



Published in final edited form as:

Cornea. 2017 April ; 36(4): 419–424. doi:10.1097/ICO.0000000000001148.

Precision of Epithelial Defect Measurements

Purak C. Parikh, MD¹, Nita G. Valikodath, MS¹, Christopher B. Estopinal, MD¹, Roni M. Shtein, MD, MS¹, Alan Sugar, MD, MS¹, Leslie M. Niziol, MS¹, and Maria A. Woodward, MD, MS¹

¹Department of Ophthalmology and Visual Sciences, W.K. Kellogg Eye Center, University of Michigan Medical School, Ann Arbor, MI, USA

Abstract

Purpose—To ensure optimal care of patients, corneal specialists measure corneal features, including epithelial defects (ED), with slit lamp calipers. However, caliper measurements are subject to inter-physician variability. We examined the extent of variability in ED measurements between cornea specialists and discuss the potential clinical impact.

Methods—A total of 48 variably sized EDs were created in pig eyes. Three corneal specialists measured the maximum vertical and horizontal ED lengths to the nearest tenth of a millimeter using slit lamp microscopy. An absolute difference in ED measurement between corneal specialists of 0.5mm was chosen to be the *a priori* threshold for clinical significance and was evaluated by Wilcoxon signed-rank tests. Inter-rater reliability was assessed by intra-class correlation coefficients (ICC).

Results—The average absolute difference in vertical ED length between pairs of examiners ranged from 0.54-0.63mm, and horizontal ED length ranged from 0.44-0.46mm. These differences in ED measurement were not significantly different from 0.5mm (all $p > 0.06$). However, pairs of examiners differed in vertical ED length measurements by >0.5 mm in 44-52% of EDs and by >1.0 mm in 13-17% of EDs. Pairs of examiners differed in horizontal ED length measurements by >0.5 mm in 31-40% of EDs and by >1.0 mm in 10-15% of EDs. ICC was 0.85 (95% confidence interval, CI=0.77-0.91) for vertical and 0.84 (95% CI=0.74-0.90) for horizontal ED measurements.

Conclusion—Cornea specialists showed good reliability in measured EDs; however, depending on the threshold for clinical significance, a nontrivial percentage of cases have high inter-examiner clinical variability.

Keywords

Epithelial defect; corneal abrasion; measurement variability

Corresponding Author: Maria A. Woodward MD, MS, University of Michigan, Department of Ophthalmology and Visual Sciences, 1000 Wall Street, Ann Arbor, MI 48105, Phone: 734-763-5506, mariawoo@umich.edu.
Address for reprints: As above for corresponding author

Conflicts of Interest: All other authors have no significant conflicts of interest or sources of funding to report.

Introduction

Corneal epithelial defects (ED) are one of the most common problems encountered in ophthalmology. One study found that 13.7% of all eye-related emergency department visits were due to corneal abrasions.¹ In addition to corneal abrasions, EDs are a hallmark of other potentially sight-threatening pathology such as infectious keratitis,² corneal burns,³ neurotrophic keratopathy,⁴ or Stevens Johnson Syndrome/Toxic Epidermal Necrolysis.⁵ In these conditions, the ED may persist for days to weeks and accurate measurement of the ED size is critical in evaluating appropriate healing.^{6, 7} If the ED size appears to be worsening or not healing at a satisfactory rate, additional or more aggressive measures may be performed dependent on the underlying disease process. The decision to escalate treatment hinges upon the ability to compare the current dimensions of an ED to its previous appearance.

Physicians use textual description, illustrations (via paper charts or electronic medical record drawing tools), or slit lamp microscopy measurements to document corneal features. For these measurements, corneal specialists have different slit-lamp techniques to measure EDs either by vertical and horizontal lengths or by the long axis of the ED. Moreover, ED measurement at the slit lamp has subjective elements due to variable lighting and magnification, even when a standardized approach is used. Variability in measurement may be enhanced with patient movement, use of different slit lamps at each visit, or timing and technique of fluorescein dye instillation. Variability can affect treatment decisions that are based on ED size, especially in the setting of slowly healing EDs when multiple clinicians provide care over time.⁶ In order to understand the degree of variability in greater detail, we studied the inter-observer variability in ED measurements between three experienced board-certified cornea specialists in a controlled, artificial environment.

