

骨伤科临床常见结局指标的“统计学意义”与“临床意义”

王晓涛^{1,2,3} 潘艳芳^{1,2,3} 崔学军^{1,2,3} 叶秀兰^{1,2,3} 徐保平^{1,2,3}
李振军^{1,2,3} 朱森^{1,2,3} 王拥军^{1,2,3} 施杞^{1,2,3△}

【关键词】 统计学意义;临床意义;最小临床显著差异(MCID);骨伤科;结局指标;量表

【中图分类号】 R68 【文献标志码】 B 【文章编号】1005-0205(2017)01-0069-04

在临床研究中,统计学的显著性差异广泛应用于推断药物或医疗器械等的治疗效果。然而,越来越多的学者意识到,统计学推断可能会对疗效的评估产生误导作用。首先,疗效的统计学差异不能表明该差异一定有临床意义。统计显著性只能推断出疗效的存在与否,而无法具体得知疗效的高低。并且,只要样本量或者样本的变异性足够大,都能表现出统计显著性差异。其次,治疗组与安慰剂组疗效对比的统计学显著性从本质上忽略了患者个体间的异质性。例如,在治疗疼痛的试验中,统计学推断出疼痛缓解程度具有统计学差异,而实际上很多患者疼痛程度并未减轻^[1,2]。

骨伤科的临床研究常需测量疼痛、功能状态等,会用到各种量表和问卷,涉及到治疗组与对照组疗效的比较,明确骨伤科临床研究中各种结局指标在评价疗效中的统计学意义与临床意义,对指导骨伤科的临床实践和研究都具有重要的意义。本文将首先介绍统计学意义、临床意义和最小临床差异的基本概念,然后综述骨伤科临床研究中的一些常见的结局指标的临床意义。

1 统计学意义、临床意义和最小临床差异

1.1 统计学意义

“统计学意义”是相对于统计推断而言的,而统计推断,即“假设检验(亦称为显著性检验)”,是从问题的对立面(H_0),利用反证法去检验要解决的问题(H_1)是

否成立。实际过程中,常取检验标准 $\alpha=0.05$,当 $P<\alpha$,则结论按所取的假设检验水准,拒绝 H_0 ,接受 H_1 ,差异具有统计学意义(统计学结论)。从统计学理论不难发现,只要样本量足够大,那么结局指标评分的任何细微变化都可以拒绝 H_0 ,使结果具有统计学差异,即可称之为“具有统计学意义”,尽管对患者而言,可能并不具有什么临床意义^[3]。所以,统计学的结论只是建立在假设检验的基础上,而确定临床结论,除了统计学结论的支持外,还要结合相应的临床推断。

1.2 临床意义

临床意义,所评估的是治疗方式对临床具有重要意义的改善作用,应来源于临床实践。目前,临床意义的评估方式通常是采用基于某些外部标准的患者报告,患者自我报告结局,诸如对治疗的满意度、自我感觉的改善程度等,是临床常用的方法之一^[1]。临床研究中结局指标的临床意义,即指某项研究的结果在临床诊断、治疗及预后等方面具有指导意义,能为临床治疗决策提供依据^[4]。

在临床研究中,只有明确具体治疗方案对病情改善所具有的临床意义,才能为最优化的临床治疗方案提供决策依据;在科研方面,明确结局指标变化的具体临床意义,才能为后续研究确定必须样本量提供充分的数据支持^[5-9]。

1.3 最小临床显著差异与测量误差

最小临床显著差异(Minimum Clinically Important Difference, MCID)是治疗后改善的直观感觉的变化阈值,籍以评估是否达到临床疗效,旨在克服统计学意义的缺陷。若患者直观感觉的改善程度达到或超过 MCID,则表明具有临床意义。同时,也证明临床医生的治疗决策是正确有效的,进一步为临床的操作实践提供依据。对病人自身而言,则认为该治疗方法有意义、有价值,值得考虑继续该治疗手段,故 Jaeschke 等将其定义为“对病人有益的,最小的分值差异”,相当于患者报告结局中,自我感觉较治疗前状态“稍好一些”的改善程度^[4,5,10-13]。

基金项目:国家自然科学基金(81330085,81373666)

国家科技部,国家中医药管理局中医药行业科研专项(201407001-2)

国家中医临床研究基地业务建设科研专项(国家中医药管理局)基金资助项目(JDZX2012118)

