

Distinctions Between Diagnostic and Classification Criteria?

Rohit Aggarwal¹, Sarah Ringold², Dinesh Khanna³, Tuhina Neogi⁸, Sindhu R. Johnson⁴, Amy Miller⁵, Hermine I. Brunner⁶, Rikke Ogawa⁷, David Felson⁸, Alexis Ogdie⁹, Daniel Aletaha¹⁰, Brian M. Feldman¹¹

¹Department of Medicine, University of Pittsburgh; ²Department of Pediatrics, Seattle Children's Hospital; ³Department of Medicine, University of Michigan; ⁴Division of Rheumatology, Toronto Western and Mount Sinai Hospitals, University of Toronto; ⁵American College of Rheumatology; ⁶Department of Pediatrics, Cincinnati Children's Hospital and Medical Center; ⁷University of California Los Angeles; ⁸Department of Medicine, Boston University School of Medicine; ⁹Department of Medicine, University of Pennsylvania; ¹⁰Department of Rheumatology, Medical University of Vienna; ¹¹Toronto Scleroderma Program, Toronto Western and Mount Sinai Hospitals, University of Toronto; The Hospital for Sick Children, University of Toronto.

Acknowledgement: This work was supported by the American College of Rheumatology Classification and Response Criteria Subcommittee of the Committee on Quality of Care.

Review Article and literature search

Manuscript word: 4005

Corresponding author:

Rohit Aggarwal, MD MS

Assistant Professor of Medicine

3601 5th Ave, Suite 2B,

Pittsburgh, PA, 15217

Phone: 4123838861

Fax: 4123838864

Email: aggarwalr@upmc.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/acr.22583

© 2015 American College of Rheumatology

Received: Dec 30, 2013; Revised: Feb 27, 2015; Accepted: Mar 10, 2015

Rheumatologists face unique challenges in discriminating between rheumatologic and non-rheumatologic disorders with similar manifestations, and in discriminating among rheumatologic disorders with shared features. The majority of rheumatic diseases are multisystem disorders with poorly understood etiology; they tend to be heterogeneous in their presentation, course, and outcome, and do not have a single clinical, laboratory, pathological, or radiological feature that could serve as a “gold standard” in support of diagnosis and/or classification. Thus, the development of criteria for use in routine clinical care and in clinical research has been an important focus in rheumatology. Improved understanding of disease pathogenesis and new diagnostic tools have led to reexamination of existing classification and diagnostic criteria with updated classification criteria for some diseases being endorsed recently (1, 2).

The American College of Rheumatology (ACR) Subcommittee on Classification and Response Criteria is responsible for guiding the development and validation of new classification and response criteria that are eventually considered for ACR endorsement. This includes review of proposals for the development of new criteria sets and providing the ACR leadership with recommendations for development and approval of new classification and response criteria sets (1, 3-5). The Subcommittee has previously published a guidance paper for the development of classification and response criteria (6). This prior work has provided details about the rationale for the ACR’s position on classification criteria, but clarification around the issue of diagnostic criteria was lacking. Indeed, the ACR endorsed preliminary diagnostic criteria for fibromyalgia (7) in 2010, which prompted discussions about whether the Subcommittee should also support the development and ACR endorsement of diagnostic criteria, in addition to that of classification and response criteria.

The primary objectives of this current article, by former and current members of the Subcommittee on Classification and Response Criteria, are to compare diagnostic and classification criteria, using specific examples from the published literature, and to clarify the ACR’s position on both types of criteria.

Diagnostic criteria

Diagnosis may be defined as the determination of the cause or nature of an illness by evaluation of the signs, symptoms and supportive tests in an individual patient. Diagnostic criteria are a set of signs, symptoms, and tests for use in routine clinical care to guide the care of individual patients.

Diagnostic criteria are generally broad and must reflect the different features of a disease (heterogeneity), with a view to accurately identify as many people with the condition as possible. Given this complexity, the development and validation of diagnostic criteria can be quite challenging. The Diagnostic and Statistical Manual of Mental Disorders (DSM) is likely the best-known example of diagnostic criteria. Its initial development was prompted by the observation of extremely poor agreement among providers regarding patients’ psychiatric diagnoses. There are only a few validated diagnostic criteria in rheumatology, and clinicians commonly establish a diagnosis

based on subjective combination of clinical signs/symptoms, available clinical tests, and knowledge about the epidemiology of their geographical area.

