RESEARCH ARTICLE

The effect on absence of clinical history and demographic data of genitourinary cases in discovering agreement between inter- and intra-observers

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

[1]*Centre of Medical Laboratory Technology, Faculty of Health Sciences, Universiti Teknologi MARA, Cawangan Selangor, 42300 Puncak Alam, Selangor, Malaysia; [2]Department of Basic Sciences, Faculty of Health Sciences, Universiti Teknologi MARA, Cawangan Selangor, 42300 Puncak Alam, Selangor, Malaysia*

*Corresponding Author

Mohd Nazri Abu
Email: nazri669@salam.uitm.edu.my

Abstract:

In this genitourinary cytology study, glass slides were provided to the observers without demographic data and clinical history during cases review. The aim of this study is to evaluate inter- and intra- observer reliability in genitourinary cases without support of clinical history and demographic data information. Five slide observers were selected among students to screen 26 genitourinary cases using light microscope. Average of Fleiss Kappa and Cohen Kappa for inter- and intra-observer agreement was in between 0.1 to 0.5 indicating 'fair agreement'. The results concluded that Medical Laboratory Technology UiTM students have abilities in recognizing the genitourinary cytology slides based on morphological characteristics even without the support of demographic data information and history.

Keywords: Clinical history, diagnostic accuracy, reliability test, inter- and intra-reliability, genitourinary

1. INTRODUCTION

Cytology laboratory services consist of processing, evaluating and performing diagnostic interpretation on submitted specimens. All specimens are examined and interpreted by cytotechnologists and evaluated by pathologists for final diagnosis. Specimens received in the cytology laboratory must be attached with laboratory request form that will help cytotechnologists to make the diagnosis due to the presence of all the patients' data such as relevant clinical history and demographic data. Incomplete and false information from the laboratory request forms that can significantly impact the quality of laboratory results, patient care and safety.

Based on the previous study by Raab et al (2000), the absence of clinical history leads to diagnostic inaccuracy as the pathologists tend to report as reactive changes rather than malignant. Presence of clinical history contributes to increase diagnostic accuracy [1]. Competency of the participants in diagnosing cases may accurately evaluated. Reliability tests were conducted involving inter-observer (agreement within the same observer) and intra-observer (agreement between different observers) by using the kappa value [2].

Cytology laboratories play a crucial role to ensure better prognosis for patients. So, any mistakes as misdiagnosis should be prevented to avoid any unfavorable situation happen. Louise Nutt et.al (2008) stated that laboratory data influences 70% of medical diagnoses. Success of overall treatment to the patient is significantly depend on correct or complete data provided from the laboratory [3].

In this research, slide observers screened the slides without information on the clinical history and demographic data in cytology laboratory, Medical Laboratory Technology Department of UiTM Puncak Alam Selangor, Malaysia. The slide cases were selected based on cell morphological characteristics that can lead to misdiagnosis. This limitation will affect the inter- and intra-observer reliability in genitourinary cases diagnosis.

Reliability of slide observers may decrease thus the observer might face difficulty to make a perfect diagnosis with an absence of clinical history [4]. A slide observer will face uncertainty that will lead themselves to misdiagnose when there is no presence of clinical findings [5]. The extent of agreement test will be conducted to prove that UiTM students have skills, ability in genitourinary cytology

diagnosis and have the competency as a real slide observer in future.

2. MATERIALS AND METHODS

2.1 Study Design

A correlational study was conducted at cytology laboratory, Medical Laboratory Technology Department of UiTM Puncak Alam Selangor from March 2020 until July 2020. Correlation in this study is to investigate the relationship between absences of clinical history and demographic data with inter and intra reliability among slide observers in genitourinary cases.

2.2 Case Selection

26 slides of genitourinary cases were selected conveniently based on morphological characteristics. The amount of cases selected is measured by using the sample size formula. The confidence interval used is 95% confidence interval and α value of less than 0.05. The slides will be selected based on the slide's selection criteria. These slides were re-labelled by using number started from case 1 until 26.

