

The Impact of Image Quality and Trachomatous Inflammation on Using Photography for Trachoma Prevalence Surveys

Michelle Odonkor¹, Fahd Naufal¹, Harran Mkocho², Nicodemus Funga², Beatriz Muñoz¹, and Sheila K. West¹

¹ Dana Center for Preventive Ophthalmology, Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA

² Kongwa Trachoma Project, Kongwa, Tanzania

Correspondence: Sheila West, Wilmer Room 155, Johns Hopkins Hospital, 1800 Orleans Street, Baltimore, MD 21287, USA. e-mail: shwest@jhmi.edu

Received: September 24, 2021

Accepted: February 21, 2022

Published: March 10, 2022

Keywords: trachoma; prevalence surveys; photography; quality; follicular trachoma

Citation: Odonkor M, Naufal F, Mkocho H, Funga N, Muñoz B, West SK. The impact of image quality and trachomatous inflammation on using photography for trachoma prevalence surveys. *Transl Vis Sci Technol.* 2022;11(3):11, <https://doi.org/10.1167/tvst.11.3.11>

Purpose: Graded images can be used for trachoma prevalence surveys, but there is concern for mismatch between image and field grades of the upper tarsal conjunctiva. We aimed to determine if poor photograph quality and/or inflammation may contribute to differential grading of trachomatous inflammation—follicular (TF) between field and photograph graders.

Methods: We developed a simplified and expanded image quality grading tool. Agreement was assessed using kappa statistic. We included 5417 eyes with both field and image grades for TF. Eyes where the field and adjudicated photograph TF grades did not match were identified (mismatched) and assigned an image quality (IQ) score and a potential mismatch reason. We also assigned IQ scores to a stratified random sample of 60 eyes with matching field and photograph TF grades (matched).

Results: There were 5240 eyes that had matching grades, whereas 177 eyes (3.3%) were mismatched. Overall quality was high, even in mismatched eyes. There was no difference in overall or specific IQ metrics between eyes with matching grades and eyes with mismatched grades ($P = 0.59$). Mismatched eyes had worse inflammation compared to matched eyes ($P = 0.048$). The primary reason for calling TF in the field but not in the photographs appeared to be the number of follicles observed.

Conclusions: Image quality did not explain mismatch between field grades and image grades from this prevalence survey. Inflammation made mismatch more likely.

Translational Relevance: Our quality grading scheme rapidly identifies image quality issues for training. Standardizing TF grading in the presence of inflammation will improve field and photograph grading.

Introduction

Trachoma, a chronic conjunctivitis caused by repeated episodes of infection with *Chlamydia trachomatis*, is a major public health concern, particularly in developing countries.¹ The global effort to eliminate trachoma has had considerable success with several countries having declared elimination following World Health Organization guidelines.¹ One of the two criteria for elimination is a prevalence of trachomatous inflammation—follicular (TF) of less

than 5% in children ages 1 to 9 years, as determined by 2 district-level population-based surveys conducted at least 2 years apart, in the absence of mass drug administration (MDA).²

For these surveys, countries typically rely on trained field graders who have been standardized using live cases.³ As trachoma prevalence declines, training and re-certifying field graders becomes increasingly more difficult and expensive, in some cases, necessitating travel to endemic countries. This problem will become more acute as more countries undertake surveys seeking to validate the elimination of trachoma. One

solution is to consider the use of photography and grading images for trachoma as a replacement for field grading. Clinical trials and research using large-scale prevalence surveys have long used this approach. Image capture and masked grading avoids bias that can result from knowledge of the TF status of the patient's other eye, family members, endemicity in the village, prior MDA status, and other characteristics that may potentially affect live grading.

However, there are a number of issues raised by the use of photography.⁴ One is concern for the number of ungradable images that can occur, which in some surveys was reported as quite large.⁵ As in other ophthalmic uses of photography, this problem can be mitigated by training, certification, and supervision of photographers. Second, there is no standardized approach to grading the quality of images, a key metric that would be beneficial for use in training photographers and providing them feedback on their performance. Third, there is concern for disagreements between field graders and image graders over the presence or absence of TF in the same upper tarsal conjunctiva. Studies of the agreement often use field grading as the "gold standard" despite evidence that field graders are also prone to over- or under-calling TF.⁶

The purposes of this study were twofold: (i) to develop a simple, standardized scheme for grading the quality of images of the upper tarsal conjunctiva, and (ii) to examine the possible reasons for mismatches between field grades and image grades using an expanded grading scheme for the assessment of inflammation and an expanded quality grading scheme that outlined detailed criteria for assessing image quality. We hypothesized that differential grading of TF between field and image graders was due to two factors: poor quality of images and the presence of inflammation.

