

· 指南解读 ·

骨关节炎临床研究中疗效评价指标的选择与方法学考虑 ——《膝骨关节炎中医诊疗指南(2020 年版)》解读

曹月龙

(上海中医药大学附属曙光医院, 上海 201203)

摘要 《膝骨关节炎中医诊疗指南(2020 版)》(以下简称《指南》)已正式发布,其中疗效评价指标的推荐是亮点之一。《指南》推荐了体现症状体征/中医证候、生活质量、关节结构改变和远期转归 4 类指标。本文结合《指南》推荐的膝骨关节炎疗效评价指标,从疗效评价指标的选择、常用疗效评价指标概述和疗效评价指标的方法学考虑 3 个方面,对骨关节炎临床研究中疗效评价指标的相关问题进行了探讨,以期临床和科研工作者开展骨关节炎临床研究时选择合适的疗效评价指标提供指导。

关键词 骨关节炎;膝;结果评价;卫生保健;方法;指南

骨关节炎(osteoarthritis, OA)是一种复杂的多因素关节疾病,以局部软骨退化、骨赘形成、软骨下骨改变为主要病理特征,以关节疼痛、肿胀、僵硬及功能障碍为主要临床表现^[1-2]。相关研究表明,OA 的患病率呈逐年递增趋势^[3-4]。OA 属中医“痹证”“痿证”范畴,是中医骨伤科治疗的优势病种之一。随着 OA 临床研究的日益增多,OA 临床研究疗效评价指标的选择成为骨科界关注的重要问题。2020 年 10 月,中国中医药研究促进会骨伤科分会发布了《膝骨关节炎中医诊疗指南(2020 年版)》^[5](以下简称《指南》)。《指南》用于指导和规范临床实践和相关研究,其中膝骨关节炎(knee osteoarthritis, KOA)疗效评价指标的推荐是亮点之一。本文结合《指南》推荐的 KOA 的疗效评价指标,从疗效评价指标的选择、常用疗效评价指标概述和疗效评价指标的方法学考虑 3 个方面,对 OA 临床研究中疗效评价指标的相关问题进行了探讨,以期临床和科研工作者开展骨关节炎临床研究时选择合适的疗效评价指标提供指导。

1 OA 临床研究中疗效评价指标的选择

OA 临床研究中的疗效评价指标包括主要结局指标和次要结局指标。主要结局指标又称主要终点指标,是与研究目的存在直接的、本质的联系且能确切反映干预措施有效性或安全性的观察指标。次要结局指标又称次要终点指标,是指与研究目的相关的辅

助性指标。一项临床研究通常只有一个结局指标,但可以有多个次要结局指标。主要结局指标用于回答与研究目的直接相关的问题,次要结局指标能够回答与研究目的间接相关的问题。在临床研究中,主要结局指标与次要结局指标均应在设计方案中明确定义,并对这些指标在解释研究结果时的作用及相对重要性加以说明。

目前,在临床研究中,研究人员在疗效指标的选择、测量和报告等方面普遍存在不一致、不规范、发表偏倚等问题,导致同类研究的结果不能合并与比较^[6]。《指南》推荐根据研究周期的长短,选择设置主要疗效指标、次要疗效指标、次要终点指标和/或主要终点指标。近期、中期疗效评价应当关注患者症状体征/中医证候、关节功能评分及生活质量的改善,远期疗效评价应当关注患者影像学改变及手术干预时间。临床研究中的疗效评价指标既要体现研究性质,又要体现干预措施的直接或间接结果^[7]。因此,除了根据研究周期外,还需要根据研究目的、干预措施的特点有针对性地选择疗效评价指标。

1.1 基于研究周期选择疗效评价指标 OA 是一种慢性疾病,在短周期临床研究中,应重点关注患者症状和体征的变化,如随访观察时间在 3 个月以内的研究,可以将疼痛视觉模拟量表(visual analogue scale, VAS)或西安大略和麦克马斯特大学(Western Ontario and McMaster Universities, WOMAC)骨关节炎指数^[8]作为主要结局指标;在长周期临床研究中,应重点关注关节的结构改变,如随访观察时间在 12 个月以上的研究,所选择的结局指标应能够体现关节功能和结

基金项目:国家自然科学基金(81373665, 81973874);上海市慢性筋骨病临床医学研究中心项目(20MC1920600)

