ORIGINAL ARTICLE

# Risk of malignancy and diagnostic accuracy of salivary gland fine-needle aspiration cytology categorized by the Milan System, with surgical histopathology as the reference standard: a retrospective study from Saudi Arabia

Mohamed Al Essa[1], Talal Alharbi[1], Abdullah Alkhaldi[1], Reem Althwanay[1], Asma Alomary[2*], Enar Alotaibi[3], Laila Aldokhail[3], Abdullah AlKarni[1]

## ABSTRACT

**Purpose:** This study aimed to estimate the risk of malignancy (ROM) of each Milan System for Reporting Salivary Gland Cytopathology (MSRSGC) category and the diagnostic accuracy of fine-needle aspiration cytology (FNAC) categorized by this system, with surgical histopathology as the reference standard.

**Methods:** This retrospective study included 143 patients with preoperative FNAC and surgical excision of a salivary gland lesion at a tertiary center in Riyadh, Saudi Arabia, between 2004 and 2024. ROM was calculated for each MSRSGC category. Diagnostic accuracy used category VI as the positive threshold. Wilson 95% confidence intervals were reported.

**Results:** Histopathology showed 116 benign (81.1%) and 27 malignant (18.9%) lesions. Category IV was the most frequent category (85, 59.4%). Category V was empty. ROM was 0% (95% CI, 0.0-25.9) in category I, 5.9% (1.0-27.0) in category II, 0% (0.0-35.4) in category III, 4.7% (1.8-11.5) in category IV, and 95.7% (79.0-99.2) in category VI. Sensitivity was 81.5% (63.3-91.8), specificity was 99.1% (95.3-99.8), positive and negative predictive values were 95.7% (79.0-99.2) and 95.8% (90.6-98.2), and accuracy was 95.8% (91.1-98.1). Of five missed malignancies, four were in category IV and one in category II.

**Conclusion:** FNAC categorized by the MSRSGC detected malignancy with high specificity and moderate sensitivity. A category VI result was highly predictive of malignancy. Benign and non-neoplastic categories did not exclude it. Small numbers in some categories and the absence of category V limit precision.

**Keywords:** Salivary gland neoplasms, Fine-needle aspiration cytology, Milan System for Reporting Salivary Gland Cytopathology, Risk of malignancy, Diagnostic accuracy, Histopathology.

## Introduction

Salivary gland neoplasms are uncommon and histologically diverse. Most arise in the parotid gland, and most parotid neoplasms are benign. Malignancy is proportionally more frequent in the other glands [1]. Benign and malignant lesions need different treatment. Malignant tumors may require wider resection, sacrifice of the facial nerve, or neck dissection. An accurate

**Correspondence to:** Asma Alomary
*College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia.
**Email:** asmaalomary90@gmail.com
*Full list of author information is available at the end of the article.*
**Received:** 06 October 2026 | **Accepted:** 21 October 2026

OPEN ACCESS
This is an open access article distributed in accordance with the Creative Commons Attribution (CC BY 4.0) license: https://creativecommons.org/licenses/by/4.0/) which permits any use, Share — copy and redistribute the material in any medium or format, Adapt — remix, transform, and build upon the material for any purpose, as long as the authors and the original source are properly cited. © The Author(s) 2026.
AMEM

preoperative diagnosis therefore guides surgical planning and patient counseling.

Fine-needle aspiration cytology (FNAC) is the usual first diagnostic test for a salivary gland mass. It is quick, safe, and suited to the outpatient clinic. Its performance varies between centers. A meta-analysis of parotid FNAC found moderate sensitivity, high specificity, and considerable heterogeneity between studies [2]. Before 2018, there was no uniform terminology for reporting salivary gland cytology. Similar findings were described with different terms in different laboratories. This made reports hard to interpret and results hard to compare [3,4].

The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC) was published in 2018 to standardize reporting [3,4]. It defines six diagnostic categories: I, non-diagnostic; II, non-neoplastic; III, atypia of undetermined significance (AUS); IV, neoplasm, subdivided into IVA (benign neoplasm) and IVB (salivary gland neoplasm of uncertain malignant potential, SUMP); V, suspicious for malignancy; and VI, malignant. Each category is linked to an expected risk of malignancy (ROM) and a management recommendation [4]. The second edition, published in 2023, refined the expected ROM using systematic reviews and meta-analyses. The revised estimates are 15% for category I, 11% for category II, 30% for category III, less than 3% for category IVA, 35% for category IVB, 83% for category V, and more than 98% for category VI [4].

Large registry studies and meta-analyses support the validity of the system [5–7]. Category-specific ROM still varies between institutions and study designs. This supports validation of the system in each practice setting. Published data from Saudi Arabia are limited to a few recent single-center series [1,8].

This study aimed to estimate the ROM of each MSRSGC category and the diagnostic accuracy of FNAC categorized by the MSRSGC, with surgical histopathology as the reference standard, in a tertiary center in Riyadh, Saudi Arabia.