上海市科委中医重点项目(14401970400)

施杞名中医工作室

¹ 上海中医药大学附属龙华医院(上海 200032)

² 上海中医药大学脊柱病研究所

³ 石筱山伤科学术研究中心

△通信作者 E-mail:13917715524@139.com

MCID,是一个以患者为中心的概念,同时获取了治疗后的改善程度以及患者对改观自我的估值,更直观地定义了治疗后评分变化的阈值,能为临床医生、卫生政策修订者等医务工作者提供关于疗效方面的客观参照。因此,最小临床显著差异的概念,被众多临床工作者广泛采纳^[1,14-16]。

然而如何将 MCID 与该量表的测量误差区别开,也使最小临床差异陷入争议之中。若测量工具精确度不够,实际测量过程中,会出现测得的数值变化在测量误差范围内的现象,而这并非真正的临床差异。例如,某测量工具的误差范围在 5~10 个单位,而该评估项目的 MCID 仅为 5 个单位,则该误差直接导致了量表实用性降低。即若误差范围超过了 MCID,那么该量表的实用性就值得进一步商榷。基于研究结果的统计学特征,量表的不精确性表述成标准误(SEM),作为阐述结局量表反应性的一种有效方法,已应用于一些研究中^[17]。SEM 的优势在于,它把量表的重测信度也纳入了考虑范围;同时,标准误使用实际的分数作为单位,易于同真正的分值改变作比较,如此,MCID 与标准误就可以从测量中分离出来。

2 骨伤科常用结局指标的临床意义

临床研究中,结局指标的测量值在试验前后确有变化,但其升高或者降低的幅度不大。经统计学分析后,变化具有统计显著性,但是这些改变是否也具有临床意义,对于临床治疗决策是否有指导意义,值得进一步推敲。下文就该问题选取骨伤科的常用结局指标进行举例说明。

2.1 关节活动度

例如,某研究评估不同肌肉的训练方式对膝关节骨折术后活动范围的影响,治疗组与对照组第一阶段均接受相同的理疗和肌肉训练方式,而后治疗组另增加关节终末端等长收缩训练,并保证两组患者每天的总锻炼时间相同。结果显示训练后两组患者膝关节的主动和被动屈、伸角度均有明显增加,与治疗前组内均数比较差异显著。治疗后组间比较,治疗组主动屈膝角度($108.21^{\circ} \pm 9.53^{\circ}$)比对照组($100.71^{\circ} \pm 9.58^{\circ}$)平均增大约 7.50° ,差异显著($P < 0.05$),具有统计学意义。确实,两种训练方式的结局指标,据其 P 值对比已具有统计学意义,但是,对于可以屈伸活动 160° (屈曲 145° ,过伸 15°)的膝关节^[18],这 7.50° 的改善,是否具有临床意义,那就应该根据临床进行判断。在统计学上 P 值只能说明在这两种不同的训练中关节活动度改善的区别在 7.5° 以下的概率小于 0.05,而不能具体判定这 7.5° 的区别对于临床骨折术后的康复是否具有临床意义。在日常生活中膝关节的活动所需范围如下:行走, $5^{\circ} \sim 67^{\circ}$ ±;上楼, 81° ±;下楼, 83° ±;坐位, 93° ±^[19]。那么屈曲角度恢复到 93° 以上,即基本满足日常生活,治疗效果已在可接受范围内,也就是说对照

组已经基本满足术后康复的临床标准。研究者常仅把最小临床显著差异(MCID)看作“对病人而言最小的显著改变”^[20],而忽略了治疗方式的风险和花费的效益比,然而在 MCID 的原始概念中包括了这一方面^[12]。那么在已满足日常需求的前提下,是否有必要再另做一项单独的治疗,值得进一步商榷。

同时对于关节活动度,测量时本身还涉及测量误差(标准误)和最小可测变化值(Smallest Detectable Difference, SDD)的问题。如某个手法治疗颈椎病的研究中,治疗组与对照组相比在即刻与短期效果上有明显的统计学意义。在颈椎的各项活动度上,治疗结束 10 min 后,颈椎各方向的活动度都有 $4^{\circ} \sim 5^{\circ}$ 的改善。而颈椎病的角度测量仪的标准误与 SDD 据文献报道分别为:标准误为屈曲 4.1° ,背伸 3.0° ,左旋 2.9° ,右旋 3.3° ,左侧屈 3.9° ,右侧屈 2.5° ;最小可测变化值为屈曲 9.6° ,背伸 7.0° ,左旋 6.7° ,右旋 7.6° ,左侧屈 9.1° ,右侧屈 5.9° 。在一定置信度的前提下,结合工具测量误差所确定的最小可测变化值,最小临床显著差异应该至少不低于该值才具有测量价值^[21]。由此可见,在骨伤科临床研究中,试验过程涉及关节活动度等测量时,应注意测量工具的最小可测值、测量误差以及最小临床显著差异三者之间的区别。

