Received: 27 November 2023 | Revised: 11 February 2024 | Accepted: 5 March 2024

DOI: 10.1002/cncy.22815

ORIGINAL ARTICLE

How the Milan System for Reporting Salivary Gland Cytopathology works in cytopathology practice: Meta-analysis of prospective studies and comparison with retrospective studies

Henri Lagerstam BMed[1,2] | **David Kalfert MD, PhD**[3] | **Zahra Maleki MD, MIAC**[4] | **Ivana Kholová MD, PhD, MIAC**[1,2]

[1]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

[2]Pathology, Fimlab Laboratories, Tampere, Finland

[3]Department of Otorhinolaryngology and Head and Neck Surgery, First Faculty of Medicine, University Hospital Motol, Charles University, Prague, Czech Republic

[4]Department of Pathology, Johns Hopkins Hospital, Baltimore, Maryland, USA

Correspondence
Ivana Kholová, Pathology, Fimlab Laboratories, Arvo Ylpön Katu 4, Tampere 33520, Finland.
Email: ivana.kholova@sll.fimnet.fi

Funding information
Charles University Cooperatio Program; Pirkanmaa Hospital District

Abstract

Background: The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC) is widely accepted and endorsed by professional societies. Although several studies focusing on the MSRSGC have been published, few have been prospective studies. The objective of this study was to evaluate the effectiveness of the MSRSGC in cytopathology practice.

Methods: A comprehensive literature search was conducted to identify all prospective studies on the MSRSGC. The risk of malignancy (ROM), risk of neoplasm, and diagnostic accuracy for each diagnostic category were calculated. Data were tabulated in Microsoft Excel, and analyses were performed with the Open Meta-Analyst program.

Results: Seven prospective and seven retrospective studies were identified. The total number of fine-needle aspirations (FNAs) was 1587 in the prospective studies and 1764 in the retrospective studies. The ROM values for the nondiagnostic, nonneoplastic, atypia of undetermined significance, benign neoplasm, salivary gland neoplasm of uncertain malignant potential, suspicious for malignancy, and malignant categories in prospective versus retrospective studies were 21.0% versus 26.6%, 9.4% versus 8.1%, 34.9% versus 39.6%, 2.4% versus 2.1%, 36.6% versus 31.2%, 86.0% versus 66.0%, and 97.0% versus 96.7%, respectively. Sensitivities, specificities, and diagnostic odds ratios were 83.1% (95% confidence interval [CI], 71.1%–90.8%) versus 89.1% (95% CI, 83.6%–92.9%), 98.4% (95% CI, 96.6%–99.3%) versus 94.9% (95% CI, 91.9%–96.9%), and 310.7 (95% CI, 121.2–796.6) versus 218.8 (95% CI, 107.3–438.1).

Conclusions: This meta-analysis indicated that the MSRSGC works well in FNA cytopathology practice and improves diagnostic accuracy in all diagnostic

This analysis was presented as a platform presentation at the 71st Annual Scientific Meeting of the American Society of Cytopathology; November 15–19, 2023; Austin, Texas.

categories. The ROMs of prospective studies were in concordance with the MSRSGC reference values.

KEYWORDS
fine-needle aspiration, meta-analysis, Milan System for Reporting Salivary Gland Cytopathology, salivary gland

INTRODUCTION

Fine-needle aspiration (FNA) is an effective tool for diagnosing salivary gland lesions. FNA is cost-effective and has high sensitivity and specificity, which makes it possible to distinguish malignant cases from nonmalignant cases and neoplastic cases from nonneoplastic cases. Until recently, there was no uniform standardized system for reporting salivary gland FNA specimens. To address this challenge, an international group of salivary gland cytopathology experts came together and published an evidence-based system for categorizing salivary gland FNA specimens in 2018. This classification system is called the Milan System for Reporting Salivary Gland Cytopathology (MSRSGC).[1]

The intention of the MSRSGC was to create a coherent system for reporting salivary gland FNAs to help clinicians communicate more clearly and make better decisions that improve patient care. The MSRSGC consists of six diagnostic categories: (1) nondiagnostic (ND), (2) nonneoplastic (NN), (3) atypia of undetermined significance (AUS), (4) neoplasm, divided into (4a) neoplasm-benign (BN) and (4b) salivary gland neoplasm of uncertain malignant potential (SUMP), (5) suspicious for malignancy (SM), and (6) malignant (MN). FNAs are classified into these categories on the basis of well-defined criteria. Furthermore, each category includes the associated risk of malignancy (ROM) and recommended management. Recommended management strategies include clinical and radiological correlation, clinical follow-up, repeat FNA, and surgery.[1–4]

The MSRSGC is widely used by cytopathologists in clinical practice worldwide, and numerous studies have been published on it since its development. In the present meta-analysis, we conducted a comprehensive literature search to find all prospective studies published on the MSRSGC. Our goal was to study the effectiveness of the MSRSGC in everyday practice in cytopathology laboratories. In addition, we also analyzed matched retrospective studies on the MSRSGC.

MATERIALS AND METHODS

A PubMed literature search was conducted to identify all prospective studies conducted between December 15, 2022, and January 9, 2023, that used the MSRSGC in the classification of routine salivary gland FNAs. The terms used for the PubMed search were "Milan and Salivary gland." Additionally, two relevant meta-analyses were manually examined for potential articles for our analysis.[5,6]

Title and abstract screening were performed to identify articles with potential data for the meta-analysis. The inclusion criteria for the prospective studies were as follows: (1) prospective study, (2) FNAs were classified according to the MSRSGC system, and (3) sufficient data to calculate the ROM and risk of neoplasm (RON). Retrospective studies, reviews, and meta-analyses were excluded. In the full-text readings, studies that failed to differentiate between retrospective and prospective cases were excluded. The prospective study selection process is presented in the form of a flowchart in Figure 1.

To form a control group of retrospective studies, retrospective studies were randomly selected from PubMed (the search was performed between March 13 and 23, 2023) with the same search terms as in the prospective studies search. Our aim was to compare the everyday practices of cytopathology laboratories with retrospective reclassifications by expert cytopathologists. The inclusion criteria for the retrospective studies were similar to those for the prospective studies, which included the same geographical location (three from Europe, two from the United States, and two from India), an approximately identical number of FNAs and surgical follow-up specimens, and the availability of sufficient data to calculate the ROM and RON. The selection process of retrospective studies is presented as a flowchart in Figure 2.

The data were collected with Microsoft Excel for Microsoft 365 MSO (version 2111, build 16.0.14701.20254, 64-bit). The following were tabulated: first author, publication year, study period, geographical area, type of institution, cytological processing method and type of staining, number of FNAs, number of surgical follow-ups, MSRSGC categories, calculated ROM and RON, and number of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) cases. Sensitivity and specificity were calculated for different scenarios, as previously described by Jalaly et al.[5] and Gubbiotti et al.[7] There were three different malignancy scenarios: (1) NN and BN as a negative index and SM and MN as a positive index; (2) NN and BN as a negative index and MN as a positive index for malignancy; and (3) NN as a negative index and SM and MN as a positive index. In addition, one scenario was applied when identifying neoplasms: NN as a negative index and BN, SUMP, SM, and MN as a positive index for neoplasm.

