Diagnostic accuracies of sialography and salivary ultrasonography in Sjögren’s syndrome patients: a meta-analysis

G.G. Song, Y.H. Lee

Division of Rheumatology, Department of Internal Medicine, Korea University College of Medicine, Seoul, Korea.

Abstract

Objective

The purpose of this study was to compare the diagnostic performance of sialography and salivary ultrasonography (US) for Sjögren’s syndrome (SS) patients.

Methods

We searched Medline, Embase, and the Cochran library, and performed two meta-analyses on the diagnostic accuracy of sialography and salivary US in SS patients.

Results

A total of six studies including 488 patients and 447 controls from two European and four Asian studies were available for the meta-analysis. The pooled sensitivity and specificity of sialography were 80.0% (95% confidence interval [CI] 76.4–83.2) and 89.0% (85.8–91.8), respectively, and 77.4 (73.7–80.9) and 81.5 (77.6–85.0) for US, respectively. For sialography, the PLR, NLR, and DOR were 9.296 (4.200–20.57), 0.228 (0.170–0.305), and 46.51 (16.14–134.0), respectively, and for US were 4.631 (2.707–7.864), 0.302 (0.226–0.403), and 17.48 (10.03–30.45), respectively. The area under the curve (AUC) of sialography was 0.824, and the $Q^*$ index was 0.757, while the AUC of US was 0.864, and its $Q^*$ index was 0.794, indicating that the diagnostic accuracy of US is comparable with sialography in SS patients. A subgroup meta-analysis according to the diagnostic criteria did not change the overall diagnostic accuracy.

Conclusion

Our meta-analysis of published studies demonstrates that the diagnostic accuracy of salivary US is comparable with sialography in SS patients.

Key words

Sjögren’s syndrome, sialography, ultrasonography, diagnostic accuracy, meta-analysis
**Introduction**

Sjögren’s syndrome (SS) is a chronic systemic autoimmune disease affecting the exocrine system. SS is characterised by dry eyes and dry mouth and lymphocytic infiltration of the salivary and lacrimal glands. Diagnosis of SS is made by diagnostic criteria, including the combination of several tests (1). Among these criteria, the classification criteria proposed by the American European criteria (AEC) are most often used (2). The AEC include sialography and salivary scintigraphy for the evaluation of salivary gland involvement. However, the evaluation of salivary gland involvement in SS is still controversial. Other diagnostic methods have been studied, including salivary ultrasonography. US has been considered to be the most attractive method for the assessment of salivary gland involvement because it is non-invasive, inexpensive, and does not require radiation (3).

US may come to replace conventional invasive examinations in clinical practice. However, the diagnostic accuracy of salivary US has not been clearly compared with sialography, and there is as of yet no consensus on the use of US as an alternative method for the assessment of salivary gland involvement in SS patients. Salivary US has been studied in the context of SS in comparison with sialography with respect to diagnostic accuracy. However, published results on the diagnostic accuracies of sialography and US are controversial and inconclusive (4-9). This may be due to small sample sizes, low statistical power, and/or clinical heterogeneity. In order to overcome the limitations of individual studies, resolve inconsistencies, and reduce the likelihood of random errors producing false positives or negatives (10-12), we performed this diagnostic meta-analysis on the sensitivities and specificities of sialography and salivary US for the diagnosis of SS in order to assess the diagnostic accuracies of sialography and US using published data.

**Materials and methods**

*Identification of eligible studies and data extraction*

We utilised Medline, Embase, and the Cochrane library to identify articles published in April 2013 in which salivary US and sialography were performed in SS patients and controls. In addition, all references mentioned in the identified articles were reviewed to identify studies not indexed by electronic databases. The following keywords and subject terms were used in the search: “sialography”, “ultrasonography”, “sensitivity”, “specificity”, and “Sjögren’s syndrome”. Studies were included in the analysis if: (a) they were case-control studies, (b) included sufficient data to calculate sensitivity and specificity, and (c) the study included patients diagnosed with SS as based on the classification criteria. No language restriction was applied. We excluded the following: (a) studies including overlapping data and (b) studies in which there was no control group. Data regarding the methods and results of meta-analysis were extracted from the original studies by two independent reviewers. Discrepancies between the reviewers were resolved by consensus or a third reviewer. We extracted information on author, publication year, and demographic characteristics of participants (ethnicity and diagnostic criteria) from each study. Sialography and US raw data were extracted from all primary studies to fill the four cell values of a diagnostic 2x2 table (true positives, false positives, true negatives, and false negatives). We used the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) to assess the quality of each study (13).

