Application of the Milan System for Reporting Salivary Gland Cytopathology: A systematic review and meta-analysis

Zhaoyang Wang, MD [1]; Huan Zhao, MD [2]; Huiqin Guo, MD [2,3]; and Changming An, MD [1]

BACKGROUND: The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC) is a standard, evidence-based classification system for salivary gland fine-needle aspiration (SG-FNA). Since it was published in 2018, many researchers across the world have applied this uniform reporting system to their cohorts. **METHODS:** The authors comprehensively reviewed cohort studies conducted since publication of the MSRSGC and performed a meta-analysis. The risk of neoplasm and the risk of malignancy (ROM) were calculated for each diagnostic category, and their diagnostic efficacy was evaluated. **RESULTS:** Thirty-five studies were included in the meta-analysis. The total number of SG-FNAs was 10,706, and 7168 of those had histopathologic follow-up. The ROM for each category was: nondiagnostic, 11.4%; nonneoplastic, 10.9%; atypia of undetermined significance, 30.5%; neoplasm–benign, 2.8%; neoplasm–salivary gland neoplasm of uncertain malignant potential, 37.7%; suspicious for malignancy, 83.8%; and malignant, 97.7%. Low-level heterogeneity was observed in ROM estimation. The sensitivity, specificity, and diagnostic odds ratio for differentiating malignant and benign lesions were 88.0%, 98.5% and 520.3, respectively. **CONCLUSIONS:** The reporting of SG-FNA using the MSRSGC demonstrated high diagnostic accuracy. The ROM for each category was generally concordant with the recommendations, except for the suspicious for malignancy category, which was significantly higher than the reference value. The tiered, standardized classification system would benefit the clinical management of salivary gland lesions. ***Cancer Cytopathol* 2022;130:849-859.**

KEY WORDS: cytology; fine-needle aspiration; Milan System for Reporting Salivary Gland Cytopathology; salivary gland.

INTRODUCTION

The importance of fine-needle aspiration (FNA) is widely accepted for the preoperative evaluation of salivary gland lesions. However, the diverse reporting system for salivary gland FNA (SG-FNA) across institutions has limited its effectiveness. To address this, the American Society of Cytopathology and the International Academy of Cytology organized an international panel to propose a standardized, evidence-based, internationally accepted classification system for SG-FNA. In 2018, the efforts finally produced the Milan System for Reporting Salivary Gland Cytopathology (MSRSGC).[1]

Similar to the Bethesda System for Reporting Thyroid Cytopathology, the MSRSGC comprised six tiered diagnostic categories: (1) *nondiagnostic* (ND), (2) *nonneoplastic* (NN), (3), *atypia of undetermined significance*

Corresponding Author: Changming An, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China (anchangming@cicams.ac.cn); Huiqin Guo, Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China (ghqin2006@163.com).

[1]Department of Head and Neck Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [2]Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [3]Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Hebei Cancer Hospital, Chinese Academy of Medical Sciences, Langfang, China

Received: April 19, 2022; **Revised:** May 10, 2022; **Accepted:** May 12, 2022

Published online May 30, 2022 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cncy.22604, wileyonlinelibrary.com

(AUS), (4a) *neoplasm–benign* (BN), (4b) *neoplasm–salivary gland neoplasm of uncertain malignant potential* (SUMP), (5) *suspicious for malignancy* (SFM), and (6) *malignant* (M). Each category is associated with a risk of malignancy (ROM) and recommended management. Since publication of the MSRSGC, many researchers across the world have applied this uniform reporting system to their cohorts.

In this study, we performed a systematic review and meta-analysis of all available literature since publication of the MSRSGC. Our objective was to summarize the ROM in each MSRSGC category and determine the efficacy and potential factors affecting clinical utility.

MATERIALS AND METHODS

Literature search and selection

We conducted a comprehensive literature search in the PubMed and Web of Science databases from January 2018 to November 2021. Search terms included: *salivary gland* and *MSRSGC* or *Milan system*. All searches were limited to English-language publications. In addition, we reviewed related references from the retrieved articles to identify potentially eligible studies.

After filtering the duplicates, two authors reviewed the title and abstract for the primary screening and, later, reviewed the full text of the selected studies to identify the included literature in the systematic review and meta-analysis. The inclusion criteria were: (1) cohort study with a retrospective or prospective design; (2) focus on the application of MSRSGC to SG-FNA; (3) histopathology was used as the gold standard for evaluating efficacy; (4) patient numbers >50 with histopathologic follow-up; (5) precisely described histopathology diagnoses, which could be categorized and counted as nonneoplastic, benign neoplasm, and malignant neoplasm; and (6) published in English. The exclusion criteria were: (1) articles labeled as reviews, letters, meeting abstracts, commentaries, case reports, or editorials; (2) histopathologic follow-up included <50 patients; (3) histopathology was not the only gold standard used to calculate efficacy; and (4) the research did not provide sufficient data for meta-analysis.

Data extraction and analysis

Two authors first discussed and developed a data-extraction sheet based on 10 randomly selected articles. Then, the two investigators independently extracted data from eligible studies. The data included: first author, year of publication, study design (retrospective or prospective; enrollment), study population (institution, time span, age, sex), cytology procedure (use of imaging guidance, rapid on-site evaluation [ROSE], ancillary tests [such as immunohistochemistry and fluorescence in situ hybridization]), MSRSGC recategorization method (based on reports or slide review; blind design); and total number of FNAs, FNAs with histopathologic follow-up, and relevant categories. Repeat FNAs were documented. The FNA diagnoses and their corresponding histopathology findings were used to tabulate the contingency table.

Quality assessment

Quality assessment among the eligible studies was based on the Quality Assessment of Diagnostic Studies-2 questionnaire concerning patient selection, index test, reference standard, and study flow and timing.

Statistical analysis

The software package Rstudio (R version 4.1.1)[2–4] was used to perform the meta-analysis and to create the figures.

The total numbers of FNAs, FNAs with histopathologic follow-up, each MSRSGC category, and corresponding histopathology diagnoses (nonneoplastic, benign neoplasm, or malignant neoplasm) were summed separately. The diagnostic efficacy of the MSRSGC was evaluated by comparing the preoperative FNA diagnoses with the gold-standard histopathologic follow-up diagnoses. All included literature was analyzed to estimate the risk of neoplasm (RON) and the risk of malignancy (ROM). Some studies documented all FNA cases consecutively but had a partial lack of FNA histopathologic follow-up, and those studies were analyzed to estimate the overall RON (ORON) and the overall ROM (OROM). The RON was calculated as a ratio of neoplastic histopathologic diagnoses between preoperative FNA diagnoses. Similarly, the ROM was calculated between malignant histopathologic diagnoses and FNA diagnoses. The OROM and ORON were calculated between histopathologic diagnoses and total FNA numbers with or without histopathologic follow-up. We performed the Shapiro–Wilk normality test on the original rate and on the four transformed rates (logit, log, arcsine, and Freeman–Tukey double-arcsine transformation) according to the estimation method and selected the method closest to the normal distribution to estimate the summary points and corresponding 95%

