Contact Lens and Anterior Eye 47 (2024) 102281

Contents lists available at ScienceDirect

Contact Lens and Anterior Eye

journal homepage: www.elsevier.com/locate/clae

Ocular surface parameters repeatability and agreement —A comparison between Keratograph 5M and IDRA

Ka Yin Chan [a,1], Xulin Liao [a,b,1], Biyue Guo [a], Jimmy S.H. Tse [b], Peter H. Li [a], Allen M.Y. Cheong [a,b,c], William Ngo [a,d,*], Thomas C. Lam [a,b,c,**]

[a] Centre for Eye and Vision Research (CEVR), 17W Hong Kong Science Park, Hong Kong
[b] Centre for Myopia Research, School of Optometry, The Hong Kong Polytechnic University, Hong Kong
[c] Research Centre for SHARP Vision (RCSV), The Hong Kong Polytechnic University, Hong Kong
[d] School of Optometry & Vision Science, University of Waterloo, Waterloo, Ontario, Canada

ARTICLE INFO

Keywords:
Repeatability
Tear film stability
Dry eye
NIBUT
IDRA
Keratograph 5M

ABSTRACT

Purpose: To evaluate the repeatability and agreement in dry eye measurements using Oculus Keratograph 5M (K5M) and SBM Sistemi IDRA (IDRA).
Methods: A total of 108 participants were enrolled and 108 eyes were evaluated. Tear meniscus height (TMH) and first and average non-invasive break-up time (NIBUT) were measured using the K5M and IDRA (order randomly assigned). TMH was measured using the built-in caliper tool while NIBUT was computed by the automatic algorithm of the instruments.
Results: The Bland Altman plots analysis showed a good agreement between the two instruments for TMH (95 % Limits of Agreement (LoA), −0.17 to 0.16), but not the first NIBUT (95 % LoA, −8.13 to 14.79) and average NIBUT (95 % LoA, −7.89 to 10.32). The values of the first and average NIBUT measured using IDRA were significantly shorter than in K5M (difference = median (IQR) −2.75 (−6.48- −0.28)s, $p < 0.001$ and difference = median (IQR) −1.65 (−3.97–1.89)s, $p = 0.008$ respectively). The TMH ($p = 0.037$) and NIBUT average ($p = 0.033$) measured by K5M, as well as the TMH ($p = 0.040$) measured by IDRA, exhibited unstable measurements across the three measurement times. The remaining parameters exhibited stability with three repeated measurements.
Conclusion: The NIBUT measurements are not interchangeable between IDRA and K5M, while the TMH was little difference between the two instruments. It is important to exercise caution when using different ocular surface analyzers to minimize errors in comparing multiple measurements.

1. Introduction

Dry eye disease (DED) is characterized by a loss of tear film and ocular surface homeostasis [1], with a prevalence that range from 5 % to 50 % in the adult population [2]. Clinical signs such as tear osmolarity, ocular surface staining, and tear film stability are important in aiding eye care practitioners in diagnosing DED [3,4]. Tear osmolarity and tear film stability can be quantitatively assessed using commercially available instruments [5], and tear film stability can be evaluated objectively by measuring the non-invasive tear breakup time (NIBUT) using an ocular surface analyzer or corneal topographer. In comparison with osmolarity and ocular surface staining assessment, NIBUT is relatively simple to perform and does not require consumables. However, there are multiple instruments capable of automated NIBUT measurement and they all use different algorithms for determining NIBUT [6]. These differences could translate to differing NIBUT values, which may impact the clinical diagnosis of DED.

