

Inter-Examiner Reliability in Meibomian Gland Dysfunction Assessment

Daniel R. Powell,¹ Jason J. Nichols,² and Kelly K. Nichols²

PURPOSE. We evaluated inter-examiner reliability in grading of clinical variables associated with meibomian gland dysfunction (MGD) in real-time examination versus a graded digital image.

METHODS. Meibography grading of meibomian gland atrophy and acini appearance, and slit-lamp grading of lid debris and telangiectasias were conducted on 410 post-menopausal women. Meibography and slit-lamp photos were captured digitally and saved for analysis by a masked examiner. Gland atrophy was graded as a proportion of partial glands in the lower lid, and acini appearance by the presence/absence of grape-like clusters. Lid debris and telangiectasias were graded based on severity and quantity from the same image, respectively. Observed agreement and weighted kappas (κ_w) with 95% confidence intervals (CI) determined the degree of inter-examiner reliability between grading of these clinical variables in real-time examination and digital photographs using a multiple-point categorical scale.

RESULTS. Observed agreement was determined for telangiectasias (40.6%), lid debris (50.9%), gland dropout (42.8%), and acini appearance (54.5%). Inter-examiner reliability for the four clinical outcomes ranged from fair agreement for acini appearance ($\kappa_w = 0.23$, 95% CI = 0.14–0.32) and lid debris ($\kappa_w = 0.24$, 0.16–0.32) to moderate agreement for gland dropout ($\kappa_w = 0.50$, 0.40–0.59) and telangiectasias ($\kappa_w = 0.47$, 0.39–0.55).

CONCLUSIONS. Gland dropout and potentially lid telangiectasia grading from a photograph are more representative of grading in a real-time examination compared to acini appearance and lid debris. Alternative grading scales and/or clinical variables associated with MGD should be addressed in future studies. (*Invest Ophthalmol Vis Sci.* 2012;53:3120–3125) DOI: 10.1167/iovs.12-9600

Examination methods are used frequently to detect and grade numerous diseases, including meibomian gland dysfunction (MGD). The understanding of these clinical tests

for MGD diagnosis, however, remains in their infancy. Slit-lamp examination of the ocular surface and meibomian glands is the most common technique in clinical practice used to diagnose MGD. However, the establishment of universally-accepted grading criteria (and new, more innovative diagnostic tests) that are reproducible and valid is necessary.

MGD is a significant public health problem. The International Workshop on Meibomian Gland Dysfunction (MGD Report) has described MGD as a chronic abnormality of the terminal ducts and/or glandular secretion involving the majority of the meibomian glands, which may result in tear film alterations, eye irritation symptoms, clinically apparent inflammation, and ultimately ocular surface disease.¹ As reported previously in 2007, dry eye prevalence in the general population has ranged from 5–30%,² and it is believed that MGD may comprise a substantial portion of these numbers. Many in the scientific community have identified MGD as the most likely causative factor in evaporative dry eye disease.^{3,4} However, in clinical practice, MGD more than likely is under-reported because of the lack of clearly defined and universally-accepted clinician and patient-reported diagnostic approaches.⁴

MGD prevalence has ranged from 3.5% to nearly 70% from data collected and analyzed in population- and clinical-based studies.⁴ In addition to studies that have reported MGD prevalence based on patient-reported symptoms (eye dryness, foreign body sensation, redness, and tearing^{5,6}), clinical parameters also have been used to estimate prevalence, such as lid telangiectasias,^{5,7} lid collarettes,^{7,8} and meibomian gland dropout.⁹ A significant drawback in determining MGD prevalence, however, is the absence of a universal set of grading criteria to assess these clinical outcomes. This is compounded further by the potentially multifactorial nature of disease with numerous risk factors, including older age, contact lens wear,^{10,11} menopause,^{12,13} androgen deficiency,^{13,14} rosacea,^{15–17} medication use,^{14,18} and environmental factors.¹⁹

