PostScript

MATTERS ARISING

If you have a burning desire to respond to a paper published in the *Annals of the Rheumatic Diseases*, why not make use of our "rapid response" option?

Log on to our website (www.annrheumdis.com), find the paper that interests you, and send your response via email by clicking on the "eLetters" option in the box at the top right hand corner.

Providing it isn't libellous or obscene, it will be posted within seven days. You can retrieve it by clicking on "read eLetters" on our homepage.

The editors will decide as before whether also to publish it in a future paper issue.

Classification criteria for Sjögren's syndrome

I read with great interest the editorial by R Manthorpe published in the June issue of the *Annals of the Rheumatic Diseases*.[1] The author reviewed the different criteria that have been proposed for the classification of patients with Sjögren's syndrome (SS) and, in particular, commented on the US-European Consensus Group criteria,[2] which were reported for the first time in this same issue. As the first author of the paper, I would like to discuss certain points and criticisms which he raised about our criteria set. First of all I would like to discuss briefly the meaning of "classification criteria" and the methods by which the European[3 4] and US-European criteria for SS[2] were derived.

Classification criteria are not meant to be used for diagnosis. Diagnosis is an often complex process by which a doctor arrives at the suspicion of a specific disease in a given patient, and then must collect enough clinical data to confirm that suspicion. Classification criteria, on the other hand, represent a tool for research and communication, providing uniform criteria for the scientific community to classify patients with the same disease, select patients for clinical-therapeutic trials, and make the data obtained by different researchers in different series of patients comparable. As any experienced rheumatologist knows, it is not uncommon to diagnose a specific rheumatic disease in a patient who does not meet the classification criteria proposed for that disease.

Given these considerations we can argue that:

- Classification criteria can be used for diagnostic purposes only when they have a sensitivity and specificity of 100%

- Because none of the classification criteria for systemic rheumatic diseases reach this level of sensitivity and specificity, it is evident that some patients with a given disease will fail to be classified as having it, and some normal controls may be erroneously classified as patients with that disorder

- Given the purpose of classification criteria, it is preferable to adopt criteria with a specificity approaching the optimum (100%), which would reduce to a minimum the possibility of including false positive controls, but without an excessive loss of sensitivity, which might result in the exclusion of large numbers of true patients.

The only objective method to derive classification criteria is to evaluate, in a series of patients with a given disease and in normal controls, the sensitivity/specificity ratio of different diagnostic tools for that disease, and then to select the combination of these which shows the highest accuracy in correctly classifying cases. Patients and controls should have been preliminary diagnosed on the basis of a "gold standard". Because for the systemic rheumatic diseases a "gold standard" does not exist, the only standard, which can be adopted, is the clinical diagnosis made by an experienced specialist. This in fact was the procedure adopted by the American College of Rheumatology to define the classification criteria for rheumatoid arthritis[5] and by the European Community Group to define[3] and validate[4] those for SS. This method is naturally far from perfect, because the pre-definition of the groups of true patients and true controls will invariably be influenced by the clinical data which are available at the moment of the preliminary evaluation and selection of cases. The fact that in our study the numbers of patients and controls were quite large and were collected from different centres in different countries nevertheless offers some assurance that any bias in the selection process would have been extremely diluted, and that the entire disease spectrum was covered. Despite these well known limitations, this method remains the only satisfactory one for defining and validating classification criteria.

The only alternative is to establish classification criteria based on the suggestions of a group of experts. However, these criteria would still have to be validated in clinically defined groups of patients and controls in order to determine their sensitivity and specificity.

In any case, once populations of "true patients" and "true controls" have been selected, the definition of a classification criteria set becomes a purely statistical operation—that is, one of choosing a set of diagnostic tests and finding the combination which shows the best sensitivity/specificity ratio.

If these points are kept in mind, most of the criticisms about the US-European classification criteria for SS fall to the ground. The definitions of item III (ocular involvement) and item V (salivary gland involvement) as the presence of one positive test, and that of item IV (histopathology) as the presence of a focus score = 1, are not merely definitions suggested by an expert committee. They were, on the contrary, arrived at after rigorous statistical analysis of a large series of patients and controls, and by testing the sensitivity/specificity ratio of all the possible items and combinations thereof. Moreover, the application of a purely statistical procedure guarantees that completely interdependent variables were excluded by the procedure itself. There are many data indicating that autoantibody production and lymphocyte infiltration in the minor salivary glands are related,[6] but statistically speaking the inclusion of both items in the classification criteria improved the performance of the whole set, with respect to their mutual exclusion.

The inclusion of symptoms (items I and II) allows the researcher to start with a simple questionnaire in selecting potential patients with SS, a point which is of great interest for epidemiological surveys. On the other hand, I would entirely agree that a limited number of patients with SS deny having any symptoms. To avoid the misclassification of these non-symptomatic patients, the US-European Consensus Group tested and added an additional criterion for primary SS—namely, three positive results out of the four objective items.