Classification criteria

Classification criteria are standardized definitions that are primarily intended to create well-defined, relatively homogenous cohorts for clinical research; they are not intended to capture the entire universe of possible patients, but rather to capture the majority of patients with key shared features of the condition. Hence the goal of classification differs from the intent of diagnostic criteria. Validated classification criteria are considered critical to the interpretation of study findings and comparisons of results between studies. Despite facilitating the comparison of study results, classification criteria have the potential to restrict the external validity of studies, as interventions may perform differently in the study participants who fulfill classification criteria for a disease than in the broader group of persons having been diagnosed with the same disease, i.e., those that share only some but not other disease manifestations considered in classification criteria.

Although they may provide some framework to aid diagnosis and are frequently used this way in teaching, classification criteria traditionally have high specificity (defined as proportion of patients that are known not to have the disease who will test negative for it) which generally comes at the expense of somewhat lower sensitivity (defined as proportion of people that are known to have the disease who test positive for it). Consequently, few individuals are incorrectly labeled as having a disease (false positives), but a proportion of individuals with the disease diagnosis may be “missed”, i.e., labeled as not having the disease based on the classification criteria (false negatives). This may make classification criteria inappropriate for use in routine clinical care (8).

The Continuum of Diagnosis and Classification

While *diagnostic criteria* may be different from *classification criteria* (9), at least in their intended purpose, in reality they represent two ends of a continuum (9). The “distance” between diagnostic and classification criteria on this continuum depends on various factors but includes disease prevalence, geographical area, and prevalence of ‘mimickers’, among other factors. Where disease etiology is well defined such as in gout and Lyme disease, diagnostic and classification criteria may be very similar and used interchangeably. If sufficient internal and external validity for diagnosis is demonstrated in a given population, classification criteria can be diagnostic. In theory, a diagnosis applies classification criteria to an individual patient (9). Therefore, when classification criteria have perfect sensitivity and specificity (both 100%), classification and diagnostic criteria are synonymous and would correctly identify every single individual case (10). However, because disease features are not typically identical among patients with a given disease, classification criteria are not 100% accurate, leaving a certain proportion of patients misclassified. Because of this possibility of misclassification, meeting classification criteria is not equivalent to carrying a given diagnosis. Only physicians considering features of an individual patient, beyond those represented in the classification criteria, in addition to extraneous factors (such as local prevalence of

conditions that are on the differential diagnosis) can establish a diagnosis for an individual patient.

Due to the lack of gold standards in rheumatology, any criteria (classification or diagnostic) are difficult to establish. As detailed below, compared with classification criteria, there are an array of factors that pose greater challenges and clinical implications for the development and validation of diagnostic criteria. Even in those situations in which diagnostic criteria can be established, the question remains as to whether the ACR and/or other international organizations should endorse one single set of diagnostic criteria.

Literature Review

A systematic search of articles addressing classification and/or diagnostic criteria in the rheumatic diseases was performed by an experienced librarian (RO), considering the PubMed database (1940–2011) and the Cochrane Central Register of Controlled Trials database (from 1996 to 2011). In combining the search terms for 'Diagnosis AND Classification AND Rheumatic Disease AND Methodology', 3,825 citations from PubMed and 88 from the Cochrane database were identified. Two reviewers (RA and SR) independently screened the titles and abstracts of articles for relevance to classification and diagnostic criteria in rheumatic diseases. Abstracts were screened to identify articles that defined, updated, addressed, reviewed, or commented on methodological aspects of classification or diagnostic criteria for the rheumatic diseases. Screening of titles and abstracts excluded 3,681 articles, leaving 232 articles for detailed review. This led to the identification of 97 articles that were deemed relevant for the evaluation of classification and diagnostic criteria which are considered in this manuscript (Figure 1). Relevant articles were defined as the articles that either a) illustrate differences between classification criteria and diagnostic criteria; b) identify key advantages and disadvantages of classification or diagnostic criteria; or c) evaluate performance characteristics of classification or diagnostic criteria. Eighteen additional relevant articles were identified through hand searching the bibliography of the initially identified 97 articles (Figure 1). Articles fell into one of the following six categories: 1) study of or commentary on differences and/or similarities between classification criteria and physician assessment; 2) description of the performance of classification or diagnostic criteria in various populations, geographical regions, or different practice settings; 3) proposal of original or revised classification or diagnostic criteria, 4) comparison of the performance of established classification criteria, 5) description of various cohorts using established classification criteria; 6) discussion of either classification or diagnostic criteria not otherwise related to one of the above categories.