To confirm whether clinical tests are a useful aid in the diagnosis of a disease, including MGD, it is important first to evaluate the psychometric properties of grading scales associated with the assessment used in the diagnosis. This may include assessments of validity and reliability, such as within-examiner (test-retest or intra-examiner) reliability, or between-examiner (inter-examiner) reliability. Grading scales may be graded categorically or tabulation-based (counting), for example. In a study by Nichols et al., within-reader and between-reader reliability was examined in assessing meibomian gland dropout using these two types of grading scales.²⁰ Between-reader reliability for each grading method was determined to have fair reliability, although both scales exhibited good concurrent validity.²⁰ To our knowledge, no other reports in the literature have addressed inter-examiner reliability (real-time versus digital image examination) between grading of ocular findings associated with MGD. Thus, our objective was to evaluate inter-examiner reliability in the grading of lid telangiectasias and lid debris by slit-lamp

From the ¹The Ohio State University, College of Optometry Columbus, Ohio; and ²The University of Houston, College of Optometry, Houston, Texas.

Supported by the 2011-12 American Optometric Foundation William C. Ezell Fellowship (DRP), and National Eye Institute Grants EY015539 (DRP) and EY015519 (KKN, JKN).

Submitted for publication January 31, 2012; accepted April 4, 2012.

Disclosure: **D.R. Powell**, None; **J.J. Nichols**, Alcon (F), CIBA Vision (F), Inspire Pharmaceuticals (F), Vistakon (F); **K.K. Nichols**, Alcon (F, C), Allergan (C), Bausch & Lomb (C), Celtic (C), Eleven Biotherapeutics (C), Inspire Pharmaceuticals (F, C), InSite Vision (C), ISTA (C), Pfizer (C), SarCode (C), TearLab (F, C)

Corresponding author: Kelly K. Nichols, The University of Houston, 505 J. Davis Armistead Building, Houston, TX 77204; Telephone 713-743-2418; knichols@optometry.uh.edu.

TABLE 1. Landis and Koch Interpretation of the Unweighted and Weighted κ Statistic Based on Its Calculated Value²²

κ Value (−1.00 to +1.00 Scale)	Interpretation
≤ 0	Less than chance agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Near perfect agreement
1.00	Perfect agreement

examination, and meibomian gland dropout and acini appearance by meibography.

METHODS

Our study was conducted under the tenets of the Declaration of Helsinki and approved by the Ohio State University Biomedical Science