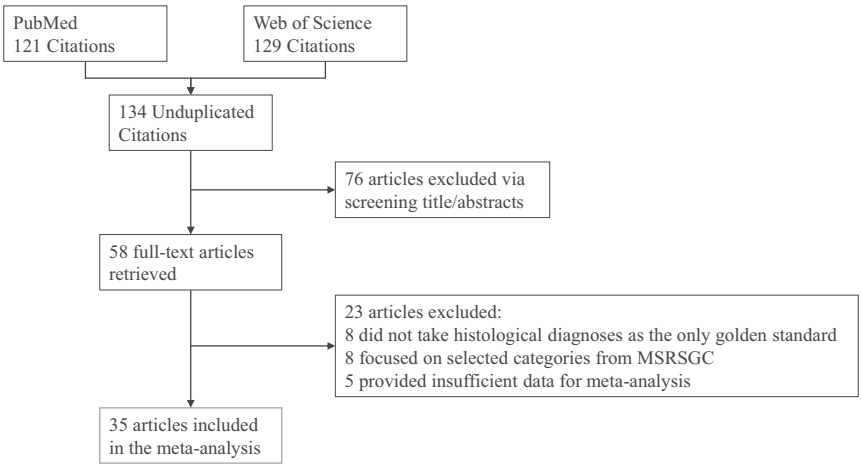

Figure 1. Flow chart for identifying the included literature. MSRSGC indicates the Milan System for Reporting Salivary Gland Cytopathology.

confidence intervals (CIs) of the RON, ROM, ORON, and OROM. The I^2 index of heterogeneity and the Cochrane Q statistic test were used to assess heterogeneity among studies. When a moderate-to-high level of heterogeneity was observed, a random-effects model was used; otherwise, a fixed-effects model was used.

Results that were true positive, true negative, false positive, and false negative were obtained from FNA diagnoses and their corresponding histopathologic follow-up. FNA diagnoses without follow-up were not included in the efficacy evaluation. The sensitivity, specificity, diagnostic odds ratio (DOR), and positive and negative likelihood ratios (PLR and NLR, respectively) in different situations were calculated using a bivariate Bayesian model. To avoid nonlinear functions, a posterior sample of 5000 was specified to compute the margins and estimate values such as the DOR, PLR, and NLR. Forest plots were generated to illustrate the sensitivity and specificity of the MSRSGC. The Spearman correlation test was used to evaluate the presence of threshold effects between sensitivity and specificity. A p value $< .05$ was considered as the presence of threshold effects.

RESULTS

Literature search and selection

After removing the duplicates, we yielded 134 articles (Figure 1). After reviewing the titles and abstracts, 54 articles were excluded. The full texts of the remaining 80 articles were reviewed, and 35 articles[5–39] were finally included in the systematic review and meta-analysis.

Characteristics of included studies

The characteristics of the included studies are summarized in Table 1. The 35 included articles comprised 10,607 patients with 10,706 FNA samples, of which four studies reported repeat FNA.[9,15,25,39] In 7168 FNAs, the diagnoses had histological follow-up. Twenty-one studies[8–10,12–17,19,20,23–26,29,31,33,34,38,39] comprised 6366 consecutive FNA samples, including 2828 cases with histologic follow-up. As reported in the demographic data, the mean patient age was 54.4 years, and the female-to-male ratio was 0.91:1. FNA sites were recorded in 29 studies: 7106 FNAs (81.5%) were from parotid glands, 1311 (15.0%) were from submandibular glands, and 305 (3.5%) were from sublingual and minor salivary glands. The time span of included cases ranged from 2000 to 2020, and the cumulative follow-up time was 221 years.

Two studies were performed prospectively,[11,31] whereas the rest were retrospective except for one multicenter study that assigned cases prospectively or retrospectively.[36] According to the research origin (sorted by the International Monetary Fund[40]), 19 studies were from developed countries, and 16 were from developing countries.

In 17 studies, all FNAs or selected FNAs were performed under image guidance, whereas one study clearly mentioned that no image guidance was applied. The use of ancillary tests for the selected cases was reported in nine studies, and ROSE was available in nine studies. Fifteen studies demonstrated that they re-categorized the presurgery FNA diagnoses into MSRSGC categories by reviewing the slides completely or partially when

TABLE 1. Characteristics of the included literature

Characteristic	No.	Characteristic	No.
Included studies	35	Re-categorized by slides review	
Only surgical cases included	14	Yes	15
Consecutive FNA cases included	21	No or unclear	20
Total no. of patients	10,607	FNA with image guidance	
Mean age, years	54.4	Yes	17
Female-to-male ratio	0.91:1	No or unclear	18
Total FNA	10,706	Implementation of MSRSGC in clinic	
FNA with follow-up	7168	Yes	7
Cumulative follow-up, years	221	No or unclear	28
FNA sites, no. (%)		Study design	
Parotid	7106 (81.5)	Retrospective	32
Submandibular	1311 (15.0)	Prospective	2
Others	305 (3.5)	Mixed	1
Territories		Setting	
Developed countries	19	Single-center	33
Developing countries	16	Multicenter	2
ROSE applied		Ancillary studies	
Yes	9	Yes	9
No or unclear	26	No or unclear	26

Abbreviations: FNA, fine-needle aspiration; MSRSGC, the Milan System for Reporting Salivary Gland Cytopathology; ROSE, rapid on-site evaluation.

