The use of digital image analysis in the histological assessment of Sjögren's syndrome salivary glands improves inter-rater agreement and facilitates multicentre data harmonisation

D. Lucchesi[1], E. Pontarini[1], V. Donati[2], M. Mandarano[3], A. Sidoni[3], E. Bartoloni[4], C. Baldini[5], A.R. Tappuni[6], M. Bombardieri[1]

[1]Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Queen Mary University of London, UK;
[2]Unit of Pathological Anatomy 2, Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy;
[3]Section of Anatomic Pathology and Histology, Department of Experimental Medicine, Medical School, University of Perugia, Italy;
[4]Rheumatology Unit, Department of Medicine, University of Perugia, Italy;
[5]Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Pisa, Italy;
[6]Centre for Immunobiology and Regenerative Medicine, Institute of Dentistry, Queen Mary University of London, UK.

Davide Lucchesi, PhD*
Elena Pontarini, PhD*
Valentina Donati, MD
Martina Mandarano, MD
Angelo Sidoni, MD
Elena Bartoloni, MD
Chiara Baldini, MD, PhD
Anwar R. Tappuni, BDS, PhD
Michele Bombardieri, MD, PhD

**These authors contributed equally.*

Please address correspondence to:
Davide Lucchesi,
Centre of Experimental Medicine and Rheumatology,
William Harvey Research Institute,
Queen Mary University of London,
Charterhouse Square,
EC1M 6BQ London, UK.
E-mail: d.lucchesi@qmul.ac.uk

Received on July 3, 2020; accepted in revised form on July 29, 2020.

Clin Exp Rheumatol 2020; 38 (Suppl. 126): S180-S188.

Key words: Sjögren's syndrome, salivary glands, histology, computer-assisted image analysis

Competing interests: none declared.

ABSTRACT

Objective. *To assess whether the use of digital image analysis (DIA) in primary Sjögren's syndrome (pSS) for the calculation of the total area of the salivary gland (SG), focus score (FS) and SG area occupied by the inflammatory infiltrate (area fraction, AF), was able to generate reproducible readings among different raters, reducing disagreement.*

Methods. *Haematoxylin and Eosin digital slides from pSS and non-specific chronic sialadenitis (NSCS) patients were analysed blindly by 4 independent raters among 3 centres. Using an open-source software (QuPath) raters were asked to provide the total area of the gland i) using a grid-based method and ii) a software-based area-calculation tool, iii) the number of inflammatory foci and iv) the total area of the inflammatory infiltrate. Collected data was used to calculate the inter-rater agreement.*

Results. *For the calculation of the total SG area, DIA generated higher agreement among raters than grid-based calculation (inter-class correlation coefficient ICC=0.85 vs. 0.98). Agreement for calculated total area of the inflammatory infiltrate (ICC=0.94) and for AF (ICC=0.94) was higher than infiltrates count number (ICC=0.54) and FS (ICC=0.56). AF achieved a 30% improvement over the FS at generating consensus among raters when used as a diagnostic cut-off.*

Conclusion. *A digital approach achieved a far superior inter-rater agreement when calculating the total area compared to a grid-based approach. The calculation of AF proved superior to FS in correctly classifying pSS vs NSCS biopsies. We suggest that digitally calculated AF should be used alongside FS for large multicentre studies to improve data harmonisation.*

Introduction

Primary Sjögren's syndrome (pSS) is a high-prevalence systemic autoimmune disease that targets mainly the exocrine glands (1-3). Despite its diffuse presence worldwide, pSS is poorly diagnosed and its pathogenesis and biology understudied, leading to a lack of therapies capable to modify, ameliorate or slow down the disease (3, 4). One of the key characteristics of the disease is the presence of inflammatory infiltrates in the target organs of the autoimmune response, particularly in the salivary glands (1, 3). An inflammatory infiltrate (or focus) of at least 50 lymphocytes per 4 mm^2 of glandular tissue (assessed via the histological analysis of a biopsy of the minor salivary glands (5)) or the presence in the serum of specific auto-antibodies (*i.e.* anti-SSA) are necessary for a diagnosis of pSS (6).

The salivary glands of pSS patients are not only a diagnostic "window" into the disease but are also able to provide priceless data on the evolution of the pathology, on patients stratification (7-11) and potentially capable to inform on treatment response. In parallel with clinical trials aiming at expanding the potential therapeutic avenues for this disease and involving biopsy-driven proof of mechanisms investigations (3, 12-15), the necessity for an international consensus on shared guidelines for the histological analysis and interpretation of salivary gland tissue has become apparent (16-18). Significant discrepancies have historically been reported not only in the interpretation of the histopathology focus score (FS) (19) which when >1 is a discriminant for a positive histopathology result (20) but also in the evaluation of other key features of salivary gland immunopathology, such as the identification

of ectopic germinal centres, leading to contrasting results (17, 18, 21-23).

The work that we present here, aimed at investigating the inter-observer performance of more detailed and quantitative analysis of the presence of focal immune cells infiltrates, the cornerstone of pSS histological assessment, implementing the guidelines suggested by the international community of Sjögren's syndrome experts (18). More specifically, we wanted to test whether the use of digital image analysis for the quantification of measurable parameters (*e.g.* the fraction of the glandular area occupied by the inflammatory infiltrate) was stringent and reproducible enough, compared to the current gold standard, the focus score, to reduce disagreement among different raters when analysing haematoxylin and eosin staining of salivary gland biopsies, the first step in the histological characterisation of the patients.

