

Application of Milan System for Reporting Salivary Gland Cytopathology: A 7 Year Study

Dr. Shikha Chopra^{*1}, Dr. Richa Jindal² and Dr. Molly Joseph³

¹MBBS, DNB Pathology Resident at St. Stephen's Hospital, New Delhi, India

²MBBS, MD, Pathology Senior Consultant, St. Stephen's Hospital, New Delhi, India

³MBBS, MD, Pathology Head of Department, St. Stephen's Hospital, New Delhi, India

Abstract

Introduction: Salivary gland shows various pathological conditions ranging from cystic, inflammatory, tumor like and neoplastic lesions. Fine needle aspiration cytology (FNAC) plays an important role in evaluating salivary gland (SG) tumors. Salivary gland tumors are one of the most heterogeneous groups of neoplasms with cytopathological features overlapping among the entities and making it difficult to assign to specific category. Due to these facts, salivary gland cytopathology is one of the most challenging areas of cytology. The lack of a uniform reported guidelines in salivary gland cytopathology leads to inter-observer variability and disagreements. The present study was undertaken to assess the degree of inter-observer reproducibility for diagnostic categorization of salivary gland lesions utilizing MSRSGC among pathologists with varying experience along with its role in providing a framework for reporting salivary gland lesions.

Materials and Methods: In this cross sectional study, total of 44 cases of salivary gland lesions subjected to FNAC over a period of 7 year were studied. The cases were critically reviewed by 2 pathologists and a pathology resident with variable experience in cytopathology using MSRSGC in our institution. Inter-observer variability was assessed by comparing the agreement between two cytopathologists and pathology resident by using Cohen's kappa statistics (κ score) and interpretation of the results was done using scale of Landis and Koch.

Results: All the salivary gland aspirates were categorized according to MSRSGC. Out of 44 cases, maximum cases 22 (50%) were classified under IVA (BN) followed by 27.27% to 29.5% cases classified under II (NN), 2.27-4.55% of cases under Category IVB (SUMP), 4.55% under category V (SM) and 6.82% cases under category VI (M). Inter-observer variability (IOV) was calculated for individual category in Milan system using Cohens kappa test, which was found to be in the almost perfect agreement range as per Landis and Koch, for categories II, IVA, V, VI (κ score 0.89- 1). Kappa score ranged from 0.645 - 1 for category I (ND), which showed substantial to an almost perfect agreement. Whereas, category IVB (SUMP) showed variable results, with substantial agreement (κ score 0.656) to no agreement (κ score 0) between different observers. The overall IOV showed an almost perfect agreement with a kappa score of 0.861 (obs1 vs 2), 0.896 (obs1 vs R), 0.965 (obs 2 vs R). The data was found to be statistically significant ($p < .0001$).

Conclusion: MSRSGC is a very efficient system and has the potential to standardize salivary gland FNA diagnoses, providing clear prognostic and management information to clinicians and surgeons. This system can be used with good reproducibility between observers with variable cytopathology experience. Placing lesions in categories using MSRSGC can result in optimal management of discordant cases without using a specific diagnosis. Application of MSRSGC has immense value for standardization of reporting of salivary gland FNAC. Hence we recommend the use of Milan system for reporting salivary gland cytopathology.

Abbreviations: MSRSGC: Milan system for reporting salivary gland cytopathology • IOV: Inter-observer variability • SG: Salivary Gland • FNAC: Fine needle aspiration cytology • κ score: kappa score

Aim and Objectives

Primary objective: To study and categorize the fine needle aspiration cytology of salivary gland lesion using the Milan system (MSRSGC) in our institution.

Secondary objective: To assess inter-observer variability to determine applicability of the system.

Introduction

Salivary gland (SG) shows various pathological conditions ranging from cystic, inflammatory, tumor like and neoplastic lesions. Salivary gland neoplasm are relatively uncommon and constitute only about 0.5% of all cancers. The global annual incidence of overall salivary gland tumors (SGT) varies approximately 0.4 to 13.5 cases per 100,000 people and 0.4-2.6 per

100,000 for malignant tumor. [1] Benign tumor accounts for 54-79% of all SGT, among them the parotid gland is the most common site. Malignant tumor accounts for 21-46% and more frequently seen in minor salivary glands and sublingual glands [2].

Fine needle aspiration (FNA) is minimally invasive, cost effective test. It can differentiate not only between neoplastic and non-neoplastic lesions, but also between low grade and high grade salivary gland lesions. [3] The accuracy of FNA to differentiate benign lesions from malignant lesions has been observed to be as high as 81% to 100%. However, the accuracy of FNA to provide a specific diagnosis has a wider range of 48% to 94%. [3-7]

Salivary gland tumors are one of the most heterogeneous groups of neoplasms with cytopathological features overlapping among the entities and making it difficult to assign to specific category. Due to these facts, salivary gland cytopathology is one of the most challenging areas of cytology. [8] These challenges are further complicated by the lack of a standardized, tiered diagnostic framework and resulted in inconsistent terminology among institutions, individual cytopathologists and lead to confusion in communication among cytopathologists. To address this need, collaborative efforts made by The American Society of Cytopathology (ASC) and International Academy of Cytology (IAC) have resulted in a evidence based tiered classification system to report salivary gland FNA. They proposed the Milan System for reporting salivary gland cytopathology (MSRSGC) which includes the six major diagnostic categories as well as a brief description of cytomorphologic features that can be applied in daily practice to accurately diagnose salivary gland lesions. [9,10]

For a diagnostic categorization system to be clinically useful, it must show good inter-observer reproducibility. The present study was undertaken to assess the degree of inter-observer reproducibility for diagnostic categorization of

***Address for Correspondence:** Dr. Shikha Chopra, Department of Pathology, St Stephen's Hospital, Tis hazari, New Delhi -110054, India; Tel: 091- 9653373325; E-mail: drshikhachopra@gmail.com

Copyright: © 2021 Chopra S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received 29 May 2021; **Accepted** 16 June 2021; **Published** 18 June 2021

salivary gland lesions utilizing MSRSGC among pathologists with varying experience in reporting salivary gland cytopathology.