通讯作者:曹月龙 E-mail:ningtem@126.com

构的改变。

1.2 基于研究目的选择疗效评价指标 临床研究一般可分为解释性临床研究和实用性临床研究。解释性临床研究的目的是确定干预措施是否产生了与之有关的结果或机制,该类研究通常在理想的试验条件下对严格符合受试条件的受试者进行干预,从而评价干预措施的疗效(一般以安慰剂作为对照);实用性临床研究通常用来研究一种治疗方法在日常临床实践中的疗效,该类研究多在“真实世界”的临床环境中进行。实用性临床研究以其独特的优势更适合中医药研究。对于实用性临床研究,主要结局指标通常选择具有临床意义的指标,但这类指标可能不是干预措施的直接结果。例如,评估某种膝关节锻炼对 KOA 的治疗效果,在解释性临床研究中应选择肌力或关节活动范围等与治疗方法直接相关的指标作为主要结局指标;但在实用性临床研究中,主要结局指标可能是 KOA 患者的功能表现,如 WOMAC 骨关节炎指数等。

1.3 基于干预措施选择疗效评价指标 辨证论治是中医特色,目前 OA 证型的分类尚存在分歧^[9]。针对特定证型的中药复方研究,《指南》中的“膝骨关节炎主要症状体征/中医证候分级量化评分表”是一大亮点。此表结合了症状和常见证型,可以说是“既对病也对证”的量表。在此表中,关节畏寒、关节发热、腰膝酸软、倦怠乏力则分别是寒湿痹阻证、湿热痹阻证、肝肾亏虚证和气血虚弱证的主症,采用此表有助于观察证候和证型的转归。而中医复方的治疗机制属于多靶点调控,对于患者的全身状态和生活质量均有改善作用,简明健康状况调查表(short form 36 health survey questionnaire, SF-36)是常用的结局指标。此外,当研究中采用药物等干预措施控制 OA 发作期的症状时,主要结局指标可不包括影像学评价指标;而采用截骨术等针对特定关节结构进行干预时,主要结局指标应选择影像学评价指标,如 Kellgren - Lawrence 影像分级^[10]、MRI 膝骨关节炎评分^[11]、全关节核磁影像评分^[12]等。目前,关于 OA 治疗药物研发的临床研究,一般倾向于采用同时能够评价关节结构和症状改善的结局指标。总之,疗效评价指标的选择必须以服务患者为中心,避免出现只关注指标而不注重患者实际需求的情况。

2 OA 临床研究中常用疗效评价指标概述

OA 临床研究的结局指标应具有实际意义,应能

够反映患者生活质量的改善。OA 疗效评价指标可分为基于测量的客观类指标和基于患者报告的主观类指标。客观类指标是对患者特定任务的完成情况的评估,如测量患者行走的距离或行走一段距离的时间、从坐位到站位的时间、上下一段楼梯的时间等;主观类指标是以患者的主观感受为中心,反映了患者对自身功能、疼痛或其他身体状态改善情况的感知。国际骨关节炎研究学会推荐在 OA 临床研究中采用基于患者报告的疗效评价指标^[13],如 WOMAC 骨关节炎指数、膝关节损伤和骨关节炎结局评分(knee injury and osteoarthritis outcome score, KOOS)^[14]、髌关节伤残和骨关节炎评分^[15]以及患者报告结局测量信息系统(patient - reported outcomes measurement information system, PROMIS)^[16]。

KOA 主要症状体征/中医证候分级量化评分表、利克特量表^[17]、VAS、WOMAC 骨关节炎指数、KOOS 及间歇性和持续性骨关节炎疼痛问卷表^[18]主要用于评价疼痛改善和功能恢复。KOA 主要症状体征/中医证候分级量化评分表侧重于评价 KOA 患者的关节疼痛、肿胀以及伴随症状的程度。关节炎影响测量量表^[19]、下肢功能量表^[20]也可用于评价患者功能恢复情况。SF-36、PROMIS、生活健康评估量表、日常生活活动量表^[21]等能够从躯体疼痛、关节功能活动、日常生活活动、生理职能、社会功能、情感职能、心理健康、远期生活质量等多方面评价患者的生活质量和健康状况。Kellgren - Lawrence 影像分级基于 X 线检查对膝关节的结构改变(关节间隙、骨赘、软骨骨化等)进行评价。该方法将 KOA 分为 0 到 4 级,0 级为正常,级别越高表明疾病越严重。MRI 是 OA 临床研究中常用的影像学检查方法,能够清晰地显示关节软骨、滑膜、半月板、肌肉、肌腱等组织的形态结构和病理变化,相较于 X 线和 CT 具有明显优势^[22-23]。常用 OA 临床疗效评价指标见表 1、表 2。

