Agreement and repeatability of objective systems for assessment of the tear film

Joaquín Fernández, MD, PhD;[1,2] Manuel Rodríguez-Vallejo, OD, PhD;[1][*] Javier Martínez, OD;[1] Ana Tauste, OD, PhD;[1] Javier García-Montesinos, MD;[1] David P Piñero, OD, PhD[3,4]

[1]Department of Ophthalmology (Qvision), Vithas Virgen del Mar Hospital, 04120, Almería, Spain
[2]Department of Ophthalmology, Torrecárdenas Hospital Complex, 04009, Almería, Spain
[3]Department of Optics, Pharmacology and Anatomy, University of Alicante, Alicante, Spain
[4]Department of Ophthalmology (OFTALMAR), Vithas Medimar International Hospital, Alicante, Spain

[*]*Corresponding author: david.pinyero@ua.es (Tel +34686500808)*

Number of Figures: 1   Number of Tables: 2

Submitted: February 15, 2018

1. Abstract.

Purpose: To assess the agreement and repeatability of two objective systems for measuring the tear film stability.

Methods: Retrospective analysis of the tear film stability including 99 healthy right eyes measured with a videokeratoscope (VK) and the Optical Quality Analysis System (OQAS, Visiometrics). Two consecutive measures (trials) were taken with both systems, with an interval of 10 minutes between devices. Variables included in the study were first and mean non-invasive break up times (NIBUT and MNIBUT) for VK and mean and standard deviation of the optical scattering index (OSIm and OSIsd) for OQAS. The agreement and repeatability of grading scales provided by both devices were also evaluated using the Cohen's k with quadratic weights. The ocular surface disease index (OSDI) questionnaire was also passed out to all subjects. Correlations and associations between subjective and objective metrics were analyzed.

Results: Mean significant differences were found between trials for NIBUT (p=0.04) and MNIBUT (p=0.01) but not for differences OSIm (p=0.11) and OSIsd (p=0.50). Grading scales resulted in a fair (k=0.20) or poor agreement (k=0.04) between systems if the first or second trial was considered. The repeatability of the grading scale was good for OQAS (k=0.59) and fair for VK (k=0.37). No significant correlations or associations were found between OSDI and any of the metrics obtained with both devices (p≥0.36).

Conclusions: The two devices evaluated cannot be used interchangeably for the assessment of tear film stability. Good intrasession repeatability was obtained for tear film grading of the OQAS whereas it is fair for VK.

Key Words: tear film stability, objective systems, ocular scatter index, non-invasive break-up time, laser refractive surgery.

2. Introductory Text

Dry eye after corneal laser refractive surgery (LRS) is one important concern for anterior segment surgeons and represents one reason for exclusion criteria.[1] It is suggested that almost all the patients have at least mild symptoms after their procedure, with persistence of symptoms at 6 months postoperatively in 20-55% of patients.[2] In fact, it has been reported that 21.1% of subjects with complaints after some LRS procedures are due to dry eye.[3] The main reason for the generation of this problem is the decrease in the corneal sensation due to the surgical dissection of corneal nerve fibers.[4] Consequently, the blinking rate is reduced, the basal tear secretion decreases and some alterations of the tear film appear, such as reduced tear menisci,[5] which can be recovered with corneal reinnervation after a period of time.[6]

The lack of association between signs and symptoms in patients with dry eye disease has been widely addressed[7] and some authors have reported the presence of symptomatology without alteration of objective clinical signs of tear insufficiency after LRS.[8] In fact, Farahi et al[9] reported a rate of 15.9% of abnormal subjects according to McMonnies questionnaire after LRS, whereas the double (30.8%) were classified as abnormal according to the fluorescein break-up time. Recently, a less invasive technique named SMILE (small incision lenticule extraction) has been proven to reduce the postoperative incidence of dry eye symptoms in comparison to LASIK, although no clear differences in objective clinical signs have been found between procedures.[10]

There are numerous diagnostic tests for dry eye that are not widely accepted yet and are often not reproducible, probably due to the complexity of the disease and the high variability between seasons, time, examiners, etc.[11] New technologies have emerged in order to assess objectively the tear film in a non-invasive way. These technologies are used by the refractive surgeons despite there is still not clear evidence about its diagnostic ability and which are the most adequate cut-off parameters for screening in refractive surgery. For creating clinical guidelines around these new technologies, it is required first to know and understand the repeatability of the metrics provided by each system considering the high variability of the tear film.[11] The aim of this study was to assess the agreement and repeatability of non-invasive break-up time measured with a commercial videokeratoscope (VK) and a double pass system (OQAS, optical quality analysis system), and of the optical scattering index (OSI) provided by the double pass device in a sample of subjects during the screening exam for refractive surgery.