Materials and Methodology

In this cross sectional study, total of 44 cases of salivary gland lesions subjected to FNAC over a period of 7 year (Jan 2013 to Dec 2019) were studied at Department of Pathology, St. Stephen's Hospital, Delhi.

Sample Size & Sample Technique:

The study of SA Thirayai et al. [11] observed that according to Milan system, 10.5% were categorized into Malignant. Taking this value as reference, the minimum required sample size with 10% margin of error and 5% level of significance is 37 patients. To reduce margin of error, total sample size taken is 44.

Formula used is: $N \geq (p(1-p))/(ME/Z\alpha)^2$

Where $Z\alpha$ is value of Z at two sided alpha error of 5%, ME is margin of error and p is proportion of patients categorized as Malignant.

Calculations: $n \geq ((.105*(1-.105))/(.1/1.96))^2 = 36.10 = 37$ (approx.)

The estimated sample size comes out to be 37 cases. However, to reduce margin of error, 44 cases were included in the study.

Methodology: Detailed clinical history regarding age, sex, site, increase or decrease in size of swelling, associated symptoms was taken from the patient. All the patients were clinically examined in detail as per the study preformed & a careful palpation of the salivary gland was done to locate the site of aspiration. After brief explanation about the procedure to the patient, consent was taken. The procedure was done with the patient in supine or sitting position. Using the first finger and thumb of the free hand, the swelling was held firmly in place while the other hand was used to smoothly and rapidly insert the needle into the swelling. Under aseptic conditions, aspiration was done using 23 G needle attached to a 10cc syringe, followed by multiple short, quick back and forth movements with the needle for approximately 5-10 seconds without significantly changing the direction of the needle. Material was collected in the needle, which was withdrawn and transferred to clean labelled glass slides and smears were prepared. Smears were air dried, fixed in methanol for Giemsa staining as per standard techniques.

For retrospective cases, the slides were retrieved from the archives and the cytological diagnoses were reviewed and retrospectively classified by two senior cytopathologists and pathology resident. For prospective cases, slides were studied under light microscopy. The slides were independently and blindly reviewed by two cytopathologist (observer 1 and observer 2) and a pathology resident. All the cases were categorized by using the MSRSGC.

The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC) consists of 6 broad categories.

1. Non Diagnostic (ND)
2. Non Neoplastic (NN)
3. Atypia of Undetermined Significance (AUS)
4. Neoplasm: (a) Benign (BN); (b) Salivary Gland Neoplasm of Uncertain Malignant Potential (SUMP)
5. Suspicious For Malignancy (SM)
6. Malignant (M)

At the end of review diagnostic categories assigned by the observers were tallied. Inter-observer variability was assessed by comparing the agreement between two cytopathologists and pathology resident by using Cohen's kappa statistics (κ score) and interpretation of the results was done using scale of Landis and Koch, it was interpreted as [12,13] (**Table 1**)

Statistical Analysis: Categorical variables were presented in number and percentage (%) and continuous variables were presented as mean \pm

Table 1:

Kappa statistic	Accuracy
<0	Less than chance agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Almost perfect agreement

SD and median. Normality of data was tested by Kolmogorov-Smirnov test. If the normality was rejected then non parametric test was used. Statistical tests were applied as follows:

1. Quantitative variables were compared using Kruskal Wallis Test (as the data sets were not normally distributed) between the Milan category.
2. Qualitative variables were associated using Fisher's exact test.
3. Inter-rater kappa agreement was used to assess the strength of agreement between observers.

A p value of <0.05 was considered statistically significant.

The data was entered in MS Excel spreadsheet and analysis was done using Statistical Package for Social Sciences (SPSS) version 21.0.

Observations and Results

This is a cross sectional study of application of Milan system for reporting salivary gland lesions, which was conducted at St. Stephen's hospital, Delhi. During January 2013 to Dec 2019, a total of 44 cases of salivary gland lesions were included in the study. Detailed cytological features were evaluated in each case. The cases were identified on the basis of clinical features i.e. complaints of swelling including size and site of swelling

The youngest patient in the study was 16 years old and the oldest was 75 years old. Maximum number of patients were in age group 51-60 years (22.73%) followed by more than 60 years (20.45%). Mean age of the cases at presentation was 45.93 years with standard deviation of ± 16 . (**Table 2 and Chart 1**).

In our study, Males dominated the study group i.e. 29 cases (65.91%), when compared to females i.e. 15 cases (34.09%). (Table 3 and Chart 2). The male to female ratio is 1.93:1. (**Table 3 and Chart 2**)

Out of the 44 patients, 22 patients presented with complaint of parotid swelling and 13 patients with complaint of submandibular swelling. So, parotid region (50 %) was the most common site of presentation followed by submandibular region (29.55%). (**Table 4 and Chart 3**)

In our study, the size of the lesion ranged from 1cm -10cm with a mean size of 3.24cm. (**Table 5 and Chart 4**)

Out of 44 lesions, 35 lesions (79.55%) were firm in consistency followed by 7 lesions (15.91%) with soft consistency and 2 (4.55%) with hard consistence. (**Table 6 and Chart 5**)

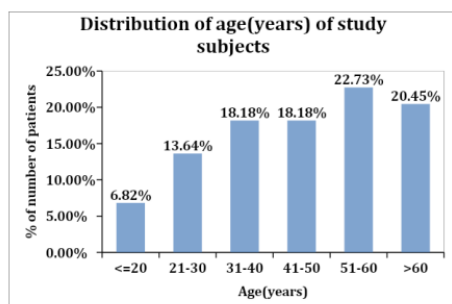
Out of 44 cases, most of the lesions (25/44) were mobile in nature with an incidence of 56.82% followed by immobile lesions (43.18%). (**Table 7 and Chart 6**)

