INTRODUCTION

Trachoma is the leading infectious cause of blindness, and the World Health Organization (WHO) aims to eliminate it by the year 2020. If this is to be realized, research to refine control strategies is urgently required.

For clinical assessment, the WHO simplified trachoma grading system (Table 1) is widely used in both research and control programs, even though it was developed only to aid assessment of trachoma by non-specialist personnel. The system is believed to have good reproducibility. However, field assessments are almost impossible to mask. In research, this makes it difficult to exclude the possibility of bias.

West and Taylor examined the use of still photographs for verifying field diagnoses of the signs trachomatous inflammation-follicular (TF), trachomatous inflammation-intense (TI), and trachomatous conjunctival scarring (TS) (Table 1). ASA 25 slide film and a macro lens sufficient to provide 1:1 magnification were used; a single exposure of the everted tarsal conjunctiva was taken from both eyes of each of 136 subjects. Slides were later examined on a light box by the clinical grader. Twenty-three (8.5%) of 272 photographs were found to be ungradable because of poor focus, inadequate eyelid eversion, shadowing, or obscuration of the central tarsal plate by the flash reflex. In the remaining 249 photographs, there was good correlation between clinical grading under field conditions and subsequent photograph grading, with kappa scores of 0.71 for TF, 0.74 for TI, and 0.73 for TS. No other formal analyses of the reliability of photographs for grading trachoma have been published.

Many subsequent studies have used photographs for the purposes of validating field data, explaining positive laboratory results for individuals graded clinically as not having active trachoma, or as the single means of assessing clinical status. In view of this reliance on photographs and their potential application in our own studies, we have undertaken further assessment of the process.

SUBJECTS AND METHODS

Research methods conformed to the tenets of the Declaration of Helsinki. Ethical approval for the study was obtained from the ethics committees of the Kilimanjaro Christian Medical Center (Moshi, Tanzania) and the London School of Hygiene and Tropical Medicine (London, United Kingdom). Written informed consent was obtained from all adult participants and all parents or guardians of children.

The study took place in Kahe Mpya sub-village in the Rombo District of Tanzania. Before commencing fieldwork, the field grader (PAM, an ophthalmic nurse with extensive trachoma field experience) was evaluated against another experienced, validated grader (DCWM). Masked to the other’s assessment, each independently examined the right eyes of the same fifty 5–7 year-old children. According to the reference grader, the prevalences of TF, TI, and TS were 10/50 (20%), 3/50 (6%), and 4/50 (8%), respectively. Agreement was 100%, 96%, and 96%, giving kappas of 1.00 (perfect agreement), 0.73, and 0.73.

In July 2000, we invited all residents of Kahe Mpya to participate in a longitudinal study. Clinical grades and photographs used here were obtained at baseline before any interventions against trachoma.

The everted right tarsal conjunctiva of each participant was evaluated against the simplified WHO system criteria using × 2.5 binocular loupes. Grading was undertaken in sunlight

---

* Address correspondence to Anthony W. Solomon, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom. Written informed consent was obtained from all adult participants and all parents or guardians of children.

† The kappa statistic is an index of intra-observer or inter-observer reliability for categoric data. It is the difference between the observed and chance values of the proportion of agreement between two sets of observations of the same variable, expressed as a proportion of this difference’s maximum value. Kappa therefore has possible values between −1 and +1, with −1 indicating complete disagreement, +1 complete agreement, and 0 the level of agreement expected by chance. Divisions for describing the relative strength of agreement associated with this measurement have been (arbitrarily) defined as poor = ≤ 0.00; slight = 0.00–0.20; fair = 0.21–0.40; moderate = 0.41–0.60; substantial = 0.61–0.80; and almost perfect = 0.81–1.00.
World Health Organization simplified clinical grading scheme for trachoma®

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF</td>
<td>Trachomatous inflammation-follicular: the presence of five or more follicles at least 0.5 mm in diameter in the central part of the upper tarsal conjunctiva</td>
</tr>
<tr>
<td>TI</td>
<td>Trachomatous inflammation-intense, pronounced inflammatory thickening of the upper tarsal conjunctiva obscuring more than half the normal deep tarsal vessels</td>
</tr>
<tr>
<td>TS</td>
<td>Trachomatous conjunctival scarring: the presence of easily visible scars in the tarsal conjunctiva</td>
</tr>
<tr>
<td>TT</td>
<td>Trachomatous trichiasis: at least one eyelash rubs on the eyeball, or evidence of recent removal of in-turned eyelashes</td>
</tr>
<tr>
<td>CO</td>
<td>Corneal opacity: easily visible corneal opacity over the pupil so dense that at least part of the pupil margin is blurred when viewed through the opacity</td>
</tr>
</tbody>
</table>

RESULTS

At enumeration, there were 978 individuals living in Kahe Mpya. We examined 956® and photographed 948 (age range = 8 days–101 years). No photographs were available from eight individuals because of a temporary camera malfunction; these subjects were excluded. Based on field diagnoses, the prevalence of right eye TF was 12% (117 of 948) and the prevalence of right eye TI was 12% (110 of 948). Of the eight excluded individuals, one (13%) had TF and one (13%) had TI.