3. Subjects and Methods

a. Subjects

This study was approved by the local ethics committee of research and was performed in adherence to the tenets of the Declaration of Helsinki. Data from 99 right healthy eyes measured during a screening for refractive surgery at Qvision (Department of Ophthalmology, Virgen del Mar Hospital, Almería, Spain) were extracted from our historical database. Exclusion criteria for the inclusion of cases were any ocular disease that may affect to eye optical quality, such as keratoconus and other ectasias, pterygium, etc. Furthermore, eyes with spherical refractive error from -8 D to +5 D and astigmatisms under

2.75 D were excluded from the analysis with the OQAS parameters but were included in the repeatability analysis with the VK. Measurements with both devices were always obtained in the following order according to the habitual protocol in our clinical practice: first, two consecutives measures were taken with the VK and after 10 minutes, two consecutive measures were taken with the OQAS system. Finally, subjects answered the Ocular Surface Disease Index questionnaire (OSDI) and continued other screening procedures not relevant for the current study.

b. Devices and variables

The VK used in this study was the Keratograph 5M system (Oculus, Optikgerate, Germany), which provides a measure of the non-invasive break-up time (NIBUT) by means of detecting perturbations of the Placido rings projected on the tear film. Each subject was asked to maintain the eyes opened after blinking two times in a low-light room (1 lux) during the measure with infrared light. For some subjects, the blink can terminate the measurement before a potential break-up of the tear film. In cases like these, the trial was repeated if the system provided the message "too short time" or was accepted, classifying the tear film as stable and recording the blinking time (BT). In the remaining cases, the system provided two variables, the first (NIBUT) and the mean (MNIBUT) non-invasive break-up times. The MNIBUT was calculated considering all broken segments and the stability was classified in a grade scale as stable (level 0: break-up average $\geq 14s$), critical (level 1: break-up average ≥ 7 bis $<14s$) or unstable (level 2: break-up average $<7s$). For uniform classification of ordinal variables with the OQAS, level 0 was considered as grade 1 in this study, and consequently grades with this device were the level offered by the manufacturer plus one (grade = level + 1).

The double pass system used was the OQAS HD Analyzer (Visiometrics SL, Terrassa, Spain) which provides a measurement of the optical effect of tear film instability using the double-pass retinal images of a point source and calculating the Ocular Scattering Index (OSI).[12] During the measurement, the subject's refractive error was corrected internally by an optometer (ranging from −8.00 to +5.00 diopters [D] in the instrument) prior to the OSI tear film measurement. Besides that, astigmatism above 0.50 D was corrected using trial lenses. Measurements were taken in a low light room (5 lux) different from the VK. The system allowed the clinician to obtain three types of measures: (1) without blinking (2) after blinking each 3 seconds, "baseline mode" and (3) after blinking each 9 seconds, "stress mode". With these three measurement modes, the system recorded double-pass images every 0.5 seconds during a 20-second period. Thus, the system recorded 40 images showing the evolution of the eye optical quality during the 20 seconds of measurement. The procedure used in the current study was the "baseline mode", with the patient blinking each 3 seconds after a beep emitted by the device. The variables provided by the system were the mean OSI (OSIm) and the standard deviation of the OSI (OSIsd). The baseline and stress modes also provide an ordinal classification of the measure depending on the increment (Δ) of the OSI after each blinking: (1) Plateau, ΔOSI ≤ 0.5; (2) P/S, ΔOSI > 0.5 and ΔOSI < 1.0; (3) Seesaw, ΔOSI ≥ 1.0 and OSI recovered after blinking; (4) S/L, ΔOSI ≥ 1.0; (5) Ladder, ΔOSI ≥ 1.0 and OSI not recovered after blinking.[13]

