Best Practice & Research Clinical Rheumatology 28 (2014) 921–934

Contents lists available at ScienceDirect

Best Practice & Research Clinical Rheumatology

journal homepage: www.elsevierhealth.com/berh

7

# The use and abuse of diagnostic/classification criteria

Rayford R. June [a, 1], Rohit Aggarwal [b, *]

[a] Department of Medicine, Division of Rheumatology, Penn State College of Medicine, 500 University Drive, P.O. Box 850, Hershey, PA 17033, USA
[b] Department of Medicine, Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA, USA

*Keywords:*
Diagnostic criteria
Classification criteria
Rheumatology

A B S T R A C T

In rheumatic diseases, classification criteria have been developed to identify well-defined homogenous cohorts for clinical research. Although they are commonly used in clinical practice, their use may not be appropriate for routine diagnostic clinical care. Classification criteria are being revised with improved methodology and further understanding of disease pathophysiology, but they still may not encompass all unique clinical situations to be applied for diagnosis of heterogenous, rare, evolving rheumatic diseases. Diagnostic criteria development is challenging primarily due to difficulty for universal application given significant differences in the prevalence of rheumatic diseases based on geographical area and clinic settings. Despite these shortcomings, the clinician can still use classification criteria for understanding the disease as well as a guide for diagnosis with a few caveats. We present the limits of current classification criteria, their use and abuse in clinical practice, and how they should be used with caution when applied in clinics.



* Corresponding author. Division of Rheumatology, University of Pittsburgh Medical Center, Biomedical Science Tower S 725, 3500 Terrace Street, Pittsburgh, PA 15217, USA. Tel.: +1 412 383 8100; fax: +1 412 383 8864.
*E-mail addresses:* rjune@hmc.psu.edu (R.R. June), aggarwalr@upmc.edu (R. Aggarwal).
[1] Tel.: +1 717 531 5384; fax: +1 717 531 8274.

http://dx.doi.org/10.1016/j.berh.2015.04.004
1521-6942/

## Introduction

Rheumatology is not a field of black and white, but a specialty full of gray. Multisystem clinical syndromes and diseases in rheumatology attract clinicians and researchers who seek to unify different shades of "gray" into a single diagnosis or classification criteria. While understanding of the pathophysiology in each disease has advanced, single laboratory tests with high sensitivity and specificity sufficient to make a diagnosis still do not exist for most of the rheumatic diseases. As opposed to a positive blood culture in infectious disease suggestive of bacteremia or a fasting blood glucose in endocrinology suggestive of diabetes mellitus, even the most common and well-studied clinical conditions in rheumatology such as rheumatoid arthritis (RA) can have significant diagnostic uncertainty of so-called seronegativity up to 30% of the time [1]. Despite making significant technological advances with diagnostic tests such as anti-cyclic citrullinated peptides (CCPs), diagnosis is still imperfect given the lack of 100% specificity for RA and, even worse, sensitivity [2]. This diagnostic uncertainty has led to the development of multiple sets of disease classification criteria for use in research on disease characterization, epidemiology, prognosis, and design of clinical trials for therapeutic investigation [3]. Although designed for clinical research, classification criteria are used and abused in clinical practice for patient care. This article will help define both classification and diagnostic criteria, and describe limitations of current classification criteria and how their use in clinical practice, while not sufficient alone for diagnosis, can be an aid or aide-mémoire in making a diagnosis.

## Statistical principles

Prior to further discussion of classification and diagnostic criteria, a review of certain statistical principles is necessary to clarify differences between classification and diagnostic criteria. *Sensitivity* is the percentage of true positives with the disease. A highly sensitive test is useful for ruling out a disease with a negative test but not necessarily ruling in the disease. Conversely, *specificity* is the percentage of true negatives without disease, and it is useful for ruling in a positive test (if high specificity) but not necessarily ruling out a disease. In the setting of a highly sensitive and specific test, whereas sensitivity is easily understood (if you do not have the test positive, then the disease is not present), specificity leads to confusion because, rather than the focus being on having the disease, the focus is on not having the disease [4]. Highly specific tests have low false-positive rates, and highly sensitive tests have low false-negative rates. For instance, anti-CCP antibodies have been shown to have a high, >90%, specificity for RA in established RA cohorts, whereas it has a moderate sensitivity of 66% [5]. For knowing the true clinical applicability of sensitivity and specificity for a given test, the population in which it is studied or developed is important. For example, *CCP is useful for ruling in RA in subjects with polyarthritis secondary to its high specificity in this particular population* [6,7]. Without knowing the population in which CCP specificity is attributed to, the meaning of the specificity is lost. For example, CCP is positive in many types of non-inflammatory arthritis including infections [2]. Therefore, the *sensitivity* and *specificity* of any diagnostic or classification criteria are dependent on the reference gold standard used for its development as well as target population it is intended for. For example, the 2010 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) RA classification criteria were developed for use on early RA cohorts and therefore not intended to be used on burned-out deforming nodular RA.

Sensitivity and specificity are on a continuum with an inverse relationship where perfect sensitivity (close to 100%) will lead to loss in specificity and vice versa. This is more evident in rheumatology where the sensitivity and specificity of any criteria depend on multiple disease variables [4]. When one gold-standard test is used for diagnosis, as in acute gout or septic arthritis [8], both sensitivity and specificity can remain high. However, as the number of variables needed for a disease classification increase, that is, elevated C-reactive protein, number of swollen joints, and seropositivity, the specificity in classification criteria increases, but sensitivity decreases, and vice versa. The receiver operator curve (ROC) is the statistical and graphical description of this process showing the equilibrium between sensitivity and specificity [9]. This same continuum is found when describing the sensitivity and specificity of any classification and/or diagnostic criteria [4].