THE INTRA- AND INTER-GRADER RELIABILITY OF MEIBOMIAN GLAND LOSS MEASUREMENT

NUR AIZZAH HANAN KAMARUL ZAMAN, BOptom.
DEPARTMENT OF OPTOMETRY AND VISUAL SCIENCE, KULLIYYAH OF ALLIED
HEALTH SCIENCES, INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA, JALAN
SULTAN AHMAD SHAH, BANDAR INDERA MAHKOTA, 25200 KUANTAN, PAHANG.

aizzah_93@yahoo.com

MOHD ZULFAEZAL BIN CHE AZEMIN, PhD (CORRESPONDING AUTHOR)
DEPARTMENT OF OPTOMETRY AND VISUAL SCIENCE, KULLIYYAH OF ALLIED
HEALTH SCIENCES, INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA, JALAN
SULTAN AHMAD SHAH, BANDAR INDERA MAHKOTA, 25200 KUANTAN, PAHANG.

zulfaezal@iium.edu.my

ABSTRACT

Introduction: Previous work employed digital image analysis using a fully-automated computer software to quantify changes in MG, which is meibomian gland loss. However, semi-automated software is more favorable for clinical applications as it allows clinicians to manually delete undesired noise or artifacts. *Purpose:* To study the intra- and inter-grader reliability of meibomian gland loss (MGL) analysis using a semi-automated computer software. *Methodology:* This is a retrospective study as participants' data were readily available from previous study. A total of 192 individual meibography images of meibomian gland (MG) from 48 subjects were loaded into ImageJ software. The MGL areas per total areas were manually traced and reported in percentage. The agreement within- and between-examiners were analyzed using intra-class correlation (ICC) and Bland-Altman plot. *Results:* ICC results showed only right and left upper eyelid were significantly correlated and showed excellent reliability for both intra- (RE: 0.944) (LE: 0.948) and inter-grader (RE: 0.831) (LE: 0.725) with *p*-value of less than 0.01. Meanwhile for the lower eyelid, the intra-grader result showed that only right lower eyelid had fair reliability with ICC of 0.525 (p = 0.005). The rest of the results showed poor reliability and the result was not significant. *Conclusion:* This method of measuring MGL area was only reliable in measuring the upper eyelid for both intra- and inter-grader. The definition of MG, and the interpretation of MGL area should be standardized. It is time consuming and might not be suitable to be used in routine eye examination.

INTRODUCTION

Reliability as defined by Joppe (2000) in Golafshani (2003) is 'the extent to which results are consistent over time and an accurate representation of the total population under study'. In a simpler word, it means that a test is considered reliable when the results remain the same throughout the study. Therefore, this study will be conducted to measure the reliability of changes that occurred in meibomian gland (MG) between- and within-examiner using semi-automated computer software. Semi-automated here means while using the software to quantify the changes in MG, there is also some interference from the clinician.

Arita *et al.* (2013) employed digital image analysis using a fully-automated computer software to quantify changes in MG, which is meibomian gland loss (MGL). In other words, the software did the grading of MGL by itself by using various methods and filters to come out with a processed MG images from meibography. They opted for fully-automated software as the inter-examiner variability might happen because there might be some differences arise between the examiners or clinicians while drawing the gland region.

Although this method is more convenient, but semi-automated software is more favorable for clinical applications as it allows clinicians to manually delete undesired noise or artifacts.

MATERIALS AND METHODS

This is a retrospective study design which was conducted at IIUM Kuantan Campus, International Islamic University Malaysia, Kuantan in which the participant's information and data are readily available from previous study. Therefore, written consent is not required. A total of 192 individual data images from a study (Abdul Rahman, 2015) were included in this study. ImageJ software, an open source imaging processing program was used to measure the MGL area.

A total of 192 individual images data of MG were loaded into ImageJ software. Examiner will trace the area of MGL per total MG area. ImageJ software analyzed the cropped region of MGL in percentage. Then, intra- and inter-examiner (2 examiners) measurements were evaluated (intra-examiner after 3 weeks). The data was then analyzed by using intra-class correlation together with Bland-Altman to find the agreement between two different observers.

RESULTS

For inter-examiner grading, only RE and LE upper eyelid shows significance finding based on value of ICC with ICC values of 0.831 and 0.725 respectively.

