

OPERATIONAL EVALUATION OF THE USE OF PHOTOGRAPHS FOR GRADING ACTIVE TRACHOMA

ANTHONY W. SOLOMON,* RICHARD J. C. BOWMAN, DAVID YORSTON, PATRICK A. MASSAE, SALESIA SAFARI, BRIAN SAVAGE, NEAL D. E. ALEXANDER, ALLEN FOSTER, AND DAVID C. W. MABEY

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom; Kilimanjaro Christian Medical College, Tumaini University, Moshi, Tanzania; Huruma Hospital, Rombo District, Tanzania; Moorfields Eye Hospital, London, United Kingdom

Abstract. We evaluated the reliability of photographs to verify field diagnoses of active trachoma. We examined 956 residents of a trachoma-endemic village for signs of trachoma using the World Health Organization simplified grading system. Two photographs of the right eye of 948 persons were independently graded (masked to field assessment) by the field examiner and two other experienced graders. There was only moderate agreement between field assessment and the subsequent photographic evaluations by the three graders. When we counted ungradable photographs as disagreements, mean kappa scores for the signs trachomatous inflammation (follicular [TF]) and trachomatous inflammation (intense [TI]) were 0.44 and 0.51, respectively. There was also only fair-to-moderate agreement between the three assessments (by different examiners) of the photographs. Either the signs TF and TI themselves are not as reliable as previously believed, or photographs should be used for their diagnosis only when reliability testing demonstrates better agreement than found here.

INTRODUCTION

Trachoma is the leading infectious cause of blindness,¹ and the World Health Organization (WHO) aims to eliminate it by the year 2020.² If this is to be realized, research to refine control strategies is urgently required.³

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.⁴ The system is believed to have good reproducibility.^{4,5} However, field assessments are almost impossible to mask. In research, this makes it difficult to exclude the possibility of bias.

West and Taylor⁶ 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.⁶ 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,^{9–16} explaining positive laboratory results for individuals graded clinically as not having active trachoma,¹⁷ or as the single means of assessing clinical status.^{18–21} 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.²² Before commencing fieldwork, the field grader (PAM, an ophthalmic nurse with extensive trachoma field experience) was evaluated against another experienced, validated²³ 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.^{22,24} Clinical grades and photographs used here were obtained at baseline²² before any interventions against trachoma.

The everted right tarsal conjunctiva of each participant was evaluated against the simplified WHO system criteria⁴ 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, Keppel Street, London WC1E 7HT, United Kingdom. E-mail: anthony.solomon@lshtm.ac.uk

† The kappa statistic is an index of intra-observer or inter-observer reliability for categorical 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.⁷ 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 = 0.81–1.00.⁸

TABLE 1

World Health Organization simplified clinical grading scheme for trachoma⁴

TF	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
TI	Trachomatous inflammation-intense: pronounced inflammatory thickening of the upper tarsal conjunctiva obscuring more than half the normal deep tarsal vessels
TS	Trachomatous conjunctival scarring: the presence of easily visible scars in the tarsal conjunctiva
TT	Trachomatous trichiasis: at least one eyelash rubs on the eyeball, or evidence of recent removal of in-turned eyelashes
CO	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

whenever possible; when the conjunctiva was inadequately illuminated, a torch was used.

To increase the likelihood of obtaining at least one satisfactory picture of each eye, we took two photographs of the tarsal conjunctiva of each individual. We used an EOS-300 single lens reflex camera (Canon, Tokyo, Japan), an EF 100mm f2.8 macro lens (Canon), a $\times 2$ teleconverter, Macro-Lite ML-3 ring flash (Canon), and 100ASA color print film (FujiFilm, Tokyo, Japan). The camera was hand-held at a focal length of approximately 30 cm and manually focused on the central tarsal plate. Aperture was set at $f/19$. One photographer took all photographs; before this study, he had taken approximately 2,000 conjunctival photographs in trachoma-endemic villages.