Blood tinged aspirate was obtained in most of the cases (38/44) with an incidence of 86.36%. (**Table 8 and Chart 7**)

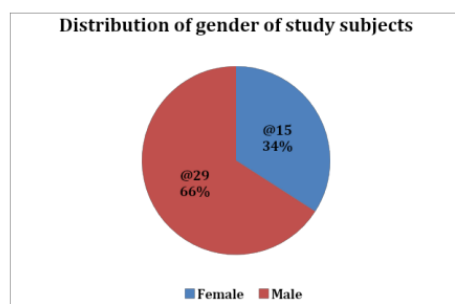
Out of 44 cases, maximum cases 22 (50%) were classified under IVA i.e. Benign Neoplasm followed by 27.27% to 29.5% cases classified under II i.e. Non-neoplastic. Category IVB i.e., SUMP seen in almost 2.27- 4.55% of the cases. (**Table 9 and Chart 8**)

Table 2. Distribution of age(years) of study subjects.

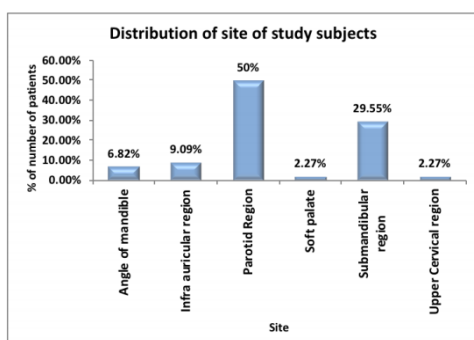
Age(years)	Frequency	Percentage
<=20	3	6.82%
21-30	6	13.64%
31-40	8	18.18%
41-50	8	18.18%
51-60	10	22.73%
>60	9	20.45%
Mean \pm Stdev	45.93 \pm 16	
Median(IQR)	47(32-57.25)	
Range	16-75	

**Chart 1:** Distribution of age(years) of study subjects.**Table 3.** Distribution of gender of study subjects.

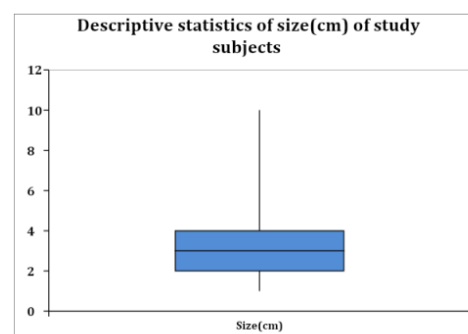
Gender	Frequency	Percentage
Female	15	34.09%
Male	29	65.91%
Total	44	100.00%

**Chart 2:** Distribution of gender of study subjects.**Table 4.** Distribution of site of study subjects.

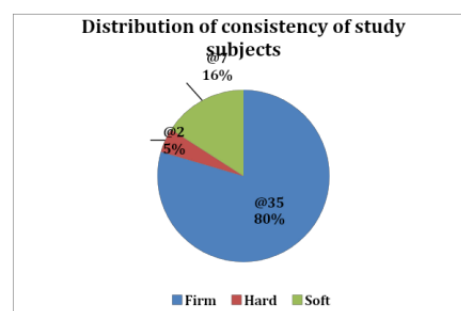
Site	Frequency	Percentage
Angle of mandible	3	6.82%
Infra auricular region	4	9.09%
Parotid Region	22	50.00%
Soft palate	1	2.27%
Submandibular region	13	29.55%
Upper Cervical region	1	2.27%
Total	44	100.00%

**Chart 3:** Distribution of site of study subjects.**Table 5.** Descriptive statistics of size(cm) of study subjects.

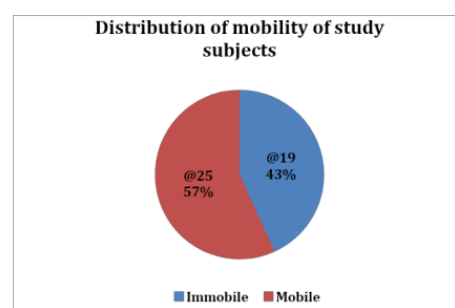
Variable	Mean \pm Stdev	Median(IQR)	Range
Size(cm)	3.24 \pm 1.76	3(2-4)	1-10

**Chart 4:** Descriptive statistics of size(cm) of study subjects.**Table 6.** Distribution of consistency of study subjects.

Consistency	Frequency	Percentage
Firm	35	79.55%
Hard	2	4.55%
Soft	7	15.91%
Total	44	100.00%

**Chart 5:** Distribution of consistency of study subjects.**Table 7.** Distribution of mobility of study subjects.

Mobility	Frequency	Percentage
Immobile	19	43.18%
Mobile	25	56.82%
Total	44	100.00%

**Chart 6:** Distribution of mobility of study subjects.

A total of 44 cases were categorized according to MSRSGC by two cytopathologists i.e. observer 1 and observer 2. At the end of the review diagnostic categories assigned by the observers were tallied and concordance level among them was calculated using Cohen's Kappa statistics (κ). The κ score was calculated as 0.861, that falls in the almost perfect agreement range. (Table 10). The findings were statistically significant ($p < .0001$). (Table 10)

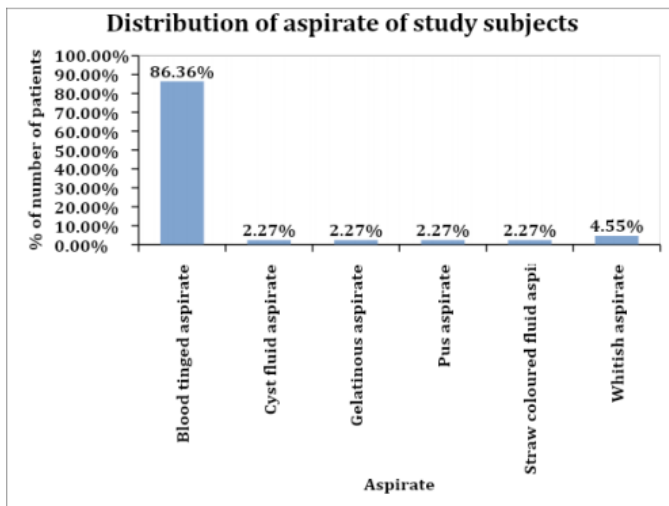


Chart 7: Distribution of aspirate of study subjects.

