



Published in final edited form as:

Cornea. 2018 March ; 37(3): 331–339. doi:10.1097/ICO.0000000000001488.

Novel image-based analysis for reduction of clinician-dependent variability in measurement of corneal ulcer size

Tapan P. Patel, MD, PhD¹, N. Venkatesh Prajna, MBBS², Sina Farisu, PhD³, Nita G. Valikodath, MD¹, Leslie M. Niziol, MS¹, Lakshey Dudeja, MBBS², Kyeong Hwan Kim, MD^{1,4}, and Maria A. Woodward, MD, MS^{1,5}

¹Department of Ophthalmology and Vision Science, W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan, United States

²Department of Cornea and Refractive Services, Aravind Eye Care System, Madurai, India

³Departments of Biomedical Engineering and Ophthalmology, Duke University, Durham, NC, Unites States

⁴Department of Ophthalmology, Inje University College of Medicine; and Haeundae Paik Hospital, Busan, South Korea

⁵Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan, United States

Abstract

Purpose—To assess variability in corneal ulcer measurements between ophthalmologists and reduce clinician-dependent variability by using semi-automated segmentation of the ulcer from photographs.

Methods—Three ophthalmologists measured 50 patients' eyes for epithelial defect (ED) and stromal infiltrate (SI) size using slit lamp (SL) calipers. SL photographs were obtained. An algorithm was developed for semi-automatic segmenting (SAS) of the ED and SI in the photographs. SAS was repeated 3 times by different users (2 ophthalmologists and 1 trainee). Clinically significant variability was assessed with intraclass correlation coefficients (ICC) and the percentage of pair-wise measurements differing by 0.5 mm. SAS measurements were compared to manual delineation of the image by a cornea specialist (gold standard) using Dice similarity coefficients.

Results—Ophthalmologists' reliability in measurements by SL calipers had an ICC from 0.84–0.88 between examiners. Measurements by SAS had an ICC from 0.96–0.98. SL measures of ulcers by clinical versus SAS measures differed by 0.5 mm in 24%–38% vs. 8%–28% (ED height); 30%–52% vs. 12%–34% (ED width); 26%–38% vs. 10%–32% (SI height); and 38%–58% vs. 14%–34% (SI width), respectively. Average Dice similarity coefficients between manual and repeated SAS ranged from 0.83–0.86 for ED and 0.78–0.83 for SI.

Correspondence: Maria A. Woodward, MD, MS, Department of Ophthalmology and Vision Science, W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan, United States, 1000 Wall St., Ann Arbor, MI 48105, Tel: +1-734-763-5506, Fax: +1-734-232-2332, mariawoo@med.umich.edu.

Conflict of interest: The authors have no proprietary or commercial interest in any of the materials discussed in this article.

Conclusion—Variability exists when measuring corneal ulcers, even amongst ophthalmologists. Photography and computerized methods for quantifying ulcer size could reduce variability while remaining accurate and impact quantifying measurement endpoints.

Keywords

Corneal ulcer; inter-examiner variability; semi-automated measurement; random forest segmentation

Introduction

Microbial keratitis, better known as corneal ulcers, is a leading cause of corneal blindness in developing nations.^{1,2} The prevalence of corneal ulceration varies greatly from country to country, and even from one region to another, depending on factors such as trauma, contact lens wear, hygiene, availability, and general standards of eye care. Estimates are difficult to generate. Several retrospective studies estimate incidence of 113 ulcers per 100,000 persons annually in south India,³ 799 per 100,000 in Nepal,⁴ and similar rates in Bhutan⁵ and Burma.⁶ In the United States, there were almost 1 million clinic visits for keratitis in 2010, with an estimated health care burden of \$175 million.⁷ If left untreated or inappropriately managed, corneal ulcer can cause blindness, corneal perforation, or endophthalmitis. Optimal care of corneal ulcer patients includes early diagnosis and accurate assessment of response to treatment. Key parameters when evaluating ulcer severity are the size of the overlying epithelial defect (ED) and the size of the stromal infiltrate (SI).⁸ Ophthalmologists measure ED and SI at each visit to guide testing, to manage medications, and to escalate interventions when needed.

In the care of patients with vision-threatening corneal diseases, like corneal ulcers, cornea specialists must accurately describe the size and location of the ulcer. Along with pathogen type and antimicrobial sensitivities, presenting ulcer features including ulcer size, location, and initial visual acuity are critical when predicting corneal ulcer outcomes.^{8–11} While visual acuity measures have some standardization, ophthalmologists use heterogeneous methods to record ulcer characteristics including drawings, descriptive text, and caliper measurements. Further, caliper measurements have inherent examiner-dependent variability. Often, the treating cornea specialist assumes he or she will provide all of the care and can ‘remember’ the previous encounter appearance or document only using descriptive text. Thus, sometimes quantified measurements are not even performed, and ulcer patients will therefore have little to no quantitative measurements. The lack of quantified measures does not always affect patient care. In healthcare settings with shared provider models or with transfer of care between sites, different ophthalmologists may examine the patient on subsequent encounters adding variability to ulcer assessments. Previous work has shown up to 17% of measurements differed by 1.0 mm between specialists.⁹ We hypothesize that similar or even greater variability exists between providers in a true clinical setting.

Many subfields of ophthalmology utilize routine imaging and quantitative computer-aided image analysis for accurate diagnosis,^{13,14} enhanced prognosis,^{15,16} and providing better patient care, such as employing optical biometry for intraocular lens selection.¹⁷

Ophthalmologists' methods to quantify corneal pathology includes topography and tomography for keratoconus and pachymetry for corneal edema, yet a computerized quantification tool does not exist for corneal ulcers. Previous research on computer-aided ulcer quantification has relied on manually tracing the boundaries of an ED from digital external photographs using commercially available software like Image Pro Plus,¹⁸ Adobe Photoshop,¹⁹ or SigmaScan.²⁰ More recently, Toutain-Kidd et al. described an easy to use, online, computer software to assess digital corneal photographs of fungal keratitis.²¹ In that study, both ophthalmologists and non-ophthalmologists (medical students) manually traced the boundaries of the stromal infiltrate from digital photograph and achieved very good agreement in clinically relevant variables, such as area of infiltrate and proximity to the visual axis. We propose a semi-automated quantitative corneal measurement (QCM) method for segmentation and measurements of corneal ulcer from external photographs.

By incorporating imaging and computerized analysis, we hypothesize that variability in ulcer measurement can be reduced. To assess the validity of this hypothesis, the purpose of this study was two-fold: 1) to assess corneal ulcer measurement variability amongst ophthalmologists in a true clinical setting, and 2) to reduce measurement variability by developing a semi-automated QCM software package.