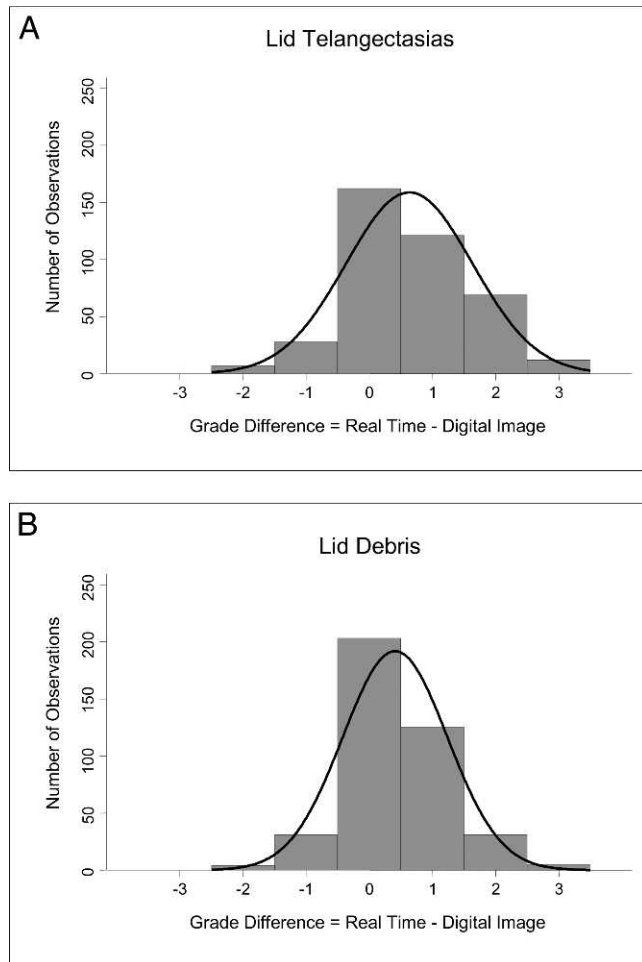


FIGURE 1. Distribution of grade differences for slit-lamp examination findings between real-time examination and from a digital image examination. (A) Lid telangiectasias: the mean grade difference (\pm SD) was 0.63 ± 1.00 of a grade with a data distribution that was left-skewed. The real-time examiner had a tendency to assign a higher grade compared to digital image examiner. (B) Lid debris: the mean difference was 0.41 ± 0.83 of a grade. Overall, a higher grade was assigned by the real-time examiner compared to those assigned by the image examiner, and this is exemplified by the left-skewedness of the data distribution.

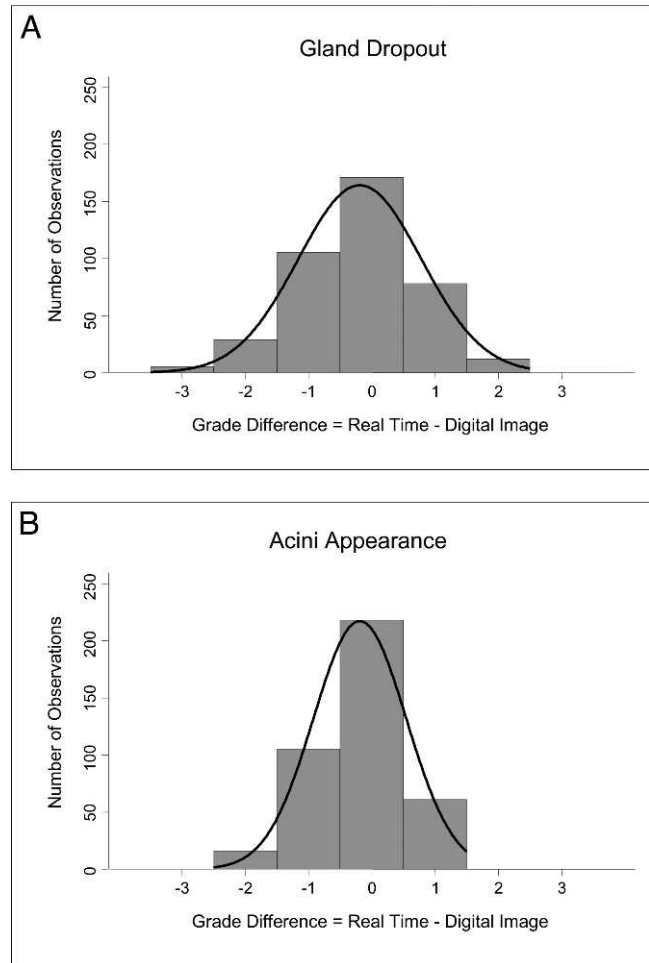


FIGURE 2. Distribution of grade differences for meibography findings between real-time and image examination. (A) Gland dropout: the mean difference was -0.19 ± 0.97 of a grade. The image examiner graded slightly higher than the real-time examiner overall, but the data follow a normal distribution. (B) Acini appearance: the mean difference was -0.20 ± 0.73 of a grade. The image examiner also assigned a slightly higher grade overall compared to grades assigned by real-time examination, although the data distribution is not normal (right-skewed).

Institutional Review Board. Before participation, all subjects signed informed consent and HIPPA privacy documents following explanation of the study procedures. There were 410 menopausal women (defined as at least one year since last menses) enrolled in the study. No presumptions were made regarding the presence or absence of MGD as an inclusion criterion.

