

Comparison of subjective grading of lid wiper epitheliopathy with a semi-objective method

Authors: Carolina M.E. Kunnen,¹ James Wolffsohn,^{1,2} Eric Ritchey¹

Affiliations: ¹The Ocular Surface Institute, College of Optometry, University of Houston, Houston, Texas, United States of America; ²Ophthalmic Research Group, Aston University, Birmingham, United Kingdom

Abstract

Purpose: To validate a semi-objective method of grading lid wiper epitheliopathy (LWE) compared to subjective assessment.

Methods: Twenty upper and 20 lower eyelid margins of patients with LWE were photographed after instillation of fluorescein and lissamine green. The images were graded by two observers using a 0-3 grading scale for height (%) and width (mm) of the lid staining. The images were also processed using custom designed software in MATLAB. After manual delineation of the staining area, width and perpendicular height were automatically measured throughout the selected area. The height as a proportion of the lid margin width and width measures were then categorized into the same bins as in the grading scale.

Results: Repeatability of the image analysis system showed a mean difference (95% limits of agreement) between repeats of -0.01mm (0.03 and -0.05mm) for LWE height, 0.04mm (1.16 and -1.08mm) for LWE width, and -0.11mm² (0.32 and -0.53mm²) for LWE area. The mean difference (95% limits of agreement) between image analysis and human grading for LWE height was -0.84 grades (0.54 and -2.21 grades), for LWE width was 0.31 grades (1.22 and -0.59 grades), and for the final grade (mean height and width) was -0.26 (0.44 and -0.96 grades) (all $p < 0.001$).

Conclusion: Human observers tend to overestimate the height and underestimate the width of LWE staining. Lid wiper region is not well defined, thus, it might be a difficult process for human observers to judge the stained region as a proportion of the lid wiper total region.

Keywords:

Lid Wiper Epitheliopathy, LWE, Accuracy, Grading, Image Analysis, Repeatability

1. Introduction

Lid wiper epitheliopathy (LWE) is an alteration of the marginal conjunctival epithelium of the upper or lower eyelid [1,2]. LWE is believed to be caused by mechanical trauma of the lid wiper area with the ocular surface, as a result of insufficient lubrication [3–7]. LWE has received increased attention in the last few years, due to its reported association with symptoms and ocular dryness during contact lens wear [3–5,7–10], as well as its relationship with aging [2].

Currently, the most common way to assess LWE [7] is with the use of lissamine green [11,12] or with a mixture of fluorescein and lissamine green [1,2,4,13–15] in combination with a subjective grading scale [3]. The most common grading scale used to assess LWE [1–3,5,6,9,10,16,17] incorporates the average between the height and the width grade of the stain on the eyelid margin.

Subjective grading depends highly on the ability and accuracy of the clinician to make judgments about the condition in light of its severity and complexity [18]; when the observer has to grade two features (height and width), errors can occur while grading each of these aspects and in how they combine them, which can result in high variability. This limits the value in clinical studies due to factors such as the need for larger sample sizes. Only one paper has attempted to analyse LWE using image analysis, however it reports only the area following manual outlining of the lissamine green staining on the lid margin including the line of Marx [19] which is not considered to be part of LWE [3].

The purpose of this study was to validate a semi-objective method of grading LWE compared to subjective assessment.

2. Methods

The study was approved by the Institutional Research Board of the University of Houston and was carried out in accordance with the guidelines of the Declaration of Helsinki. All participants who had previously been screened for the presence of LWE, gave informed consent prior to enrolment and were aged between 23 and 60 years of age (mean \pm SD: 34.5 \pm 11.9) at the time of enrolment.

2.1 Subjective grading

Twenty upper and twenty lower eyelid margins were photographed after instillation with a solution of 2% fluorescein and 1% lissamine green (compounded ophthalmic drops - Greenpark Compounding Pharmacy, Houston, TX). Two drops of each \sim 20 μ l were instilled in the conjunctival sac approximately 5 minutes apart. Images were captured 1 minute after the second drop was instilled. The 40 images were selected from 116 images of a larger study to encompass the full range of LWE grades (final score $>$ 1), and the images were graded by two masked experienced observers using the 0-3 grading scale for height (%) and width (mm) of the lid staining as represented in Table 1 [3]. Images were presented on a TV screen and settings were kept constant while both observers graded the images separately. For each image, the LWE height was estimated by the observers as a proportional staining height relative to the anatomical lid wiper region, as described by Korb *et al.* [3,5]. The observers then graded the absolute width of the staining. An average of the height and width grading was calculated as the final score, rather than imposing an integer scale on the data in this study. An average final score of the two observers was used in the analysis.

Table 1. LWE grading scale for height (%) and width (mm)[3,5].

Grading	Height (%)	Grading	Width (mm)
0	<25% of the lid wiper height	0	<2
1	25 - <50% of the lid wiper height	1	2-4