2.3 Participant Selection

The participants are mainly the fourth-year Medical Laboratory Technology student that were selected based on the inclusion criteria for this study which are firstly the observers must have the knowledge about genitourinary cases and able to recognize cells. Second, the observers must have the skills in using light microscope. The third one is the observers must have screening cytological slides at least with one year of experience [6].

2.4 Screening Session

Five slide observers had been briefly explained to screen and diagnose twenty-six genitourinary cases at different time and date for twice screening session. Slide observers were given diagnosis worksheet to record their final diagnosis after a whole slide viewing and classified each case into benign, atypical or malignant. A month of gap between the first screening session and the second screening session was done to compare the agreement within the same slide observer for intra-observer reliability test [7,8]. Inter-observer reliability test was done by comparing the agreement among different slide observer [8].

2.5 Statistical Analysis

All data collected will be analysed using SPSS software (Statistical Package for Social Science for Windows version 24.0, IBM Corp, Armonk, New York, USA)

3. RESULT AND DISCUSSION

3.1 Inter-Observer Agreement (Fleiss Kappa)

Inter observer reliability was conducted to observe an agreement within the same slide observer which were selected according to inclusion and exclusion criteria. Fleiss kappa, k was used to analyze the inter- observers' agreement. The variables being assessed were in categorical scales which data represent categories such as malignant, atypical and benign. A range of values from -1 to 1 was interpreted. The value of -1 referred to as perfect disagreement and the value of 1 referred to as perfect agreement [9, 10].

Inter observer reliability measure the extent to which the independent slide observers produce similar diagnosis in the same genitourinary cases without demographic data and clinical history. It often expressed as a correlation coefficient. According to McHugh (2012), k value <0.20 indicated as 'Poor' agreement, 0.21-0.40 as 'Fair' agreement, 0.41-0.60 as 'Moderate' agreement, 0.61-0.80 as 'Good' agreement and 0.81-1.00 as 'Very good' agreement [11].

Table 1: Overall Fleiss Kappa in first screening session (A) and second screening session (B)

Screen session	Kappa	95% Confidence Interval (C.I.)	
		Lower	Upper
First (A)	0.295	0.186	0.403
Second (B)	0.347	0.244	0.450

Table 2: Kappa for individual categories in first screening session (A)

Rating Category	Kappa	95% Confidence Interval (C.I.)	
		Lower	Upper
Benign	0.391	0.269	0.512
Atypical	0.223	0.102	0.345
Malignant	0.097	0.024	0.219

Table 3: Kappa for individual categories in second screening session (B)

Rating Category	Kappa	95% Confidence Interval (C.I.)	
		Lower	Upper
Benign	0.375	0.251	0.499
Atypical	0.239	0.115	0.363
Malignant	0.581	0.457	0.705

Based on the result attained, the overall kappa outcomes showed 'fair agreement' with 0.295 for first screening session (A) and 0.347 for second screening (B). The 95% confidence interval (C.I) showed the value of Fleiss kappa as true or valid between 0.186 and 0.403 of marginal distribution during first screening session (A) and between 0.244 and 0.450 during second screening session (B). Since there was different in marginal distribution between the first and second sessions, the average value to both screening sessions could not be evaluated due to this statistical analysis limitation. The individual kappa was calculated for each variable which are malignant, atypical and benign.

Although there was no demographic data and clinical history being provided during both screening sessions, majority of the diagnosis being made had a fair agreement among them. However, the slide observers had difficulty in diagnosing benign cases as they misdiagnosed them into atypical, suspicious and malignant. Afterall, slide observers were able to diagnose malignant cases and majority of the benign cases.