Summary of the published literature:

1. Examples of the differential performance of classification criteria in relation to physician assessment. No examples of direct comparison of diagnostic criteria to classification criteria were identified in the literature review. The performance of classification criteria as diagnostic tools has been assessed in a handful of studies. Because of the lack of gold standards for diagnosis and classification, performance of

criteria sets was compared to physician judgment in these studies. For example, Rao et al. assessed the measurement properties of the 1990 ACR vasculitis classification criteria when used as diagnostic criteria relying on the treating rheumatologist's final diagnosis as the criteria standard (11). In this study only 38/51 (75%) of patients with vasculitis fulfilled the ACR classification criteria for one or more types of vasculitis, and 31/147 (21%) (a Specificity of 79%) of patients without vasculitis also fulfilled these criteria. This illustrates that the 1990 ACR classification criteria had a relatively low sensitivity and specificity for predicting the presence of a specific type of vasculitis in an individual patient seen in routine practice when compared against physician diagnosis. A separate study reported that the Chapel Hill Consensus Conference (CHCC) classification criteria for vasculitis correctly identified only 30% of patients with the disease, when compared with physician's assessment (gold standard) (12). Likewise, Patarroyo et al. reported that 65.8% of patients with histopathologically proven vasculitis from a single center failed to fit into a discrete type of vasculitis as defined by the CHCC criteria (13).

Clinicians have voiced concern that too high a number of patients diagnosed with SLE in clinical practice fail to meet the SLE classification criteria (14, 15). Similarly, the diagnosis of knee osteoarthritis (OA) made by community physicians is only in fair agreement with the ACR classification of knee OA (Kappa coefficient = 0.28) (16). Other examples of diseases in which the classification criteria do not perform well when compared with clinical diagnosis of treating rheumatologists or experts include juvenile idiopathic arthritis and systemic sclerosis (17-19). These examples may not only reflect differences between physician decision-making versus classification criteria, but also that older classification criteria may require revision; several have now been updated or are in the process of being updated (e.g., RA, systemic sclerosis – (1, 2); in development: vasculitis (20), gout (21)).

Nonetheless, these examples suggest that classification differs from diagnosis and may generally underreport the presence of a disease because they capture a narrower range of disease severity than that treated in routine clinical practice since classification criteria tend to identify a uniform population for participation in clinical trials at the expense of excluding some patients with less common phenotypes as suggested by above examples (22). The rationale for perhaps favoring specificity over sensitivity for classification criteria in the setting of clinical trials is to avoid exposing patients who may not have the disease to undue risks of experimental interventions.

2. Need for Revision of Classification Criteria. Newer revised classification criteria may perform better than some older classification criteria in terms of sensitivity and/or specificity with use of better datasets and methodology (1, 5, 23). The more favorable risk-benefit profiles of therapeutics and recognition that early therapy may affect long term prognosis has prompted trials in individuals with rheumatic diseases who do not yet meet thresholds for traditional classification. Some recent classification criteria have, consequently, had more of a focus on improving sensitivity of criteria (1, 23). For example, the 1987 ACR Rheumatoid Arthritis (RA) Criteria were believed to miss early disease (lack sensitivity) that led to development of new 2010 RA criteria (better

sensitivity) so that we could identify patients early for intervention studies. It should be considered, however, that this may increase the chance of false positives in the absence of gold standard tests, and likely has implications for prevalence estimates of these diseases as well as for clinical practices. On the other hand, lack of sufficient specificity in criteria (i.e., false positives) also has bearing on enrollment of patients into trials of agents with unclear safety: efficacy profiles; as a result, some recent classification criteria have aimed to improve specificity (21).