2 50 - <75% of the lid wiper height

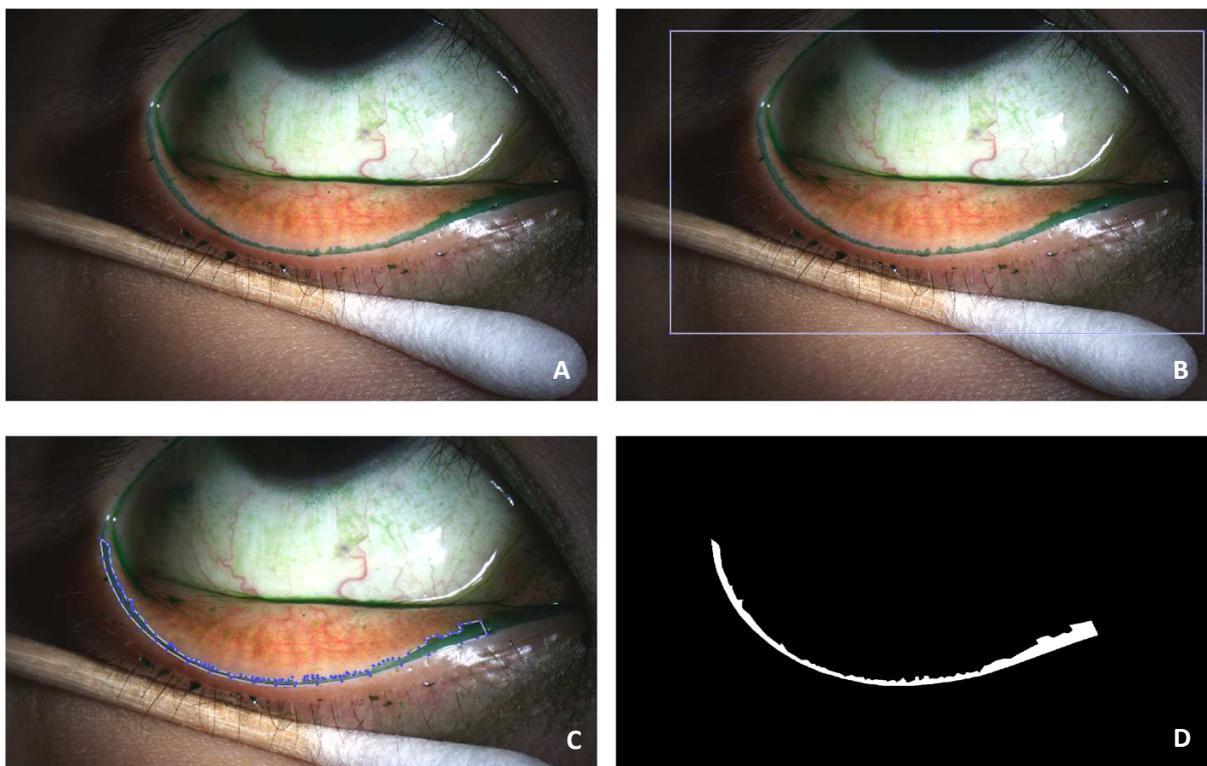
3 >75% of the lid wiper height

2 5-9

3 >10

2.2 Objective Measurement

The same 40 images were also processed using custom designed software in MATLAB (Figure 1). After a region of interest of 12.1 x 21.2mm was selected, the staining area was manually selected by the operator and after delineation automatically masked by transforming it to a black and white image. The operator then manually selected evenly spaced points throughout the masked area. The system automatically drew a spline interpolation curve between the manually selected points and at each midpoint between two adjacent points on the spline curve a perpendicular bisector to the spline was calculated.



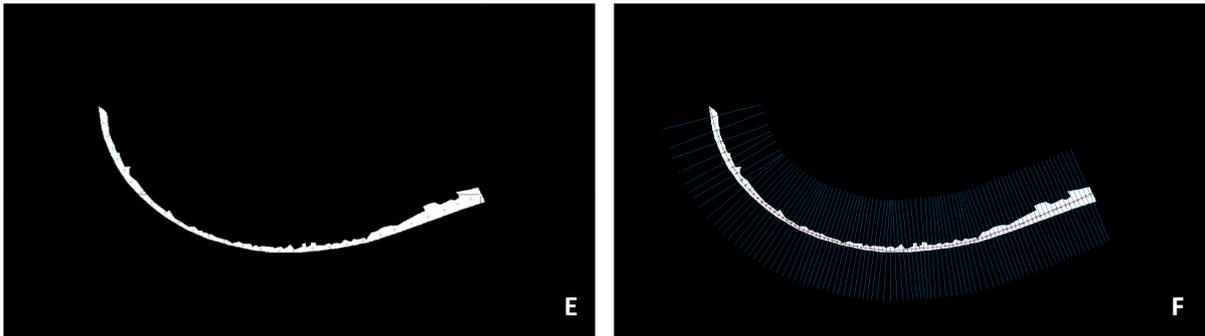


Figure 1. Process of the image with the algorithm; Original image (A). A region of interest of 12.1 x 21.2mm was selected (B). Manual selection of the staining area (C). Masking of the staining area (D). Manual selection of points throughout the masked area (E). Automatic spline interpolation curve was placed between points and at each midpoint between two adjacent points on the spline function, the perpendicular bisector to the spline was calculated (F).

2.3 Data extraction

The height of the staining area for each line was calculated and an average of the height (mm) was used for analysis. The length of the spline curve was used as the width of the staining area (mm).

In some images, the staining area was spread out as several 'patches' (Figure 2). In those cases, each of the staining areas were analysed and a total LWE width (mm) and the overall average height (mm) was used in the analysis.

To make comparisons between the image analysis outcomes and the subjective grading, it was necessary to calculate the height of the LWE as a proportion of the individual lid wiper region. The anatomical lid wiper region was manually computed by measuring the height of the lid wiper area at three locations, approximately 1.5 mm apart, in the centre of the eyelid (Figure 3). The final relative LWE height was calculated by the mean LWE height divided by the average individual lid wiper area height.

