ORIGINAL ARTICLE

Effect of Absence Clinical History in Diagnostic Accuracy of Thyroid Fine Needle Aspiration Cytology

Najwa Nadeera Roslan[1], Mohd Nazri Abu[1], Nur Nadirah Abd Malek [1], Nur Amirah Roslan[1], Nur Adlina Alihad[1], Siti Norbaya Mohamad[1], Khairil Anuar Md. Isa[2]

[1] Centre of Medical Laboratory Technology, Faculty of Health Sciences, Universiti Teknologi MARA (UiTM) Selangor, Puncak Alam Campus, 42300 Selangor.
[2] Department of Basic Sciences, Faculty of Health Sciences, Universiti Teknologi MARA (UiTM) Selangor, Puncak Alam Campus, 42300 Selangor.

ABSTRACT

Introduction: Slides without clinical history can influence diagnostic interpretation accuracy, contributing to misdiagnosis and impair consistency. The goal is to evaluate the new screeners' performance through blinded cytology screening of thyroid FNAC cases by demonstrating reliability and diagnostic accuracy between inter-and intra-observers. This study conducted a correlational research design at Cytology Laboratory, Faculty of Health Sciences, Universiti Teknologi MARA (UiTM) Selangor, Malaysia. **Methods:** Five new screeners were chosen to blindly screen slides without unprovided clinical history to measure competency. Diagnoses were obtained using a light microscope from two screening sessions, with 40 cases, respectively. The inter-and intra-observer reliability testing was measured using the kappa value of Fleiss' and Cohen's kappa value, respectively, while the diagnostic accuracy without a clinical history was determined by the receiver operating characteristic (ROC) curve. **Results:** The inter-reliability kappa value (multiple new screeners) unveiled as 'Fair' agreement (κ) = 0.598, while intra-reliability (single new screener) showed 'Almost Perfect' agreement with (κ) = 0.856. The overall diagnostic accuracy was an average of 83.50 %, and the area under the curve (AUC) of ROC was 0.837 (83.7%). Operating parameters show an average value of 87.38% for sensitivity, 79.98% for specificity, 81.26% for positive predictive value (PPV), 86.62% for negative predictive value (NPV), and 23.73 for likelihood ratio (LR). **Conclusion:** Most new screeners have ample knowledge and skills in cytological screening. This blinded screening could be a practice to access a new screener's true ability in the future.

Key words : Clinical history, diagnostic accuracy, inter and intra-observer, new screener, reliability test, thyroid FNAC cases.

Corresponding Author:
Dr Mohd Nazri Abu PhD
Email: nazri669@uitm.edu.my
Tel:: +603-3258 4433

INTRODUCTION

Most of the teaching laboratory in universities, including the Centre of Medical Laboratory Studies, Faculty of Health Sciences, Universiti Teknologi MARA (UiTM), received donation slides from other institutions for teaching purposes without complete documentation, including demographic data and clinical history. These procedures may result in a diagnostic missed interpretation (1, 2). Evans et al., in 2011, discovered an increase of 17% to 30% in false-negative error during the screening session were due to the absence of clinical history (3). Therefore, the screener's ability to assess a cytomorphology-based diagnosis alone is critical to ensure the precise diagnosis to achieve a high level of accuracy during the interpretation of the result (1).

The absence of clinical history can influence diagnostic precision, contributing to misdiagnosis and impair consistency. Diagnostic accuracy and clinical history correlate with each other as the absence of clinical history lead to lower diagnostic accuracy (1). Slide screening based on morphological characteristics can only impact the diagnostic interpretation, leading to misdiagnosis (1). Raab et al., in 2000, carried out a study

which revealed that 89.2% of accuracy in diagnosis occurs when the slide observers are aware of clinical history while only about 74.0% of diagnostic accuracy when the slide observers are unaware of the patient's clinical history (1). The availability of these data will help the slide observers learn through cases before making a diagnosis.