Statistical analysis was performed with the OpenMeta[Analyst] program (http://www.cebm.brown.edu/openmeta/). The pooled data were calculated with a random-effects model and the DerSimonian and Laird method. A random-effects model was used because of the differences in the study population. The presence or absence of

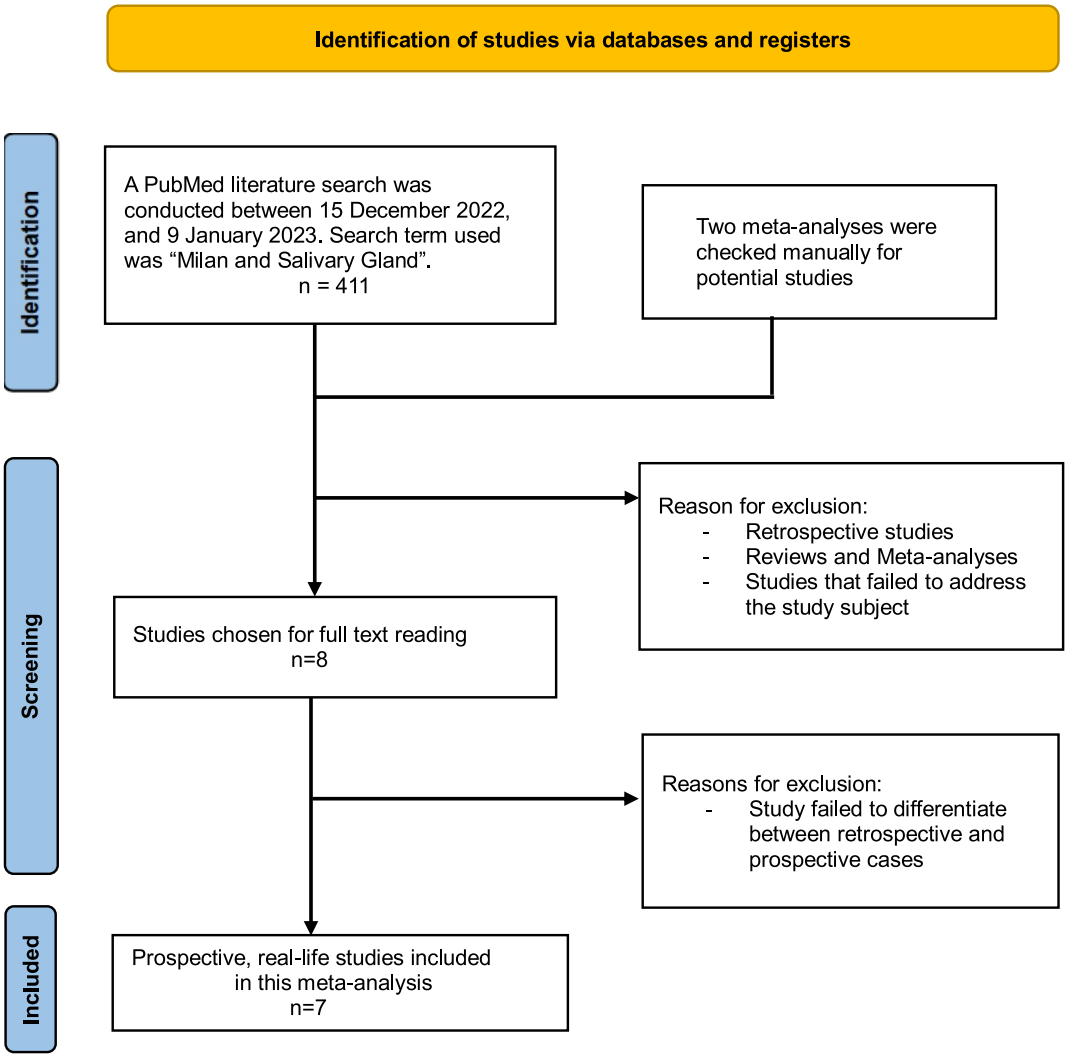

FIGURE 1 Flow diagram describing the process of identifying prospective studies.

heterogeneity was assessed via the I^2 (inconsistency index) and χ^2 statistics. The Egger test was used to identify publication bias, and a forest plot was constructed. For the pooled effect measure, a p value of <.05 was considered statistically significant. Sensitivity and specificity, the positive likelihood ratio and negative likelihood ratio, and the diagnostic odds ratio (DOR) were measured with corresponding 95% confidence intervals (CIs). Our meta-analysis followed the PRISMA 2020 checklist items.[8]

RESULTS

Overview of the studies

Seven prospective studies were identified. Three[9–11] of these were from Europe, two[12,13] were from India, and two[14,15] were from the United States. The studies were published between 2019 and 2022, and together represent 13.8 study years. The prospective studies included 1587 FNAs, with 861 cases (54.3%) involving surgical follow-up. In addition, seven retrospective studies were identified from the same geographical areas (three from Europe,[16–18] two from India,[19,20] and two from the United States[21,22]). The retrospective studies were published between 2018 and 2020, and represent 33 study years. The retrospective studies consisted of 1764 FNAs, with 815 cases (46.2%) involving surgical follow-up. Table 1 provides a comprehensive overview of the included studies.

MSRSGC categories

In summary, Tables 2 and 3 show the classification of FNAs according to the MSRSGC and of surgical follow-ups from the prospective and retrospective studies. Of the 1587 total FNAs in the prospective studies, 278 (17.5%) were classified as ND, 202 (12.7%) as NN, 146 (9.2%) as AUS, 603 (38.0%) as BN, 175 (11.0%) as SUMP, 40 (2.5%) as SM, and 143 (9.0%) as MN. A total of 54.3% (range, 40.1%–100.0%) of cases involved surgical follow-up in the prospective studies. Cases with surgical follow-up in the prospective studies were classified as