Methods

Design of Image Quality Grading Scheme

Designing the quality grading metrics was an iterative process that consisted of identifying the various independent factors affecting quality and determining the size of their impact on the quality of the image. The specific factors included focus, lighting and shadows, glare and tear film, degree of eyelid eversion, eyelid blanching, and debris and obstruction in the field of interest (Table 1). These factors affecting image quality could all be addressed by a photographer with minimal effort, and quality issues could generally be resolved

Table 1. Definition of Factors Affecting Image Quality

Quality Factor	Definition
<i>Focus</i>	Blurring present due to the focal point of the photograph being situated in front of or behind the plane of the tarsal conjunctiva. Out-of-focus images may occur when the lid is everted poorly with several curvilinear features or when the autofocus feature (or the photographer) concentrates on the eyelashes instead of the tarsal plate. (Fig. 1A).
<i>Lighting and shadows</i>	Suboptimal lighting can result from overexposure of the tarsal plate, presence of shadows, or lack of adequate lighting.
<i>Glare and tear film</i>	Regions of increased reflectivity in the grading area caused by increased tearing or moisture. Gentle removal of tears prior to taking the image should resolve this problem.
<i>Eyelid eversion</i>	Inadequate eversion of the eyelid results in inability to grade the entire eyelid. Whereas less of a problem if the visible part of the eyelid contains 5 follicles of the correct size, it is difficult to be certain that TF is not present if there are fewer than 5 or no follicles on the visible part of the eyelid (Fig. 1B).
<i>Blanching of the eyelid</i>	Blockage of blood supply to the conjunctiva results from prolonged pressure from eversion, which turns part or most of the eyelid pale, making follicles difficult to discern (Fig. 1C).
<i>Debris and obstruction</i>	Inadequate clearance of debris (e.g. eyelashes, mucus, and foreign bodies) or obstructions (e.g. eyelid flipper's finger) that block the grading area. These do not include eyelid pathologies that cannot be cleared or removed by the photographer.

by retaking the photograph. For instance, we included factors, such as glare or tear film, which could be wiped away, but excluded eyelid pathologies like tumors, nevi, discolorations, injuries, surgical scars, inflammation, and scarring, as they could not be cleared or



Figure 1. Images illustrating quality issues that were used as reference standards. (A) Sample tarsal conjunctiva with a moderate focus issue (grade = 2) and insignificant glare spots. (B) Sample tarsal conjunctiva with a moderate eversion issue (grade = 2). (C) Sample tarsal conjunctiva with a moderate blanching issue (grade = 2).

resolved by a photographer and would be encountered by and would be similarly challenging for field graders.

The grading metrics took all modifiable quality factors impacting the grading area (area of the tarsal conjunctiva, as described by Solomon et al.) into account.³ Focus generally affects the entire photograph and grading area and therefore took

priority in the metrics. Focus issues were categorized as none/mild, moderate, or severe. Moderate focus issues were determined by comparison to a reference image (Fig. 1A). A photograph with substantially worse focus was then graded as having a severe focus issue, whereas a photograph with better focus was graded as having no or a mild focus issue. Following categorization of image focus, the combined grading area affected by the remaining quality factors was assessed. Adequate eversion of the eyelid assumed maximum possible visibility of the conjunctival grading area. Complete eversion was achieved if the vertically oriented upper deep tarsal vessels covered approximately the top third of the eyelid and the lower vessels covered approximately the bottom two-thirds of the eyelid.

We developed a simple image quality grading scheme that separated good-quality images from moderate- or poor-quality images based on focus and how much of the grading area was impacted by the other quality factors (Table 2).

We also developed an expanded grading scheme that allowed us to determine the impact of each of the quality metrics on the problem of disagreement between field grades and image grades (Table 3). For each quality issue, we have three different levels of concern — significant, moderate, and no/mild issue.

Expansion of World Health Organization Grading of Trachomatous Inflammation—Severe

For assessment of the impact of inflammation on disagreements between image and field graders, we expanded the World Health Organization (WHO) simplified grading scheme³ to allow for no to mild inflammation to be defined as <33% of vessels obscured by inflammation, moderate inflammation as 33 to <50% of vessels obscured by inflammation, and severe inflammation as 50% or more vessels obscured. Our criterion for severe inflammation is the WHO definition of trachomatous inflammation—severe (TI; Table 4).

Image Quality and Inflammation Agreement Analysis

To test inter-grader agreement, a set of 75 images, including a range of levels of quality, was created by a grader who was not part of the agreement analysis. Two graders, who had been trained to grade quality

Table 2. Simplified Image Quality Metrics

Quality Grade	Definition and Criteria
<i>Good quality</i>	No to mild focus issue (grade = 1) AND <25% of the grading area impacted by a combination of poor lighting and shadows, glare and tear film, incomplete eversion, blanching, and debris and obstruction.
<i>Moderate quality</i>	1) Moderate focus issue (grade = 2), OR 2) 25–67% of the grading area impacted by a combination of poor lighting and shadows, glare and tear film, incomplete eversion, blanching, and debris and obstruction.
<i>Poor quality</i>	1) Severe focus issue (grade = 3), OR 2) >67% of the grading area impacted by a combination of poor lighting and shadows, glare and tear film, incomplete eversion, blanching, and debris and obstruction, OR 3) Moderate focus issue (grade = 2) AND >50% of the grading area impacted by a combination of poor lighting and shadows, glare and tear film, incomplete eversion, blanching, and debris and obstruction.