Meanwhile, accuracy and reliability are checked between inter- and intraobserver to determine the efficacy of a diagnosis (4). The observation made by the same observer is referred to as intra-observer. Meanwhile, inter-observer is referred to the observation made by different observers. In this research, reliability among the new screener was evaluated to measure the ability of new screeners to make an accurate diagnosis through their expertise and skills. Unavoidable variations resulting from inter-and intra-observer variations may be considered an inherent part of the reporting system. Failure to accurately measure the variation observed in a study resulted in error interpretation and affected by the reporting laboratory's performance quality and patient management (5). Thus, it is important to assess the slide observer's reliability to make the same measurements under the same diagnosis consistently. Hence, rescreening is a tool to reduce variations and enhance screening quality. The aims are to evaluate a new screener's performance through blinded thyroid FNAC screening cases by demonstrating reliability and diagnostic accuracy between and intra-observers by depending on their knowledge and skills.

MATERIALS AND METHODS

The correlational research design was performed in this study as it tests the relationship between two or more variables using statistical analysis. The link to be tested was the association between the lack of clinical history in interpreting FNAC thyroid cases between new screeners by determining the reliability test's agreement and evaluating the overall diagnostic accuracy.

Forty FNAC thyroid cases were conveniently selected from all the total slides available in the cytology laboratory of Centre of Medical Laboratory Technology, Faculty of Health Sciences, Universiti Teknologi MARA, Malaysia, used as teaching slides during the practical session.

The sample size was determined using the Raosoft software. Each case consisted of a pair of Papanicolaou (PAP) and May Grunwald Giemsa (MGG) stained slides. All 40 cases were blindly screened without regard to clinical history by five new screeners among the students in this study. All new screeners were chosen based on these inclusion criteria. All participants could recognise the cell morphology of thyroid cells, had at least one year of experience in screening cytological slides, use light microscopes, and practice daily maintenance. Participants were primarily fourth-and fifth-year medical laboratory technology students who had all the inclusion requirements.

The screening session consisted of two parts, the first screening session (A) and the second screening session (B), where the same cases were observed on a different interval. The time gap between the first and second screening sessions was one month. The reason for conducting the first screening session was to obtain an agreement between new screeners for an inter-observer reliability test, while the second screening session was conducted to obtain the results from the same new screeners but at different times for an intra-observer reliability test. A sample of 40 FNAC thyroid cases was split into four sets, each of which had 10 cases at a time to be screened. Participants made a diagnosis for each case they had screened. The diagnostic response form was given, and participants were only allowed to select the final diagnosis after screening and classified based on the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) classification.

The inter-observer and intra-observer reliability agreement was calculated using the value of the kappa coefficient. The Kappa value will reflect an agreement between observers, either good or otherwise. For other parameters, the data were tabulated into the contingency table to measure the true positive (TP), the true negative (TN), the false positive (FP), and the false negative (FN) among the new screeners. Data were quantified using operating parameters; sensitivity (%), specificity (%), positive predictive value (PPV), negative predictive value (NPV), and likelihood ratio (LR). The receiver operating curve (ROC) was plotted using the sensitivity and specificity value. Thus, the diagnostic accuracy is determined by the area's value under the ROC curve. All the data collected were analysed using SPSS software (Statistical Package for Social Science for Windows version 25.0, IBM Corp, Armonk, New York, USA).

RESULT

Each new screener interpreted 40 FNAC thyroid cases according to TBSRTC classification. The distribution was divided into unsatisfactory (n=2), benign (n=19), atypia of undetermined significance or follicular lesion of undetermined significance (n=8), follicular neoplasm or suspicious for a follicular neoplasm (n=5), suspicious for malignancy (n=0), and malignant (n=6).

Fleiss' Kappa was performed to determine the level of agreement between the five new screeners in diagnosis 40 cases of FNAC thyroid with no clinical history in screening session A (Table I). Each new screener recognised and interpreted cases based on the classification of TBSRTC. Fleiss' Kappa's value was (κ) = 0.541; this shows a moderate agreement between the inter-observers. Besides, the individual kappa results

for each TBSRTC classification were shown in Table I for access to the degree of agreement between the new screener during screening session A. Most new screeners can differentiate unsatisfactory and malignant cases with 'Very Good' agreement with 0.904 and 0.863 kappa values, respectively. Next, all cases in the category Benign and Atypia of Undetermined Significance or Follicular Lesion of Undetermined Significance were reported with kappa values of 0.548 and 0.407, respectively. Besides, follicular neoplasm or suspicious for a follicular neoplasm with a 'fair' agreement of 0.360 kappa value. Meanwhile, for suspicious malignancy, there was no agreement with a negative kappa value of -0.036.