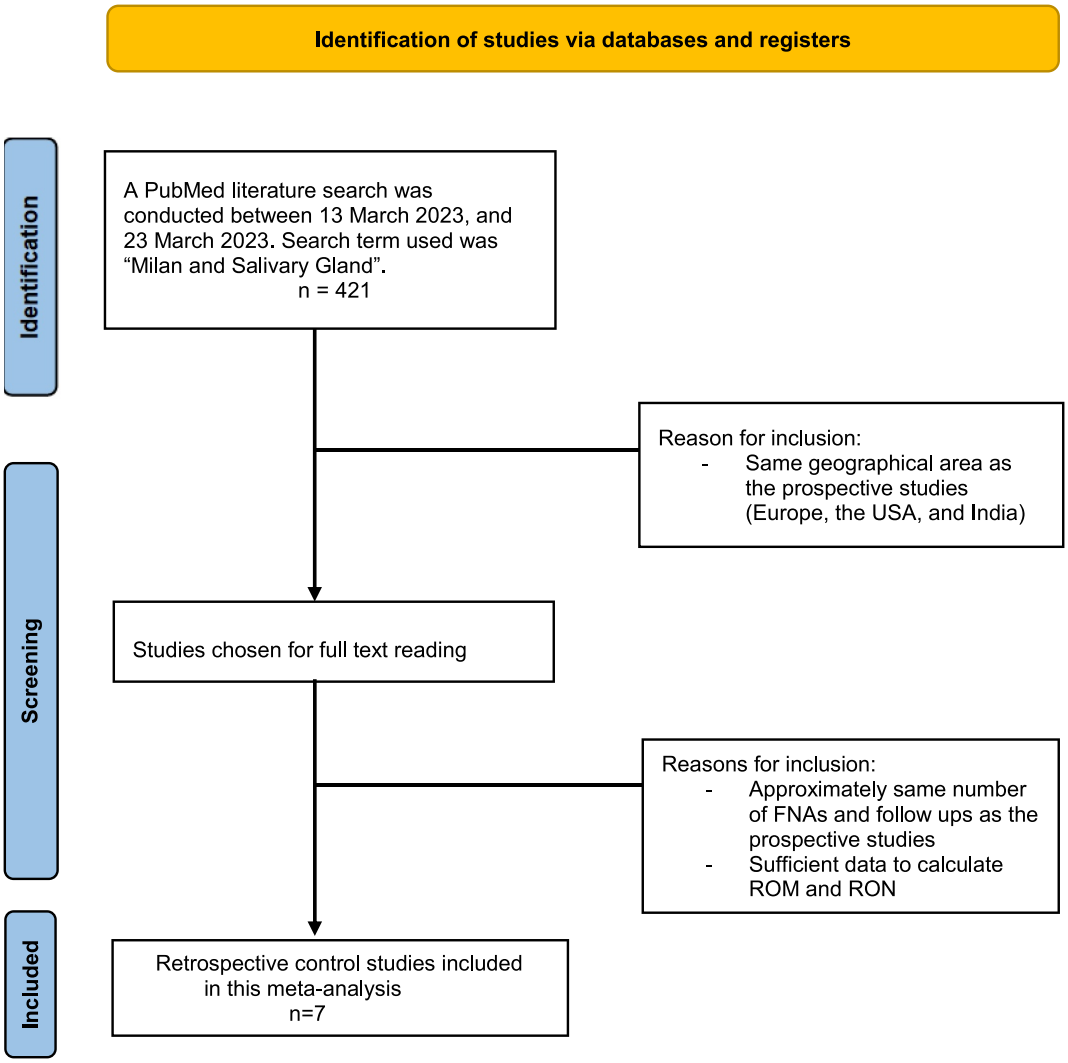

FIGURE 2 Flow diagram describing the process of identifying retrospective studies. FNA indicates fine-needle aspiration; ROM, risk of malignancy; RON, risk of neoplasm.

ND in 113 cases (13.1%), NN in 135 cases (15.7%), AUS in 61 cases (7.1%), BN in 305 cases (35.4%), SUMP in 118 cases (13.7%), SM in 27 cases (3.1%), and MN in 102 cases (11.8%) (Table 2).

In 1764 retrospective study cases, 256 (14.5%) of FNAs were classified as ND, 470 (26.6%) as NN, 57 (3.2%) as AUS, 654 (37.1%) as BN, 72 (4.1%) as SUMP, 78 (4.4%) as SM, and 177 (10.0%) as MN. A total of 46.2% (range, 28.1%–65.9%) of cases involved surgical follow-up in the retrospective studies. Cases with surgical follow-up in the retrospective studies were classified as ND in 76 cases (9.3%), NN in 73 cases (9.0%), AUS in 27 cases (3.3%), BN in 406 cases (49.8%), SUMP in 41 cases (5.0%), SM in 61 cases (7.5%), and MN in 131 cases (16.1%) (Table 3).

ROM

The pooled prevalence of the ROM was calculated. In the prospective studies, the ROM was 21.0% (95% CI, 8.6%–33.5%; I^2, 55.98%) for ND, 9.4% (95% CI, 1.7%–17.0%; I^2, 42.42%) for NN, 34.9% (95% CI, 18.2%–51.5%; I^2, 52.03%) for AUS, 2.4% (95% CI, 0.7%–4.1%; I^2, 0%) for BN, 36.6% (95% CI, 28.0%–45.2%; I^2, 0%) for SUMP, 86.0% (95% CI, 74.0%–98.0%; I^2, 0%) for SM, and 97.0% (95% CI, 93.8%–100.0%; I^2, 0%) for MN.

In the retrospective studies, the pooled prevalence of the ROM was 26.6% (95% CI, 17.1%–36.1%; I^2, 0%) for ND, 8.1% (95% CI, 2.1%–14.0%; I^2, 0%) for NN, 39.6% (95% CI, 21.1%–58.0%; I^2, 18.24%) for AUS, 2.1% (95% CI, 0.7%–3.5%; I^2, 0%) for BN, 31.2% (95% CI, 17.5%–44.8%; I^2, 0%) for SUMP, 66.0% (95% CI, 49.8%–82.2%; I^2, 52.24%) for SM, and 96.7% (95% CI, 93.7%–99.7%; I^2, 0%) for MN. Individual forest plots of the pooled prevalence of ROM are available in Figures S1–S14.

RON

The pooled RON prevalence was calculated. In the prospective studies, the RON was 58.0% (95% CI, 45.8%–70.3%; I^2, 32.5%) for ND, 42.6% (95% CI, 16.3%–69.0%; I^2, 93.6%) for NN, 71.3% (95% CI,

TABLE 1 Overview of the included studies.

Study	Study design	Year of publication	Country	Time period, years	Institution	Cytological method and staining
Archondakis[11]	Prospective	2021	Greece	2.0	Private medical laboratory	Conventional & liquid-based, Papanicolaou & May-Grunwald Giemsa
Dubucs[17]	Retrospective	2019	France	4.0	Department of pathology and cytology, university hospital	Conventional air-dried, May-Grunwald Giemsa
Hollyfield[22]	Retrospective	2018	USA	7.0	Department of pathology, university hospital	Conventional, Papanicolaou alcohol-fixed & Diff-Quik air-dried
Hosseini[14]	Prospective	2021	USA	2.1	Department of pathology, university hospital	Conventional, Papanicolaou alcohol-fixed & Diff-Quik air-dried
Kala[19]	Retrospective	2019	India	5.0	Department of pathology, university hospital	Conventional air-dried, May-Grunwald Giemsa
Karuna[13]	Prospective	2022	India	3.2	Department of pathology, university hospital	Conventional, Papanicolaou alcohol-fixed & Leishman and Giemsa
Manucha[15]	Prospective	2021	USA	2.0	Department of pathology, university hospital	Conventional, Papanicolaou alcohol-fixed & Diff-Quik air-dried
Mishra[12]	Prospective	2019	India	1.5	Department of pathology, university hospital	Cytospin, Papanicolaou alcohol-fixed & Leishman and Giemsa direct air-dried
Mullen & Gibbons[18]	Retrospective	2020	Ireland	5.0	Department of pathology, university hospital	Liquid-based (ThinPrep), Papanicolaou
Pujani[20]	Retrospective	2019	India	3.0	Department of pathology, university hospital	Conventional, Papanicolaou alcohol-fixed & Giemsa air-dried
Rivera Rolon[21]	Retrospective	2020	USA	6.0	Department of pathology, university hospital	Conventional, Papanicolaou alcohol-fixed & Romanowsky
Sadullahoğlu[16]	Retrospective	2019	Turkey	3.0	Department of pathology, university hospital	Liquid-based (SurePath), Papanicolaou
Tommola[9]	Prospective	2019	Finland	1.0	Department of pathology, university hospital	Papanicolaou alcohol-fixed, Cytospin
Tommola[10]	Prospective	2021	Finland	2.0	Department of pathology, university hospital	Papanicolaou alcohol-fixed, Cytospin

60.4%–82.2%; I^2, 0%) for AUS, 98.9% (95% CI, 97.8%–100.1%; I^2, 0%) for BN, 97.8% (95% CI, 95.2%–100.3%; I^2, 0%) for SUMP, 89.3% (95% CI, 78.8%–99.7%; I^2, 0%) for SM, and 97.3% (95% CI, 94.3%–100.4%; I^2, 0%) for MN.