Materials and methods

Rationale

Salivary gland tissue was collected from 17 Sjögren's syndrome and 7 non-specific chronic sialadenitis (NSCS) patients attending a specialist Rheumatology/Oral medicine clinic at Barts Health NHS Trust. Demographic, clinical and laboratory data of the patients enrolled in this study is provided in Table I. All procedures were approved by the local Ethics Committee (LREC 05/Q0702/1 and LREC 17/WS/0172) and performed after informed written consent.

Four raters scored a total of 24 digital slides obtained from minor labial SG biopsies, one slide per patient. For each sample, more than one lobule were present on the slide and were analysed, in order to achieve the recommended minimum glandular area of 8 mm^2 (18). Only 2 out of 24 samples did not fulfil this criterion, with an area of approximately 4-5 mm^2. The raters were from 3 different centres (2 in Italy, 1 in the UK) and had different backgrounds (2 anatomical pathologists and 2 immunobiologists) with extensive but not identical experience in pSS histology.

The slides to be analysed were selected from the full biobank, using as selection criteria a good quality H&E staining and the presence of at least a 25% of NSCS.

The digital slides were disseminated to the 4 raters together with video tutorials explaining how to install and operate the software. Each rater was asked to provide i) the total area of the gland calculated using the classic grid-based method, ii) the total area of the gland calculated using the software area calculation capabilities by manually or semi-manually determining the perimeter of the salivary gland tissue, iii) the number of inflammatory foci (peri-ductal leukocytic infiltrates with more than 50 lymphocytes) and iv) the area of the inflammatory infiltrate. The raters were provided with short video tutorials to get acquainted with the use of Qupath and its tools for the specific tasks that they had to perform. No instruction on how to score the tissue was provided, as all the raters agreed to use the recommendations published in 2017 (18).

Tissue preparation, staining and digital image acquisition

The slides were obtained from FFPE minor salivary glands biopsies. 3 μm sections were cut, mounted on polarised microscope slides (VWR) and baked at 60°C for 20 minutes. Tissue sections were deparaffinised in sequential changes of xylene and ethanol before being washed, stained with haematoxylin and eosin and then dehydrated by sequential changes in ethanol and xylene before mounting in DPX medium (Sigma-Aldrich). Slides were digitalised using a Hammamatsu NanoZoomer S60 Slide Scanner at 20x magnification.

The digital analysis of the slides was performed with the open software Qupath 2.0-m9 (24). Each rater installed the same version of the software on their personal computer. Instructions were provided on how to annotate the slides and extract information such as the total area of the gland, the number of infiltrates and the area of the infiltrates. For the manual quantification of the total area of the gland, the raters were asked to superimpose a digital grid to the slides using the native Qupath function and count the squares occupied by salivary tissue in order to provide the total area in mm^2. The annotations and manual area calculations were then sent as separate spreadsheets to the lead centre in London, where they were collated and cleaned for analysis.

Data analysis and statistics

The data from the digital image analysis and manual area calculation were imported, analysed and plotted using R version 3.6.3 (2020-02-29) for Linux. Non-normal distribution of the data was confirmed using a Shapiro-Wilk test. For inter-rater and intra-rater pairwise correlations, the Spearman correlation test was used, and *R*-values are reported with their *p*-values. In the Bland-Altman plots, for every pair of measures assessed the mean of the differences (blue solid line) and the confidence interval (red dashed lines, mean of the differences ± 1.96xSD (differences)) are reported. For the inter-class correlation analysis, the *icc* function from the *irr* package (v. 0.84.1) was used, applying a "two-way" model for non-repeated measurements and we reported the absolute agreement coefficient. For the receiving operating characteristics (ROC) curves, we used the *pROC* package (v. 1.16.2) to calculate Specificity, Sensitivity, Accuracy, Area under the Curve (AUC) and best threshold using Younden's method. All the plots were produced using *ggplot2* package (v. 3.3.0).

Results

Digital gland surface quantification is more robust than manual quantification and reduces inter-raters bias

Every rater performed the analyses of the same batch of images on different machines using the same version of Qupath (v. 2.0-m9) (24). We chose this software due to its versatility, the fact that is platform-independent, it does not require above-average CPU and RAM to run, it is under constant development and debugging with a wide user based and it is free.