The OSDI is a validated questionnaire to assess the symptoms related to dry eye disease and their effect on vision.[14] It is comprised by a scale of 12 items. Each of the 12 items is graded on a scale from 0 to 4, where 0 indicates none of the time; 1, some of the time; 2, half of the time; 3, most of the time; and 4, all of the time. The total OSDI score is then

6

calculated on the basis of the formula: OSDI = [(sum of scores for all questions answered) × 100] / [(total number of questions answered) × 4]. The OSDI punctuation can range from 0 to 100, with higher scores representing greater disability. A patient can be classified as normal (0-12), mild dry eye (13-22), moderate dry eye (23-32), and severe dry eye (33-100) depending on the obtained score.[15]

c. Statistical analysis

The normality of data distributions of the variables included in the study was tested with the Kolmogorov-Smirnov test. The paired t-test was used for comparisons with differences normally distributed between both trials, whereas the Wicoxon test was used for non-normally distributed variables. Since the differences between trials were normally distributed for the VK variables, the Bland-Altman plot was used for the analysis of repeatability.[16] For comparison purposes, we decided to use the same plot with the OQAS, even though the non-normal distribution of the differences, because a non-normal distribution of differences may not be as serious here as in other statistical contexts following the guidelines of Bland and Altman.[17] The agreement and reproducibility for the grading scales were computed using the Cohen's k with quadratic weights.[16] The association of the OSDI with the grading scales was tested with the Chi-square test and the correlations with the Spearman rho. The SPSS version 20 (SPSS Inc, Chicago, Illinois, USA) and the MedCalc (ver. 12.7; MedCalc Inc., Belgium) were used for the statistical analysis. The significance was accepted at the $p < 0.05$ level.

4. Results

The total sample retrospectively retrieved was comprised of 99 right eyes of 99 subjects (51 men and 48 women) with mean age of 30 ± 6 years old. Analysis of the data depended on the BT obtained with the VK as for BTs lower than NIBUT, the system resulted in a missing value and an ordinal classification of grade 1. Considering this, a total of 61 eyes were included in the repeatability analysis of the NIBUT and MNIBUT using the Bland-Altman method, whereas 99 eyes were considered for the analysis of the ordinal classification provided by the system.

Regarding OQAS measurements, 16 eyes were excluded from the analysis of the tear film with OSIm and OSIsd as the values obtained were extreme outliers. These cases corresponded to eyes with refractive errors outside the range measurable by the device, from -8 D to +5 D for sphere and astigmatism above 2.75 D. While the median in ordinal scale was grade 1 for the sample evaluated, the median was grade 3 for the excluded cases despite being compensated with trial lenses. Therefore, 83 eyes were included in the repeatability analysis with the OQAS. These eyes were also considered in the evaluation of the agreement between devices in terms of grading classification in the sample of subjects who completed the measurement procedure first with the VK and afterwards with the OQAS.

Correlation of objective measurements with symptoms and agreement between devices

The percentages of subjects classified according to the OSDI scale as normal, mild, moderate and severe dry eye were 75.8%, 11.1%, 4.0% and 9.1% in the VK sample (n=99), respectively, and 74.7%, 10.8% , 3.6% and 10.8% in the OQAS sample (n=83). Table 2 shows the percentage of subjects who were simultaneously classified in a particular OSDI level and the grade obtained with each device. No significant associations or correlations were found between any of the parameters obtained with both devices and the classification of the subject according to the OSDI guidelines (Table 1). As metrics provided by the VK and OQAS are completely different, only an agreement analysis in terms of grading scale was feasible. A fair agreement was obtained between devices in the first trial (k=0.20 [CI95%: -0.02 to 0.42]) and poor in the second trial (k=0.04 [CI95%: -0.17 to 0.26]).

Table 1

Videokeratoscope repeatability

A total of 38.4% of subjects (n=38) had lower BT than NIBUT, either in the first, the second or both trials. These subjects resulted in missing values for NIBUT and MNIBUT and this was the reason why these cases were not included in analysis of Table 2. From these 38.4%, only 11.1% (n=11) obtained a lower BT in both trials simultaneously. This means that 27.3% of eyes had lower BT in the first or in the second trial. The mean BT for these subjects was 21.14 ± 2.87 s for the first trial and 21.14 ± 2.44 s for the second trial. Subjects with lower BT than NIBUT were classified as grade 1. In contrast, subjects with no lower BT than NIBUT in any of the two trials were classified as grade 2 (first trial:

26.1%, mean NIBUTs 10.82 ± 5.68 s; second trial: 30.8%) and grade 3 (3.8 %, mean NIBUTs 11.74 ± 6.47 s).