General Role

All five new screeners were reached in 14 out of 40 cases had no disagreement, which had been correctly classified into unsatisfactory (n=2), benign (n=6), Atypia of Undetermined Significance (n=2), and malignant (n=4). Besides, a one-category disagreement, which means that there were two diagnoses for one case, was noted with 21 out of 40 cases. The cases were benign vs follicular neoplasm (n=5); benign vs atypia of undetermined significance (n=8); benign vs suspicious malignancy (n=2); atypia of undetermined significance vs follicular neoplasm (n=2); atypia of undetermined significance vs suspicious malignancy (n=1); follicular neoplasm vs suspicious malignancy (n=1); suspicious malignancy vs malignant (n=2). Whereas in 5 cases, the disagreement overlapped more than one classification among five new screeners. It was unsatisfactory vs benign vs Atypia of Undetermined Significance (n=1); benign vs Atypia of Undetermined Significance vs follicular neoplasm (n=2); Atypia of Undetermined Significance vs follicular neoplasm vs suspicious malignancy (n=1); and Atypia of Undetermined Significance vs follicular neoplasm vs malignant (n=1).

Table I: Overview of results reliability testing

Inter-observer reliability (Fleiss' Kappa)			Intra-observer reliability (Cohen's Kappa)	
Overall kappa value	Individual categories (TBSRTC)	Kappa value	New screener (NS)	Kappa value
0.541	BC I	0.904	NS1	0.905
	BC II	0.548	NS2	0.852
	BC III	0.407	NS3	0.927
	BC IV	0.360	NS4	0.896
	BC V	-0.036	NS5	0.702
	BC VI	0.863		

Bethesda classification (BC); BC I: Unsatisfactory; BC II: Benign; BC III: Atypia of Undetermined Significance or Follicular Lesion of Undetermined Significance; BC IV: Follicular neoplasm or suspicious for follicular neoplasia; BC V: suspicious for malignancy; BC VI: malignant. *Kappa value is interpreted as follows: κ value ≤ 0: 'No agreement'; 0.01 – 0.20: 'None to slight agreement'; 0.21 – 0.40: 'Fair agreement'; 0.41 – 0.60: 'Moderate agreement; 0.61 – 0.80: 'Substantial agreement' and 0.81 – 1.00: 'Almost perfect agreement' .

For intra-observer reliability tests, the data collected during the first (A) and second (B) screening sessions were analysed and shown as Cohen's kappa value (Table I). The new screener diagnosed the same FNAC thyroid cases twice at different intervals within one month. All new screeners that were NS1, NS2, NS3, NS4 had an 'almost perfect' agreement after comparing the diagnosis between the two screening sessions as κ =0.905, 0.852, 0.927, and 0.896, respectively. Nonetheless, a substantial agreement was recorded for NS5 (κ = 0.702). The average value for each operating parameter was obtained by cumulatively sum up each measured value and calculating the average value among the five new screeners (Table II). It showed a significant value of sensitivity (87.38%), specificity (79.98%), PPV (81.26%), NPV (86.62%), LR (23.73) and diagnostic accuracy (83.50%). This result indicated a significant value in the diagnostic interpretation of thyroid FNAC cases without clinical history.

DISCUSSION

Clinical history is one of the critical data in which each cyto-screeners requires to produce precise results (3, 4). Indeed, examining unknown slides or cases is a valuable tool for evaluating and comparing screening and diagnostic performance, although it does not entirely mimic routine screening activity (5). A significant discrepancy is characterised as a difference in the evaluation series between participants. In 2000 previous research by Jones & Davey defined error as any abnormality and used a second material evaluation as a gold standard (5). Nonetheless, the monitoring of individual responses is encouraged as part of the competence assessment laboratories. Regarding the previous report, error rates are 1-2% higher for new participants, indicating their educational value (5).

Table II: Operating parameters to evaluate overall diagnostic accuracy for each new screener without providing clinical history

NS	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Likelihood ratio	Diagnostic accuracy (%)	Area under curve (AUC)
NS1	73.7	52.4	58.3	68.8	2.876	62.5	0.630
NS2	89.5	95.2	94.4	90.9	34.224	92.5	0.924
NS3	94.7	95.2	94.7	95.2	39.476	95.0	0.950
NS4	89.5	85.7	85.0	90.0	25.440	87.5	0.876
NS5	89.5	71.4	73.9	88.2	16.634	80.0	0.805
Average	**87.38**	**79.98**	**81.26**	**86.62**	**23.73**	**83.50**	**0.837**

NS: New screener; PPV: Positive Predictive Value; NPV: Negative Predictive Value; LR: Likelihood Ratio.