A rigorous statistical method was also followed to define the sequence of items in classification tree procedure. I agree that to perform the autoantibody determination (item VI) before lip biopsy (item IV) appears more logical from the clinical point of view and more acceptable for the patient. However, this was not suggested by the statistical results in order to obtain the best performance of the procedure as whole.

Keeping in mind the statistically derived European classification criteria and using the European database for new statistical analysis, the US-European Consensus Group decided to introduce some modifications in the criteria set. These modifications were particularly designed to (*a*) more precisely define the individual criteria items; (*b*) revise the list of exclusion criteria for primary SS; and (*c*) attempt to improve the specificity of the criteria.

Manthorpe's conclusion that the US-European classification criteria can only correctly classify a subgroup of patients with SS is not confirmed by the results of our statistical analysis. By testing previously proposed criteria for SS in our European populations of patients and controls,[4] we showed that the accuracy of the European classification criteria was significantly higher than all of the others. The accuracy of the Copenhagen criteria,[7] for example, was found to be 85.6% compared with 96.9% for the European set (p<0.005), a difference that can be ascribed to the low sensitivity (71.4%) rather than the quite good specificity (93.5%) of the former. This means that more than a quarter of patients with clinically defined SS cannot be correctly classified using the Copenhagen criteria. It is worth noting that the modifications introduced in the proposed US-European version of the criteria do not absolutely reduce their accuracy with respect to that of the European criteria[2]; the analysis of the receiver operating characteristic curve showed a slight increase in specificity with a corresponding loss of sensitivity. This in fact represents an improvement when the purpose of classification criteria is considered.

C Vitali
Department of Internal Medicine and Section of Rheumatology, "Villamarina" Hospital, Piombino, Italy

Correspondence to: Dr C Vitali, Department of Internal Medicine and Section of Rheumatology, "Villamarina" Hospital, via Forlanini, 57025 Piombino, Italy; hyqprgbv@tin.it

References

1 **Manthorpe R**. Sjögren's syndrome criteria. American-European and Japanese groups' criteria compared and contrasted. Ann Rheum Dis 2002;61:482–4.
2 **Vitali C**, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, *et al*. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis 2002;61:554–8.
3 **Vitali C**, Bombardieri S, Moutsopoulos HM, Balestrieri G, Bencivelli W, Bernstein RM, *et al*. Preliminary classification criteria for Sjögren's syndrome. Results of a prospective concerted action supported by the European Community. Arthritis Rheum 1993;36:340–7.
4 **Vitali C**, Bombardieri S, Moutsopoulos HM, Coll J, Gerli R, Hatron PY, *et al*. Assessment of the European classification criteria for Sjögren's syndrome in a series of clinically defined cases. Results of a prospective multicentre study. Ann Rheum Dis 1996;55:116–21.
5 **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.
6 **Gerli R**, Muscat C, Giansanti M, Danieli MG, Sciuto M, Gabrielli A, *et al*. Quantitative assessment of salivary gland inflammatory infiltration in primary Sjögren's syndrome: its relationship to different demographic, clinical and serological features of the disorder. Br J Rheumatol 1997;36:969–75.
7 **Manthorpe R**, Oxholm P, Prause JU, Schiödt M. The Copenhagen criteria for Sjögren's syndrome. Scand J Rheumatol Suppl 1986;61:19–21.

Author's response

Claudio Vitali has provided a valuable historical background to his comments on the European Sjögren's syndrome (SS) criteria. To this part there is only a little to add. Although classification criteria are not meant to be used for diagnosis, discussions and talks with colleagues at various congresses world wide have shown that this is unfortunately not so in practice, neither in scientific trials nor in publications. Under ideal circumstances there should be only small differences—if any—between classification criteria and diagnostic criteria. It is not an easy task to tell a patient that she has SS when participating in this scientific project but otherwise she is not considered to have SS—or vice versa. It was with the same arguments that most Europeans discarded the terminology probable/definite SS. Progress within clinical science—including SS—is usually a continuous process but occasionally bigger strides are made. When leading SS scientists from America (US) and Europe (Eur) formed a consensus group to propose a new set of criteria for SS it was expected that they would include the latest news within the SS area—or that at least the news would be discussed and commented upon. The consensus group failed in this important aspect and this was the reason, therefore, why my leader had the subtitle: "American-European and Japanese Groups' criteria compared and contrasted", especially as the Japanese SS researchers came up with rather different results, which were based upon data from more patients/cases.[1] From a clinical point of view the Japanese III criteria are of great importance and seemingly more relevant.

One important factor of the Japanese III criteria is that they do not operate with or include subjective symptoms because their statistical calculations showed that it did not improve their results.[1] This is in contrast with the US-Eur consensus group which continues to include ocular (item I) and oral (item III) dry symptoms—unchanged from 1993. Research has shown that even though the cornea is the most densely innervated organ, there are no nerves which can register dryness. To include dry eyes in the criteria is therefore inappropriate. (Dry eyes is an iatrogenic expression which some patients are very quick to adopt.)

