Measurements, composite scores and the art of 'cutting-off'

Pedro M Machado

Measuring is an essential part of medicine, both in research as well as in clinical practice. Clinical reasoning itself is a mental exercise based on a succession of tests and assessments (signs and symptoms and their quantification, impact on the patient, findings from the physical examination, results from laboratory and imaging investigations, among others) that result in making a diagnosis and making therapeutic decisions.¹ Lord Kelvin's (1824-1907) quote from 1883 extracted from a lecture on 'Electrical Units of Measurement' also applies to modern medicine: 'When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be'.

The use of general or disease-specific measuring tools has become routine in rheumatology. In addition to the critical need of these tools for drug development, they are also valuable for patient management. These tools vary from simple visual analogue scales or numerical rating scales, for example, for pain or global assessment of disease activity, to patient-reported outcomes based on multiple questions, such as the Health Assessment Questionnaire the (HAQ) or Bath Ankylosing Spondylitis Functional Index, to composite indices such as the 28-joint count Disease Activity Score (DAS28) and the Ankylosing Spondylitis Disease Activity Score, which combine patient-reported and/or physician-reported variables with the laboratory measurement of acute phase reactants (C reactive protein (CRP) erythrocyte sedimentation or rate). Composite measures can be particularly helpful and are more likely to give more

complete and reliable information about a certain health outcome such as disease activity.

Cut-offs increase the interpretability of a measurement or score, making it more meaningful and likely to be applied both in clinical practice as well as in the research setting. Ultimately, if the same score is being used across different settings, this allows for data to be compared and pooled, which can be useful for example for clinical benchmarking and for performing meta-analyses. Furthermore, an early diagnosis, determination of early therapeutic response and monitoring therapeutic response have now become increasingly important because very effective therapies are available for several rheumatic inflammatory diseases and they may even be more effective if used in early disease stages. Treat-to-target strategies have been proposed in diseases such as rheumatoid arthritis and spondyloarthritis, including ankylosing spondylitis and psoriatic arthritis (PsA),²⁻⁴ and composite scores can facilitate these treat-to-target strategies. The recently published Tight Control of Psoriatic Arthritis (TICOPA) trial confirmed that a treat-to-target approach can improve clinical outcomes for patients with early PsA.⁵

PsA is a heterogeneous disease, characterised by the involvement of peripheral joints, entheses, axial joints, skin and nails. The degree of involvement of these different domains varies and changes over time in individual patients. In PsA, several measures of disease activity states and therapeutic response have been proposed, including composite measures (table 1). $^{6-10}$ Some of these measures have been adapted from other diseases (Disease Activity Score (DAS), DAS28, European League Against Rheumatism (EULAR) response criteria and American College of Rheumatology (ACR) response criteria, all adapted from rheumatoid arthritis (RA), and the Disease Activity in PsA (DAPSA) score, adapted from reactive arthritis), while others have been specifically developed for use in PsA (Psoriatic Arthritis Response Criteria, Psoriatic Arthritis Disease Activity Score, Psoriatic Arthritis Joint Activity Index and Composite Psoriatic Disease Activity Index (CPDAI), modified versions of the CPDAI, Arithmetic Mean of Desirability Function, GRAppa Composite Exercise Index and Minimal Disease Activity).