The first question that we addressed, was whether there was a significant

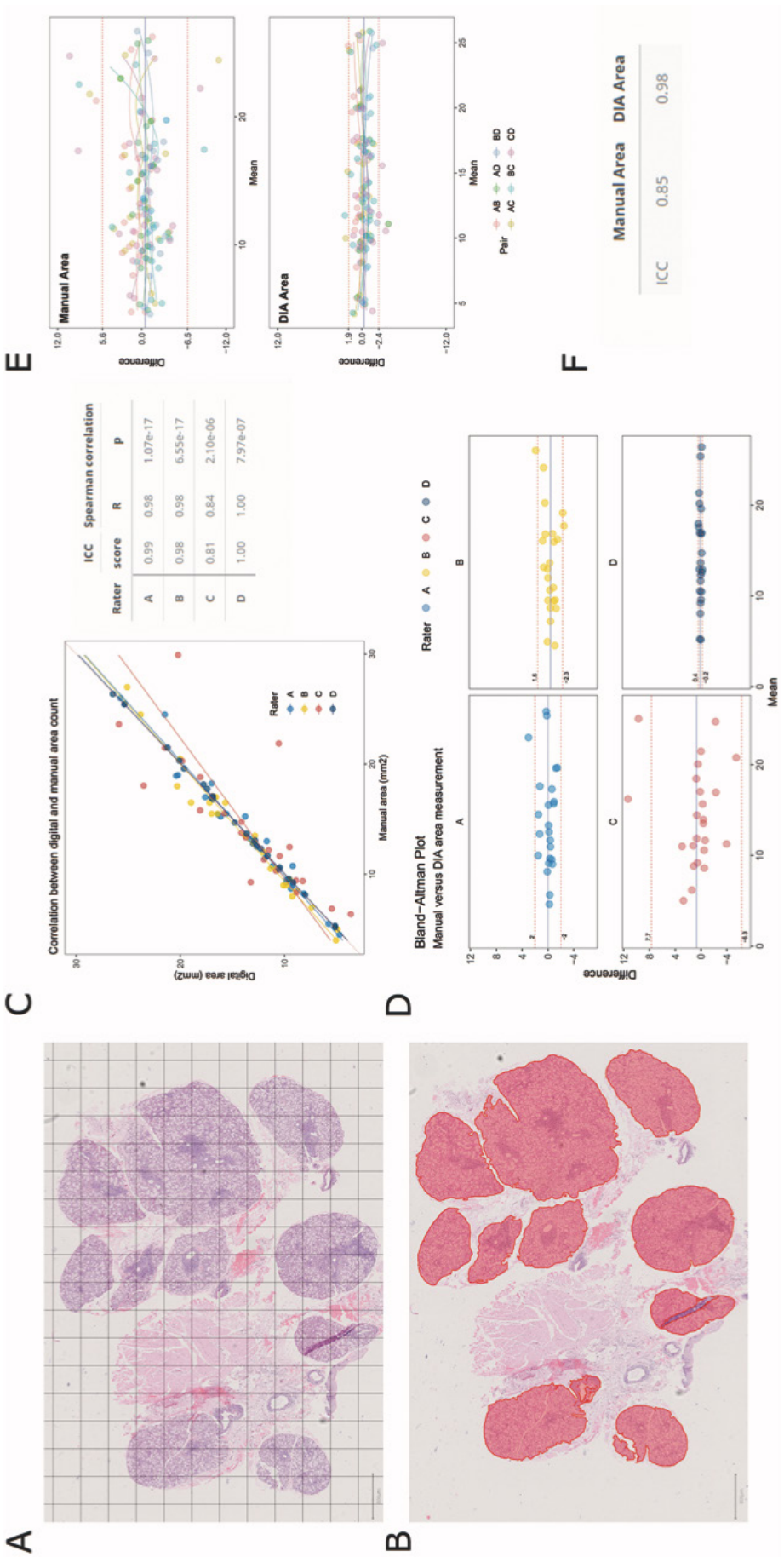

Fig. 1. Comparison between the manual grid-based quantification of the total gland surface and the digital quantification methods.
A-B representative H&E slides showing the two different approaches to the quantification of the total glandular surface. In **A**, a representative grid spaced 0.5 mm per side has been superimposed to the image to allow the rater to count the squares occupied by tissue. In **B**, the perimeter of the glandular tissue of the same slide has been semi-manually marked to allow the software to determine the area. **C** Scatterplot showing manual and digital area distribution for the various raters and relative linear regression lines, the table shows the interclass correlation coefficient and the Spearman R with its p-value for digital and manual quantification for the various raters. In **D**, Bland-Altman plot showing mean and difference for manual and digital areas for the various raters (blue horizontal line represents the average difference, dashed red lines represent the confidence intervals, see Material and Methods). **E** Bland-Altman plots showing the pairwise difference and mean for the manual area (top) and digital area (bottom) quantification for each pair of raters (blue horizontal line represents the average difference, dashed red lines represent the confidence intervals, see Material and Methods) and (**F**) interclass correlation coefficient for manual and digital quantification among the 4 raters.

difference in the quantification of the total area of the gland when using the grid-based manual method (5) or a more modern approach that mixes the experience of the rater with the precision of digital analysis. In Figure 1A-B the same representative image is shown with a superimposed grid (Fig. 1A, example grid side 0.5 mm, raters were left freedom to choose the grid size) or with the glandular area manually annotated and highlighted by the software in red (Fig. 1B). The Spearman correlation *R* and inter-class correlation coefficient (ICC) for the intra-rater agreement between manual and digital quantification were generally good going from a moderate agreement for rater C to an almost perfect agreement in rater D (Fig. 1C and Table I, Spearman *R*=0.84 and 1 and ICC=0.81 and 1, respectively for rater C and D). The variation in raters internal agreement is even clearer when plotting the mean and difference between manual and digital area calculation in a Bland-Altman plot (Fig. 1D). These plots clearly show how the confidence interval varies substantially among raters, with rater D having the narrower and rater C the widest and raters A and B sitting in between. This observation is well reflected in the inter-rater agreement for manual and digital area measurements: in Figure 1E the Bland-Altman plot for all the possible pairs of raters shows how using the manual calculation for the total area of the gland produces a large confidence interval wider than 10 mm^2 while using the digital quantification this is reduced to about 4.5 mm^2. This was also confirmed by the ICC which was excellent for the digital image analysis (DIA) calculation agreement and only moderate for the Manual area calculation agreement (Fig. 1F).

Overall, this first analysis allows to conclude that when calculating the total area of the gland, the use of a software-assisted approach where the operator annotates the perimeter of the tissue and the area is automatically calculated provided both a lower intra-observer variation and a higher inter-observer agreement when compared with the classic grid-based method.

Table I. Demographic of the population used for this work.

