Comparison of Smartphone Photography, Single-Lens Reflex Photography, and Field-Grading for Trachoma

John M. Nesemann,[1,2]* Michael I. Seider,[1,3] Blake M. Snyder,[1] Robi N. Maamari,[1,4] Daniel A. Fletcher,[4] Berhan A. Haile,[5] Zerihun Tadesse,[5] Nicole E. Varnado,[1] Sun Y. Cotter,[1] Elizabeth Kelly Callahan,[6] Paul M. Emerson,[6] Todd P. Margolis,[7] Thomas M. Lietman,[1,3,8,9] and Jeremy D. Keenan[1,3]

[1]*Francis I Proctor Foundation, University of California, San Francisco, San Francisco, California;* [2]*David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California;* [3]*Department of Ophthalmology, University of California, San Francisco, San Francisco, California;* [4]*Department of Bioengineering, University of California, Berkeley, Berkeley, California;* [5]*The Carter Center Ethiopia, Addis Ababa, Ethiopia;* [6]*The Carter Center, Atlanta, Georgia;* [7]*Department of Ophthalmology and Visual Sciences, Washington University School of Medicine in St. Louis, St. Louis, Missouri;* [8]*Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California;* [9]*Institute for Global Health Sciences, University of California, San Francisco, San Francisco, California*

Abstract. Conjunctival examination for trachomatous inflammation—follicular (TF) guides public health decisions for trachoma. Smartphone cameras may allow remote conjunctival grading, but previous studies have found low sensitivity. A random sample of 412 children aged 1–9 years received an in-person conjunctival examination and then had conjunctival photographs taken with 1) a single-lens reflex (SLR) camera and 2) a smartphone coupled to a 3D-printed magnifying attachment. Three masked graders assessed the conjunctival photographs for TF. Latent class analysis was used to determine the sensitivity and specificity of each grading method for TF. Single-lens reflex photo-grading was 95.0% sensitive and 93.6% specific, and smartphone photo-grading was 84.1% sensitive and 97.6% specific. The sensitivity of the smartphone-CellScope device was considerably higher than that of a previous study using the native smartphone camera, without attachment. Magnification of smartphone images with a simple attachment improved the grading sensitivity while maintaining high specificity in a region with hyperendemic trachoma.

INTRODUCTION

The WHO aims to eliminate trachoma as a public health problem.[1] The WHO guidelines suggest trachoma programs perform in-person conjunctival examinations to guide public health interventions and confirm elimination. Specifically, a population-based sample of children aged 1–9 years should be graded for trachomatous inflammation—follicular (TF) using the simplified WHO system, with a threshold of 5% used to identify areas requiring active interventions.[2]

In-person trachoma examinations have disadvantages. Current guidelines suggest graders achieve sufficient agreement with a reference-standard grader during in-person examinations, but the lack of trachoma in locations nearing elimination makes this difficult. Grading has variable reproducibility and cannot be audited.[3] Conjunctival photography with a single-lens reflex (SLR) camera overcomes some of these issues.[4] But using SLR cameras is complicated and expensive. Smartphone cameras would be cheaper and easier to use, but a 2012 study found a native smartphone camera less sensitive than an SLR, perhaps because of an inadequate macro lens.[5]

In this study, we evaluate an enhanced smartphone system with a novel attachment to magnify smartphone images. We assess the agreement of smartphone images with SLR images and test diagnostic accuracy relative to a latent class, hypothesizing that the smartphone attachment will improve smartphone sensitivity for TF.

METHODS

A random sample of 40 children aged 1–9 years from each of 13 Ethiopian communities was invited to participate in this sub-study of a previously reported trial; a total of 412 children ultimately participated.[6] Communities had received ≥ 4 annual rounds of mass azithromycin distribution. Each right superior tarsal conjunctiva was field-graded for TF and trachomatous inflammation—intense (TI) using the WHO's simplified grading system, and then had conjunctival photographs taken with a digital SLR camera and smartphone.[7]

A Nikon D-series camera and 105/2.8f macro lens were used for SLR imaging (aperture priority, f/40, ISO 400, native flash, and automatic white balance). An iPhone 4S coupled to a Corneal CellScope attachment was used for smartphone imaging (autofocus enabled and engaged by tapping the screen). Corneal CellScope components included a +25-diopter lens, two light-emitting diode light sources for external illumination, and a 3D-printed housing with a rotating piece that brought the image into focus when placed on the subject's orbital rim.[8,9] Although the CellScope did not change the resolution of the overall photograph, the magnification allowed more pixels per image to be devoted to the conjunctiva.