Schoels and colleagues¹¹ present the results of a study developing cut-offs for disease activity states and response criteria according to the DAPSA and the clinical DAPSA (cDAPSA). The investigators started by retrieving 30 patient profiles from an observational data set and performed an email-based survey among 44 international rheumatology experts asking them to classify the disease activity state of each patient (remission (REM), low disease activity (LDA), moderate disease activity (MDA) and high disease activity (HDA)) based on their 66/68 joint counts. patient global assessments (visual analogue scale; VAS), patient pain scores (VAS) and CRP values. Distributions of DAPSA/ cDAPSA scores within each state were analysed and cut-offs determined by calculating the respective 25th and 75th perof DAPSA/cDAPSA, centiles with numerical differences between the 75th percentile of the lower and the 25th percentile of the adjacent higher disease activity state reconciled by calculating their mean (if necessary). The following DAPSA cut-offs were proposed for REM, DAPSA≤4; for LDA, 4<DAPSA≤14; for MDA. $14 < DAPSA \le 28$; for HDA, DAPSA>28. Derived cDAPSA cut-offs were similar and the investigators arbitrarily proposed to reduce the cDAPSA cut-off for MDA and HDA by one point compared with the DAPSA, to account for the putative higher levels of CRP in patients with these levels of disease activity: for REM, cDAPSA≤4; for LDA, MDA, 4<cDAPSA<13; for 13<cDAPSA≤27; for HDA, cDAPSA>27.

To define minor, moderate and major treatment response, the investigators used peak values of Cohen's κ agreement to detect the highest agreement between DAPSA percentage response and ACR20/ 50/70 response in three randomised controlled trials. After summarising all the analyses, the following DAPSA response criteria were proposed: minor response, 50% change in DAPSA; moderate response, 75% change; major response, 85% change. Discriminative validity was assessed using χ^2 statistics.

This study represents an important contribution to the field of PsA and it may stimulate the use of the DAPSA/cDAPSA among practising rheumatologists and PsA researchers. Of note, the DAPSA/cDAPSA has a simple arithmetic formula and can be mentally calculated in clinic, which offers an advantage compared with other more complex composite measures. This



Centre for Rheumatology Research & MRC Centre for Neuromuscular Diseases, University College London, London, UK

Correspondence to Dr Pedro M Machado, Centre for Rheumatology Research & MRC Centre for Neuromuscular Diseases, Box 102, 8-11 Queen Square, London WC1N 3BG, UK; p.machado@ucl.ac.uk, pedrommcmachado@gmail.com

| Table 1 Summary of composite scores proposed to assess disease activity in psoriatic arthritis and variables/domains included in the composite scores | | | | | | | | | | | | |
|---|-----|---------------------|-------|-------|--------|-------|--------|-----------|--------|--------|----------------------|-----|
| | ACR | DAS/DAS28/ EULAR | PSaRC | DAPSA | cDAPSA | CPDAI | mCPDAI | CPDAI-JED | PsAJAI | PASDAS | ADMDF/GRACE index | MDA |
| TJC/SJC | Х | Х | Х | Х | Х | Х | Х | Х | Х | Х | Х | Х |
| PGA | Х | х | Х | Х | Х | | | | Х | Х | Х | Х |
| PhGA | Х | | Х | | | | | | Х | Х | | |
| Patient pain | Х | | | Х | Х | | | | Х | | | Х |
| Dactylitis | | | | | | Х | Х | Х | | Х | | |
| Enthesitis | | | | | | Х | Х | Х | | Х | | Х |
| Spinal involvement | | | | | | Х | | | | | | |
| Skin involvement | | | | | | Х | Х | | | | Х | Х |
| Acute phase reactants | Х | Х | | Х | | | | | Х | Х | | |
| Physical function (HAQ) | Х | | | | | Х | Х | | Х | | Х | Х |
| HRQoL | | | | | | Х | | | | Х | Х | |
| PGA of joint involvement | | | | | | | | | | | Х | |
| | | | | | | | | | | | | |

ACR, American College of Rheumatology response criteria; AMDF, Arithmetic Mean of Desirability Function; cDAPSA, clinical version of the DAPSA; CPDAI, Composite Psoriatic Disease Activity Index; CPDAI-JED, CPDAI joints, entheses, dactyliti; DAPSA, Disease Activity in Psoriatic Arthritis; DAS, Disease Activity Score; DAS28, 28-joint Disease Activity Score; EULAR, European League Against Rheumatism response criteria; GRACE Index, GRAppa Composite Exercise Index (GRACE index=(1-AMDF)); HAQ, Health Assessment of Questionnaire; HRQoL, Health-related quality of life; mCPDAI, modified CPDAI; MDA, minimal disease activity; PASDAS, Psoriatic Arthritis Disease Activity Score; PGA, Patient Global Assessment; PhGA, Physician Global Assessment; PsAJAI, Psoriatic Arthritis Joint Activity Index; PsARC, Psoriatic Arthritis Response Criteria; TJC/SJC, Tender Joint Count/Swollen Joint Count.