## Subjects and Methods

### *Study design and setting*

This retrospective diagnostic accuracy study was conducted at a tertiary referral center in Riyadh, Saudi Arabia. The index test was preoperative FNAC categorized according to the MSRSGC. The reference standard was the final histopathological diagnosis of the excised specimen. The study is reported in line with the Standards for Reporting of Diagnostic Accuracy Studies 2015 statement [9].

### *Participants*

We reviewed the medical records of all patients who underwent FNAC and surgical excision of a salivary gland lesion in the Department of Otolaryngology, Head and Neck Surgery between 2004 and 2024. Patients were eligible if they had a preoperative FNAC of the lesion, underwent excision at the study center, and had a final histopathological diagnosis. Every included patient therefore had both an FNAC result and a histopathological diagnosis. Patients who did not undergo excision were not eligible, because they had no reference standard. Patients with incomplete medical records were excluded. A non-probability consecutive sampling approach was used, and all eligible patients were included.

All MSRSGC categories were eligible. Non-diagnostic aspirates (category I), lesions that were non-neoplastic on cytology or histology, and lymphomas were included.

### *Fine-needle aspiration cytology and Milan categorization*

FNAC was performed as part of routine clinical care, and the cytology reports were retrieved from the hospital health information system. Data were extracted by the study investigators using a standardized data collection sheet. Patient names and medical record numbers were not collected.

Reports issued with MSRSGC terminology were used as reported. Reports issued before the MSRSGC was introduced in 2018 were assigned to MSRSGC categories from the wording of the original report during data extraction. Slides were not re-reviewed. Category assignment was therefore limited to the information recorded in each report. It was not blinded to the final histopathological diagnosis, and interobserver agreement was not assessed. The implications of this approach are considered in the Discussion.

Category IV aspirates were analyzed as one group because the IVA and IVB subcategories were not recorded separately.

### *Reference standard*

Final diagnoses were taken from the surgical pathology reports. Each lesion was classified as benign or malignant. The benign group included non-neoplastic lesions and benign neoplasms. The malignant group included primary salivary carcinomas, lymphomas, and other malignant tumors.

### *Variables*

Data collected were age at diagnosis, sex, lesion site, lesion size, MSRSGC category, and final histopathological diagnosis. Lesion size was the largest dimension recorded in the medical record.

### *Definition of a positive test*

The primary analysis compared malignant with benign lesions. Category VI was a positive test result. Categories I to IV were negative test results. A true positive (TP) was a category VI aspirate with a malignant final diagnosis. A false positive (FP) was a category VI aspirate with a benign final diagnosis. A true negative (TN) was a category I to IV aspirate with a benign final diagnosis. A false negative (FN) was a category I to IV aspirate with a malignant final diagnosis.

Category VI was chosen as the threshold because it is the only category that gives a definitive cytological diagnosis

of malignancy. Many studies combine categories V and VI as the positive threshold. No aspirate in this series was assigned to category V. The category VI threshold and the combined V and VI threshold therefore gave identical results. Non-diagnostic aspirates were counted as negative in the primary analysis, because no malignant diagnosis was made before surgery. A sensitivity analysis excluded category I aspirates.

### *Statistical analysis*

Data were entered in Microsoft Excel and analyzed with IBM SPSS Statistics, version 22.0 (IBM Corp., Armonk, NY, USA). The normality of continuous variables was checked with the Shapiro-Wilk test, histograms, and quantile-quantile (Q-Q) plots. Normally distributed variables are reported as mean ± standard deviation (SD). Non-normally distributed variables are reported as median and interquartile range (IQR). Categorical variables are reported as counts and percentages. Age was compared between the benign and malignant groups with the independent samples t-test without assuming equal variances. Lesion size was compared with the Mann-Whitney U test. Categorical variables were compared with the Pearson chi-square test or Fisher's exact test, as appropriate.

ROM for each category was the number of lesions that were malignant on histopathology divided by the number of lesions in that category. Only excised lesions were included, so ROM reflects the risk among resected lesions. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated from the 2 × 2 table. Ninety-five percent confidence intervals (CIs) for all proportions were calculated with the Wilson score method [10]. A two-tailed p value below 0.05 was considered statistically significant.

### *Ethical approval*

The study was approved by the Institutional Review Board of King Abdullah International Medical Research Center (KAIMRC), Riyadh, Saudi Arabia (IRB No. 0000079824). The requirement for informed consent was waived because of the retrospective design. Patient confidentiality was protected throughout data collection and analysis. The study was conducted in accordance with the 1964 Declaration of Helsinki and its later amendments.

## Results

### *Study population*

A total of 143 patients were included. Mean age at diagnosis was 44.1 ± 16.3 years. There were 75 women (52.4%) and 68 men (47.6%). Age did not differ significantly between patients with benign and malignant lesions (44.9 ± 15.6 vs. 40.8 ± 19.1 years; p = 0.310). Sex distribution also did not differ (p = 0.224) (Table 1).