3. Effects of geographic area, practice setting and race on criteria performance.

The performance of any criteria (classification or diagnostic) is dependent on the prevalence of the disease in a given geographical area or clinical setting (e.g., community clinic versus tertiary care facility). While sensitivity and specificity are functions of the screening test or criteria set and are not influenced by disease incidence or prevalence, the predictive validity changes with the prevalence of the disease. The performance of criteria depends on both the pre-test probability of the disease (which reflects the prevalence of the disease as well as potential ‘mimickers’), and the sensitivity and specificity of the criteria themselves. Given the low prevalence of certain rheumatologic diseases, the positive predictive value (defined as the proportion of positive test results that are true positives) of any criteria set will generally be low. This is probably the reason why there are few diagnostic criteria sets in rheumatology. For example, in areas where Behçet’s disease is endemic (high pre-test probability), patients with recurrent oral ulcers may be accurately diagnosed and treated with few supporting criteria, whereas in the United States, where the disease prevalence is low, any set of diagnostic criteria will have a low positive predictive value. Similarly, the performance of the European Spondyloarthropathy Study Group (ESSG) classification criteria for diagnosis for spondyloarthropathies varies across private practices, outpatient departments and patients admitted to the hospital, depending on prevalence of SpA in each region (24). For example, in Spanish rheumatology services where SpA prevalence is 49%, the likelihood of a patient under 35 years of age fulfilling the ESSG criteria for SpA is 87%. Conversely, in clinics in France where the age-adjusted prevalence of SpA is lower at 22.5%, the likelihood of fulfilling ESSG criteria is estimated to be 70% (24). While classification and diagnostic criteria perform differently in different clinical and geographic settings, this difference is less pronounced for classification criteria as their intended purpose is to identify patients with similar disease features for studies within different populations, rather than guiding clinicians in establishing diagnoses and making treatment decisions.

The same concept applies to patients of different races or ethnicities within a geographic area. Indeed, the measurement properties of classification criteria can differ markedly when used in populations other than those used for classification criteria development. As one example, a study conducted in Asia found that only 12/71 (17%) of patients with Henoch Schoenlein Purpura fulfilled both the ACR vasculitis classification criteria and those of the Chapel Hill Consensus Conference (25). The variable performance of both classification and diagnostic criteria in different settings highlights the differences between these types of criteria and further illustrates the difficulty of developing diagnostic criteria with consistent performance across

populations. Diagnostic criteria will typically need to be based on local prevalence of the disease and of other diseases in the differential diagnosis, which is not be practical given the vast differences in epidemiology of the most rheumatic diseases in different clinical settings and geographical areas. Performance of classification criteria are also affected by their application to patients other than the intended target population, for example if 2010 RA classification criteria applied to burnt out deforming nodular RA, when it was intended for early active RA.

4. Well-defined disease phenotypes. One of the main differences between classification and diagnostic criteria is that classification criteria aim to assemble a study sample that is well-defined and representative of the vast majority of individuals with the disease. In contrast, diagnostic criteria aim to identify all individuals with the disease, including those with unusual features or presentations. Achieving a relatively homogeneous disease population is important for any classification criteria so that multiple studies and populations can be compared or combined. On the other hand, to be highly sensitive while preserving acceptable specificity, diagnostic criteria have to allow for all the heterogeneous manifestations of the disease (which may be difficult to achieve in rheumatic diseases). SLE is a prototypical example of a disease with heterogeneous presentations. Although SLE classification criteria can support an SLE diagnosis, clinicians still must diagnose SLE based upon the totality of patients' disease manifestations (26, 27). Classification criteria for SLE perform reasonably well for making a diagnosis in academic medical centers that attract more severe or advanced disease, and are typically those with a higher pretest probability of having the diagnosis (26). However, the SLE classification criteria may fail to recognize individuals with milder phenotypes or uncommon presentations of the disease. Classification criteria tend to include phenotypic features that have sufficiently high prevalence, whereas low prevalence features that may be very specific and helpful in diagnosis are typically not included in criteria sets due to the expected low yield from including such a feature.