*Evaluation of statistical associations*

We used two meta-analysis methods to assess the overall diagnostic ability of sialography and US. Within- and between-study variations and heterogeneities were assessed using Cochran’s Q-statistic. Cochran’s Q-statistic test assesses the null hypothesis that all studies evaluated the same effect. The effect of heterogeneity was quantified using I² with a range between 0 and 100%, representing the proportion of between-study variability attributable to heterogeneity rather than to chance (14). I² values of 25%, 50%, and 75% were nominally assigned as low, moderate, and high estimates, respectively.
The fixed effects model assumes that a genetic factor has a similar effect on disease susceptibility across all studies investigated and that observed variations among studies are caused by chance alone (15). The random effects model assumes that different studies show substantial diversity and assesses both within-study sampling error and between-study variance (16). When study groups are homogenous, the two models are similar. If the study groups lack homogeneity, the random effects model usually provides wider CIs than the fixed effects model. The random effects model is most appropriate in the presence of significant between-study heterogeneity (16). A fixed effects or random effects model was used to combine sensitivity, specificity, positive and negative likelihood ratios (PLR, NLR), and diagnostic odds ratio (DOR) estimates, and summary receiver-operating characteristic curves (SROC) were analysed. DOR is a unitary measure of diagnostic performance that encompasses both sensitivity and specificity or both PLR and NLR, and DOR is regarded as a suitable global measure of accuracy for comparing the overall diagnostic accuracies of different tests (17).

Because sensitivity and specificity are inter-dependent variables, independent calculations may sometimes underestimate both. SROC curve analysis is more appropriate because it accounts for this mutual dependence. The area under the curve (AUC) – in this case, area under the SROC curve – presents an overall summary of test performance and displays the trade-off between sensitivity and specificity, and an AUC of 1.0 (100%) indicates perfect discriminatory ability for a diagnostic test (13). In addition, the Q* index is another useful global estimate of test accuracy for comparing SROC curves. The Q* index is defined at the point where sensitivity equals specificity on an SROC curve, and is the point on a SROC curve intersected by the anti-diagonal. A Q* value of 1.0 indicates 100% accuracy – i.e. sensitivity and specificity of 100% – (13). Statistical manipulations for this meta-analysis were performed using MetaDiSc, version 1.4 (Hospital Universitario Ramon y Cajal, Madrid, Spain) (16).

**Results**

**Studies included in the meta-analysis**

We identified 57 studies by electronic and manual searching, and 13 were selected for full-text review based on title and abstract (3-9, 18-23). Seven of these were excluded: six had no control data (18-23), and one was a review (3) (Fig. 1). Thus, six studies that reported on the diagnostic accuracies of sialography and US met our study inclusion criteria.

**Table I. Characteristics of individual studies included in meta-analysis.**

<table>
<thead>
<tr>
<th>Author</th>
<th>Ethnicity</th>
<th>Diagnosis criteria</th>
<th>n</th>
<th>Sialography</th>
<th>Ultrasonomography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takagi, 2010</td>
<td>Asian</td>
<td>AEC (2)</td>
<td>177</td>
<td>172</td>
<td>146  31  42  141</td>
</tr>
<tr>
<td>Obinata, 2010</td>
<td>Asian</td>
<td>Japanese revised criteria (5)</td>
<td>32</td>
<td>37  30  2  6  35</td>
<td>28  8  8  29</td>
</tr>
<tr>
<td>Poul, 2009</td>
<td>European</td>
<td>AEC (2)</td>
<td>37</td>
<td>15  35  2  10  13</td>
<td>38  4  7  11</td>
</tr>
<tr>
<td>Salafti, 2008</td>
<td>European</td>
<td>AEC (2)</td>
<td>68</td>
<td>79  56  12  21  67</td>
<td>58  13  19  66</td>
</tr>
<tr>
<td>Yonetsu, 2002</td>
<td>Asian</td>
<td>Japanese criteria (7)</td>
<td>151</td>
<td>123</td>
<td>149  2  30  121</td>
</tr>
<tr>
<td>Yoshiura, 1997</td>
<td>Asian</td>
<td>EC (24)</td>
<td>23</td>
<td>21  23  0  1  21</td>
<td>11  1  13  21</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>488</td>
<td>447</td>
<td></td>
</tr>
</tbody>
</table>