### *Final histopathology*

Histopathology showed 116 benign lesions (81.1%) and 27 malignant lesions (18.9%). Pleomorphic adenoma was the most frequent diagnosis (67 lesions, 46.9% of all lesions), followed by Warthin tumor (20, 14.0%) and

***Table 1.*** *Demographic and clinical characteristics of the 143 patients, by final histopathological diagnosis.*

| Characteristic | Total (*n* = 143) | Benign (*n* = 116) | Malignant (*n* = 27) | *p* value |
|---|---|---|---|---|
| Age at diagnosis, years, mean ± SD | 44.1 ± 16.3 | 44.9 ± 15.6 | 40.8 ± 19.1 | 0.310[a] |
| Sex, n (%) | | | | 0.224[b] |
| Female | 75 (52.4) | 58 (50.0) | 17 (63.0) | |
| Male | 68 (47.6) | 58 (50.0) | 10 (37.0) | |
| Lesion site, n (%) | | | | 0.096[c] |
| Parotid gland | 124 (86.7) | 100 (86.2) | 24 (88.9) | |
| Submandibular gland | 15 (10.5) | 14 (12.1) | 1 (3.7) | |
| Sublingual gland | 1 (0.7) | 0 (0.0) | 1 (3.7) | |
| Minor salivary glands | 3 (2.1) | 2 (1.7) | 1 (3.7) | |
| Milan system category, n (%) | | | | < 0.001[c] |
| I Non-diagnostic | 11 (7.7) | 11 (9.5) | 0 (0.0) | |
| II Non-neoplastic | 17 (11.9) | 16 (13.8) | 1 (3.7) | |
| III AUS | 7 (4.9) | 7 (6.0) | 0 (0.0) | |
| IV Neoplasm, benign or SUMP[d] | 85 (59.4) | 81 (69.8) | 4 (14.8) | |
| V Suspicious for malignancy | 0 (0.0) | 0 (0.0) | 0 (0.0) | |
| VI Malignant | 23 (16.1) | 1 (0.9) | 22 (81.5) | |
| Lesion size, cm, median (IQR)[e] | 3.3 (2.5-4.3) | 3.5 (2.4-4.4) | 3.1 (2.5-4.0) | 0.867[f] |

Percentages are column percentages.

[a]Independent samples t-test, equal variances not assumed; [b]Pearson chi-square test; [c]Fisher's exact test; [d]Category IV was not subdivided into IVA (benign neoplasm) and IVB (SUMP); [e]Largest dimension; [f]Mann-Whitney U test

AUS atypia of undetermined significance, SUMP salivary gland neoplasm of uncertain malignant potential, IQR interquartile range, SD standard deviation.

lymphoepithelial cyst (9, 6.3%). The malignant lesions included mucoepidermoid carcinoma (n = 6), adenoid cystic carcinoma (n = 6), acinic cell carcinoma (n = 4), and lymphoma (n = 4). The remaining seven malignancies were rare types, including secretory carcinoma, epithelial-myoepithelial carcinoma, angiosarcoma, and lymphoepithelial carcinoma.

### *Lesion site and size*

The parotid gland was involved in 124 patients (86.7%), the submandibular gland in 15 (10.5%), minor salivary glands in 3 (2.1%), and the sublingual gland in 1 (0.7%). Site did not differ significantly between benign and malignant lesions (p = 0.096). Median lesion size was 3.3 cm (IQR 2.5-4.3). Size was similar in benign lesions (3.5 cm, IQR 2.4-4.4) and malignant lesions (3.1 cm, IQR 2.5-4.0; p = 0.867).

### *Distribution of Milan categories*

Category IV was the most frequent category (85 aspirates, 59.4%). It was followed by category VI (23, 16.1%), category II (17, 11.9%), category I (11, 7.7%), and category III (7, 4.9%). No aspirate was assigned to category V (Table 1; Figure 1). The distribution differed significantly between benign and malignant lesions (p < 0.001). Most benign lesions were in category IV (81 of 116, 69.8%). Most malignant lesions were in category VI (22 of 27, 81.5%).

### *Risk of malignancy*

ROM by category is shown in Table 2 and Figure 2. ROM was highest in category VI (95.7%; 95% CI 79.0-99.2). It was 5.9% (1.0-27.0) in category II and 4.7% (1.8-11.5) in category IV. Categories I and III contained no malignant lesions. Their ROM was 0%, with upper confidence limits of 25.9% and 35.4%, respectively. ROM could not be estimated for category V.

### *Diagnostic performance*

With category VI as the positive threshold, there were 22 TP, 1 FP, 115 TN and 5 FN results (Table 3; Figure 3). Sensitivity was 81.5% (95% CI 63.3-91.8), and specificity was 99.1% (95.3-99.8). PPV was 95.7% (79.0-99.2), and NPV was 95.8% (90.6-98.2). Overall accuracy was 95.8% (91.1-98.1). In the sensitivity analysis without category I aspirates (n = 132), specificity was 99.0% (94.8-99.8), NPV was 95.4% (89.7-98.0), and accuracy was 95.5% (90.4-97.9). Sensitivity and PPV were unchanged.