In the retrospective studies, the pooled RON prevalence was 71.7% (95% CI, 54.4%–89.1%; I^2, 59.8%) for ND, 26.8% (95% CI, 12.5%–41.2%; I^2, 58.9%) for NN, 67.4% (95% CI, 48.0%–86.9%; I^2, 34.2%) for AUS, 99.1% (95% CI, 98.2%–100.0%; I^2, 0%) for BN, 87.1% (95% CI, 77.7%–96.5%; I^2, 0%) for SUMP, 92.8% (95% CI, 86.7%–98.9%; I^2, 0%) for SM, and 97.8% (95% CI, 95.5%–100.3%; I^2, 0%) for MN. Individual forest plots of the pooled RON prevalence are available in Figures S15–S28.

Diagnostic accuracy

The diagnostic accuracy of the MSRSGC was evaluated for the detection of malignancies in three different scenarios. In scenario 1, the sensitivity, specificity, and DOR were 83.1% (95% CI, 71.1%–90.8%), 98.4% (95% CI, 96.6%–99.3%), and 310.7 (95% CI, 121.2–796.6) for the prospective studies and 89.1% (95% CI, 83.6%–92.9%), 94.9% (95% CI, 91.9%–96.9%), and 218.8 (95% CI, 107.3–438.1) for the retrospective studies, respectively. The forest plots of the specificity and sensitivity of scenario 1 are displayed in Figures 3 and 4. In scenario 2, the sensitivity, specificity, and DOR were 78.2% (95% CI, 65.5%–87.1%), 98.9% (95% CI, 97.3%–99.6%), and 343.4 (95% CI, 117.9–1000.1) for the prospective studies and 85.5% (95% CI, 78.7%–90.4%), 98.0% (95% CI, 95.9%–99.1%), and 382.5 (95% CI, 148.1–988.4) for the retrospective studies, respectively. In scenario 3, the sensitivity, specificity, and DOR were 88.0% (95% CI, 80.2%–93.0%), 93.7% (95% CI, 86.8%–97.1%), and 123.0 (95% CI, 42.8–353.6) for the prospective studies and 93.1% (95% CI, 87.3%–96.3%), 72.8% (95% CI, 55.5%–85.1%), and 51.1 (95% CI, 17.4–150.3) for the retrospective studies, respectively.

In addition, the diagnostic accuracy of the MSRSGC in the identification of neoplasms was assessed. The sensitivity, specificity, and DOR for the prospective studies were 95.3% (95% CI, 87.8%–98.3%), 84.1% (95% CI, 58.8%–95.2%), and 111.0 (95% CI, 17.0–725.0),

TABLE 2 MSRSGC categorization of FNA cases in prospective studies.

Study	FNAs, No.	Cases with surgical follow-up, No. (%)	Categorization of FNA cases applying the MSRSGC, No. (%)							Categorization of FNA cases with surgical follow-up in the MSRSGC, No. (%)						
			ND	NN	AUS	BN	SUMP	SM	MN	ND	NN	AUS	BN	SUMP	SM	MN
Archondakis[11]	102	102 (100.0)	3 (2.9)	18 (17.6)	1 (1.0)	65 (63.7)	2 (2.0)	1 (1.0)	12 (11.8)	3 (2.9)	18 (17.6)	1 (1.0)	65 (63.7)	2 (2.0)	1 (1.0)	12 (11.8)
Hosseini[14]	328	162 (49.4)	21 (6.4)	24 (7.3)	46 (14.0)	142 (43.3)	57 (17.4)	6 (1.8)	32 (9.8)	8 (4.9)	2 (1.2)	19 (11.7)	63 (38.9)	43 (26.5)	4 (2.5)	23 (14.2)
Karuna[13]	329	132 (40.1)	8 (2.4)	55 (16.7)	29 (8.8)	152 (46.2)	35 (10.6)	11 (3.3)	39 (11.9)	5 (3.8)	35 (26.5)	13 (9.8)	39 (29.5)	16 (12.1)	6 (4.5)	18 (13.6)
Manucha[15]	160	92 (57.5)	30 (18.8)	17 (10.6)	5 (3.1)	59 (36.9)	21 (13.1)	3 (1.9)	25 (15.6)	11 (12.0)	9 (9.8)	3 (3.3)	27 (29.3)	16 (17.4)	2 (2.2)	24 (26.1)
Mishra[12]	119	119 (100.0)	3 (2.5)	66 (55.5)	0 (0.0)	30 (25.2)	2 (1.7)	3 (2.5)	15 (12.6)	3 (2.5)	66 (55.5)	0 (0.0)	30 (25.2)	2 (1.7)	3 (2.5)	15 (12.6)
Tommola[9]	183	90 (49.2)	74 (40.4)	6 (3.3)	20 (10.9)	50 (27.3)	21 (11.5)	6 (3.3)	6 (3.3)	31 (34.4)	1 (1.1)	9 (10.0)	28 (31.1)	12 (13.3)	5 (5.6)	4 (4.4)
Tommola[10]	366	164 (44.8)	139 (38.0)	16 (4.4)	45 (12.3)	105 (28.7)	37 (10.1)	10 (2.7)	14 (3.8)	52 (31.7)	4 (2.4)	16 (9.8)	53 (32.3)	27 (16.5)	6 (3.7)	6 (3.7)
Total	1587	861 (54.3)	278 (17.5)	202 (12.7)	146 (9.2)	603 (38.0)	175 (11.0)	40 (2.5)	143 (9.0)	113 (13.1)	135 (15.7)	61 (7.1)	305 (35.4)	118 (13.7)	27 (3.1)	102 (11.8)

Abbreviations: AUS, atypia of undetermined significance; BN, benign neoplasm; FNA, fine-needle aspiration; MN, malignant neoplasm; MSRSGC, Milan System for Reporting Salivary Gland Cytopathology; ND, nondiagnostic; NN, nonneoplastic; SM, suspicious for malignancy; SUMP, salivary gland neoplasm of uncertain malignant potential.

TABLE 3 MSRSGC categorization of FNA cases in retrospective studies.