Materials and Methods

Study design

A sample of 50 patients with corneal ulcers were recruited from a cornea clinic at Aravind Eye Hospital, Madurai, India from June 14, 2016 to December 19, 2016. The study was approved by the institutional review board at Aravind Eye Hospitals. Participant consent was obtained. The patients selected to participate in the study were present in the clinic on days that all three examiners were available. The ulcer pathogen was typically unknown at the time of presentation. If patients required immediate surgical intervention, they were excluded from the study. Examiners were one cornea-trained, board-certified ophthalmologist (NR, abbreviated as E1 throughout the text) and two board-certified cornea clinical fellows (HSS and HS, abbreviated as E2 and E3). The examiners measured the largest vertical distance (height) and largest horizontal distance (width) of the ED and of the SI to the nearest tenth of a millimeter using the slit-lamp biomicroscope calipers (Haag-Streit International, BX 900, Wedel, Germany). In cases where patients presented with multiple lesions (e.g. fungal keratitis), the largest (primary) corneal lesion was measured. Each examiner was masked to the measurements obtained by the other examiners (Figure 1). All examiners first measured the cornea and then imaging was performed. Additionally, imaging was performed prior to corneal scraping, if necessary for clinical care.

Image analysis

A cornea ophthalmology fellow took an external, diffuse light, slit lamp photograph of the ulcerated eye under white light and cobalt blue light illumination (with fluorescein) prior to any corneal scraping (Canon EOS 7D camera) (Figure 2A). Photographs were taken with a Canon EOS 7D camera mounted on a Haag-Steit International BX 900 model slit lamp biomicroscope. Ambient room lighting was used for the study and a diffuse beam at

maximal width (30mm) of the white light source and maximal light intensity as tolerated by the patient were used. Photographs were taken at the 10× magnification setting. If fluorescein had been used during the clinical examination, the patient waited 15 minutes prior to photography and topical artificial tears were used prior to initial imaging. Fluorescein was then instilled in the eye for cobalt blue excitation filter photographs. The photographer was instructed to image the primary lesion in the case of multiple lesions.

Photographs were analyzed in two ways: 1) manually traced by a single ophthalmologist, and 2) using the proposed QCM algorithm for semi-automated segmentation. For each image, manual measurements for height and width of ED and SI were obtained by tracing the boundary of the ulcer and using the ruler tool from ImageJ software (National Institutes of Health, Bethesda, MD; available at: <http://imagej.nih.gov/ij>). *A priori* white-to-white measurements were not obtained on patients, therefore a horizontal white-to-white distance of 11.7 mm, the mean white-to-white distance of south Indian populations,²² was used to calibrate the size of an image pixel (pixel pitch in millimeters). An ophthalmologist who was not involved in examining the patient at the slit-lamp or acquisition of photographs (KHK) performed all manual measurements.

A QCM algorithm was developed to facilitate semi-automatic segmentation (delineation) of ED and SI from external photography (Figure 2A). The algorithm was written in MATLAB version R2014b (The Mathworks, Inc., Natick, MA) and utilizes the Insight Segmentation and Registration Toolkit²³ and the command-line subroutines of Convert 3D for random forest classifier implementation.²⁴ Given the wide range of corneal ulcer appearances and lack of distinct edge boundary between an infiltrate and normal cornea, we adopted a random forest classification scheme to first generate a probability map of the foreground (SI or ED) and background.^{25,26} In order to train the classifier, the user first initializes seed regions in the foreground (ED or SI) and the surrounding background regions, providing the classifier with ground truths for each class (Figure 2B). Training parameters were: 50 trees in a forest, maximum depth of 30 for each tree, and a 5 pixel radius of patch used to generate mean intensity, and pixel coordinate features. These parameters were empirically tuned for speed and accuracy of training. Once the classifier is trained, the remaining steps of the QCM analysis pipeline were performed. (Figure 2C–F). ED and SI height and width measurements were obtained from the segmentation, similar to manual measurement. For illustration purposes, Figure 2 shows QCM pipeline for measurement of SI dimensions; similar approach was carried out for measurement of ED dimensions using a photograph of the eye under cobalt blue illumination after fluoresceine dye instillation. Similar to the manual analysis, the pixel pitch in millimeters was estimated based on a horizontal white-to-white distance of 11.7 mm. Since our semi-automated method relies on user selection of foreground and background regions, a repeated analysis with a different seed region may produce different segmentation result, and hence different measurements of ED and SI dimensions. To test inter-observer repeatability, each photograph was segmented via the semi-automated method by three different users (TPP, an ophthalmology trainee, and MAW & PANC, ophthalmologists, abbreviated as A1-A3 through the text).

Statistical Analysis

Descriptive statistics of the measured horizontal or vertical length of ED and SI were calculated, including mean, standard deviation (SD), range, and median, and stratified by examiner (E1–E3) and method of segmentation measurement (manual or semi-automated). Scatterplots plots were used to assess the agreement in measurement between pairs examiners and between manual and semi-automated methods, and the degree of linear association was assessed with Pearson's correlations (ρ). Absolute differences in ED and SI measurements between pairs of examiners and between manual and semi-automated methods were investigated and displayed with histograms. A threshold of 0.5 mm absolute difference in measurement length was deemed a clinically significant difference. The absolute differences in ED and SI measurements between examiners and between methods were tested for deviations from 0.5 mm with Wilcoxon signed rank tests. Dice similarity coefficient was computed to determine the spatial overlap of semi-automated segmentation of ED and SI, compared to manual segmentation by a cornea specialist, which served as the gold standard.²⁷ More specifically, this statistic is the proportion of twice the number of pixels that are identified as ulcer in both images (intersection of ulcer identified by manual segmentation and by semi-automated segmentation) by the sum of pixels identified as ulcer in each image (Figure 3). Dice coefficient ranges from 0 (no overlap) to 1 (perfect overlap between manual and computerized segmentations; in general, Dice coefficient > 0.8 is considered very good). Reliability of measurements between examiners and between repeated segmentations with the QCM algorithm was assessed with intraclass correlation coefficients (ICCs) and reported with 95% confidence intervals (CIs). ICC analysis was performed with SAS software version 9.4 (SAS Institute, Cary, NC) and Dice coefficient was measured with MATLAB version R2014b (The MathWorks Inc., Natick, MA).

Results

The average age of the cohort was 52.4 ± 14.4 years (mean \pm SD), of which 37 (74%) were males. The best corrected visual acuity in the affected eye ranged from Snellen 20/20 to light perception (mean Snellen = 20/600; mean logMAR visual acuity= 1.49 ± 1.00).