Materials and Methods

Two study personnel (NV, CE) created variably sized EDs in 24 pig eyes using a #15 blade. One temporal ED and one nasal ED were created on each eye for a total of 48 EDs. The study team deliberately created abrasions of varying sizes to examine a spectrum of measurements. The eyes were stained with fluorescein and a cobalt blue light was used to check for presence of the ED. Eyes were then mounted on polystyrene foam heads and balanced salt solution (BSS) drops were applied to prevent desiccation. Just prior to measurement, eyes were re-stained with fluorescein to visualize the EDs and held for viewing at a slit lamp (Haag-Streit BQ 900, Köniz, Switzerland). The three corneal specialists, not present at ED creation, measured the EDs at the point of maximum horizontal and vertical length to the nearest tenth of a millimeter by slit lamp microscopy. The horizontal and vertical measurements were chosen to minimize variability compared to the use of “long” and “short” axis measurements, which may be subject to greater variability in choice of the axis. The calipers on the slit lamp biomicroscope indicating the length of the light beam were used to make measurements. Examiners took measurements with the epithelial defect in fine focus using the slit lamp. Measurements for each eye were completed within a 30-minute window.

Statistical Analysis

Descriptive statistics of ED size were calculated, including mean, standard deviation (SD), range, and median, and stratified by examiner and horizontal or vertical length. Absolute differences in ED measurements between pairs of examiners were investigated and displayed with histograms. We hypothesized that a discrepancy in ED measurement between examiners of 0.5mm was a clinically significant difference. The absolute differences in ED measurements between examiners were tested for deviations from 0.5mm with Wilcoxon signed-rank tests. The proportion of ED measurements that differed by 0.5mm between pairs of examiners was also reported, including 95% Wilson confidence intervals (CI). Inter-rater reliability of abrasion measurements were assessed with intra-class correlation coefficients (ICC) and reported with 95% CIs. CIs are reported to provide reliability of our estimates in the absence of a power analysis.^{8,9} All analysis was performed with SAS software version 9.4 (SAS Institute, Cary, NC).

Results

The average maximum vertical length of EDs was 3.4 ± 1.5 mm (mean \pm SD), 3.7 ± 1.6 mm, and 3.6 ± 1.4 mm, for examiners 1-3, respectively. Similarly, the average maximum horizontal length of abrasions was 2.9 ± 1.1 mm, 3.2 ± 1.1 mm, and 3.2 ± 1.1 mm, for examiners 1-3, respectively. Table 1 shows the descriptive statistics of ED measurements stratified by examiner and direction of measurement (horizontal or vertical). Comparison of ED measurement between examiners is displayed in scatterplots (Figure 1). A strong, positive linear trend was observed for all pairs of examiners with respect to ED measurements.

The mean absolute difference in vertical ED measurements between examiners was 0.63 ± 0.69 mm for examiner 1 versus 2, 0.56 ± 0.56 mm for examiner 1 versus 3, and 0.54 ± 0.45 mm for examiner 2 versus 3 as shown in Table 2. The mean absolute difference in horizontal ED measurements between examiners was 0.44 ± 0.35 mm for examiner 1 versus 2, 0.46 ± 0.52 mm for examiner 1 versus 3, and 0.46 ± 0.42 mm for examiner 2 versus 3 (Table 2). The absolute measurement differences in vertical and horizontal ED measurements between pairs of examiners were not significantly different from 0.5mm (all $p>0.69$ for vertical length of ED; all $p>0.06$ for horizontal length of ED). However, large measurement differences between examiners greater than or equal to 0.5mm were noted (Figure 2 and Table 3). For vertical ED length, examiners differed in their measurement by 0.5mm in 43.8% to 52.1% of abrasions and by 1.0mm in 12.5% to 16.7%. For horizontal ED length, examiners differed in their measurement by 0.5mm in 31.2% to 39.6% of abrasions and by 1.0mm in 10.4% to 14.6%.