It can be seen on Figure 1, the regression slope does not show a trend, suggesting that there is no effect of the measurement magnitude between Examiner 1 and 2.
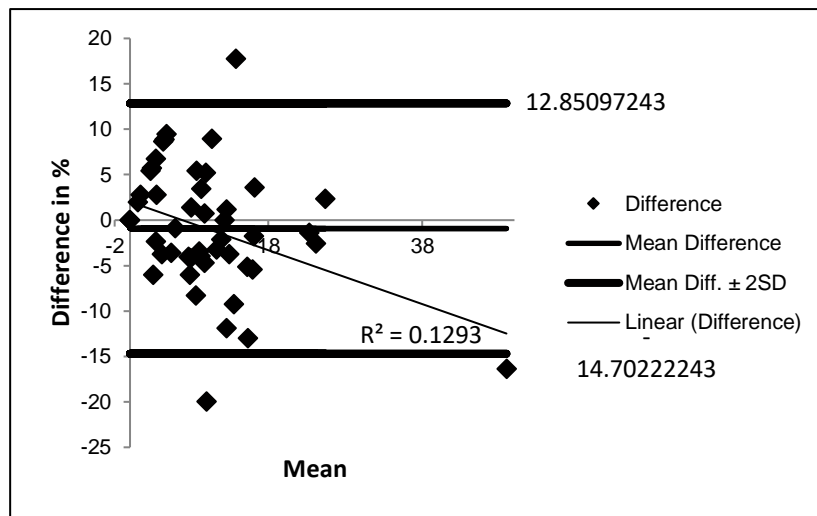
Figure 1 Bland-Altman plot of % of MGL for RE upper eyelid (inter-grader)

In Figure 2, a positive trend can be seen on the slope, suggesting that with increase percentage, the differences in measurement between Examiner 1 and 2 was larger, and that the measurement of Examiner 2 was consistently higher than Examiner 1.
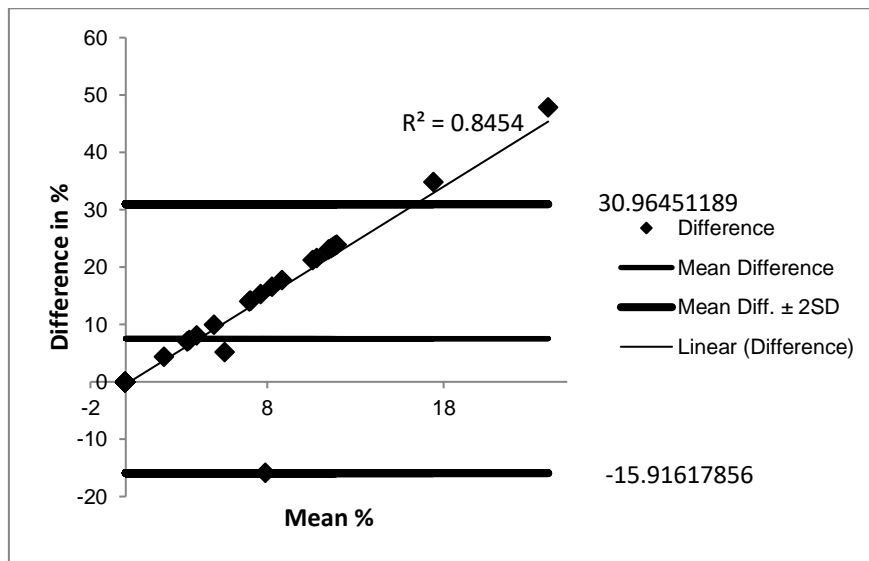

Figure 2 Bland-Altman plot of % of MGL for RE lower eyelid (inter-grader)

