Comparison of core needle biopsy and fine-needle aspiration in diagnosis of malignant salivary gland neoplasm: Systematic review and meta-analysis

Jungheum Cho MD1 | Junghoon Kim MD1 © | Ji Sung Lee PhD2 | Choong Guen Chee MD3 | Youngjune Kim MD4 | Sang Il Choi MD, PhD1

1Department of Radiology, Seoul National University Bundang Hospital, Seongnam, South Korea
2Clinical Research Center, Asan Medical Center, Seoul, South Korea
3Department of Radiology, Asan Medical Center, Seoul, South Korea
4Aerospace Medical Group, Air Force Education and Training Command, Jinju, Korea

Abstract
Background: In this meta-analysis, we compared the risk of obtaining non-diagnostic results and the diagnostic accuracy for detection of salivary gland malignancy between core needle biopsy (CNB) and fine-needle aspiration (FNA).

Methods: All published English-language studies comparing CNB and FNA diagnostic accuracy for salivary gland masses through December 2019 were searched. Pooled risk ratios (RRs) of nondiagnostic results, sensitivities, and specificities of CNB and FNA for salivary gland malignancy diagnosis were determined. Complication rates were compared.

Results: Six studies (1924 procedures) were quantitatively analyzed. CNB yielded significantly fewer nondiagnostic results ($P<.001$) and had significantly higher pooled sensitivity ($P<.001$) and specificity ($P=.002$) than FNA for differentiating malignant and benign salivary gland neoplasms. Hematoma occurred in 0.3% of CNB, while no complication occurred in FNA procedures.

Conclusion: CNB yielded fewer nondiagnostic results and had superior diagnostic performance compared with FNA for detecting salivary gland malignancies.

Keywords
core needle biopsy, fine-needle aspiration, meta-analysis, neoplasm, salivary glands

1 | INTRODUCTION

Salivary gland cancer accounts for about 6% of all head and neck malignancies.1 However, the majority of salivary lesions are non-neoplastic or benign tumors; the malignancy rate of salivary gland tumors has been reported to be 14% to 27%.2-4 Therefore, differentiating neoplastic from non-neoplastic lesions and malignant from benign neoplasms of the salivary gland is crucial to avoid unnecessary surgery or conversely to plan appropriate surgery.5-7

Fine-needle aspiration (FNA) is commonly used as an initial diagnostic method for salivary gland masses. FNA has the advantages of minimal recovery time and low risk of complications; however, nondiagnostic results are occasionally reported due to insufficient aspiration or the innate limitation of distinguishing between benign and malignant results in the obtained cytology.8 In contrast, core needle biopsy (CNB) is capable of obtaining sufficient tissue retaining the histological architecture and can provide more accurate histological results, and may allow utilization of immunohistochemical techniques.
However, the risk of complication such as bleeding, pain, or tumor seeding of CNB is known to be higher than that of FNA.9,10

As nondiagnostic or inaccurate results of salivary gland biopsy may lead to unnecessary surgery or delays in surgical intervention,11 it is important to determine which biopsy method provides a more accurate diagnosis. To date, a few studies have compared the diagnostic yield of CNB and FNA of salivary gland lesions.10,12-19 However, to the best of our knowledge, there have been no meta-analyses or large cohort studies on this topic. Therefore, we undertook a meta-analysis of comparative studies of CNB and FNA for salivary gland masses, aiming to compare the diagnostic value of salivary gland CNB and FNA for the diagnosis of salivary gland malignancy.

2 | MATERIALS AND METHODS

This systematic review was performed in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.20 The study was registered to the International Prospective Register of Systematic Reviews, or PROSPERO, with a registration number of CRD42020138307.

2.1 | Literature search strategy

A systematic search was performed in PubMed/MEDLINE, Embase, and Cochrane Library for all relevant articles that compared the diagnostic accuracy of FNA and CNB for identifying salivary gland tumors, particularly malignant neoplasms. We used the keywords: “salivary gland,” “parotid,” “submandibular,” “sublingual,” “fine needle,” “core needle,” and “core biopsy.” The latest literature search was performed on 31 December 2019. The studies were limited to those written in English.