Study	FNAs, No.	Cases with surgical follow-up, No. (%)	Categorization of FNA cases applying the MSRSGC, No. (%)							Categorization of FNA cases with surgical follow-up in the MSRSGC, No. (%)						
			ND	NN	AUS	BN	SUMP	SM	MN	ND	NN	AUS	BN	SUMP	SM	MN
Dubucs[17]	328	216 (65.9)	84 (25.6)	27 (8.2)	4 (1.2)	145 (44.2)	15 (4.6)	16 (4.9)	37 (11.3)	47 (21.8)	9 (4.2)	2 (0.9)	105 (48.6)	11 (5.1)	16 (7.4)	26 (12.0)
Hollyfield[22]	134	77 (57.5)	20 (14.9)	31 (23.1)	15 (11.2)	44 (32.8)	7 (5.2)	3 (2.2)	14 (10.4)	8 (10.4)	12 (15.6)	9 (11.7)	26 (33.8)	6 (7.8)	3 (3.9)	13 (16.9)
Kala[19]	293	172 (58.7)	18 (6.1)	112 (38.2)	8 (2.7)	98 (33.4)	6 (2.0)	7 (2.4)	44 (15.0)	4 (2.3)	20 (11.6)	5 (2.9)	90 (52.3)	6 (3.5)	7 (4.1)	40 (23.3)
Mullen & Gibbons[18]	192	73 (38.0)	30 (15.6)	31 (16.1)	1 (0.5)	97 (50.5)	4 (2.1)	3 (1.6)	26 (13.5)	5 (6.8)	5 (6.8)	0 (0.0)	42 (57.5)	2 (2.7)	3 (4.1)	16 (21.9)
Pujani[20]	150	64 (42.7)	7 (4.7)	63 (42.0)	3 (2.0)	65 (43.3)	2 (1.3)	2 (1.3)	8 (5.3)	1 (1.6)	10 (15.6)	2 (3.1)	40 (62.5)	2 (3.1)	2 (3.1)	7 (10.9)
Rivera Rolon[21]	208	84 (40.4)	23 (11.1)	53 (25.5)	11 (5.3)	77 (37.0)	13 (6.3)	7 (3.4)	24 (11.5)	2 (2.4)	8 (9.5)	4 (4.8)	46 (54.8)	7 (8.3)	6 (7.1)	11 (13.1)
Sadullahoğlu[16]	459	129 (28.1)	74 (16.1)	153 (33.3)	15 (3.3)	128 (27.9)	25 (5.4)	40 (8.7)	24 (5.2)	9 (7.0)	9 (7.0)	5 (3.9)	57 (44.2)	7 (5.4)	24 (18.6)	18 (14.0)
Total	1764	815 (46.2)	256 (14.5)	470 (26.6)	57 (3.2)	654 (37.1)	72 (4.1)	78 (4.4)	177 (10.0)	76 (9.3)	73 (9.0)	27 (3.3)	406 (49.8)	41 (5.0)	61 (7.5)	131 (16.1)

Abbreviations: AUS, atypia of undetermined significance; BN, benign neoplasm; FNA, fine-needle aspiration; MN, malignant neoplasm; MSRSGC, Milan System for Reporting Salivary Gland Cytopathology; ND, nondiagnostic; NN, nonneoplastic; SM, suspicious for malignancy; SUMP, salivary gland neoplasm of uncertain malignant potential.

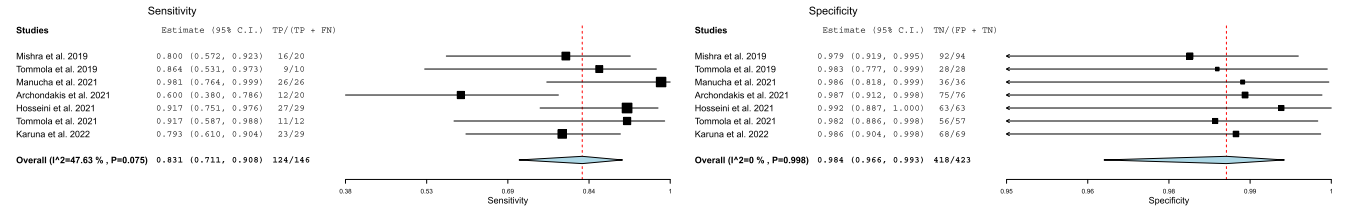

FIGURE 3 Forest plots showing sensitivity and specificity in prospective studies. Nonneoplastic and benign neoplasms were regarded as a negative index, and suspicious for malignancy and malignant were regarded as a positive index for malignancy. Sensitivity was 83.1% (95% CI, 71.1%–90.8%; I^2, 47.6%), and specificity was 98.4% (95% CI, 96.6%–99.3%; I^2, 0%). CI indicates confidence interval; FN, false-negative; FP, false-positive; I^2, inconsistency index; TN, true-negative; TP, true-positive.

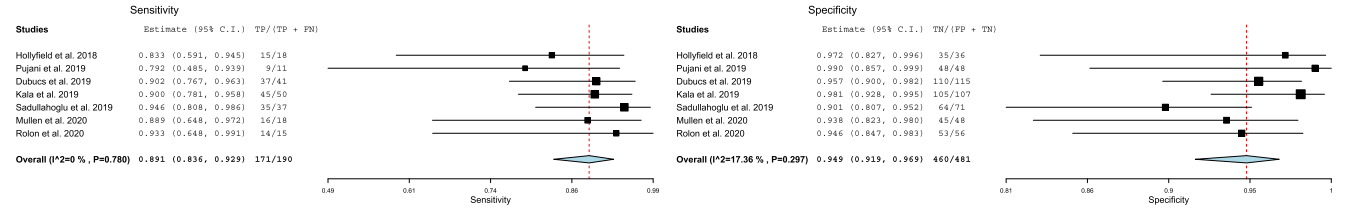

FIGURE 4 Forest plots showing sensitivity and specificity in retrospective studies. Nonneoplastic and benign neoplasms were regarded as a negative index, and suspicious for malignancy and malignant were regarded as a positive index for malignancy. Sensitivity was 89.1% (95% CI, 83.6%–92.9%; I^2, 0%), and specificity was 94.9% (95% CI, 91.9%–96.9%; I^2, 17.4%). CI indicates confidence interval; FN, false-negative; FP, false-positive; I^2, inconsistency index; TN, true-negative; TP, true-positive.

respectively. The sensitivity, specificity, and DOR for the retrospective studies were 96.1% (95% CI, 93.9%–97.6%), 77.0% (95% CI, 61.4%–87.5%), and 148.2 (95% CI, 62.5–351.4), respectively.

An overview of the diagnostic accuracy of the MSRSGC in the identification of malignancies and neoplasms is provided in Table 4. Individual forest plots of diagnostic accuracy are presented in Figures S29–S50.

Prospective versus retrospective studies

The most significant differences in the classification of FNAs between the prospective and retrospective studies were in the NN, AUS, SUMP, and SM categories. The prospective studies contained more FNAs in the AUS (9.2% vs. 3.2%) and SUMP (11.0% vs. 4.1%) categories. On the contrary, the retrospective studies had more than two times as many FNAs in the NN category (12.7% vs. 26.6%) and almost twice as many FNAs in the SM category (2.5% vs. 4.4%).