Parameter	Value
Gender (n) F/M	20/4
Diagnosis (n) Sjögren's/Sicca	17/7
Age in years mean -median (SD) [min - max]	55.9 - 57.5 (16.9) [25-88]
Disease duration in years mean -median (SD) [min - max]	4.9 - 3 (6.3) [1 - 30]
ESSDAI mean -median (SD) [min - max]	4.3 - 3.5 (4.3) [0 - 18]
Anti-Ro positive % of total	53.3%
Anti-La positive % of total	21%
Rheumatoid factor above 15 g/L % of total	37.5%

Raters agreement is significantly improved when calculating the total area of the infiltrate compared to the number of inflammatory foci

Next, we investigated which measure of infiltration, the total number of inflammatory *foci* per slide or the total size of the infiltrate per slide, performed better in terms of inter-observer agreement. The pairwise Bland-Altman plots in Figure 2A and B suggested that the agreement would be higher for the size of the infiltrate rather than the number of *foci*. This is clear when plotting the number of *foci* per slide and the total size of the infiltrate per slide, plotting the values reported by each individual rater (Fig. 2C-D). The variation among the raters is visibly higher for the infiltrates number (Fig. 2C) than it is for the total infiltrate area (Fig. 2D). In order to account for the inherent difference in measurement units (an integer for the number of *foci* and an area in mm^2 for the infiltrate size), we plotted the coefficient of variation (the ratio between the standard deviation and the mean of the measurements of each rater per slide) for the number of infiltrates and the infiltrate size for each slide, setting an arbitrary tolerance threshold at 30%: Figure 2E shows that in these conditions the infiltrate area suffered a lower variation when compared to the number of infiltrates. This was further confirmed by ICC (Fig. 2F).

Immune cell area fraction is more robust than the focus score for the characterisation of the salivary gland infiltrate

Both Focus Score (FS) and Area Fraction (AF) are dependent on the total area of the gland, but where FS is a function of the number of infiltrates, AF is calculated using the total infiltrate area (Fig. 3A). When comparing FS and AF as indicators of the level of glandular infiltration, the two indices correlated well, as shown by the Spearman *R* values in Figure 3B. Furthermore, when we tested whether AF was able to discriminate pSS from NSCS samples in our cohort, the ROC calculated using the AF values by each individual rater or the average AF all provided a good area under the curve (Fig. 3C). We were also able to identify the best diagnostic thresholds for each rater and for the average of the raters by using Younden's method (Fig. 3C). Bases on this analysis we identified an AF threshold of 0.61 as the averaged best discriminator among the 4 different raters.

Area fraction threshold largely improves inter-rated diagnostic agreement compared to the focus score

As we showed in Figure 1, different operators generally have a good agreement when determining the total gland surface area value, in particular when using the DIA approach. Thus, the robustness of FS and AF as indices of infiltration