Tear meniscus height (TMH) is a parameter that indirectly assesses the tear volume to enable sub-classification of DED [3]. Although tear volume can be assessed using anterior segment optical coherence tomography (OCT) [7–9], conventional methods of measuring the inferior tear meniscus via imaging is still commonly used in clinical practice. The

* Corresponding author at: OPT 255, School of Optometry & Vision Science, University of Waterloo, Waterloo, Ontario, Canada.
** Corresponding author at: HJ537, School of Optometry, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
E-mail addresses: william.ngo@uwaterloo.ca (W. Ngo), thomas.c.lam@polyu.edu.hk (T.C. Lam).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.clae.2024.102281
Received 25 February 2024; Received in revised form 25 May 2024; Accepted 22 July 2024
Available online 2 August 2024

use of Oculus Keratograph 5M (K5M) and SBM Sistemi IDRA (IDRA) allow the capture of digital images and measurements are made using a calibrated digital caliper. Similar to the measurement of NIBUT, differences in algorithm and illumination system of instrument may influence the TMH values [10]. As such, evaluating the agreement of NIBUT and TMH values obtained from different instruments and understand whether measurements could be compared between instruments is warranted. The aim of this study was to evaluate the repeatability and agreement of NIBUT and TMH measurements between two commercially available tear film analyzers.

2. Method

2.1. Subjects

This study was part of a large-scale dry eye population study in Hong Kong and the enrollment criteria adhered to the population study. A total of 108 participants who had resided in Hong Kong continuously for 3 years were recruited. The inclusion criteria involved individuals aged 18 years or older, while the exclusion criteria encompassed individuals who wore contact lenses or used any eyedrops within 2 hours prior to the visit. The analysis was conducted using data from the right eye only. Participants were asked not to wear contact lenses and instill any eyedrops on the day of the visit. Eligible participants who attended between September 2021 and September 2023 were included in the study. The study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of The Hong Kong Polytechnic University.

2.2. Procedures

The Ocular Surface Disease Index (OSDI), and 5-Item Dry Eye Questionnaire (DEQ-5) were administered to evaluate the severity of DED symptoms [11]. Tear profiles (i.e., NIBUT and TMH) were evaluated in a random order in a temperature and humidity-monitored clinic cubicle using the Oculus Keratograph 5M (K5M) (Oculus, Wetzlar, Germany) and IDRA (Sbm Sistemi, Inc., Torino, Italy, Full version). The NIBUT (first and average) and TMH were measured using both instruments. During NIBUT measurement, participants were instructed to maintain fixation and refrain from blinking. A video was recorded during the measurement and the recording was stopped if either of the conditions were met: 1) after 25 s had elapsed, or 2) when the software detected a blink a break in the tear film. Three measurements were taken and averaged. 2 min break was given between each NIBUT measurement to allow the tear film to re-stabilize. The results of the first break up time (NIBUT first) and average break up time (NIBUT average) were computed by the automatic algorithm of the instruments. TMH was measured from a static captured image of tear meniscus of the lower eyelid, and the averaged TMH was determined at the centre of the lower lid using the built-in caliper tool of the software. Three TMH measurements were obtained and averaged. All the measurements within the study were performed by one trained practitioner to maintain measurement consistency. Measurements were performed on both eyes but only data of the right eye is analyzed.

2.3. Statistical analysis

For sample size calculation, at least 73 eyes were required to detect a difference of 2 s in NIBUT with a standard deviation of 6 s, achieving 80 % power at a two-sided statistical significance level of 5 %. Statistical analyses were performed using SPSS Software (version 26.0, SPSS, Inc, Chicago, IL, USA) and GraphPad Software (version 8.0.1, CA, USA). The Shapiro–Wilk test was used to assess TMH, NIBUT first and NIBUT average data for normality [12]. The data of TMH, NIBUT first, and NIBUT average were not normally distributed ($p < 0.05$). However, the TMH data obtained from the IDRA showed a normal distribution ($p = 0.130$). Therefore, non-parametric tests were employed for statistical analysis. The differences between the three repeated measurements are analyzed using the Friedman test [13] (Table 2). Wilcoxon signed-rank test [14] was used to analyze the differences in the average of 3 measurements results between two machines for the same subject (Table 3). Non-parametric Bland Altman plots were used to assess the differences between instruments against their means [15]. A p-value of less than 0.05 was used to indicate statistical significance.