### *Discordant cases*

One benign lesion was classified as category VI and was the only FP result. Five malignant lesions were placed in lower categories and were the FN results. One was classified as category II, and four were classified as category IV. The histological type of each discordant case was not captured in the extracted dataset.

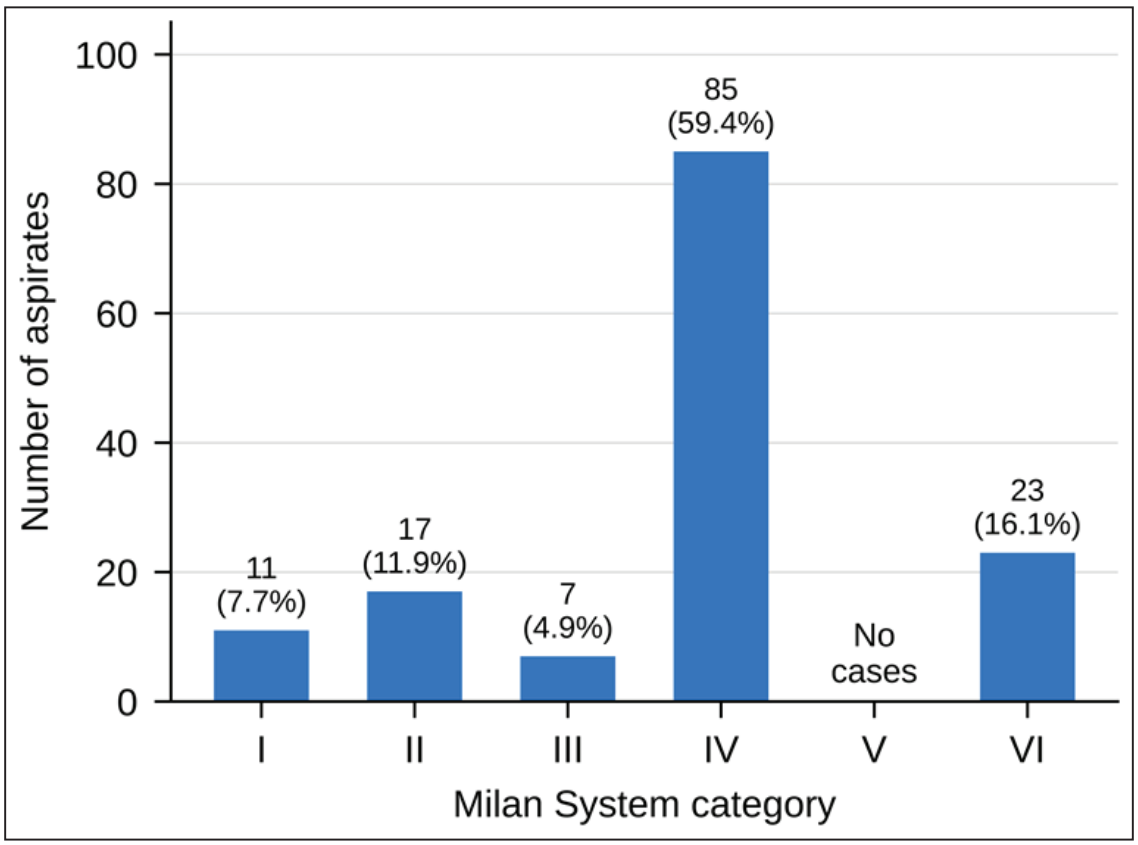



***Figure 1.*** *Distribution of the 143 fine-needle aspirates across Milan System categories. Labels give the number and percentage of aspirates in each category. No aspirate was assigned to category V.*

***Table 2.*** *Risk of malignancy for each Milan System category, with surgical histopathology as the reference standard.*

| Milan System category | Aspirates, n | Malignant on histopathology, n | ROM, % (95% CI)[a] |
|---|---|---|---|
| I Non-diagnostic | 11 | 0 | 0.0 (0.0-25.9) |
| II Non-neoplastic | 17 | 1 | 5.9 (1.0-27.0) |
| III AUS | 7 | 0 | 0.0 (0.0-35.4) |
| IV Neoplasm, benign or SUMP[b] | 85 | 4 | 4.7 (1.8-11.5) |
| V Suspicious for malignancy | 0 | 0 | NA[c] |
| VI Malignant | 23 | 22 | 95.7 (79.0-99.2) |

[a]Wilson score 95% confidence interval; [b]Category IV was not subdivided into IVA (benign neoplasm) and IVB (SUMP); [c]Not applicable: no aspirate was assigned to category V, so ROM could not be estimated.

AUS: atypia of undetermined significance, SUMP: salivary gland neoplasm of uncertain malignant potential, CI: confidence interval, NA: not applicable, ROM: risk of malignancy.