The height and width measures were then categorized into the same bins (grade 0-3) as described in the grading scale [3]. The LWE area (mm^2) was calculated from the number of white pixels (representing the LWE area). This was done by multiplying the surface area of 1 pixel with the total number of pixels in the LWE area. To test repeatability of the objective measurement technique, 40 images (20 upper and 20 lower eyelids) were reanalysed on a separate occasion two months after the original analysis by the original investigator, masked to the original readings.

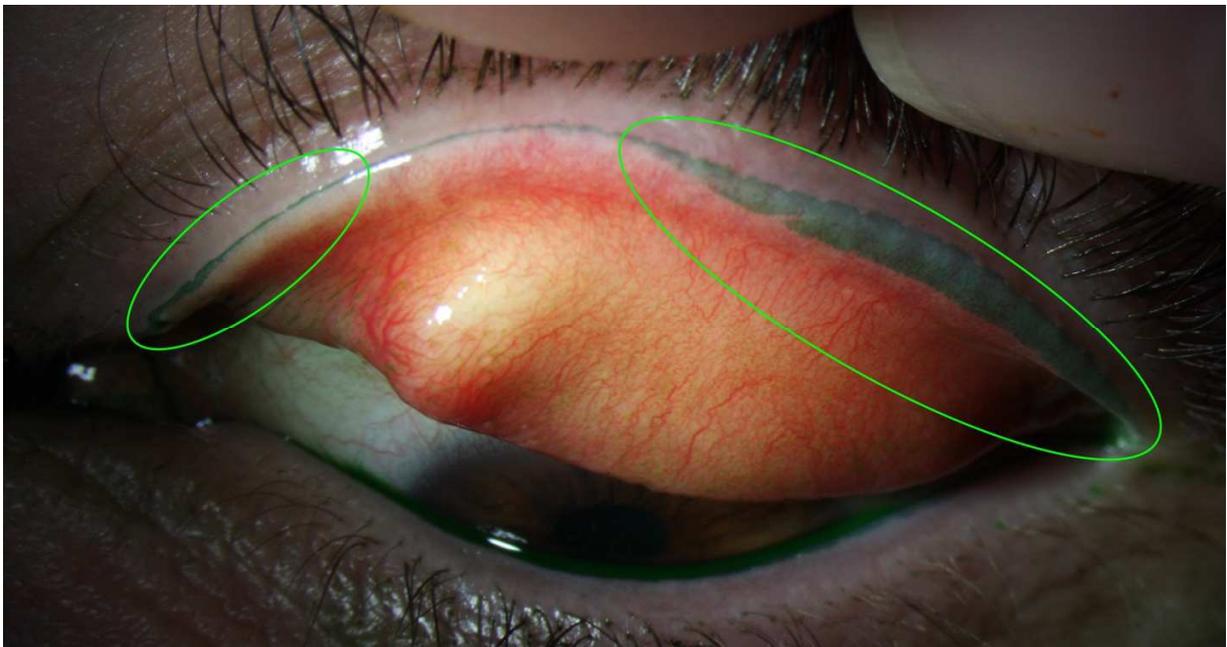


Figure 2. Example of two LWE 'patches' (circled in green).

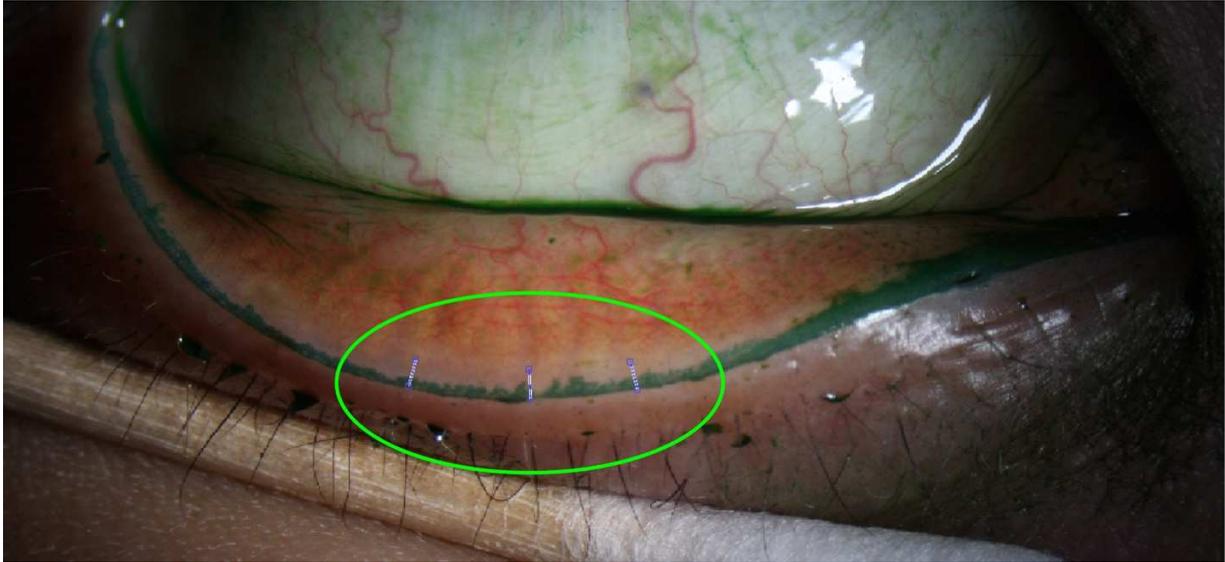


Figure 3. Example of the measurement of the anatomical lid wiper region at three locations, approximately 1.5 mm apart, in the centre of the eyelid (circled in green).