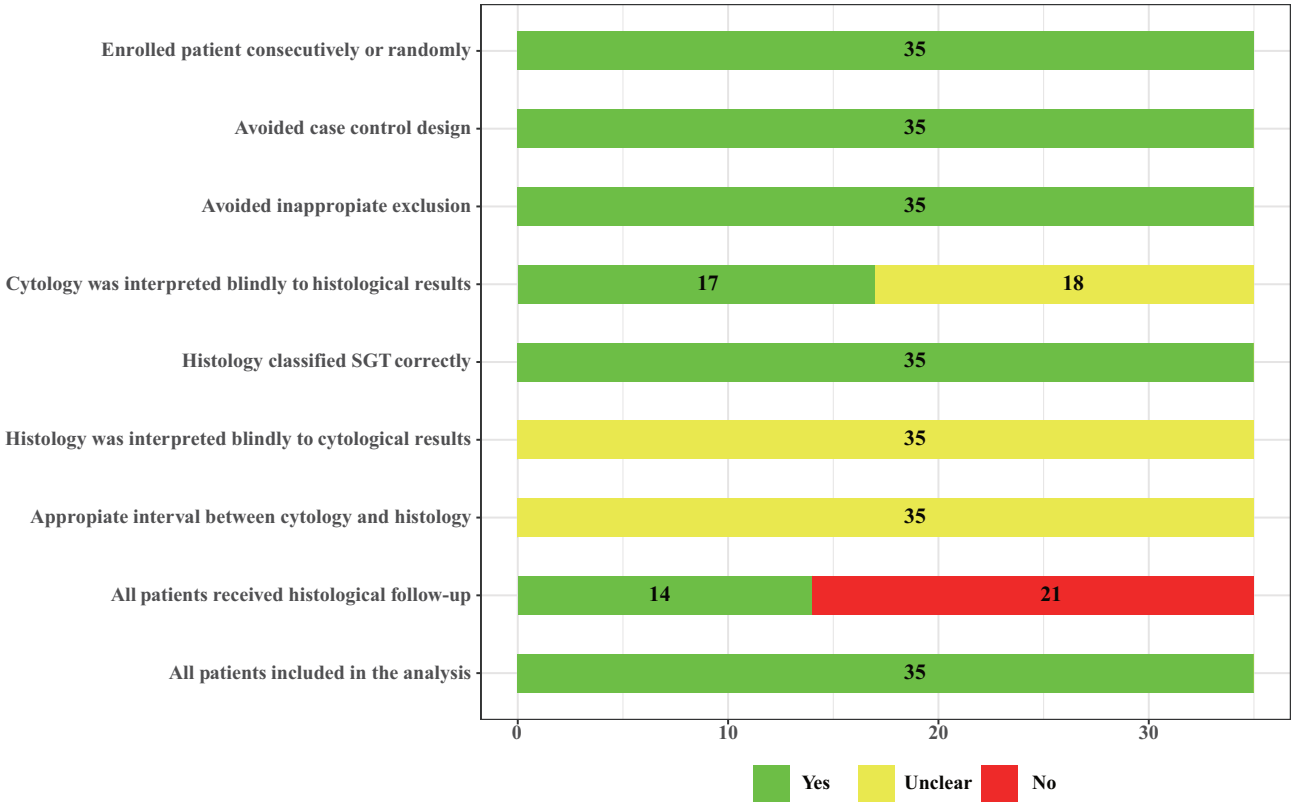

Figure 2. Summary of the QUADAS-2 questionnaire among 35 studies. QUADAS indicates Quality Assessment of Diagnostic Studies; SGT, salivary gland tumor.

necessary. The other 20 studies did not mention the detailed approach for re-categorization or only reviewed the FNA reports. Seven studies[5,9,11,15,31,36,39] reported that the MSRSGC classification had been implemented in their centers.

Figure 2 summarizes the overall result of the Quality Assessment of Diagnostic Studies-2 questionnaire of the included studies. None of the studies were considered to have a high risk of bias in four evaluation domains.

Data synthesis and meta-analysis

The 7168 FNAs that had corresponding histologic follow-up are summarized in Table 2. In total, 811 FNAs (11.3%) were classified as ND, 549 (7.7%) were classified as NN, 498 (6.9%) were classified as AUS, 3371 (47%) were classified as BN, 708 (9.9%) were classified as SUMP, 278 (3.9%) were classified as SFM, and 953 (13.3%) were classified as M. The estimated summary points of ROM and RON in each of the MSRSGC categories are shown in Table 2.

Table 3 lists the 21 consecutive FNA studies, which comprised 6366 FNA cases and 2828 with histologic follow-up. The surgical rate among the salivary gland lesions was 44.4%. The OROM and the ORON are detailed in Table 3.

To evaluate the capability of differentiating salivary gland neoplasms from nonneoplastic lesions, neoplastic lesions in the BN, SUMP, SFM, and M categories were considered true positive, and nonneoplastic lesions in the NN category were considered true negative (Situation 1). The combined forest plot is shown in Figure 3. The sensitivity and specificity were 99.2% (95% CI, 98.8%-99.5%) and 74.8% (95% CI, 66.8%-82.1%), respectively. Then, we added the category AUS as the positive index test for Situation 2; the sensitivity, specificity, DOR, PLR, and NLR are summarized in Table 4.

Next, we tested the accuracy of differentiating benign from malignant lesions. The benign lesions included NN and BN. The following situations were considered in estimating the efficacy: Situation 3, NN and BN as the negative index test and SFM and M as the positive test; Situation 4, NN and BN as the negative index test and SUMP, SFM, and M as the positive test; and Situation 5, NN and BN as the negative index test and AUS, SUMP, SFM, and M as the positive test. The combined forest plots of Situation 3 are provided in Figure 4, and the parameters are summarized in Table 4.

Sixty-nine patients underwent repeat FNA from four studies.[9,15,25,39] The cytologic diagnoses changed in 26 of 69 patients (37.7%) after repeat FNA: 39 cases in the ND category changed 15 times to a conclusive diagnosis (NN, BN, SFM, or M) and five times to an undetermined diagnosis (AUS or SUMP); 23 cases from the AUS and SUMP categories changed five times to a conclusive diagnosis; seven cases from NN, BN, or SFM; and one NN case changed to M.

Subgroup analyses were performed to investigate the possible causes of heterogeneity. We tested six potential

TABLE 2. Summary of all presurgery fine-needle aspiration diagnoses with corresponding postsurgery histologic diagnoses and risk of malignancy/risk of neoplasm across categories of the Milan System for Reporting Salivary Gland Cytopathology classification

Cytology/histology	Total	Nondiagnostic	Nonneoplastic	AUS	Neoplasm-benign	SUMP	Suspicious for malignancy	Malignant
Total no. (%)	7168 (100.0)	811 (11.3)	549 (7.7)	498 (6.9)	3371 (47.0)	708 (9.9)	278 (3.9)	953 (13.3)
Nonneoplastic no. (%)	824	263 (32.4)	400 (72.9)	113 (22.7)	21 (0.6)	23 (3.2)	9 (3.2)	4 (0.4)
Benign neoplasm, no. (%)	4450	409 (50.4)	89 (16.2)	233 (46.8)	3255 (96.6)	418 (59.0)	36 (12.9)	18 (1.9)
Malignant, no. (%)	1874	139 (17.1)	60 (10.9)	152 (30.5)	95 (2.8)	267 (37.7)	233 (83.8)	931 (97.7)
RON [95% CI], %	88.1 [85.0–90.7]	65.6 [55.9–74.8]	24.8 [17.4–34.1]	85.1 [81.7–88.6]	99.4 [99.0–99.6]	96.8 [95.2–97.8]	96.8 [93.9–98.3]	99.9 [99.7–100.0]
ROM [95% CI], %	25.7 [23.2–28.5]	11.4 [8.6–14.4]	10.9 [8.6–13.8]	30.5 [26.6–34.7]	2.8 [2.3–3.4]	37.7 [34.2–41.3]	83.8 [79.0–87.7]	97.7 [96.5–98.5]
I^2 index of heterogeneity, %	88.0	34.0	0.0	0.0	2.0	0.0	0.0	0.0
MSRSGC ROM, %	–	25.0	10.0	20.0	<5.0	35.0	60.0	90.0

Abbreviations: AUS, atypia of undetermined significance; FNA, fine-needle aspiration; MSRSGC, the Milan System for Reporting Salivary Gland Cytopathology; ROM, risk of malignancy; RON, risk of neoplasm; SUMP, salivary gland neoplasm of uncertain malignant potential.