Rules: • To categorize focus, [Figure 1](#) should be used as reference for a moderate focus issue (grade = 2), and anything worse should be classified as a severe focus issue (grade = 3).
• The amount of eversion is estimated based on comparison to a fully everted eyelid where upper deep tarsal vessels cover the top third and lower vessels cover the bottom two-thirds of the eyelid.

on a different set of images, used our simplified quality grading scheme to independently assign quality grades to this test set. Agreement in assessment of image quality was measured using a weighted Cohen's kappa statistic where weights of 0.5 were assigned to adjacent cells.

For determining agreement in assessment of TI, a set of 74 images that included a mix of inflammation levels was created from a library of images maintained at the Dana Center for Preventive Ophthalmology from prior population-based surveys. None of the images

from the training set used for training image graders was part of the test set. Two trained graders independently assigned inflammation grades to the test set, and agreement was measured using Cohen's kappa statistic.

Survey Used to Obtain Matched and Mismatched Grades From Image and Field Graders

The images graded in this analysis were obtained from a survey of 3118 children ages 1 to 9 years old from a formerly trachoma hyperendemic district (Kongwa, Tanzania), 2 years after the cessation of MDA. For 404 children (808 eyes), a camera was not available to use in the field, and the rest had images taken of the upper tarsal plate using a handheld Nikon D40 digital SLR camera in manual setting, with a 105 mm f/2.8D Auto Focus Micro Nikkor lens (fully extended). Of the 2714 children (5428 eyes), 11 images were ungradable according to our simplified grading criteria. Thus, our analysis is based on the 5417 eyes with gradable images.

A field grader, trained and standardized in the use of the WHO simplified grading scheme,³ assigned field grades for the presence or absence of TF and was also the photographer who obtained tarsal conjunctival photographs of both eyes in each child. The photographer was originally trained in early 2000 for clinical trials that had masked assessment of trachoma as an end point. Training involved on site evaluation of use of the SLR digital camera, and the acquisition of gradable images from both eyes in at least five children. Each year, there is refresher review of image capture practice. Photographs of the tarsal conjunctiva are taken with manual focus under natural light conditions while shielded from direct sunlight. The entirety of the conjunctiva must be in frame while the thumb used for eyelid eversion must not obstruct the grading area. Care is taken to wipe away tears or excess moisture prior to image capture and no flash or artificial source of light is used to prevent glare. The camera lens is angled to be directly perpendicular to the conjunctiva to reduce shadows and glare. All images taken during the survey were sent to the Dana Center where two standardized trachoma image graders independently graded each image. Any disagreements were openly adjudicated with a senior grader.

We identified the eyes where the field and adjudicated photograph TF grades matched and took a stratified random sample of 60 eyes, with strata based on presence or absence of TF ("match" eyes). We also identified all eyes where the field and photograph

Table 3. Expanded Image Quality Grading Scheme

	Image Quality Criteria					
	Lighting and Shadows	Focus	Glare and Tear Film	Eyelid Eversion	Blanching	Debris and Obstruction
Significant issue (grade = 3)	Poor lighting, or shadows covering >50% of grading area	Grading area significantly out of focus	Glare covering >33% of grading area	Eyelid incompletely everted – missing >33% of grading area	>33% of grading area is blanched	>33% obstruction of grading area
Moderate issue (grade = 2)	Moderate overall lighting, or shadows covering 33% to 50% of grading area	Grading area moderately out of focus	Glare covering 10% to 33% of grading area	Eyelid partially everted – missing 10% to 33% of grading area	10% to 33% of grading area is blanched	10% to 33% obstruction of grading area
No/mild issue (grade = 1)	Good overall lighting and shadows covering <33% of grading area	Grading area in focus or mildly out of focus	Glare covering <10% of grading area	Eyelid mostly everted – missing <10% of grading area	<10% of grading area is blanched	<10% obstruction of grading area

Table 4. Criteria for Inflammation Grade

Significant inflammation (Grade = 3)	>50% of deep tarsal vessels obscured
Moderate inflammation (grade = 2)	33% to 50% of deep tarsal vessels obscured
No/mild inflammation (grade = 1)	<33% of deep tarsal vessels obscured

grades did not match (“mismatch” eyes, $n = 177$). All eyes in both samples were assigned a quality score for each metric of quality, using the expanded system. In addition, the image graders assigned a grade of inflammation to the image. Finally, all mismatched eyes were re-reviewed by the two image graders who, without knowledge of the direction of the mismatch, assigned a potential reason for the mismatch just based on review of the photographs. Potential reasons for mismatch included different interpretation of follicle number (<5 vs. 5 or more), different assessment of follicle size (<0.5 mm or 5 mm or greater), or no obvious reason for the mismatch.

The survey was approved by the Institutional Review Board of Johns Hopkins Medicine and the National Institute for Medical Research in Tanzania, and conducted under the tenets of the Declaration of Helsinki.