3 OA 临床研究中疗效评价指标的方法学考虑

3.1 样本量 对于 OA 的临床研究,样本量需控制在一定范围内,以避免干预措施导致的差异是偶然因素造成的。通常情况下,样本量应基于 I 型错误和 II 型错误的概率进行计算。样本量过小会导致研究的统计学权重不足而增加 II 型错误的概率,而样本量过大会导致资源浪费。因此,临床研究中样本量的确定应以主要结局指标的最小临床有意义改善(minimum

clinical important improvement, MCII) 作为参考, 并将次要结局指标的需求考虑进来。

3.2 MCII 测量的 OA 结局指标主要有 3 类: 症状体征变化、结构改变、生活质量改善。MCII 是指测量的结局指标发生具有临床意义的最小变化值。临床研究中应避免只从统计学角度解读数值结果的差异。

例如, 在采用 VAS 评价病情转归时, 患者 2 次测量结果的距离差值在 5 mm 时, 就可能在统计学上表现出差异, 但这种差异并不具有临床意义, 即不能达到 MCII 值。KOA 临床研究中公认的常用结局指标的 MCII 值见表 3^[24]。

表 1 常用主观类骨关节炎临床疗效评价指标

评价指标	评价方法	评价内容	条目数量	指标应用时间窗	完成时间	评分原则	获取方式
西安大略和麦克马斯特大学骨关节炎指数	问卷	疼痛、僵硬、功能活动	24 条	48 h	5 ~ 10 min	0 ~ 96 分, 评分越高疾病越严重	http://www.womac.com/
膝关节损伤和骨关节炎结局评分	问卷	疼痛、日常生活活动、日常运动、生活质量	42 条	1 周	< 10 min	0 ~ 100 分, 评分越高疾病越严重	http://www.koos.nu/
间歇性和持续性骨关节炎疼痛问卷表	问卷	间歇性疼痛、持续性疼痛	11 条	1 周	< 10 min	0 ~ 44 分, 评分越高疼痛越严重	http://www.oarsi.org/research/outcome-measures#questionnaire
简明健康状况调查表	问卷	生理、心理健康	32 条或 12 条	1 周或 4 周	5 ~ 10 min	0 ~ 100 分, 评分越高生活状态越好	http://www.sf-36.org/
患者报告结局测量信息系统	问卷	生理、心理、社会健康	不确定	不确定	不确定	采用工具不同分值不同, 分数越高问题越严重	http://www.nihpromis.org/
生活健康评估量表	问卷	残疾或受限、主观不适、药物不良反应、医疗保健使用、人口、生活方式、健康行为	不确定	1 周或 4 周或 6 周	20 ~ 30 min	0 ~ 100 分, 评分越高健康状况越差	http://aramis.stanford.edu/HAQ.html
改良生活健康评估量表	问卷	功能、疼痛、整体健康	22 条	1 周或不受限制	< 5 min	0 ~ 100 分, 评分越高健康状况越差	http://aramis.stanford.edu/HAQ.html
KOA 主要症状体征/中医证候分级量化评分表	问卷	关节疼痛、肿胀、畏寒、发热、腰膝酸软、倦怠乏力	24 条	即时	< 5 min	0 ~ 18 分, 评分越高症状越严重	《膝骨关节炎中医诊疗指南(2020 年版)》

表 2 常用客观类骨关节炎临床疗效评价指标

评价指标	评价方法	评价内容	拍摄体位	评分原则	获取方式
Kellgren - lawrence 影像分级	X 线	关节结构(间隙、骨赘、软骨骨化)	正位、侧位	0 ~ 4 级, 等级越高表明关节结构破坏越严重	《Radiological assessment of osteo - arthrosis》
全关节核磁共振影像评分	MRI	软骨、软骨下骨、关节周围软组织	冠状位、矢状位、横断位	评分越高表明疾病越严重	《Whole - Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis》

表 3 膝骨关节炎常用结局指标的最小临床有意义改善值

结局指标	最小临床有意义改善值			
	绝对变化 ³⁾	95% CI	相对变化 ⁴⁾ /%	95% CI
疼痛 ¹⁾	-19.9 mm	-21.6 ~ -17.9	-40.8	-44.8 ~ -36.1
患者病情总体评价 ¹⁾	-18.3 mm	-19.8 ~ -16.7	-39.0	-45.8 ~ -30.6
标准化西安大略和麦克马斯特大学骨关节炎指数 ²⁾	-9.1 分	-10.5 ~ -7.5	-26	-28.6 ~ -23.3

1) 均采用视觉模拟量表(0 ~ 100 mm)评价; 2) 标准化评分为 0 ~ 100 分; 3) 绝对变化 = 治疗后测量数值 - 治疗前测量数值; 4) 相对变化 = (治疗后测量数值 - 治疗前测量数值)/治疗前测量数值。