Clinical Evaluation

Slit-lamp examination was conducted at 10 \times magnification with moderate illumination using a Haag-Streit BX 900 biomicroscope (Haag-Streit, K oniz, Switzerland). In addition to a complete health evaluation of the ocular surface and surrounding structures, lid telangiectasia quantity and lid debris severity of the lower lid of each eye also were evaluated and graded. Grading of lid telangiectasias was based on the number present on a 5-point scale (0 = none present, 1 = one present, 2 = 2-5 are present, 3 = >5 are present, and 4 = 100% lower lid involvement),²¹ while lid debris also was graded on a 5-point scale (0 = none, 1 = mild debris, 2 = moderate debris, 3 = severe debris, and 4 = very severe).²¹

TABLE 2. Between Examiner-Reader Observations for Overall Lid Telangiectasias Grading of Severity (Cell Data Represents Number of Observations)

	Masked Reader					Row Totals (%)
	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	
Examiner						
Grade 0	95	14	5	0	0	116 (29.1)
Grade 1	37	26	9	2	0	74 (18.5)
Grade 2	40	42	36	5	0	123 (30.8)
Grade 3	11	21	36	3	0	71 (17.8)
Grade 4	0	1	8	6	0	15 (3.8)
Column totals (%)	185 (46.4)	104 (26.1)	94 (23.5)	16 (4.0)	0 (0.0%)	399

Grade 0, none present; Grade 1, one present; Grade 2, 2-5 present; Grade 3, >5 present; Grade 4, 100% lid involvement.²¹ Overall agreement = 160/399 = 40.1%.

A Topcon BG-4M non-contact meibography unit containing a Topcon SL-D7 biomicroscope (Topcon, Oakland, NJ) equipped with a Sony XC-E150 camera (Sony Electronics, Park Ridge, NJ) and an internal filter (650-700 nm wavelength), or a Haag Streit BX 900 slit-lamp with a built-in AVT Stingray camera (Allied Vision Technologies, Newburyport, MA), were used to visualize the slightly everted lower lid of each eye to grade the central 2/3 (~ 15 glands) of the lower meibomian glands. (During protocol development, the images from both instruments were judged clinically equivalent.) Normal meibomian glands appear as hypoilluminant grape-like clusters with hyperilluminant ducts and orifices that traverse the lid 3-4 mm from the orifice. Glands that do not traverse the lid completely or are missing entirely indicate gland dropout. If gland dropout was present, severity was determined by the proportion of these glands that were partially or completely missing. The degree of meibomian gland dropout was determined on a 4-point scale based on the visualization of the individual acini within the glands (1 = no partial or missing glands, 2 = <25% are partial glands, 3 = 25-75% are partial glands, and 4 = >75% are partial glands).²⁰ The absence of the grape-like appearance may indicate dysfunctional glands and was graded on a 3-point scale (1 = grape-like clusters, or normal; 2 = stripes, or individual architecture is difficult to discern visually; and 3 = acini are not visible).²¹

Image Capture and Analysis

All images were saved with a coded identification number and graded later by a masked examiner who was trained to read the archived images. Two static images of the central lower lids for each subject were captured using a photo slit-lamp (color images) and during the meibography procedure. The image examiner analyzed the first of two images captured from the right eye, and only in the event that the first photo was of substandard image quality (poor focus, dim illumination, or poor centration) was the second image then opened, evaluated, and graded. The image examiner did not grade any images from the left eye.

Therefore, all data used in the statistical analyses were from the right eye only.

Statistical Analyses

Unweighted and weighted kappa (κ) statistics with 95% confidence intervals (CI) were used to examine the reliability of grading between real-time examination and the digital image. Unlike the unweighted κ , which penalizes any difference in grading, the weighted κ takes into account the degree of disagreement between individual grades between the examiners. Unweighted and weighted kappas are expressed as a number on a continuous scale from -1.00 to +1.00. The interpretation of the calculated value is arbitrary; however, most reports of reliability testing in the scientific literature generally rely on the guidelines suggested by Landis and Koch²² (Table 1). Overall agreement (proportion of all paired grades where the examiner and the masked reader graded the clinical variable of interest the same) was calculated for all four outcomes as well. Histograms were constructed to illustrate the distribution of differences in grading between real-time and image examination for each clinical variable, along with an assessment for asymmetry in the data distribution.