2.4 Statistical analysis

Bland-Altman plots for absolute staining area height and width were used to examine repeatability of the objective measurement technique. The mean of the difference between replicates ± 1.96 SD of the differences represents the 95% limits of agreement. Bland-Altman plots were also used to test the repeatability of the lid wiper area height. Student t-tests were used to compare the normally distributed (Kolmogorov-Smirnov test) upper and lower height and width LWE staining. Bland-Altman plots comparing the height and width measures image analysis outcomes with the average score of the 2 observers were used to examine grading accuracy. Comparisons between height, width and final grade between the image analysis and subjective grading were made using Wilcoxon Signed Ranks Tests. Correlations between the final subjective grade and the relative proportion of LWE were made using Spearman Rank correlation analysis.

3. Results

3.1 Subjective grading

The mean \pm standard deviation (SD) of the final LWE scores of the twenty upper and twenty lower eyelids are shown in Figure 4. The average SD of LWE by the two observers was 1.98 grades (range: 1.00 - 3.00). For LWE height and width the average standard deviation was 1.94 grades (range: 1.00 - 3.00) and 2.01 grades (range: 0.50 - 3.00), respectively (data not displayed).

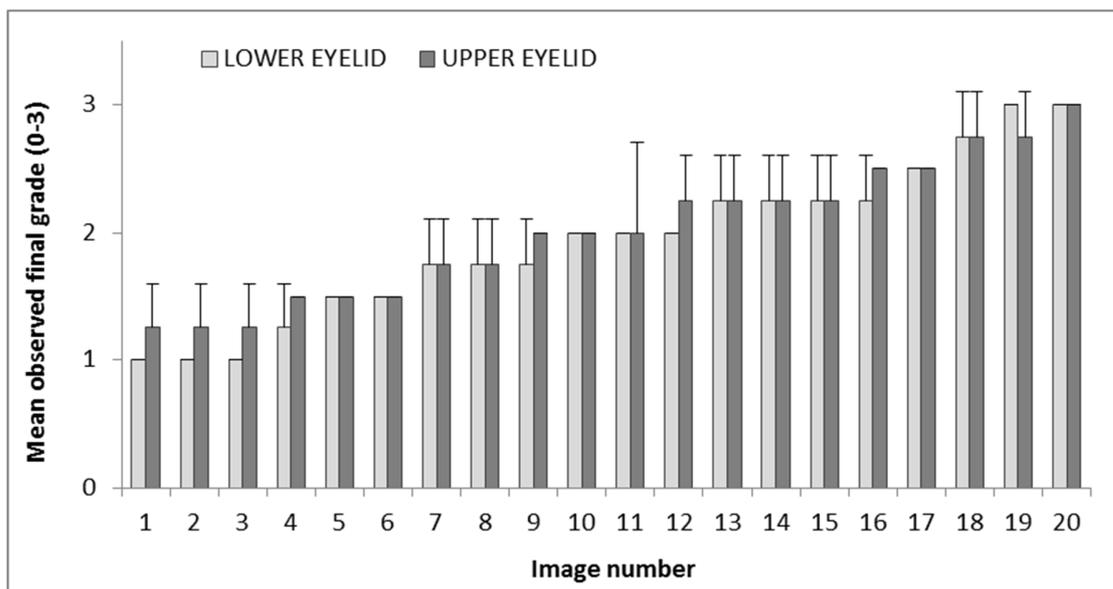


Figure 4. Mean observed final LWE score in each of the 20 images for the lower and upper eyelid from two observers (Error bars = 1 SD). Order of images are presented in terms of severity from low to severe.

3.2 Absolute Image Analysis Results

The mean absolute LWE height (mm) measured with image analysis was not significantly different between the lower and the upper eyelid (0.27 ± 0.06 vs. 0.30 ± 0.13 , $p = 0.27$) (Figure 5A). However, when the LWE height was calculated as relative proportion to the lid wiper area height, a significantly

larger staining height for the lower eyelid was observed (46.33 ± 11.94 vs. 32.89 ± 12.92 , $p = 0.002$) (Figure 5B). LWE width and LWE area were not significantly different between the lower and upper eyelid (10.76 ± 5.50 vs. 10.20 ± 4.39 , $p = 0.72$ and 3.00 ± 2.01 vs. 3.36 ± 2.60 , $p = 0.62$, respectively) (Figure 5C and D). The average anatomical lid wiper region height (Figure 3) of the lower eyelid was 0.59 ± 0.07 mm, whereas the average lid wiper of the upper eyelid was 0.92 ± 0.13 mm.

Image analysis showed that in more than half of the eyelids LWE was distributed over 2 or more areas, or 'patches' (Figure 6). For the lower eyelid, the staining was in 1 or 2 patches, whereas for the upper eyelid the staining area was distributed over 3 or more 'patches' in some cases.