AEC: American-European criteria; EC: European criteria; TP: true positive; FP: false positive; FN: false negative; TN: true negative; 1 means 100% in sensitivity and specificity.
criteria, and these studies included 488 patients and 447 controls (4-9). These studies consisted of two European and four Asian studies. Three studies used the American-European criteria for SS diagnosis (2) and three employed other criteria (5, 7, 24). The characteristic features of the participants in the studies included in the meta-analysis are given in Table I, and the quality assessments of the diagnostic accuracy of the studies are shown in Figure 2.

**Diagnostic accuracy of sialography and US**

When all six studies were considered together, the sensitivity estimates of sialography ranged from 77.7% to 95.8% and the specificity estimates ranged from 82.0% to 100% (Fig. 3). On the other hand, the sensitivity estimates of US ranged from 45.8% to 84.4% and the specificity estimates ranged from 70.9% to 95.5% (Fig. 4). The pooled sensitivity and specificity of sialography were 80.0% (95% CI 76.4–83.2) and 89.0% (85.8–91.8), respectively, and those of US were 77.4 (73.7–80.9) and 81.5 (77.6–85.0) (Table II, Figures 3 and 4). In summary, sialography PLR, NLR, and DOR were 9.296 (4.200–20.57), 0.228 (0.170–0.305), and 46.51 (16.14–134.0), respectively, and those for US were 4.631 (2.707–7.864), 0.302 (0.226–0.403), and 17.48 (10.03–30.45), respectively (Table II). Figure 4 shows the performance of sialography US testing in the form of SROC curves. The AUC of sialography was 0.824 and the Q’ index was 0.757, indicating modest accuracy, while the AUC of US was 0.864 and the Q’ index was 0.794, indicating that the diagnostic accuracy of US is comparable with sialography in SS patients (Table III).

**Diagnostic accuracy of sialography and US in SS patients diagnosed by the AEC**

When all three studies were considered together, the pooled sensitivity and specificity of sialography were 76.5% (95% CI 71.3–81.1) and 83.1% (78.0–87.4), respectively, and those for US were 80.6 (75.8–84.9) and 74.8 (69.1–79.9) (Table II). In summary, sialography PLR, NLR, and DOR were 4.526 (3.434–5.963), 0.285 (0.231–0.305), and 41.91 (28.21–59.01), respectively, and those for US were 4.613 (2.973–7.069), 0.289 (0.232–0.348), and 17.48 (10.03–30.45), respectively (Table II). Figure 4 shows the performance of sialography US testing in the form of SROC curves. The AUC of sialography was 0.824 and the Q’ index was 0.757, indicating modest accuracy, while the AUC of US was 0.864 and the Q’ index was 0.794, indicating that the diagnostic accuracy of US is comparable with sialography in SS patients (Table III).
Diagnostic accuracies of sialography and ultrasonography in SS patients / G.G. Song & Y.H. Lee

and 15.95 (10.64–23.91), respectively, and those for US were 3.165 (2.555–3.922), and 12.35 (8.410–18.16), respectively (Table II). The AUC of sialography was 0.869 and the Q* index was 0.800, indicating modest accuracy. The AUC of US was 0.855 and the Q* index 0.786, which indicated that the diagnostic accuracy of US is comparable with sialography in SS patients satisfying the AEC (Table III).

Heterogeneity
We found heterogeneity in the meta-analyses of sialography and US in the overall groups. However, most of the heterogeneity was resolved by a subgroup meta-analysis according to the diagnostic criteria.

Discussion
Studies of the diagnostic accuracies of sialography and salivary US for the diagnosis of SS have reported inconsistent findings (4-9). These inconsistent findings may be due to false positives, false negatives, or low statistical power due to small sample size. Meta-analysis integrates previous research, and increases statistical power and resolution by pooling the results of independent analyses (25), and thus provides a powerful means of overcoming the small sample size problem and inadequate statistical power.