Mean NIBUT and MNIBUT were significantly higher in the second trial than in the first (Table 2). Figure 1 shows the mean differences versus the average between trials in a Bland-Altman plot for NIBUT (Figure 1A) and MNIBUT (Figure 1B). Limits of agreement were slightly higher for NIBUT than for MNIBUT and the standard deviation of the difference was not uniform(Figure 1A). The quadratic Kappa agreement between trials for the grading scale was k=0.37 [CI95%: 0.17 to 0.59].

Table 2

Figure 1

Optical Quality Analysis System Repeatability

The mean spherical equivalent for the 83 eyes included in this analysis was -2.42 ± 2.35 D, ranging from -8.00 to +5.00 D, and mean astigmatism was -0.60 ± 0.57 D, ranging from 0 to -2.75 D. A total of 52 eyes were measured with their best subjective refraction, whereas in 31 eyes compensation was not required. No significant differences were found between compensated or non-compensated eyes in the two trials, neither for OSIm or OSIsd (p > 0.05). The median for OSIm and OSIsd was not significantly different in the second trial compared to the first (Table 2). Figure 1 shows the mean differences versus the average between trials in a Bland-Altman plot for OSIm (Figure 1C) and OSIsd (Figure 1D). The standard deviation of the differences was not uniform (Figure 1D). The quadratic Kappa agreement between trials for the grading scale was k=0.59 [CI95%: 0.44 to 0.74].

5. Discussion

The NIBUT can be considered as an additional useful technique for the evaluation of the tear film stability in clinical practice, although it has not still been adopted by many eye care professionals. For instance, in New Zealand, 92% of ophthalmologists referred to use the fluorescein tear break-up time (FBUT) as tear stability assessment technique versus 4% that used NIBUT and 4% that used other techniques.[18] Differences between FBUT and NIBUT are controversial. While some authors have reported that the NIBUT is lower than the FBUT,[19–21] other authors have reported higher NIBUT than FBUT.[22–24] Particularly, Tian et al reported higher NIBUT than FBUT in dry eye subjects, but the opposite in healthy subjects.[25] The evidence of this suggests that these high discrepancies may be due to the complexity of the tear film evaluation.[11]

The cut-off value for dry eye classification according to the NIBUT also varies between studies. A sensitivity of 84.1% and specificity of 75.6% has been reported for the cutoff point of 2.65 s for the NIBUT.[19] A sensitivity of 86.1 % and a specificity of 81.1 % has been also reported for a cut-off value of 6.2 s.[21] This is in agreement with Wang et al[26] who reported a mean of 6.32 ± 2.49 s for eyes grouped as dry eye according to an OSDI value higher than 13 points (mild dry eye). Considering both cut off values of 2.65 s and 6.2 s, we calculated the percentage of subjects that could be classified as dry eye considering the inherent limitation of device repeatability. We found 5 cases below 2.65 s in the second trial and above in the first, only one case below in the first and above in the second, and no cases below 2.65 s in both trials simultaneously. This means that in a 6% of

cases (6 from 99) the subject may have been classified as dry eye considering a cut off value of 2.65 s because measurements were repeated twice. This suggests that measures below this cut-off value should be repeated more than once for obtaining a consistent diagnosis, always waiting a period of time between measures. For a cut off value of 6.2 s, the diagnosis of dry eye increased considerably, with a percentage of 44%. Our results suggest that caution should be taken if this higher cut off value is assumed because there exists a great probability of classifying incorrectly patients as normal or dry eye due to the poor repeatability of the measurement procedure.