Images were graded for TF and TI by two photo-graders uninvolved in the field activities, masked to clinical information, camera type, and each other's grades. Discrepancies were adjudicated by a third masked grader. Field and photo-graders had to pass a minimum standard before being allowed to grade, defined as a Cohen's kappa ≥ 0.6 relative to the consensus TF and TI grades from a panel of three trachoma experts.

Images deemed ungradable by a consensus of photo-graders were excluded from analysis. Inter-method agreement was assessed with an intra-class correlation coefficient (ICC). Given the absence of a gold standard, the sensitivity, specificity, and predictive values of SLR photo-grading, smartphone photo-grading, and field-grading were estimated with latent class analysis. Accuracy metrics were computed relative to a reference latent class constructed from the observed data of the three methods (for both the TF and TI

* Address correspondence to John M. Nesemann, Francis I Proctor Foundation, 113 South Ferry Dr., Lake Mills, WI 53551. E-mail: jnesemann@mednet.ucla.edu

models, the two-class model fit [i.e., Bayesian information criterion] was better than the one-class model and $\chi^2 <$ 0.001).[10–12] Methods were compared by calculating the mean difference and its bootstrapped 95% CI. As a secondary analysis, the community-level TF and TI prevalence was calculated for each grading method. Statistical significance was determined with a McNemar test for individual-level data and the Wilcoxon signed-rank test for community-level data.[13] Analyses were performed with R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria).[14]

The study received ethical approval from the University of California, San Francisco, and the Ethiopian Ministry of Science and Technology. Caregivers provided verbal consent because of high illiteracy levels in the study area.

RESULTS

Of 412 children (218 boys and 194 girls; mean age 5.6 years and SD 3.6) photographed, two had ungradable smartphone images and were excluded from further analysis. Field-grading, SLR photo-grading, and smartphone photo-grading results are presented in Table 1 and Supplemental Figure 1.

Table 1 shows estimates of agreement between the different grading methods as well as estimates of diagnostic test accuracy (i.e., sensitivity, specificity, and positive and negative likelihood ratios). Agreement between smartphone and SLR grades was excellent for TF (ICC: 0.73, 95% CI: 0.62 to 0.84) and TI (ICC: 0.78, 95% CI: 0.60 to 0.96), and was higher than the agreement observed between either of the photography methods and field grades (Table 1). Sensitivity and specificity were relatively high for each of the methods: 84% and 98% for smartphone photography, 95% and 94% for SLR photography, and 90% and 88% for in-field grading, respectively.

Table 2 summarizes the differences in prevalence and diagnostic test accuracy. For TF, SLR photo-grading was 5.4%

(95% CI: –2.3% to 10.4%, $P = 0.157$) more sensitive and 5.4% (95% CI: –3.0% to 13.0%, $P = 0.058$) more specific than field-grading. Smartphone photo-grading was 4.0% more specific than SLR photo-grading (95% CI: 7.8% to 0.6%, $P = 0.022$) but 10.9% less sensitive (95% CI: –5.8% to –20.8%, $P = 0.004$). Analogous information for TI is shown in Table 2; both cameras were significantly more sensitive than field-grading for TI, but not significantly different in specificity.

A community-level prevalence of TF and TI was calculated for each of the 13 communities in the study. Overall, smartphone grading underestimated the prevalence of TF relative to field-grading by a mean of 8.8% (95% CI: 5.4% to 16.4%, $P = 0.014$) and to SLR grading by a mean of 5.7% (95% CI: 4.1% to 11.3%, $P = 0.009$). Single-lens reflex–based TF prevalence was similar to field-grading (SLR prevalence on average 3.1% lower, 95% CI: –8.5% to 1.9%, $P = 0.367$). Figure 1 compares community-level TF and TI prevalence estimates for photo grades versus field grades. All 13 communities had a TF prevalence of $\geq$ 5% by all three methods, demonstrating agreement that no community reached the WHO elimination target. Ocular chlamydia was detected in only three communities, and all three grading methods estimated a high TF prevalence (e.g., above 25%) in each of these communities (Figure 1). By contrast, the TI prevalence estimates of the two photography methods were consistently higher than those of the field-grade method in the three chlamydia-positive communities (e.g., > 10%; Figure 1).