Results

Agreement was good between two independent graders when using the simplified quality grading scheme, with a weighted kappa of 0.67 (95% confidence interval [CI] = 0.50–0.84). The two graders agreed on quality for 82.7% of images in the mixed sample of 75 eyes (Table 5).

The agreement between two graders using the grading scheme for inflammation was also good, with a weighted kappa of 0.72 (95% CI = 0.59–0.85). The two graders agreed on inflammation in 81% of the 74 images (Table 6).

Of the 5428 images available in the survey, only 11 (or 0.2%) were deemed ungradable. Of the remaining 5417 eyes, field and photograph graders disagreed over the presence or absence of TF in 177 (3.2%) eyes (i.e.

Table 5. Grader Agreement using the Simplified Grading Scheme for Quality of Trachoma Images

	Grader 2			Total
	Good Quality	Moderate Quality	Poor Quality	
Grader 1				
Good quality	44	1	0	45
Moderate quality	12	15	0	27
Poor quality	0	0	3	3
Total	56	16	3	75

Weighted kappa = 0.67, 95% confidence interval (CI) = 0.50–0.84.

Table 6. Grader Agreement when determining the Degree of Inflammation in Trachoma Images

	Grader 2			Total
	No/Mild Inflammation	Moderate Inflammation	Severe Inflammation	
Grader 1				
No/mild inflammation	26	4	2	32
Moderate inflammation	1	14	3	18
Significant inflammation	1	3	20	24
Total	28	21	25	74

Weighted kappa = 0.72, 95% confidence interval (CI) = 0.59–0.85.

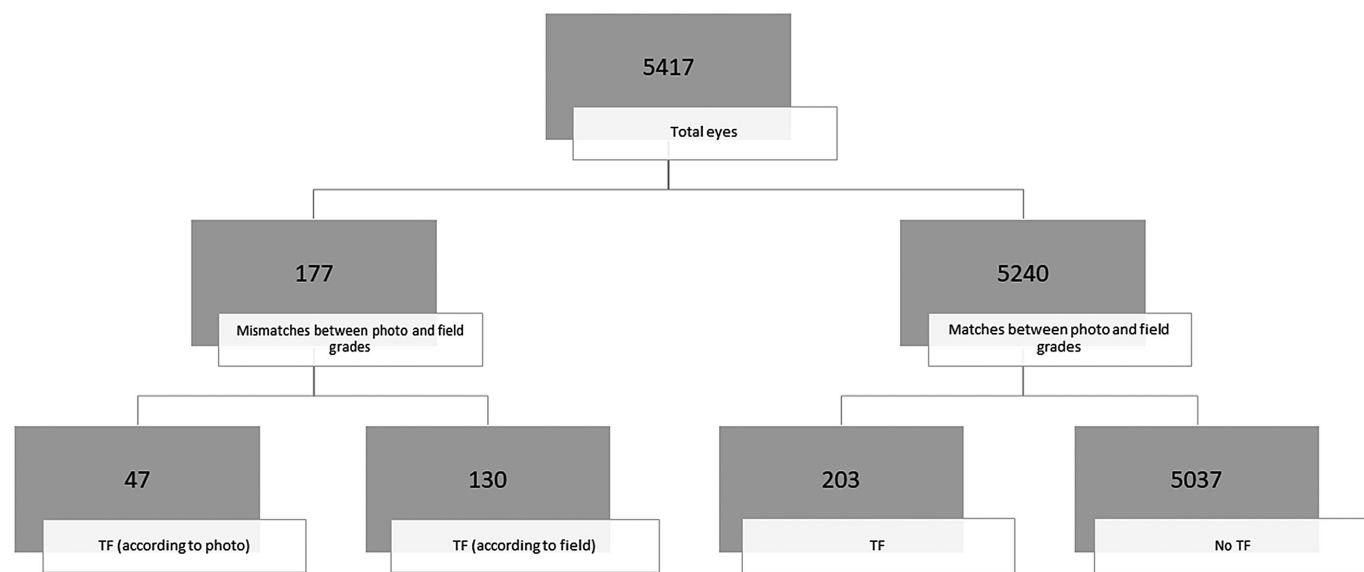


Figure 2. Total number of eyes with both field and photograph grades.

mismatched eyes). When there was a mismatch, the field graders assigned a field grade of TF ($n = 130$ eyes, 73.4% of the mismatched eyes) more frequently than the photograph graders ($n = 47$ eyes, 26.6% of the mismatched eyes; Fig. 2).