Another important contrast between the US-Eur consensus group and the Japanese expert group is that the latter requires at least two abnormal ocular tests for the function of the lachrymal gland to confirm the diagnosis keratoconjunctivitis sicca and two abnormal oral tests for the function of the salivary glands to confirm the diagnosis stomatitis sicca.[1] However, sialography can stand alone.[1] The US-Eur group requires only one test, which in practice is an abnormal Schirmer-I test ($\leqslant 5$ mm in 5 minutes performed without anaesthesia and with closed eyes) for the lachrymal gland and an abnormal unstimulated whole sialometry ($\leqslant 1.5$ ml in 15 minutes performed without tobacco, eating, and drinking during the preceding two hours) for the salivary gland. (It is usually customary to get a "confirmatory" test result when the findings are abnormal, as in HIV.) In the leader I expressed concern about the proposal that three positive results out of four objective items in an asymptomatic patient should automatically be called primary SS. If the abnormal items are IV, V, and VI, there is no proof that the lachrymal gland is also affected.

Probably the greatest "negative" scientific point of discussion was the lack of comments on the observation previously published in this journal that the number of cigarettes smoked per week may have a tremendous effect on the result of the focus score in lower lip biopsy (item IV) as well as on the level of anti-SSA/B autoantibodies (item VI).[2] In historical non-smokers the results in item IV and VI were statistically significant compared with those found in present and/or past smokers.[2] In the last group it did not matter if the date at which they stopped was recent or several years (decades) previously. The smoking effect was highly dose dependent, with the threshold around 21 cigarettes a week.[2] Consider the number of people with irritation of the eyes and dryness in the mouth who are/have been smokers and thus might not fulfil item IV and item VI of the US-Eur consensus group. If they nevertheless have at least two abnormal functional test results from both the lachrymal and the salivary glands, I find it today medically, ethically, and morally wrong not to accept that these patients have primary SS. Research in the autoimmunity/tobacco area seems only to be in its infancy.[3]

In conclusion, I cannot advise colleagues to start using the US-Eur consensus group criteria for SS uncritically. A big step towards obtaining longlasting international SS criteria was taken at the VIII International SS Symposium in Kanasawa, Japan 2002, when great acclamation was given to a proposal to form a big international SS consensus group. It is to be hoped, that this group of SS researchers from Japan, China, Europe, and America will some day, and the sooner the better, deliver their view(s)—unless we could have an earlier 100% diagnostic test in our hands—valid for smokers, ex-smokers, and "never" smokers.

R Manthorpe
Sjögren's Syndrome Research Centre, Department of Rheumatology, Malmö University Hospital, SE-205 02 Malmö, Sweden

Correspondence to: Dr R Manthorpe; Manthorpe@inet.uni2.dk

References

1 **Fujibayashi T**, Sugai S, Miyasaki N, Tojo T, Miyawaki S, Ichikawa Y, *et al*. Criteria for the diagnosis of Sjögren's syndrome (Japanese criteria III). Supported by the Ministry of Health and Welfare of Japan. *1999 Annual reports of the research group of autoimmune disease* 1999;35–8.
2 **Manthorpe R**, Benoni C, Jacobsson L, Kirtava Z, Larsson Å, Liedholm R, *et al*. Lower frequency of focal lip sialadenitis (focus score) in smoking patients. Can tobacco diminish the salivary gland involvement as judged by histological examination and anti-SSA/Ro and anti-SSB/La antibodies in Sjögren's syndrome? Ann Rheum Dis 2000;59:54–60.
3 **Sopori M**. Effects of cigarette smoke on the immune system. Nat Rev Immunol 2002;2:372–7.

Etanercept treatment of rheumatoid arthritis in the "real world"

A recent paper in the *Annals of the Rheumatic Diseases* attempted to examine the "real world" experience of etanercept treatment by examining the incidence of flares of rheumatoid arthritis (RA) disease activity in a cohort of patients with RA who had started treatment with etanercept before September 1999.[1] The number of flares and patients experiencing flares within the first year of etanercept treatment was compared with that seen in the same cohort of patients the year before they started etanercept. It is well documented that the withdrawal rate from disease modifying antirheumatic drug (DMARD) treatment in RA increases with the length of time the patient has been receiving the drug and that a number of these withdrawals relate to loss of efficacy.[2–4] Therefore, it is reasonable to assume that the number of disease flares will increase the longer a patient with RA is receiving treatment, particularly if that treatment is failing to control the disease activity. The fact that treatment of this cohort of patients with RA was changed to etanercept suggests that their current DMARD treatment was failing to control their disease. Therefore it is likely that there would be an increased number of disease flares in this group before starting etanercept treatment. Although not stated in the paper, it is reasonable to assume that this cohort of patients with RA had been receiving their previous DMARD treatment for some time before changing to etanercept treatment. Therefore what the authors of this paper have compared is the number of RA disease flares in a cohort of patients with RA in their first year of etanercept treatment with the number of RA flares in their last year of (failing) DMARD treatment. The results are predictably in favour of the new treatment.

Would the authors have found the same results in favour of etanercept treatment if they had conducted a "fairer" comparison and compared the number of flares in this RA cohort during their first year of the previous DMARD treatment, especially if it was methotrexate, with the first year of etanercept treatment?