DISCUSSION

Single-lens reflex photography was highly sensitive and specific for assessing clinically active trachoma in a hyperendemic region of Ethiopia. Photo grades from an iPhone 4S/CellScope exhibited high agreement with SLR photo grades although they were slightly less sensitive than SLR for grading TF.

TABLE 1
Agreement and diagnostic test accuracy of smartphone photography, SLR photography, and field-grading

Metric*	TF	TI
Prevalence		
Smartphone	125/410 (30.5%)	51/410 (12.4%)
SLR	151/410 (36.8%)	60/410 (14.6%)
Field grades	158/410 (38.5%)	35/410 (8.5%)
Intra-class correlation coefficient (95% CI)		
Smartphone vs. SLR	0.73 (0.62–0.84)	0.78 (0.60–0.96)
Smartphone vs. field-grading	0.63 (0.52–0.74)	0.51 (0.29–0.73)
SLR vs. field-grading	0.66 (0.56–0.77)	0.50 (0.29–0.69)
Specificity (95% CI)		
Smartphone	97.6% (89.0–99.3%)	98.3% (92.2–99.5%)
SLR	93.6% (88.0–96.6%)	96.7% (91.3–98.5%)
Field grades	88.2% (83.0–91.7%)	97.4% (94.3–98.6%)
Sensitivity (95% CI)		
Smartphone	84.1% (74.7–90.0%)	93.4% (5.6–99.8%)
SLR	95.0% (75.2–98.8%)	99.9% (99.7–100.0%)
Field grades	89.7% (80.8–94.2%)	53.6% (38.3–67.9%)
Positive likelihood ratio (95% CI)		
Smartphone	38.2 (23.8–188.1)	57.9 (39.9–267.7)
SLR	15.2 (11.1–97.3)	27.9 (22.0–99.9)
Field grades	8.2 (6.1–16.7)	22.3 (15.8–239.8)
Negative likelihood ratio (95% CI)		
Smartphone	0.15 (0.11–0.24)	0.043 (0.00–0.38)
SLR	0.05 (0.02–0.07)	0.001 (0.00–0.06)
Field grades	0.10 (0.02–0.15)	0.458 (0.37–0.69)

TF = trachomatous inflammation—follicular; TI = trachomatous inflammation—intense; SLR = single-lens reflex.
* When calculated, bootstrap CIs were resampled by field-grader identifier to account for potential intra-grader correlation because field-graders were not masked to community; 9,999 replications.

TABLE 2
Mean differences in trachoma prevalence, sensitivity, and specificity between three different grading methods

	TF		TI	
Comparison	Difference (95% CI)	*P*-value	Difference (95% CI)	*P*-value
Prevalence				
SLR minus field	−3.1% (−8.5% to 1.9%)	0.367	−5.6% (−9.9% to −2.0%)	0.018
Field minus smartphone	8.8% (5.4% to 16.4%)	0.014	−3.3% (−6.8% to 0.5%)	0.155
SLR minus smartphone	5.7% (4.1% to 11.3%)	0.009	2.3% (0.9% to 4.3%)	0.014
Sensitivity				
SLR minus field	5.4% (−2.3% to 10.4%)	0.157	46.4% (37.0% to 67.8%)	< 0.001
Smartphone minus field	−5.5% (−19.7% to 1.8%)	0.117	39.8% (23.5% to 54.5%)	< 0.001
Smartphone minus SLR	−10.9% (−20.8% to −5.8%)	0.004	−6.6% (−30.0% to 0.0%)	0.157
Specificity				
SLR minus field	5.4% (−3.0% to 13.0%)	0.058	−0.8% (−5.1% to 2.2%)	0.394
Smartphone minus field	9.4% (4.0% to 16.1%)	< 0.001	0.8% (−1.0% to 2.8%)	0.439
Smartphone minus SLR	4.0% (0.6% to 7.8%)	0.022	1.6% (−1.4% to 4.6%)	0.108

TF = trachomatous inflammation—follicular; TI = trachomatous inflammation—intense; SLR = single-lens reflex.