2.2 疼痛量表

疼痛类结局指标测量方面也存在上述类似的争议。在某项关于手法和物理疗法治疗颈椎病的研究中,采用视觉模拟评分法(VAS 评分)观察即刻治疗效果时,两种方法对疼痛皆有明显改善作用。其中即刻 VAS 评分(10 分制),治疗组(3.5 ± 0.8)、对照组(4.5 ± 1.3),组间比较 $P < 0.05$,具有统计学差异。这 1.0 分的差别是否具有临床意义,国外有一些学者开展这方面研究,发现 VAS 评分改变只有达到 1.3 分以上时,病人主观感受的疼痛变化才具有最低程度的临床意义^[22]。也就是说,视觉模拟评分法的灵敏度是有一定局限性的。

有学者更是发现,VAS 评分表上,不同初始疼痛程度的受试者,达到具有临床意义的缓解,其所要求的缓解程度是不同的。美国一所医学院曾做过一项专门的临床研究,试验过程中,对受试者创伤处引起的疼痛进行起始的 VAS 评分,随后每隔 30min 评估其疼痛程度并予以记录,直至受试者疼痛消失,或者从记录开始,2 h 后作为时间截点。结果发现起始分低于 34 分的受试者,需改变(13 ± 14)分(mean±SD)才有临床意义;而起始分值在 67 分以上的受试者,疼痛评分平均改变(28 ± 21)分才有显著临床差异^[23]。

同样,疼痛数字评分量表也并非等距量表,而且其分值的变化也存在着最小临床差异,在一项针对肌肉骨骼系统疼痛的 MCID 研究中,发现临床上骨骼肌疼痛的患者,NRS 量表分值减少超过 15 分以上,甚或一

个疼痛点的消失,才具有最小临床显著差异^[24]。

所以对疼痛的评估,不能单靠视觉模拟评分法或者疼痛数字评分法来测定。在 2007 年的临床试验方法、测定及疼痛评估组织 (IMMPACT) 共识会议上曾提出,对于一个慢性疼痛随机对照试验结局指标的测定,最好采用两种或者两种以上的不同方法来评估具有临床意义的疼痛加剧或缓解,这样才能改善镇痛试验方法的测定灵敏度^[25]。

2.3 功能量表

功能量表方面,具有临床意义的分值改变,其实际意义到底如何理解,一直是临床医生的困惑。因为一个涉及多项目结局指标测量的功能量表,每个项目分值的改变并不是同等效力的。不同项目相对值的变化又对应不同方面的健康状况的改变;即便是相同的总分值,一个项目上 3 个单位的改变可能比 3 个项目上 1 个单位的变化显得更具有临床意义。

例如,某患者在行走与拿持重物能力方面发生的巨大改变和另一名患者在睡眠质量与性功能方面的变化分值相等,但个体的偏好与预期可能完全不同。Gross 等^[26]在研究加拿大工伤康复中的患者的 SF-36 量表参考值时发现,索赔人群 SF-36 表上都表现出显著的局限性,但主要集中在情绪量表和身痛评分方面。

与其他量表相似,功能量表也存在一定的误差。以被认为等距量表的颈残障指数 (NDI) 举例,作为公认的一维量表,也不乏有研究对其质疑。某个专门对 NDI 的系统评价得出,一般 NDI 的最小可测值在 5/50 左右,而对颈部反射性疼痛的最小可测值则达到 10/50,对于不同的症状,临床差异的变化值也从 5/50 到 19/50 不等。故而在使用 NDI 评分时,不同的疾病可能会面临 MCID 的不同^[27]。其次,由于量表都是跨文化翻译,难免在不同语言文化中存在信度与效度差异的问题。有研究指出,同一版本的 NDI 量表,阿拉伯语、意大利语和泰语版的译本质量明显高于其他语种^[28]。