3. Results

A total of 120 subjects were recruited and underwent the dry eye assessment. Among them, 12 of them were excluded because of various reasons such as wearing contact lenses on the day of measurement (n = 3), prosthesis in the right eye (n = 1) and missing pertinent data due to instrument failure or patients' intolerance to the tests (n = 8). In the 108 subjects analyzed, 38 of them (35.2 %) were asymptomatic, 37 (34.3 %), 19 (17.6 %) and 14 (13.0 %) of them had mild, moderate, and severe symptoms, respectively, based on the OSDI scores [3]. The participant demographics, symptomology and tear film parameters are presented in Table 1.

Table 2 displays the inter-session repeatability of tear film parameters. The Friedman test was employed to assess the differences between three measurements from the same machine. The results revealed that the TMH of K5M measured three times (0.23 (0.18–0.28)mm, 0.23 (0.19–0.27)mm, and 0.23 (0.19–0.28)mm respectively, with $p = 0.037$) and IDRA (0.23 (0.20–0.28)mm, 0.25 (0.20–0.29)mm, and 0.24 (0.20–0.26)mm respectively, with $p = 0.040$), as well as the average NIBUT (12.34 (8.13–17.67)s, 14.08 (8.93–19.46)s, and 12.31 (7.16–17.72)s respectively, with $p = 0.033$) in K5M, displayed inconsistent measurements across the three measurements.

Table 3 illustrates the agreement between the two instruments on tear film measurements. Use the paired-samples Wilcoxon signed-rank test to analyze differences in the average of 3 measurements results between two machines for the same subject. The TMH results showed no significant difference. However, the NIBUT first in IDRA (5.63 (4.99–7.88) s) was significantly shorter (difference = −2.75 (−6.48-−0.28)s, $p < 0.001$) than the NIBUT first in K5M (9.26 (6.33–13.92) s). Additionally, the NIBUT average in IDRA (13.00 (8.87–14.43) s) was also significantly shorter (difference = −1.65 (−3.97–1.89) s, $p = 0.008$) than the NIBUT average in K5M (13.00 (8.88–18.07) s). Non-parametric Bland Altman plots indicated satisfactory agreement in TMH (95 % Limits of Agreement (LoA), −0.17 to 0.16) (Fig. 1, A) but revealed poor agreement in NIBUT first (95 % LoA, −8,13 to 14.79) (Fig. 1, B) and average (95 % LoA, −7.89 to 10.32) (Fig. 1, C) between the two devices.

4. Discussion

This study investigated the repeatability and agreement of ocular surface parameters, making a comparison between the K5M and IDRA. It is the first study conducted on the repeatability and agreement of ocular surface parameters of these two machines in the Chinese population. This research findings indicate differences in NIBUT results between the

Table 1
Demographic and ocular surface characteristics of the eligible participants.

Parameters	
Number of patients	108
Number of eyes	108
Age (years)	46.80 ± 14.80
Female (proportion)	77 (71 %)
Ocular Surface Disease Index	16.50 (8.00–27.00)
Dry Eye Questionnaire-5	7.00 (5.00–10.00)

Age is presented as mean ± standard deviation; Ocular Surface Disease Index and Dry Eye Questionnaire-5 are presented as median (interquartile range).

Table 2
Inter-session repeatability of tear film parameters. A total of 108 eyes were examined, with each eye undergoing three measurements. The comparisons between the three measurements were made using the #Friedman test.