In this study, inter-observer reliability was performed

to reveal the magnitude of individual variability in diagnosis made between new screeners and their consistency with the actual diagnosis. According to the previous study, inter-observer interpretation variability would affect the final diagnosis. If reliability is low, accurate measurements will not be possible due to significant variations during measures (6). The study was done by Gupta, Komaromy-Hiller, Raab, & Nath in 2001 had proved that a high incidence of error in interpretation might occur when there is high inter-observer variability (7). Therefore, consistency between and within evaluators is essential to evaluate and implement results (8). As a tool to assess competency among cytotechnologists, Salmi, Toth, & Kong in 2015 had published a report on mock gynaecologic cytology proficiency testing (9).

Consequently, in this study, Fleiss' Kappa was carried out to determine the degree of agreement between five new screeners in the diagnosis of 40 thyroid FNAC cases without the clinical history in screening session A. Each new screener examined and interpreted the cases based on the TBSRTC classification (10). Therefore, the variables were measured in the categorical scale, and there was no overlap (mutually exclusive) between the six thyroid FNAC categories. Based on Table I, the kappa value obtained was κ= 0.541; this indicates that all students had a moderate level of reliability agreement in diagnosing thyroid FNAC cases without clinical history. Generally, a kappa value greater than 0.750 indicates an excellent deal, while a low of 0.40 suggests poor agreement (11,12).

Meanwhile, low levels of inter-rater reliability in healthcare are not acceptable, leading to poorer patient outcomes. According to McHugh's study in 2012, it claimed that having 40% of disagreements for a clinical laboratory would be a severe quality problem (4). Therefore, it suggested that 80% agreement is the minimum acceptable interrater agreement.

Furthermore, the kappa value for each category of thyroid FNAC cases according to the TBSRTC classification in screening session A also had been identified. The majority of new screeners can classify with certainty a benign (κ=0.904) and malignant (κ=0.863) diagnosis. However, they had difficulties in categorising indeterminate thyroid biopsy, which was three categories; atypical (κ = 0.407), follicular neoplasm (κ= 0.360), and suspicious thyroid biopsy (κ = -0.036), thus misclassified them into either benign or malignant. This condition arose because the characteristics of abnormal/atypical cells in the sample were insufficient to be diagnosed as cancer. Nevertheless, due to abnormal cells' presence, the sample cannot be diagnosed as benign. According to Alexander et al., in 2012, a definitive diagnosis of malignancy or benign can only be made in 70-85% of the nodules, with 15-30% falling into the "indeterminate" category (13). These indeterminate nodules are usually removed by surgery, and about 10-30% of those in the atypical or follicular neoplasm group may be cancerous. The negative Kappa (κ = -0.036) was obtained in Table I because there is less agreement than expected, despite the marginal distribution of ratings. In this context, malignancy suspects are not distributed. Therefore, if the new screener identifies the case as suspicious for malignancy, the particular category would have negative Kappa. Observing relying only on morphological characteristics can cause misdiagnosis. The slide observer tends to either show in false-positive, which shows the positive result when it is an actual negative, or false-negative result, which shows the negative result when it is positive when blindly screening without the clinical history. Poor diagnosis is made due to the absence of clinical information. Therefore, further explained as a slide observer will process the sample received and make the diagnosis based on their skills and knowledge.

In some cases, a slide observer will screen clinical history withholding to prevent bias in result interpretation. Therefore, this screening technique indeed requires high skills and lots of experience. Thus, a screener's ability to determine a diagnosis based on morphological characteristics alone is crucial (1).

Intra-observer reliability was statistically defined as the degree of agreement in the diagnostic test between a single rater's repetitive administration. This represents whether the observers' agreement or diagnosis is similar, consistent, or otherwise. This variation will affect the performance quality of reporting laboratory and patient management (14). Therefore, it is essential to evaluate slide observer reliability to produce the exact measurements under the same diagnosis consistently. Rescreening is a method to reduce variations and enhance screening efficiency (14). The data obtained from the first and second screening sessions were evaluated and defined by Cohen's kappa value (Table I). All new screeners had also diagnosed the same FNAC thyroid cases twice at various intervals within one month to assess the agreement. The arrangement number of cases was reshuffled. This was done to minimise the chance of obtaining bias outcomes from each new screener because they could remember the previous session's final diagnosis. The value of the reliability test ranges between -1 and 1. The value close to 1 shows perfect agreement, while the value closed to -1 shows perfect disagreement (15).