2.2 | Study selection

Studies were included if they compared diagnostic performance of FNA and CNB for salivary gland lesions, particularly with a view to differentiating malignancy, as an attempt to minimize the intrastudy and interstudy biases. Only studies that provided sufficient data to calculate the number of nondiagnostic results, true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of FNA and CNB, as compared with the final diagnosis, were included. The final diagnosis was established on the basis of surgical histopathology or clinical follow-up.
TABLE 1  Characteristics of the included studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Location</th>
<th>Candidates</th>
<th>Needles</th>
<th>Image guidance</th>
<th>Site</th>
<th>No. of passes</th>
<th>Performed by</th>
<th>Complications</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang et al</td>
<td>2011</td>
<td>Taiwan</td>
<td>Surgical pathology + clinical follow-up (&gt;1 year)</td>
<td>CNB: 14-18G, spring-loaded (Bard Magnum; C.R. Bard Inc., Covington, GA or Temno; Carefusion Co., San Diego, CA) FNA: 21G, fine needle (Precision Glide; Becton Dickinson, Franklin Lakes, NJ)</td>
<td>CNB, FNA: USG</td>
<td>All parotid</td>
<td>CNB: 1-5</td>
<td>N/A</td>
<td>CNB: 1 hematoma (surgical intervention)</td>
<td>Retrospective</td>
</tr>
<tr>
<td>Song et al</td>
<td>2015</td>
<td>South Korea</td>
<td>Surgical pathology</td>
<td>CNB: 18 G, double-action spring-activated (AceCut; TSK, Tochigi, Japan) FNA: 23G</td>
<td>CNB: USG FNA: Blind</td>
<td>CNB: Parotid 171, SMG 56, SLG 1 FNA: Parotid 329, SMG 40, SLG 2</td>
<td>N/A</td>
<td>CNB: Radiologists FNA: Pathologists</td>
<td>None</td>
<td>Retrospective</td>
</tr>
<tr>
<td>Eom et al</td>
<td>2015</td>
<td>South Korea</td>
<td>Surgical pathology + clinical follow-up (&gt;1 year)</td>
<td>CNB: 18G, double-action spring-activated (AceCut) or semi-automatic needles (Stericut; TSK, Tochigi, Japan) FNA: N/A</td>
<td>CNB, FNA: USG</td>
<td>Parotid, SMG</td>
<td>CNB: 1-3</td>
<td>CNB, FNA: Radiologist</td>
<td>None</td>
<td>Retrospective</td>
</tr>
</tbody>
</table>

(Continues)
up for more than 1 year. Studies with the following features were excluded: case reports, abstracts, and review articles; evaluation of only one index test (FNA or CNB); not providing data separately for FNA and CNB. Titles and abstracts of the studies were independently reviewed for eligibility by two authors (J.K., J.C.), followed by a full-text review. Discrepancies regarding eligibility were resolved by discussion with a third evaluator (C.G.C.).

### 2.3 Quality assessment of the studies

The methodological qualities of the included studies were evaluated using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) criteria. These criteria are intended to determine the risk of bias and the applicability of the included studies.

### 2.4 Data extraction

The numbers of nondiagnostic results such as TP, TN, FP, and FN results of FNA and CNB as compared with the final diagnosis were extracted. In addition to the quantitative data, the following information for each study were extracted: overall study characteristics, method of the final diagnosis, biopsy needle size, image guidance, biopsy site, number of passes, performer, and complication. Two authors (J.K., J.C.) independently extracted data from studies using a standard form. Disagreements were resolved by discussion with a third reviewer (Y.K.).

### 2.5 Statistical analysis

A radiologist (J.K., with 7 years’ experience) and a biostatistician (J.S.L, with 18 years’ experience) planned and performed the statistical analyses. All statistical analyses were performed using MetaDisc version 1.4 (Hospital Universitario Ramón y Cajal, Madrid, Spain) and STATA version 15.0 (StataCorp, College Station, Texas).

Risk ratios (RRs) of nondiagnostic results were subjected to meta-analysis using a random effects model. Study-specific estimates of pooled sensitivity and specificity, with 95% confidence intervals (CIs), were calculated using the Mantel-Haenszel method. Higgins’ $I^2$ statistic was used to assess the proportion of total variation in the reported results, according to intrastudy heterogeneity. An $I^2$ score of 25% to 49% represents low heterogeneity, a score of 50% to 74% is considered to indicate moderate heterogeneity, and a score from 75% onwards indicates high heterogeneity. To explore the potential source of
heterogeneity observed in the meta-analysis, subgroup analysis and meta-regression were conducted. Summary receiver operating characteristic (SROC) curves were calculated to show how the different sensitivities and specificities of the studies were related to each other, and to convey the diagnostic performance of FNA and CNB. The area under the SROC curve (AUC) was also calculated; an AUC score of 1.0 indicated that the diagnostic test was accurate and verified. A Z-test was used to compare the diagnostic accuracy of CNB and FNA, and a P-value <.05 was considered statistically significant.