	Measurement	Median (IQR)	*p-value*#
TMH (mm)			
	I	0.23 (0.20–0.28)	
IDRA	II	0.25 (0.20–0.29)	0.040
	III	0.24 (0.20–0.26)	
	I	0.23 (0.18–0.28)	
K5M	II	0.23 (0.19–0.27)	0.037
	III	0.23 (0.19–0.28)	
NIBUT first (s)			
IDRA	I	5.36 (4.36–5.88)	
	II	5.52 (4.84–7.28)	0.302
	III	5.56 (4.93–6.65)	
	I	7.97 (5.36–7.96)	
K5M	II	8.92 (5.45–8.92)	0.302
	III	8.41 (4.59–13.38)	
NIBUT average (s)			
	I	12.63 (7.78–14.31)	
IDRA	II	12.72 (8.45–14.64)	0.221
	III	13.29 (8.76–14.53)	
	I	12.34 (8.13–17.67)	
K5M	II	14.08 (8.93–19.46)	0.033
	III	12.31 (7.16–17.72)	

Abbreviation
TMH, tear meniscus height. NIBUT, non-invasive tear break-up time; K5M, Keratograph 5 M. IQR, interquartile range; mm, millimeter; s, second.

Table 3
Inter-instrument agreement of tear film parameters.

	IDRA	K5M	Difference	*p-value**
	Median (IQR)	Median (IQR)	Median (IQR)	
TMH (mm)	0.24 (0.21–0.28)	0.22 (0.19–0.28)	0.02 (−0.05–0.05)	0.642
NIBUT_first (s)	5.63 (4.99–7.88)	9.26 (6.33–13.92)	−2.75 (−6.48–0.28)	<0.001
NIBUT_average (s)	13.00 (8.87–14.43)	13.00 (8.88–18.07)	−1.65 (−3.97–1.89)	0.008

*Paired samples Wilcoxon signed rank test. **Abbreviation:** TMH, tear meniscus height. NIBUT, non-invasive tear break-up time, K5M, Keratograph 5 M.

two devices, with IDRA showing a shorter first and average NIBUT compared to the K5M. Additionally, there was poor repeatability in the three measurements of TMH and NIBUT average for the K5M, while the IDRA exhibited poor repeatability in TMH.

NIBUT, as one of the key criteria for assessing the objective dry eye disease, play a crucial role in diagnosing DED. The accuracy of NIBUT, especially when the NIBUT first was less than 10 s, was highly significant in the diagnosis of dry eye disease [3]. This research indicated that the NIBUT measured by the IDRA device, both in terms of the first and average values, was shorter than the values measured by the K5M. This implied that IDRA is more likely to generate a dry eye disease diagnosis based on this parameter. The difference in NIBUT may be attributed to differences in algorithms and sensitivity or by the possibility that subjects were less able to tolerate the illumination by IDRA compared to the infrared light from K5M, leading to earlier tear break-up. Another possible factor contributing to the different results is that K5M has 22 Placido disk rings [16], while IDRA has approximately 8 rings based on the examination images of the patients. The results from Singh et al [17] is also aligned with this research findings, showing that NIBUT measured by IDRA were lower than the K5M. The difference was that their study found good repeatability and reproducibility for NIBUT with K5M and IDRA in the normal group, while in the dry eye group, the opposite was observed. In contrast, this research results indicated that K5M had poor repeatability, while IDRA had good repeatability in NIBUT measurement. The discrepancies in results may have stemmed

