。严重不良事件应与不良事件的严重程度区别开来,如剧烈而不需住院的头痛是程度较严重的不良事件,但通常将危及生命、导致住院或延长原住院时间、致残、致癌、致畸、致出生缺陷、对器官功能产生永久损伤等的事件定义为严重不良事件^[6]。

1.3 卫生经济学指标

常用的对临床干预措施进行卫生经济学评价的指标主要有以下3种:成本-效果分析(CEA)、成本-效用分析(CUA)和成本-效益分析(CBA)。在比较多种

不同干预措施时,如果临床结局相同,可选用CEA,其表示方法为每一效果单位所耗费的成本(成本效果比)。如果临床结局不同,成本效果之间不可比,因此UEA多用于比较同一疾病不同治疗方法的投入和产出比,来确定哪种方法更划算。比较不同疾病治疗方法的成本时,可以选用CUA,它将不同的临床结局转换成相同的效果单位,一般采用质量调整生命年(quality adjusted life years, QALYs)来测量效用。CUA可以用来确定投入哪些疾病可以获得最大的整体健康回报。CBA将投入与产出用可以直接比较的统一的货币单位来估算。因为把卫生保健效益赋予货币价值非常困难,常用的测定方法是人力资本法和支付意愿法。它可以比较投入医疗卫生服务和社会其他领域的回报,从而合理地分配有限的社会资源^[1,7]。

2 临床试验结局效应的测量

临床试验结局效应的测量方式主要根据结局指标的数据类型和临床意义来决定。常用的临床结局指标主要有两种数据类型,即二分类变量和连续性变量。二分类变量是指结局只有两种可能,非此即彼,它可以是对定性资料的分类,如治疗有效与无效、病情好转与恶化、检查结果阳性与阴性、副作用有与无等,也可以是对定量资料的划分,如对高血压进行降压治疗,以舒张压降至90mmHg以下为有效,反之则无效。可以用连续变化的数值来表示的临床结局,如血压、血清胆固醇含量、体重、体温等,被认为是连续性变量,比如舒张压下降至85或75mmHg,相对于二分类变量的简单划分,连续性变量来表示结局程度会更加准确。

2.1 二分类变量的效应指标

二分类变量的结局指标通常以其中一个作为临床事件,分别计算其在治疗组和对照组的发生率,如生存率、死亡率、复发率、症状改善发生率等。然后,可以通过比较治疗组和对照组的事件发生率的相对和绝对差异来表达结局效应^[8]。

2.1.1 相对效应指标 常用的相对效应指标有相对危险度(relative risk, RR)、比值比(odds ratio, OR)及其95%的可信区间(confidence interval, CI)。相对危险度是治疗组的结局事件发生率(experiment event rate, EER)与对照组结局事件发生率(control event rate, CER)的比值,即 $RR = EER / CER$ 。如附表所示, $EER = a / a + b$, $CER = c / c + d$ 。例如,银杏达莫注射液联合视网膜激光光凝术治疗糖尿病视网膜病变,治疗组视力提高占26.6%,对照组视力提高占11.3%^[9],相对危险度是235%,即26.6%/11.3%,意指联合治疗组的视力提高率是单独光凝术治疗组的235%倍。

比值比是治疗组结局事件发生几率(a/b) 与对照组结局事件发生几率(c/d) 的比($OR = ad/bc$)。比值比与相对危险度是对同一干预结局的不同表达方式, 在结局事件发生率较低时, 如低于10%, 二者的值会较为接近。

附表 二分类结局指标的四格表

组别	结局事件	无结局事件	合计
治疗组	a	b	a+b
对照组	c	d	c+d

2.1.2 绝对效应指标 常用的绝对效应指标为绝对危险度, 亦称为绝对危险度降低(absolute risk reduction, ARR), 以及需要治疗的病例数(number needed to treat, NNT)。ARR 是指对照组结局事件率与试验组结局事件率之差的绝对值, 即 $ARR = CER - EER$ 。NNT 是绝对危险度的倒数, 即 $NNT = 1/ARR$, 是指要使一个患者从治疗中得益所需要治疗的患者总数, 例如, 上面提到的银杏达莫注射液联合视网膜激光光凝术治疗糖尿病视网膜病变, 绝对危险度是15.3%, 即[11.3% - 26.6%], 是指与单纯视网膜激光光凝术比较, 联合治疗组的视力提高率增加了15.3%; NNT 为6.5, 即1/15.3%, 是指与单纯视网膜激光光凝术比较, 联合治疗约6.5个患者可使其中1个患者视力得到提高。