TABLE 3. Summary of the 21 consecutive fine-needle aspiration (FNA) studies showing the total number of FNAs, the proportion of surgery, the overall risk of malignancy, and the overall risk of neoplasm across the categories of the Milan System for Reporting Salivary Gland Cytopathology classification

Cytology/histology	Total	Nondiagnostic	Nonneoplastic	AUS	Neoplasm–benign	SUMP	Suspicious for malignancy	Malignant
Total no. (%)	6366 (100)	1006 (15.8)	1399 (22.0)	404 (6.3)	2212 (34.7)	481 (7.6)	188 (3.0)	676 (10.6)
Follow-up, no.	2828	298	244	197	1216	303	135	435
Surgery proportion, %	44.4	29.6	17.4	48.8	55.0	63.0	71.8	64.3
Nonneoplastic, no.	352	109	173	48	6	9	6	1
Benign neoplasm, no.	1612	126	37	80	1178	167	16	8
Malignant, no.	864	63	34	69	32	127	113	426
ORON [95% CI], %	40.2 [35.2–45.4]	17.7 [13.6–23.1]	3.6 [2.6–4.9]	47.8 [37.9–60.3]	58.4 [51.0–66.9]	61.1 [56.7–65.4]	71.3 [64.7–77.6]	68.5 [61.7–76.1]
OROM [95% CI], %	12.3 [10.4–14.6]	6.3 [4.9–7.9]	2.4 [1.7–3.4]	17.1 [13.7–21.1]	1.4 [1.0–2.0]	22.6 [18.2–27.2]	61.6 [54.5–68.4]	66.7 [59.9–74.3]

Abbreviations: AUS, atypia of undetermined significance; OROM, overall risk of malignancy; ORON, overall risk of neoplasm; SUMP, salivary gland neoplasm of uncertain malignant potential.

modalities in Situations 1 and 3 (Tables 5 and 6). There was no significant improvement in efficacy with the aid of ROSE, ancillary studies, image-guided FNA, or implementation of the MSRSGC, although differences were observed in territories of studies and slide reviews for Situation 1. The studies in developing regions demonstrated higher specificity and DORs. Re-categorization by reviewing the slides had worse specificity. The ratio in the ND group also was examined (Table 7), and the differences were observed in territories and image guidance.

DISCUSSION

SG-FNA has been established as an effective tool for the preoperative evaluation of salivary gland lesions. However, there was lack of a uniform system for reporting SG-FNA, which created obstacles to communication and patient care. This void prompted the development of the MSRSGC. Studies concerning the MSRSGC have increased since it was published in 2018.[1] To investigate application of the MSRSGC in a setting close to the real world, we chose cohort studies that had strict inclusion and exclusion criteria for our meta-analysis. Thirty-five studies were finally included.

Basically, the included studies prospectively or retrospectively collected cases over a period of time; however, not all patients who performed SG-FNA would undergo surgery or histologic follow-up. First, we summarized all cases across the literature to calculate the ROM and RON (Table 2). The ROM values in the ND, NN, AUS, BN, SUMP, SFM, and M categories were 11.4%, 10.9%, 30.5%, 2.8%, 37.7%, 83.8%, and 97.7%, respectively. However, the values of ROM and RON in cohorts that underwent surgery may have been overestimated. Therefore, the OROM and ORON were obtained from consecutive SG-FNA cases as a reference (Table 3). Although using the same classification system could not eliminate observer bias, low-level heterogeneity was observed in ROM calculation in our research. The ROM of each category in this meta-analysis was generally in concordance with MSRSGC recommendations; however, the ROMs of the AUS and SFM categories were higher than the reference values. Specifically, the ROM of the SFM category was 83.8%, which was significantly higher than the reference value of 60%. We believe that the 60% ROM for the SFM category may underestimate the malignancy risk and creates obstacles for clinical management. However, an ROM of approximately 80% may meet clinical requirements, and the studies in our

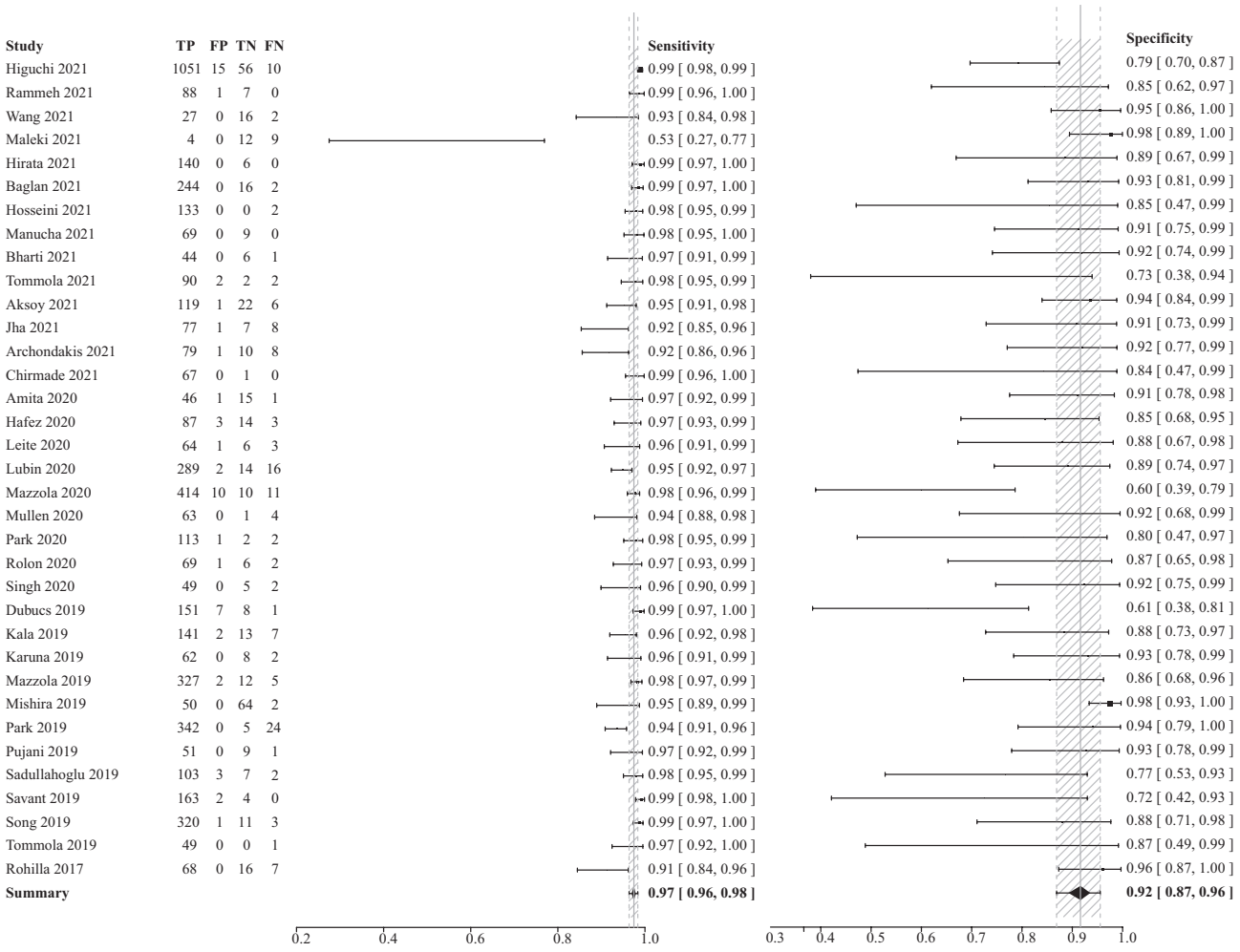

Figure 3. Forest plots of the combined sensitivity and specificity of the Milan System for Reporting Salivary Gland Cytopathology in differentiating salivary glands neoplasms from nonneoplastic lesions. The neoplastic lesions in the categories benign, salivary gland neoplasm of uncertain malignant potential, suspicious for malignancy, and malignant were considered as true positive (TP), and nonneoplastic lesions were considered as true negative (TN) (Situation 1). FN indicates false negative; FP, false positive.