We used the expanded grading scheme to determine if there were specific quality metrics that contributed to

mismatch between field versus photo grades in the 177 mismatch eyes, compared to the sample of eyes where there was no mismatch (Table 7). The most common image quality issue was blanching in the image (7.3% of mismatch images and 5.0% of matching-grade images). The frequency of image quality issues worse than grade one was very low in both the mismatch and matched

Table 7. Detailed Analysis of Image Quality in Eyes where Field and Photograph Graders had Matched and Mismatched Grades for TF

		Lighting and Shadows Grade > 1	Focus Grade > 1	Glare and Tear Film Grade > 1	Eyelid Eversion Grade > 1	Blanching Grade > 1	Debris and Obstruction Grade > 1	Overall Quality Grade > 1*
	N	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Mismatched	177	1 (0.6)	5 (2.8)	4 (2.3)	6 (3.4)	13 (7.4)	1 (0.6)	9 (5.1)
Matched	60	0 (0.0)	1 (1.7)	1 (1.7)	1 (1.7)	3 (5.0)	0 (0.0)	1 (1.7)
P value	-	1.00	1.00	1.00	0.68	0.77	1.00	0.46

*Overall quality grade of 1 in this table was defined as: lighting and focus were both grade 1 **and** <33% of the upper tarsal conjunctiva was obscured by a combination of other image quality metrics.

Table 8. Severity of Inflammation in Eyes where Field and Photograph Graders had Matched and Mismatched Grades for TF

	N	Inflammation			P Value (Test for Trend)
		Grade = 1 N (%)	Grade = 2 N (%)	Grade = 3 N (%)	
Mismatched	177	98 (55.4)	47 (26.5)	32 (18.1)	0.048
Matched	60	45 (75.0)	6 (10.0)	9 (15.0)	

eyes, and although the frequencies were slightly higher in the mismatch eyes, there was no significant difference between the two groups. Over 98% of the random sample of matching-grade eyes were evaluated as high-quality images, and 95.5% of the mismatch eyes were evaluated as high-quality images.

However, there was a significant difference in the degree of inflammation between mismatch eyes and matched eyes, with mismatch eyes being more likely to have higher-grade inflammation than matched eyes

(Table 8). Within the 177 mismatch eyes, inflammation influenced both photograph graders and the field grader, as the degree of grade 3 inflammation (TI) was no different in the 47 eyes called TF only by the photograph graders (7/47 or 14.9%) compared to the 130 eyes called TF only by the field grader (25/130 or 19.2%, $P = 0.51$).

For mismatch eyes where the field grader assigned a grade of TF and the photograph graders did not ($n = 130$), a review of the images suggested the main reason

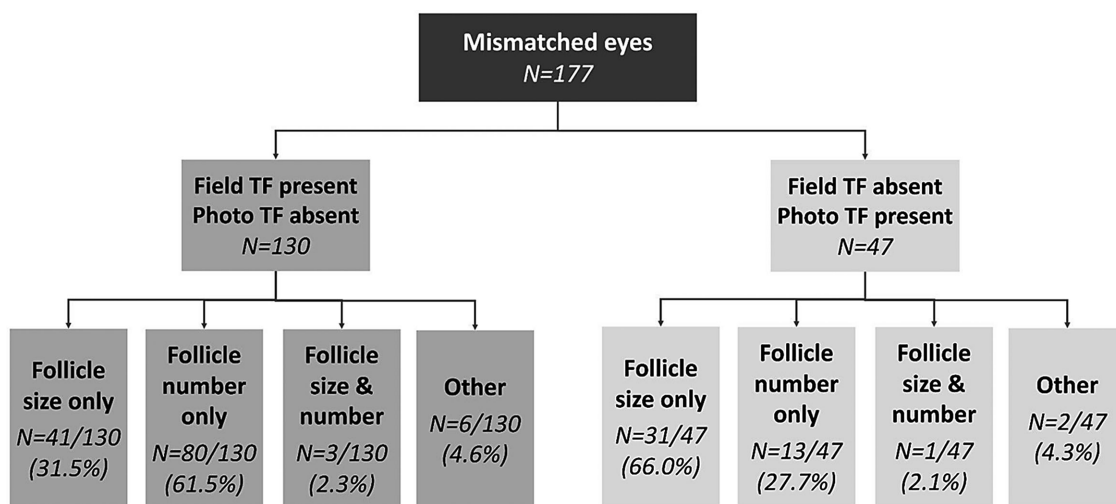
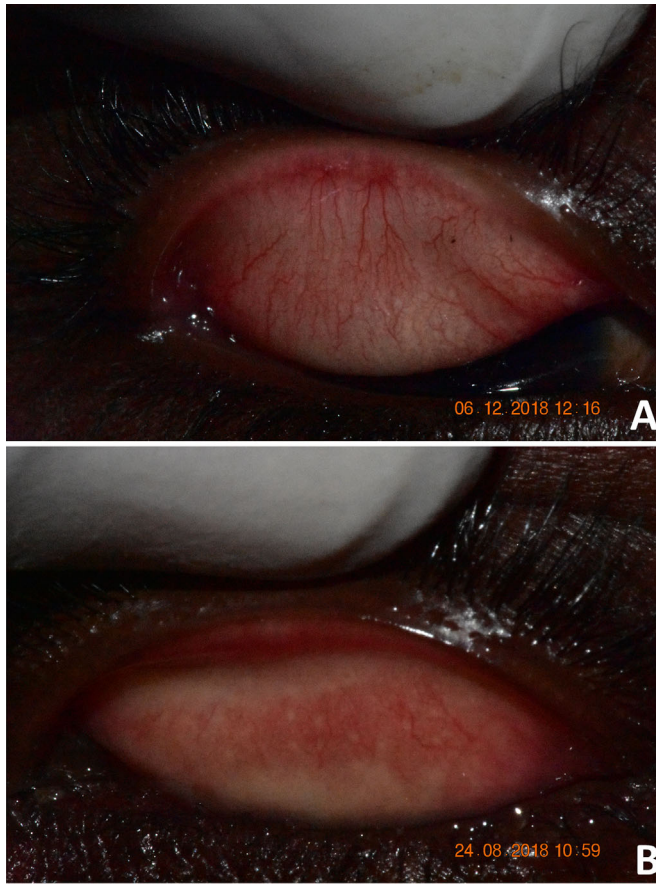