All new screeners had an 'almost perfect' agreement after comparing the diagnosis of both screening sessions as κ=0.905 (NS1), 0.852 (NS2), 0.927 (NS3), and 0.896 (NS4). However, a substantial agreement had been reported for NS5 (κ=0.702). Overall, each of the five raters demonstrated a good relationship between screening session ratings. Hence, the high intra-rater shows excellent consistency in providing a single observer diagnosis. Low intra-rater suggests that the rater had the inconsistency due to their failure to identify

thyroid FNAC cells. Apart from that, the most crucial factor is experience, which means recognising pitfalls and appearances relevant to specific sites and clinical presentations (2). According to Sharma in 2015, having a better understanding of the pitfall by analysing the false positive and false negative diagnoses would help avoid pitfalls and improve patient care (16).

Table II demonstrated the new screener's overall performance in the diagnostic interpretation of thyroid FNAC cases without clinical history. According to the study in 2004 done by Renshaw, Young, & Holladay, the average sensitivity of the blinded screening cases was reported to be approximately 70-80% (17). In this study, most new screeners were able to classify atypical/abnormal cells (TP) with an average sensitivity value of 87.38%, whereas benign/normal cells (TN) with an average specificity of 79.98%. High sensitivity leads to a lower specificity as both are inversely proportional. Sensitivity is specified to recognise and classify all abnormal cases (18). Identifying and classifying mild cases is known as specificity (18). Therefore, both parameters were useful to assess the new screener's ability to correctly classify cases as either benign/normal cells or atypical/abnormal cells. In this finding, the sensitivity value (87.38%) is higher than the specificity (79.98%), which indicates that few atypical/abnormal cases have been misinterpreted as benign (false negative). Many mild cases were misclassified as abnormal (false positive). This phenomenon is known as a pitfall.

Nonetheless, new screeners had difficulty classifying atypical cases as they misclassified them as either atypical/abnormal or benign cases. Screening tasks are challenging, with low target prevalence (3). Non-expert searchers show decreased false-positive results when targets are rare and increase false-positive results compared to results when the target is common.
(3). Cytology audits indicate that low numbers of abnormal cells in slides are associated with false-negative reporting.

The value of sensitivity and specificity was significant as the relationship can be expressed as a receiver operating characteristic (ROC) curve by plotting sensitivity (TP) versus 1-specificity (FP) (1). The area under the curve (AUC), which generally ranges from 0.5 to 1.0, can be measured to generate diagnostic accuracy for each slide observer (1). The AUC corresponding to an area of 0.5 indicates that the diagnostic accuracy was less attainable, while the area of 1.0 was a perfect diagnostic test (19, 20, 21). The further the curves from the 45-degree angle, the higher the diagnostic accuracy (19). In this study, the AUC value of the ROC curve showed a high value of 0.837 (83.7%). In Table II, all new screeners had a high overall diagnostic accuracy value except for NS1 (AUC = 0.630; 63%) as represented by the AUC value. Although the diagnostic accuracy can be determined from ROC, predictive value and likelihood ratio also can be measured to support the diagnostic accuracy result (22). In the previous study, the diagnostic accuracy is lower in screening slides without clinical history (1). However, a study done by Venrick & Sidawy, shows a high percentage of up to 90% diagnostic accuracy had been reported by certified cyto-screeners in screening slides without clinical history compared to this study, in which new screener exhibited an average diagnostic accuracy of up to 83.50% (Table II) (23). Nevertheless, most of the previous research focused on certified cyto-screeners with experiences working in cytology laboratory only, and there is a limited study among new screeners on the effect of diagnostic accuracy in screening without clinical history.

CONCLUSIONS

In conclusion, all new screeners from the Centre of Medical Laboratory Technology, Faculty of Health Sciences, UiTM Malaysia were able to diagnose thyroid cases accurately by 83.5% even without the presence of clinical history. Thus, it reflects that the process, technique, and skills introduced to each student are equal and easily understood in our facility. The blinded screening teaching method may continuously practice in the future to polish a new screener's true capabilities or competency.