专门针对颈椎前路减压术后评估的颈椎结局问卷 (Cervical Spine Outcomes Questionnaire, CSOQ) 调查显示,颈椎减压内固定术后,最小的具有临床差异为 13 分 (针对功能残疾评分) 到 24 分 (针对肩臂部疼痛) 不等,而实质上的具有临床意义的改善程度,其分值改变须 20 到 30 分^[29]。同时,功能量表又往往会掩盖患者实质性的改观。比如,一组慢性腰痛的患者或许在 Oswestry 功能障碍指数问卷表 (ODI 量表) 上有平均 15~20 个单位的显著改善,但仍可能由于严重的疼痛而丧失行动能力,无法工作^[30]。类似的结局指标还有步伐速度^[31]、JOA^[32]、SF-36^[33]、欧洲五维健康量表 (EQ-5D)^[34] 等等。

结局指标,尤其是患者自我报告结局的量表等结局指标,已经成为对疗效的标准化评估方法,因此对于

其临床意义的研究与理解显得尤为重要。

3 结语

国外学者在临床意义方面已有比较深入的认识和研究,国内学者亦有阐发过临床意义与统计学意义方面的思考^[35,36],但目前国内关于一个领域的结局指标的研究并不多见,故本文就骨伤科领域的结局指标的临床意义作一综述性研究。

在科研、临床实践中,只有了解相关结局指标的数值变化及其对应的临床环境,明确区分结局指标的统计学意义与临床意义,将统计学结论与专业有机地结合起来,才能得出相对客观、符合实际的临床结论^[37]。

总而言之,在骨伤科临床实践及研究中,只有真正明确“临床意义”,全面考虑、深入理解结局指标的灵敏度、测量误差及最小临床显著差异等因素,才能更好地指导临床治疗最优方案的选定以及科研方案更优化的设计,进而更好地指导科学研究及临床实践。

当然,尽管 MCID 是目前公认的,判断试验是否具备临床意义的阈值,但因为各试验研究的方法学、对试验假设的分析、使用的量表以及试验观察周期等的不同,产生的各结局指标的 MCID 可能不尽相同;同时,在研究中,对患者的评估,因于观察员的个体偏倚,往往会不够精确^[24]。因此,对于疼痛、功能等量表,在临床背景下,确定籍以评估临床显著差异是否具有临床意义的标准值,对于提升研究质量具有重要意义,值得进一步深入研究。

参考文献

- [1] Hedayat AS, Wang J, Xu T. Minimum clinically important difference in medical studies[J]. Biometrics, 2015, 71(1): 33-41.
- [2] Younger J, McCue R, Mackey S. Pain outcomes: a brief review of instruments and techniques[J]. Current pain and headache reports, 2009, 13(1): 39-43.
- [3] Fortin PR, Stucki G, Katz JN. Measuring relevant change: an emerging challenge in rheumatologic clinical trials[J]. Arthritis Rheumatism, 1995, 38(8): 1027-1030.
- [4] Copay AG, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales[J]. The Spine Journal, 2008, 8(6): 968-974.
- [5] Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation[J]. Controlled clinical trials, 1991, 12(4): S142-S158.
- [6] Beaton DE. Understanding the relevance of measured change through studies of responsiveness[J]. Spine, 2000, 25(24): 3192-3199.
- [7] Bombardier C, Kerr MS, Shannon HS, et al. A guide to interpreting epidemiologic studies on the etiology of back pain[J]. Spine, 1994, 19(18): 2047S-2056S.
- [8] Lydick E, Epstein RS. Interpretation of quality of life changes