Figure 3. Photograph grader assessment of the possible reasons for mismatch between photo and field grades for TF, according to review of images where the field grade was TF ($N = 130$) or the photograph grade was TF ($N = 47$).



Figures 4. Examples of images where there was a mismatch in TF grading due to differences in interpretation of follicle number and follicle size. **(A)** Mismatch due to follicle number, with the field grader assigning a grade of TF and the photograph graders assigning a grade of no TF. **(B)** Mismatch due to follicle size, with the photograph graders assigning a grade of TF and the field grader assigning a grade of no TF.

for disagreement was due to differences in whether five follicles were present (Fig. 3). In 61.5% of the 130 eyes, the field grader observed at least 5 follicles, but the photograph graders did not find 5 follicles in the image (Fig. 4a). The greater frequency of the perceived follicle number as the reason for mismatch in eyes where the field grader called TF versus eyes where the photograph graders called TF — 61.5% versus 27.7% — was statistically significant ($P = <0.0001$).

However, for mismatch eyes where the photograph graders assigned a grade of TF and the field grader did not, the main reason for disagreement was due to differences in interpretation of follicle size (66.0%), where the photograph graders felt that the follicles present were of sufficient size and the field grader did not (Fig. 4b). The lesser frequency of perceived follicle size as the reason for the mismatch in eyes where the field grader called TF versus eyes where the photograph graders called TF — 31.5% versus 66.0% — was

statistically significant ($P = <0.0001$). In about 4% of images, regardless of direction of the mismatch, the photograph graders could not discern any likely reason for the mismatch between field and photograph grades for TF.

The presence of inflammation also appeared to affect whether or not the size or the number of follicles was the likely reasons for the mismatch in field versus photograph grade. As the score of inflammation increased, the likelihood that the size of the follicle was the problem increased, from 39% in grade 1 inflammation, to 50% in grade 3 (Table 9). As inflammation increased, it was less likely that the number of follicles was the issue in the mismatch. None of the trends were statistically significant.

Discussion

Previous studies using photography to grade eyelids for trachoma have shown reasonably good correlation with field grading.^{4,7-9} However, some studies have had difficulties with high rates of ungradable images from 11% to 78%.^{5,9,10} Commonly reported factors contributing to images being ungradable included improper focus on the grading area, inadequate coverage of the grading area, excess light reflection, or shadows obscuring the grading area. Quality assessment of photographs at an early stage in photographer training and surveying may help prevent data loss and wasted effort in the field, but there is no standardized assessment for quality of images of the tarsal conjunctiva. In this study, we defined a series of quality metrics to use when grading the quality of images of the everted upper eyelid. This detailed assessment was quite time-consuming to implement to determine the degree of quality for each metric, requiring almost 5 minutes per image. A simplified overall quality grading scheme was then developed, which was much easier to use and had reasonable inter-grader agreement. Such a scheme could be rapidly deployed to measure the quality of images and provide feedback to photographers. The more detailed scheme might be used where feedback on the reasons for poor quality is needed. If the grade of images is the primary end point for a survey or a research study, then quality assessment can be built in at the outset by the review of images by the photograph graders. The simple method for quality assessment could also be taught to photography supervisors who could review a sample of images each day during a survey to provide feedback. The timing of review is critical for the training utility of assessing

Table 9. Percentage of Mismatched Eyes with either Number of Follicles or Size of Follicles as the likely Reasons for the Mismatch, by Degree of Inflammation

Reason for Mismatch	N = 177	Inflammation		
		Grade = 1 N = 96	Grade = 2 N = 47	Grade = 3 N = 32
Size of follicle*	76	37 (38%)	23 (49%)	16 (50%)
Number of follicles†	101	57 (58%)	24 (51%)	16 (50%)

*Test for trend, $P = 0.14$.†Test for trend, $P = 0.35$.

quality, as it is not helpful if performed near the end of the survey and the quality is found lacking.