Inter-rater reliability for the measurement of maximum vertical ED length between the three examiners had an ICC of 0.85 (95% CI=0.77-0.91). For maximum horizontal ED length, the ICC was 0.84 (95% CI=0.74-0.90).

Discussion

The results of this study show that the horizontal and vertical measurements of corneal EDs were relatively consistent when measured separately by three different specialist examiners

in a controlled, artificial environment. A priori, we decided that measurement differences greater than 0.5mm would be meaningful, as it would indicate clinically significant variability between examiners. The absolute measurement differences were not significantly different than 0.5mm, but 31-52% of ED measurements had greater than 0.5mm variation between examiners when vertical and horizontal ED size measurements were pooled with 95% CIs showing this effect exists even with a small sample size.

While it is reassuring that there was not a statistically significant difference in measurements greater than the standard of 0.5mm, it is of concern that the CI was wide, with 10-17% of measurements differing by more than 1.0 mm. These results indicated that consistent corneal ED measurements are difficult even in a controlled environment where patient cooperation was not a factor. For instance, certain measurements (Figure 1) have significant discrepancies. For vertical ED length measurements, Examiner 1 was very different than Examiners 2 and 3 (6.0 mm versus 2.8mm and 3.0 mm, respectively). We made an effort to control for external sources that could contribute to variability in measurement between examiners in our study by mounting eyes on standard artificial heads and using the same examination room and lighting conditions. Potential sources of variability are also present in clinical practice. Therefore, we suspect that with real patient encounters, ED size measurement variability would be greater though such large deviations would hopefully not be present.

Our study has implications when multiple providers are involved in a patient's care, such as in academic centers with trainees or large practices with shared patient management. Multiple providers with varying practice patterns or levels of experience would likely have more variable measurements of EDs, potentially influencing treatment decisions. Shared management of complex patients is a known issue when performing patient care. Outside of the field of ophthalmology, studies support the notion that an increased number of providers caring for the same patient during an acute illness may worsen clinical outcomes.¹⁰ Poor outcomes have been shown to occur more often when a “covering” physician was involved in the care of the patient.¹¹ This situation is analogous to ophthalmology when an “on-call” physician (e.g. during a weekend or holiday) is tasked with examining a patient requiring frequent follow-up. It stands to reason that in the case of prolonged EDs with multiple providers, similar variability may occur. This can be problematic given that the ED size is a key marker of improvement or worsening in clinical course of a corneal ulcer, for example.

Inter- and intra-observer variability in the field of ophthalmology has been demonstrated in other eye examination findings, such as the measurement of cup to disc ratio,¹² corneal white-to-white diameter,¹³ tear break-up time,¹⁴ retinal arterio-venous ratio,¹⁵ and the clinical staging of diabetic retinopathy.¹⁶ Advanced technology in the field has helped decrease this variability with the advent of automated visual fields and optical coherence tomography of the retinal nerve fiber layer and macula, which are now considered the standard of care.¹⁷⁻²² In addition, for intraocular lens calculations, physicians have transitioned from manual keratometry and ultrasound biometry to automated or optical biometry. These transitions have been shown to improve refractive outcomes after cataract surgery.^{23, 24} We did not find significant variability at the threshold for concern; however, the percentage of cases with high variability highlights need for more investigation in this

area. We suspect that the advent of photographic and computerized image analysis of EDs would improve the precision and accuracy of these measurements, but it is yet to be determined whether these are practical for clinical application.

The main limitation of our study was the use of a post-mortem pig-eye model. Pig eyes are susceptible to corneal desiccation between examinations. To help prevent desiccation, each individual eye was examined by all examiners within a 30-minute window with frequent BSS wetting. In addition, examining a mounted eye does not perfectly replicate a real clinical examination. Our assumption was that the environment with pig eyes was more controlled, thus less variable, than a clinical environment. We were not able to assess intra-observer variability. We felt that the short duration of the study (all measurements within 30 minutes) did not allow us to mask graders effectively and prolonged times between measurements would affect the corneas. A “gold-standard” to record the epithelial defect size was also not defined because manually measuring the size of the defect on the eye itself with handheld calipers was found to be unreliable based on preliminary testing. For future work, we will take photographs of all eyes and use imaging software to measure the size of the defects relative to the horizontal white-to-white measurements. Finally, we examined only small to medium sized EDs, and the variance might be larger with very small or very large defects.

