

Minor salivary gland biopsy and Sjögren's syndrome: comparative analysis of biopsies among different Italian rheumatologic centres

A. Tavoni¹, C. Baldini², W. Bencivelli¹, L. Cavazzini³, M. Covelli⁴, S. De Vita⁵, R. Caporali⁶, R. Gerli⁷, M. Giansanti⁸, M. Govoni⁹, V. Donati¹¹, C.M. Montecucco⁶, P. Morbini¹², F.P. Pilato¹³, L. Resta¹⁴, C.A. Scott¹⁵, M. Mosca², C. Vitali¹⁶, S. Bombardieri²
From the Italian Study Group on Sjögren's Syndrome of the Italian Society of Rheumatology.

¹Clinical Immunology Unit, Department of Internal Medicine, and ²Rheumatology Unit, Dept. of Internal Medicine, University of Pisa, Pisa, Italy; ³Section of Anatomy Histology and Pathological Cytology, University of Ferrara, Italy; ⁴Rheumatology Section, University of Bari, Italy; ⁵Rheumatology Clinic, DPMSC, University of Udine, Udine, Italy; ⁶Dept. of Rheumatology, University of Pavia, Pavia, Italy; ⁷Rheumatology Unit, Dept. of Clinical and Experimental Medicine, University of Perugia; ⁸Dept. of Pathology, General Hospital, Perugia; ⁹Rheumatology Unit, Dept. of Clinical and Experimental Medicine, University of Ferrara, Ferrara, Italy; ¹⁰Osteo-Articular Department, Rheumatic Disease and Internal Medicine Unit, Hospital and University of Parma, Italy; ¹¹Unit of Pathological Anatomy (II), Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy; ¹²Dept. of Pathology, University of Pavia, Pavia, Italy; ¹³Dept. of Pathology and Laboratory Medicine, University of Parma, Parma, Italy; ¹⁴Dept. of Pathology, University Medical School of Bari, Bari, Italy; ¹⁵Institute of Pathology, DRMM, University of Udine, Udine, Italy; ¹⁶Dept. of Internal Medicine and Section of Rheumatology, Villamarina Hospital, Piombino, Italy.

Abstract Objective

The minor salivary gland biopsy (MSGB) is widely considered an important component of the diagnostic algorithm of primary Sjögren's syndrome (pSS) and is mentioned in all the classification criteria sets for the disease. The aim of this study, coordinated by the Italian Society of Rheumatology, was to verify the inter-observer agreement on the evaluation of MSGB among different experienced Italian rheumatologic centres, in order to better standardise the diagnostic methodology.

Methods

Seven centres participated in the study, providing a total of 50 MSGB samples. Each center blindly classified all the samples according to the Chisholm and Mason (CM) grading. The results were collected and analysed.

Results

The inter-observer agreement was satisfactory when the samples were stratified as consistent and non-consistent with the final diagnosis of pSS (median $\kappa = 0.75$; mean $\kappa = 0.70$). Nonetheless, significant discrepancies in the histopathologic evaluation of MSGB emerged when the agreement was assessed on the single scores. Considering the modal CM grading for each sample as the correct grading, upon re-examination, a potential bias in the final clinical diagnosis was detected in 7 out of 50 samples.

Conclusion

This study has shown significant discrepancies in the evaluation of MSGB among different rheumatologic centres in the same country. Greater standardisation of the procedure is clearly necessary, both to improve the diagnostic performance and scientific communication.