We applied the detailed metric of quality assessment in the context of a previous survey where we had both field and image grades for TF to determine if there was a particular feature of the image that might explain the mismatch in grades between the field and image grade for TF. The rate of ungradable images was very low, 0.2% of eyes. Overall, the rate of mismatch eyes was very low as well, 177 (3.3%) of 5417 images. There was generally good agreement on the absence of TF by both field and image graders. However, in the possible presence of TF, we found that the rate of assigning a grade of TF was not equal between the field grader and the image graders in the same eyes. As [Figure 2](#) shows, in the 333 total eyes where the field grader found TF, 39% (130) of those images were not called TF by photograph graders. In contrast, of the 250 total eyes where the photograph graders found TF, only 19% (47) were not also called TF in the field.

We sought to determine the role of image quality as a reason for the mismatch between field and image grades. Because the possible presence of TF may be a confounder in assessing quality, we stratified the random sample of comparator eyes (i.e. eyes where the field and photograph grades for TF matched) by the presence of TF to be certain that close to half of the sample had TF. In fact, the overall image quality was high in both the samples of eyes where the field grades agreed with image grades and the full sample of mismatched eyes. The most common problem overall was blanching, caused by prolonged eversion of the eyelid, which makes ascertainment of follicles difficult. Blanching of greater than 10% of the grading area of the upper eye lid occurred in 7% of the mismatched eyes compared to 5% of the matched eyes, a difference that was not significant. We note that with the low rate of image quality issues, coupled with the small sample of mismatched eyes, we had limited power to detect significant differences in quality. We argue that even had the sample size been larger, it is not clear from our

**Figure 5.** Image with inflammation where the field grade was TF but the photograph grade was not TF.

data that quality issues explain much of the difference found between field and image grades.

However, the presence of inflammation, as graded on images, does appear to explain some of the differences between the matched and mismatched eyes. The matched eyes were more likely to have no or mild inflammation on image review, whereas the mismatched eyes were more likely to have higher-grade inflammation. Inflammation can be severe enough to obscure follicles, which would impact the assessment of the number of follicles, as well as cause encroachment of tissue around the follicle, leading to apparent diminution in size. An analysis of the impact of inflammation on the reasons for the mismatch suggested that at least an effect on the size of the follicles might be an issue. An example of such a problem may be seen in [Figure 5](#), which was called TF by the field grader but not TF by the photograph graders.

Currently, the sign of TI is not included in the assessment of active trachoma, which relies solely on the sign of TF. It is not clear if graders compensate for the presence of TI by lowering the threshold for the size of follicles or by presuming that if three or four follicles are visible that more may be hidden under the inflamed

tissue. The impact of inflammation on field grading and image grading needs further discussion because, although inflammation may be a relatively rare sign in general, it was present in close to half of the eyes where mismatch of TF grades occurred. The argument that TI cannot be graded reliably, at least for images, was not the case in this study, and agreement on grading TI in the field has also reportedly been very good.⁴

If the field grader had been more likely to compensate for the presence of inflammation than the image graders, we would have expected much higher rates of inflammation in the 130 eyes where only the field grader called TF than in the eyes where only the image graders called TF. However, the rates of grade three inflammation were not different between these two groups. The rate of inflammation in the 130 eyes called TF by only the field grader was 19.2%, compared to 14.9% in the eyes called TF by only the image graders. Although there was some indication of a higher rate in mismatch eyes with field grades of TF, the difference was not large and not statistically significant.

Without knowledge of the direction of mismatch (i.e. whether the field or photograph graders called TF a mismatch eye), the photograph graders were asked to re-review the 177 mismatched images again and speculate on the possible reasons for mismatch. In particular, was it likely that the size of the follicles was an issue, or the number of follicles, or both reasons? If they could not discern an obvious reason, this was also noted. There are obvious limitations to this approach, primarily because the determination of reason for mismatch was made solely by the photograph graders without the thought process of the field grader and a re-examination of the eye in the field was not possible, which may have allowed for reconciliation of the image and field grades. The determination was based solely on the image, which, for example, could have had small areas of glare that obscured a follicle. Similarly, variations in follicle shape may have contributed to some ambiguity. Mismatched field and TF grades may have also represented recording errors in the field. Our data suggested that the problem of determining the number of follicles present was the primary reason the field grader called TF when the image graders did not. It is tempting to assume that the field grade should be the “gold standard,” as it represented review of the actual everted lid that could be assessed from multiple angles, whereas the photograph graders had only the image to assess. However, there is a risk of overcalling TF in situations where the rate of TF is low, as was the case in this survey, so we cannot entirely rule out field grader error.^{11,12} The data suggest that a difference in determination of the size of the follicles was likely a more common reason

for the photograph graders to call TF when the field grader did not. For instance, the lid flipper in the survey did not have a thumb marker to assist in determining size, whereas photograph graders can standardize size and account for magnification with a ruler. Overall, the exercise of determining potential reason for mismatch was useful for at least two reasons. First, it points to the difficulty of categorizing borderline cases in a survey, which must be included in any live training of field graders and for training of image graders.¹³ Second, the findings again highlight that when comparing field and photograph grades of the same eye, we should not assume the field grade is the gold standard. In our data, we had photographic evidence of eyes with five follicles of the correct size that were not called TF by the field grader.