TABLE 4. Diagnostic efficacy of the Milan System for Reporting Salivary Gland Cytopathology in different situations

Variable	Sensitivity (95% CI), %	Specificity (95% CI), %	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)
Situation 1[a]	97.5 (96.4–98.4)	91.6 (86.9–95.6)	12.6 (7.0–23.7)	0.03 (0.02–0.04)	472.0 (234.0–934.2)
Situation 2[b]	97.6 (96.7–98.4)	79.3 (69.3–88.0)	5.1 (3.1–8.7)	0.03 (0.02–0.04)	174.3 (91.0–336.7)
Situation 3[c]	88.0 (84.9–90.9)	98.5 (97.9–99.0)	61.5 (40.3–97.9)	0.12 (0.09–0.16)	520.3 (294.8–902.6)
Situation 4[d]	89.6 (86.7–92.1)	90.3 (88.0–92.3)	9.3 (7.5–11.8)	0.12 (0.09–0.15)	81.6 (60.7–109.4)
Situation 5[e]	90.5 (88.0–92.7)	86.4 (82.9–89.6)	6.8 (5.2–8.8)	0.11 (0.08–0.14)	62.3 (45.8–86.1)

Note: Situations 1 and 2 tested the efficacy for differentiating between neoplastic and nonneoplastic (NN) lesions. Situations 3, 4, and 5 tested the efficacy for differentiating between malignant and benign lesions.

Abbreviations: CI, confidence interval; DOR, diagnostic odds ratio; NLR, negative likelihood ratio; PLR, positive likelihood ratio.

[a]Situation 1: The categories benign neoplasm (BN), salivary gland neoplasm of uncertain malignant potential (SUMP), suspicious for malignancy (SFM), and malignant (M) were considered as the positive index test, with NN as the negative test.

[b]Situation 2: The categories atypia of undetermined significance (AUS), BN, SUMP, SFM, and M were considered as the positive index test, with NN as the negative test.

[c]Situation 3: The categories SFM and M were considered as the positive index test, with NN and BN as the negative test.

[d]Situation 4: The categories SUMP, SFM, and M were considered as the positive index test, with NN and BN as the negative test.

[e]Situation 5:The categories AUS, SUMP, SFM, and M were considered as the positive index test, with NN and BN as the negative test.

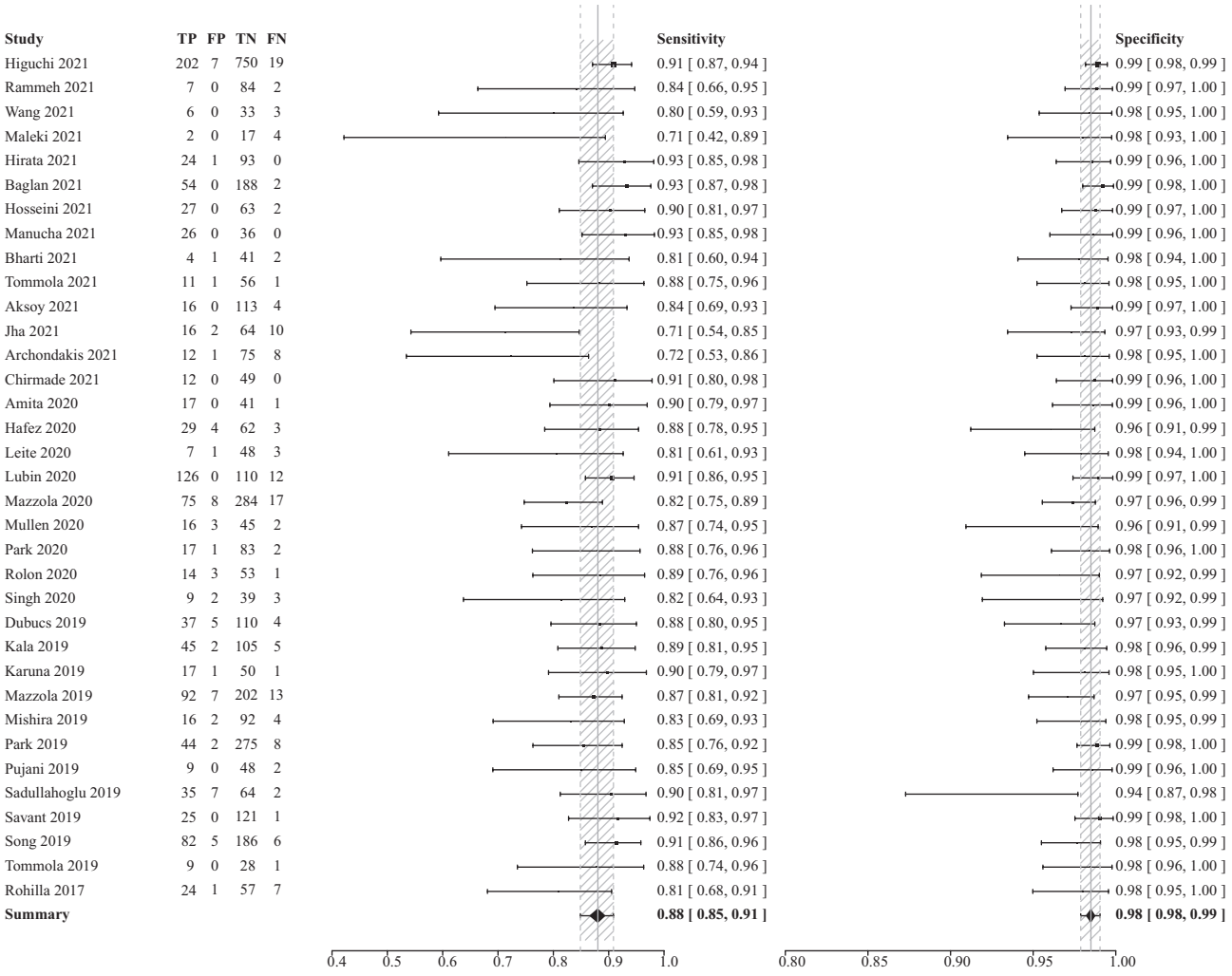

Figure 4. Forest plots of the combined sensitivity and specificity of the Milan System for Reporting Salivary Gland Cytopathology in differentiating malignant from benign salivary gland lesions. The malignant neoplasms in the suspicious for malignancy and malignant categories were considered as true positive (TP), and benign lesions in the nonneoplastic and benign categories were considered as true negative (TN) (Situation 3). FN indicates false negative; FP, false positive.

meta-analysis supported such risk stratification. This should be considered in future revisions of the MSRSGC.

When we considered FNA as an informed test for differentiating between neoplastic and nonneoplastic salivary gland lesions (Situation 1), the sensitivity, specificity, and DOR were 97.5%, 91.6% and 472.0, respectively. However, if we added suspicious diagnoses (AUS) as a positive index test for neoplasm, the specificity and DOR decreased without a relatively significant improvement in sensitivity. The clinical management of AUS in the MSRSGC includes re-biopsy and surgical resection. Our data suggest that a classification as AUS should not be easily considered a neoplastic lesion for surgery because the diagnostic specificity decreased without improved sensitivity. Therefore, we believe that repeat FNA is more reasonable for AUS cases.

In differentiating between malignant and benign masses, when considering definite FNA diagnoses (Situation 3), the specificity and PLR were 98.5% and 61.5, respectively, which indicated that the SFM and M categories of FNA diagnoses had high predictive value for a diagnosis of malignancy. However, the sensitivity of 88.0% and the NLR of 0.12 were moderately informative for ruling out a malignancy diagnosis. As expected, when we added SUMP (Situation 4) or SUMP and AUS (Situation 5) as the positive index test, the specificity and PLR decreased significantly. However, little improvement was observed in sensitivity or NLR. These results indicate that the SUMP and AUS categories should not be considered as SFM in clinical management.