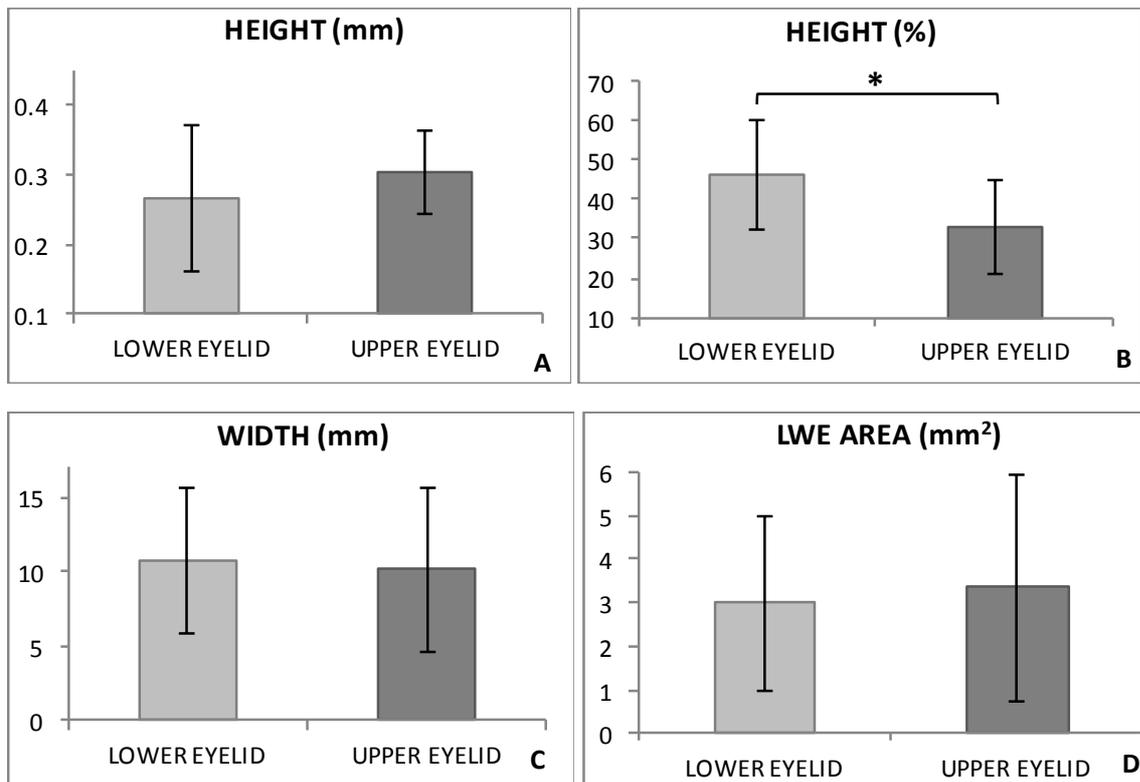


Figure 5. Absolute image analysis results: mean absolute staining height (mm) for the lower and upper eyelid (A); Mean relative proportion of staining to the lid wiper area (%) (B); Absolute width of the staining area (mm) for the lower and upper eyelid (C); Absolute area of staining (mm²) for the lower and upper eyelid (D).

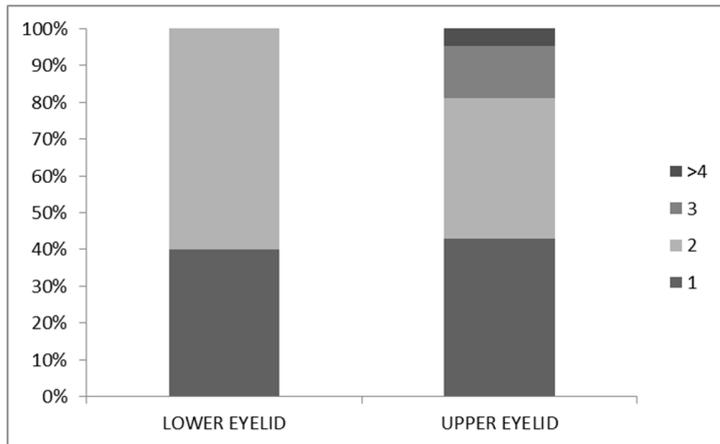


Figure 6. Distribution of staining areas ('patches') per eyelid.

3.1 Image analysis repeatability

The Bland-Altman plots in Figure 7A, B and C show the repeatability of the image analysis for the absolute staining area height, width and LWE area. The mean difference between repeats for the height was -0.01 mm and 95% limits of agreement were between 0.03 and -0.05 mm. The mean difference between repeats for the width was 0.04 mm and 95% limits of agreement were between 1.16 and -1.08 mm. The mean difference between repeats for LWE area was -0.11 mm² and 95% limits of agreement are between 0.32 and -0.53 mm².