Table 8. Distribution of aspirate of study subjects.

Aspirate	Frequency	Percentage
Blood tinged aspirate	38	86.36%
Cyst fluid aspirate	1	2.27%
Gelatinous aspirate	1	2.27%
Pus aspirate	1	2.27%
Straw coloured fluid aspirate	1	2.27%
Whitish aspirate	2	4.55%
Total	44	100.00%

Table 9. Distribution of cases under Milan category by different observers.

Milan Category given	Observer 1	Observer 2	Resident
I (ND)	2 (4.55%)	4 (9.09%)	4 (9.09%)
II (NN)	13 (29.55%)	13 (29.55%)	12 (27.27%)
IVA (BN)	22 (50.00%)	22 (50.00%)	22 (50.00%)
IVB (SUMP)	2 (4.55%)	0 (0%)	1 (2.27%)
V (SM)	2 (4.55%)	2 (4.55%)	2 (4.55%)
VI (M)	3 (6.82%)	3 (6.82%)	3 (6.82%)
TOTAL	44 (100.00%)	44 (100.00%)	44 (100.00%)

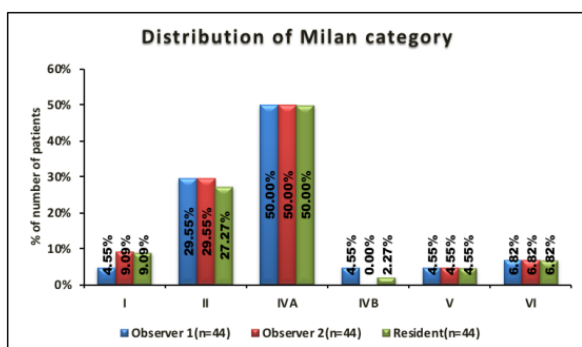


Chart 8: Distribution of cases under Milan Category by different observers.

The inter-observer variability was calculated between observer 1 and Pathology Resident. The κ score was calculated as 0.896, which represents an almost perfect agreement according to Cohen's kappa. (Table 11). The findings were statistically significant ($p < .0001$). (Table 11)

Table 10. Inter-rater kappa agreement of Milan Category given by observer 1 and Milan category given by observer 2.

Milan Category given by observer 1	Milan category given by observer 2					Total	P value	Kappa
	I(n=4)	II(n=13)	IVA(n=22)	V(n=2)	VI(n=3)			
I	2 (4.55%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (4.55%)	<.0001	0.861
II	1 (2.27%)	12 (27.27%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	13 (29.55%)		
IVA	1 (2.27%)	0 (0.00%)	21 (47.73%)	0 (0.00%)	0 (0.00%)	22 (50.00%)		
IVB	0 (0.00%)	1 (2.27%)	1 (2.27%)	0 (0.00%)	0 (0.00%)	2 (4.55%)		
V	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (4.55%)	0 (0.00%)	2 (4.55%)		
VI	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (6.82%)	3 (6.82%)		
Total	4 (9.09%)	13 (29.55%)	22 (50.00%)	2 (4.55%)	3 (6.82%)	44 (100.00%)		

Table 11. Inter-rater kappa agreement of Milan Category given by observer 1 and Milan category given by Resident Table 10.

Milan category given by observer 2	Milan category given by Resident						Total	P value	Kappa
	I(n=4)	II(n=12)	IVA(n=22)	IVB(n=1)	V(n=2)	VI(n=3)			
I	4 (9.09%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4 (9.09%)	<.0001	0.965
II	0 (0.00%)	12 (27.27%)	0 (0.00%)	1 (2.27%)	0 (0.00%)	0 (0.00%)	13 (29.55%)		
IVA	0 (0.00%)	0 (0.00%)	22 (50.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	22 (50.00%)		
V	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (4.55%)	0 (0.00%)	2 (4.55%)		
VI	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (6.82%)	3 (6.82%)		
Total	4 (9.09%)	12 (27.27%)	22 (50.00%)	1 (2.27%)	2 (4.55%)	3 (6.82%)	44 (100.00%)		

The inter-observer variability was calculated between observer 2 and Pathology Resident. The κ score was calculated as 0.965, which represents an almost perfect agreement. 3. The findings were statistically significant ($p < .0001$). (Table 12)

The table shows the chance correct agreement (Cohen's kappa statistics) for each category. Using the scale of Landis and Koch, an almost perfect agreement was seen between observers for categories II, IVA, V, VI. Agreement ranged from substantial to almost perfect for category I. Category IVB showed variable results, with substantial agreements between observer 1 vs. resident to no agreement between observer 1 vs 2 and observer 2 vs. resident. (Table 13)

ND- Non diagnostic, **NN-** Non neoplastic, **BN** - Benign Neoplasm, **SUMP** - Salivary gland neoplasm of uncertain malignant potential, **SM** - Suspicious of malignancy, **M** - Malignant (Figures 1-8).