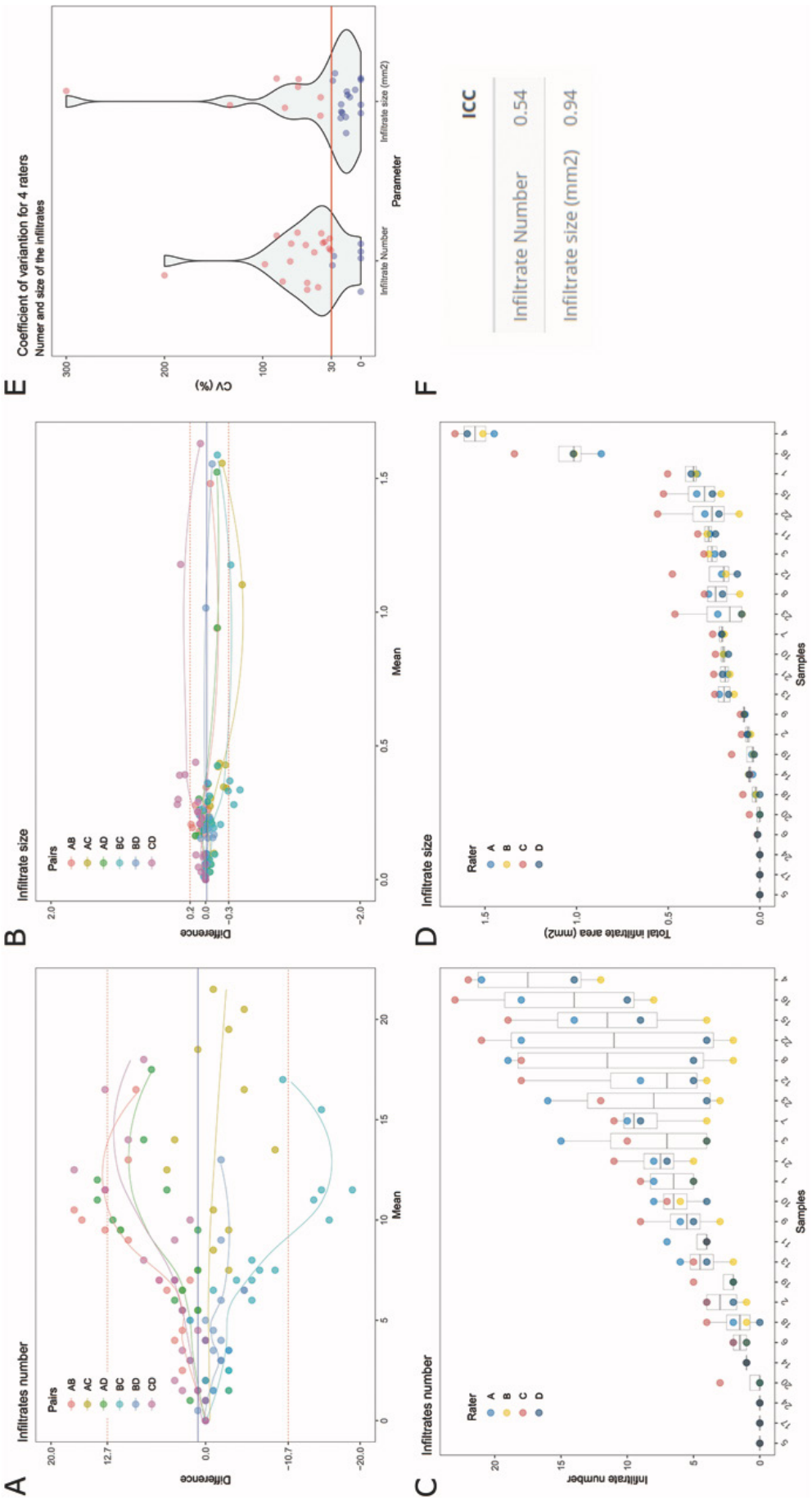

Fig. 2. Raters agreement for the quantification of the number of infiltrates and their total area.
Bland-Altman plots showing the pairwise difference and mean for the number of infiltrates (**A**) and total area of the infiltrate (**B**) quantification for each pair of raters. The number of infiltrates per slide (**C**) and the total area of the infiltrate per slide (**D**) as quantified by each rater show that the disagreement among raters is higher for the infiltrate number than it is for their size. This is also clear by the distribution of the percentage coefficient of variation (CV, **E**) and confirmed by the ICC (**F**).

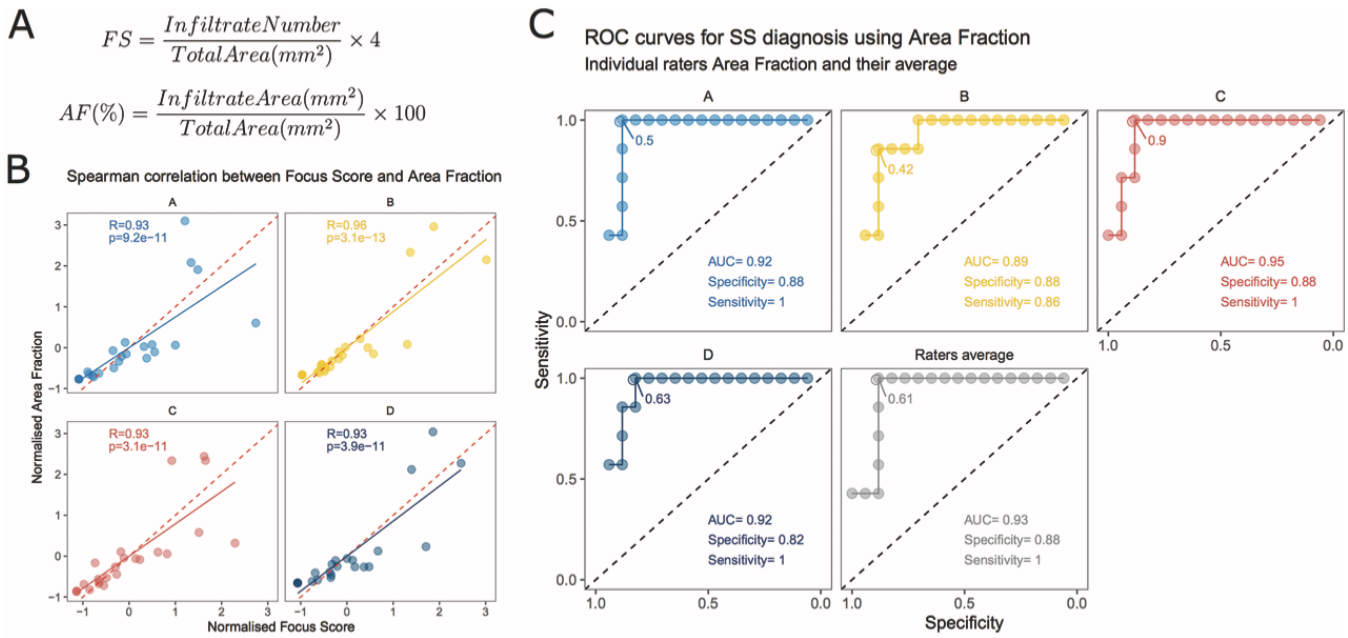

Fig. 3. Area fraction and focus score correlate and area fraction can discriminate Sjögren's from NSCS patients.
Focus score and percentage area fraction were calculated using the formulas in **A**. Dot plot showing the normalised area fraction and focus score for the 4 raters (**B**). Spearman *R* and its *p*-value are reported for each rater in the relevant plots. In **C**, the ROC curves for the capacity of AF measure to discriminate between pSS and NSCS are reported for each rater and for the raters average AF. For each plot, the best Younden's threshold is plotted in the top-left corner and the relative point in the curve is indicated by the connecting line. Each plot also reports the Area Under the Curve (AUC) and the specificity and sensitivity for the Younden's threshold.

largely depends on the performance of the number of infiltrates and the total infiltrate size, respectively (Fig. 3A). Unsurprisingly, given the high inter-rater variation for the number of infiltrates shown in Figure 2, the FS value also suffered a high variation among raters, which was observed at lower levels for the AF value (Fig. 4A-B). Consequently, when plotting the coefficient of variation per slide for the two measures, most of the AF values fell below the arbitrary 30% CV threshold whereas most of the FS values were above threshold (Fig. 4C) and the ICC was excellent for AF and poor for FS (Fig. 4D).