Key words

Sjögren's syndrome, minor salivary gland biopsy, sicca syndrome

Antonio Tavoni, MD,
Chiara Baldini, MD,
Walter Bencivelli, PhD, Assoc. Professor
Luigi Cavazzini, MD, Professor
Michele Covelli, MD
Salvatore De Vita, MD, Assoc. Professor
Roberto Caporali, MD
Roberto Gerli, MD, Professor
Michele Giansanti, MD
Marcello Govoni, MD, Assoc. Professor
Valentina Donati, MD
Carlo Maurizio Montecucco, Professor
Patrizia Morbini, MD
Francesco Paolo Pilato, MD
Leonardo Resta, MD
Cathryn Anne Scott, MD
Marta Mosca, MD
Claudio Vitali, MD
Stefano Bombardieri, MD, Professor

Please address correspondence to:

Chiara Baldini, MD,
Dipartimento di Reumatologia,
Dipartimento di Medicina Interna,
Università di Pisa,
Via Roma 67,
56126 Pisa, Italy.

E-mail: chiara.baldini74@gmail.com

Received on September 14, 2012; accepted
in revised form on October 2, 2012.

© Copyright CLINICAL AND
EXPERIMENTAL RHEUMATOLOGY 2012.

Introduction

Primary Sjögren's syndrome (pSS) is a chronic, systemic inflammatory disorder characterised by a progressive hyposalivation of the salivary and lacrimal glands, frequently associated with a variety of scattered extraglandular manifestations, including malignant lymphoproliferative disorders (1-3).

While novel biomarkers are being investigated (4-10) the minor salivary gland biopsy (MSGB) is widely considered a key tool for the diagnosis of primary Sjögren's syndrome (pSS) and represents an integral component of the several diagnostic algorithms proposed in the last 30 years (11-15).

However, a recent systematic review has pointed out a lack of information about the diagnostic value of MSGB with a sensitivity ranging from 63.5% to 93.7% and a specificity ranging from 61.2% to 100% (16). More specifically, it has been shown that, similarly to what happens for serological tests (17), the reproducibility of biopsy grade in pSS is one of the major drawbacks of the test, due to technical factors such as deficiency in the size of the samples, glandular atrophy, number and depths of sections examined (18-25). Finally, very few studies have verified to which extent the MSGB histopathological interpretation itself might influence the reproducibility of the test (26).

This study was, therefore, specifically aimed at verifying the agreement of qualified tertiary Italian rheumatologic centres on the grading of minor salivary gland biopsies collected from histopathological archives in a retrospective study. This "exercise" was performed in order to explore any potential pitfall in the interpretation of the MSGB histopathology in routine clinical practice.

Materials and methods

Seven rheumatology centres in northern, central and southern Italy, participated in the study. Each centre provided 7 MSGB samples (except one which provided 8 samples) which could be consistent or non-consistent with the diagnosis of pSS. All the samples were pooled together and a total of 50 biopsies were collected. MSGBs were performed by specialists in oral medicine

according to the standard procedure described by Daniels TE (27). The area of the gland tissue was measured using a light microscope equipped with an ocular micrometer (at a magnification 100 x). The presence/absence of diffuse/focal chronic inflammation was recorded by the providing centre using a modification of the system originally introduced by Chisholm and Mason (CM) (28, 29). All the samples were pooled together and blindly re-evaluated by each pathologist. The samples were not re-sectioned or modified at any stage. Clinical and serological data of the patients were separately registered, and therefore, all the reviewers proceeded without any knowledge of the patients' clinical data or working clinical diagnosis and without knowing the baseline histologic interpretation. The results were collected and statistically analysed.

Statistical analysis

The biopsy results from each centre were compared to those of the other centres. The assessment of inter-observer agreement was carried out using different statistical tests:

- First, the samples were stratified into positive and negative (according to CM grading) and the agreement among centres was evaluated by the Cohen's kappa coefficient (1 representing full agreement and 0 fully casual agreement) assuming the latter as a statistical measure of inter-rater reliability when observing or coding qualitative / categorical variables. A Cohen's kappa >0.70 was considered satisfactory.
- Then, the samples were assigned an index score based on the concordance among centres in their grading according to the stratified values into positive and negative of the CM. This score ranged from 1 (if only 2/7 centres agreed in the grading) to 21 (if all 7 centres assigned the same grade). An index score ≥10 that was at least the concordance of five out of seven centres with the two other centres discordant when compared with each other, was considered to be a good degree of agreement.
- Thirdly for each sample the absolute value of the difference between the