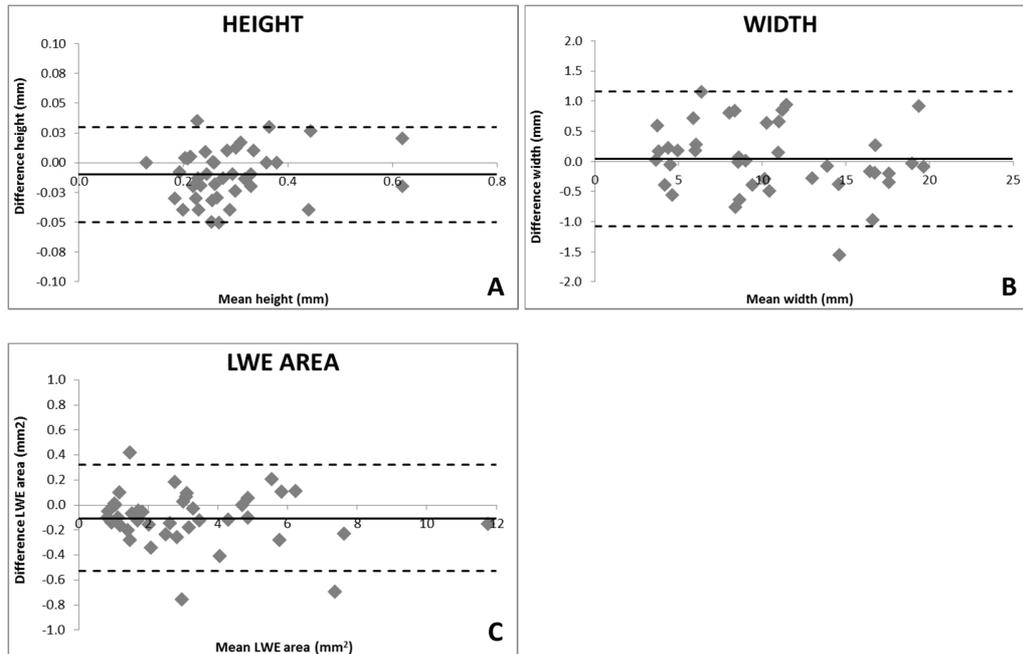


Figure 7. Bland–Altman plots for LWE height (A), width (B) grade, and LWE area (C) between repeats. The solid line shows the mean difference between repeats, the dashed lines show the 95% limits of agreement.

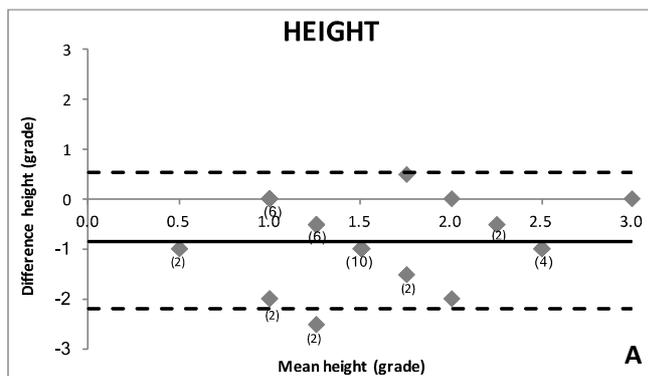
3.3 Comparison of subjective grading vs image analysis

The median and interquartile range (IQR) for LWE height, width and final grade is shown in Table 2. The absolute height, width and final values from objective measurement system were transformed into grades (using the bins described by Korb *et al.* [3,5]) to be able to make comparisons with the subjective grades. The LWE height was graded significantly higher than the height measured with image analysis ($p < 0.001$). The LWE width however was graded significantly lower than the width measured with image analysis ($p < 0.001$). This resulted in a significantly higher subjective final grade than final LWE measured with image analysis ($p < 0.001$). This was also shown in the Bland-Altman plots in Figure 8A, B, and C. The mean difference between methods for the height was -0.84 grades and 95% limits of agreement were

between 0.54 and -2.21 grades. The mean difference between methods for the width was 0.31 grades and 95% limits of agreement were between 1.22 and -0.59 grades. The mean difference between methods for the final grade was -0.26 grades and 95% limits of agreement were between 0.44 and -0.96 grades. Also presented in Table 2 is the coefficient of repeatability (COR) for the image analysis grades. The COR for LWE height, width and final grade was 0.69, 0.43, and 0.41 respectively. Subjective grading was performed once by two observers in this study. The variability between observers for height, width, and final grade was 1.08, 11.12, and 0.76 respectively.

Table 2. Median and interquartile range (IQR) of the height, width, and final grade for image analysis and subjective grading. Significance was set for $p < 0.05$.

	Image Analysis		Subjective Grading		Z	p value
	Median	IQR	Median	IQR		
Height					-3.57	
t	1.00	1.00-1.00	2.00	1.50-2.50		<0.001
Width	2.50	2.00-3.00	2.00	1.50-2.75	-4.93	<0.001
Final	1.75	1.50-2.00	2.00	1.50-2.25	-3.84	<0.001



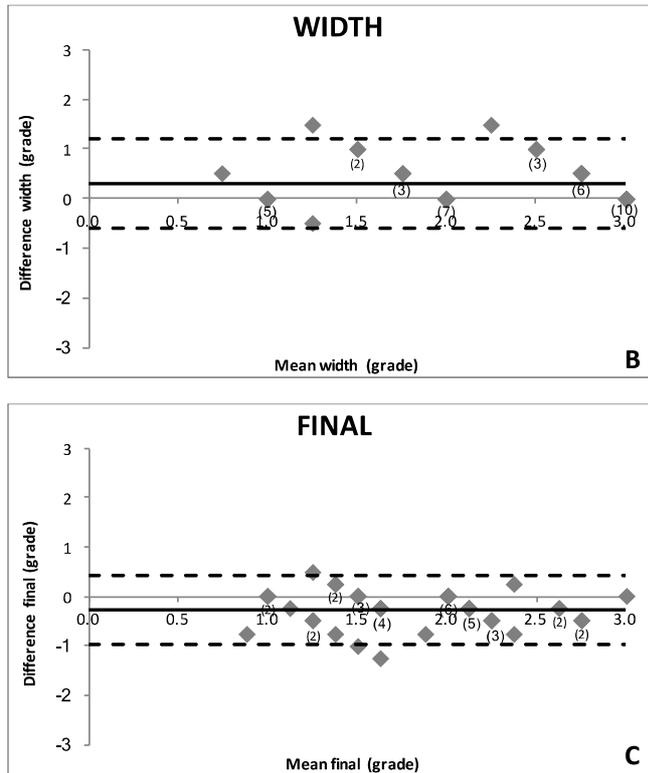


Figure 8. Bland–Altman plots for LWE height (A), width (B), and final (C) grade between image analysis and subjective grading. The solid line shows the mean difference between methods, the dashed lines show the 95% limits of agreement.