Despite the fact that Tian et al[25] reported a limit of repeatability in healthy subjects of 5.24 s for NIBUT and 5.60 s for MNIBUT, we found in our study considerable higher limits of agreement of 13.46 s for NIBUT and 10.34 s for MNIBUT. However, our results are in agreement with those reported by Cox et al[24] who found better reproducibility with MNIBUT than with NIBUT and FBUT, although it is important to note that they used the Keratograph 4 and they really measured the reproducibility in two different days instead of the intrasession repeatability. In contrast to earlier findings,[27] we found significantly higher NIBUT and MNIBUT in the second measure. Possibly, the use of a more recent version of the system and the potential increase of the tear film reflex during the second measure as a consequence of maintaining the eyes opened a long time without blinking may account for this discrepancy. The repeatability of the grading scale showed a fair agreement with a Quadratic weighted Kappa index of 0.37.[28]

The application of OQAS for tear film assessment is more recent than NIBUT, and therefore there is a limited number of previous studies performed with this device, being our study the first one that compares OSI and NIBUT in the same sample of subjects. Tan

et al.[13] reported that the ΔOSI after blinking differs between control, mild dry eyes and severe dry eyes. Despite our procedure differed from this previous study as they measured ΔOSI with a free blink and we measured this parameter with a periodic blinking each 3 seconds, the grading scale used in our study was also computed considering this increment, as in the previous study. Interestingly, we found a fair agreement between the grading scale of the VK and the OQAS for the first trial (K=0.2) and a poor agreement for the second (K=0.04). Considering that the repeatability of the grading scale of the OQAS showed a good agreement with a Quadratic weighted Kappa index of 0.59,[28] our results suggests that two consecutive measures are not recommendable to be performed in clinical practice, especially for the NIBUT, and that a minimal period of time should be waited before conducting a second measure in the same session to ensure the tear film stabilization.

Symptoms measured with the OSDI are not usually correlated with clinical signs.[7] In spite of the fact that changes in the OSDI and in some clinical signs can be produced after refractive surgery[10] or after treatments,[29] there are not usually correlations between objective measures and subjective questionnaires. Our experiments are consistent with these previous findings, either for VK or OQAS, even though some authors have reported significant non-linear correlations between NIBUT and OSDI.[19, 21, 26]

We are aware that our research has limitations, some of them related to inherent limitations from devices. For instance, a lower sample was included in the Bland-Alman plots for the VK than for the OQAS, but this was unavoidable as in some patients, despite of the long time maintaining the eyes opened without blinking, the device did not provide a NIBUT time value since they blinked before the device was able to obtain a measurement. This limitation was not presented in the grading scale because all these eyes were classified as

grade 1. On the other hand, as OQAS is more sensitive to the correction of the high refractive errors, some eyes were excluded as they might be classified as poor tear film due to an increase of the OSI attributable to optical aberrations induced but trial lenses, poor fixation, etc. Finally, our results are only applicable to healthy subjects and are not applicable to dry eye subjects. Future studies should be performed to analyze the repeatability of both systems in subjects with consistent diagnosis of dry eye.

In conclusion, there is no agreement between OQAS and Keratograph devices in terms of tear film stability assessment. Good repeatability was obtained for the tear film grading provided by the OQAS whereas the repeatability for the grading system of the VK was fair. The use of NIBUT to evaluate tear film stability is recommendable in high refractive errors and high astigmatisms, where the optical aberrations induced by trial lenses might act as confounding variables leading to false poor tear stability with the OQAS. In the rest of cases, the OQAS seems to be preferable to evaluate tear film stability due to its better intrasession repeatability. Finally, it is important to note that this is a precision study and does not describe the accuracy of both devices in terms of dry eye diagnosis. Future studies should be performed in order to determine the cut off values for the screening of dry eye in refractive surgery candidates, considering the repeatability limitations obtained in our study.

6. Funding

7. Conflict of Interest

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements). Dr. Joaquín Fernández has participated as invited speaker in events organized by Oculus Optikgeräte GmbH, Wetzlar, Germany. Remaining authors declare non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

8. Ethical approval.

This study was approved by the local ethics committee of research and was performed in adherence to the tenets of the Declaration of Helsinki. For this type of study formal consent is not required.