TABLE 5. Subgroup analysis for differentiating between salivary gland neoplastic and nonneoplastic lesions

Situation 1	Sensitivity (95% CI), %	Specificity (95% CI), %	PLR (95%CI)	NLR (95%CI)	DOR (95%CI)
Territories					
Developed	97.6 (96.0–98.7)	86.1 (77.5–93.2)*	7.85 (4.2–15.6)	0.03 (0.01–0.05)	297.8 (117.7–665.0)**
Developing	97.3 (95.4–98.6)	94.8 (90.3–97.9)*	22.6 (9.2–54.3)	0.03 (0.01–0.05)	868.4 (297.7–2224.4)**
Image-guided FNA					
Yes	96.2 (93.9–97.9)	91.7 (84.7–96.6)	13.7 (6.0–31.1)	0.04 (0.02–0.07)	355.7 (131.9–840.0)
No or unclear	98.2 (97.0–99.0)	91.5 (84.2–96.6)	13.9 (6.0–32.0)	0.02 (0.01–0.03)	730.4 (266.9–1823.6)
ROSE					
Yes	95.5 (91.5–98.0)	93.1 (84.7–97.9)	18.7 (5.8–52.8)	0.05 (0.02–0.10)	435.9 (105.0–1248.1)
No or unclear	97.9 (96.8–98.8)	90.9 (84.7–95.7)	12.3 (6.1–25.4)	0.02 (0.01–0.04)	559.5 (235.3–1275.3)
Ancillary studies					
Yes	96.9 (93.8–98.7)	86.8 (73.6–95.5)	9.4 (3.5–25.2)	0.04 (0.01–0.07)	300.2 (75.9–884.6)
No or unclear	97.6 (96.3–98.5)	92.9 (87.8–96.7)	15.5 (3.8–31.3)	0.03 (0.02–0.04)	621.1 (267.2–1351.6)
Re-categorized by slides review					
Yes	98.0 (96.5–99.0)	86.2 (76.2–93.7)*	8.2 (3.9–16.7)	0.02 (0.01–0.04)	380.3 (142.8–916.4)
No or unclear	96.9 (95.1–98.3)	94.5 (89.7–97.8)*	21.3 (9.0–50.6)	0.03 (0.02–0.05)	695.7 (247.9–1766.5)
Implementation of MSRSGC in clinic					
Yes	97.9 (95.4–99.2)	91.1 (77.5–97.9)	16.8 (4.0–55.6)	0.02 (0.01–0.05)	872.8 (157.7–3239.5)
No or unclear	97.3 (95.9–98.3)	91.6 (86.2–96.0)	13.0 (6.9–25.7)	0.03 (0.02–0.05)	460.5 (213.4–951.8)

Note: A p value < .05 was considered significant.
Abbreviations: CI, confidence interval; DOR, diagnostic odds ratio; FNA, fine-needle aspiration; MSRSGC, the Milan System for Reporting Salivary Gland Cytopathology; NLR, negative likelihood ratio; PLR, positive likelihood ratio; ROSE, rapid on-site evaluation.
*$p < .01$.
**$p < .05$.

TABLE 6. Subgroup analysis for differentiating salivary gland malignant and benign lesions

Situation 3	Sensitivity (95% CI), %	Specificity (95% CI), %	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)
Territories					
Developed	89.1 (85.0–92.5)	98.6 (97.8–99.2)	67.3 (37.7–121.3)	0.11 (0.07–0.16)	636.0 (287.3–1299.4)
Developing	86.3 (80.7–90.9)	98.4 (97.3–99.2)	59.4 (30.5–118.5)	0.14 (0.09–0.21)	452.8 (183.1–1052.9)
Image–guided FNA					
Yes	86.1 (80.6–90.6)	98.2 (97.9–99.1)	53.9 (28.3–105.5)	0.14 (0.09–0.20)	397.0 (164.4–881.6)
No or unclear	89.4 (85.3–92.8)	98.6 (97.6–99.2)	71.5 (40.3–129.2)	0.11 (0.07–0.16)	695.3 (319.8–1422.4)
ROSE					
Yes	85.5 (77.2–91.5)	97.4 (95.3–98.8)	36.7 (16.8–77.2)	0.15 (0.08–0.24)	272.2 (85.4–686.4)
No or unclear	88.8 (85.1–92.0)	98.7 (98.1–99.2)	74.8 (46.1–128.5)	0.11 (0.08–0.15)	682.0 (345.2–1331.7)
Ancillary studies					
Yes	87.4 (79.8–93.0)	98.1 (96.4–99.2)	53.8 (22.2–122.7)	0.13 (0.07–0.22)	467.6 (130.8–1313.7)
No or unclear	88.1 (84.4–91.3)	98.6 (97.9–99.2)	67.0 (40.1–113.7)	0.12 (0.08–0.16)	578.0 (289.5–1107.7)
Re-categorized by slides review					
Yes	87.7 (82.5–91.9)	98.9 (98.2–99.4)	86.3 (47.3–159.1)	0.12 (0.08–0.18)	738.5 (314.5–1626.5)
No or unclear	88.3 (83.9–91.9)	97.9 (96.9–98.8)	45.5 (26.8–79.4)	0.12 (0.08–0.17)	399.3 (184.9–798.7)
Implementation of MSRSGC in clinic					
Yes	88.8 (81.2–94.3)	99.1 (98.0–99.7)	119.4 (40.4–313.8)	0.11 (0.05–0.20)	1212.9 (278.5–3867.0)
No or unclear	87.8 (84.0–91.0)	98.3 (97.5–98.9)	53.9 (34.2–88.4)	0.12 (0.09–0.17)	447.0 (235.1–823.9)

Note: A p value < .05 was considered significant.
Abbreviations: CI, confidence interval; DOR, diagnostic odds ratio; FNA, fine-needle aspiration; MSRSGC, the Milan System for Reporting Salivary Gland Cytopathology; NLR, negative likelihood ratio; PLR, positive likelihood ratio; ROSE, rapid on-site evaluation.

Many variables were analyzed to investigate heterogeneity among the studies. The studies in developing regions demonstrated higher specificity and DOR in differentiating between neoplastic and nonneoplastic lesions (Situation 1). Moreover, the ratio of ND FNAs was lower in developing regions. This difference could be attributed to the finding that patients from developing regions may have larger or advanced salivary gland lesions; however, the included literature provided insufficient information to verify our conjectures. Some studies demonstrated that ROSE and ancillary studies could improve the accuracy of SG-FNA and reduce the ratio of ND FNAs,[41–43] but this was not observed in our studies.