3.4 Correlation between final grade and the relative LWE area

Correlation analysis show that the subjective final grade (average between 2 observers) was strongly correlated ($R^2 = 0.68$, $p < 0.001$) with the LWE area (created with image analysis) (Figure 9) with a linear analysis and more strongly with a quadratic ($R^2 = 0.77$, $p < 0.001$).

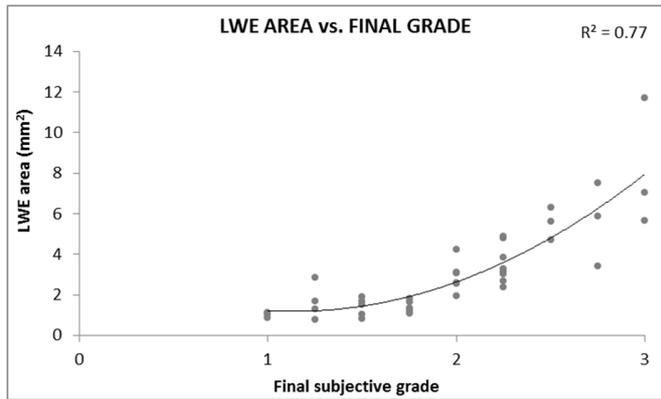


Figure 9. Correlation between the final subjective grade (average of 2 observers) and the LWE area (mm²) calculated with image analysis.

4. Discussion

This study built on the previously reported semi objective image analysis of the LWE [19], providing data on height, height relative to the lid margin width, and width as well as area. The LWE area found in this study was very similar to the area of upper lid LWE reported with the previous methodology [19]. This semi-automatic image analysis system developed proved to be highly repeatable and hence a step towards truly automated objective grading of LWE. However, issues remain such as how to subtract the Marx line and how to cope with the often observed patches of LWE along the lid margins.[20]

This study found that human observers tend to overestimate the height and underestimate the width of LWE staining when comparing their grading with a semi-automatic image analysis system. A larger variance between automatically generated and the subjective height grades was observed compared to the absolute width grade, which may indicate that the height staining relative to the lid wiper area height is a more challenging task than grading the absolute width grade of the staining. One of the explanations might be that the lid wiper region is not well defined, thus, it might be a difficult process

for human observers to judge the stained region as a proportion of the lid wiper total region. Staining can also be spread across several sections on the eyelid, and 'interpreting' (and mentally adding) those areas together to assess the overall LWE width may be a challenging subjective task. Even improving on grading by averaging relative height and width 'bins' rather than making them a combined integer, the 95% confidence interval compared to subjective grading is nearly 1 unit of the 4-point scale (25%) [21] compared to just 10.7% with repeated image analysis for width, 13.8% for height, and 13.0% for area measures. It is also interesting to note the quadratic fit of subjectively grading LWE compared to objective image analysis, which has been observed before with some grading scales [22], caused in this case by the increasing width bands (2mm between grades 1 and 2; 3mm between grades 2 and 3; 5mm between grades 3 and 4) with increasing grade. Although the final subjective grade is higher, the correlation between the LWE area by image analysis and final subjective grade is strong, giving validity to the objective grading methodology.

A previous study found the prevalence of lower LWE was greater than upper lid LWE [2], but that does not fit with the theory that LWE is caused by friction of the lid against the ocular or contact lens surface and hence should be greater on the upper lid which travels more with each blink. However, recent modelling by Pult and colleagues [23] suggests that once the resting friction is overcome, the speed of the upper lid blink results in aqua-planing with minimal friction for the rest of the blink motion. The LWE height between the upper and lower lids were similar in this study, so it is the reduced margin width of the lower compared to the upper lid that results in the relative percentage LWE height being greater for the lower lid.

In conclusion, human observers tend to overestimate the height and underestimate the width of LWE staining when compared to a semi-objective technique. The lid wiper region is not well defined, thus, it might be a difficult process for human observers to judge the stained region as a proportion of the lid

wiper total region. Staining can also be spread across several sections on the eyelid, and ‘interpreting’ those areas together to assess the LWE width may be a challenging task. Further studies are needed to determine which aspects of LWE such as patches, height, relative height, width and even intensity, correlate best with clinical problems such as discomfort and complications.

Acknowledgements

This work was partly supported by Alcon Laboratories Inc. The authors would like to thank Dr. Daniel Powell and Dr. Jessica Mathew for their assistance with the project.

References

- [1] B. Golebiowski, K. Chim, J. So, I. Jalbert, Lid margins: sensitivity, staining, meibomian gland dysfunction, and symptoms., *Optom. Vis. Sci.* 89 (2012) e1443–e1449.
doi:10.1097/OPX.0b013e3182693cef.
- [2] A. Shiraishi, M. Yamaguchi, Y. Ohashi, Prevalence of upper- and lower-lid-wiper epitheliopathy in contact lens wearers and non-wearers., *Eye Contact Lens.* 40 (2014) 220–4.
doi:10.1097/ICL.0000000000000040.
- [3] D.R. Korb, J. V Greiner, J.P. Herman, E. Hebert, V.M. Finnemore, J.M. Exford, T. Glonek, M.C. Olson, Lid-wiper epitheliopathy and dry-eye symptoms in contact lens wearers., *CLAO J.* 28 (2002) 211–6. doi:10.1097/01.ICL.0000029344.37847.5A.
- [4] D.R. Korb, J.P. Herman, C.A. Blackie, R.C. Scaffidi, J. V Greiner, J.M. Exford, V.M. Finnemore, Prevalence of lid wiper epitheliopathy in subjects with dry eye signs and symptoms., *Cornea.* 29 (2010) 377–83. doi:10.1097/ICO.0b013e3181ba0cb2.
- [5] D.R. Korb, J.P. Herman, J. V Greiner, R.C. Scaffidi, V.M. Finnemore, J.M. Exford, C.A. Blackie, T.