RESULTS

Grades from real-time examination, and images from slit-lamp examination and meibography were collected from 410 subjects. The overall average age was 62.3 ± 8.8 years with a median age of 61 years (46-91 years, 100% women). Of the 410 images collected from the right eye, 11 (2.5%) from the slit-lamp examination and 10 (2.4%) from meibography were deemed not gradable because of poor image quality (illumination, or out of focus), equipment malfunction (not saved, or power failure), or not enough lid eversion to appreciate the meibomian glands during the meibography procedure.

TABLE 3. Between Examiner-Reader Observations for Overall Lid Debris Grading of Severity (Cell Data Represents the Number of Observations)

	Masked Reader				Row Totals (%)
	Grade 0	Grade 1	Grade 2	Grade 3	
Examiner					
Grade 0	163	24	4	0	191 (47.9)
Grade 1	110	31	6	0	147 (36.8)
Grade 2	30	13	7	1	51 (12.8)
Grade 3	5	1	2	2	10 (2.5)
Column totals (%)	308 (77.2)	69 (17.3)	19 (4.8)	3 (0.7)	399

Grade 0, none present; Grade 1, mild debris; Grade 2, moderate debris; Grade 3, severe debris; Grade 4, very severe debris (not included as no observations were recorded with this grade).²¹ Overall agreement = 203/399 = 50.9%.

TABLE 4. Between Examiner-Reader Observations for Overall Meibomian Gland Dropout Grades of Severity (Cell Data Represents the Number of Observations)

	Masked Reader				Row Totals (%)
	Grade 1	Grade 2	Grade 3	Grade 4	
Examiner					
Grade 1	15	47	12	5	79 (19.8)
Grade 2	7	74	25	17	123 (30.7)
Grade 3	1	49	32	33	115 (28.8)
Grade 4	0	11	22	50	83 (20.7)
Column totals (%)	23 (5.8)	181 (45.2)	91 (22.8)	105 (26.2)	400

Grade 1, no partial or missing glands; Grade 2, <25% are partial glands; Grade 3, 25–75% are partial glands; Grade 4, >75% are partial glands.²⁰ Overall agreement = 171/400 = 42.8%.

The inter-examiner reliability data associated with the grading scales for their respective clinical outcomes are illustrated in Tables 2–5. The overall agreement between real-time and digital image grading was 40.1% for lid telangiectasias (Table 2), 50.9% for lid debris (Table 3), 42.8% for gland dropout (Table 4), and 54.5% for acini appearance (Table 5). The unweighted and weighted κ values (with 95% CI) for lid telangiectasias, lid debris, gland dropout, and acini appearance are presented in Table 6. For grading of gland dropout and lid telangiectasias, there was fair-to-moderate reliability between the real-time and image examination. Reliability for lid debris and acini appearance fared worse, as only slight-to-fair reliability was appreciated between the real-time and image examiner.

The mean difference \pm SD in overall grades was 0.63 \pm 1.00 (median = 1) and 0.41 \pm 0.83 (median = 0) of a grade for lid telangiectasias (Fig. 1A) and debris (Fig. 1B), respectively, and -0.19 ± 0.97 (median = 0) and -0.20 ± 0.73 (median = 0) of a grade for gland dropout (Fig. 2A) and acini appearance (Fig. 2B), respectively. (A positive mean difference value indicates that the real-time examiner overall graded a clinical outcome at a higher grade than the image examiner. A negative mean difference indicates that the image examiner overall graded at a higher grade than the real-time examiner.) Data for gland dropout followed a normal distribution ($P = 0.22$); however, data distribution for lid telangiectasias and lid debris were left-skewed (all $P < 0.0001$) and right-skewed for acini appearance ($P = 0.02$).