SG-FNA is a fast, cost-effective, and safe diagnostic method. Most common tumors, such as pleomorphic adenoma, Warthin tumor, and adenoid cystic carcinoma with cytomorphologic features, can be effectively differentiated through FNA. However, the cytomorphologic diversity and the lack of common language in reports

TABLE 7. Subgroup analysis for the nondiagnostic ratio in consecutive fine-needle aspiration studies

ND ratio (no. of studies)	ND ratio [95% CI], %	ND ratio (no. of studies)	ND ratio [95%CI], %
Territories		Ancillary studies	
Developed (11)	19.2 [13.9–25.9]*	Yes (6)	18.7 [10.7–30.1]
Developing (10)	7.1 [4.6–10.7]*	No or unclear (15)	10.2 [6.9–14.7]
Image-guided FNA		Recategorized by slides review	
Yes (4)	16.0 [10.9–22.9]*	Yes (7)	9.3 [4.6–18.1]
No or unclear (17)	7.8 [4.9–12.2]*	No or unclear (14)	13.7 [9.5–19.4]
ROSE		Implementation of MSRSGC in clinic	
Yes (5)	16.9 [12.6–22.2]	Yes (4)	22.3 [10.4–41.5]
No or unclear (16)	10.8 [6.9–16.4]	No or unclear (17)	10.5 [7.6–14.5]

Note: I^2 index of heterogeneity = 94.0%; a p value < .05 was considered significant.
Abbreviations: CI, confidence interval; FNA, fine-needle aspiration; MSRSGC, the Milan System for Reporting Salivary Gland Cytopathology; ND, nondiagnostic; ROSE, rapid on-site evaluation.
*$p < .01$.

may confuse patients and clinicians, which can impede the clinical decision and peer communication.

The current meta-analysis comprehensively reviews the cohort studies that have been performed since publication of the MSRSGC. Reporting SG-FNA results using the MSRSGC demonstrated high diagnostic accuracy and significance in the clinical management of salivary gland masses. The included studies demonstrated ROMs that generally were similar to those in the MSRSGC recommendations with low-level heterogeneity; however, the ROM of the SFM category was higher than the reference value, and we hope these data can be considered in future MSRSGC revisions.

FUNDING SUPPORT

This Beijing Hope Run Special Fund of Cancer Foundation of China: No. LC2018L06: No. LC2020A19. Non-profit Central Research Institute Fund of Chinese Academy of Medical Science:2019-RC-HL-004.

AUTHOR CONTRIBUTIONS

Zhaoyang Wang: Conceptualization, methodology, data curation, investigation, formal analysis, software programming, and writing–original draft. **Huan Zhao:** Data curation, investigation, formal analysis, and writing–review and editing. **Huiqin Guo:** Conceptualization, supervision, and writing–review and editing. **Changming An:** Conceptualization, methodology, funding acquisition, project administration, supervision, and writing–review and editing.

CONFLICTS OF INTEREST

The authors made no disclosures.

REFERENCES

1. Faquin WC, Rossi ED, Baloch Z, et al, eds. The Milan System for Reporting Salivary Gland Cytopathology. Springer International Publishing AG; 2018.
2. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2021. Accessed April, 2022. https://www.R-project.org/
3. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. 2019;22(4):153–160. https://doi.org/10.1136/ebmental-2019-300117
4. Guo J, Riebler A. meta4diag: Bayesian bivariate meta-analysis of diagnostic test studies for routine practice. *J Stat Software*. 2018;83(1):1–31. https://doi.org/10.18637/jss.v083.i01
5. Archondakis S, Roma M, Kaladelfou E. Two-year experience of the implementation of the Milan for Reporting Salivary Gland Cytopathology at a private medical laboratory. *Head Neck Pathol*. 2021;15(3):780–786. https://doi.org/10.1007/s12105-020-01278-1
6. Park JH, Cha YJ, Seo JY, Lim JY, Hong SW. A retrospective cytohistological correlation of fine-needle aspiration cytology with classification by the Milan System for Reporting Salivary Gland Cytopathology. *J Pathol Transl Med*. 2020;54(5):419–425. https://doi.org/10.4132/jptm.2020.06.09
7. Hafez NH, Abusinna ES. Risk assessment of salivary gland cytological categories of the Milan system: a retrospective cytomorphological and immunocytochemical institutional study. *Turk Patoloji Derg*. 2019;36(2):142–153. https://doi.org/10.5146/tjpath.2019.01469
8. Singh S, Singh P, Auplish R, Khanna SP, Verma K, Aulakh SK. Application of Milan System for Reporting of Salivary Gland Pathology and risk stratification: an institutional experience. *J Oral Maxillofac Pathol*. 2020;24(2):266–272. https://doi.org/10.4103/jomfp.JOMFP_6_20
9. Manucha V, Golzalez MF, Akhtar. Impact of the Milan System for Reporting Salivary Gland Cytology on risk assessment when used in routine practice in a real-time setting. *J Am Soc Cytopathol*. 2021;10(2):208–215. https://doi.org/10.1016/j.jasc.2020.08.005
10. Karuna V, Gupta P, Rathi M, Grover K, Nigam JS, Verma N. Effectuation to cognize malignancy risk and accuracy of fine needle aspiration cytology in salivary gland using "Milan System for Reporting Salivary Gland Cytopathology": a 2 years retrospective study in academic institution. *Indian J Pathol Microbiol*. 2019;62(1):11–16. https://doi.org/10.4103/IJPM.IJPM_380_18
11. Mishra S, Ray S, Sengupta M, Sengupta A. A cytohistological correlation in salivary gland swelling with special reference to the proposed Milan system. *Indian J Pathol Microbiol*. 2019;62(3):379–383. https://doi.org/10.4103/IJPM.IJPM_662_17
12. Kala C, Kala S, Khan L. Milan System for Reporting Salivary Gland Cytopathology: an experience with the implication for risk of malignancy. *J Cytol*. 2019;36(3):160–164. https://doi.org/10.4103/JOC.JOC_165_18
13. Amita K, Rakshitha HB, Singh A, Vijay Shankar S. Evaluation of accuracy of Milan System for Reporting Salivary Gland Cytology: review of morphology and diagnostic challenges in each