We next modelled the performance of an AF threshold of 0.61 to the traditional FS threshold of 1 to investigate the level of diagnostic accuracy defined as the level of misclassified samples across the 4 different raters. As shown in Figure 4A-B, AF displayed significantly higher robustness o compared to the FS, whereby the AF 0.61 threshold generated complete agreement among all raters in 75% of the slides, whereas the use of the diagnostic FS threshold (FS=1), even when implemented with DIA tools, led to a complete agreement in only in 45.8% of the cases.

Discussion

The use of minor salivary gland biopsies histopathology has long been used for the diagnosis and classification of pSS in clinical practice. A correct and thorough interpretation of histological salivary gland specimens obtained to confirm a pSS diagnosis can also provide a vast amount of additional information on disease stage, its potential evolution and even treatment response (17). However, long recognised pitfalls and limitations in the analysis and reporting of minor salivary gland biopsy histopathology including the use of cumbersome traditional methodology described over 50 years ago has hindered any significant progress in the field over and above its routine diagnostic use (19). Recent recommendations from an international consensus group of pSS experts have provided guidelines towards the definition of a minimal set of requirements for a more refined quantitative interpretation of salivary gland immunopathology, particularly in the context of clinical trials (18).

Here we presented a modern computer-based approach to the analysis of digitalised high-definition slides using a platform-independent free software such as Qupath that allowed a precise quantification of histological features implementing the recommendations described in Fisher *et al*. (18). Because the FS is inherently influenced by an accurate analysis of the total glandular area, we first investigated whether DIA could reduce the variability in the calculation of the total salivary gland area. We showed that, our approach was highly reproducible across different independent observers and could accurately calculate the total area of the gland in a vastly superior way compared to the traditional manual, grid-based, total area calculation. Inclusion of fat degeneration and fibrotic areas present within the salivary gland tissue was discussed among the raters. We again followed the 2017 recommendations and included the whole salivary gland area in the denominator of AF and FS. For specimens where the degeneration affects a large proportion of the gland, including the whole area in the calculation of AF/FS might generate misleading results (17). Suggesting what would be the best practice in these cases of widespread degeneration is not in the scope of this work, but we recognise that this represent one of the points