3.2 Intra Observer Agreement (Cohen Kappa)

Intra observer reliability was conducted to observe an agreement between different slide observers which were selected according to inclusion and exclusion criteria. Value obtained by each slide observer during first and second screening session was calculated using Cohen's kappa value. Intra-observer reliability is the degree of agreement among repeated screening sessions of genitourinary cases performed by a single observer. The same genitourinary cases were diagnosed twice by the same slide observers within a month of gap. Consistency of the diagnosis reached by an observer was evaluated at least twice from the same genitourinary cases with the same method in order to show the reliability of the observers. According to Mchugh, 2012; the Kappa result are interpreted as follows: values ≤ 0; indicating no agreement and 0.01–0.20 as none to slight agreement, 0.21–0.40 as fair agreement, 0.41– 0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as almost perfect agreement [11].

Table 4: Cohen's kappa value in intra-observer reliability

Slide observer	Cohen's Kappa value	Strength of agreement
SO1A SO1B	-0.430	No agreement
SO2A SO2B	0.314	Fair agreement
SO3A SO3B	-0.156	No agreement
SO4A SO4B	0.217	Fair agreement
SO5A SO5B	0.262	Fair agreement

Based on the result obtained, three slide observers (SO2, SO4 and SO5) had 'fair agreement' after comparing diagnosis between first and second screening sessions while two slide observers, SO1 and SO3 had 'no agreement' with -0.430 and -0.156. It showed that the slides observers were not consistent with their own diagnosis due to absent of the demographic data and history.

4. CONCLUSION

In conclusion, this study showed that slide observers have only fair agreement in inter and intra reliability for the diagnosis of genitourinary cases. The slide observers were able to diagnose almost all cases without demographic data and clinical history. This proved that most of the slide observers have enough knowledge, experience and skills in screening genitourinary cases. The presence of demographic data and clinical history might be helpful for slide observers to diagnose genitourinary cases with a perfect diagnosis.

ACKNOWLEDGEMENTS

REFERENCES

[1] Raab, S. S., Oweity, T., Hughes, J. H., Salomao, D. R., Kelley, C. M., Flynn, C. M., Cohen, M. B. (2000). Effect of Clinical History on Diagnostic Accuracy in The Cytologic Interpretation of Bronchial Brush Specimens. American Journal of Clinical Pathology, 114(1), 78–83.

[2] Hayen, A., Dennis, R. J., & Finch, C. F. (2007). Determining The Intra- and Interobserver Reliability of Screening Tools Used in Sports Injury Research. Journal of Science and Medicine in Sport, 10(4), 201–210.

[3] N. Louise, E.Z Annalise E and T. E Rajiv T (2008), Incomplete laboratory request forms: the extent and impact on critical results at a tertiary hospital in South Africa

[4] Plebani, M., & Carraro, P. (1997). Mistakes in a Stat Laboratory: Types and Frequency. 1351, 1348-1351.

[5] Liu, K., Layfield, L. J., Coogan, A. C., Ballo, M. S., Bentz, J. S., & Dodge, R. K.(1999). Diagnostic Accuracy in Fine-needle Aspiration of Soft Tissue and Bone Lesions: Influence of Clinical History and Experience. American Journal of Clinical Pathology, 111(5), 632–640.

[6] Daniel W.W. (1999). Biostatistics: A Foundation for Analysis in The Health Sciences, New York: John Wiley & Sons, 7(4), 141-142.

[7] Sim, J., & Wright, C.C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Physical Therapy. 85(3), 257-268

[8] Bujang, M.A. & Baharum, N. (2017). Guidelines of The Minimum Sample Size Requirements for Cohen's Kappa. Epidemiology Biostatistics and Public Health. 14(2), 1-10.

[9] Altman, Douglas G. 1999. Practical Statistics for Medical Research. Chapman; Hall/CRC Press.

[10] Landis JR, Koch GG. 1977. "The Measurement of Observer Agreement for Categorical Data" 1 (33). Biometrics: 159–74.

[11] Mchugh, M.L. (2012). Interrater Reliability: The Kappa Statistic. Biochemia Medica. 22(3), 276-282.