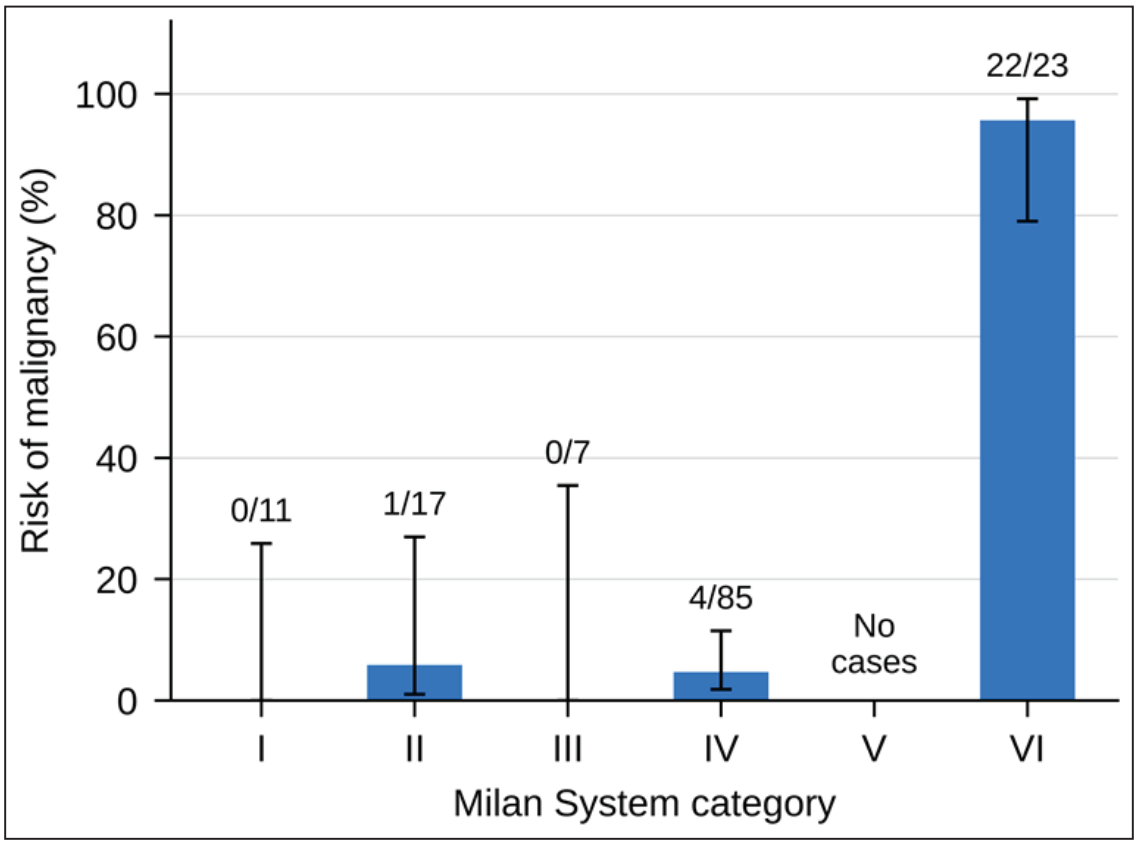



***Figure 2.*** *Risk of malignancy in each Milan System category, with surgical histopathology as the reference standard. Bars show the observed risk, and error bars show Wilson 95% confidence intervals. Labels give the number of malignant lesions over the number of aspirates in each category. No aspirate was assigned to category V.*

## Discussion

In this series, FNAC categorized by the MSRSGC was highly specific for malignancy. A category VI result was confirmed as malignant in 22 of 23 patients. Sensitivity was moderate. About one in five malignancies was placed in a lower category before surgery, most often category IV. These results support the use of the system for preoperative risk stratification. They also show that a benign or non-neoplastic category does not exclude malignancy.

The ROM of category VI (95.7%) was close to the expected value of more than 98% in the second edition of the MSRSGC [4]. It was also close to the pooled estimates of 99% and 97.7% in two meta-analyses [6,7] and to the 99.3% reported in a Dutch registry study of 12,898 parotid aspirates [5]. The lower point estimate in our series reflects one FP result, and its CI includes these values. The ROM of category IV (4.7%) lay between the expected values for benign neoplasm (less than 3%) and SUMP (35%) [4]. Category IV was not subdivided in our data, so this group mixed lesions with very different expected risks. In the Dutch study, ROM was 2.3% for category IVA and 28.6% for category IVB [5]. A meta-analysis reported 2.8% and 37.7% for the same subcategories [7]. A combined ROM of 4.7% is consistent with a group made up mainly of benign neoplasms. It also suggests that a few malignancies with

***Table 3.*** *Diagnostic performance of fine-needle aspiration cytology categorized by the Milan System for malignancy, with category VI as the positive threshold (n = 143).*

| Measure | Definition | n/N | Value, % (95% CI)[a] |
|---|---|---|---|
| Sensitivity | TP/(TP + FN) | 22/27 | 81.5 (63.3-91.8) |
| Specificity | TN/(TN + FP) | 115/116 | 99.1 (95.3-99.8) |
| Positive predictive value | TP/(TP + FP) | 22/23 | 95.7 (79.0-99.2) |
| Negative predictive value | TN/(TN + FN) | 115/120 | 95.8 (90.6-98.2) |
| Diagnostic accuracy | (TP + TN)/N | 137/143 | 95.8 (91.1-98.1) |

TP: true positive (category VI, malignant on histopathology; *n* = 22), FP: false positive (category VI, benign; *n* = 1), TN: true negative (categories I to IV, benign; *n* = 115), FN: false negative (categories I to IV, malignant; *n* = 5).