When the pooled ROM prevalence was calculated, the most notable differences between the prospective and retrospective studies were in the categories ND (21.0%, $p = .034$ vs. 26.6%, $p = .608$), AUS (34.9%, $p = .064$ vs. 39.6%, $p = .295$), SUMP (36.6%, $p = .857$ vs. 31.2%, $p = .790$), and SM (86.0%, $p = .984$ vs. 66.0%, $p = .051$).

The most significant differences between the prospective and retrospective studies in the pooled prevalence of RON were in the ND (58.0%, $p = .180$ vs. 71.7%, $p = .021$), NN (42.6%, $p \leq .001$ vs. 26.8%, $p = .024$), and SUMP (97.8%, $p = .784$ vs. 87.1%, $p = .717$) categories.

In terms of diagnostic accuracy in detecting malignancies, the prospective studies had more FN cases than the retrospective studies. However, the sensitivity was higher in all scenarios in the retrospective studies. In scenario 1, the sensitivities were 83.1% ($p = .075$) and 89.1% ($p = .780$), in scenario 2 they were 78.2% ($p = .134$) and 85.5% ($p = .919$), and in scenario 3 they were 88.0% ($p = .520$) and 93.1% ($p = .529$). Meanwhile, there were more FP cases in the retrospective studies than in the prospective studies, especially in the SM category. The specificity in scenario 1 was 98.4% ($p = .984$) in the prospective studies and 94.9% ($p = .297$) in the retrospective studies. In scenario 3, there was an even more significant difference in specificity between the prospective studies and the retrospective group (93.7%, $p = .431$ vs. 72.8%, $p = .075$). There was also a significant difference in identifying neoplasms, with a specificity of 84.1% ($p = .068$) in the prospective studies and 77.0% ($p = .257$) in the retrospective studies.

DISCUSSION

Since its publication in 2018,[1] the MSRSGC has been widely accepted and endorsed by professional societies.[4] Salivary gland FNA is a fast and cost-effective method for preoperative diagnostic workups. Six categories of the MSRSGC can be applied to various salivary gland lesions. Each category is accompanied by the ROM and management to facilitate communication with clinicians.

The present meta-analysis aimed to examine the effectiveness of the MSRSGS in everyday cytopathology practice worldwide. Surprisingly, there were only a few prospective studies on the MSRSGC. This is because it is a relatively new system, and therefore most of the published studies have been retrospective and based on the expert reclassification of previously diagnosed specimens. We analyzed 1587 cases in seven prospective studies and compared

TABLE 4 Summary of the diagnostic accuracy of the MSRSGC and in Jalaly et al.,[5] Wang et al.,[6] and Gubbiotti et al.[7] studies.

Type of analysis	Type of study	Sensitivity, % (95% CI)	Specificity, % (95% CI)	DOR (95% CI)	NLR (95% CI)	PLR (95% CI)
Detecting malignancy, scenario 1	Prospective	83.1 (71.1–90.8)	98.4 (96.6–99.3)	310.7 (121.2–796.6)	0.15 (0.07–0.32)	50.7 (23.5–109.4)
	Retrospective	89.1 (83.6–92.9)	94.9 (91.9–96.9)	218.8 (107.3–438.1)	0.09 (0.06–0.13)	17.1 (11.0–26.5)
	Jalaly[5]	87.6 (85.1–89.6)	98.5 (97.7–99.0)	464.9 (303.5–712.2)	0.12 (0.11–0.15)	58.9 (38.7–89.6)
	Wang[6]	88.0 (84.9–90.9)	98.5 (97.9–99.0)	520.3 (294.8–902.6)	0.12 (0.09–0.16)	61.5 (40.3–97.9)
	Gubbiotti[7]	86.7 (85.2–88.1)	98.3 (97.9–98.6)	ND	0.14 (0.12–0.15)	50.4 (41.7–60.9)
Detecting malignancy, scenario 2	Prospective	78.2 (65.5–87.1)	98.9 (97.3–99.6)	343.4 (117.9–1000.1)	0.18 (0.07–0.46)	69.1 (27.4–174.6)
	Retrospective	85.5 (78.7–90.4)	98.0 (95.9–99.1)	382.5 (148.1–988.4)	0.13 (0.06–0.28)	43.3 (20.7–90.4)
	Jalaly[5]	84.7 (81.5–87.3)	99.5 (99.1–99.7)	1106.0 (596.5–2050.7)	0.15 (0.13–0.19)	170.5 (91.2–318.8)
Detecting malignancy, scenario 3	Prospective	88.0 (80.2–93.0)	93.7 (86.8–97.1)	123.0 (42.8–353.6)	0.11 (0.06–0.23)	10.1 (4.3–23.8)
	Retrospective	93.1 (87.3–96.3)	72.8 (55.5–85.1)	51.1 (17.4–150.3)	0.07 (0.03–0.15)	3.0 (1.9–4.9)
	Gubbiotti[7]	94.0 (92.8–95.0)	90.0 (88.0–91.7)	ND	0.07 (0.06–0.08)	9.4 (7.8–11.3)
Detecting neoplasm	Prospective	95.3 (87.8–98.3)	84.1 (58.8–95.2)	111.0 (17.0–725.0)	0.07 (0.02–0.18)	4.4 (1.8–10.8)
	Retrospective	96.1 (93.9–97.6)	77.0 (61.4–87.5)	148.2 (62.5–351.4)	0.03 (0.02–0.05)	3.7 (2.2–6.2)
	Jalaly[5]	97.2 (95.9–98.0)	88.2 (81.6–92.6)	255.4 (149.8–435.5)	0.03 (0.02–0.05)	8.2 (5.2–13.0)
	Wang[6]	97.5 (96.4–98.4)	91.6 (86.9–95.6)	472.0 (234.0–934.2)	0.03 (0.02–0.04)	12.6 (7.0–23.7)

Note: Detecting malignancy, scenario 1: nonneoplastic and benign neoplasm as negative index and suspicious for malignancy and malignant as positive index for malignancy. Detecting malignancy, scenario 2: nonneoplastic and benign neoplasm as negative index and malignant as positive index for malignancy. Detecting malignancy, scenario 3: nonneoplastic as negative index and suspicious for malignancy and malignant as positive index for malignancy. Detecting neoplasm: nonneoplastic as negative index and benign neoplasm and salivary gland neoplasm of uncertain malignant potential, suspicious for malignancy, and malignant as positive index for neoplasm.

Abbreviations: CI, confidence interval; DOR, diagnostic odds ratio; ND, not determined; NLR, negative likelihood ratio; PLR, positive likelihood ratio.

them to 1764 cases in seven geographically matched retrospective studies.

Two meta-analyses of the MSRSGC have been previously published: Jalaly et al.,[5] who analyzed 37 studies, and Wang et al.,[6] who analyzed 35 studies. However, most of the studies included in these two meta-analyses were retrospective.

MSRSGC categorization

Most of the cases were placed in the BN category in both the prospective and retrospective studies (38.0% and 37.1%, respectively). There was a higher number of cases in both new undetermined categories (AUS and SUMP) in the prospective studies (9.2% and 11.0%) compared to only 3.2% and 4.1% in the retrospective studies. AUS cases were placed in the NN category in the retrospective studies and SUMP cases were placed in the SM category, which were more common in the retrospective studies. Stricter criteria for the NN category could also lead to its higher use in prospective studies.