DISCUSSION

Diagnostic tests, photographic interpretations, and physical examination findings rely frequently on a degree of subjective interpretation by clinicians. To determine the usefulness of a

TABLE 5. Between Examiner-Reader Observations for Overall Meibomian Gland Acini Appearance Grading of Severity (Cell Data Represents the Number of Observations)

	Masked Grader			Row Totals (%)
	Grade 1	Grade 2	Grade 3	
Examiner				
Grade 1	22	51	16	89 (22.3)
Grade 2	46	171	55	272 (68.0)
Grade 3	0	14	25	39 (9.7)
Column totals (%)	68 (17.0)	236 (59.0)	96 (24.0)	400

Grade 1, grape-like clusters; Grade 2, stripes; Grade 3, not visible.²¹ Overall agreement = 218/400 = 54.5%.

clinical outcome measure, reliability and validity must be established. Overall agreement is a method used to evaluate reliability. Although this provides a measurement of agreement, it does not take into account chance agreement or disagreement made by examiners. If examiners are agreeing or disagreeing solely by chance, then no real true measurement of reliability is taking place. This issue can be addressed by the calculation of the κ coefficient.

The κ statistic is the most commonly used statistic for measuring agreement between ratings of two or more examiners (inter-examiner) or by the same examiner on two or more occasions (intra-examiner). The unweighted κ indicates the proportion of agreement that is above what is expected by chance, although it does not differentiate between differences that are due to chance or an examiner’s inconsistent grading pattern (systematic error or bias). Unweighted κ values provide a true measure of agreement between or within examiners. Weighted κ coefficients offer a benefit by penalizing disagreements based on the degree by which they are different compared to unweighted kappas, where all disagreements are penalized equally (no value assigned).

The large differences between unweighted and weighted κ values are likely the result of the number of paired observations that disagreed by only one grade. As seen in Tables 2 and 3, more than 37% (149/399) of lid telangiectasias and 39% (156/399) of lid debris observations from the slit-lamp examination were within one grade of each other. The real-time examiner graded one grade higher than the image examiner, approximately 80% (121/149) for lid telangiectasias and 125/156 for lid debris) of the time (See Figs. 1A, 1B). This most likely explains the significant left-skewedness of the data distribution as seen in both of these Figures. Likewise, in Tables 4 and 5, nearly 46% (183/400) and 42% (166/400) of meibomian gland dropout and visible acini observations, respectively, fell within one grade of true agreement. The image examiner graded one grade higher than the real-time examiner just over half the time (57%, or 105/183) for gland dropout (Fig. 2A), and nearly two-thirds of the time (64%, or 106/166) for acini appearance (Fig. 2B). These findings likely

TABLE 6. Unweighted and Weighted κ Values for Evaluating Agreement of the Four Clinical Variables between Paired Grades of the Real-Time and Image Examiner

Clinical Variable	Unweighted (95% CI)	Weighted (95% CI)
Lid telangiectasias	0.19 (0.14–0.25)	0.47 (0.39–0.55)
Lid debris	0.12 (0.06–0.19)	0.24 (0.16–0.32)
Gland dropout	0.22 (0.16–0.27)	0.50 (0.40–0.59)
Acini appearance	0.15 (0.08–0.22)	0.23 (0.14–0.32)

explain the normal data distribution for the mean difference in grades for gland dropout and a right-skewed distribution (image examiner graded generally higher than the real-time examiner) for acini appearance.

The high percentage of grades for each clinical variable differing by one grade accounted for the greater weighted κ compared to the unweighted κ . Application of these kappa values might depend on the scenario in which they are being used; weighted κ might be relevant to the use and interpretation of these values in the clinical care of patients, where absolute agreement might not be necessary, for example. Likewise, it might be more important to consider the unweighted κ in the evaluation of an outcome in a clinical trial or epidemiological study, where absolute agreement might be more important as it impacts the validity of the study findings.