ED and SI measurements obtained by examiners at the slit-lamp and by semi-automated and manual segmentations of photographs is shown in Table 1. ED height measured by examiners at the slit-lamp was on average smaller than corresponding measurements taken from photographs (range over examiners: 2.6–2.7mm; manual measurement from photos: 3.3mm; range of semi-automated measurements from photos over 3 users: 2.9–3.2mm). Discrepancy in measurements was also observed for ED width, and SI width and height. Scatterplots displaying ED and SI measurement between examiners, between manual and semi-automated methods, and between repeated measurements from semi-automated segmentation show a strong positive linear relationship for all pairwise comparisons (Figure 4A–C). Correlation in measurement between pairs of examiners ranged from $\rho=0.78$ to $\rho=0.94$, between manual and semi-automated methods ranged from $\rho=0.93$ to $\rho=0.98$, and between separate users of the semi-automated methods ranged from $\rho=0.95$ to $\rho=0.99$ (all $p<0.0001$).

Variability in measurements of ulcer size between clinician-examiners

Median absolute differences in ulcer measurement between pairs of ophthalmologists were not significantly greater than 0.5mm. (Table 2) Ophthalmologists 1 and 2 had the least difference in measurements, with median difference of ED and SI dimensions all less than 0.5 mm (all $p < 0.05$). ED height measurements between pairs of examiners all had statistically significant differences under the threshold of 0.5mm. SI height showed similar results, albeit with some comparisons between examiners not achieving statistical significance. Alternatively, width measurements of ED and SI had median absolute differences between pairs of examiners that were mostly larger than those observed for height measurements and were not significantly lower than 0.5mm in 4 of 6 comparisons.

Although the median absolute difference of ED and SI measurements between examiner pairs was not found to be significantly greater than 0.5 mm, a non-trivial percentage of individual cases had a difference in measurement greater than 0.5 mm or 1.0 mm (Figure 5A). For ED height, examiner pairs differed by 0.5 mm in 24% to 38% of ulcers; for ED width, examiners differed by 0.5 mm in 30% to 52% of cases. Similarly, for SI height, pairs of examiners differed by 0.5 mm in 26% to 38% of ulcers; for SI width, pair of examiners differed by 0.5 mm in 38% to 58% of ulcers. The percentage of pairwise measurements that differed by 1.0 mm, ranged from 6% to 12% for ED height, 14% to 26% for ED width, 10% to 16% for SI height, and 6% to 30% for SI width.

Ulcer measurement between examiners showed good reliability (Figure 5). ICC for ED height was 0.87 (95% CI, 0.80 – 0.92), for ED width was 0.89 (95% CI, 0.83 – 0.93), for SI height was 0.88 (95% CI, 0.81 – 0.92), and for SI width was 0.84 (95% CI, 0.75 – 0.90).

Validation of semi-automated segmentation

Dice similarity coefficients were computed to compare the surface area overlap in segmentations obtained manually (gold-standard) to those by the semi-automated method (Figure 3). Comparing manual segmentation to the first user of the semi-automated segmentation method (A1), average Dice similarity coefficients were 0.84 (95% CI, 0.82 – 0.87) for ED surface area and 0.83 (95% CI, 0.81 – 0.86) for SI surface area. Semi-automated segmentation was repeated by two more users (A2, and A3). Average Dice similarity coefficients comparing manual to A2 were 0.86 (95% CI, 0.83 – 0.89) for ED surface area and 0.83 (95% CI, 0.81 – 0.85) for SI surface area. Comparing manual to A3 segmentation, average Dice coefficients were on average 0.83 (95% CI, 0.81 – 0.86) for ED surface area and 0.78 (95% CI, 0.75 – 0.81) for SI surface area.

Manual measurement of ulcer surface area showed good agreement with the semi-automated algorithm measurements (Figure 4B). Median absolute difference measurements of ED and SI dimensions between manual and semi-automated segmentation methods were not significantly greater than 0.5mm for all comparisons (range of medians 0.18 – 0.61, Table 2). The percentage of measurements that differed by 0.5 mm ranged from 20% to 38% for ED height, 22% to 32% for ED width, 24% to 54% for SI height, and 34% to 58% for SI width. The percentage of measurements that differed by 1.0 mm ranged from 2% to 10%

for ED height, 2% to 10% for ED width, 8% to 20% for SI height, and 10% to 22% for SI width (Figure 5B).

Variability in measurements of ulcer size with a semi-automated segmentation algorithm

Median absolute difference of ED and SI dimension between measurements with semi-automated segmentation from different users were significantly less than 0.5mm for all but one comparison (range of median absolute differences 0.13–0.46mm, Table 2). The percentage of measurements that differed by ≥ 0.5 mm ranged from 8% to 28% for ED height, 12% to 34% for ED width, 10% to 32% for SI height, and 14% to 34% for SI width. The percentage of measurements that differed by ≥ 1.0 mm ranged from 0% to 2% for ED height, 0% to 4% for ED width, 0% to 2% for SI height, and 0% to 8% for SI width (Figure 5C).

Ulcer measurement between repeated semi-automated segmentations by different users showed excellent reliability and was substantially better than that seen between examiners (Figure 6). ICC for ED height was 0.98 (95% CI, 0.97 – 0.99), for ED width was 0.98 (95% CI, 0.97 – 0.99), for SI height was 0.96 (95% CI, 0.94 – 0.98), and for SI width was 0.97 (95% CI, 0.96 – 0.98).

Discussion

In this study, we demonstrated small median differences in corneal ulcer measurements between ophthalmologists, albeit with some clinically meaningful differences. There was reduced variability in measurement when using computerized methods to measure corneal ulcers from photographs and process with image analysis. The usefulness of digital images and image analysis has been demonstrated in other disciplines of medicine and ophthalmology. In clinical settings, computerized imaging methods are used for automated classification of skin lesions,²⁸ grading of diabetic retinopathy,²⁹ and for monitoring progression of macular degeneration.³⁰ Automated tools can potentially increase precision (reducing measurement error), and images serve as permanent, standardized records of clinical findings for analysis. Ophthalmologists are exploring the use of automated imaging to extend care via imaging in telemedicine programs.^{31,32}

In our previous work, cornea specialists had good reliability when measuring epithelial defects in a controlled, artificial environment. However, cornea specialists differed in measurement length by ≥ 0.5 mm in a non-trivial percentage of cases (31% to 52%).¹² We anticipated similar or even greater inter-ophthalmologist measurement differences in a real world scenario, given the added complexity of patient positioning or movement. For both studies, we decided *a priori* that inter-examiner measurement differences ≥ 0.5 mm to be clinically significant, as a difference in measurement of this size could affect treatment decisions. This decision was based on the authors' clinical expertise and the hypothesis that measurement errors in the range of 5–10% (0.0585 – 1.17 mm for a 11.7 mm cornea horizontal diameter) are clinically meaningful. A well-designed clinical investigation looking at ulcers over time (to measure change) would be necessary to prove this tool's utility for a clinical setting. We found that, overall, measurements showed good reliability between ophthalmologists, with ICC ranging from 0.84 to 0.89. However, when comparing

measurements between pairs of ophthalmologist, between 24% to 52% of ED measurements differed by 0.5 mm, similar to the 31% to 52% reported in the controlled-environment study.¹² Variability in height measurements (both for ED and SI) was less than width measurements, perhaps because the slit beam is oriented vertically and ophthalmologists are more comfortable measuring pathology in the vertical direction.