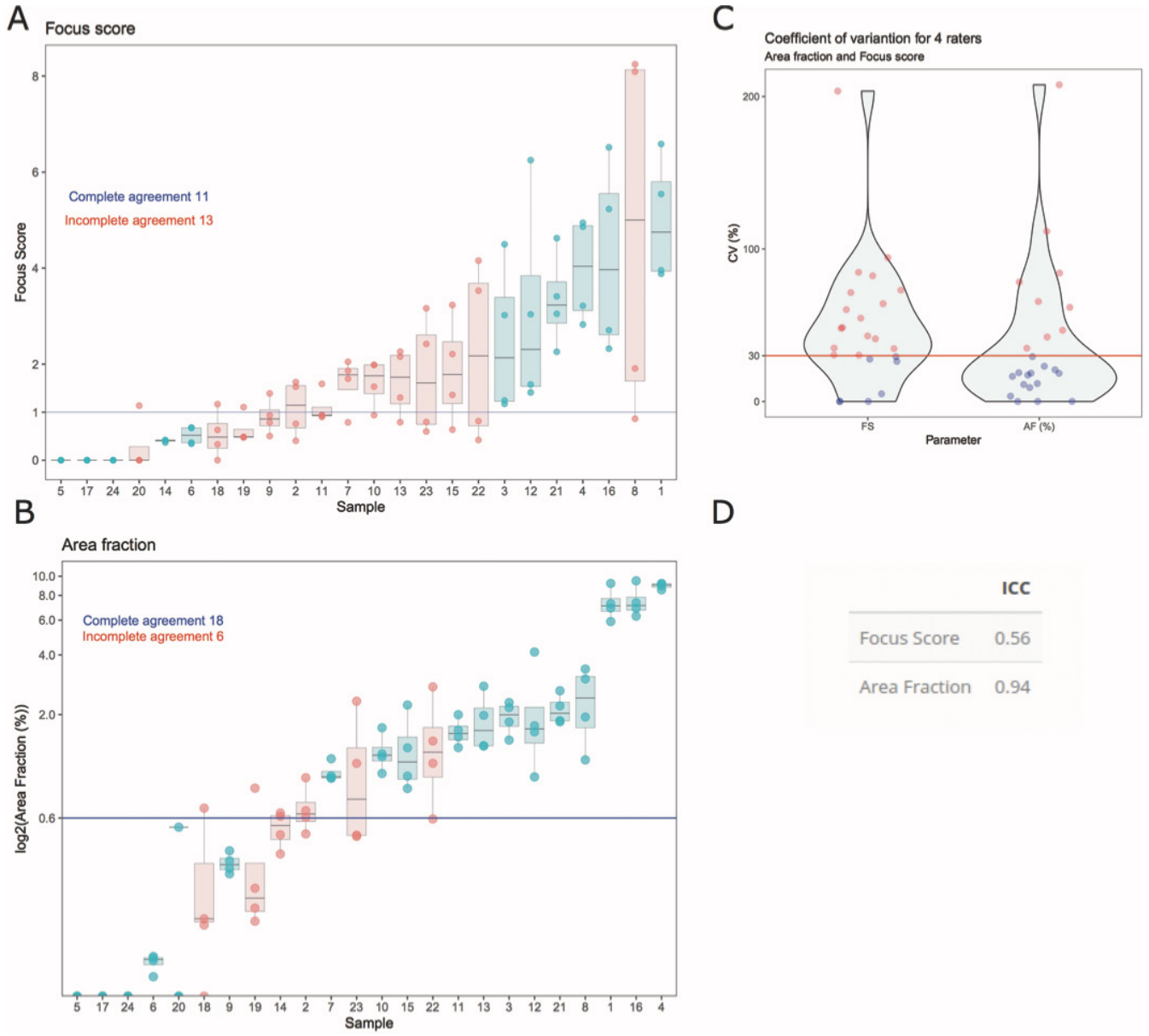

Fig. 4. Area fraction shows a lower level of variation and higher agreement among raters compared to the focus score.
The FS per slide (**A**) and the AF per slide (**B**) as quantified by each rater show that the disagreement among raters is higher for the FS than it is for the AF. This is also clear by the distribution of the percentage coefficient of variation (CV, **C**) and confirmed by the ICC (**D**). Moreover, it was more frequent to achieve complete agreement among the 4 raters using AF and the diagnostic threshold calculated with the ROC curves in Figure 3 than by using the diagnostic FS threshold at FS=1 (**A-B**, blue lines represent the relevant threshold for FS=1 and AF=0.6, red colours histograms and dots highlight incomplete diagnostic agreement, blue histograms and dots highlight complete agreement among raters).

still unclear in the analysis of salivary gland biopsies that the international Sjögren's community should address soon. Next, we evaluated the accuracy of the evaluation of the total number of inflammatory foci in each digitalised salivary gland sample by independent raters. Unsurprisingly, we showed that the quantitative assessment of the total number of foci per gland was extremely variable among raters leading to an extremely poor inter-rater agreement. As expected, samples with a higher number of inflammatory infiltrates led to a much larger degree of variability in the total count of foci among the different observers. This is not surprising as largely infiltrated glands frequently present with confluent inflammatory foci leaving a high degree of subjective interpretation in the definition of the spatial limit of each inflammatory focus. Another area where subjective interpretation can lead to discording results, is whether plasma cells should be included in the calculation of the foci and the infiltrate area. It has been reported for a long time in literature that a decrease of the proportion of IgA+ against IgG+/IgM+ plasma cells can be a diagnostic marker for pSS (17, 25-27). Nevertheless, including plasma cells in the focal infiltrate calculation as determined by H&E staining alone, without a precise immunostaining for the immunoglobulin classes, might introduce a strong bias, as plasma cells are numerous also in non-pSS biopsies (17, 25-27). Since the identification and

exclusion of plasma cells within dense lymphocytic aggregates might be difficult, we suggest excluding the plasma cells-only clusters that are often found in the intra-acinar space and are clearly devoid of other lymphocytes. Notably, we were able to overcome the limitation of the number of foci in the assessment of the degree of salivary gland immune cell infiltration through i) the calculation of the total area covered by the inflammatory infiltrates and ii) the calculation of the area fraction defined as the total area covered by the inflammatory infiltrates divided by the total glandular area. Accordingly, the derivative scores of the number of infiltrates vs their total size and of the focus score vs area fraction, showed similar results in terms of inter-rater agreement, with area fraction providing a much higher ICC when compared to focus score.

In order to quantify the superior diagnostic accuracy of the area fraction compared to the focus score, we calculated the area fraction threshold that could separate pSS from NSCS patients with the highest sensitivity and specificity based solely on the available histology. We performed this analysis on the area fraction measurements for each individual rater and for the mean area fraction, obtained using the individual measurements to calculate an average per slide. By performing this procedure, we showed that not only the individual thresholds could separate pSS from NSCS with high accuracy, but this was true also for the threshold calculated using the area fraction averages. Using this threshold as a cut-off for the individual area fraction measures from the different raters, we showed that it was far more likely that all 4 raters were in agreement in scoring each individual slide above or below the area fraction threshold than when we used the focus score at the diagnostic cut-off of FS=1. Strikingly, the use of the FS in our setting led to a misclassified FS diagnostic value in over 50% of the samples, highly similar to the 53% diagnostic revision rate reported by Vivino *et al*. (19). Conversely, the area fraction analysis using a cut-off AF=0.6 more than halved the rate of histological misclassification. Overall, these data strongly advocate the use of the area fraction as vastly superior to the focus score for the accurate classification of pSS minor salivary gland biopsies across different centres. This observation has also profound implications for the evaluation of the prognostic role of salivary gland histopathology in disease evolution in large longitudinal multicentre initiatives such as HarmonicSS (https://www.harmonicss.eu/) and Necessity (https://www.imi.europa.eu/projects-results/project-factsheets/necessity). Whereas the use of area fraction might not be practical for the day to day routine of pathology laboratories, the data we showed here suggests that its use would allow correct harmonisation of histopathology data across different sites, in the context of multi-centre clinical trials and international consortia. Consistently, all the three centres involved in the current study are active members of the HarmonicSS consortium.

In conclusion, the purpose of this study was to show that the use of a modern digital approach to the histological scoring of the minor salivary gland biopsies and with the adoption of the area fraction calculation can provide a distinct advantage in the study of pSS. This study does not have the necessary power nor was it designed to prove the superiority of AF to FS for the clinical diagnosis of pSS which needs validation in larger cohorts. Nevertheless, our data strongly advocate the routine use of the AF score alongside the routine FS during the assessment of the salivary gland biopsies and in particular in the context of multicentre clinical studies and clinical trials in pSS patients.