In conclusion, our study shows that inter-examiner measurements of ED size have good precision. However, despite cornea-trained specialists performing the measurements in a controlled environment, 10-17% were 1.0mm different between examiners. Variability between providers can change clinical management decisions. Future studies should be aimed towards studying the variability between examiners measuring corneal pathology in human participants, such as EDs in corneal ulcers, corneal injury with trauma, or burn injuries, as these may represent situations where prolonged EDs and shared patient management is most likely.

Acknowledgments

Source of Funding: Maria A. Woodward receives a grant (K23 Mentored Clinical Scientist Award K23EY023596) from National Institutes of Health, Bethesda, MD. The funding organizations had no role in the design or conduct of this research.

References

1. Channa R, Zafar SN, Canner JK, et al. Epidemiology of Eye-Related Emergency Department Visits. *JAMA Ophthalmol.* 2016; 134:312–319. [PubMed: 26821577]
2. Herretes S, Wang X, Reyes JM. Topical corticosteroids as adjunctive therapy for bacterial keratitis. *Cochrane Database Syst Rev.* 2014; 10 CD005430.
3. Roper-Hall MJ. Thermal and chemical burns. *Trans Ophthalmol Soc U K.* 1965; 85:631–653. [PubMed: 5227208]
4. Bonini S, Lambiase A, Rama P, et al. Topical treatment with nerve growth factor for neurotrophic keratitis. *Ophthalmology.* 2000; 107:1347–1352. [PubMed: 10889110]
5. Sotozono C, Ang LP, Koizumi N, et al. New grading system for the evaluation of chronic ocular manifestations in patients with Stevens-Johnson syndrome. *Ophthalmology.* 2007; 114:1294–1302. [PubMed: 17475335]