Prints (15 cm \times 10 cm) were prepared by a professional London photograph laboratory (giving a final magnification of $\times 5$, c.f. $\times 2.5$ for field grading), then assessed independently by two ophthalmologists, both of whom were experienced trachoma graders. Photographs were also graded independently by the field grader. Photograph grading was undertaken without magnifying loupes, using the simplified WHO system, and masked to field assessments and assessments of the other photograph graders.

Each set of two photographs (taken of one conjunctiva) was considered a pair, allowing graders to obtain as much visual information as had been recorded for that eye. Because most research studies use photographs to assess for active trachoma, only the signs TF and TI are considered here. For each subject, graders could record grades for these two signs or, if the photographs did not provide sufficient information, record that the two photographs were ungradable. No time limit was imposed.

Data were double-entered into Microsoft (Redmond, WA) Access (2002, SP3) and analyzed using Stata 7 (Stata Corporation, College Station, TX). For primary analyses, when a photograph grader wrote ungradable, but the comparison grader (in the field or examining photographs) made a diagnosis, this counted as a disagreement. Analyses were repeated with exclusion of ungradable photographs.⁶

RESULTS

At enumeration, there were 978 individuals living in Kahe Mpya. We examined 956^{22,24} 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 photograph graders A, B, and C respectively (mean = 0.46). The kappa statistics comparing the three photograph 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 agreement⁸ 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

TABLE 2

Comparability between the field grader and three photograph graders for the signs trachomatous inflammation—follicular (TF) and trachomatous inflammation—intense (TI) in 948 eyes from 948 subjects. Figures are kappa scores, counting ungradable photographs as disagreements (in **bold** type) and excluding ungradable photographs (in normal type)*

	Photograph grader			
	A	B	C	Mean
TF	0.49 (0.45–0.54) 0.79 (0.72–0.86)	0.34 (0.29–0.39) 0.34 (0.29–0.39)	0.49 (0.43–0.54) 0.58 (0.52–0.65)	0.44 0.57
TI	0.41 (0.36–0.45) 0.58 (0.52–0.64)	0.58 (0.52–0.65) 0.59 (0.52–0.65)	0.53 (0.48–0.59) 0.62 (0.56–0.69)	0.51 0.60

* Values in parentheses are 95% confidence intervals.

pressure to maintain high throughput, and many individuals (particularly children) are unable to cooperate fully with the examination process. Conversely, the conjunctivae may be examined from multiple angles and are always in focus; illumination can be adjusted if required. The photograph graders in this study, conversely, saw only two views of each conjunctiva, and image quality relied on the subject's cooperativeness, the photographer's skill and patience, and the nature of the photographic medium. It is difficult to take clear, close photographs of a small, irregularly curved, reflective, and often camera-shy surface. Furthermore, particularly when thickened and inflamed, the conjunctiva is a three-dimensional structure, and can only be imperfectly represented by a two-dimensional photograph. Although the camera was manually refocused from a slightly different vantage for the two pictures of each eye (in an effort to provide two slightly different views), the amount of information available to the photographic examiners was considerably less than that available to the field examiner.