3.3 多重性分析 在临床研究中,主要结局指标和次要结局指标的选择在一定程度上会影响研究结果。一项临床研究可以包括一个或多个次要结局指标,但次要结局指标的测量不应干扰主要结局指标的测量。对于验证性临床研究,特别是Ⅲ期临床试验,只有主要结局指标有统计学意义,次要结局指标的统计分析结果才有参考价值;对于探索性临床研究,主要结局指标和次要结局指标的结果均可为进一步的临床试验设计提供线索。一般情况下,一项临床研究只采用一个主要结局指标。对于只有一个主要结局指标且不进行期中分析的临床研究,一般不涉及多重性分析问题;但当临床研究中采用了多个主要结局指标或进行期中分析或进行多组间比较时,基本都会涉及多重性分析问题。

多重性分析是较复杂的统计学问题,临床研究的多重性分析是指多次采用假设检验。在进行单个假设检验时,我们会先确定好显著性水平(α 通常取 0.05),代表单次检验中可能出现假阳性的概率为 5%。若同时检验多个假设,我们需要控制好 I 型错误的概率。多次假设检验会增加 I 型错误发生的风险,从而导致试验结果不准确^[25]。例如,4 组之间进行两两比较,如果每次检验的显著性水平均为 0.05,则 6 次假设检验至少出现 1 次假阳性的概率会高达 26.5%。因此,当需要采取合适的方法将总的 I 型错误控制在预定的检验水准之下时,就会牵涉多重检验校正问题。例如,在包含 5 个亚组分析或对比 5 个疗效评价指标的研究中,采用 Bonferroni 法将每项分析的检验水准调整为 0.01 (0.05/5),以保证总体 I 型错误的概率为 0.05。对于有 ≥ 2 个主要结局指标的临床研究,如果所有主要结局指标的差异有统计学意义才可以认为研究结果有效,则无需校正检验水准;而如果任意主要结局指标的差异有统计学意义即可认为研究结果有效,则需要校正检验水准。

此外,对于涉及同一疗效评价指标的多亚组之间或多时间点之间差异比较的验证性临床研究,建议在注册随机对照试验的原始资料中予以说明。在随机对照试验中进行亚组分析常需要较大的样本量,而通常没有足够的样本进行次要结局指标的检测及交互效应分析。这时就需要明确相关因素的分析是否适当和可行、是否具有合适的显著性水平,可参考临床试验报告统一标准进行判断^[26]。

4 小 结

《指南》在一定程度上和范围内规范了 KOA 中药临床研究的疗效评价体系,提出了适合不同分期及病证结合状态的评价方法,对于指导中医药治疗 KOA 的临床实践与研究具有重要意义。《指南》推荐根据研究周期选择合适的主要结局指标和次要结局指标。此外,对于 OA 临床研究中的疗效评价指标,还需要根据研究目的、干预措施的特点有针对性地进行选择;同时应该从方法学角度充分考虑样本量、MCII 及多重性分析等因素,从而能够更加规范、合理地开展相关临床研究。目前,国际上关于 OA 临床研究中疗效评价指标的研究较为细致,取得了一定的共识,但探索能够更好地体现中医药疗效特点且形成规范方法学依据的中医药治疗 OA 临床研究中的疗效评价指标仍任重道远。

参考文献

- [1] GLYN - JONES S, PALMER A J, AGRICOLA R, et al. Osteoarthritis[J]. Lancet, 2015, 386(9991): 376 - 387.
- [2] HUNTER D J, BIERMA - ZEINSTRAS S. Osteoarthritis[J]. Lancet, 2019, 393(10182): 1745 - 1759.
- [3] MANDL L A. Osteoarthritis year in review 2018: clinical[J]. Osteoarthritis Cartilage, 2019, 27(3): 359 - 364.
- [4] ZHANG Y, XU L, NEVITT M C, et al. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: the Beijing osteoarthritis study[J]. Arthritis Rheum, 2001, 44(9): 2065 - 2071.
- [5] 中国中医药研究促进会骨伤科分会. 膝骨关节炎中医诊疗指南(2020 年版)[J]. 中医正骨, 2020, 32(10): 1 - 14.
- [6] ZHANG L, ZHANG J, CHEN J, et al. Clinical research of traditional chinese medicine needs to develop its own system of core outcome sets[J]. Evid Based Complement Alternat Med, 2013[2021 - 05 - 27]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3838829/>. DOI: 10. 1155/2013/202703.
- [7] THORPE K E, ZWARENSTEIN M, OXMAN A D, et al. A pragmatic - explanatory continuum indicator summary (PRECIS): a tool to help trial designers [J]. CMAJ, 2009, 180(10): e47 - e57.
- [8] BELLAMY N, BUCHANAN W W, GOLDSMITH C H, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to