Competing interests: none declared.

original biopsy grades (CM score 0, 1, 2, 3, 4) assigned by two different centres was calculated in order to quantify the entity of the disagreement among the centres; this operation was repeated for each different pair of centres and the values were summed. A summed value ≤ 12 was assumed to be indicative of an acceptable degree of concordance as it means at least the scoring agreement of four out of seven centres with the three other centres concordant when compared each other and differing of one unit from the former four centres.

- Finally, we evaluated the modal CM grading for each sample, *i.e.* the score assigned by the largest number of centres for a given sample, and assumed it as the correct grading (gold standard). The disagreement between the original interpretation of each sample as given by the providing centre and the modal CM grading was assessed and the misclassification rate was evaluated.

Results

A total of 50 biopsies were collected (specimens' area varying from 6 to 24.57 mm²). According to this baseline evaluation, 64% of the specimens was characterised by a lymphocytic infiltration, while the remaining 36% showed a non-specific chronic sialoadenitis. Clinical and serological data of the patients are summarised in Table I.

Inter-observer agreement evaluated by the Cohen's kappa coefficient

The inter-observer agreement was firstly assessed using the Cohen's kappa coefficient (30, 31). When the MSGB results were stratified as positive (CM grade 3 or 4) or negative (CM grade 0, 1 or 2), the Cohen's kappa coefficient ranged from 0.44 to 0.92 (median 0.75, mean 0.70).

Inter-observer agreement evaluated by the index score based on the concordance among centres in their grading

As previously stated, we generated an index score based on the probability of concordance between centres in their grading (Table II). An index score ≥ 10

Table I. Clinical and serological characteristics of the population.

Number of samples	50
Female : male	49:1
Age	50.10 years (range 24–81)
Onset of dry eyes/dry mouth	31.82 months (range 3–168)
Antinuclear autoantibodies	80%
anti-Ro/SSA autoantibodies	56%
anti-La/SSB autoantibodies	36%
Rheumatoid Factor	54%
Hypergammaglobulinaemia	60%
Cryoglobulins	2%

Antinuclear antibodies were determined by indirect immunofluorescence (cut-off value for positive cases $\geq 1:160$); Antibodies anti-Ro/SSA and anti-La/SSB were determined by controimmunoelectrophoresis; RF was determined by Latex test.

(*i.e.* agreement among 5 out of 7 centres with the 2 other centres discordant among them), was observed in the 48% of the samples. The greatest agreement among the centres in grading was obtained for biopsy samples with a CM grade 4 (11/50), while the worst was obtained for CM grade 2 (*i.e.* abnormal findings but no foci).

Inter-observer agreement evaluated according to the absolute value of the difference between the original CM biopsy grades

To quantify the entity of the disagreement we analysed the absolute value of the difference between the original biopsy grading assigned by two different centres repeated for each different pair of centres and summed. In other terms, for each sample, the absolute value of the difference between the biopsy scores assigned by two different centres was calculated; this operation was repeated for each different pair of centres and the values were summed. A summed value of 0 (*i.e.* complete agreement) was found only in the 11/50 samples with CM grade 4. An acceptable degree of concordance (summed value ≤ 12) was reached in 29/50 (58%) of the samples.

Inter-observer agreement evaluated by the modal CM grading

Finally, we assessed the modal CM grading for each sample (Table III). As expected, the maximum agreement was on modal grade 4; nonetheless, a good agreement was also observed on modal grade 1. To sum up, considering the modal CM grading as the cor-

rect grading, upon re-examination a disagreement in the scoring assessment was observed for 18 out of 50 samples, and for 7 of them this implied clinical consequences on the final diagnosis of pSS. In detail, the original grading of 4 of these biopsies resulted to be a false positive, while in the other 3 cases it was a false negative. The change of the initial diagnosis was correlated to the grading of the sample and the majority of sustained revision was registered for the CM grade 2 and grade 3.