category. *J Cytol*. 2020;37(1):18–25. https://doi.org/10.4103/JOC.JOC_191_18
14. Chirmade J, Kothari K, Naik L, Agnihotri M. Utility of the Milan System for Reporting Salivary Gland Cytopathology: a retrospective 5 years study. *Diagn Cytopathol*. 2021;49(4):500–508. https://doi.org/10.1002/dc.24697
15. Hosseini SM, Resta IT, Baloch ZW. Diagnostic performance of Milan System for Reporting Salivary Gland Cytopathology: a prospective study. *Diagn Cytopathol*. 2021;49(7):822–831. https://doi.org/10.1002/dc.24748
16. Pujani M, Chauhan V, Agarwal C, Raychaudhuri S, Singh K. A critical appraisal of the Milan System for Reporting Salivary Gland Cytology (MSRSGC) with histological correlation over a 3-year period: Indian scenario. *Diagn Cytopathol*. 2018;47(5):382–388. https://doi.org/10.1002/dc.24109
17. Rivera Rolon M, Schnadig VJ, Faiz S, Nawgiri R, Clement CG. Salivary gland fine-needle aspiration cytology with the application of the Milan system for risk stratification and histological correlation: a retrospective 6-year study. *Diagn Cytopathol*. 2020;48(11):1067–1074. https://doi.org/10.1002/dc.24478
18. Savant D, Jin C, Chau K, et al. Risk stratification of salivary gland cytology utilizing the Milan system of classification. *Diagn Cytopathol*. 2018;47(3):172–180. https://doi.org/10.1002/dc.24063
19. Lubin D, Buonocore D, Wei XJ, Cohen JM, Lin O. The Milan System at Memorial Sloan Kettering: utility of the categorization system for in-house salivary gland fine-needle aspiration cytology at a comprehensive cancer center. *Diagn Cytopathol*. 2020;48(3):183–190. https://doi.org/10.1002/dc.24350
20. Rammeh S, Romdhand E, Ksentini M, et al. Accuracy of fine-needle aspiration cytology in the diagnosis of salivary gland masses according to the Milan reporting system and to an in-house system. *Diagn Cytopathol*. 2021;49(4):528–532. https://doi.org/10.1002/dc.24682
21. Leite AA, Vargas PA, Dos Santos Silva AR, et al. Retrospective application of the Milan System for Reporting Salivary Gland Cytopathology: a cancer center experience. *Diagn Cytopathol*. 2020;48(9):821–826. https://doi.org/10.1002/dc.24464
22. Baglan T, Sak DSD, Ersoz CC, Ceyhan K. Contribution of small tissue biopsy and flow cytometry to preoperative cytological categorization of salivary gland fine needle aspirates according to the Milan System: single center experience on 287 cases. *Diagn Cytopathol*. 2021;49(4):509–517. https://doi.org/10.1002/dc.24698
23. Sadullahoglu C, Tildirim H, Nergiz D, et al. The risk of malignancy according to Milan reporting system of salivary gland fine-needle aspiration with Becton Dickinson SurePath liquid-based processing. *Diagn Cytopathol*. 2019;47(9):863–868. https://doi.org/10.1002/dc.24214
24. Aksoy Altinboga A, Yildirim F, Ahsen H, Kiran MM, Kesici GG, Yuce G. The effectiveness of the Milan system for risk stratification of salivary gland lesions: the 10-year cytohistopathological correlation results of salivary gland FNA cytology at a tertiary center. *Diagn Cytopathol*. 2021;49(8):928–937. https://doi.org/10.1002/dc.24768
25. Mullen D, Gibbons D. A retrospective comparison of salivary gland fine needle aspiration reporting with the Milan System for Reporting Salivary Gland Cytology. *Cytopathology*. 2020;31(3):208–214. https://doi.org/10.1111/cyt.12811
26. Song SJ, Shafique K, Wong LQ, LiVolsi VA, Montone KT, Baloch Z. The utility of the Milan System as a risk stratification tool for salivary gland fine needle aspiration cytology specimens. *Cytopathology*. 2019;30:91–98. https://doi.org/10.1111/cyt.12642
27. Bharti JN, Elhence P, Rao M, Nalwa A, Khera S. Risk stratification by application of Milan System for Reporting Salivary Gland Cytopathology: a tertiary care experience. *Cytojournal*. 2021;18:19. doi:https://doi.org/10.25259/cytojournal_26_2020
28. Hirata Y, Higuchi K, Tamashiro K, et al. Application of the Milan System for Reporting Salivary Gland Cytopathology: a 10-year experience in a single Japanese institution. *Acta Cytol*. 2021;65(2):123–131. https://doi.org/10.1159/000510990
29. Jha S, Mitra S, Purkait S, Adhya KA. The Milan System for Reporting Salivary Gland Cytopathology: assessment of cytohistological concordance and risk of malignancy. *Acta Cytol*. 2021;65(1):27–39. https://doi.org/10.1159/000510720
30. Mazzola F, Gupta R, Luk PP, Palme C, Clark JR, Low THH. The Milan System for Reporting Salivary Gland Cytopathology—proposed modifications to improve clinical utility. *Head Neck*. 2019;41(8):2566–2573. https://doi.org/10.1002/hed.25732
31. Tommola E, Kalfert D, Hasko-Makinen H, Kholova I. The contributory role of cell blocks in salivary gland neoplasm fine needle aspirations classified by the Milan System for Reporting Salivary Gland Cytology. *Diagnostics*. 2021;11(10):1778. https://doi.org/10.3390/diagnostics11101778
32. Park W, Bae H, Park MH, et al. Risk of high-grade malignancy in parotid gland tumors as classified by the Milan System for Reporting Salivary Gland Cytopathology. *J Oral Pathol Med*. 2019;48(3):222–231. https://doi.org/10.1111/jop.12816
33. Maleki Z, Allison DB, Butcher M, Kawamoto S, Eisele DW, Pantanowitz L. Application of the Milan System for Reporting Salivary Gland Cytopathology to cystic salivary gland lesions. *Cancer Cytopathol*. 2021;129(3):214–225. https://doi.org/10.1002/cncy.22363
34. Wang H, Weiss VL, Borinstein SC, et al. Application of the Milan System for Reporting Pediatric Salivary Gland Cytopathology: analysis of histologic follow-up, risk of malignancy, and diagnostic accuracy. *Cancer Cytopathol*. 2021;129(7):555–565. https://doi.org/10.1002/cncy.22415
35. Rohilla M, Singh P, Rajwanshi A, et al. Three-year cytohistological correlation of salivary gland FNA cytology at a tertiary center with the application of the Milan system for risk stratification. *Cancer Cytopathol*. 2017;125(10):767–775. https://doi.org/10.1002/cncy.21900
36. Higuchi K, Urano M, Akiba J, et al. A multi-institutional study of salivary gland cytopathology: application of the Milan System for Reporting Salivary Gland Cytopathology in Japan. *Cancer Cytopathol*. 2022;130(1):30–40. https://doi.org/10.1002/cncy.22505
37. Mazzola F, Tomasoni M, Mocellin D, et al. A multicenter validation of the revised version of the Milan System for Reporting Salivary Gland Cytology (MSRSGC). *Oral Oncol*. 2020;109:104867. https://doi.org/10.1016/j.oraloncology.2020.104867
38. Dubucs C, Basset C, D'Aure D, Courtade-Saidi M, Evrard SM. A 4-year retrospective analysis of salivary gland cytopathology using the Milan System for Reporting Salivary Gland Cytology and ancillary studies. *Cancers (Basel)*. 2019;11(12):1912. https://doi.org/10.3390/cancers11121912
39. Tommola E, Tommola S, Porre S, Kholova I. Salivary gland FNA diagnostics in a real-life setting: one-year-experiences of the implementation of the Milan System in a tertiary care center. *Cancers (Basel)*. 2019;11(10):1589. https://doi.org/10.3390/cancers11101589
40. International Monetary Fund (IMF). World Economic and Financial Surveys. World Economic Outlook. Database—WEO Groups and Aggregates Information. IMF; 2020. Accessed April, 2022. https://www.imf.org/external/pubs/ft/weo/2020/01/weodata/groups.htm
41. Barats R. Evrard S, Collin L, Vergez S, Gellee S, Courtade-Saidi M. Ultrasound-guided fine-needle capillary cytology of parotid gland masses coupled with a rapid-on-site evaluation improves results. *Morphologie*. 2018;102(336):25–30. https://doi.org/10.1016/j.morpho.2017.06.003
42. Farahani SJ, Baloch Z. Retrospective assessment of the effectiveness of the Milan System for Reporting Salivary Gland Cytology: a systematic review and meta-analysis of published literature. *Diagn Cytopathol*. 2019;47(2):67–87. https://doi.org/10.1002/dc.24097
43. Kakkar A, Kumar M, Subramanian P, et al. Utility of the Milan System for Reporting Salivary Gland Cytopathology during rapid on-site evaluation (ROSE) of salivary gland aspirates. *Cytopathology*. 2021;32(6):779–788. https://doi.org/10.1111/cyt.13038