Depending on the clinical variable evaluated, in some instances the real-time examiner assigned a higher grade than the image examiner (lid telangiectasias and debris), while in other situations, the digital image examiner assigned a grade higher than the real-time examiner (gland dropout and acini visibility). The findings from the slit-lamp examination (lid debris and, to a lesser extent, lid telangiectasias) seemed to have indicated that either the real-time examiner over-graded the actual grade compared to the digital image examiner, or the digital image examiner may have underestimated overall the actual grade compared to the real-time examination. On the contrary, for the outcomes evaluated by meibography (acini appearance and, to a lesser extent, gland dropout), results have revealed assignment of an overall lower grade by the real-time examiner compared to the digital image examiner or vice versa. Besides the non-normal distribution of the mean difference data (lid telangiectasias, lid debris, and acini appearance), the real-time examinations could have benefited from the dynamic examination process to aid accurate grading, whereas single-photo grading eliminates gestalt impressions. Should photographic techniques be used in a clinical trial, multiple images may be needed from different angles or illuminations to simulate the real-time examination.

A limitation of this study was that the digital image examiner was able to grade only a clinical variable based on a single, static electronic image. Any deviations from the standard protocol (image selection, contrast, illumination) as well as image resolution may correspond to increased variability and, thus, a lower κ coefficient. In addition, the digital image examiner was unable to evaluate each subject in a real-time environment as was the examiner. The digital image examiner expressed the difficulty in evaluating for lid debris based on its three-dimensional structure imaged in a two-dimensional photograph. This may explain any under-grading of this variable by the digital image examiner. In contrast to still digital images, video recordings may be more representative as an alternative to grading images in a real-time setting (although challenges would remain in accurately capturing data from video imaging). For our study, however, video recording was not feasible nor possible given the electronic storage capabilities required to archive these recordings for all the subjects who were examined. This would have been more realistic given a smaller sample size, although the inability to detect a significant κ value or provide a CI of a desired precision with this strategy would become more likely. One advantage that the digital image examiner has over the real-time examiner is that more time can be vested in evaluating an image, while the real-time examiner has much shorter time to evaluate and grade a clinical variable because of time constraints and potential subject fatigue. Also, there were no opportunities to re-evaluate each of the outcomes as the subject was seen at a single visit.

Multiple examiners were used to evaluate the clinical outcomes in real-time. All were trained on the standard

operating procedures for biomicroscopy and meibography (subject alignment, illumination and width of slit, and magnification), including the grading of clinical parameters before examination of any enrolled subject. Training included instruction on the grading scales for each variable, reviewing samples of each clinical variable and group interaction on assignment of the appropriate grade, protocol for proper image capture (lighting, field of view, focus, and magnification) and electronic storage of static images for future retrieval by the digital image examiner. One explanation for the fair unweighted κ values between the real-time and digital image examiner may lie in variability of grading from one examiner to the next (systematic error) despite the training that each examiner received before examining subjects in this study. Within-examiner analyses of agreement were not conducted because each subject was examined by a single examiner.

In summary, we demonstrated that inter-examiner reliability is at least slight-to-moderate for grading of lid margin telangiectasias, with fair-to-moderate agreement for meibomian gland dropout compared to only slight-to-fair agreement for lid debris and acini appearance. It seems, however, that the degree of agreement may be affected by several factors, including variability within examiners, quality of digital images graded by the examiner (substandard resolution, different field of view for grading clinical variable compared to that of the examiner), and grading bias by the real-time examiners because of their knowledge of each subject's medical and ocular history (presence of meibomian gland dysfunction). The digital image examiner was not privy to any subject information besides what was provided in the images. It appears that assessment of gland dropout may be the best clinical outcome in the assessment of MGD. Future studies investigating new grading schemes, modification of existing grading schemes, and the assessment of additional clinical variables used in MGD diagnosis, may be beneficial in elevating inter-examiner reliability.

Acknowledgments

Andrew Emch, Anupam Laul, Kathleen Reuter, Michele Hager, and Aaron Zimmerman, of The Ohio State University College of Optometry, served as the real-time examiners. Gregory Hopkins, of The Ohio State University College of Optometry, was the image examiner. Topcon provided loan of the BG-4M meibography slit-lamp system.