[a]Wilson score 95% confidence interval.

CI confidence interval.

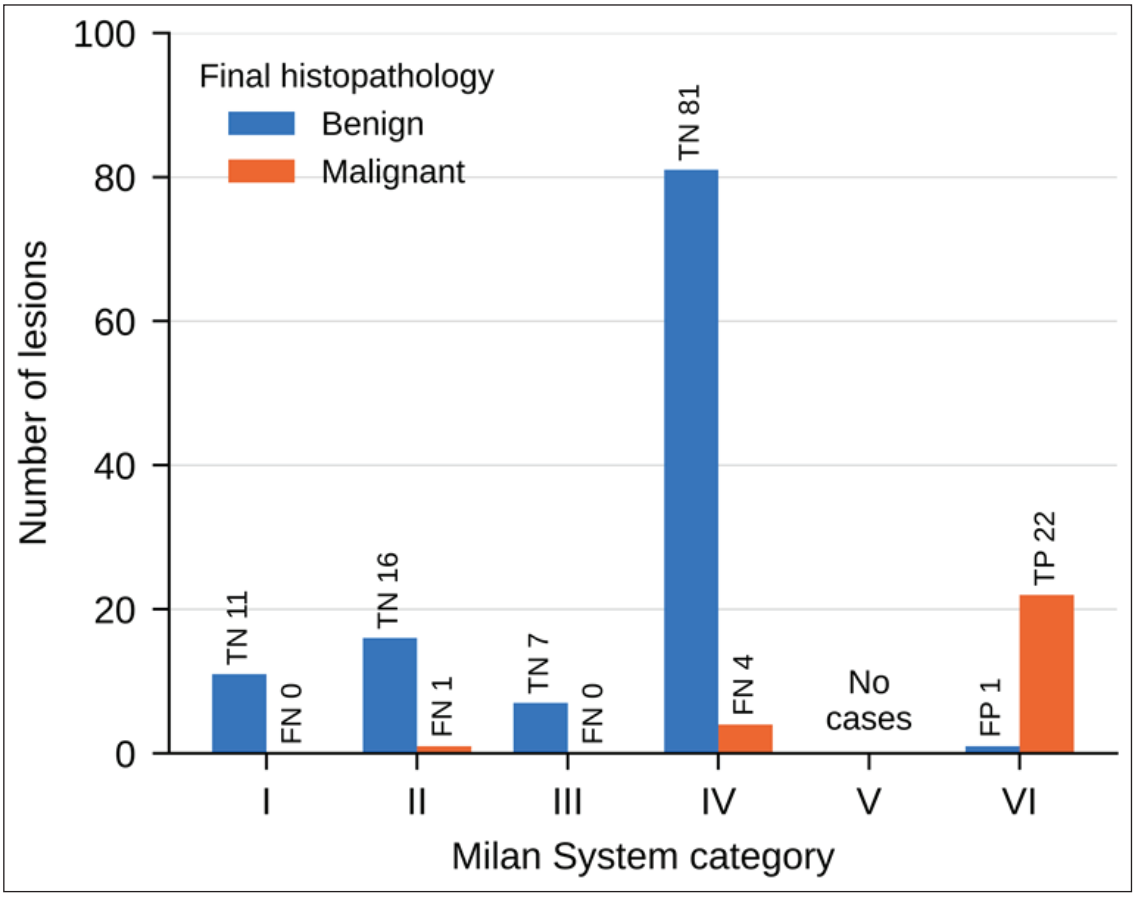



***Figure 3.*** *Final histopathological diagnosis within each Milan System category. Bars show the number of benign and malignant lesions. Labels give the diagnostic outcome with category VI as the positive threshold. TP: true positive, FP: false positive, TN: true negative, FN: false negative.*

bland or SUMP-like features were reported within this group.

The ROM of categories I, II, and III (0%, 5.9%, and 0%) was lower than the expected values of 15%, 11%, and 30% [4]. The Dutch study reported 12.5%, 10.3%, and 29% for these categories [5]. These differences should be interpreted with caution. Categories I to III together contained only 35 aspirates. Their CIs were wide and included the reference values. Selection may also contribute. Only excised lesions were included. Lesions in these categories are often followed up or resampled rather than excised [4]. The excised subset may therefore differ from all aspirates in these categories.