MSRSGC sensitivity and specificity

The MSRSGC is an effective tool in daily cytopathology practice. Its sensitivity for detecting malignancies was slightly lower in the prospective studies than in the meta-analyses and in the retrospective group in this study. The analyses can be performed in three different scenarios (see the Materials and Methods section). In scenario 1, Jalaly et al.,[5] Wang et al.,[6] and Gubbiotti et al.[7] reported comparable sensitivities of 87.6%, 88.0%, and 86.7%. In the retrospective group in this study, the sensitivity was 89.1%, whereas it was only 83.1% in the prospective study group. However, the specificity was similar to those reported in previously published meta-analyses (98.4% in the present prospective cohort vs. 98.5%,[5] 98.5%,[6] and 98.3%[7]). In

scenario 2, which was only applied by Jalaly et al., the meta-analysis sensitivities were slightly lower than in scenario 1.[5] Jalaly et al.[5] reported a sensitivity of 84.7%, whereas the sensitivity was 78.2% in the prospective group and 85.5% in the retrospective group in the present meta-analysis. Specificities were slightly improved and a little higher than in scenario 1 because only the malignant group was considered positive for malignancy. The specificity in Jalaly et al.'s study[5] was 99.5%, and the specificity for the prospective group in scenario 2 was 98.9%. There were issues in the retrospective group identifying malignancy in the SM category in scenario 1, and thus the specificity was higher in scenario 2 (94.9% vs. 98.0%). The third scenario was only applied by Gubbiotti et al.[7] In that scenario, the sensitivities were the highest. In this meta-analysis, the sensitivities in the prospective studies were lower than the retrospective group values and those reported by Gubbiotti et al. (88.0% vs. 93.1% and 94.0%[7]). In contrast, the prospective studies had a higher specificity than Gubbiotti et al. (93.7% vs. 90.0%[7]). The retrospective group had a significantly lower specificity (72.8%) because of the low number of histologically verified cases in the NN category and problems identifying malignancies in the SM category. In summary, scenarios 1 and 2 exhibited better specificity (up to 98.9% in the prospective studies analyzed in the present meta-analysis) but scenario 3 had superior sensitivity (up to 94.0%) (Table 4).

According to the current meta-analysis, the MSRSGC has good ability to identify neoplasms. The sensitivities in the prospective group (95.3%) and the retrospective group (96.1%) were almost identical to those reported in two previously published meta-analyses (97.2%[5] and 97.5%[6]). However, the specificities in the present meta-analysis were lower (84.1% in the prospective group and 77.0% in the retrospective group) than in two previous meta-analyses (88.2%[5] and 91.6%[6]).

MSRSGC ROM

The ROM was calculated for each MSRSGC diagnostic category. It is an important measure for guiding clinical management. A summary of the ROMs in the present meta-analysis and in previously published analyses is presented in Table 5. In the ND category, the ROMs were 21.0% in the prospective studies and 26.6% in the retrospective studies. Two previous meta-analyses and Gubbiotti et al. reported lower ROMs (17.3%,[5] 11.4%,[6] and 17.1%[7]) for ND. The higher prevalence of insufficiently sampled malignant cases in the ND category may explain the higher ROM values.

In the NN category, there were no significant differences between the present meta-analysis and previous ones (8.1% and 9.4% vs. 10.7%[5] and 10.9%[6]). However, Gubbiotti et al.[7] reported a higher ROM (14.3%) in the NN category.

In the AUS category, the present meta-analysis showed slightly higher ROMs than two other meta-analyses. Jalaly et al.,[5] Wang et al.,[6] and Gubbiotti et al.[7] reported ROMs of 32.2%, 30.5%, and 32.7% ROM for AUS, respectively. In the present meta-analysis, the ROM in the prospective studies was 34.9% and even higher (39.6%) in the retrospective studies, even though fewer cases were categorized as AUS. Lymphomas are often categorized as AUS.

In the BN category, the ROMs were almost the same in all groups (2.4% and 2.1% vs. 3.0%,[5] 2.8%,[6] and 2.2%[7]). The prospective studies had similar ROMs for SUMP to the two previous meta-analyses (36.6% vs. 37.2%[5] and 37.7%[6]). On the contrary, the ROM for SUMP was lower (31.2%) in the retrospective studies. Gubbiotti et al.[7] reported a notably higher ROM for the SUMP category (46.4%).

Retrospective studies probably tend to upgrade lesions to the SM category, which leads to lower ROMs in retrospective studies than in prospective studies, as in Jalaly et al., Wang et al., and Gubbiotti et al. (66.0% vs. 86.0%, 84.0%,[5] 83.8%,[6] and 80.0%,[7] respectively).

ROMs were highest in the MN category, and all groups reported similar values (97.0% and 96.7% vs. 98.0%,[5] 96.7%,[6] and 97.1%[7]).

The ROM values for each MSRSGC diagnostic category in the prospective studies were consistent with the reference values from the second edition of the MSRSGC.[23] The most significant difference was found in the ND category. In the second edition of the MSRSGC, the ROM for ND is estimated to be 15%, whereas it was estimated to be 25% in the first edition of the MSRSGC. The ROM for ND was found to be 21.0% in the present meta-analysis of prospective studies. This may be associated with the higher occurrence of insufficiently sampled malignancies. The ROM value for ND in the

TABLE 5 Risk of malignancy in the present study and in studies by Jalaly et al.,[5] Wang et al.,[6] and Gubbiotti et al.[7]

	Risk of malignancy according to the MSRSGC categories						
Study	**ND, %**	**NN, %**	**AUS, %**	**BN, %**	**SUMP, %**	**SM, %**	**MN, %**
Prospective studies	21.0	9.4	34.9	2.4	36.6	86.0	97.0
Retrospective studies	26.6	8.1	39.6	2.1	31.2	66.0	96.7
Jalaly[5]	17.3	10.7	32.2	3.0	37.2	84.0	98.0
Wang[6]	11.4	10.9	30.5	2.8	37.7	83.8	96.7
Gubbiotti[7]	17.1	14.3	32.7	2.2	46.4	80.0	97.1

Abbreviations: AUS, atypia of undetermined significance; BN, benign neoplasm; MN, malignant neoplasm; MSRSGC, Milan System for Reporting Salivary Gland Cytopathology; ND, nondiagnostic; NN, nonneoplastic; SM, suspicious for malignancy; SUMP, salivary gland neoplasm of uncertain malignant potential.

first version of the MSRSGC is more in concordance with the results of the present meta-analysis, whereas the meta-analyses of Jalaly et al.[5] and Wang et al.[6] are more in agreement with the second edition of the MSRSGC.

The most significant differences in the proposed ROMs between the first and second editions of the MSRSGC were in the AUS (20% and 30%), SM (60% and 83%), and MN (90% and 98%) categories. The present meta-analysis and those of Jalaly et al.[5] and Wang et al.[6] support these modifications.