5. Rheumatic diseases where both diagnostic and classification criteria are feasible. Single sets of criteria that serve for both classification and diagnosis appear feasible for diseases for which there is a "gold standard". The presence of monosodium urate crystals (MSU) in synovial fluid during an episode of acute arthritis is widely considered diagnostic for gout. In fact, compared to the presence of MSU crystals, the 1977 ACR classification criteria for acute gout have shown limited diagnostic accuracy (28-30). In one study, diagnosis by primary care physicians correctly identified 93.5% of patients with positive MSU crystals, indicating that clinician diagnosis of gout can be at least as sensitive as acute gout classification criteria (28). A similar argument for diagnostic criteria can be applied to any rheumatic disease where the pathology is well understood and/or the etiology is well defined. Infectious arthritides, such as septic arthritis, can be diagnosed based on gold standard tests, and a diagnostic criteria set can be devised. For such diseases, diagnostic criteria are also suited to guide subject identification for research studies because diagnostic criteria perform as well as classification criteria in terms of sensitivity and specificity.

6. Resources and feasibility. Feasibility, acceptability and available resources are other limitations in establishing universally accepted diagnostic criteria. Clinicians may be faced with limited access to or affordability of testing in certain geographical regions, patients' own financial and/or insurance limitations, patient preferences, and overall health condition, among others, when deciding on strategies to establish a diagnosis. This could necessitate making a diagnosis and subsequently initiating treatment solely on clinical grounds. More stringent diagnostic criteria that require a particular laboratory or imaging test or surgical procedure could constitute a hurdle for patients and clinicians and has the potential to postpone the initiation of effective therapy.

7. Health priorities of a country or geographical area. In a malaria endemic area, a doctor can empirically diagnose malaria in a patient with high grade fever and chills, and start empiric treatment (9). In most cases, the initiation of malaria treatment, without waiting for the results of confirmatory test, prevents serious complications and even death. The above clinical approach in these malaria endemic regions outweighs the harm of over-diagnosis and over-treatment with a relatively non-toxic medication in patients without malaria. Conversely, a similar approach in a Nordic country would likely be unacceptable and irrational. Thus, the health priorities and conditions in different countries/geographical areas often dictate the diagnostic approaches to be used, which suggests that a single universal diagnostic criteria set cannot always be applied equally in different regions of the world. However, classification criteria for the purpose of enrollment into clinical trials and epidemiological studies may be used across the globe, with high specificity even if a few cases are missed without affecting the internal validity of the study.

8. Legal, financial and treatment implications. Unlike classification of a disease for research purposes, the accurate diagnosis of a particular disease has important implications for a patient's treatment as well as billing and reimbursement. Highly specific diagnostic criteria will leave some patients undiagnosed (as no criteria will ever be 100% sensitive). This means that they may be denied treatment coverage if insurance companies and government agencies use the diagnostic criteria as a standard for reimbursement. Similarly, patients incorrectly diagnosed with an illness (as 100% specificity is difficult) can encounter difficulty in obtaining health insurance or life insurance, and may be exposed unnecessarily to incorrect, potentially harmful, therapies.

9. Undifferentiated Rheumatic Diseases. Many rheumatic disease patients present to their physicians at an undifferentiated stage, which may later evolve into more established disease. While classification criteria are typically applied at a given time point, they can be reevaluated as individuals may fulfill criteria as the disease manifestations evolve over time. Unlike acute infections, many rheumatic diseases evolve over time and cross-sectional application of any criteria as either "disease present or absent" is too simplistic. Moreover, some cases may never evolve into well-established disease and others may transform from one presumed condition to another with time. Strict universal diagnostic criteria may limit the ability to make a clinical diagnosis and treat undifferentiated diseases based on symptoms. This was emphasized in recent studies on the outcome of early arthritis in which 32 to 53% of

patients remained unclassified after one year of observation (31, 32). Similarly, in a 3-year follow up study of 270 patients with early arthritis, the diagnosis remained unclear in 61/270 (23%), and changed between the first and last examination in 96 of the remaining 209 cases (46%) (33).