Variability, even between high-volume, experienced ophthalmologists is not surprising. Inter-observer variability has been characterized in other studies in ophthalmology, including measurement of cup to disc ratio,³³ quantifying endothelial cell density by microscopy technicians,³⁴ measurement of intraocular pressure by Goldmann applanation between technicians,³⁵ and variability in interpretation of digital fundus image for diabetic retinopathy screening between ophthalmologists.³⁶ Ophthalmologists now use digital measurement tools to minimize measurement error and to provide quantified measures longitudinally for many diseases.

A computerized QCM algorithm and corresponding software package were developed to measure corneal ulcers from standardized imaging in the care of patients with corneal opacities and compare to the variability of measurements between clinicians. The proposed segmentation algorithm was semi-automatic because it required delineation of seed-points by the user to initialize segmentation. However, aside from providing seed-point, no manual correction was applied on the resulting segmentation. The hypothesis was that digital imaging and image analysis could reduce ED and SI measurement error. The QCM software showed good validity between manual (the imaging gold standard) and semi-automated segmentation (average Dice coefficients ranged from 0.78 to 0.86). Further, measurements of ED and SI dimensions obtained by the semi-automated algorithm used by different ophthalmologists had better reliability (higher ICC) than measures by 3 ophthalmologists' clinical gradings. An added feature is that semi-automated segmentation can acquire surface area measurements. Surface area may be a more meaningful metric to characterize an asymmetric, arbitrarily shaped corneal ulcer and will be the scope of future work.

A limitation of our study is the lack of individual corneal white-to-white diameter for each patient. Use of imaging software requires a measured length to correlate to pixel distance. Without available white-to-white measurements, we used an average horizontal white-to-white diameter of 11.7 mm for all patients.³⁷ As a result, comparing measurements between the ophthalmologists and the imaging methods has intrinsic error. We cannot say if differences are from differences in the methods (like the differences in corneal thickness measurement with optical coherence tomography vs. pachymetry) or intrinsic problems with the technique. This weakness will be remedied in future work.

Computerized image analysis requires good, consistent quality of the photographs. We acknowledge that lighting conditions, patient cooperation during image acquisition, and experience and comfort of the photographer can all have an impact on the quality of the photograph. We did not explicitly examine how photographs acquired under different light conditions or by different photographers would impact our semi-automated segmentation and measurement variability. However, we anticipate that since we did not make assumptions about background intensity and did not use predefined thresholds for

segmentation, small changes in imaging conditions will likely result in small perturbation in quantitative measurements. A analysis of the effects of the imaging condition on our quantitative segmentation will require a separate prospective study.

Future studies will refine QCM for corneal ulcers to improve a standardized approach to ulcer measurement. With advances in image analysis methods, future work will focus on creating a fully automated method for segmentation and co-registration of images between clinic encounters. Image-based measurement may even provide an opportunity for care coordination for patients with limited access to cornea specialists. This study highlights the extent of variability in measurements of “simple” height and width dimensions of ulcers, even between experienced ophthalmologists. By reducing human variability using automated tools, we hope to standardize and elevate the care of patients with potentially sight threatening corneal diseases.

Acknowledgments

We especially thank Drs. Naveen Radhakrishnan, Hitha Sara Sajeev, and Hem Shah who participated in measuring the eyes for corneal ulcer dimensions at Aravind and Dr. Paula Anne Newman-Casey who participated in performing semi-automated segmentation.

Disclosure of funding: This work was supported by a grant from the National Institutes of Health, Bethesda, MD (MAW; K23 Mentored Clinical Scientist Award K23EY023596) and the Inje University (KHK; Project No. 20150894). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Resnikoff S, Pascolini D, Etya'ale D, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ.* 2004; 82:844–851. [PubMed: 15640920]
2. Bourne RR, Stevens GA, White RA, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob Health.* 2013; 1:e339–349. [PubMed: 25104599]
3. Gonzales CA, Srinivasan M, Whitcher JP, et al. Incidence of corneal ulceration in Madurai district, South India. *Ophthalmic Epidemiol.* 1996; 3:159–166. [PubMed: 8956320]
4. Upadhyay MP, Karmacharya PC, Koirala S, et al. The Bhaktapur eye study: ocular trauma and antibiotic prophylaxis for the prevention of corneal ulceration in Nepal. *Br J Ophthalmol.* 2001; 85:388–392. [PubMed: 11264124]
5. Getshen K, Srinivasan M, Upadhyay MP, et al. Corneal ulceration in South East Asia. I: a model for the prevention of bacterial ulcers at the village level in rural Bhutan. *Br J Ophthalmol.* 2006; 90:276–278. [PubMed: 16488943]
6. Nemet AY, Nemet P, Cohn G, et al. Causes of blindness in rural Myanmar (Burma): Mount Popa Taung-Kalat Blindness Prevention Project. *Clin Ophthalmol.* 2009; 3:413–421. [PubMed: 19684864]
7. Collier SA, Gronostaj MP, MacGurn AK, et al. Estimated burden of keratitis—United States, 2010. *MMWR Morb Mortal Wkly Rep.* 2014; 63:1027–1030. [PubMed: 25393221]
8. Krachmer, JH., Mannis, MJ., Holland, EJ. *Cornea.* 2nd. Philadelphia: Elsevier/Mosby; 2005.
9. Kim RY, Cooper KL, Kelly LD. Predictive factors for response to medical therapy in bacterial ulcerative keratitis. *Graefes Arch Clin Exp Ophthalmol.* 1996; 234:731–738. [PubMed: 8986444]
10. Srinivasan M, Mascarenhas J, Rajaraman R, et al. Visual recovery in treated bacterial keratitis. *Ophthalmology.* 2014; 121:1310–1311. [PubMed: 24612976]
11. Fernandes M, Vira D, Dey M, et al. Comparison Between Polymicrobial and Fungal Keratitis: Clinical Features, Risk Factors, and Outcome. *Am J Ophthalmol.* 2015; 160:873–881 e872. [PubMed: 26210867]