Specificity in our series (99.1%) was similar to published estimates. Sensitivity (81.5%) was lower than in some reports. A meta-analysis of studies using the MSRSGC reported a pooled sensitivity of 88.0% and specificity of 98.5% [7]. A study of 117 patients from western Saudi Arabia reported a sensitivity of 87.9% and specificity of 99.5% [1]. An earlier analysis of 138 patients from our institution reported a sensitivity of 71.4% and specificity of 97.3% [8]. In an Italian surgical series of 554 parotid aspirates, sensitivity was 59%, and specificity was

99% [11]. An earlier meta-analysis of parotid FNAC also found moderate sensitivity and high specificity [2]. Three factors may explain our lower sensitivity. The first is the positive threshold. Some studies count category V, and occasionally category IVB, as positive. A broader threshold raises sensitivity at the cost of specificity. With no category V aspirates and no IVB subcategory, our positive group was limited to definitive malignant diagnoses. The second is case mix. Acinic cell carcinoma, secretory carcinoma, epithelial-myoepithelial carcinoma, and lymphoma accounted for at least 10 of our 27 malignancies. These tumors are often low grade or cytologically bland, and they are harder to recognize on aspiration [11]. The third is precision. Sensitivity was estimated from only 27 malignant lesions. Its CI was wide (63.3%-91.8%) and included most published estimates.

No aspirate was assigned to category V. Other series use this category regularly, with ROM of 83% in the Dutch registry study and 83.8% in a meta-analysis [5,7]. Its absence here may reflect local reporting practice. Borderline aspirates may have been reported as category IV or category VI. Reports issued before 2018 had no suspicious category, and their wording may not have mapped to category V. This has two consequences. ROM could not be estimated for category V. The usual threshold of categories V and VI could not be tested separately from category VI alone. Aspirates that other centers would call suspicious may have been absorbed into category IV, where four malignancies were found.

Missed malignancy is the main clinical risk of salivary gland FNAC. Four of our five FN results were in category IV, and one was in category II. Several mechanisms are recognized. Low-grade carcinomas, such as acinic cell carcinoma, secretory carcinoma, epithelial-myoepithelial carcinoma, and low-grade mucoepidermoid carcinoma, may contain bland cells that resemble a benign neoplasm [4,11]. Cystic tumors may yield fluid with few lesional cells, which can lead to a non-diagnostic or non-neoplastic category. Low-grade lymphomas may resemble reactive lymphoid tissue when flow cytometry or other ancillary tests are not performed [4]. Sampling of only part of a heterogeneous tumor can also miss the malignant component. A recent surgical cohort classified with the updated MSRSGC also reported limitations of FNAC in detecting parotid malignancy [12]. In practice, a benign or non-neoplastic category should not override clinical or radiological suspicion. Rapid growth, pain, facial nerve weakness, fixation or suspicious imaging should prompt repeat FNAC with ancillary testing or core needle biopsy [1,4].

The single FP result also matters clinically. A category VI report may lead to more extensive surgery, such as facial nerve sacrifice or neck dissection. When the clinical and radiological picture does not fit a malignant report, confirmation before radical surgery is reasonable.

Our results support the management framework of the MSRSGC [3,4]. A category VI result supports planning of definitive oncological surgery. A category IV result supports conservative excision, with awareness of a small residual risk of malignancy. Category I and III results call for repeat sampling or further investigation rather than reassurance.

Data on the MSRSGC from Saudi Arabia remain limited. Two recent single-center series reported similar specificity and a sensitivity of 71.4%-87.9% [1,8]. The present cohort overlaps with that earlier analysis, which drew on the same institutional records under the same ethics approval [8]. The two analyses differ in the handling of cases without a recorded Milan category and in histopathological counts. Their estimates should therefore not be combined. Our study adds category-level CIs, the numerator and denominator of each estimate, and a sensitivity analysis for non-diagnostic aspirates.

This study has strengths. Every included aspirate had histopathological confirmation. All estimates are reported with CIs and with their numerators and denominators. The study also has limitations. First, it was a retrospective single-center study. Second, only excised lesions were included. This introduces partial verification bias. ROM and NPV therefore apply to resected lesions and may differ for all aspirates. Third, categories I and III were small, and category V was empty. Estimates for these categories are imprecise or unavailable. Fourth, category IV was not subdivided into IVA and IVB. Fifth, reports issued before 2018 were assigned to Milan categories retrospectively from their wording, and the slides were not re-reviewed. This has several consequences. Free-text reports may not contain the detail needed to separate AUS from a neoplastic category or a benign neoplasm from SUMP. Pre-2018 reports had no suspicious category, which may partly explain the absence of category V. A cytopathologist reviewing the original slides might have assigned some aspirates to a different category. Because the assignment was not blinded to histopathology, it may also have been influenced by the final diagnosis, which would tend to inflate agreement. Interobserver agreement was not assessed. Category-level estimates should therefore be interpreted with caution. Sixth, the histological type of the discordant cases was not available for case-level review. Finally, the sample was too small for analysis by gland or by histological type.