10. Complex decision making for diagnosis. Finally, clinicians perform a complex multi-step process in order to make a diagnosis of a rheumatic disease. This includes balancing the post-test probability of the disease with thresholds for further action based on factors such as severity of disease, risks of further testing, side effects of treatment and ruling out other conditions in the differential diagnosis (e.g., infections and malignancies). It is difficult to establish diagnostic criteria that may satisfactorily perform this complex multi-step process.

Role of the American College of Rheumatology.

Classification criteria have demonstrated utility for identifying well-defined, relatively homogenous groups of patients for clinical research purposes across different regions, and have some utility as teaching tools in the clinical setting; however, they may not capture all physician-assigned diagnoses. Conversely, diagnostic criteria appear to be more impacted by practice setting than classification criteria, and the performance characteristics of diagnostic criteria may vary significantly due to differences in disease prevalence and the severity and manifestations of disease in different settings. Given these differences, concerns regarding the challenges in generating diagnostic criteria with consistent performance properties, and the legal and financial implications of diagnostic criteria, the ACR will provide approval only for classification criteria and will no longer consider funding or endorsement of diagnostic criteria. However, the ACR recognizes the importance of diagnostic tools to aid rheumatologists in their clinical practice and encourages their development. The ACR anticipates that both types of criteria will continue to evolve as the pathogenesis of the rheumatic diseases becomes better understood and as comparative effectiveness studies gain increased emphasis.

Conclusion: Diagnostic and classification criteria play central roles in clinical rheumatology practice. Unfortunately, existing criteria for rheumatic diseases are not always properly applied, most often due to confusion between the two types of criteria. Classification criteria are used as a standardized means of including a well-defined set of patients in research studies to ensure comparability across studies. Given the heterogeneous nature of rheumatic diseases, it is difficult to capture the full range of disease presentations by any single set of criteria. Therefore, any criteria would be expected to fail to capture some cases of a disease by capturing a more homogenous population and narrower range of disease severity than that treated in routine clinical practice. Nonetheless, classification criteria are critically important for advancing research in the field of rheumatology, enabling the conduct of clinical trials and epidemiologic studies with well-defined patient populations. The process of diagnosis, particularly for complicated multisystem involvement typical of rheumatic diseases, is a highly complex cognitive process that requires synthesis of many data points, typically beyond a simple algorithm-based set of criteria.

Summary table: Key points of the manuscript

Background Question	Should ACR support the development and endorsement of diagnostic criteria?	
Objectives	1. To compare diagnostic and classification criteria 2. To clarify the ACR's position on ACR endorsement of diagnostic criteria	
Methods	Systematic review of literature, recommendation of ACR Classification and Response Criteria Subcommittee, and subsequent ACR Quality of Care Committee and ACR Board of Directors approval	
Results	Classification Criteria	Diagnostic Criteria
	Classification criteria are standardized definitions that are primarily intended to enable clinical studies to have uniform cohorts for research	Diagnostic criteria are a set of signs, symptoms, and tests developed for use in routine clinical care to guide the care of individual patients.
	Need to define (relatively) homogenous group that can be compared across studies and geographic regions	Need to be broad and must reflect the all possible different features and severity of a disease (heterogeneity).
	Very high specificity is required, even if some loss in sensitivity	Both specificity and sensitivity need to be very high, approaching 100%, which is difficult to achieve
	Single universal classification criteria can be applied to different geographical regions, race and ethnicities	Single universal diagnostic criteria cannot be used for making diagnosis due to different disease prevalence in different geographic areas, race and ethnicities
	Classification criteria are possible for disease with and without true "gold standards" (e.g., MSU crystals in gout)	Diagnostic criteria are possible for disease with a true "gold standard" like MSU crystals in gout. For such diseases, the classification and diagnostic criteria can be very similar. For diseases without gold