On the other hand, intra-examiner result also shows high reliability on the RE (ICC = 0.944) and LE (ICC = 0.948) upper eyelids and the values of ICC were significant (P < 0.001). Figure 3 shows similar result as Figure 1 with no effect of increase in magnitude detected.
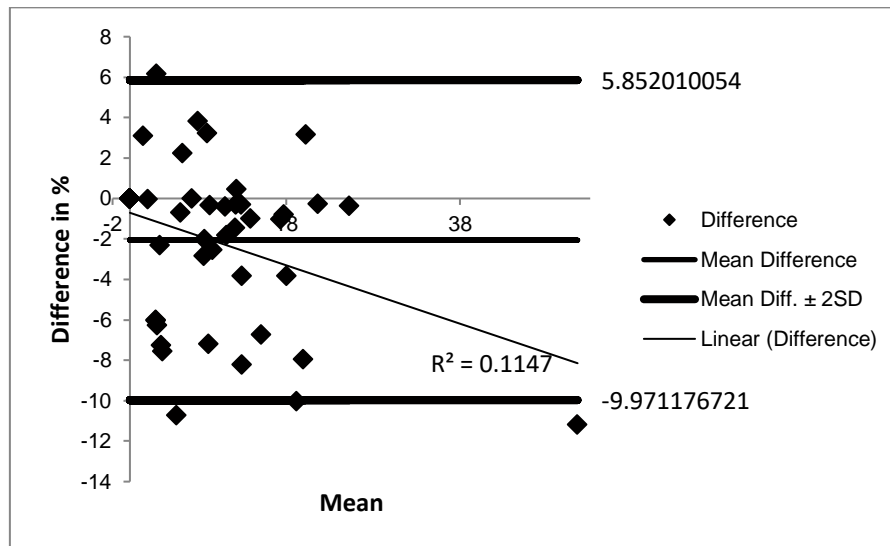
Figure 3 Bland-Altman plot of % of MGL for RE upper eyelid (intra-grader)

For lower eyelids, only RE shows fair reliability (ICC = 0.525) and the rest of the result shows poor reliability for both inter- and intra-examiner and the results were not significant.

DISCUSSION

In this study, a semi-automated software was used to measure the inter- and intra-grader reliability by identifying the MGL area from meibography images. A study (Pult & Riede-Pult, 2013) has been conducted to analyse the repeatability in meibography assessment by using subjective and objective method. The subjective method used was several examiners need to identify the MGL area based on the grading scale while objective method represents a semi-objective method using ImageJ software, which is same as this study. And the results show that the inter- and intra-grader agreement was much better with the objective method than with the subjective.

The data that was collected in this study was analyzed by using intra-class correlation (ICC) together with Bland-Altman to find the agreement between two different examiners.

A study by Cicchetti and Sparrow in 1981 had developed guidelines on how to classify the reliability of ICC. The guideline states that, 'when the reliability coefficient is below 0.40, the level of clinical significance is poor; when it is between 0.40 and 0.59, the level of clinical significance is fair; when it is between 0.60 and 0.74, the level of clinical significance is good; and when it is between 0.75 and 1.00, the level of clinical significance is excellent'.

Therefore, the inter- and intra-examiner reliability of area of MGL can be graded. The MGL area that were being traced are on the right eye (RE) upper eyelid, left eye (LE) upper eyelid, RE lower eyelid, and LE lower eyelid. For inter-examiner grading, only RE and LE

upper eyelid shows excellent significance level based on the value of ICC. This may be because the tarsal plates are more noticeable on the upper eyelid, despite the presence of some noise and artifacts.

Similarly, it can be seen on the Bland-Altman plot for RE and LE upper eyelid that the regression slope was not showing a trend, suggesting that there is not much difference in the measurement between Examiner 1 and 2. In contrast with the RE and LE lower eyelid, a positive trend can be seen on the regression slope, suggesting that with increase percentage, the differences in measurement between Examiner 1 and 2 was larger. As reported in Knop et al. (2011), the tarsal plate in the lower eyelid is 'smaller and forms a strip of rather equal length' and also, the MG in the lower eyelid tend to be wider compared to the upper eyelid, which can make the contrast of MG in lower eyelid not as good as the upper eyelid's MG. This can be the reason why the lower eyelid shows poor reliability.

Also, the definition of MG and the understanding of what is the loss area might differ between the observers, leading to differences in the result as they need to manually define the MGL region. Arita *et al*. (2013) had mentioned that inter-examiner variability may happen because there may be some differences arise between the examiners while drawing the gland region.

Next, for intra-examiner, the RE and LE upper eyelid result also shows high reliability based on the ICC values and Bland-Altman plot. For lower eyelid, only RE shows fair reliability and the rest of the result shows poor reliability for both inter- and intra-examiner. In spite of the differences in examiners' interpretation of what is loss, this might be due to some eyelid were not properly inverted, the images were not sharp, and also the presence of some noise during the images were taken. The almost similar problem happens in a study (Arita et al., 2013) whereby there is strong reflection of light illumination on the MG image which makes the tracing becomes difficult as there is no MG existed in the area. This needs an expert judgment to identify the MGL area.