study also defined new potential (articular) treatment targets in PsA, namely REM (a more ambitious target) and LDA (a less ambitious target) according to the DAPSA. The cut-off for MDA could potentially be used as a criterion for inclusion of PsA patients in clinical trials of biological therapies, but this use still requires further validation.

Some methodological aspects of this study are of interest to debate. Conceptually, the option of allowing the same patient to be included in more than one disease activity state according to the external construct (opinion of the expert) is a questionable approach because it artificially inflates the number of patients in each disease activity state. However, the fact that in a sensitivity analysis, with patient profiles assigned to a particular disease activity state only, if a majority (>50%) of experts adjudicated that state, resulted in similar results is somehow reassuring. It would have also been of interest to incorporate the patients' perspective in the determination of the cut-offs by using patient-reported disease activity states as the external construct and assess how these compared with the physicians' perspective.

Regarding the methodology used for cut-off determination of disease activity states (25th and 75th percentiles of the distribution of DAPSA/cDAPSA scores), this is a valid but not the only possible approach to cut-off determination. Other methods commonly used for establishing 'optimal' cut-points are the point on the receiver operating characteristic curve (ROC curve) closest to (0, 1), ie. the point closest to perfect differentiation, and the Youden index, that is, the point farthest from none differentiation (figure 1).¹² These two ROC curve points do not necessarily agree and the Youden index is generally favoured in the literature. Furthermore, clinical reasoning should also be part of the 'art of cutting-off' and depending on the clinical scenario and the need for stringency in the cut-off being established, the use of alternative criteria such as the ROC curve based cut-offs of 90% specificity/sensitivity or the score distribution cut-offs of the 10th/90th percentile may be more appropriate.¹³

Importantly, the cut-off selection procedure should always be an informed decision that takes into account the clinical situation (eg, potential treatment implications of the cut-off) and the relative consequences of false-negative and false-positive test results (which may differ according to different clinical decision-making situations). A step further in cut-off determination that has rarely been taken because of its complexity is the optimisation of cut-offs with regard to the cost implications of false-positive and false-negative results.¹² ¹⁵

On the topic of cut-off determination, it should also be borne in mind that the cut-off values for diagnostic tests are usually derived using different methods among which the Gaussian (normal) distribution method is the most commonly used. Based on this method, a cut-off value is defined as the mean plus two SD of the negative reference sample. The rationale of the two SD approach is to establish a cut-off value providing a specificity of 97.5%, although this method may not be adequate if the test values follow a skewed or multimodal distribution.¹⁵

Regarding the methodology used by Schoels and colleagues for cut-off determination of response criteria for the DAPSA, the use of the ACR20/50/70 as external criteria is supported by the fact that they have been extensively used to assess response in PsA clinical trials (albeit being adapted from RA). However, the ACR20/50/70 criteria focus on the joint



Figure 1 Simulation of a receiver operating characteristic (ROC) curve and classical 'optimal' cut-points (adapted from ref ¹⁰). The vertical lines and reference arcs identify the Youden index (solid lines) and the closest point to (0,1) (dashed lines) and their corresponding cut-points (red dot and green dot, respectively); the cut-point for the Youden index corresponds to the point in the ROC curve with the maximum value of (sensitivity —(1—specificity)); the cut-point for the closest point to (0, 1) corresponds to the point in the ROC curve with the minimum value of $((1-\text{sensitivity})^2+(1-\text{specificity})^2)$.

