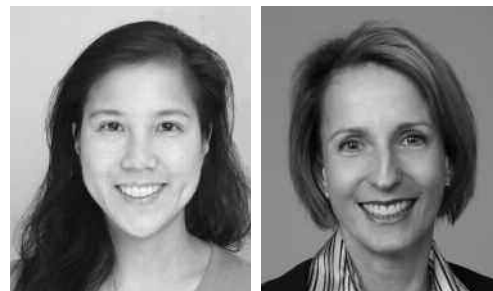# All That Glitters Is Not Gold — Standardizing Diagnosis in Rheumatoid Arthritis Studies

The newly minted 2010 rheumatoid arthritis (RA) classification criteria[1] were developed using a combination of rigorous, well established and newer state-of-the art methodology. This international effort led by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) resulted in a set of criteria that set the standard for criteria development. Now that the criteria have been published, it is time to see how well they perform in independent cohorts.

Varache and colleagues in this issue of *The Journal* present their study on the criteria's diagnostic accuracy in a 2 year cohort[2]. At the leading edge of many important studies assessing the diagnostic validity of the new RA criteria, patients with undifferentiated arthritis < 1 year were followed forward 2 years for their eventual diagnosis. The gold standard definition of RA in the study was "diagnosis of RA by an office-based rheumatologist after 2 years combined with disease modifying antirheumatic drug (DMARD) or glucocorticoid treatment." The authors compared the performance characteristics of 5 criteria sets including the 1987 ACR criteria[3] and the 2010 ACR/EULAR criteria[1] against their gold standard definition of RA. They concluded that there was no significant difference between the diagnostic accuracy of the 2010 ACR/EULAR criteria and those classified as RA by the 1987 ACR criteria.

A particularly interesting characteristic of this study is that the cohort was examined at a time prior to the development of biologic agents (1995–1997), such as tumor necrosis factor (TNF) inhibitors, in contrast to the cohorts used for criteria development (post-2000). Therefore the course of the disease in the subsequent 2 years may be different from the cohorts used for the initial development and validation. This study validated the 2010 criteria in a truly independent cohort derived during a different era in terms of RA treatment. Even with these differences the study suggests that the new criteria had good generalizability, classifying RA as well as the existing 1987 ACR criteria. However, some may wonder why the criteria did not perform better. Part of the answer may lie in how the gold standard was defined.

The gold standard used for validation of the criteria sets by Varache and colleagues limits the interpretability of the study. Their choice of gold standard, a diagnosis of RA by a rheumatologist and DMARD or glucocorticoid treatment, differs from the gold standards used to develop the 1987 and 2010 criteria. Deciding upon a gold standard for RA ascertainment was one of the key challenges during the development of the 2010 ACR/EULAR criteria. To understand why this is the case, it is important to understand the history of the RA criteria sets preceding the 2010 RA Criteria.

The first widely used criteria for RA were developed in 1956 by the American Rheumatism Association (now ACR). A subcommittee of 5 rheumatologists were tasked to create diagnostic criteria that would allow for studies to relate "prevalence, incidence, manifestations, course, treatment or other features of the disease."[4] The selection of criteria to diagnose RA was formed by consensus as to what constituted "manifestations of RA which might have sufficient diagnostic value to be worthy of consideration." The criteria defined 3 categories, definite RA, probable RA, possible RA, and those who did not have RA. To improve sensitivity and specificity of the criteria, the same committee made revisions in 1958 by adding an additional category of "classical" rheumatoid arthritis[5]. These criteria were widely applied for the next 29 years until the publication of the 1987 ARA (now ACR) revised criteria[3].

The 1987 ACR criteria had 2 features that had to be addressed during the development of the 2010 RA Criteria. The first was that the gold standard for development of the 1987 criteria was a "clinical diagnosis of RA without regard to the presence or absence of specific criteria." A gold standard based on physician diagnosis to create a classification criteria inevitably results in some degree of circularity. Rheumatologists have been using the 1958 criteria to diagnose RA for almost 3 decades and therefore the previous criteria would almost certainly influence which clinical characteristics were important in deciding who was considered to have RA. The 1987 criteria were considered to be a simplified

See Diagnostic accuracy of ACR/EULAR 2010 criteria for RA, *page 1250*.

and updated subset of the more comprehensive criteria used in 1958. The second feature was that the 1987 ACR criteria were no longer meant for diagnosis and were designed to be used for research purposes only for selection into clinical studies.

Twenty-three years later, in developing the 2010 ACR/EULAR classification criteria, a committee of rheumatologists was again tasked to develop first and foremost classification criteria for research purposes that could also be used for diagnosis. The issue of a gold standard for RA arose again. Now, almost all rheumatologists would, purposefully or not, refer to the 1958 or 1987 RA criteria when deciding upon whether a patient has RA. Therefore, expert opinion on what constituted RA would mirror the components of these criteria. The circularity of the situation led to the search for a new gold standard during development of the 2010 criteria. In answer to this dilemma, the committee decided on a surrogate marker of the disease RA, the decision to begin treatment with methotrexate, as the new gold standard[6].

Now that an alternate gold standard has been agreed upon, studies validating the 2010 ACR/EULAR criteria should include methotrexate treatment for inflammatory arthritis as one of the comparison groups to report the sensitivity, specificity, and positive predictive value. Although Varache and colleagues diligently compared several criteria sets from the literature to their cohort, their standard for comparison was a hybrid of gold standards used in previous studies. Thus, it would have been informative to see how the 2010 criteria performed as compared to the 2 standards separately — (1) methotrexate use (gold standard in the 2010 criteria) and (2) rheumatologist diagnosis (gold standard in the 1987 criteria). Further comparison using the same gold standards for the 1987 criteria would provide a more direct comparison with the 2010 criteria. In addition, comparing and contrasting the performance of different criteria sets against rheumatologist diagnosis and methotrexate use separately can provide information on how differently patients would be classified using one standard versus the other.

The field of RA classification is strewn with different gold standards. It is important not only in development, but also in validation of criteria to carefully select the gold standard to allow for meaningful comparisons. Currently, we suggest at least 2 standards to validate or compare criteria sets: (1) methotrexate use for inflammatory arthritis and (2) rheumatologist diagnosis of RA. Validation studies, such as the report from Varache and colleagues, provide important information when new classification criteria are required, which, if history repeats itself, may be in the next 20 years.

KATHERINE P. LIAO, MD, MPH,
Division of Rheumatology,
Immunology and Allergy,
Brigham and Women's Hospital;
VIVIAN BYKERK, MD, FRCP,
Brigham and Women's Hospital,
Boston, MA, USA.

Address correspondence to Dr. Bykerk; E-mail: vbykerk@partners.org

REFERENCES

1. 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 Rheum 2010;62:2569-81.
2. Varache S, Cornec D, Morvan J, Devauchelle-Pensec V, Berthelot J-M, Le Henaff-Bourhis C, et al. Diagnostic accuracy of ACR/EULAR 2010 criteria for rheumatoid arthritis in a 2-year cohort. J Rheumatol 2011;38:1250-7.
3. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
4. Bennett GA, Cobb S, Jacox R, Jessar RA, Ropes MW. Proposed diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis 1956;7:121-4.
5. Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA. 1958 Revision of diagnostic criteria for rheumatoid arthritis. Bull Rheum Dis 1958;9:175-6.
6. Funovits J, Aletaha D, Bykerk V, Combe B, Dougados M, Emery P, et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: methodological report phase I. Ann Rheum Dis 2010;69:1589-95.