Grader A found that 106 (11%) of 948 sets of photographs were inadequate for grading. Graders B and C (the field grader grading the photographs) found 1 set (0.1%) and 35 sets (4%) inadequate, respectively. Based on graded sets only, graders A, B, and C recorded TF prevalences of 11%, 38%, and 9%, respectively, and TI prevalences of 23%, 15% and 14%, respectively.

Inter-observer agreements are shown in Table 2 for the comparison of photographic and field grading. Kappa statistics for agreement between the three photographic graders were 0.32 (95% confidence interval [CI] = 0.29–0.35) for TF and 0.52 (95% CI = 0.49–0.55) for TI or, after exclusion of ungradable photographs, 0.37 (95% CI = 0.33–0.41) for TF and 0.66 (95% CI = 0.62–0.70) for TI.

Prevalence of active trachoma is highest in young children, and WHO recommends using prevalence of TF in 1–9-year-old children as the key sign for control programs. In field diagnoses in this study, the prevalence of TF in the 322 1–9-year-old children who had conjunctival photographs taken was 31%. Kappa statistics for the comparison of photographic grading versus field grading of TF for these children were 0.56 (95% CI = 0.47–0.65), 0.38 (95% CI = 0.29–0.47), and 0.43 (95% CI = 0.34–0.53) for photographic graders A, B, and C respectively (mean = 0.46). The kappa statistics comparing the three photographic graders’ diagnoses of TF in children were 0.34 (95% CI = 0.29–0.40) or 0.41 (95% CI = 0.34–0.48) after exclusion of ungradable photographs.

DISCUSSION

These results are disappointing, inasmuch as they decrease our confidence in the utility of photographs for validating field diagnosis of trachoma. Using photographs, three very experienced trachoma graders had only fair-to-moderate agreement® with the field grader and with each other for each sign. Mean kappa scores for the three photographic versus field comparisons were 0.44 and 0.51 for TF and TI, respectively.

Examining individuals in a village and reading photographs in an office are very different activities. In the field, there is...
On the basis of our results, we conclude that either an independent investigator, or that photographs are problematic for their di-

Sixty percent specifies the film speed. None provides

standardized against a gold standard grader, we did not have

the opportunity to validate our two other photograph graders against the gold standard, each other, or the field grader. Our work’s limitations, however, mirror those of most published studies that use photographs. Of 13 studies using photographs to validate or replace field trachoma grading cited in this paper’s introduction, only two13,16 state that slide film was used, and only one10 specifies the film speed. None provides sufficient information to determine the image magnification ratio between field and photographic examination. Most reports give no information as to who graded the photogra-

graphs10,11,17–19 or identify them only as a trained grader,12 a trained reader,20,21 an independent investigator,13 or clini-

cians.16 On the basis of our results, we conclude that either the signs TF and TI are less reproducible than previously believed1,2 or that photographs are problematic for their di-

agnosis.

Could better pictures be obtained? Digital imaging, now recommended for fundus photography in diabetic retinopathy screening,25 could potentially reduce the proportion of sub-

jects for whom no useful pictures are delivered to the remote examiner because it allows quality control through immediate image review.26 Until recently, however, digital images were of lower resolution than those generated by conventional photography, and in any case the difficulties presented by the irregular curvature of the conjunctiva, the limited number of views that can be taken of each eye, and the two-dimensional representation of three-dimensional epithelium would persist.

A recent trial of latrine provision for trachoma control used a combination of slide and digital photography.16 Sixty percent of 2,489 slide photographs were gradable and 72% of 986 digital images were gradable. The proportion of images that were out of focus, too bright, too dark, or which otherwise provided inadequate views was approximately the same for the two media; the only difference was that 353 slide photo-

graphs were rendered useless by problems (such as untimely camera opening) that affected whole rolls of film (Emerson PM and others, unpublished data). Digital photography does at least minimize the risk of the latter type of error. Unless reliability testing demonstrates better agreement than seen here, however, for trachoma studies, we believe that photo-

graphs should not be used for diagnosing TF or TI.

Received May 18, 2005. Accepted for publication September 28, 2005.

Acknowledgments: We thank the village and sub-village chairmen, elders, and villagers of Kahe for their enthusiastic participation; our field team for help with data collection; Dr. Paul Emerson for helpful conversations about the use of photographs; and Professor John Shao and the members of the research steering committee.

Financial support: This study was supported by grants from the Wellcome Trust/Burroughs Wellcome Fund (059134) and the Edna McConnell Clark Foundation (99100).

Authors’ addresses: Anthony W. Solomon, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom, Telephone: 44-20-7958-8336; Fax: 44-20-7958-8317, E-mail: anthony.solomon@lshtm.ac.uk. Richard J. C. Bowman, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom, Telephone: 44-20-7958-8359; Fax: 44-20-7958-8317, E-mail: richardbowman@intafrica.com. David C. W. Mabey, Clinical Research Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom, Telephone: 44-20-7958-8359; Fax: 44-20-7958-8317, E-mail: richardbowman@intafrica.com.
REFERENCES