In this meta-analysis, we combined evidence of the diagnostic accuracies of sialography and US for the diagnosis of SS. This meta-analysis of six studies including 488 patients and 447 controls showed that the sensitivity, specificity, PLR, NLR, and DOR of US are similar to that of sialography. SROC analysis was used to compare the two modalities. When sensitivity and specificity were considered simultaneously, the AUC of sialography was 0.824, whereas that of US was 0.864. The diagnostic performance of US was not significantly different from that of sialography. These results indicated that the diagnostic accuracy of US is comparable with sialography in SS patients. A subgroup meta-analysis according to the diagnostic criteria did not change the overall diagnostic accuracy.

There is no single test that has sufficient accuracy to diagnose SS (2, 26). The diagnostic criteria of the AEC are those most often used for the diagnosis of SS (2). The AEC are reliable and widely accepted, and sialography is a conventional and reliable method for salivary gland involvement and has been included in the AEC. However, sialography has limitations, as it is invasive and requires radiation exposure. Cannulation of the main duct may sometimes be difficult and can cause complications such as sialadenitis (6). Sialectasia may also occur due to parotitis, and sialography cannot be used in patients with infection, inflammation, or allergy to iodine, while US can be used in these patients (5).

As with other imaging methods, sali-
Diagnostic accuracies of sialography and ultrasonography in SS patients / G.G. Song & Y.H. Lee

Varying US is the most attractive method because US is non-invasive, concise, inexpensive, and provides real-time results (27). In addition, the diagnostic accuracy of US is similar to that of sialography. Salivary US can be used as an alternative to sialography for the diagnosis of SS. In addition, considering its convenience, non-invasiveness, and inexpensiveness, US has advantages over sialography. Salivary US may replace sialography in the criteria of the AEC.

However, the present study has several limitations that should be considered. First, between-study heterogeneity was encountered in this meta-analysis. This between-study heterogeneity may have affected the results of this meta-analysis, which may be compounded by the limited information provided on clinical status and disease severity in the populations involved. We tried to overcome this limitation by using a random-effects model that incorporates uncertainties arising due to between-study variation and by doing subgroup analysis based on the diagnostic criteria. Second, there are varying levels of disease severity, and the SS severity level was unclear. The diagnostic accuracy of US may be different in the advanced sialographic stage (8). Further research is required to examine how the diagnostic accuracies of both studies are changed due to the activity or clinical features of the disease.

Nevertheless, this meta-analysis also has its strengths. The number of SS patients from individual studies ranged from 32 to 177. However, this pooled analysis included a total of 488 SS patients and 447 controls. In comparison with an individual study, we were able to provide more accurate data on the diagnostic tests by increasing the statistical power and resolution by pooling the results of independent analyses.

Our meta-analysis of published studies demonstrates that the diagnostic accuracy of salivary US is comparable with sialography in SS patients. We conclude that non-invasive salivary US may be an alternative to sialography in the diagnosis of SS, and could also be a potential replacement for sialography in the diagnostic criteria of the AEC.

Table III. Estimates of summary receiver operating characteristic curve parameters.

<table>
<thead>
<tr>
<th>Test</th>
<th>Population</th>
<th>Study n</th>
<th>AUC</th>
<th>SE(AUC)</th>
<th>Q'</th>
<th>SE(Q')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sialography</td>
<td>All</td>
<td>6</td>
<td>0.824</td>
<td>0.029</td>
<td>0.757</td>
<td>0.026</td>
</tr>
<tr>
<td>Ultrasonography</td>
<td>All</td>
<td>6</td>
<td>0.864</td>
<td>0.014</td>
<td>0.794</td>
<td>0.014</td>
</tr>
<tr>
<td>Sialography</td>
<td>AEC</td>
<td>3</td>
<td>0.869</td>
<td>0.038</td>
<td>0.800</td>
<td>0.037</td>
</tr>
<tr>
<td>Ultrasonography</td>
<td>AEC</td>
<td>3</td>
<td>0.855</td>
<td>0.021</td>
<td>0.786</td>
<td>0.020</td>
</tr>
</tbody>
</table>

AEC: American-European criteria; AUC: area under the curve; SE: standard error.

Fig. 5. SROC curves for sialography (A) and ultrasonography (B) for the diagnosis of Sjogren’s syndrome. Solid circles represent individual studies included in this meta-analysis. The curve shown is a regression line that summarizes overall diagnostic accuracy. SE (AUC), standard error of AUC; Q', an index defined by the point on the SROC curve where the sensitivity and specificity are equal; SE (Q'), Q' index standard error.
References