- antirheumatic drug therapy in patients with osteoarthritis of the hip or knee [J]. J Rheumatol, 1988, 15 (12): 1833 – 1840.
- [9] 曹月龙. 膝骨关节炎中医药临床研究的问题与启示[J]. 中医正骨, 2018, 30(9): 1 – 3.
- [10] KELLGREN J H, LAWRENCE J S. Radiological assessment of osteo – arthrosis [J]. Ann Rheum Dis, 1957, 16 (4): 494 – 502.
- [11] HUNTER D J, GUERMAZI A, LO G H, et al. Evolution of semi – quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score) [J]. Osteoarthritis Cartilage, 2011, 19(8): 990 – 1002.
- [12] PETERFY C G, GUERMAZI A, ZAIM S, et al. Whole – Organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis [J]. Osteoarthritis Cartilage, 2004, 12(3): 177 – 190.
- [13] MCALINDON T E, DRIBAN J B, HENROTIN Y, et al. OARSI clinical trials recommendations: design, conduct, and reporting of clinical trials for knee osteoarthritis [J]. Osteoarthritis Cartilage, 2015, 23(5): 747 – 760.
- [14] ROOS E M, ROOS H P, LOHMANDER L S, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self – administered outcome measure [J]. J Orthop Sports Phys Ther, 1998, 28(2): 88 – 96.
- [15] JACOBS C A, PEABODY M R, DUNCAN S T, et al. Development of the HOOSglobal to assess patient-reported outcomes in patients undergoing hip preservation procedures [J]. Am J Sports Med, 2018, 46(4): 940 – 946.
- [16] HAYS R D, SPRITZER K L, SCHALET B D, et al. PROMIS® – 29 v2.0 profile physical and mental health summary scores [J]. Qual Life Res, 2018, 27(7): 1885 – 1891.
- [17] KRZYCH L J, LACH M, JONIEC M, et al. The Likert scale is a powerful tool for quality of life assessment among patients after minimally invasive coronary surgery [J]. Kardiochir Torakochirurgia Pol, 2018, 15(2): 130 – 134.
- [18] SIT R W S, CHAN D C C, WONG W, et al. Translation, cross – cultural adaptation and validation of the traditional Chinese intermittent and constant osteoarthritis pain (ICO-AP) questionnaire for knee osteoarthritis [J]. BMJ Open, 2019, 9(3) [2021 – 05 – 27]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6475224/>. DOI: 10.1136/bmjopen-2018-026006.
- [19] MEENAN R F, GERTMAN P M, MASON J H. Measuring health status in arthritis. The arthritis impact measurement scales [J]. Arthritis Rheum, 1980, 23(2): 146 – 152.
- [20] PUA Y H, COWAN S M, WRIGLEY T V, et al. The lower extremity functional scale could be an alternative to the Western Ontario and McMaster Universities Osteoarthritis Index physical function scale [J]. J Clin Epidemiol, 2009, 62(10): 1103 – 1111.
- [21] BANDE J M, PAPASIDERO S B, MEDINA M A, et al. Validation of the HAQ – UP – A (Health Assessment Questionnaire – Upper Limbs – Argentine Version) in patients with rheumatoid arthritis [J]. Reumatol Clin (Engl Ed), 2021, 17(1): 20 – 24.
- [22] HUNTER D J, GUERMAZI A, ROEMER F, et al. Structural correlates of pain in joints with osteoarthritis [J]. Osteoarthritis Cartilage, 2013, 21(9): 1170 – 1178.
- [23] PETERFY C, WOODWORTH T, ALTMAN R. Workshop for consensus on osteoarthritis imaging: MRI of the knee [J]. Osteoarthritis and Cartilage, 2006, 14(Suppl 1): 44 – 45.
- [24] TUBACH F, RAVAUD P, BARON G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement [J]. Ann Rheum Dis, 2005, 64(1): 29 – 33.
- [25] LI G, TALJAARD M, VAN DEN HEUVEL E R, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how [J]. Int J Epidemiol, 2017, 46(2): 746 – 755.
- [26] ALTMAN D G, SCHULZ K F, MOHER D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration [J]. Ann Intern Med, 2001, 134(8): 663 – 694.

(收稿日期: 2021-06-20 本文编辑: 吕宁)