6. American Academy of Ophthalmology. 2015-2016 Basic and Clinical Science Course (BCSC), Section 8: External Disease and Cornea. American Academy of Ophthalmology; 2015. City
7. Krachmer, JH., Mannis, MJ., Holland, EJ. Cornea Fundamentals, Diagnosis and Management. 3rd. St. Louis, MO: Mosby; 2010.
8. Levine M, Ensom MH. Post hoc power analysis: an idea whose time has passed? *Pharmacotherapy*. 2001; 21:405–409. [PubMed: 11310512]
9. Colegrave N, Ruxton GD. Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. *Behavioral Ecology*. 2002; 14:446–447.
10. Cohen MD, Hilligoss PB. The published literature on handoffs in hospitals: deficiencies identified in an extensive review. *Qual Saf Health Care*. 2010; 19:493–497. [PubMed: 20378628]
11. Petersen LA, Brennan TA, O'Neil AC, et al. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med*. 1994; 121:866–872. [PubMed: 7978700]
12. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc*. 1976; 74:532–572. [PubMed: 867638]
13. Baumeister M, Terzi E, Ekici Y, et al. Comparison of manual and automated methods to determine horizontal corneal diameter. *J Cataract Refract Surg*. 2004; 30:374–380. [PubMed: 15030827]
14. Nichols JJ, Nichols KK, Puent B, et al. Evaluation of tear film interference patterns and measures of tear break-up time. *Optom Vis Sci*. 2002; 79:363–369. [PubMed: 12086302]
15. Heitmar R, Kalitzeos AA, Patel SR, et al. Comparison of subjective and objective methods to determine the retinal arterio-venous ratio using fundus photography. *J Optom*. 2015; 8:252–257. [PubMed: 26386537]
16. Gonzalez ME, Gonzalez C, Stern MP, et al. Concordance in diagnosis of diabetic retinopathy by fundus photography between retina specialists and a standardized reading center. Mexico City Diabetes Study Retinopathy Group. *Arch Med Res*. 1995; 26:127–131. [PubMed: 7620277]
17. Strøm C, Sander B, Larsen N, et al. Diabetic macular edema assessed with optical coherence tomography and stereo fundus photography. *Invest Ophthalmol Vis Sci*. 2002; 43:241–245. [PubMed: 11773037]
18. McDonald HR, Williams GA, Scott IU, et al. Laser scanning imaging for macular disease: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2007; 114:1221–1228. [PubMed: 17466377]
19. Virgili G, Menchini F, Dimastrogiovanni AF, et al. Optical coherence tomography versus stereoscopic fundus photography or biomicroscopy for diagnosing diabetic macular edema: a systematic review. *Invest Ophthalmol Vis Sci*. 2007; 48:4963–4973. [PubMed: 17962446]
20. Prum, BE., Jr, Rosenberg, LF., Gedde, SJ., et al. [Accessed September 6, 2016] American Academy of Ophthalmology Primary Open-Angle Glaucoma Preferred Practice Pattern® Guidelines. 2015. [AAO web site]. Available at: <http://www.aao.org/preferred-practice-pattern/primary-open-angle-glaucoma-ppp-2015>
21. Olsen, TW., Adelman, RA., Flaxel, CJ., et al. [Accessed September 6, 2016] American Academy of Ophthalmology Age-Related Macular Degeneration Preferred Practice Pattern® Guidelines. 2015. [AAO web site]. Available at: <http://www.aao.org/preferred-practice-pattern/age-related-macular-degeneration-ppp-2015>
22. Olsen, TW., Adelman, RA., Flaxel, CJ., et al. [Accessed September 6, 2016] American Academy of Ophthalmology Diabetic Retinopathy Preferred Practice Pattern® Guidelines. 2016. [AAO web site]. Available at: <http://www.aao.org/preferred-practice-pattern/diabetic-retinopathy-ppp-updated-2016>
23. Landers J, Goggin M. Comparison of refractive outcomes using immersion ultrasound biometry and IOLMaster biometry. *Clin Experiment Ophthalmol*. 2009; 37:566–569. [PubMed: 19702705]
24. Vogel A, Dick HB, Krummenauer F. Reproducibility of optical biometry using partial coherence interferometry: intraobserver and interobserver reliability. *J Cataract Refract Surg*. 2001; 27:1961–1968. [PubMed: 11738911]

Abbreviations

ED	epithelial defect
SD	standard deviation
ICC	intra-class correlation coefficients
BSS	balanced salt solution