In the only other published evaluation of imaging for trachoma,⁶ West and Taylor achieved better agreement than we did. Their study had some limitations. First, the same expert examiner examined subjects clinically and graded the slides. Although masked to the clinical grade, the examiner would have known the approximate prevalence of each sign. In our study, two other highly experienced graders (in addition to the field examiner) evaluated the photographs. Second, in the study of West and Taylor, there was agreement in clinical trachoma status between right and left eyes in 91% of the 136 subjects. It is not clear from their report if right and left eye slides of each patient were examined sequentially; if they were, a potential bias was introduced. In our study, only right eyes were included. Third, West and Taylor excluded the 8.5% of photographs believed to inadequately represent the conjunctiva. We believe that photographs considered ungradable are more likely to be of conjunctivae for which the diagnosis is borderline. If a photograph provides an in-focus, free-of-flash-reflex view of three-quarters of the central tarsal conjunctiva, and six follicles are visible in that area, the grader will assign a diagnosis of TF. If no follicles are seen, the grader may be comfortable assigning a diagnosis of no TF. If three follicles are seen, the examiner's task is difficult. Similarly, in the field, when three follicles are noted in the first three-quarters of the central tarsal conjunctiva examined, the grader needs to evaluate the rest of the conjunctiva very carefully. Such eyes are the ones for which verification of field diagnoses are most important: excluding them from an evaluation of photographs may be unhelpful. We calculated primary kappa scores counting ungradable photographs as disagreements. If these photographs are excluded, the mean of three kappas for the photograph versus field comparisons increases slightly (from 0.44 to 0.57 for TF and 0.51 to 0.60 for TI), but agreement remains only moderate.⁸

On some counts, our study can be criticized in relation to the previous work. We used 100ASA film rather than high resolution 25ASA film.⁶ We used photographic prints, while they used slides, which have better color reproduction. Our photographic graders saw photographed tissues at twice the magnification used in the field, while West and Taylor's examiner had the same magnification in each setting. In addition, although our field grader (photograph grader C) was 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 two^{13,16} state that slide film was used, and only one¹⁶ 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 photographs^{9,10,15,17-19} or identify them only as a trained grader,¹² a trained reader,^{20,21} an independent investigator,¹³ or clinicians.¹⁶ On the basis of our results, we conclude that either the signs TF and TI are less reproducible than previously believed,^{4,5} or that photographs are problematic for their diagnosis.

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

44-20-7927-2297, Fax: 44-20-7637-4314, E-mail: david.mabey@lshtm.ac.uk. David Yorston, Moorfields Eye Hospital, 162 City Road, London EC1V 2PD, United Kingdom, Telephone: 44-20-7253-3411, Fax: 44-20-7253-4696, E-mail: dhyorston@enterprise.net. Patrick A. Massae, Rombo Trachoma Research Project, Huruma Hospital, PO Box 202, Mkuu, Rombo, Tanzania, Telephone: 255-27-275-7230, E-mail: patrick.massae@iwayafrica.com. Salesia Safari, Huruma Hospital, PO Box 948 Moshi, Tanzania, Telephone: 255-27-275-7136, Fax: 255-27-275-7341, E-mail: huruma.hospital@iwayafrica.com. Brian Savage, 91 Drewry Lane, Derby DE22 3QS, United Kingdom, Telephone: 44-1332-242-635, E-mail: Able4@btinternet.com. Neal D. E. Alexander, Infectious Diseases Epidemiology Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom, Telephone: 44-20-7927-2483, Fax: 44-20-7636-8739, Email: neal.alexander@lshtm.ac.uk. Allen Foster, International Centre for Eye Health, 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: allenfoster@compuserve.com.