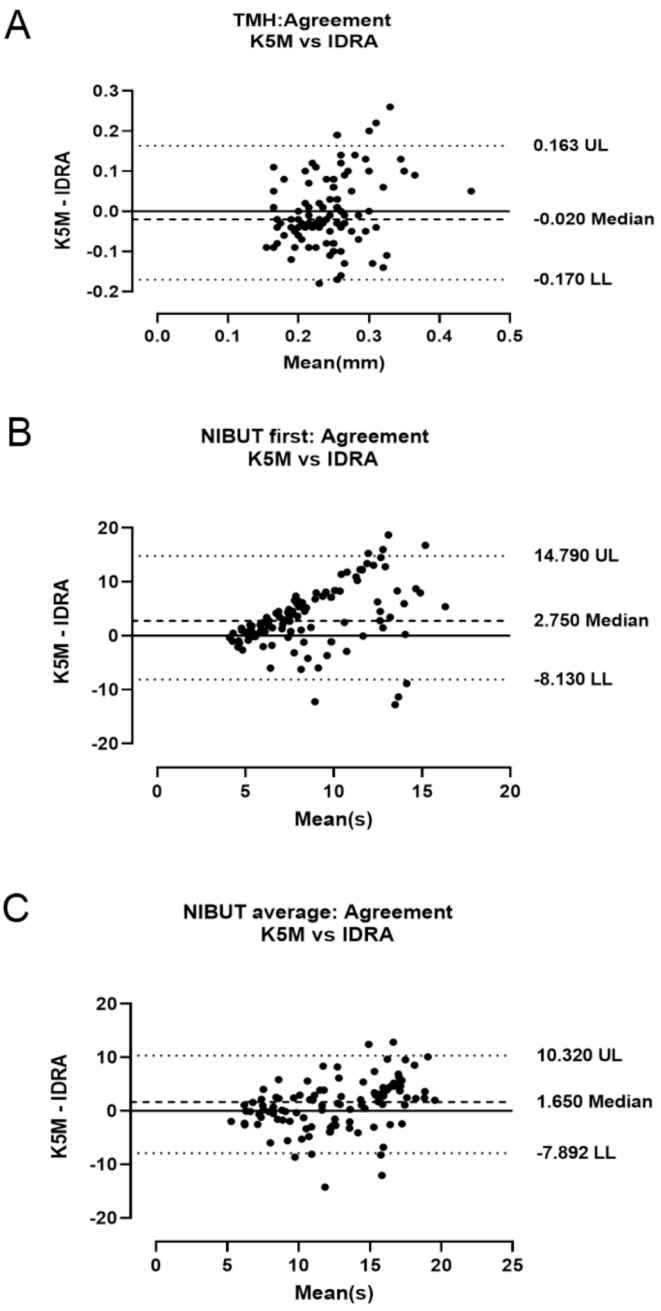

Fig. 1. Non-parametric Bland and Altman's plots (n = 108) of between-instrument difference against average of two instruments of (A) tear meniscus height (TMH), (B) first non-invasive tear breakup time (NIBUT first) and (C) average non-invasive tear breakup time (NIBUT average). The bilateral blacked dotted lines show the upper (UL) and lower (LL) 95 % limits of agreement.

from differences in this study participants were not grouped compared to theirs. This might also be a limitation of this study. The study results are valuable for guiding the operation of dry eye testing machines for complex clinical patients. In future research, different groups of the population will be studied to determine whether the results differ. The study by Ryan Lee et al. [18] found that there is also low agreement on NIBUT between the Tomey RT-7000 Auto Refractor-Keratometer (Tomey Corporation, Nagoya, Japan) and K5M. It appears that this issue was common across various instruments. Therefore, when using instruments for automated NIBUT assessment, it would be helpful to understand the values relative to the instruments and their limitations.

In this study, it was observed that K5M showed inconsistent results in

three consecutive measurements for TMH and NIBUT averages. Similar to this research findings, Jose et al. [19] reported that NIBUT measurements in K5M have low repeatability among examiners, even when accounting for factors such as sex, age, and dry eye disease diagnosis. They attribute this low repeatability not only to the device but also primarily to the inherent variability of the tear film. In contrast, according to Singh et al., K5M showed good repeatability in NIBUT when tested on healthy rabbits [20], which is different from the results of another set of experiments they conducted on a dry eye subjects [17]. The K5M is a commonly-used instrument that evaluate dry eye disease and holds significant value for guiding clinical decisions [21], however, further investigation is necessary to understand variations in the data between different groups or populations. This work had limitation. Continuous eye-opening in the NIBUT testing may have caused participants to experience fatigue after multiple measurements, which might have contributed variability to the repeatability of the tests.