CONCLUSION

In conclusion, this method of measuring MGL area was reliable in measuring for the upper eyelid only for both intra- and inter-grader as the result shows high reliability. In contrast, one of the causes for the poor reliability results is due to the different understanding of what is the MGL area, therefore it can be suggested that the definition of MG, and the interpretation of MGL area should be standardized among the examiners. The examiners might also sit together in analyzing the result to achieve a consensus on the loss area.

In the future, the number of examiners might need to be increased in order for the result to get better reliability than it is now as we can compare the result between many different examiners instead of two. And lastly, even though this method can be used to diagnose dry eyes and MGD, it is time consuming and might not be suitable to be used in routine eye examination.

REFERENCES

Arita, R., Itoh, K.,Inoue, K. & Amano, S. (2008). Noncontact Infrared Meibography to Document Age-Related Changes of the Meibomian Glands in a Normal Population. *Ophthalmology, 115*(5), 911-915.

Arita, R., Itoh, K., Inoue, K., Kuchiba, A., Yamaguchi, T. & Amano, S. (2009). Contact Lens Wear Is Associated with Decrease of Meibomian Glands. *Ophthalmology, 116*(3), 379-384.

Arita, R., Suehiro, J., Haraguchi, T., Shirakawa, R., Tokoro, H. & Amano, S. (2013). Objective image analysis of the meibomian gland area. *Br J Ophthalmol, 0,* 1-10.

Bron, A. J., Benjamin, L. & Snibson, G. R. (1991). Meibomian Gland Disease. Classification and Grading of Lid Changes. *Eye, 5,* 395-411.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report, 8*(4), 597-607. Green-Church, K. B., Butovich, I., Willcox, M., Borchman, D., Paulsen, F., Barabino, S. &

Glasgow, B. J. (2011).   The International Workshop on Meibomian Gland Dysfunction: Report of the Subcommittee on Tear Film Lipids and Lipid–Protein Interactions in Health and Disease. *Investigative Ophthalmology & Visual Science (IOVS), 52*(4), 1979-1993.

Knop, E., Knop, N., Millar, T., Obata, H. & Sullivan, D. A. (2011). The International Workshop on Meibomian Gland Dysfunction: Report of the Subcommittee on Anatomy, Physiology, and Pathophysiology of the Meibomian Gland. *Investigative Ophthalmology & Visual Science (IOVS), 52*(4), 1938-1978.

N. Farhana A. R. (2015). Comparison of Meibomian Glands Loss among Contact Lens Wearer and Non-Wearer.

Nichols, J. J., Berntsen, D. A., Mitchell, G. L. & Nichols, K. K. (2005). An Assessment of Grading Scales for Meibography Images. *Cornea, 24*(4), 382-388. Nichols, K. K., Foulks, G. N., Bron, A. J., Glasgow, B. J., Dogru, M., Tsubota,

K.,….Sullivan, D. A. (2011). The International Workshop on Meibomian Gland Dysfunction: Executive Summary. *Investigative Ophthalmology & Visual Science (IOVS), 52*(4), 1922-1929.

Powell, D. R., Nichols, J. J. & Nichols, K. K. (2012). Inter-Examiner Reliability in Meibomian Gland Dysfunction Assessment. *Investigative Ophthalmology & Visual Science (IOVS), 53*(6), 3120-3125.

Pult, H. & Nichols, J. J. (2012). A Review of Meibography. *Optometry and Vision Science, 89*(5), E760-E768.

Pult, H. & Riede-Pult, B. (2011). Non-contact meibography: Keep it simple but effective. *Contact Lens & Anterior Eye, 35*(2012), 77-80.

Pult, H. & Riede-Pult, B. (2012). Comparison of subjective grading and objective assessment in meibography. *Contact Lens & Anterior Eye, 36*(2013), 22-27.

Remington, L. A. (1998). Physiology of the Eye. USA: Butterworth-Heinemann.
Sirigu, P., Shen, R. L. & Da Silva, P. P. (1992). Human Meibomian Glands: The Ultrastructure of Acinar Cells as Viewed by Thin Section and Freeze-Fracture Transmission Electron Microscopies. *Investigative Ophthalmology & Visual Science (IOVS), 33*(7), 2284-2292.

Zhao, Y., San Tan, C. L. & Tong, L. (2015). Intra-observer and inter-observer repeatability of ocular surface interferometer in measuring lipid layer thickness. *BMC Ophthalmology, 15*(53), 1-7.