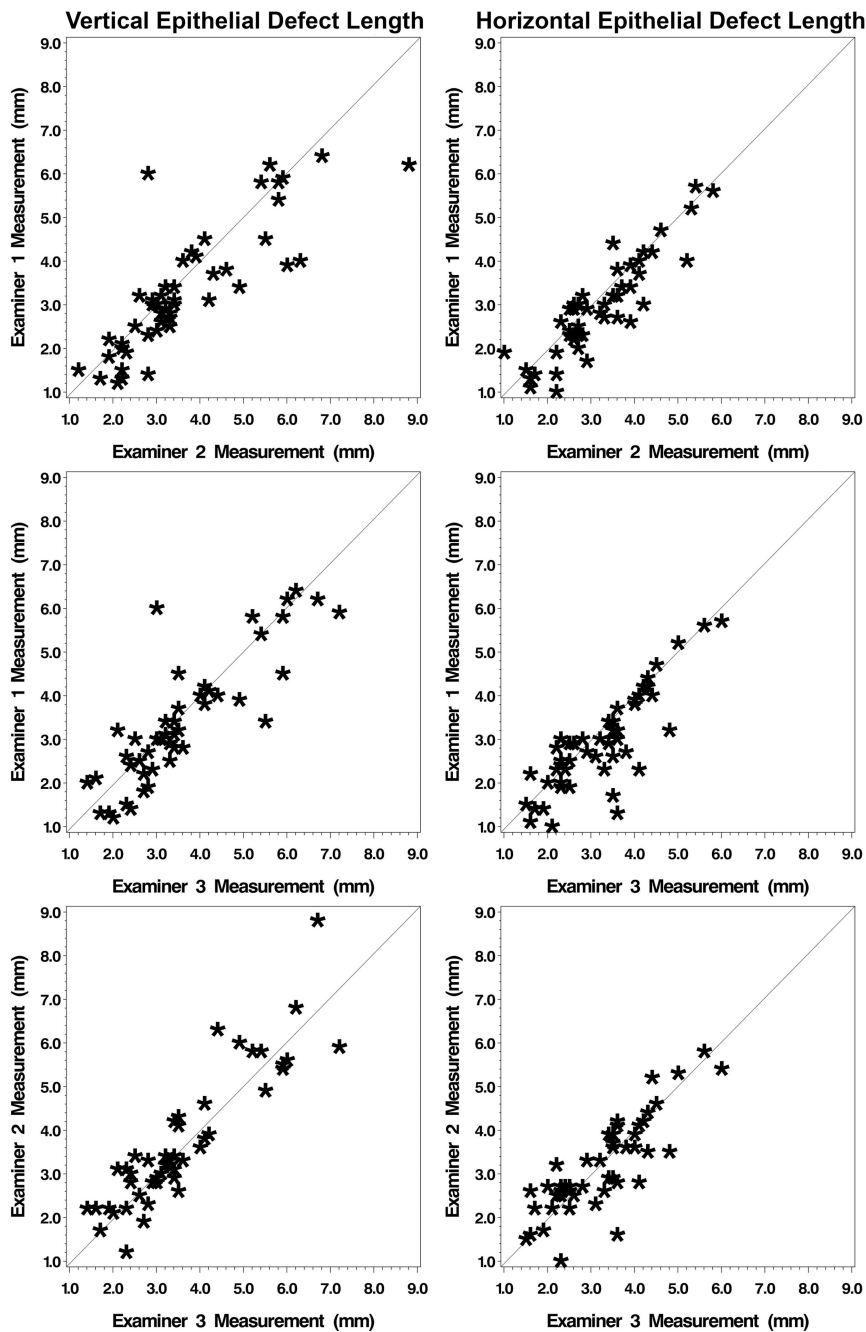


Figure 1. Comparison of ED measurement between examiners. A strong, positive linear trend was observed for all pairs of examiners with respect to abrasion measurement.