12. Parikh PC, Valikodath NG, Estopinal CB, et al. Precision of Epithelial Defect Measurements. *Cornea*. 2017; 36:419–424. [PubMed: 28129296]
13. Farsi S, Chiu SJ, O'Connell RV, et al. Quantitative Classification of Eyes with and without Intermediate Age-related Macular Degeneration Using Optical Coherence Tomography. *Ophthalmology*. 2014; 121:162–172. [PubMed: 23993787]
14. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography: A pilot study. *Arch Ophthalmol*. 1995; 113:586–596. [PubMed: 7748128]
15. Allingham MJ, Mukherjee D, Lally EB, et al. A Quantitative Approach to Predict Differential Effects of Anti-VEGF Treatment on Diffuse and Focal Leakage in Patients with Diabetic Macular Edema: A Pilot Study. *Transl Vis Sci Technol*. 2017; 6:7–7.
16. Allingham MJ, Nie Q, Lad EM, et al. Semiautomatic segmentation of rim area focal hyperautofluorescence predicts progression of geographic atrophy due to dry age related macular degeneration. *Invest Ophthalmol Vis Sci*. 2016; 57:2283–2289. [PubMed: 27127926]
17. Lee AC, Qazi MA, Pepose JS. Biometry and intraocular lens power calculation. *Curr Opin Ophthalmol*. 2008; 19:13–17. [PubMed: 18090891]
18. Mukerji N, Vajpayee RB, Sharma N. Technique of area measurement of epithelial defects. *Cornea*. 2003; 22:549–551. [PubMed: 12883349]
19. VanRoekel RC, Bower KS, Burka JM, et al. Anterior segment measurements using digital photography: a simple technique. *Optom Vis Sci*. 2006; 83:391–395. [PubMed: 16772897]
20. Wilhelmus KR, Mitchell BM, Dawson CR, et al. Slitlamp biomicroscopy and photographic image analysis of herpes simplex virus stromal keratitis. *Arch Ophthalmol*. 2009; 127:161–166. [PubMed: 19204233]
21. Toutain-Kidd CM, Porco TC, Kidd EM, et al. Evaluation of fungal keratitis using a newly developed computer program, Optscore, for grading digital corneal photographs. *Ophthalmic Epidemiol*. 2014; 21:24–32. [PubMed: 24467559]
22. 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]
23. Yoo TS, Ackerman MJ, Lorensen WE, et al. Engineering and algorithm design for an image processing Api: a technical report on ITK—the Insight Toolkit. *Stud Health Technol Inform*. 2002; 85:586–592. [PubMed: 15458157]
24. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006; 31:1116–1128. [PubMed: 16545965]
25. Breiman L. Random Forests. *Machine Learning*. 2001; 45:5–32.
26. Criminisi, A., Shotton, J. *Decision forests for computer vision and medical image analysis*. London ; New York: Springer; 2013.
27. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004; 11:178–189. [PubMed: 14974593]
28. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542:115–118. [PubMed: 28117445]
29. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016; 316:2402–2410. [PubMed: 27898976]
30. Wintergerst MWM, Schultz T, Birtel J, et al. Algorithms for the Automated Analysis of Age-Related Macular Degeneration Biomarkers on Optical Coherence Tomography: A Systematic Review. *Transl Vis Sci Technol*. 2017; 6:10.
31. Ting DSW, Tan GSW. Telemedicine for Diabetic Retinopathy Screening. *JAMA Ophthalmol*. 2017; 135:722–723. [PubMed: 28520885]
32. Hark LA, Katz LJ, Myers JS, et al. Philadelphia Telemedicine Glaucoma Detection and Follow-up Study: Methods and Screening Results. *Am J Ophthalmol*. 2017
33. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc*. 1976; 74:532–572. [PubMed: 867638]

34. Rand GM, Kwon JW, Gore PK, et al. Technician Consistency in Specular Microscopy Measurements: A “Real-World” Retrospective Analysis of a United States Eye Bank. *Cornea*. 2017; 36:1172–1177. [PubMed: 28692574]
35. Dielemans I, Vingerling JR, Hofman A, et al. Reliability of intraocular pressure measurement with the Goldmann applanation tonometer in epidemiological studies. *Graefes Arch Clin Exp Ophthalmol*. 1994; 232:141–144. [PubMed: 8188062]
36. Ruamviboonsuk P, Teerasuwanajak K, Tiensuwan M, et al. Interobserver agreement in the interpretation of single-field digital fundus images for diabetic retinopathy screening. *Ophthalmology*. 2006; 113:826–832. [PubMed: 16650679]
37. Gharaee H, Abrishami M, Shafiee M, et al. White-to-white corneal diameter: normal values in healthy Iranian population obtained with the Orbscan II. *Int J Ophthalmol*. 2014; 7:309–312. [PubMed: 24790875]

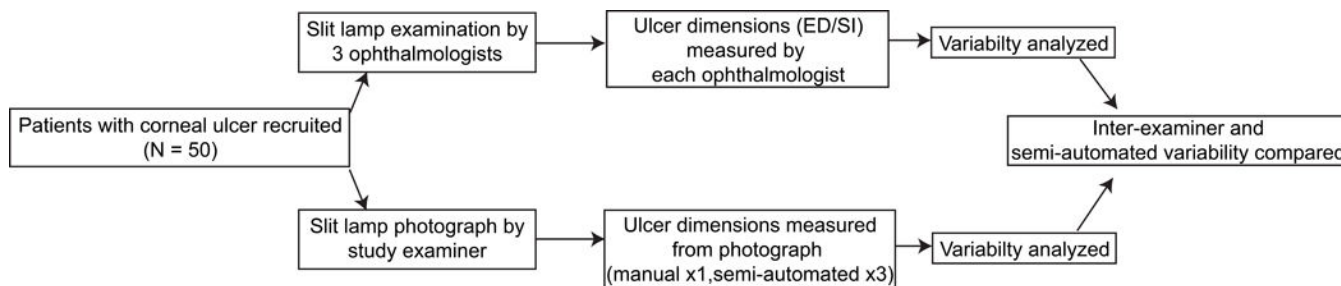


Figure 1. Study design: 50 patients with corneal ulcers were recruited. Three ophthalmologists measured the size of ED and SI at the slit-lamp. A cornea fellow took slit-lamp photographs of each ulcer. The image was then analyzed by quantitative corneal monitoring (QCM) software for the size of the ED and SI by manual and automated segmentation techniques. Measurements were analyzed for variability.