A previous study of an iPhone 4 without external modification demonstrated a sensitivity of 41% for the detection of TF.[5] The present study demonstrated that adding external magnification greatly improved the sensitivity of smartphone TF
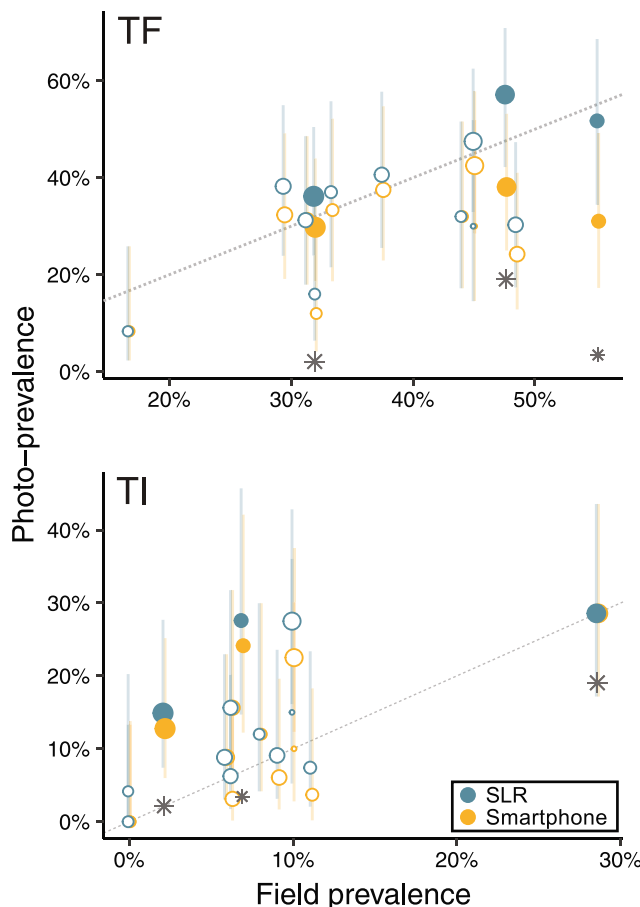

FIGURE 1. Community-level agreement of three methods of clinical trachoma grading. The scatterplot depicts trachoma prevalence from single-lens reflex (SLR) photography (blue) and smartphone photography (red) across the range of prevalence estimates observed from field-grading in each of 13 communities. Ocular chlamydia was detected in only three communities (solid SLR and smartphone markers; chlamydia prevalence depicted as gray stars). The lines represent 95% CI of the prevalence in a community. Points above the dotted identity line represent overestimates relative to field-grading, and those below, underestimates. This figure appears in color at www.ajtmh.org.

grading, to 84.1%, while still maintaining an acceptably high specificity of 97.6%. Agreement between smartphone and SLR grading was high, but smartphone photo-grading was slightly less sensitive, suggesting design modifications or smartphone camera improvements may be necessary to reach a sensitivity equivalent to SLR photography. Any changes must minimize reductions in specificity, since verification of trachoma elimination will require a highly specific test to minimize false positives in low-prevalence areas.

The smartphone's lower sensitivity resulted in systematic underestimation of community-level TF prevalence. Although statistically significant, this underestimation would not have changed programmatic activities because TF prevalence estimates were above 5% in all 13 communities, regardless of grading method. The community-level implications could be different in a hypo- or meso-endemic area, where an underestimate might lead to cessation of mass antibiotic distributions. However, there is no perfect gold standard, and it is possible that photo-grading might be more accurate than field-grading.[10] Evidence for this possibility was found in the TI prevalence assessment because photo-grading appeared to be a more reliable indicator of community-level chlamydial infection than field-grading.

Several limitations should be noted. Photo-graders differed from the field-graders, which could have introduced bias in comparisons of field versus photo grades. The study area had hyperendemic trachoma and had received multiple mass antibiotic distributions. Findings might differ in other settings. The aim of the study was to test smartphone photography in general, but this was implemented by testing a specific, older-version iPhone coupled to a specific external attachment. Although the generalizability of the study to newer devices is unknown, the iPhone 4S had an eight megapixel camera with $f/2.4$ aperture and 4.28 mm focal length—specifications not so different from more recent devices—and in our experience has provided image quality superior to newer budget smartphones. Moreover, the CellScope design has not changed since this study was performed, and the 3D-printed housing can easily be adapted to fit other cellular devices. Thus, the findings in this study should be relevant to newer smartphones. At the very least, this study provides a lower benchmark of smartphone performance that newer smartphones may exceed.