References

1. MARIETTE X, CRISWELL LA: Primary Sjögren's syndrome. *N Engl J Med* 2018; 378: 931-9.
2. QIN B, WANG J, YANG Z *et al*.: Epidemiology of primary Sjögren's syndrome: A systematic review and meta-analysis. *Ann Rheum Dis* 2015; 74: 1983-9.
3. BRITO-ZERÓN P, BALDINI C, BOOTSMA H *et al*.: Sjögren syndrome. *Nat Rev Dis Prim* 2016; 2: 1-20.
4. CAFARO G, CROIA C, ARGYROPOULOU OD *et al*.: One year in review 2019: Sjögren's syndrome. *Clin Exp Rheumatol* 2019; 37 (Suppl. 118): S3-15.
5. CHISHOLM DM, MASON DK: Labial salivary gland biopsy in Sjögren's disease. *J Clin Pathol* 1968; 21: 656-60.
6. SHIBOSKI CH, SHIBOSKI SC, SEROR R *et al*.: 2016 Classification Criteria for primary Sjögren's Syndrome: a consensus and data-driven methodology involving three international patient cohorts. *Arthritis Rheumatol* 2017; 69: 35-45.
7. NOCTURNE G, PONTARINI E, BOMBARDIERI M, MARIETTE X: Lymphomas complicating primary Sjögren's syndrome: from autoimmunity to lymphoma. *Rheumatology* 2019; 8: e59868.
8. HAACKE EA, VEGT B VAN DER, VISSINK A, SPIJKERVET FKL, BOOTSMA H, KROESE FGM: Germinal centres in diagnostic labial gland biopsies of patients with primary Sjögren's syndrome are not predictive for parotid MALT lymphoma development. *Ann Rheum Dis* 2017; 76: 1783-6.
9. THEANDER E, VASAITIS L, BAECKLUND E *et al*.: Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary Sjögren's syndrome. *Ann Rheum Dis* 2011; 70: 1363-8.
10. CARUBBI F, ALUNNO A, CIPRIANI P *et al*.: Is minor salivary gland biopsy more than a diagnostic tool in primary Sjögren's syndrome? Association between clinical, histopathological, and molecular features: A retrospective study. *Semin Arthritis Rheum* 2014; 44: 314-24.
11. RISSELADA AP, KRUIZE AA, GOLDSCHMEDING R, LAFEBER FPJG, BIJLSMA JWJ, VAN ROON JAG: The prognostic value of routinely performed minor salivary gland assessments in primary Sjögren's syndrome. *Ann Rheum Dis* 2014; 73 :1537-40.
12. BROWN S, NAVARRO COY N, PITZALIS C *et al*.: The TRACTISS Protocol: A randomised double blind placebo controlled clinical TRial of anti-B-cell therapy in patients with primary Sjögren's syndrome. *BMC Musculoskelet Disord* 2014; 15: 554.
13. FISHER BA, SZANTO A, NG W-FF *et al*.: Assessment of the anti-CD40 antibody iscalimab in patients with primary Sjögren's syndrome: a multicentre, randomised, double-blind, placebo-controlled, proof-of-concept study. *Lancet Rheumatol* 2020; 9913: 1-11.
14. FISHER BA, EVERETT CC, ROUT J *et al*.: Effect of rituximab on a salivary gland ultrasound score in primary Sjögren's syndrome: results of the TRACTISS randomised double-blind multicentre substudy. *Ann Rheum Dis* 2018; 77: 412-6.
15. MCHUGH J: Blocking CD40–CD40L interactions in Sjögren syndrome. *Nat Rev Rheumatol* 2019; 15: 252.
16. FISHER BA, BROWN RM, BOWMAN SJ, BARONE F: A review of salivary gland histopathology in primary Sjögren's syndrome with a focus on its potential as a clinical trials biomarker. *Ann Rheum Dis* 2015; 74: 1645-50.
17. KROESE FGM, HAACKE EA, BOMBARDIERI M: The role of salivary gland histopathology in primary Sjögren's syndrome: Promises and pitfalls. *Clin Exp Rheumatol* 2018; 36 (Suppl. 112): S222-33.
18. FISHER BA, JONSSON R, DANIELS T *et al*.: Standardisation of labial salivary gland

histopathology in clinical trials in primary Sjögren's syndrome. *Ann Rheum Dis* 2017; 76: 1161-8.

19. 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.
20. 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.
21. SÈNE D, ISMAEL S, FORIEN M *et al*.: Ectopic germinal center–like structures in minor salivary gland biopsy tissue predict lymphoma occurrence in patients with primary Sjögren's syndrome. *Arthritis Rheumatol* 2018; 70: 1481-8.
22. HILLEN MR, BARONE F, RADSTAKE TRDJ, VAN ROON JAG: Towards standardisation of histopathological assessments of germinal centres and lymphoid structures in primary Sjögren's syndrome. *Ann Rheum Dis* 2016; 75: e31.
23. CARUBBI F, ALUNNO A, CIPRIANI P *et al*.: Different operators and histologic techniques in the assessment of germinal center-like structures in primary Sjögren's syndrome minor salivary glands. *PLoS One* 2019; 14: 1-10.
24. BANKHEAD P, LOUGHREY MB, FERNÁNDEZ JA *et al*.: QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017; 7: 16878.
25. ZANDBELT MM: The synergistic value of focus score and IgA% score of sublabial salivary gland biopsy for the accuracy of the diagnosis of Sjögren's syndrome: a 10-year comparison. *Rheumatology* 2002; 41: 819-23.
26. BODEUTSCH C, DE WILDE PCM, KATER L *et al*.: Quantitative immunohistologic criteria are superior to the lymphocytic focus score criterion for the diagnosis of Sjögren's syndrome. *Arthritis Rheum* 1992; 35: 1075-87.
27. DE WILDE PCM, VOOYS GP, BAAK JPA *et al*.: Quantiative immunohistologic and histomorphometric diagnostic criteria for Sjögren's syndrome. *Pathol Res Pract* 1989; 185: 778-80.