domain, which limits their value as a measure of change in a multidimensional disease such as PsA. Furthermore, the use of the κ statistic to define levels of response according to the DAPSA also has limitations. Since the scores on the external construct are dichotomous, the use of the area under the ROC curve would be the preferred method with sensitivity and specificity being the preferred parameters.¹⁶¹⁷ It would have also been of interest to use a 'global rating of change' (GRC) as the external construct and to compare the results with the ACR20/50/ 70 external construct. The GRC is a Likert-type scale scored for change by the patient (eg. 'much better', 'better', 'unchanged', 'worse' and 'much worse'). The GRC is a useful external criterion for ROC-curve analysis.^{18–20} using

Finally, the use of a measure that omits some of the PsA disease domains focusing mainly on peripheral joint activity is not consensual. The authors argue that "given the relatively low frequency of entheseal involvement and the availability of separate tools for both entheseal and spinal involvement, it seems reasonable to focus on joint activity and systemic levels of inflammation" and that the various domains respond differently to the various therapies, a comment that is particularly relevant regarding the skin versus articular manifestations. Conversely, one could also argue that by not taking the various disease domains into account we are missing information, which may be particularly relevant in terms of defining disease REM. Of note, criteria to define a REM-like state ('minimal disease activity', allowing minor disease activity and not necessarily 'complete REM'), taking multiple domains into account, have been previously proposed and validated in PsA. According to Coates and colleagues,²¹ patients can be classified as having 'minimal disease activity' if they fulfil five of seven outcome measures: tender joint count ≤ 1 ; swollen joint count ≤ 1 ; psoriasis activity and severity index (PASI) ≤ 1 or body surface area ≤3; patient pain VAS score ≤ 15 (0–100 scale); patient global disease activity VAS score ≤20 (0-100 scale); HAQ score ≤ 0.5 and tender entheseal points ≤ 1 . Interestingly, the 'minimal disease activity' criteria were the criteria used as therapeutic target in the TICOPA trial.

In summary, diseases activity states and response criteria according to the DAPSA/ cDAPSA have been developed and are ready to be rolled out into clinical practice and clinical trials. Further validation and comparative studies, namely with other measures and disease activity states and therapeutic responses that have been proposed in PsA, are required until a final consensus can be reached about the best measure(s) and treatment target(s) to use in PsA. This discussion is ongoing²²⁻²⁴ and being fuelled by recent publications on the topic.⁵¹¹ In the future, further validation of these newly developed cut-offs should be undertaken in larger cohorts of patients from observational studies as well as further assessment of the performance of the score in different subgroups of PsA patients. It would also be of interest to explore the concept of 'flare' according to the DAPSA/cDAPSA as well as the prognostic validity of DAPSA/cDAPSA cut-off levels with regard to structural damage and disability.

Acknowledgements PMM is supported by the National Institute for Health Research (NIHR) Rare Diseases Translational Research Collaboration (RD TRC) and by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR or the Department of Health.

Contributors PMM drafted the article and revised it critically for important intellectual content.

Funding National Institute for Health Research.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.



To cite Machado PM. *Ann Rheum Dis* 2016;**75**:787–790.

Received 8 November 2015 Revised 27 December 2015 Accepted 31 December 2015 Published Online First 22 January 2016



► http://dx.doi.org/10.1136/annrheumdis-2015-207507

Ann Rheum Dis 2016;**75**:787–790. doi:10.1136/annrheumdis-2015-208274

REFERENCES

- Landewé RB, van der Heijde DM. Principles of assessment from a clinical perspective. *Best Pract Res Clin Rheumatol* 2003;17:365–79.
- 2 Smolen JS, Aletaha D, Bijlsma JW, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631–7.
- 3 Smolen JS, Braun J, Dougados M, et al. Treating spondyloarthritis, including ankylosing spondylitis and psoriatic arthritis, to target: recommendations of

an international task force. *Ann Rheum Dis* 2014;73: 6–16.