相同的绝对效应指标在不同的危险度水平时其临床意义可能会有很大的差异, 例如绝对危险度是2%时, 58%到60%的变化可能没有什么临床意义, 此时相对危险度约是97%, 但是从1%到3%则在自身比例上变化较大, 可能具有一定的临床意义, 此时相对危险度是33%。因此, 绝对效应指标和相对效应指标是从不同的角度对同一结局效应的不同表达方式, 在报告临床结局的时候, 应根据临床意义选择适合的效应测量方法, 予以说明。

2.2 连续变量的效应指标

连续变量效应指标是治疗组和对照组之间的均数差(mean difference, MD) 及其95%CI。均数差即试验组均数与对照组均数的差值, 表明了干预措施导致的结局指标的平均变化大小。

可信区间(CI) 是测量研究结果准确度(或不确定性) 的一个指标, 表示对某效应值估计的把握度, 通常用95%表示 CI 的范围。95%CI 表明有95%的把握说明该效应值差异是在此区间的范围内, 即据此下结论出现错误的可能性只有5%。若 RR 和 OR 的95%CI 的数值范围包含1, MD 的95%CI 数值范围包括0, 则表明

试验与对照干预措施比较的效果差异在统计学上不具有显著性意义($P > 0.05$); 而当 RR 和 OR 值的95%CI 范围大于或小于1, MD 的95%CI 范围不包括0, 则表明相对和绝对效果差异具有统计学上的显著性($P < 0.05$)。通常, 临床试验样本量越大, 效应值的可信区间范围越窄, 准确性越高。反之, 试验样本量越小, 可信区间范围越宽, 准确性越低。目前临床试验报告国际标准(CONSORT) 要求报告可信区间, 相对于 P 值, 可信区间可以提供更多的信息^[10], 因此, 建议在研究报告中同时报告可信区间和 P 值。

参考文献

[1]Muir Gray, 唐金陵.循证医学:循证医疗卫生决策.北京:北京大学医学出版社,2004:167~237.

[2]郭新峰,赖世隆,梁伟雄.中医药临床疗效评价中结局指标的选择与应用.广州中医药大学学报,2002,19(12):251~255.

[3]郭新峰,朱泉,赖世隆.替代指标和中间指标及其在中医药疗效评价研究中应用价值的思考.中国中西医结合杂志,2005,25(7):585~590.

[4]Bengtsson M, Ohlsson B, Ulander K. Development and psychometric testing of the Visual Analogue Scale for Irritable Bowel Syndrome (VAS-IBS). BMC Gastroenterol, 2007, 7:16.

[5]刘凤斌,方积乾,王建华.中医药临床疗效评价的探讨.中药新药与临床药理,2004,15(4):290~292.

[6]胡晓民,吴磊.临床试验中不良事件的监测和管理体会.广州医药,2005,36(5):66~68.

[7]王家良主编.临床流行病学—临床科研设计、衡量与评价.2版.上海:上海科学技术出版社,2001:345~347.

[8]刘建平.临床试验效应的表达—指标及其意义.中国中西医结合杂志,2004,24(8):677~679.

[9]魏光川,刘名县,温转,等.银杏达莫注射液联合视网膜激光光凝术治疗糖尿病视网膜病变近期疗效观察.中国综合临床,2005,21(9):833~835.

[10] 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.

(收稿日期:2007-07-06)

•本刊启事•

本刊已加入《万方数据数字化期刊群(www.periodicals.net.cn)》、《中国期刊全文数据库(www.cnki.net)》、《中文科技期刊数据库(www.cqvip.com)》、《中国科学引文数据库(www.sciencechina.ac.cn)》、《中国核心期刊(遴选)数据库》。凡被本刊录用的稿件即认定为作者已同意将该论文的复制权、发行权、信息网络传播权、翻译权、汇编权等权利在全世界范围内转让给本刊。稿件一经刊用,将一次性支付作者著作权使用报酬(即包括印刷版、光盘版和网络版各种使用方式的报酬)。作者如有特殊要求请在投稿时说明。