Discussion

To our knowledge this is the first study specifically aimed at quantifying the inter-observer agreement on MSGB grading in routine clinical practice. We found that the agreement was overall acceptable when MSGB were stratified as consistent or not consistent with a diagnosis of SS. Nonetheless, significant discrepancies emerged when the scores were more thoroughly analysed. Independently of the test employed to assess the inter-observer agreement, the maximum level of concordance was assessed only for samples with CM grade 4. Our data are comparable with the existing literature which has emphasised the potential limited reproducibility in the interpretation of samples characterised by moderate non-specific inflammation or by mononuclear inflammatory sialadenitis with only one lymphocytic focus (26). A potential limitation of our study is represented by the fact that MSGB samples were classified according to the older, less specific and reproducible CM grading (28, 29) rather than according to the

Table II. Index score based on the concordance between centers in their CM grading.

Index score	n° sample / 50	(%)	Explanation
21	11/50	22%	Agreement of all the centres
15	2/50	4%	Agreement of 6 out of 7 centres
11	6/50	12%	Agreement of 5 out of 7 centres with the 2 other centres concordant among them
10	5/50	10%	Agreement of 5 out of 7 centres with the 2 other centres discordant among them
9	8/50	16%	Agreement of 4 out of 7 centers with the 3 other centres concordant among them
7	10/50	20%	Agreement of 4 out of 7 centers with the other 3 centres: 2 concordant and 1 discordant among them
6	3/50	6%	Agreement of 4 out of 7 centers with the 3 other centres concordant among them OR 3 centres concordant among them, 3 centres concordant among them, 1 discordant with all the others
5	2/50	4%	Agreement of 3 out of seven centres with the other 4 concordant pair to pair
4	1/50	2%	Agreement of 3 out of seven centres with of the other 4 centres two concordant among them and two discordant among them and with the others
3	2/50	4%	Agreement of three pair of centres among them with the last discordant with all the others

The index score was based on the concordance among centers in their grading according to the stratified values into positive and negative of the CM. An index score ≥ 10 (data reported above the line) was considered to be a good degree of agreement.

Table III. Agreement on the modal CM grading for each sample (gold standard).

n° of centre in agreement on the modal CM grade	n° of samples out of 50	Categories of the gold standard CM grading
Agreement of 7 centres	11/50	4
Agreement of 6 centres	2/50	1.4
Agreement of 5 centres	11/50	0.1.3.4
Agreement of 4 centres	19/50	0.1.2.3.4
Agreement of 3 centres	5/50	1.2.3
Agreement of 2 centres	2/50	0.2

The table shows the agreement among the centres regarding the CM gold standard grading. As expected, the complete agreement among all the centres was on modal grade 4; a good agreement (6 out of 7 centres) was also observed on modal grade 1.

currently recommended focus scoring system (13). Nonetheless, our choice was motivated by the fact that when the study was designed, CM grading was still routinely employed in most of the Italian Units. Moreover, CM grading allowed the oral pathologists to well stratify the biopsies with absent, slight or moderate infiltration. Despite its limitations, this study provides the evidence that the diagnostic power of the MSGB could be much improved by standardised methods of scoring at national and at international levels. It is to be clearly remarked that this work does

not challenge the validity of well established and validated scoring methods, but rather their "tolerance" to the deviations from the main correct protocols that are often inevitable in a busy clinical environment. It would be advisable that scientific societies fostered the collection of reference MSGB specimens. These specimens should be assessed by experienced leading centres and utilised to promote training "quality control exercises in sample interpretation" in order to overcome the potential pitfalls of the MSGB scoring and to improve scientific communication.