- Douglass, Lid wiper epitheliopathy and dry eye symptoms., *Eye Contact Lens*. 31 (2005) 2–8.
doi:10.1097/01.ICL.0000140910.03095.FA.
- [6] C.W. McMonnies, Incomplete blinking: Exposure keratopathy, lid wiper epitheliopathy, dry eye, refractive surgery, and dry contact lenses, *Contact Lens Anterior Eye*. 30 (2007) 37–51.
doi:10.1016/j.clae.2006.12.002.
- [7] N. Efron, N.A. Brennan, P.B. Morgan, T. Wilson, Lid wiper epitheliopathy, *Prog. Retin. Eye Res*. 53 (2016) 140–174. doi:10.1016/j.preteyeres.2016.04.004.
- [8] G. Foulks, R. Chalmers, N. Keir, C.A. Woods, T. Simpson, R. Lippman, W. Gleason, D.A. Schaumber, M.D.P. Willcox, I. Jalbert, The TFOS International Workshop on Contact Lens Discomfort: Report of the Subcommittee on Clinical Trial Designs and Outcomes, *Invest Ophthalmol Vis Sci*. 54 (2013) TFOS71-97. doi:10.1167/iovs.13-13226.
- [9] B. Yeniad, M. Beginoglu, L.K. Bilgin, Lid-Wiper Epitheliopathy in Contact Lens Users and Patients With Dry Eye, *Eye Contact Lens Sci. Clin. Pract.* 36 (2010) 140–143.
doi:10.1097/ICL.0b013e3181d94e82.
- [10] H. Pult, C. Purslow, P.J. Murphy, The relationship between clinical signs and dry eye symptoms., *Eye (Lond)*. 25 (2011) 502–10. doi:10.1038/eye.2010.228.
- [11] W.M. Alghamdi, M. Markoulli, B.A. Holden, E.B. Papas, Impact of duration of contact lens wear on the structure and function of the meibomian glands, *Ophthalmic Physiol. Opt.* 36 (2016) 120–131. doi:10.1111/opo.12278.
- [12] I. Jalbert, S. Rejab, Increased numbers of demodex in contact lens wearers, *Optom. Vis. Sci.* 92 (2015) 671–678. doi:10.1097/OPX.0000000000000605.

- [13] K.-C. Yoon, S.-K. Im, H.-G. Kim, I.-C. You, Usefulness of double vital staining with 1% fluorescein and 1% lissamine green in patients with dry eye syndrome., *Cornea*. 30 (2011) 972–6. doi:10.1097/ICO.0b013e31820687dd.
- [14] H. Pult, C. Purslow, M. Berry, P.J. Murphy, Clinical Tests for Successful Contact Lens Wear: relationship and predictive potential., *Optom. Vis. Sci.* 85 (2008) E924–E929. doi:10.1097/OPX.0b013e3181888909.
- [15] J. Varikooty, S. Srinivasan, L. Jones, Atypical manifestation of upper lid margin staining in silicone hydrogel lens wearers with symptoms of dry eye., *Contact Lens Anterior Eye*. 31 (2008) 44–46. doi:10.1016/j.clae.2007.07.001.
- [16] H. Pult, P.J. Murphy, C. Purslow, A novel method to predict the dry eye symptoms in new contact lens wearers., *Optom. Vis. Sci.* 86 (2009) E1042-50. doi:10.1097/OPX.0b013e3181b598cd.
- [17] Z. Mohammed, M. Hanif, H. Maidin, B. Mohd-ali, Lid wiper epitheliopathy in subjects with and without dry eye symptoms in Kuala Lumpur – a pilot study., *Arch. Des Sci.* 65 (2012) 585–590.
- [18] G. Cardona, C. Serés, Grading contact lens complications: the effect of knowledge on grading accuracy., *Curr. Eye Res.* 34 (2009) 1074–81. doi:10.3109/02713680903340746.
- [19] M. Navascues-Cornago, C. Maldonado-Codina, R. Gupta, P.B. Morgan, Characterization of Upper Eyelid Tarsus and Lid Wiper Dimensions., *Eye Contact Lens*. 42 (2016) 289–94. doi:10.1097/ICL.0000000000000230.
- [20] J. Varikooty, S. Srinivasan, L. Subbaraman, C.A. Woods, D. Fonn, T.L. Simpson, L.W. Jones, Variations in observable lid wiper epitheliopathy (LWE) staining patterns in wearers of silicone hydrogel lenses, *Contact Lens Anterior Eye*. 38 (2015) 471–476. doi:10.1016/j.clae.2015.05.004.

- [21] R.C. Peterson, J.S. Wolffsohn, Objective Grading of The Anterior Eye, *Optom. Vis. Sci.* 86 (2009) 273–278. doi:10.1097/OPX.0b013e3181981976.
- [22] J.S. Wolffsohn, C. Purslow, Clinical monitoring of ocular physiology using digital image analysis., *Cont. Lens Anterior Eye.* 26 (2003) 27–35. doi:10.1016/S1367-0484(02)00062-0.
- [23] H. Pult, S.G.P. Tosatti, N.D. Spencer, J.M. Asfour, M. Ebenhoch, P.J. Murphy, Spontaneous Blinking from a Tribological Viewpoint, *Ocul. Surf.* 13 (2015) 236–249. doi:10.1016/j.jtos.2014.12.004.