9. References

1.   Torricelli AM, Bechara SJ, Wilson SE (2014) Screening of refractive surgery candidates for LASIK and PRK. Cornea 33:1051–5

2.   Levitt AE, Galor A, Weiss JS, et al (2015) Chronic dry eye symptoms after LASIK: parallels and lessons to be learned from other persistent post-operative pain disorders. Mol Pain 11:1–12

3.   Jabbur NS, Sakatani K, O'Brien TP (2004) Survey of complications and recommendations for management in dissatisfied patients seeking a consultation after refractive surgery. J Cataract Refract Surg 30:1867–1874

4.   Li M, Zhao J, Shen Y, et al (2013) Comparison of dry eye and corneal sensitivity between small incision lenticule extraction and femtosecond LASIK for myopia. PLoS One 8:e77797

5.   Hu L, Xie W, Liu J, et al (2015) Tear Menisci and Corneal Subbasal Nerve Density in Patients After Laser In Situ Keratomileusis. Eye Contact Lens Sci Clin Pract 41:51–57

6.   Tao A, Shen M, Wang J, et al (2010) Upper and lower tear menisci after laser in situ keratomileusis. Eye Contact Lens 36:81–85

7.   Nichols KK, Nichols JJ, Mitchell GL (2004) The lack of association between signs and symptoms in patients with dry eye disease. Cornea 23:762–770

8.   Tuisku IS, Lindbohm N, Wilson SE, Tervo TM (2007) Dry eye and corneal sensitivity after high myopic LASIK. J Refract Surg 23:338–42

9.   Farahi A, Hashemi H, Mehravaran S, et al (2014) Tear function evaluation in candidates of corneal laser refractive surgery for myopia. Eye Contact Lens 40:91–4

10.   Shen Z, Shi K, Yu Y, et al (2016) Small incision lenticule extraction (SMILE) versus femtosecond laser-assisted In situ keratomileusis (FS-LASIK) for myopia: a systematic review and meta-analysis. PLoS One 11:e0158176

11.   Zeev MS, Miller DD, Latkany R (2014) Diagnosis of dry eye disease and emerging technologies. Clin Ophthalmol 8:581–90

12.   Artal P, Benito A, Pérez GM, et al (2011) An objective scatter index based on double-pass retinal images of a point source to classify cataracts. PLoS One 6:1–7

13.   Tan CH, Labbe A, Liang Q, et al (2015) Dynamic change of optical quality in patients with dry eye disease. Invest Ophthalmol Vis Sci 56:2848–2854

14.   Schiffman RM (2000) Reliability and Validity of the Ocular Surface Disease Index. Arch Ophthalmol 118:615

15.   Miller KL, Walt JG, Mink DR, et al (2010) Minimal clinically important difference for the ocular surface disease index. Arch Ophthalmol (Chicago, Ill 1960) 128:94–101

16.   Kottner J, Audigé L, Brorson S, et al (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 64:96–106

17. Bland JM, Altman DG (1999) Statistical methods in medical research. Stat Methods Med Res 8:135–160

18. Xue AL, Downie LE, Ormonde SE, Craig JP (2017) A comparison of the self-reported dry eye practices of New Zealand optometrists and ophthalmologists. Ophthalmic Physiol Opt 37:191–201

19. Hong J, Sun X, Wei A, et al (2013) Assessment of tear film stability in dry eye with a newly developed keratograph. Cornea 32:716–721

20. Hong J, Liu Z, Hua J, et al (2014) Evaluation of age-related changes in noninvasive tear breakup time. Optom Vis Sci 91:1

21. Bhandari V, Reddy JK, Relekar K, et al (2016) Non-invasive assessment of tear film stability with a novel corneal topographer in Indian subjects. Int Ophthalmol 36:781–790

22. Lan W, Lin L, Yang X, Yu M (2014) Automatic noninvasive tear breakup time (TBUT) and conventional fluorescent TBUT. Optom Vis Sci 91:1412–1418

23. Abdelfattah NS, Dastiridou A, Sadda SR, Lee OL (2015) Noninvasive imaging of tear film dynamics in eyes With ocular surface disease. Cornea 34 Suppl 1:S48-52

24. Cox SM, Nichols KK, Nichols JJ (2015) Agreement between automated and traditional measures of tear film breakup. Optom Vis Sci 92:e257–e263

25. Tian L, Qu JH, Zhang XY, Sun XG (2016) Repeatability and reproducibility of noninvasive keratograph 5m measurements in patients with dry eye disease. J Ophthalmol 2016:

26. Wang X, Lu X, Yang J, et al (2016) Evaluation of dry eye and meibomian gland dysfunction in teenagers with myopia through noninvasive keratograph. J Ophthalmol 2016:6761206

27. Best N, Drury L, Wolffsohn JS (2012) Clinical evaluation of the oculus keratograph. Contact Lens Anterior Eye 35:171–174

28. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical

data. Biometrics 33:159

29.   Caglar C, Senel E, Sabancilar E, Durmus M (2017) Reduced ocular surface disease index (OSDI) scores in patients with isotretinoin treatment. Int Ophthalmol 37:197–202

10. Tables

Table 1. Contingence table showing the percentage of subjects that match a particular level of OSDI and a particular grade of classification with videokeratoscopy and Optical Quality Analyzer System.

	Ocular Surface Disease Index Percentage 1st trial / 2nd trial				Association and correlation	
	Normal	Mild	Moderate	Severe		
VK Classification						
Grade 1	41.4 / 42.4	5.1 / 4.0	2.0 / 2.0	4.0 / 5.1	1st trial	χ^2 = 3.13, p = 0.81
Grade 2	27.3 / 26.3	4.0 / 7.1	2.0 / 2.0	3.0 / 3.0		ρ= 0.09, p = 0.36
Grade 3	7. 1 / 7.1	2.0 / 0.0	0.0 / 0.0	2.0 / 1.0	2nd trial	χ^2 = 4.03, p = 0.64
						ρ= 0.105, p = 0.63
OQAS Classification						
Grade 1	43.4 / 43.4	6.0 / 7.2	1.2 / 2.4	6.0 / 3.6	1st trial	χ^2 = 7.25, p = 0.23
Grade 2	15.7 / 24.1	4.8 / 1.2	2.4 / 1.2	1.2 / 4.8		ρ= 0.24, p = 0.83
Grade 3	15.7 / 7.2	0.0 / 2.4	0.0 / 0.0	3.6 / 2.4	2nd trial	χ^2 = 5.56, p = 0.41
						ρ= 0.10, p = 0.89

χ^2: Associations between OSDI and objective tests evaluated with the chi-square test.
ρ: Correlations between OSDI and objective tests evaluated with the spearman rho.

Table 2. Results for both consecutive measures taken with the Videokeratoscope and the Optical Quality Analyzer System.

Variables	Consecutive measures mean ± SD; median (IQR)		test value, p-value
Videokeratoscope	trial 1	trial 2	
NIBUT	8.33 ± 5.33	10.15 ± 6.10	t= -2.08, p = 0.04
	5.93 (6.09)	9.94 (8.27)	
MNIBUT	11.62 ± 5.00	13.35 ± 4.97	t= -2.56, p = 0.01
	10.32 (8.57)	12.98 (6.47)	
Classification grade	1.59 ± 0.69	1.55 ± 0.64	χ^2= 0.54 , p = 0.80
	1 (1)	1 (1)	
Optical Quality Analyzer System	trial 1	trial 2	
OSIm	0.99 ± 0.51	0.98 ± 0.51	z= -1.60, p = 0.11
	0.85 (0.59)	0.81 (0.59)	
OSIsd	0.14 ± 0.11	0.13 ± 0.09	z= -0.68, p = 0.50
	0.11 (0.14)	0.11 (0.08)	
Classification grade	1.63 ± 0.79	1.55 ± 0.70	χ^2= 2.17 , p = 0.34
	1 (1)	1 (1)	

11. NIBUT: first non-invasive break-up time; MNIBUT: mean non-invasive break-up time;
12. OSIm: mean ocular scatter index; (D) OSIsd: standard deviation of ocular scatter index.
13. t: paired t-test for normal distribution of the differences.
14. z: Wilcoxon signed rank test for non-normal distribution of the differences.
15. χ^2: Associations between trials and the classification grade with the chi-square test.

16.   Figure legends.

Figure 1. Bland-Altman plots for repeatability. Difference between trials versus the mean for (A) first non-invasive break-up time (NIBUT); (B) mean non-invasive break-up time (MNIBUT); (C) mean ocular scattering index (OSIm); (D) standard deviation of ocular scattering index (OSIsd).