Authors' contributions

AT participated in the design of the study and drafted the manuscript; CB participated in the design of the study and drafted the manuscript; WB performed the statistical analysis; LC performed the histopathological analysis; MC, SDV, RC, RG, MG, MM and CMM participated in case collection; MG, VD, PM, FPP, LR and CAS performed the histopathological analysis; CV participated in the design of the study; SB coordinated the study and participated in its design.

Acknowledgements

We wish to thank Wendy Doherty for her valuable contribution in reviewing the text.

References

1. RAMOS-CASALS M, TZIOUFAS AG, FONT J: Primary Sjögren's syndrome: New clinical and therapeutic concepts. *Ann Rheum Dis* 2005; 64: 347-54.
2. ALBRECHT J, ATZENI F, BALDINI C *et al.*: Skin involvement and outcome measures in systemic autoimmune diseases. *Clin Exp Rheumatol* 2006; 24: S52-9.
3. BALDINI C, PEPE P, LUCIANO N *et al.*: A clinical prediction rule for lymphoma development in primary Sjögren's syndrome. *J Rheumatol* 2012; 39: 804-8.
4. TANDON M, GALLO A, JANG SI, ILLEI GG, ALEVIZOS I: Deep sequencing of short rnas reveals novel micrnas in minor salivary glands of patients with Sjögren's syndrome. *Oral Dis* 2012; 18: 127-31.
5. BALDINI C, GIUSTI L, BAZZICHI L, LUCACCHINI A, BOMBARDIERI S: Proteomic analysis of the saliva: A clue for understanding primary from secondary Sjögren's syndrome? *Autoimmun Rev* 2008; 7: 185-91.
6. GIUSTI L, BALDINI C, BAZZICHI L, BOMBARDIERI S, LUCACCHINI A: Proteomic diagnosis of Sjögren's syndrome. *Expert Rev Proteomics* 2007; 4: 757-67.
7. FERRACCIOLI G, DE SANTIS M, PELUSO G *et al.*: Proteomic approaches to Sjögren's syndrome: A clue to interpret the pathophysiology and organ involvement of the disease. *Autoimmun Rev* 2010; 9: 622-6.
8. BALDINI C, GIUSTI L, CIREGIA F *et al.*: Proteomic analysis of saliva: A unique tool to distinguish primary Sjögren's syndrome from secondary Sjögren's syndrome and other sicca syndromes. *Arthritis Res Ther* 2011; 13: R194.
9. BALDINI C, GALLO A, PEREZ P, MOSCA M, ALEVIZOS I, BOMBARDIERI S: Saliva as an ideal milieu for emerging diagnostic approaches in primary Sjögren's syndrome. *Clin Exp Rheumatol* 2012; 30: 785-790.
10. GALLO A, BALDINI C, TEOS L, MOSCA M, BOMBARDIERI S, ALEVIZOS I: Emerging trends in Sjögren's syndrome: Basic and