3 | RESULTS

3.1 | Literature search and selection

A systematic search of PubMed/MEDLINE, Embase, and the Cochrane Library identified 443 studies, and after de-duplication, 322 studies remained. Of these, 311 studies were excluded by reading titles and abstracts. Upon full-text review, a further four articles were removed due to insufficient information for setting up a 2-by-2 contingency table and one article was excluded for its unequal and small sample size, which was not appropriate for performing meta-analysis (Figure 1). Finally, six studies, describing a total of 1924 procedures, were included in this meta-analysis.10,14,16-19 The basic data of these studies are shown in Table 1.

3.2 | Quality assessment of the studies

Risk of bias and applicability concerns for all included studies are summarized in Table 2, according to the QUADAS-2 criteria.21 All of the articles had unknown or high risk of bias for a reference standard, because none of these studies clearly suggested that the standard results were interpreted without knowledge of the index test results. There was one study that showed a high risk for patient selection bias, because patients were identified from a pathology database; this could not ensure consecutiveness.16

3.3 | Nondiagnostic results of CNB and FNA

The data extracted from the studies that compared the diagnostic performance of CNB and FNA for the diagnosis of malignant salivary gland masses are summarized in

---

**Table 2** Methodological quality of the included studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>Risk of bias</th>
<th>Concerns about applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient selection</td>
<td>Index test</td>
</tr>
<tr>
<td>Huang et al14</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Song et al17</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Eom et al10</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Haldar et al16</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Novoa et al18</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Park et al19</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Table 3** Summarized data of included studies

<table>
<thead>
<tr>
<th>Reference</th>
<th>CNB</th>
<th>FNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diag</td>
<td>Nondiag</td>
</tr>
<tr>
<td>Huang et al14</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>Song et al17</td>
<td>222</td>
<td>6</td>
</tr>
<tr>
<td>Eom et al10</td>
<td>247</td>
<td>8</td>
</tr>
<tr>
<td>Haldar et al16</td>
<td>300</td>
<td>13</td>
</tr>
<tr>
<td>Novoa et al18</td>
<td>104</td>
<td>7</td>
</tr>
<tr>
<td>Park et al19</td>
<td>40</td>
<td>4</td>
</tr>
</tbody>
</table>

Abbreviations: CNB, core needle biopsy; Diag, diagnostic; FN, false negative; FNA, fine-needle aspiration; FP, false positive; TN, true negative; TP, true positive.

*aAll data are number of cases.*
Table 3. In comparison with FNA, CNB was associated with statistically significant fewer nondiagnostic results, with a pooled RR of 0.85 (95% CI: 0.82-0.88, \( P < .001 \)). There was evidence of significant heterogeneity \( (I^2 = 96.8\%, \chi^2 P < .001) \). The forest plot of the RRs of nondiagnostic results is presented in Figure 2.

### 3.4 | Diagnostic performance of CNB and FNA

The pooled sensitivity and specificity of CNB for the diagnosis of malignant salivary gland neoplasms were 92% (95% CI: 88%-96%) and 100% (95% CI: 99%-100%), respectively.
respectively. The pooled sensitivity and specificity of FNA were 65% (95% CI: 55%-73%) and 97% (95% CI: 94%-98%), respectively. The sensitivity and specificity of CNB were significantly superior to those of FNA ($P < .001$, and $P = .002$, respectively) (Figure 3).

Comparing the SROC curves, the AUC values of CNB and FNA were 1.00 (95% CI: 0.99-1.00) and 0.80 (95% CI: 0.76-0.83), respectively, which suggests an overall higher diagnostic performance of CNB ($P = .049$, Figure 4).

![SROC curves comparison](image)

**FIGURE 4** Summarized receiver operating characteristic (SROC) curves of (A) core needle biopsy (CNB) and (B) fine-needle aspiration (FNA). SROC of CNB had a significantly higher area under the curve (AUC) compared with FNA (1.00 vs 0.80, $P = .049$), indicating a higher overall diagnostic performance of CNB in the diagnosis of malignant salivary gland neoplasm.