- [J]. Quality of life Research, 1993, 2(3): 221-226.
- [9] Shekelle PG, Andersson G, Cherkin D, et al. A brief introduction to the critical reading of the clinical literature[J]. Spine, 1994, 19(18): 2028S-2031S.
- [10] Huskisson EC. Measurement of pain[J]. Lancet, 1974, 304(7889): 1127-1131.
- [11] Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods[J]. The Spine Journal, 2007, 7(5): 541-546.
- [12] Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference[J]. Controlled clinical trials, 1989, 10(4): 407-415.
- [13] Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments [J]. Journal of chronic diseases, 1987, 40(2): 171-178.
- [14] Food and Drug Administration, Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims [J]. Fed Regist, 2009, 74(235): 65132-65133.
- [15] Patrick DL, Burke LB, Powers JH, et al. Patient - Reported Outcomes to Support Medical Product Labeling Claims: FDA Perspective[J]. Value in Health, 2007, 10(s2): S125-S137.
- [16] McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients[J]. JAMA, 2014, 312(13): 1342-1343.
- [17] Hägg O, Fritzell P, Nordwall A. The clinical importance of changes in outcome scores after treatment for chronic low back pain[J]. European Spine Journal, 2003, 12(1): 12-20.
- [18] 王和鸣. 中医骨伤科学[M]. 北京: 中国中医药出版社, 2013: 30
- [19] 王亦聰. 骨与关节损伤[M]. 北京: 人民卫生出版社, 2011: 1302
- [20] Stratford PW, Binkley JM, Riddle DL, et al. Sensitivity to change of the Roland-Morris back pain questionnaire: part 1[J]. Phys Ther, 1998, 78(11): 1186-1196.
- [21] Reid SA, Callister R, Katekar MG, et al. Effects of Cervical Spine Manual Therapy on Range of Motion, Head Repositioning, and Balance in Participants With Cervicogenic Dizziness: A Randomized Controlled Trial[J]. Arch Phys Med Rehabil, 2014, 95(9): 1603-1612.
- [22] Todd KH, Funk KG, Funk JP, et al. Clinical significance of reported changes in pain severity[J]. Annals of emergency medicine, 1996, 27(4): 485-489.
- [23] Bird SB, Dickson EW. Clinically significant changes in pain along the visual analog scale[J]. Annals of emergency medicine, 2001, 38(6): 639-643.
- [24] Salaffi F, Stancati A, Silvestri CA, et al. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale[J]. European Journal of Pain, 2004, 8(4): 283-291.
- [25] Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations[J]. The Journal of Pain, 2008, 9(2): 105-121.
- [26] Gross DP, Algarni FS, Niemeläinen R. Reference Values for the SF-36 in Canadian Injured Workers Undergoing Rehabilitation[J]. Journal of occupational rehabilitation, 2015, 25(1): 116-126.
- [27] Macdermid JC, Walton DM, Avery S, et al. Measurement properties of the neck disability index: a systematic review [J]. Journal of orthopaedic & sports physical therapy, 2009, 39(5): 400-C12.
- [28] Yao M, Sun YL, Cao ZY, et al. A Systematic Review of Cross-Cultural Adaptation of the Neck Disability Index [J]. Spine(Phila Pa 1976), 2015, 40(7): 480-490.
- [29] Skolasky RL, Albert TJ, Maggard AM, et al. Minimum Clinically Important Differences in the Cervical Spine Outcomes Questionnaire[J]. The Journal of Bone & Joint Surgery, 2011, 93(14): 1294-1300.
- [30] Boos N. Comment to "The clinical importance of changes in outcome scores after treatment for chronic low back pain" by O. Hägg et al[J]. European Spine Journal, 2003, 12(1): 21-21.
- [31] Bohannon RW, Glenney SS. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review[J]. Journal of evaluation in clinical practice, 2014, 20(4): 295-300.
- [32] Tetreault L, Nouri A, Kopjar B, et al. The Minimum Clinically Important Difference of the Modified Japanese Orthopaedic Association Scale in Patients with Degenerative Cervical Myelopathy[J]. Spine(Phila Pa 1976), 2015, 40(21): 1653-1659.
- [33] Spratt KF. Minimal Clinically Important Difference based on Clinical Judgment and Minimally Detectable Measurement Difference: A rationale for the SF-36 Physical Function scale in the SPORT Intervertebral Disc Herniation Cohort[J]. Spine(Phila Pa 1976), 2009, 34(16): 1722.
- [34] Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review [J]. Expert review of pharmacoeconomics & outcomes research, 2014, 14(2): 221-233.
- [35] 李光伟. 统计学意义与临床意义距离有多远? [J]. 中华内分泌代谢杂志, 2006, 22(1): 1-2.
- [36] 李河. 正确理解医学统计推断中的“有统计学意义”与“有实际专业意义”[J]. 循证医学, 2004, 4(1): 54-55.
- [37] Curran-Everett D, Taylor S, Kafadar K. Fundamental concepts in statistics: elucidation and illustration[J]. J Appl Physiol, 1998, 85(3): 775-786.