		standard, development of diagnostic criteria poses significant challenges
	Differences in resources and feasibility has limited effect on classification criteria	Differences in resources and feasibility significantly effect development of diagnostic criteria
	Classification criteria are for research, and therefore, should have no or little impact on billing and reimbursement	Diagnostic criteria are for diagnosis, and therefore, have implications for billing and reimbursement
	Classification criteria have no treatment implications for patients	Diagnostic criteria have treatment implications for patients
	Health care priorities of different geographical areas do not influence classification criteria	Health care priorities of different geographic areas may influence diagnostic criteria
	Classification of a disease can be accomplished by a set of criteria with reasonable sensitivity and specificity	Diagnosis is a complex multi-step process by a physician, which is difficult to accomplish with a single set of criteria
ACR recommendation	Given the difficulty in establishing a uniform diagnostic criteria as noted above, the ACR will only provide approval for classification criteria and will no longer consider funding or endorsement of diagnostic criteria.	

References:

1. Neogi T, Aletaha D, Silman AJ, Naden RL, Felson DT, Aggarwal R, et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. *Arthritis and rheumatism*. 2010;62(9):2582-91.
2. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis and rheumatism*. 2013;65(11):2737-47.
3. Dasgupta B, Cimmino MA, Kremers HM, Schmidt WA, Schirmer M, Salvarani C, et al. 2012 Provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. *Arthritis and rheumatism*. 2012;64(4):943-54.
4. Johnson SR, Goek ON, Singh-Grewal D, Vlad SC, Feldman BM, Felson DT, et al. Classification criteria in rheumatic diseases: a review of methodologic properties. *Arthritis and rheumatism*. 2007;57(7):1119-33.
5. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis and rheumatism*. 2010;62(9):2569-81.
6. Singh JA, Solomon DH, Dougados M, Felson D, Hawker G, Katz P, et al. Development of classification and response criteria for rheumatic diseases. *Arthritis and rheumatism*. 2006;55(3):348-52.
7. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. *Arthritis Care Res (Hoboken)*. 2010;62(5):600-10.
8. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis and rheumatism*. 1991;34(10):1218-27.
9. Yazici H. Diagnostic versus classification criteria - a continuum. *Bulletin of the NYU hospital for joint diseases*. 2009;67(2):206-8.
10. Fries JF, Hochberg MC, Medsger TA, Jr., Hunder GG, Bombardier C. Criteria for rheumatic disease. Different types and different functions. The American College of Rheumatology Diagnostic and Therapeutic Criteria Committee. *Arthritis and rheumatism*. 1994;37(4):454-62.
11. Rao JK, Allen NB, Pincus T. Limitations of the 1990 American College of Rheumatology classification criteria in the diagnosis of vasculitis. *Annals of internal medicine*. 1998;129(5):345-52.
12. Sorensen SF, Slot O, Tvede N, Petersen J. A prospective study of vasculitis patients collected in a five year period: evaluation of the Chapel Hill nomenclature. *Annals of the rheumatic diseases*. 2000;59(6):478-82.
13. Patarroyo PA, Restrepo JF, Rojas SA, Rondon F, Matteson EL, Iglesias-Gamarra A. Are classification criteria for vasculitis useful in clinical practice? Observations and lessons from Colombia. *J Autoimmune Dis*. 2009;6:1.
14. Fries JF, Siegel RC. Testing the 'preliminary criteria for classification of SLE'. *Annals of the rheumatic diseases*. 1973;32(2):171-7.
15. Michet CJ, Jr., McKenna CH, Elveback LR, Kaslow RA, Kurland LT. Epidemiology of systemic lupus erythematosus and other connective tissue diseases in Rochester, Minnesota, 1950 through 1979. *Mayo Clinic proceedings Mayo Clinic*. 1985;60(2):105-13.
16. Peat G, Greig J, Wood L, Wilkie R, Thomas E, Croft P. Diagnostic discordance: we cannot agree when to call knee pain 'osteoarthritis'. *Fam Pract*. 2005;22(1):96-102.