References

1. Nichols KK. The international workshop on meibomian gland dysfunction: introduction. *Invest Ophthalmol Vis Sci.* 2011; 52:1917–1921.
2. The epidemiology of dry eye disease: report of the Epidemiology Subcommittee of the International Dry Eye WorkShop (2007). *Ocul Surf.* 2007;5:93–107.
3. The definition and classification of dry eye disease: report of the Definition and Classification Subcommittee of the International Dry Eye WorkShop (2007). *Ocul Surf.* 2007;5:75–92.
4. Schaumberg DA, Nichols JJ, Papas EB, Tong L, Uchino M, Nichols KK. The international workshop on meibomian gland dysfunction: report of the subcommittee on the epidemiology of, and associated risk factors for, MGD. *Invest Ophthalmol Vis Sci.* 2011;52:1994–2005.
5. Jie Y, Xu L, Wu YY, Jonas JB. Prevalence of dry eye among adult Chinese in the Beijing Eye Study. *Eye (Lond).* 2009;23:688–693.
6. Lin PY, Tsai SY, Cheng CY, Liu JH, Chou P, Hsu WM. Prevalence of dry eye among an elderly Chinese population in Taiwan: the Shihpai Eye Study. *Ophthalmology.* 2003;110:1096–1101.

7. Lekhanont K, Rojanaporn D, Chuck RS, Vongthongsri A. Prevalence of dry eye in Bangkok, Thailand. *Cornea*. 2006;25:1162-1167.
8. Schein OD, Muñoz B, Tielsch JM, Bandeen-Roche K, West S. Prevalence of dry eye among the elderly. *Am J Ophthalmol*. 1997;124:723-728.
9. Uchino M, Dogru M, Yagi Y, et al. The features of dry eye disease in a Japanese elderly population. *Optom Vis Sci*. 2006;83:797-802.
10. Arita R, Itoh K, Inoue K, Kuchiba A, Yamaguchi T, Amano S. Contact lens wear is associated with decrease of meibomian glands. *Ophthalmology*. 2009;116:379-384.
11. Ong BL, Larke JR. Meibomian gland dysfunction: some clinical, biochemical and physical observations. *Ophthalmic Physiol Opt*. 1990;10:144-148.
12. Mathers WD, Stovall D, Lane JA, Zimmerman MB, Johnson S. Menopause and tear function: the influence of prolactin and sex hormones on human tear production. *Cornea*. 1998;17:353-358.
13. Sullivan DA, Sullivan BD, Evans JE, et al. Androgen deficiency, meibomian gland dysfunction, and evaporative dry eye. *Ann N Y Acad Sci*. 2002;966:211-222.
14. Sullivan BD, Evans JE, Krenzer KL, Reza Dana M, Sullivan DA. Impact of antiandrogen treatment on the fatty acid profile of neutral lipids in human meibomian gland secretions. *J Clin Endocrinol Metab*. 2000;85:4866-4873.
15. Akpek EK, Merchant A, Pinar V, Foster CS. Ocular rosacea: patient characteristics and follow-up. *Ophthalmology*. 1997;104:1863-1867.
16. Alvarenga LS, Mannis MJ. Ocular rosacea. *Ocul Surf*. 2005;3:41-58.
17. Zengin N, Tol H, Gündüz K, Okudan S, Balevi S, Endoğru H. Meibomian gland dysfunction and tear film abnormalities in rosacea. *Cornea*. 1995;14:144-146.
18. Pinna A, Piccinini P, Carta F. Effect of oral linoleic and gamma-linolenic acid on meibomian gland dysfunction. *Cornea*. 2007;26:260-264.
19. Fenga C, Aragona P, Cacciola A, et al. Meibomian gland dysfunction and ocular discomfort in video display terminal workers. *Eye (Lond)*. 2008;22:91-95.
20. Nichols JJ, Berntsen DA, Mitchell GL, Nichols KK. An assessment of grading scales for meibography images. *Cornea*. 2005;24:382-388.
21. Bron AJ, Benjamin L, Snibson GR. Meibomian gland disease. Classification and grading of lid changes. *Eye (Lond)*. 1991;5:395-411.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.