There are some limitations to this study, in addition to those noted above. The quality of the images in the survey was overall very good, and very few were ungradable. We have historically used a well-trained photographer using an SLR camera for our surveys and recognize that comparisons using other camera systems in other surveys may not yield the same result. Thus, quality of images may be a more important factor in producing grading mismatches than we were able to discern from our dataset. However, by monitoring the quality of images using the simple quality assessment scheme, where we demonstrated reliable agreement, image quality should be enhanced in general. The higher rate of mismatches where the field graded TF may also be a function of the overall low rate of TF prevalence in the survey, where the cases may not be as severe. Where TF prevalence is high and the cases are more florid, there may be less mismatch. Ideally, we would have had multiple field graders for each eye for this study to help clarify the field grade, as we did for the image grade.

In summary, we developed a useful tool to provide a rapid and reliable assessment of the quality of images of the upper tarsal conjunctiva that can be used to monitor photographers in the field. If the image quality is substandard, a more detailed assessment to determine the precise issue and institute re-training can be undertaken. Whereas we initially thought that the quality of the image, or metrics of quality, might be lower in eyes that had mismatched field and image TF grades, that was in fact not the case. A more significant problem was the presence of inflammation in the eye, a physical sign which affects both field and photograph grading. Training of both image and field graders needs to be standardized to either ignore inflammation or provide some accommodation in terms of follicle presence and size. Ideally, reconsideration of including the sign of TI in the determination of active trachoma

might be valuable, as it does provide some information on ocular disease in the presence of TF¹⁴ and can be graded reliably.

Acknowledgments

Funding Sources: Funds for the survey were provided by a grant from the Task Force for Global Health. Additional funds were provided by the El Maghraby chair at the Wilmer Institute.

Disclosure: **M. Odonkor**, None; **F. Naufal**, None; **H. Mkocha**, None; **N. Funga**, None; **B. Muñoz**, None; **S.K. West**, None

References

- World Health Organization. World Report on Vision. Geneva, Switzerland: World Health Organization; 2019.
- World Health Organization. Design parameters for population-based trachoma prevalence surveys. Geneva, Switzerland: World Health Organization; 2018.
- Solomon AW, Kello AB, Bangert M, et al. The simplified trachoma grading system, amended. *Bulletin of the World Health Organization*. 2020;98(10):698.
- Naufal F, West SK, Brady CJ. Utility of Photography for Trachoma Surveys: A Systematic Review [published online ahead of print August 20, 2021]. *Surv Ophthalmol*, <https://doi.org/10.1016/j.survophthal.2021.08.005>.
- Butcher RM, Sokana O, Jack K, et al. Low prevalence of conjunctival infection with *Chlamydia trachomatis* in a treatment-naïve trachoma-endemic region of the Solomon Islands. *PLoS Neglected Tropical Dis*. 2016;10(9):e0004863.
- Odonkor M, Naufal F, Munoz B, et al. Serology, infection, and clinical trachoma as tools in prevalence surveys for re-emergence of trachoma in a formerly hyperendemic district. *PLoS Neglected Tropical Dis*. 2021;15(4):e0009343.
- West SK, Taylor HR. Reliability of photographs for grading trachoma in field studies. *Br J Ophthalmol*. 1990;74(1):12–13.
- Sheehan JP, Gebresillase S, Shiferaw A, et al. School-based versus community-based sampling for trachoma surveillance. *Am J Tropical Med Hygiene*. 2018;99(1):150.
- Emerson PM, Lindsay SW, Alexander N, et al. Role of flies and provision of latrines in trachoma control: cluster-randomised controlled trial. *Lancet*. 2004;363(9415):1093–1098.
- Solomon AW, Bowman RJ, Yorston D, et al. Operational evaluation of the use of photographs for grading active trachoma. *Am J Tropical Med Hygiene*. 2006;74(3):505.
- Harding-Esch EM, Sillah A, Edwards T, Burr SE, Hart JD, Joof H, Partnership for Rapid Elimination of Trachoma (PRET) study group. Mass treatment with azithromycin for trachoma: when is one round enough? Results from the PRET Trial in the Gambia. *PLoS Neglected Tropical Dis*. 2013;7(6):e2115.
- Martin DL, Bid R, Sandi F, et al. Serology for trachoma surveillance after cessation of mass drug administration. *PLoS Neglected Tropical Dis*. 2015;9(2):e0003555.
- Gaynor BD, Amza A, Gebresailassie S, et al. Importance of including borderline cases in trachoma grader certification. *Am J Tropical Med Hygiene*. 2014;91(3):577.
- Zambrano AI, Munoz BE, Mkocha H, et al. Measuring trachomatous inflammation-intense (TI) when prevalence is low provides data on infection with *Chlamydia trachomatis*. *Invest Ophthalmol Vis Sci*. 2017;58(2):997–1000.