## Conclusion

FNAC categorized by the MSRSGC showed high specificity and moderate sensitivity for salivary gland malignancy in this tertiary center in Saudi Arabia. A category VI result was highly reliable. Lower categories, especially category IV, did not exclude malignancy. Clinical and radiological suspicion should guide further testing when cytology is reassuring. Estimates for categories with few patients, and for the absent category V, remain uncertain. Larger multicenter studies with prospective MSRSGC reporting and slide review, covering all categories including IVA, IVB and V, are needed to refine local estimates.

**Conflict of interest**

The authors have no relevant financial or non-financial interests to disclose.

## Consent to participate

The requirement for informed consent was waived because of the retrospective design and the use of anonymized data. Patient names and medical record numbers were not collected.

## Consent to publish

Not applicable. The manuscript contains no identifiable patient information.


## Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.


## Author contributions

All authors actively participated on study conception and design, data collection, writing, and analysis. Mohammed Alessa supervised the study. All authors contributed to data interpretation and to drafting and revising the manuscript. All authors read and approved the final manuscript.

## Data availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

## Ethical approval

This retrospective chart review was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Declaration of Helsinki and its later amendments. The study was approved by the Institutional Review Board of King Abdullah International Medical Research Center (KAIMRC), Riyadh, Saudi Arabia (IRB No. 0000079824).


## Author details

Mohamed Al Essa[1], Talal Alharbi[1], Abdullah Alkhaldi[1], Reem Althwanay[1], Asma Alomary[2], Enar Alotaibi[3], Laila Aldokhail[3], Abdullah AlKarni[1]

1. Department of Otolaryngology, Head and Neck Surgery, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia
2. College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
3. College of Medicine, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia


*Supplementary content (if any) is available online.*

## References


1. Saharti SN, Aljohani R. Performance characteristics of Milan System for reporting salivary gland cytopathology: an experience from western Saudi population. Diagn Cytopathol. 2026;54(4):264–9. https://doi.org/10.1002/dc.70028
2. Liu CC, Jethwa AR, Khariwala SS, Johnson J, Shin JJ. Sensitivity, specificity, and posttest probability of parotid fine-needle aspiration: a systematic review and meta-analysis. Otolaryngol Head Neck Surg. 2016;154(1):9–23. https://doi.org/10.1177/0194599815607841
3. Rossi ED, Faquin WC. The Milan System for reporting salivary gland cytopathology: the clinical impact so far. considerations from theory to practice. Cytopathology. 2020;31(3):181–4. https://doi.org/10.1111/cyt.12819
4. Rossi ED, Baloch Z, Barkan G, Foschini MP, Kurtycz D, Pusztaszeri M, et al. Second edition of the Milan System for reporting salivary gland cytopathology: refining the role of salivary gland FNA. Cancer Cytopathol. 2024;132(1):10–21. https://doi.org/10.1002/cncy.22753
5. Reerds STH, Van Engen–Van Grunsven ACH, van den Hoogen FJA, Takes RP, Marres HAM, Honings J. Accuracy of parotid gland FNA cytology and reliability of the Milan System for reporting salivary gland cytopathology in clinical practice. Cancer Cytopathol. 2021;129(9):719–28. https://doi.org/10.1002/cncy.22435
6. Kumar A, Chandra S, Singh B, Sharma S, Tandon A, Shahi AK. Assessment of risk of malignancy of fine-needle aspiration cytology in salivary gland lesions using the Milan System for reporting salivary gland cytopathology categorization: a systematic review and meta-analysis. J Contemp Dent Pract. 2022;23(10):1039–56. https://doi.org/10.5005/jp-journals-10024-3424
7. Wang Z, Zhao H, Guo H, An C. Application of the Milan System for reporting salivary gland cytopathology: a systematic review and meta-analysis. Cancer Cytopathol. 2022;130(11):849–59. https://doi.org/10.1002/cncy.22604
8. Al-Essa M. Diagnostic accuracy of the milan classification system for salivary gland cytology: a retrospective validation in a middle eastern tertiary center. Indian J Otolaryngol Head Neck Surg. 2026;78(4):2114–21. https://doi.org/10.1007/s12070-026-06403-0
9. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. https://doi.org/10.1136/bmj.h5527
10. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22(158):209. https://doi.org/10.1080/01621459.1927.10502953
11. Galli A, Tulli M, Giordano L, Biafora M, Di Santo D, Bondi S, et al. Fine needle aspiration cytology for parotid neoplasms: risk of malignancy through inconclusive results and lower grade tumors. Eur Arch Otorhinolaryngol. 2020;277(3):841–51. https://doi.org/10.1007/s00405-019-05733-w
12. Galli A, Howardson BO, Scionti C, Salerno E, Pettirossi C, Di Santo D, et al. Limitations of FNAC in detecting parotid malignancies: insights from a surgical cohort using the updated Milan System. Eur Arch Otorhinolaryngol. 2026. https://doi.org/10.1007/s00405-026-10523-2