**TABLE 4** Subgroup analysis for fine-needle aspiration results

<table>
<thead>
<tr>
<th></th>
<th>No. of studies</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>$P$ value</th>
<th>$I^2%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needle size$^{a,b}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq 21G$</td>
<td>2</td>
<td>0.71 (0.53, 0.89)</td>
<td>0.95 (0.90, 1.00)</td>
<td>&lt; .001</td>
<td>85</td>
</tr>
<tr>
<td>&lt;21G</td>
<td>2</td>
<td>0.60 (0.49, 0.72)</td>
<td>0.97 (0.95, 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biopsy site</td>
<td></td>
<td></td>
<td></td>
<td>.45</td>
<td>0</td>
</tr>
<tr>
<td>Parotid gland only</td>
<td>2</td>
<td>0.71 (0.53, 0.89)</td>
<td>0.95 (0.90, 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parotid and other salivary glands</td>
<td>4</td>
<td>0.63 (0.53, 0.72)</td>
<td>0.98 (0.96, 0.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guidance$^c$</td>
<td></td>
<td></td>
<td></td>
<td>.73</td>
<td>0</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>4</td>
<td>0.68 (0.56, 0.79)</td>
<td>0.97 (0.95, 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free hand</td>
<td>2</td>
<td>0.61 (0.47, 0.75)</td>
<td>0.97 (0.93, 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performer$^d$</td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>48</td>
</tr>
<tr>
<td>Radiologist</td>
<td>2</td>
<td>0.70 (0.51, 0.88)</td>
<td>0.98 (0.96, 1.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonradiologist</td>
<td>3</td>
<td>0.63 (0.51, 0.75)</td>
<td>0.96 (0.92, 1.00)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval.

$^a$Heterogeneity was suspected in these subgroups ($P < .05$).

$^b$Needle size was not identified in two studies.

$^c$Performer was not identified in one study.
3.5 | Heterogeneity analysis

The results related to CNB did not indicate significant heterogeneity, while the specificity of the FNA demonstrated moderate heterogeneity (\(I^2 = 59.5\%, \chi^2 P = .03\)) (Figure 3). To explore the potential source of heterogeneity of the FNA results, subgroup analysis and meta-regression were conducted, incorporating the following covariates: (a) needle size (≥ 21 G vs <21 G), (b) biopsy site (parotid gland only vs parotid and other salivary glands), (c) guidance (ultrasound vs free hand), and (d) performer (radiologist vs nonradiologist). Among the abovementioned covariates, the cause of heterogeneity was suspected to be the needle size (\(P < .001\)) and performer (\(P = .02\)) (Table 4). Deeks’ funnel plot asymmetry test revealed no significant asymmetry for either CNB or FNA (\(P = .60\) and \(P = .46\), respectively), suggesting no major publication bias (Figure 5).

3.6 | Postprocedural complications

Of the 1015 CNB procedures, hematoma occurred in three cases (0.3%), of which one required surgical intervention (Table 1). No complication was associated with FNA procedures. No major complication, such as tumor seeding or permanent facial paralysis, was reported for either CNB or FNA of salivary gland masses.

4 | DISCUSSION

The main purpose of this meta-analysis was to compare CNB and FNA in terms of their diagnostic performances in the differentiation of malignant and benign salivary gland neoplasia. CNB yielded a significantly lower percentage of nondiagnostic results than did FNA in salivary gland lesions, with a pooled RR of 0.85. Moreover, our meta-analysis revealed superiority of CNB over FNA for detecting malignancy in salivary gland tumors in terms of sensitivity (92% vs 65%, \(P < .001\)) and specificity (100% vs 97%, \(P = .002\)).

FNA is a widely used, safe, technically simple, and low-cost method to diagnose salivary gland tumors. Previous studies evaluating the diagnostic performance of FNA have reported that its specificity is high, ranging from 86% to 100%, while its sensitivity varies from 64% to 90%, which is consistent with our pooled results (sensitivity of 65% and specificity of 97%). In a previous meta-analysis of the diagnostic performance of FNA in the diagnosis of malignant salivary gland lesions, the summary estimates of sensitivity and specificity were 80% and 97%. However, with high heterogeneity, it was difficult to draw a definitive conclusion. In order to improve comparability, our meta-analysis included only studies that compared CNB and FNA in the same centers. As a result, heterogeneity was relatively low except for the specificity of FNA with moderate heterogeneity (\(I^2 = 59.5\%\)), thus facilitating analysis of pooled sensitivity and specificity.