In conclusion, magnification of smartphone images with a simple attachment improved the sensitivity for TF while maintaining high specificity in a region with hyperendemic trachoma. Further research is needed to determine the utility

of smartphone photo-grading in areas with less prevalent infection.

Authors' addresses: John M. Nesemann, Geffen School of Medicine (Medical Student), University of California, Los Angeles, Los Angeles, CA, E-mail: jnesemann@mednet.ucla.edu. Michael I. Seider, Blake M. Snyder, Robi N. Maamari, Nicole E. Varnado, Sun Y. Cotter, Thomas M. Lietman, and Jeremy D. Keenan, Francis I Proctor Foundation, University of California, San Francisco, San Francisco, CA, E-mails: michael.i.seider@kp.org, blake.snyder@cuanschutz.edu, robimaamari@gmail.com, nicolestoller@gmail.com, sun.cotter@ucsf.edu, tom.lietman@ucsf.edu, and jeremy.keenan@ucsf.edu. Daniel A. Fletcher, Department of Bioengineering, University of California, Berkeley, Berkeley, CA, E-mail: fletch@berkeley.edu. Berhan A. Haile and Zerihun Tadesse, The Carter Center Ethiopia, Addis Ababa, Ethiopia, E-mails: berhanayele@burnet.edu.au and zerihun.tadesse@cartercenter.org. Elizabeth Kelly Callahan and Paul M. Emerson, The Carter Center, Atlanta, GA, E-mails: kelly.callahan@cartercenter.org and pemerson@taskforce.org. Todd P. Margolis, Department of Ophthalmology and Visual Sciences, Washington University School of Medicine in St. Louis, St. Louis, MO, E-mail: margolist@wustl.edu.

REFERENCES

1. WHO, 1998. *World Health Assembly Resolution WHA 51.11*. Geneva, Switzerland: World Health Organization.

2. WHO, 2006. *Trachoma Control: A Guide for Programme Managers*. Geneva, Switzerland: World Health Organization.

3. Miller K et al., 2004. How reliable is the clinical exam in detecting ocular chlamydial infection? *Ophthalmic Epidemiol 11:* 255–262.

4. Stare D, Harding-Esch E, Munoz B, Bailey R, Mabey D, Holland M, Gaydos C, West S, 2011. Design and baseline data of a randomized trial to evaluate coverage and frequency of mass treatment with azithromycin: the Partnership for Rapid Elimination of Trachoma (PRET) in Tanzania and The Gambia. *Ophthalmic Epidemiol 18:* 20–29.

5. Bhosai SJ, Amza A, Beido N, Bailey RL, Keenan JD, Gaynor BD, Lietman TM, 2012. Application of smartphone cameras for detecting clinically active trachoma. *Br J Ophthalmol 96:* 1350–1351.

6. Keenan JD et al., 2018. Mass azithromycin distribution for hyperendemic trachoma following a cluster-randomized trial: a continuation study of randomly reassigned subclusters (TANA II). *PLoS Med 15:* e1002633.

7. Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR, 1987. A simple system for the assessment of trachoma and its complications. *Bull World Health Organ 65:* 477–483.

8. Snyder BM, Sie A, Tapsoba C, Dah C, Ouermi L, Zakane SA, Keenan JD, Oldenburg CE, 2019. Smartphone photography as a possible method of post-validation trachoma surveillance in resource-limited settings. *Int Health 11:* 613–615.

9. Maamari RN, Ausayakhun S, Margolis TP, Fletcher DA, Keenan JD, 2014. Novel telemedicine device for diagnosis of corneal abrasions and ulcers in resource-poor settings. *JAMA Ophthalmol 132:* 894–895.

10. See CW, Alemayehu W, Melese M, Zhou Z, Porco TC, Shiboski S, Gaynor BD, Eng J, Keenan JD, Lietman TM, 2011. How reliable are tests for trachoma? A latent class approach. *Invest Ophthalmol Vis Sci 52:* 6133–6137.

11. Linzer DA, Lewis JB, 2011. poLCA: an R package for polytomous variable latent class analysis. *J Stat Softw 42:* 1–29.

12. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KG, de Groot JA, 2014. Latent class models in diagnostic studies when there is no reference standard–a systematic review. *Am J Epidemiol 179:* 423–431.

13. Stock C, Hielscher T, 2014. *DTComPair: Comparison of Binary Diagnostic Tests in a Paired Study Design*. Vienna, Austria: Comprehensive R Archive Network.

14. R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.