AUS and SUMP are considered new entities and diagnostically challenging categories. In the SUMP category, the proposed ROM (35.0%) did not change from the first MSRSGC edition to the second edition, although its range has been reported as 0–100%.[23] The ROM in the present meta-analysis (36.6%) was consistent with the reference value as well as the findings of Jalaly et al. and Wang et al. (37.2%[5] and 37.7%[6]). Conversely, Gubbiotti et al.[7] reported a notably higher ROM (46.4%). The ROM values ranged from 25.0%[9] to 50.0% in the prospective studies.[11,12] In the AUS category, the ROM values ranged from 0% to 100% in the literature.[23] The pooled prevalence of AUS ranged from 11.1%[9] to 75.0%[11] in the present prospective study group. In addition, it has been recommended that the overall percentage of salivary gland FNA cases categorized as AUS should not exceed 10%.[23] In the present study, the percentage of AUS cases (9.2%) in the prospective group was consistent with the MSRSGC recommendation.

MSRSGC RON

In addition to the ROM, the RON was also calculated in the present meta-analysis, even though it is not a part of the MSRSGC. Benign tumors such as pleomorphic adenoma and Warthin tumor nevertheless constitute the largest proportion of salivary gland tumors. In the ND category, the RON values were 58.0% in the prospective studies and 71.7% in the retrospective studies. Wang et al. reported a RON of 65.6% for the ND category.[6] In the prospective studies, the RON was significantly higher in the NN category than in the retrospective studies and in Wang et al. (42.6% vs. 26.8% and 24.8%[6]). In the AUS category, the RON was higher in Wang et al.'s meta-analysis than in both the prospective and retrospective groups in the present meta-analysis (85.1%[6] vs. 71.3% vs. 67.4%). In the BN category, the majority of cases were neoplastic, and the RONs were 98.9% in the prospective studies, 99.1% in the retrospective studies, and 99.4% in Wang et al.'s meta-analysis.[6] In the SUMP category, the present retrospective study group had a lower RON than those in the prospective studies and Wang et al.'s meta-analysis (87.1% vs. 97.8% and 96.8%[6]). In the SM category, the RONs were 89.3% in the prospective studies, 92.8% in the retrospective studies, and 96.8% in Wang et al.'s meta-analysis.[6] In the MN category, the RONs were almost identical (97.3%, 97.8%, and 99.9%[6]).

In conclusion, the present meta-analysis showed that the MSRSGC is an effective tool for everyday cytopathology practice, although the number of prospective studies on the MSRSGC was surprisingly low. It exhibited improved diagnostic accuracy in all diagnostic categories. The ROMs of prospective studies were in concordance with the MSRSGC reference values.

AUTHOR CONTRIBUTIONS

Henri Lagerstam: Conceptualization, methodology, validation, formal analysis, investigation, data curation, visualization, and writing–original draft. **David Kalfert**: Methodology, validation, formal analysis, investigation, data curation, writing–review and editing, visualization, and funding acquisition. **Zahra Maleki**: Writing–review and editing. **Ivana Kholová**: Conceptualization, investigation, data curation, writing–original draft, writing–review and editing, supervision, project administration, funding acquisition, and resources.

ACKNOWLEDGMENTS

This study was funded by a Government Research Funding (Valtion tutkimusrahoitus (VTR)) grant from the Pirkanmaa Hospital District (to Ivana Kholová) and by the Charles University Cooperatio Program, research area "Surgical Disciplines" (to David Kalfert). Neither grant had an impact on the study design and analysis.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data are included in the article and supplementary material.

ORCID

Zahra Maleki https://orcid.org/0000-0003-3273-4802
Ivana Kholová https://orcid.org/0000-0003-4089-0450

REFERENCES

1. Faquin WC, Rossi ED, eds. *The Milan System for Reporting Salivary Gland Cytopathology*. Springer; 2018.
2. Rossi ED, Faquin WC. The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC): an international effort toward improved patient care—when the roots might be inspired by Leonardo da Vinci. *Cancer Cytopathol*. 2018;126(9):756-766. doi:10.1002/cncy.22040
3. Rossi ED, Baloch Z, Pusztaszeri M, Faquin WC. The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC): an ASC-IAC–sponsored system for reporting salivary gland fine-needle aspiration. *J Am Soc Cytopathol*. 2018;7(3):111-118. doi:10.1016/j.jasc.2018.02.002
4. Barbarite E, Puram SV, Derakhshan A, Rossi ED, Faquin WC, Varvares MA. A call for universal acceptance of the Milan System for Reporting Salivary Gland Cytopathology. *Laryngoscope*. 2020;130(1):80-85. doi:10.1002/lary.27905
5. Jalaly JB, Farahani SJ, Baloch ZW. The Milan System for Reporting Salivary Gland Cytopathology: a comprehensive review of the literature. *Diagn Cytopathol*. 2020;48(10):880-889. doi:10.1002/dc.24536
6. 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):849-859. doi:10.1002/cncy.22604
7. Gubbiotti MA, Jalaly J, Baloch Z. Making a case for the success of Milan System for Reporting Salivary Gland Cytopathology. *Diagn Cytopathol*. 2022;50(9):451-455. doi:10.1002/dc.25010

8. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med*. 2021;18(3):e1003583. doi:10.1371/journal.pmed.1003583
9. Tommola E, Tommola S, Porre S, Kholová 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*. 2019;11(10):1589. doi:10.3390/cancers11101589
10. Tommola E, Kalfert D, Hakso-Mäkinen H, Kholová I. The contributory role of cell blocks in salivary gland neoplasms fine needle aspirations classified by the Milan System for Reporting Salivary Gland Cytology. *Diagnostics*. 2021;11(10):1778. doi:10.3390/diagnostics11101778
11. 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. doi:10.1007/s12105-020-01278-1
12. 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. doi:10.4103/ijpm.ijpm_662_17
13. Karuna V, Vivek V, Singh R, Gupta P, Verma N. MSRSGC: a prospective study of heterogenous group atypia of undetermined significance. *Indian J Pathol Microbiol*. 2022;65(3):630-636.
14. 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. doi:10.1002/dc.24748
15. Manucha V, Gonzalez MF, Akhtar I. 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. doi:10.1016/j.jasc.2020.08.005
16. Sadullahoğlu C, Yıldırım HT, 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. doi:10.1002/dc.24214
17. Dubucs C, Basset C, D'Aure D, Courtade-Saïdi 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*. 2019;11(12):1912. doi:10.3390/cancers11121912
18. 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. doi:10.1111/cyt.12811
19. 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. doi:10.4103/joc.joc_165_18
20. 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*. 2019;47(5):382-388. doi:10.1002/dc.24109
21. 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. doi:10.1002/dc.24478
22. Hollyfield JM, O'Connor SM, Maygarden SJ, et al. Northern Italy in the American South: assessing interobserver reliability within the Milan System for Reporting Salivary Gland Cytopathology. *Cancer Cytopathol*. 2018;126(6):390-396. doi:10.1002/cncy.21989
23. Faquin WC, Rossi ED, eds. *The Milan System for Reporting Salivary Gland Cytopathology*. 2nd ed. Springer; 2023.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lagerstam H, Kalfert D, Maleki Z, Kholová I. How the Milan System for Reporting Salivary Gland Cytopathology works in cytopathology practice: meta-analysis of prospective studies and comparison with retrospective studies. *Cancer Cytopathol*. 2024;1-11. doi:10.1002/cncy.22815