Discussion

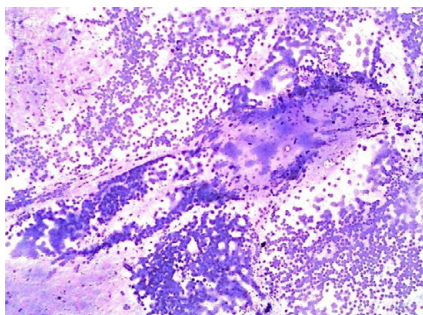
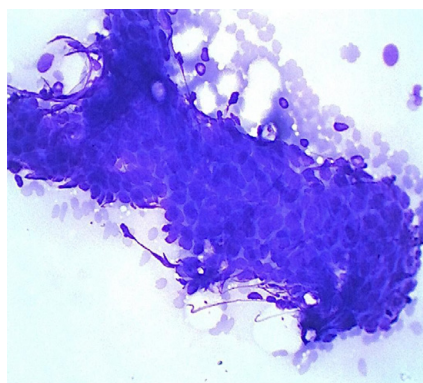
Salivary gland tumors represent about 6% of head and neck neoplasm and about 0.5% of all the malignancies in humans. [1] Recommendations and guidelines have been published for a number of body sites including the thyroid, [14] pancreaticobiliary system, [15] urinary tract [16] and respiratory system. [17] Recently a reporting system for the cytology specimens from the salivary glands has been published. [3] This system proposes categorical diagnoses analogous to those used in the Bethesda System for reporting thyroid gland cytopathology. This reporting system is intended to improve

Table 12. Inter-rater kappa agreement of Milan Category given by observer 2 and Milan category given by Resident.

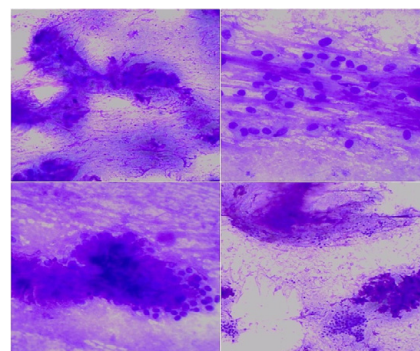
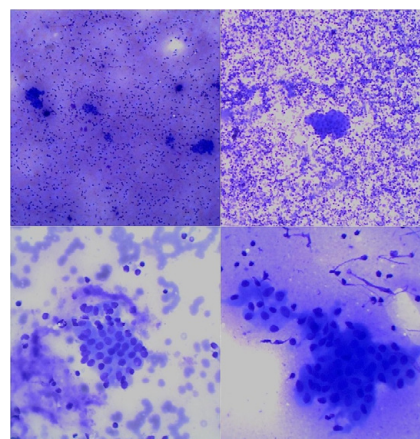
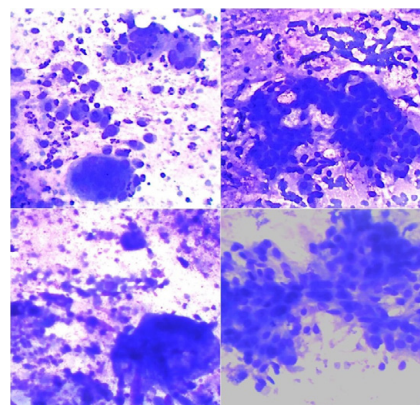
Milan category given by observer 2	Milan category given by Resident						Total	P value	Kappa
	I(n=4)	II(n=12)	IVA(n=22)	IVB(n=1)	V(n=2)	VI(n=3)			
I	4 (9.09%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4 (9.09%)	<.0001	0.965
II	0 (0.00%)	12 (27.27%)	0 (0.00%)	1 (2.27%)	0 (0.00%)	0 (0.00%)	13 (29.55%)		
IVA	0 (0.00%)	0 (0.00%)	22 (50.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	22 (50.00%)		
V	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (4.55%)	0 (0.00%)	2 (4.55%)		
VI	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	3 (6.82%)	3 (6.82%)		
Total	4 (9.09%)	12 (27.27%)	22 (50.00%)	1 (2.27%)	2 (4.55%)	3 (6.82%)	44 (100.00%)		

Table 13. Summary of agreement by category for Milan system.

Inter-rater kappa	Observer 1 vs 2	Observer 1 vs resident	Observer 2 vs resident
I (ND)	0.645	0.645	1
II (NN)	0.891	0.944	0.944
IVA (BN)	0.909	0.909	1
IVB (SUMP)	0	0.656	0
V (SM)	1	1	1
VI (M)	1	1	1

**Figure 1.** Non diagnostic (ND) - The aspirate contains blood and debris only. Insufficient for classification.**Figure 2.** Non neoplastic (NN) - This aspirate shows sheet of cytologically bland ductal epithelial cells.

communication between reporting cytopathologists and clinicians both within an institution and to external institutions. In our study, 44 cases of salivary gland lesion FNAC reported during January 2013 to Dec 2019 at St. Stephen's hospital, Delhi were reviewed and categorized according to Milan system for reporting salivary gland cytopathology. An attempt was made

**Figure 3.** Benign neoplasm (BN): a typical Pleomorphic Adenoma with plasmacytoid myoepithelial cells mixed with an intense metachromatic fibrillary matrix which stains bright magenta on MGG.**Figure 4.** Benign Neoplasm: FNA of Warthin tumor showing classic cytomorphic features consisting of background lymphocytes and group of oncocytic epithelial cells with abundant granular cytoplasm.**Figure 5.** Salivary gland of undetermined malignant potential (SUMP). This aspirate shows neoplastic cells with oncocytic cytoplasm. Few cells demonstrate eccentrically placed nuclei.

to assess the inter-observer variability to determine the applicability of this reporting system.

Demographic Data in our Study

The mean age of patients in our study was 45.9 years. This is in concurrence with the study of Rohilla et al. [18], where the mean age was 43.7 years. Similarly, Amita K et al. [19], Chen et al. [20] also observed mean age of their study sample as 48.3 years and 53 years respectively, which is in close concordance with our results.