- 4 Smolen JS, Breedveld FC, Burmester GR, *et al.* Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis* 2016;75:3–15.
- 5 Coates LC, Moverley AR, McParland L, et al. Effect of tight control of inflammation in early psoriatic arthritis (TICOPA): a UK multicentre, open-label, randomised controlled trial. *Lancet* 2015;386:2489–98.
- 6 Helliwell PS, Fitzgerald O, Mease PJ. Development of composite measures for psoriatic arthritis: a report from the GRAPPA 2010 annual meeting. J Rheumatol 2012;39:398–403.
- 7 Her M, Kavanaugh A. A review of disease activity measures for psoriatic arthritis: what is the best approach? *Expert Rev Clin Immunol* 2014;10:1241–54.
- 8 Mease PJ. Measures of psoriatic arthritis: Tender and Swollen Joint Assessment, Psoriasis Area and Severity Index (PASI), Nail Psoriasis Severity Index (NAPSI), Modified Nail Psoriasis Severity Index (mNAPSI), Mander/Newcastle Enthesitis Index (MEI) Leeds Enthesitis Index (LEI), Spondyloarthritis Research Consortium of Canada (SPARCC). Maastricht Ankylosing Spondylitis Enthesis Score (MASES), Leeds Dactylitis Index (LDI), Patient Global for Psoriatic Arthritis, Dermatology Life Quality Index (DLQI), Psoriatic Arthritis Quality of Life (PsAQOL), Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F), Psoriatic Arthritis Response Criteria (PsARC), Psoriatic Arthritis Joint Activity Index (PsAJAI), Disease Activity in Psoriatic Arthritis (DAPSA), and Composite Psoriatic Disease Activity Index (CPDAI). Arthritis Care Res (Hoboken) 2011;63 (Suppl 11):S64-85.
- 9 Machado PM, Raychaudhuri SP. Disease activity measurements and monitoring in psoriatic arthritis and axial spondyloarthritis. *Best Pract Res Clin Rheumatol* 2014;28:711–28.
- 10 Helliwell PS, FitzGerald O, Fransen J. Composite disease activity and responder indices for psoriatic arthritis: a report from the GRAPPA 2013 meeting on development of cutoffs for both disease activity states and response. J Rheumatol 2014;41:1212–7.
- 11 Schoels MM, Aletaha D, Alasti F, et al. Disease activity in psoriatic arthritis (PsA): defining remission and treatment success using the score. Ann Rheum Dis 2016;75:811–8.
- 12 Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670–5.
- 13 Machado P, Landewé R, Lie E, *et al.* Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. *Ann Rheum Dis* 2011;70:47–53.
- 14 Consolaro A, Ruperto N, Bracciolini G, et al. Defining criteria for high disease activity in juvenile idiopathic arthritis based on the juvenile arthritis disease activity score. Ann Rheum Dis 2014;73: 1380–3.
- 15 Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000;45:23–41.
- 16 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- 17 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of

measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.

- 18 Wyrvich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract* 2000;6:39–49.
- 19 Turner D, Schünemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol 2010;63:28–36.
- 20 Tubach F, Ravaud P, Beaton D, et al. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188–93.
- 21 Coates LC, Helliwell PS. Validation of minimal disease activity criteria for psoriatic arthritis using interventional trial data. *Arthritis Care Res (Hoboken)* 2010;62:965–9.
- 22 Helliwell PS, Coates LC. The definition of remission in psoriatic arthritis: can this be accurate without

assessment of multiple domains? *Ann Rheum Dis* 2015;74:e66.

- Schoels MM, Aletaha D, Smolen JS. Defining remission and treatment success using the DAPSA score: response to letter by Helliwell and Coates. *Ann Rheum Dis* 2015;74:e67.
 Coates LC, Helliwell PS. Treating to target in psoriatic
- 24 Coates LC, Helliwell PS. Treating to target in psoriatic arthritis: how to implement in clinical practice. *Ann Rheum Dis* 2016;75:640–3.