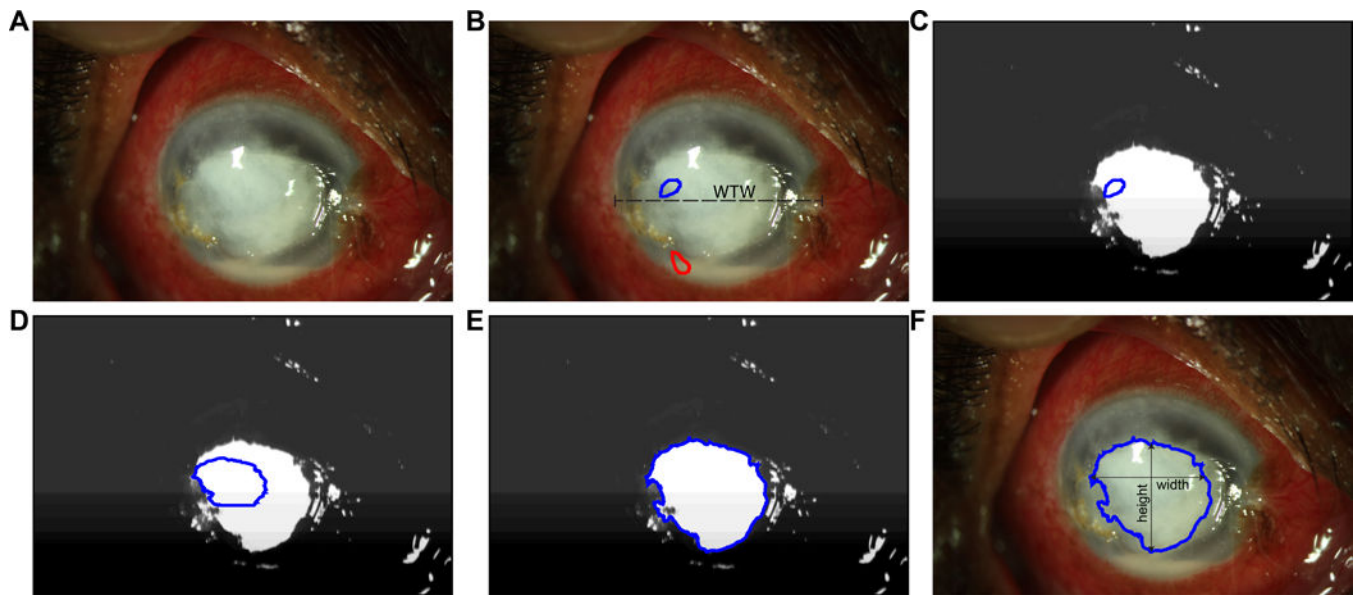


Figure 2.

QCM analysis pipeline: given a digital photograph of the corneal ulcer (A), the user draws a line to measure the horizontal white to white distance (WTW), and seed regions to denote the foreground (stromal infiltrate or epithelial defect, depicted in blue) and the background (clear cornea, depicted in red). A random forest tissue classifier generates a probability map of the foreground and background image (C). The probability map is used as a speed image for active contour evolution (D – E) and segmentation (F). The maximum vertical and horizontal distance is measured and recorded as the height and width, respectively.

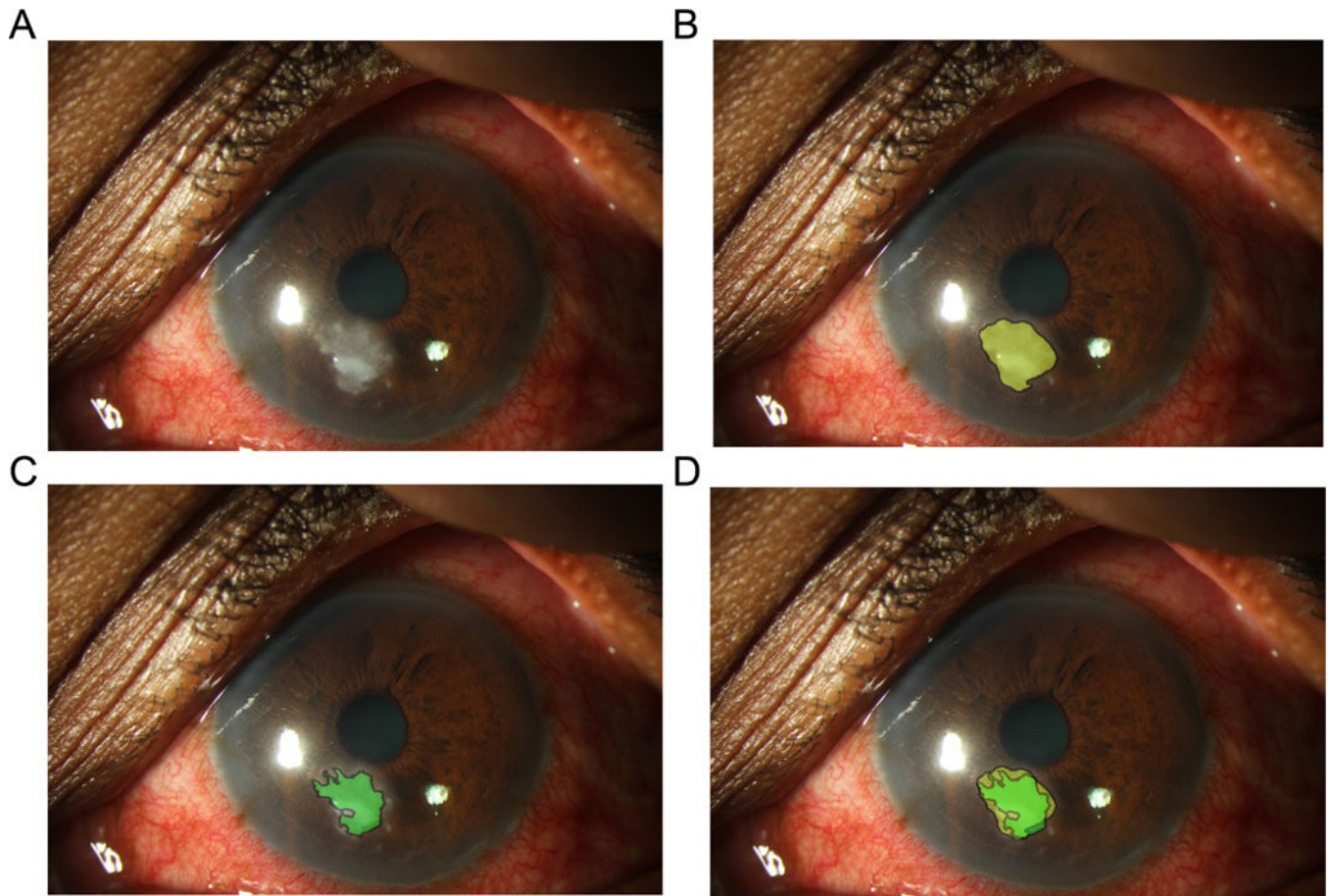


Figure 3.

Dice overlap coefficient between manual and semi-automated segmentations. Each corneal ulcer photograph (A) was both manually segmented (B, yellow) and semi-automatically segmented (C, green). The fractional spatial overlap between the manual and computerized segmentations (D) is the Dice overlap coefficient (equals 0.78 in this example).