ACKNOWLEDGMENTS
The authors would like to thank the Department of Medical Laboratory Technology, Faculty of Health Sciences, Universiti Teknologi MARA, Malaysia, to support and conduct this research in their facility.

REFERENCES

1. Raab SS, Oweity T, Hughes JH, Salomao DR, Kelley CM, Flynn CM, et al. Effect of clinical history on diagnostic accuracy in the cytologic interpretation of bronchial brush specimens. Am J Clin Pathol. 2000;114:78–83.
2. Orell SR. Pitfalls in fine-needle aspiration cytology. Cytopathology. 2003;14(4):173–82.
3. Evans KK, Tambouret RH, Evered A, Wilbur DC, Wolfe JM. Prevalence of abnormalities influences cytologists' error rates in screening for cervical cancer. Arch Pathol Lab Med. 2011;135(12):1557–60.
4. McHugh ML. Interrater reliability: The kappa statistic. Biochem Medica. 2012;22(3):276–82.
5. Jones BA, Davey DD. Quality management in gynecologic cytology using interlaboratory comparison. Arch Pathol Lab Med. 2000;
6. Hayen A, Dennis RJ, Finch CF. Determining the intra- and inter-observer reliability of screening tools used in sports injury research. J Sci Med Sport. 2007;10(4):201–10.
7. Gupta DK, Komaromy-Hiller G, Raab SS, Nath ME. Interobserver and intraobserver variability in

the cytologic diagnosis of normal and abnormal metaplastic squamous cells in Pap smears. Acta Cytol. 2001;45(5):697–703.
8. Schliep KC, Stanford JB, Chen Z, Zhang B, Dorais JK, Boiman Johnstone E, et al. Interrater and intrarater reliability in the diagnosis and staging of endometriosis. Obstet Gynecol. 2012;120(1):104–12.
9. Salmi DJ, Toth BD, Kong CS. Mock gynecologic cytology proficiency testing as a milestone assessment tool for anatomic pathology residents. J Am Soc Cytopathol. 2015;
10. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Am J Clin Pathol. 2009;132(5):658–65.
11. Liu K, Layfield LJ, Coogan AC, Ballo MS, Bentz JS, Dodge RK. Diagnostic accuracy in fine-needle aspiration of soft tissue and bone lesions: Influence of clinical history and experience. Am J Clin Pathol. 1999;111(5):632–40.
12. imundić A-M. Measures of Diagnostic Accuracy: Basic Definiti imundić, A.-M. (2009). Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC, 19(4), 203–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27683318ons. Ejifcc. 2009;
13. Alexander EK, Kennedy GC, Baloch ZW, Cibas ES, Chudova D, Diggans J, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. N Engl J Med. 2012;367(8):705–15.
14. Sushma ., Jacob R. Implications of inter-observer variability in cervical smear reporting. Int J Res Med Sci. 2017;5(9):4104–7.
15. Sim J, Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Phys Ther. 2005;85(3):257–68.
16. Sharma C. Diagnostic accuracy of fine-needle aspiration cytology of thyroid and evaluation of discordant cases. J Egypt Natl Canc Inst. 2015;27(3):147–53.
17. Renshaw AA, Young ML, Holladay EB. Blinded review of Papanicolaou smears in the context of litigation. Cancer. 2004;
18. Skaik Y. Understanding and using sensitivity, specificity, and predictive values. Indian Journal of Ophthalmology. 2008.
19. Raab SS. Diagnostic accuracy in cytopathology. Diagn Cytopathol. 1994;
20. Bulgaresi P, Cariaggi MP, Bonardi L, Carozzi MF, Confortini M, Galanti L, et al. Analysis of morphologic patterns of fine-needle aspiration of the breast to reduce false-negative results in breast cytology. Cancer. 2005.
21. Swreide K. Receiver-operating characteristic curve analysis in diagnostic, prognostic, and predictive biomarker research. J Clin Pathol. 2009;
22. Akobeng AK. Understanding diagnostic tests 1: Sensitivity, specificity, and predictive values. Acta Paediatrica, International Journal of Paediatrics. 2007.
23. Venrick MG, Sidawy MK. Cytologic evaluation of serous effusions: Processing techniques and an optimal number of smears for routine preparation. Am J Clin Pathol. 1993;99(2):182–6.