In conclusion, the research on the repeatability and agreement of ocular surface parameters using K5M and IDRA revealed that there was poor agreement in NIBUT, with IDRA showing a shorter value. Therefore, the IDRA and K5M are not interchangeable with regard to NIBUT measurement. The K5M also demonstrated poorer repeatability in TMH and NIBUT average while IDRA showed poorer repeatability in TMH. Therefore, the use of these dry eye testing instruments necessitated exercising greater caution to ensure accurate measurements and data interpretation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The study was supported by InnoHK and the Government of the Hong Kong Special Administrative Region.

References

[1] Craig JP, et al. TFOS DEWS II definition and classification report. Ocul Surf 2017;15(3):276–83.
[2] Stapleton F, et al. TFOS DEWS II epidemiology report. Ocul Surf 2017;15(3):334–65.
[3] Wolffsohn JS, et al. TFOS DEWS II diagnostic methodology report. Ocul Surf 2017;15(3):539–74.
[4] Bron AJ, Evans VE, Smith JA. Grading of corneal and conjunctival staining in the context of other dry eye tests. Cornea 2003;22(7):640–50.
[5] Lemp MA, et al. Tear osmolarity in the diagnosis and management of dry eye disease. Am J Ophthalmol 2011;151(5):792–798.e1.
[6] Bandlitz S, et al. Agreement and repeatability of four different devices to measure non-invasive tear breakup time (NIBUT). Cont Lens Anterior Eye 2020;43(5):507–11.
[7] Arriola-Villalobos P, et al. Assessment of lower tear meniscus measurements obtained with Keratograph and agreement with Fourier-domain optical-coherence tomography. Br J Ophthalmol 2015;99(8):1120–5.
[8] Baek J, Doh SH, Chung SK. Comparison of tear meniscus height measurements obtained with the keratograph and fourier domain optical coherence tomography in dry eye. Cornea 2015;34(10):1209–13.
[9] Chan HH, et al. Repeatability of tear meniscus evaluation using spectral-domain Cirrus® HD-OCT and time-domain Visante® OCT. Cont Lens Anterior Eye 2015;38(5):368–72.
[10] Baek J, Doh SH, Chung SK. Comparison of tear meniscus height measurements obtained with the keratograph and fourier domain optical coherence tomography in dry eye. Cornea 2015;34(10):1209–13.
[11] Pult H, Wolffsohn JS. The development and evaluation of the new ocular surface disease index-6. Ocul Surf 2019;17(4):817–21.
[12] Hanusz Z, Tarasinska J, Zielinski W. Shapiro–Wilk test with known mean. Revstat Stat. J. 2016;14(1):89–100.
[13] Zimmerman DW, Zumbo BD. Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. J Exp Educ 1993;62(1):75–86.
[14] Blair RC, Higgins JJ. Comparison of the power of the paired samples t test to that of Wilcoxon's signed-ranks test under various population shapes. Psychol Bull 1985;97(1):119.
[15] Bland JM, Altman DG. Statistics notes: measurement error. Br Med J 1996;312(7047):1654.
[16] Ortiz-Toquero S, et al. Repeatability of Placido-Based Corneal Topography in Keratoconus 2014;91(12):1467–73.
[17] Singh S, et al. Repeatability, reproducibility and agreement between three different diagnostic imaging platforms for tear film evaluation of normal and dry eye disease. Eye (Lond) 2023;37(10):2042–7.
[18] Lee R, et al. Agreement of noninvasive tear break-up time measurement between Tomey RT-7000 auto refractor-keratometer and oculus keratograph 5M. Clin Ophthalmol 2016;10:1785–90.
[19] García-Marqués JV, et al. Repeatability of non-invasive keratograph break-up time measurements obtained using oculus keratograph 5M. Int Ophthalmol 2021;41(7):2473–83.
[20] Singh S, et al. Repeatability of non-invasive tear film evaluation in healthy rabbit eyes. Curr Eye Res 2023;48(8):699–703.
[21] García-Montero M, et al. Repeatability of noninvasive keratograph 5M measurements associated with contact lens wear. Eye Contact Lens 2019;45(6):377–81.