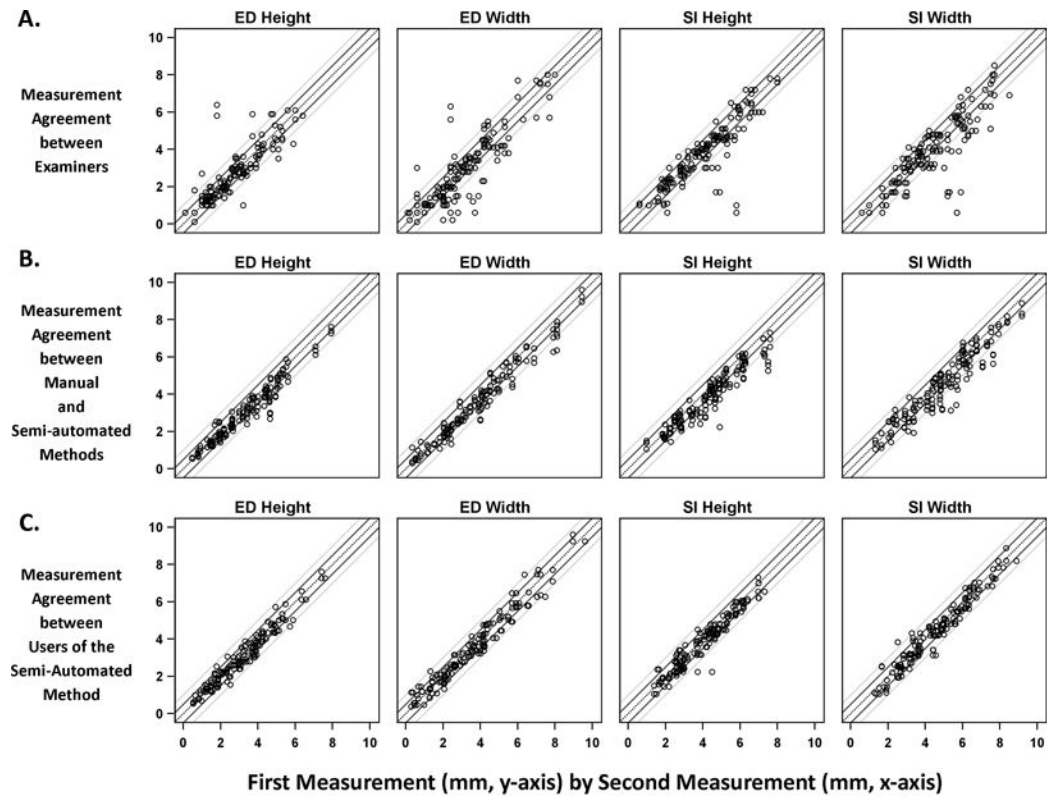


Figure 4.

Scatterplots displaying agreement in measurement of ED and SI dimensions, height and width, between a) pairs of examiners, b) manual and semi-automated segmentation methods, and c) repeated semi-automated segmentation. Note, all comparisons between pairs of examiners (E1 vs E2, E1 vs E3, E2 vs E3), manual and semi-automated segmentation (M vs A1, M vs A2, M vs A3), and repeated semi-automated methods (A1 vs A2, A1 vs A3, A2 vs A3) are represented with the same symbol on each corresponding scatterplot for ease of viewing. Reference lines for no difference (black dashed line) between two examiners and ± 0.5 mm (dark gray, solid lines) and ± 1.0 mm (light gray, solid lines) differences are displayed.

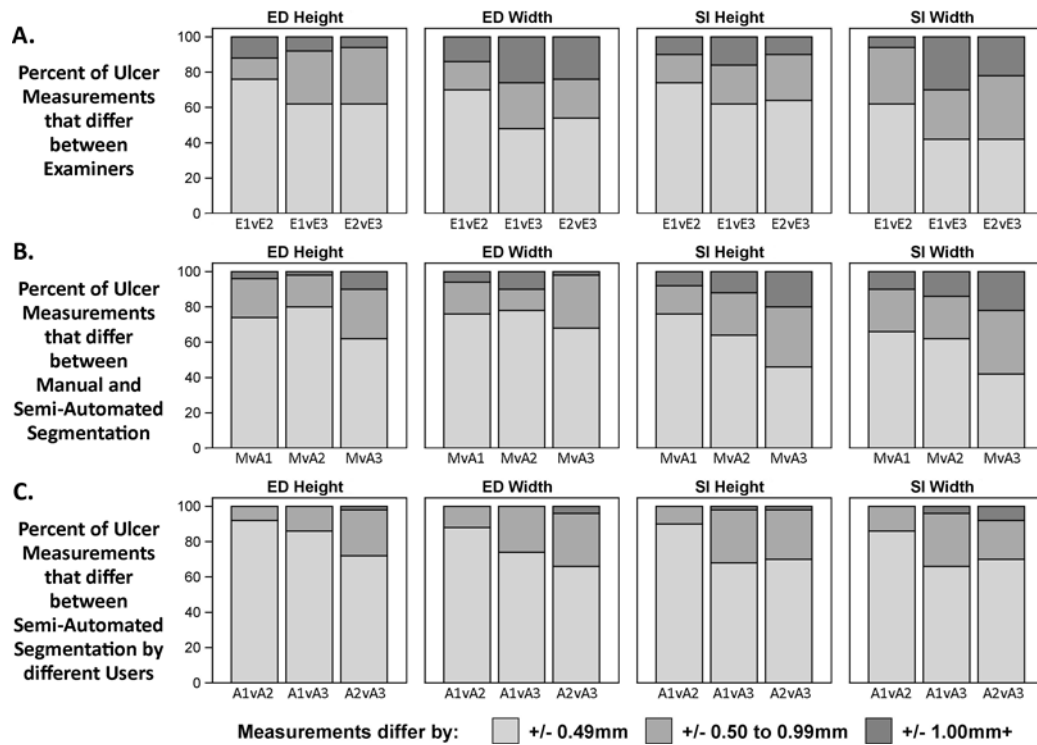


Figure 5. Histograms displaying differences in measurement of ED and SI dimensions, height and width, between a) pairs of examiners, b) manual and semi-automated segmentation methods, and c) repeated semi-automated segmentation by different users. E1-E3=Examiner1-Examiner3; M=Manual Segmentation; A1-A3=Semi-Automated Segmentation1-Semi-Automated Segmentation3; v=versus.