In our study, the incidence of salivary gland lesions in male was (65.91%) and female (34.09%). This is in concurrence with the study done by Karuna

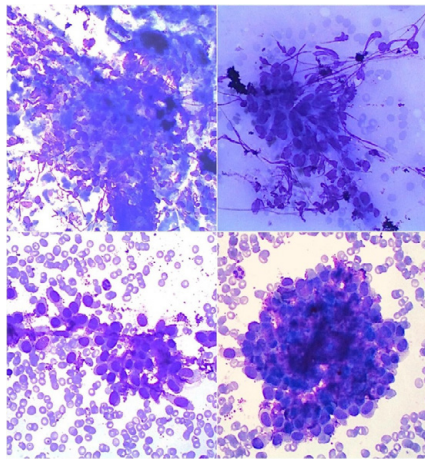


Figure 6. Suspicious for malignancy(SM):The aspirate contains bland epithelial cells with moderate amount of dense cytoplasm and well defined cell borders.

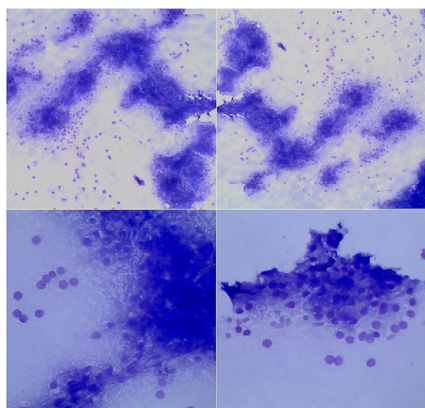


Figure 7. Malignant(M): FNA of Acinic cell carcinoma showing cellular smear with loosely cohesive groups of fragile acinar cells. Note the presence of stripped nuclei in the flocculent background and conspicuous absence of ductal cells.

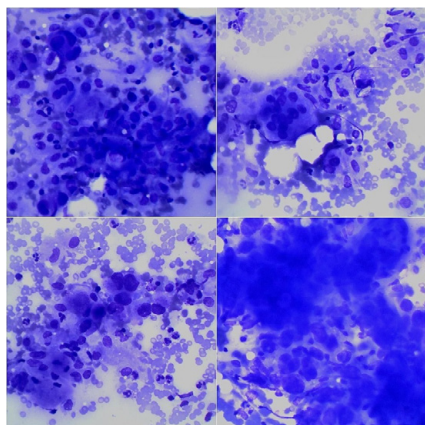


Figure 8. Malignant (M): This aspirate shows groups of atypical epithelial cells with hyperchromatic pleomorphic nuclei with abundant finely vacuolated to dense granular cytoplasm. Many bizarre nuclei and multinucleate forms can be seen.

et al. [21] who observed the incidence of 69.52% and 30.48% in males and females respectively.

Clinical Manifestations:

Parotid was the most common site to be involved (N=22; 50%) followed by Submandibular gland (N=13; 29.55%). Similarly, Kala C et al. [22] also observed an incidence of 57.7% and 27.3% in parotid and submandibular gland respectively, which is in close concordance with our study. Most of

the studies like Karuna et al. [21], Katta R et al. [23], Amita K et al. [19], Vishwanathan et al. [24] and Rohilla et al. [18] also observed parotid gland as the most common site of involvement in their studies.

Categorization and Inter-observer Variability (IOV)

Out of 44 cases, maximum cases 22 (50%) were classified under IVA i.e. Benign Neoplasm followed by 27.27% to 29.5% cases classified under II i.e. Non neoplastic. Category IVB i.e. SUMP seen in almost 2.27-4.55% of the cases. (Refer to Table 9 and Chart 8)

Discordant Cases Observer 1 vs Observer 2

IOV was seen in 4 / 44 cases, in which two cases were reported as SUMP by observer 1 and observer 2 categorized them as NN and BN respectively. Similarly two cases reported as ND by observer 2 has been categorized under NN and BN by observer 1.

Observer 1 vs Pathology Resident

IOV was observed in 3 / 44 cases. One case reported as SUMP by observer 1, the same has been reported as BN by resident. Similarly another two cases were reported as BN and NN by observer 1, that has been categorized under ND by resident.

Observer 2 vs Pathology Resident

IOV was seen in 1 / 44 cases. One case was categorized under NN by observer 2 and the same has been reported as SUMP by resident.

In our study, ND category represents 4.55-9.09% of cases. This is somewhat similar to incidence reported in other studies by Wei et al., [4] (2%), Karuna et al. [21] (2.08%), Rohilla et al. [18] (2.2%), Park et al. [25] (9.9%), Savant et al. [26] (9.2%), Viswanathan et al. [24] (12%) and Song et al. [27] (13.5%). We reported 29.55% cases under NN category, with most frequent diagnosis as sialadenosis followed by few cases of sialadenitis. This is in concurrence with study by Chen et al. [20], Song et al. [27], Wu et al. [28] and Karuna et al. [21] where they reported 23.6%, 21.5%, 16.1% and 16.20% cases respectively. The majority of our cases were classified as BN (50%). This is consistent with the frequency reported in the literature by various studies like Karuna et al. [21] reported 51.3%, Chen et al. [20] reported 45.6%, Wu et al. [28] reported 37.2%, Song et al., [27] reported 34.9%. Pleomorphic adenoma (34%) and Warthin's tumours (15%) were the most commonly diagnosed entities under BN category. SUMP represents almost 2.27- 4.55% of the cases in our study. This is in concurrence with the study of Karuna et al. [20], in which they reported 4.76% cases under SUMP category. Similar findings were recorded in study conducted by Chen et al. [20] (4.6%), Wu et al. [28] (5.9%) and Dubucs C et al. [29] (5%). We could not find any case to be categorized under category III i.e. Atypia of undetermined significance (AUS) in duration of our study period. But various studies like Karuna et al. [21], Chen et al. [20], Song et al. [27] and Wu et al. [28] reported cases under AUS category with a variable incidence of 2.85%, 3.2%, 10.8% and 37.2% respectively. In our study, SM was seldomly observed in 4.55% cases and Malignant (M) in 6.82% cases, which is consistent with the study done by Karuna et al. [21] where they reported 4.76% and 16.20% of cases under SM and M respectively. Since the release of MSRS GC in March 2018, multiple studies have been put forth in establishing the ROM (risk of malignancy) for each of the 6 categories, correlating classification with outcomes and indicating the utility of this system in day to day practice. [18,24,27] Only a few studies explored the inter-observer variability in MSRS GC. We calculated the inter-observer variability between all the three observers by using cohen's kappa statistics (κ score) & interpretation of the results was done using scale of Landis and Koch [12,13], which can range between -1 to +1. The overall kappa score between two cytopathologists (Observer 1 vs 2), Observer 1 vs resident and observer 2 vs resident was calculated as 0.861, 0.896 and 0.965 respectively, which showed an almost perfect agreement according to Landis and Koch scale. The data was statistically significant with a p value of < 0.0001. This is in concurrence with the study conducted by Garg et al. [32] where they noted an almost perfect agreement between consultant and JR, with a kappa score of 0.812. In a study conducted by