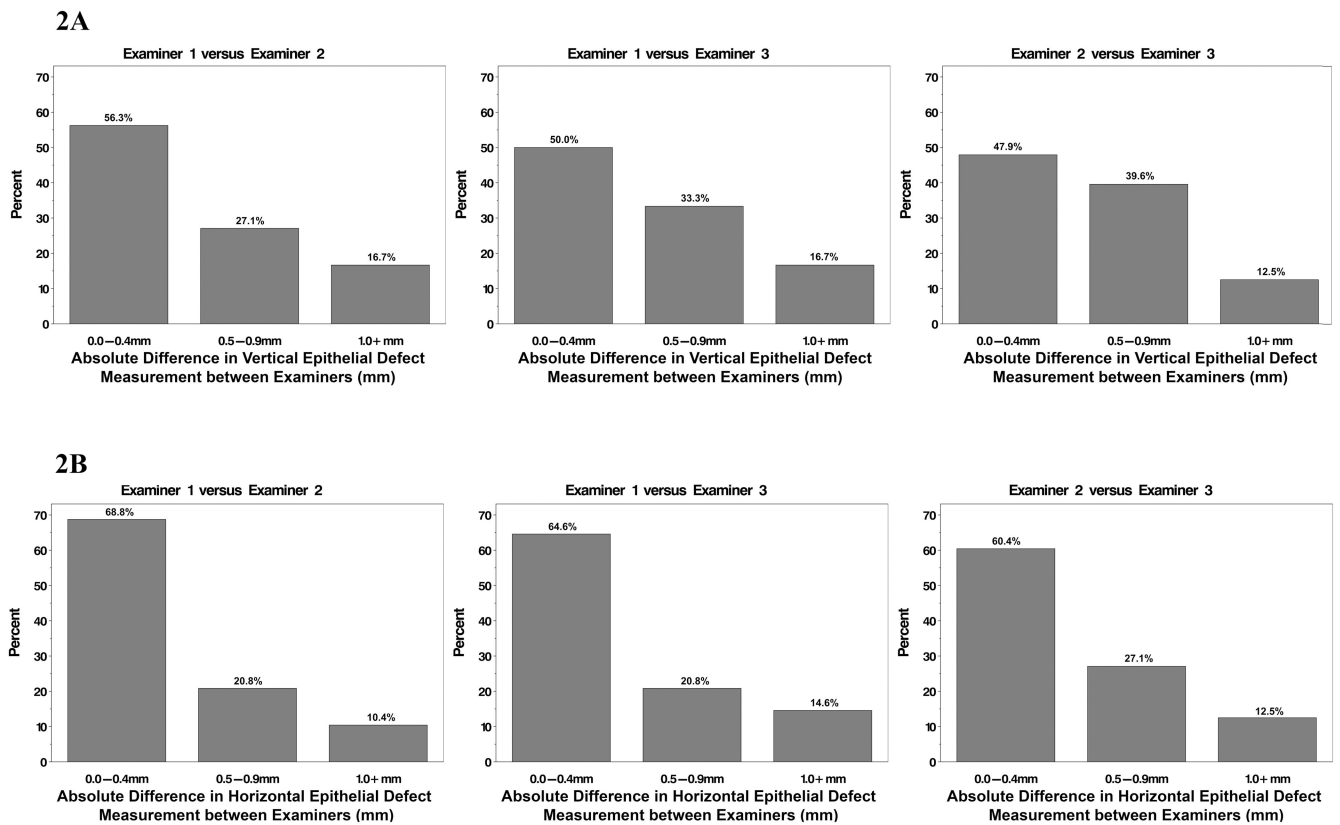


Figure 2. Measurement differences between examiners displayed as proportions grouped by the absolute difference between ED measurements. 2A displays the vertical length measurements and 2B displays the horizontal length measurements.

Table 1
Descriptive statistics of ED measurements (n=48 corneal abrasions)

Examiner	Maximum Vertical Length of ED (mm)			Maximum Horizontal Length of ED (mm)		
	Mean (SD)	Min, Max	Median	Mean (SD)	Min, Max	Median
E1	3.35 (1.45)	1.20, 6.40	3.10	2.91 (1.12)	1.00, 5.70	2.85
E2	3.67 (1.56)	1.20, 8.80	3.25	3.17 (1.06)	1.00, 5.80	2.85
E3	3.56 (1.41)	1.40, 7.20	3.30	3.22 (1.07)	1.50, 6.00	3.35

mm = millimeter, SD = standard deviation.

Table 2
Difference in epithelial defect measurements between examiners (n=48 abrasions)

Examiners	Absolute Difference in Maximum Vertical Length of ED (mm)				Absolute Difference in Maximum Horizontal Length of ED (mm)			
	Mean (SD)	Min, Max	Median	95% CI	Mean (SD)	Min, Max	Median	95% CI
E1 vs E2	0.63 ^a (0.69)	0.00, 3.20	0.40	0.43, 0.83	0.44 ^d (0.35)	0.00, 1.30	0.30	0.31, 0.61
E1 vs E3	0.56 ^b (0.56)	0.00, 3.00	0.45	0.40, 0.72	0.46 ^e (0.52)	0.00, 2.30	0.30	0.33, 0.58
E2 vs E3	0.54 ^c (0.45)	0.00, 2.10	0.50	0.41, 0.67	0.46 ^f (0.42)	0.00, 2.00	0.35	0.34, 0.54

ED = epithelial defect, E = Examiner, mm = millimeter, SD = standard deviation, CI = confidence interval; Wilcoxon signed-rank test p-value:

^a 0.69,

^b 0.96,

^c 0.81,

^d 0.08,

^e 0.06,

^f 0.12