- translational research. *Clin Exp Rheumatol* 2012; 30: 779-784.
11. VITALI C, BOMBARDIERI S, MOUTSOPOULOS HM *et al.*: Preliminary criteria for the classification of Sjögren's syndrome. Results of a prospective concerted action supported by the European community. *Arthritis Rheum* 1993; 36: 340-7.
12. VITALI C, BOMBARDIERI S, MOUTSOPOULOS HM *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. The European study group on diagnostic criteria for Sjögren's syndrome. *Ann Rheum Dis* 1996; 55: 116-21.
13. VITALI C, BOMBARDIERI S, JONSSON R *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.
14. BALDINI C, TALARICO R, TZIOUFAS AG, BOMBARDIERI S: Classification criteria for Sjögren's syndrome: A critical review. *J Autoimmun* 2012; 39: 9-14.
15. SHIBOSKI SC, SHIBOSKI CH, CRISWELL L *et al.*: American college of rheumatology classification criteria for Sjögren's syndrome: A data-driven, expert consensus approach in the Sjögren's International Collaborative Clinical Alliance Cohort. *Arthritis Care Res (Hoboken)* 2012; 64: 475-87.
16. GUELLEC D, CORNEC D, JOUSSE-JOULIN S *et al.*: Diagnostic value of labial minor salivary gland biopsy for Sjögren's syndrome: A systematic review. *Autoimmun Rev* 2012;
17. ULVESTAD E, KANESTROM A, MADLAND TM, THOMASSEN E, HAGA HJ: Clinical utility of diagnostic tests for rheumatoid factor. *Scand J Rheumatol* 2001; 30: 87-91.
18. STEWART CM, BHATTACHARYA I, BERG K *et al.*: Labial salivary gland biopsies in Sjögren's syndrome: Still the gold standard? *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2008; 106: 392-402.
19. TEPPH H, REVONTA M: A follow-up study of minimally invasive lip biopsy in the diagnosis of Sjögren's syndrome. *Clin Rheumatol* 2007; 26: 1099-103.
20. AL-HASHIMI I, WRIGHT JM, COOLEY CA, NUNN ME: Reproducibility of biopsy grade in Sjögren's syndrome. *J Oral Pathol Med* 2001; 30: 408-12.
21. MORBINI P, MANZO A, CAPORALI R *et al.*: Multilevel examination of minor salivary gland biopsy for Sjögren's syndrome significantly improves diagnostic performance of aecg classification criteria. *Arthritis Res Ther* 2005; 7: R343-8.
22. CAPORALI R, BONACCI E, EPIS O, BOBBIO-PALLAVICINI F, MORBINI P, MONTECUCCO C: Safety and usefulness of minor salivary gland biopsy: Retrospective analysis of 502 procedures performed at a single center. *Arthritis Rheum* 2008; 59: 714-20.
23. OBINATA K, SATO T, OHMORI K, SHINDO M, NAKAMURA M: A comparison of diagnostic tools for Sjögren's syndrome, with emphasis on sialography, histopathology, and ultrasonography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2010; 109: 129-34.
24. NAKAMURA H, KAWAKAMI A, IWAMOTO N *et al.*: A single centre retrospective analysis of aecg classification criteria for primary Sjögren's syndrome based on 112 minor salivary gland biopsies in a Japanese population. *Rheumatology (Oxford)* 2010; 49: 1290-3.
25. YAZISIZ V, AVCI AB, ERBASAN F, KIRIS E, TERZIOGLU E: Diagnostic performance of minor salivary gland biopsy, serological and clinical data in Sjögren's syndrome: A retrospective analysis. *Rheumatol Int* 2009; 29: 403-9.
26. VIVINO FB, GALA I, HERMANN GA: Change in final diagnosis on second evaluation of labial minor salivary gland biopsies. *J Rheumatol* 2002; 29: 938-44.
27. DANIELS TE: Labial salivary gland biopsy in Sjögren's syndrome. Assessment as a diagnostic criterion in 362 suspected cases. *Arthritis Rheum* 1984; 27: 147-56.
28. CHISHOLM DM, MASON DK: Labial salivary gland biopsy in Sjögren's disease. *J Clin Pathol* 1968; 21: 656-60.
29. GREENSPAN JS, DANIELS TE, TALAL N, SYLVESTER RA: The histopathology of Sjögren's syndrome in labial salivary gland biopsies. *Oral Surg Oral Med Oral Pathol* 1974; 37: 217-29.
30. ZANGEN SW, RYU S, ORNOY A: Alterations in the expression of antioxidant genes and the levels of transcription factor nf-kappa b in relation to diabetic embryopathy in the Cohen diabetic rat model. *Birth Defects Res A Clin Mol Teratol* 2006; 76: 107-14.
31. BECKER G: Creating comparability among reliability coefficients: The case of Cronbach alpha and Cohen kappa. *Psychol Rep* 2000; 87: 1171-82.