Vishwanathan et al. [33] inter-observer variability of indeterminate category in MSRSGC was calculated between cytopathologists & cytotechnologists. They calculated inter-observer agreement using a slightly modified version of previously published agreement scoring system described by Griffith et al. [34] into uniform, majority, divided, minimal or no agreement with an overall kappa score of 0.314 for the entire group, and 0.091-0.254 for cytotechnologists & 0.403-0.539 in case of cytopathologists. Furthermore, in our study kappa statistics for individual categories under Milan system were also calculated. There was considerable variability in agreement between categories. Some individual categories within the Milan system had excellent kappa scores within categories V (SM) and VI (M) having perfect agreement (kappa score = 1). SUMP category showed highly variable results, we found substantial agreement between observer 1 vs resident with a kappa score of 0.656. (Table 14)

Above mentioned studies had been done in various parts of the world, where they concluded the results with overall kappa score between the observers and few studies calculated kappa score only for indeterminate categories like AUS and SUMP. Whereas in our study, we have reported inter-observer variability in all the categories of Milan system and calculated three different Table 14. Comparison of different studies on inter-observer variability around the world and in India.

Reference study	Country	Year	Sample size	Number of observers	Kappascore
Hollyfield et al. [30]	North Carolina, USA	2018	134	6CP+1 R	0.217 (AUS) 0.024 (SUMP)
Layfield et al. [31]	Wisconsin, USA	2019	408	2CP +1R	0.42
Garg et al. [32]	Delhi, India	2019	150	1 CP+1R	0.812
Vishwanathan et al. [33]	New York, USA	2020	86	4 CPs+2 CTs	0.314
present study	Delhi, India	2020	44	2CP+1R	0.861 (CP1 vs 2) 0.896 (CP1 vs R) 0.965 (CP2 vs R)

values of kappa score with each value specific for variability between two observers.

It is to be noted that, before conducting the current study, the MSRSGC was not implemented at our institution; therefore, awareness among observers of this system was limited, thus a brief explanation of each category under Milan system was given before categorization.

Conclusion

MSRSGC has the potential to standardize salivary gland FNA diagnoses. This system can be used with good reproducibility between observers with variable cytopathology experience. Placing lesions in categories using MSRSGC can result in optimal management of discordant cases without using a specific diagnosis. Hence we recommend the use of Milan system for reporting salivary gland cytopathology.

Strengths of the Study

1. This is a cross sectional study.
2. Three different inter-observer variability agreements among three observers were calculated with three different kappa score values.

Limitations of the Study

1. Small number of sample size, due to study limited to one institution
2. Retrospective nature of the study.
3. No Histological correlation was done, so risk of malignancy was not calculated.

Recommendations:

1. The rate of inadequate samples can be reduced by the use of Diff - Quik staining, where the adequacy of the sample can be checked at the time of procedure.
2. Defining and implementing adequacy criteria for salivary gland FNA cases may help in standardizing pathologist interpretations of samples.
3. The ongoing discovery of genetic alterations and judicious use of assays such as immunohistochemistry and molecular testing will continue to increase the specificity of salivary gland FNA.
4. Studies of larger numbers of salivary gland fine needle aspirates will be necessary to fully establish the level of inter-observer agreement that can be expected while using MSRSGC.

References

1. Thiagarajan Shivakumar, Khuzema Fathehi, Deepa Nair and AnujaDeshmukh, et al. "Surgical morbidities and outcomes of major salivary gland neoplasms treated at a tertiary cancer center." *Indian J Cancer* 55 (2018):33.
2. Rousseau Audrey and Cécile Badoual. "Head and neck: salivary gland tumors: an overview." *Atlas Genet Cytogenet Oncol Haematol* 15(2011):533-41.
3. Faquin William C, Esther Diana Rossi, ZubairBaloach and Maria P. Foschini et al. "The Milan system for reporting salivary gland cytopathology."Switzerland: Springer International Publishing (2018):1-175.
4. Wei Shuanzeng, Lester J Layfield, Virginia ALiVolsi and Kathleen T Montone, et al. "Reporting of fine needle aspiration(FNA) specimens of salivary gland lesions: a comprehensive review." *Diag Cytopathol* 45(2017): 820-27.
5. Rammeh S, Romdhane E, Sassi A and Belhajkacem L, et al. "Accuracy of fine-needle aspiration cytology of head and neck masses." *Diagn Cytopathol* 5(2019):394-99.
6. Schmidt Robert L, Brian J Hall, Andrew R Wilson and Lester J Layfield. "A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions." *Am J Clin Pathol* 136(2011): 45-59.
7. GaikwadVaishali P, ChandaAnupriya and Leena P Naik. "Milan system for reporting salivary gland cytopathology—An experience from Western Indian Population." *J Cytol* 37(2020):93.
8. Pusztaszeri Marc, Esther D Rossi, Zubair W Baloach and William C Faquin. "Salivary gland fine needle aspiration and introduction of the Milan reporting system." *Adv Anat Pathol* 26(2019):84-92.
9. Rossi Esther Diana, Zubair W Baloach, Marc Pusztaszeri and William C. Faquin. "The Milan System for Reporting Salivary Gland Cytopathology (MSRSGC): an ASC-IAC-sponsored system for reporting salivary gland fine-needle aspiration." *Acta Cytologica* 62(2018):157-65.
10. Rossi Esther Diana, William C Faquin, Zubair Baloach, Güliz A Barkan and Maria Pia Foschini et al. "The Milan System for Reporting Salivary Gland Cytopathology: analysis and suggestions of initial survey." *Cancer Cytopathol* 125(2017):757-66.
11. Thirayai SA, YX Low, D Shelton, N Narine and D Slater, et al. "A retrospective 3-year study of salivary gland FNAC with categorization using the Milan reporting system." *Cytopathology* 29(2018):343-48.
12. McHugh, Mary L. "Inter rater reliability: the kappa statistic." *Bio chemica* 22(2012):276-82.