Although many factors were proposed to contribute to heterogeneity in previous studies, our meta-regression results identified needle size (\(P < .001\)) and performer (\(P = .02\)) as potential sources of heterogeneity: a larger-sized needle and an examination performed by a radiologist provided higher diagnostic accuracy, which may have clinical implications.

CNB is known to reduce the probability of inadequate specimens and to increase histopathological diagnostic
accuracy by providing architecture-preserved tissue samples. Consequently, our pooled analysis showed that CNB yielded a significantly lower percentage of nondiagnostic specimens than FNA in salivary gland lesions, with an RR of 0.85. In the diagnosis of malignant salivary gland neoplasm, the sensitivity and specificity of CNB were reported to be 92%-94% and 99%-100% in previous studies, which were in good agreement with our meta-analysis results (sensitivity of 92% and specificity of 100%). In this meta-analysis, CNB yielded a lower proportion of nondiagnostic results and demonstrated superior sensitivity and specificity as compared to FNA for detecting salivary gland malignancies. Therefore, CNB is expected to be more useful than FNA for preoperative diagnosis of salivary masses, helping to avoid unnecessary excision and to plan appropriate surgery.

In this meta-analysis, there was no reported case of procedure-related complications in FNA. CNB caused a small number (0.3%, 3/1015) of hematomas, with one case requiring surgical intervention (0.1%, 1/1015). Although details were not reported meticulously about the case in that report, it was the only study using a large-sized needle (14-16G), as compared to other studies (18-20G for CNB and 21-24G for FNA), which might possibly be a cause of the severe hematoma. In addition to hematomas, major complications, such as needle tract seeding or permanent facial paralysis, have been reported in salivary gland biopsy. Needle tract seeding is theoretically more likely to occur in CNB, which uses larger needles than FNA, and published data suggest that the risk of needle tract seeding are 0.0011% for CNB and 0.00012% for FNA in head and neck tumors. Facial nerve damage after a salivary gland biopsy is reported to be extremely rare and can be avoided with careful biopsy. None of these major complications occurred in the procedures included in this meta-analysis, suggesting that CNB and FNA performed in salivary glands are relatively safe.

There were some potential limitations to this review. First, the major limitation was the relatively small number of included publications that compared the diagnostic performance of CNB and FNA for salivary gland neoplasms from the same institution. However, it is expected that utilizing this inclusion criterion could reduce bias and enhance the comparability between the two biopsy methods. Second, there were limitations inherent to a meta-analysis assessing diagnostic test accuracy, such as selection bias, publication bias, and missing information in some studies. Although Deeks’ funnel plot asymmetry test revealed no major publication bias for either CNB or FNA, this does not completely exclude the possibility of publication bias. In addition, the possibility of publication bias could have been increased by excluding studies not written in English. Third, limited data were available from each study about the diagnostic performance of the tests according to factors such as needle size, number of samples obtained, and characteristics of the performer (eg, department, years of experience in performing biopsy), which could have contributed to the variability of the results. Finally, the follow-up periods for which candidates were considered as benign differed between the included studies. However, the bias might not be large, since the diagnoses in the studies were based on a follow-up of at least 1 year. Finally, our study was not able to include a comparison of societal costs and performer’s preference based on technical difficulty and the disease suspicion level between the CNB and FNA. In the adoption of CNB, a trade-off between the benefits of improved accuracy, and general cost (cost-effectiveness, patient safety, and comfort) should be considered and these issues might be the focus of future studies.

In conclusion, CNB for salivary gland lesions yielded lower rates of nondiagnostic results and demonstrated superior diagnostic performance for detecting malignant tumors, as compared to FNA. Both CNB and FNA performed in salivary glands were relatively safe. Therefore, CNB seems preferable over FNA when diagnosis of salivary gland malignancies is required.

ACKNOWLEDGMENT
We thank Division of Statistics in Medical Research Collaborating Center at Seoul National University Bundang Hospital for reviewing statistical analyses.

ORCID
Junghoon Kim https://orcid.org/0000-0001-7066-8477

REFERENCES