17. Ziswiler HR, Urech R, Balmer J, Ostensen M, Mierau R, Villiger PM. Clinical diagnosis compared to classification criteria in a cohort of 54 patients with systemic sclerosis and associated disorders. *Swiss Med Wkly.* 2007;137(41-42):586-90.
18. Maricq HR, LeRoy EC, D'Angelo WA, Medsger TA, Jr., Rodnan GP, Sharp GC, et al. Diagnostic potential of in vivo capillary microscopy in scleroderma and related disorders. *Arthritis and rheumatism.* 1980;23(2):183-9.
19. Behrens EM, Beukelman T, Gallo L, Spangler J, Rosenkranz M, Arkachaisri T, et al. Evaluation of the presentation of systemic onset juvenile rheumatoid arthritis: data from the Pennsylvania Systemic Onset Juvenile Arthritis Registry (PASOJAR). *The Journal of rheumatology.* 2008;35(2):343-8.
20. Craven A, Robson J, Ponte C, Grayson PC, Suppiah R, Judge A, et al. ACR/EULAR-endorsed study to develop Diagnostic and Classification Criteria for Vasculitis (DCVAS). *Clinical and experimental nephrology.* 2013;17(5):619-21.
21. Dalbeth N, Fransen J, Jansen TL, Neogi T, Schumacher HR, Taylor WJ. New classification criteria for gout: a framework for progress. *Rheumatology.* 2013;52(10):1748-53.
22. Hunder GG. The use and misuse of classification and diagnostic criteria for complex diseases. *Annals of internal medicine.* 1998;129(5):417-8.
23. Johnson SR, Fransen J, Khanna D, Baron M, van den Hoogen F, Medsger TA, Jr., et al. Validation of potential classification criteria for systemic sclerosis. *Arthritis care & research.* 2012;64(3):358-67.
24. Gomariz EM, del M, Guijo VP, Contreras AE, Villanueva M, Estevez EC. The potential of ESSG spondyloarthropathy classification criteria as a diagnostic aid in rheumatological practice. *The Journal of rheumatology.* 2002;29(2):326-30.
25. Murali NS, George R, John GT, Chandi SM, Jacob M, Jeyaseelan L, et al. Problems of classification of Henoch Schonlein purpura: an Indian perspective. *Clin Exp Dermatol.* 2002;27(4):260-3.
26. Perez-Gutthann S, Petri M, Hochberg MC. Comparison of different methods of classifying patients with systemic lupus erythematosus. *The Journal of rheumatology.* 1991;18(8):1176-9.
27. Panush RS, Schur PH. It is lupus? *Bulletin on the rheumatic diseases.* 1997;46(6):3-8.
28. Janssens HJ, Janssen M, van de Lisdonk EH, Fransen J, van Riel PL, van Weel C. Limited validity of the American College of Rheumatology criteria for classifying patients with gout in primary care. *Annals of the rheumatic diseases.* 2010;69(6):1255-6.
29. Malik A, Schumacher HR, Dinnella JE, Clayburne GM. Clinical diagnostic criteria for gout: comparison with the gold standard of synovial fluid crystal analysis. *Journal of clinical rheumatology : practical reports on rheumatic & musculoskeletal diseases.* 2009;15(1):22-4.
30. Taylor WJ, Fransen J, Dalbeth N, Neogi T, Schumacher HR, Brown M, et al. Performance of classification criteria for gout in early and established disease. *Annals of the rheumatic diseases.* 2014.
31. van der Horst-Bruinsma IE, Speyer I, Visser H, Breedveld FC, Hazes JM. Diagnosis and course of early-onset arthritis: results of a special early arthritis clinic compared to routine patient care. *British journal of rheumatology.* 1998;37(10):1084-8.
32. Harrison BJ, Silman AJ, Barrett EM, Scott DG, Symmons DP. Presence of psoriasis does not influence the presentation or short-term outcome of patients with early inflammatory polyarthritis. *The Journal of rheumatology.* 1997;24(9):1744-9.
33. Berthelot JM, Klarlund M, McGonagle D, Bernelot-Moens HJ, Calin A, Harrison B, et al. Lessons from an international survey of paper cases of 10 real patients from an early arthritis clinic. CRI (Club Rhumatismes et Inflammation) Group. *The Journal of rheumatology.* 2001;28(5):975-81.

Figure 1. Methods used for comprehensive literature search for articles relevant to classification and/or diagnostic criteria.