13. Landis J Richard and Gary G Koch. "The measurement of observer agreement for categorical data." *Biometrics* 33(1977):159-74.
14. Ali Syed Z and Edmund S Cibas. "The Bethesda system for reporting thyroid cytopathology." New York: Springer, 2018.
15. Pitman Martha Bishop and Lester James Layfield. "The Papanicolaou Society of Cytopathology system for reporting pancreaticobiliary cytology: definitions, criteria and explanatory notes." Springer, 2015.
16. Rosenthal Dorothy L, Eva M Wojcik and Daniel FI Kurtycz. "The Paris system for reporting urinary cytology." Cham, Switzerland: Springer International Publishing, 2016.
17. Layfield Lester J and Zubair Baloch. "The Papanicolaou Society of Cytopathology System for Reporting Respiratory Cytology." Springer Nature, 2019.
18. Rohilla M, Singh P, Rajwanshi A and Gupta N, 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* 125(2017):767-75.
19. Amita K, Rakshitha HB, Singh A, Shankar SV. "Evaluation of accuracy of milan system for reporting salivary gland cytology: Review of morphology and diagnostic challenges in each category." *J Cytol.* 37(2020):18-25.
20. Chen YA, Wu CY, Yang CS. "Application of the Milan system for reporting salivary gland cytopathology: A retrospective study in a tertiary institute." *Diagn Cytopathol* 47(2019):1160-67.
21. Karuna V, Gupta P, Rath M and Grover K, et al. "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* 62(2019):11-6.
22. Kala C, Kala S and Khan L. "Milan system for reporting salivary gland cytopathology: An experience with the implication for risk of malignancy." *J Cytol* 36(2019):160-4.
23. Katta R and Chaganti DP. "Application of the Milan system of reporting salivary cytopathology – A retrospective cytohistological correlation study." *J NTR Univ Health Sci* 8 (2019):11-7.
24. Viswanathan K, Sung S, Scognamiglio T and Yang GCH et al. "The role of the Milan system for reporting salivary gland cytopathology: A 5- year institutional experience." *Cancer Cytopathol* 126(2018):541-51.
25. Park W, Bae H, Park MH and Hwang NY, 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* 48(2019):222– 31.
26. Savant D, Jin C, Chau K and Hagan T et al. "Risk stratification of salivary gland cytology utilizing the Milan system of classification." *Diagn Cytopathol* 47(2019):172–80.
27. Song SJ, Shafique K, Wong LQ and LiVolsi VA et al. "The utility of the Milan system as a risk stratification tool for salivary gland fine needle aspiration cytology specimens." *Cytopathology* 30(2019):91-8.
28. Wu HH, Alruwaili F, Zeng BR and Cramer HM, et al. "Application of the Milan System for Reporting Salivary Gland Cytopathology: A Retrospective 12 Year Bi-institutional Study." *Am J Clin Pathol* 151(2019):613-21.
29. Dubucs C, Basset C, D'Aure D and Courtade-Saïdi M, et al. "A 4-Year Retrospective Analysis of Salivary Gland Cytopathology Using the Milan System for Reporting Salivary Gland Cytology and Ancillary Studies." *Cancers* 11(2019):1912.
30. Hollyfield JM, O'Connor SM, Maygarden SJ and Greene KG, et al. "Northern Italy in the American south: Assessing interobserver reliability within the Milan system for reporting salivary gland cytopathology." *Cancer Cytopathol* 126(2018):390-6.
31. Layfield LJ, Esebua M, Yang Z and Vatzmitsel M, et al. "The Milan System for Reporting Salivary Gland Cytopathology: a study of interobserver reproducibility." *Diagn Cytopathol* 47(2019):765-8.
32. Garg N, Diwaker P, Pathak P and Aggarwal D, et al. "Implementation of the Milan system for reporting salivary gland cytopathology: Interobserver concordance and cytohistological correlation of discordant cases." *Diagn Cytopathol* 2019;47(8):769-775.
33. Viswanathan K, Patel A, Abdelsayed M and Rosado L, et al. "Interobserver variability between cytopathologists and cytotechnologists upon application and characterization of the indeterminate category in the Milan system for reporting salivary gland cytopathology." *Cancer Cytopathol* 128(2020):828-39.
34. Griffith CC, Schmitt AC, Pantanowitz L and Monaco SE. "A pattern-based risk stratification scheme for salivary gland cytology: a multi-institutional, interobserver variability study to determine applicability." *Cancer Cytopathol* 125(2017):776-85.

How to cite this article: Shikha Chopra, Richa Jindal and Molly Joseph. "Application of Milan System for Reporting Salivary Gland Cytopathology: A 7 Year Study." *J Cytol Histol* 12 (2021): 576.