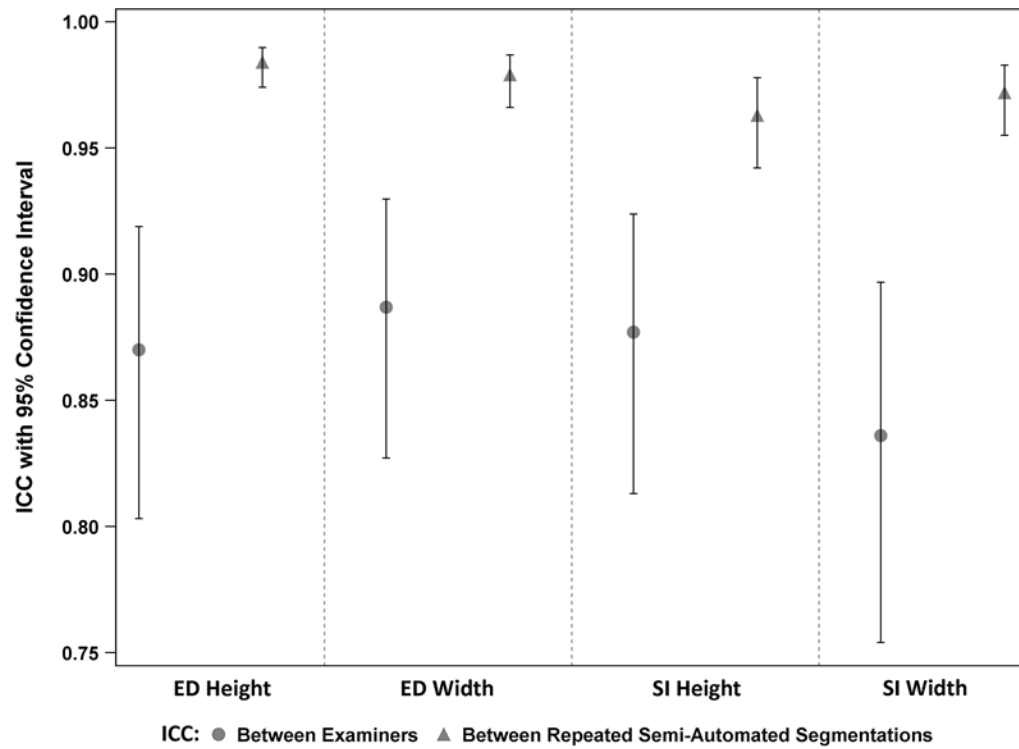


Figure 6. Forest plot displaying intraclass correlation coefficients (ICC) for reliability of measurements between examiners and between repeated semi-automated segmentations from photos. ED=epithelial defect; SI=stromal infiltrate.

Descriptive statistics of epithelial defect (ED) and stromal infiltrate (SI) measurements by examiners (abbreviated E1 to E3) and manual and semi-automated segmentations of photographs; n = 50 corneal ulcers.

Table 1

Measurer	ED Height			ED Width			SI Height			SI Width		
	Mean (SD)	Min, Max	Mean (SD)	Min, Max	Mean (SD)	Min, Max	Mean (SD)	Min, Max	Mean (SD)	Min, Max	Mean (SD)	Min, Max
E1	2.7 (1.4)	0.6, 6.0	3.1 (1.7)	0.2, 7.6	4.1 (1.6)	1.6, 8.0	4.5 (1.7)	1.7, 7.7				
E2	2.7 (1.6)	0.1, 6.4	3.2 (2.0)	0.1, 8.0	3.9 (1.7)	0.6, 7.6	4.3 (1.8)	0.6, 8.5				
E3	2.6 (1.4)	0.6, 6.1	2.8 (1.9)	0.2, 8.0	3.7 (1.7)	1.0, 7.8	3.7 (1.7)	0.6, 8.0				
Manual	3.3 (1.7)	0.5, 7.9	3.8 (2.1)	0.3, 9.4	4.3 (1.6)	1.0, 7.6	4.9 (1.9)	1.3, 9.2				
Automated 1	3.0 (1.5)	0.5, 7.4	3.5 (2.0)	0.3, 9.0	4.0 (1.5)	1.4, 7.0	4.6 (1.8)	1.5, 8.3				
Automated 2	3.2 (1.6)	0.5, 7.6	3.6 (2.0)	0.4, 9.6	3.9 (1.5)	1.5, 7.3	4.6 (1.8)	1.3, 8.9				
Automated 3	2.9 (1.4)	0.6, 7.3	3.4 (2.0)	0.5, 9.2	3.7 (1.4)	1.0, 6.5	4.3 (1.8)	1.1, 8.2				

SD=Standard Deviation, Min=Minimum, Max=Maximum

Table 2

Absolute difference in endothelial defect (ED) and stromal infiltrate (SI) measurements between examiners (abbreviated E1 to E3), manual segmentation (M), and semi-automated segmentations (A1–A3); n = 50 corneal ulcers

Measurer	Absolute Difference in ED Height (mm)			Absolute Difference in ED Width (mm)			Absolute Difference in SI Height (mm)			Absolute Difference in SI Width (mm)		
	Median (IQR)	P-value*	Median (IQR)	P-value*	Median (IQR)	P-value*	Median (IQR)	P-value*	Median (IQR)	P-value*	Median (IQR)	P-value*
E1 vs E2	0.25 (0.10–0.40)	0.002	0.25 (0.10–0.50)	0.034	0.30 (0.20–0.50)	0.007	0.35 (0.10–0.60)	0.048				
E1 vs E3	0.30 (0.10–0.60)	0.028	0.50 (0.20–1.00)	0.349	0.30 (0.10–0.70)	0.242	0.55 (0.20–1.20)	0.052				
E2 vs E3	0.30 (0.10–0.50)	0.001	0.40 (0.20–0.90)	0.391	0.30 (0.10–0.60)	0.012	0.60 (0.20–0.90)	0.112				
E1 vs M	0.60 (0.20–1.12)	0.078	0.43 (0.20–1.31)	0.383	0.40 (0.21–0.97)	0.466	0.81 (0.43–1.30)	<0.001				
E1 vs A1	0.52 (0.16–1.16)	0.080	0.38 (0.23–1.30)	0.132	0.33 (0.19–0.70)	0.578	0.61 (0.23–1.06)	0.080				
M vs A1	0.23 (0.07–0.51)	<0.001	0.27 (0.13–0.49)	0.001	0.32 (0.12–0.49)	0.019	0.36 (0.19–0.74)	0.323				
M vs A2	0.18 (0.08–0.45)	<0.001	0.19 (0.12–0.41)	0.002	0.36 (0.23–0.63)	0.087	0.36 (0.19–0.63)	0.282				
M vs A3	0.37 (0.17–0.66)	0.235	0.38 (0.18–0.64)	0.036	0.58 (0.21–0.91)	0.168	0.61 (0.29–0.93)	0.125				
A1 vs A2	0.13 (0.07–0.30)	<0.001	0.14 (0.06–0.29)	<0.001	0.17 (0.09–0.33)	<0.001	0.20 (0.10–0.32)	<0.001				
A1 vs A3	0.25 (0.07–0.44)	<0.001	0.31 (0.17–0.56)	<0.001	0.40 (0.22–0.56)	0.002	0.46 (0.21–0.60)	0.067				
A2 vs A3	0.34 (0.14–0.54)	0.001	0.37 (0.21–0.58)	0.017	0.25 (0.11–0.55)	<0.001	0.37 (0.16–0.51)	0.004				

IQR=Interquartile Range, mm=millimeter;

* Wilcoxon signed rank test for test of median absolute difference =0.5mm