REFERENCES

- Resnikoff S, Pascolini D, Etya'ale D, Kocur I, Pararajasegaram R, Pokharel GP, Mariotti SP, 2004. Global data on visual impairment in the year 2002. *Bull World Health Organ* 82: 844–851.
- World Health Organization, 1997. *Planning for the Global Elimination of Trachoma (GET): Report of a WHO Consultation*. (WHO/PBL/97.60). Geneva: World Health Organization.
- Kuper H, Solomon AW, Buchan J, Zondervan M, Foster A, Mabey D, 2003. A critical review of the SAFE strategy for the prevention of blinding trachoma. *Lancet Infect Dis* 3: 372–381.
- 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.
- Taylor HR, West SK, Katala S, Foster A, 1987. Trachoma: evaluation of a new grading scheme in the United Republic of Tanzania. *Bull World Health Organ* 65: 485–488.
- West SK, Taylor HR, 1990. Reliability of photographs for grading trachoma in field studies. *Br J Ophthalmol* 74: 12–13.
- Siegel S, Castellan NJ, 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- Landis JR, Koch GG, 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
- Taylor HR, Siler JA, Mkocho HA, Munoz B, Velez V, Dejong L, West S, 1991. Longitudinal study of the microbiology of endemic trachoma. *J Clin Microbiol* 29: 1593–1595.
- Taylor HR, Siler JA, Mkocho HA, Munoz B, West S, 1992. The natural history of endemic trachoma: a longitudinal study. *Am J Trop Med Hyg* 46: 552–559.
- Baral K, Osaki S, Shreshta B, Panta CR, Boulter A, Pang F, Cevallos V, Schachter J, Lietman T, 1999. Reliability of clinical diagnosis in identifying infectious trachoma in a low-prevalence area of Nepal. *Bull World Health Organ* 77: 461–466.
- Holm SO, Jha HC, Bhatta RC, Chaudhary JS, Thapa BB, Davis D, Pokhrel RP, Yinghui M, Zegans M, Schachter J, Frick KD, Tapert L, Lietman TM, 2001. Comparison of two azithromycin distribution strategies for controlling trachoma in Nepal. *Bull World Health Organ* 79: 194–200.
- Bowman RJ, Sillah A, van Dehn C, Goode VM, Muquit M, Johnson GJ, Milligan P, Rowley J, Faal H, Bailey RL, 2000. Operational comparison of single-dose azithromycin and topical tetracycline for trachoma. *Invest Ophthalmol Vis Sci* 41: 4074–4079.
- Laming AC, Currie BJ, DiFrancesco M, Taylor HR, Mathews JD, 2000. A targeted, single-dose azithromycin strategy for trachoma. *Med J Aust* 172: 163–166.
- Thein J, Zhao P, Liu H, Xu J, Jha H, Miao Y, Pizzarello L, Tapert L, Schachter J, Mabon M, Osaki-Holm S, Lietman T, Paxton A, 2002. Does clinical diagnosis indicate ocular chlamydial infection in areas with a low prevalence of trachoma? *Ophthalmic Epidemiol* 9: 263–269.
- Emerson PM, Lindsay SW, Alexander N, Bah M, Dibba SM, Faal HB, Lowe KO, McAdam KP, Ratcliffe AA, Walraven GE, Bailey RL, 2004. Role of flies and provision of latrines in trachoma control: cluster-randomised controlled trial. *Lancet* 363: 1093–1098.
- Taylor HR, Rapoza PA, West S, Johnson S, Munoz B, Katala S, Mmbaga BB, 1989. The epidemiology of infection in trachoma. *Invest Ophthalmol Vis Sci* 30: 1823–1833.
- West SK, Munoz B, Lynch M, Kayongoya A, Mmbaga BB, Taylor HR, 1996. Risk factors for constant, severe trachoma among preschool children in Kongwa, Tanzania. *Am J Epidemiol* 143: 73–78.
- West S, Munoz B, Lynch M, Kayongoya A, Chilangwa Z, Mmbaga BB, Taylor HR, 1995. Impact of face-washing on trachoma in Kongwa, Tanzania. *Lancet* 345: 155–158.
- Hsieh Y, Bobo LD, Quinn TC, West SK, 2001. Determinants of trachoma endemicity using Chlamydia trachomatis ompA DNA sequencing. *Microbes Infect* 3: 447–458.
- Hsieh YH, Bobo LD, Quinn TO, West SK, 2000. Risk factors for trachoma: 6-year follow-up of children aged 1 and 2 years. *Am J Epidemiol* 152: 204–211.
- Solomon AW, Holland MJ, Burton MJ, West SK, Alexander ND, Aguirre A, Massae PA, Mkocho H, Munoz B, Johnson GJ, Peeling RW, Bailey RL, Foster A, Mabey DC, 2003. Strategies for control of trachoma: observational study with quantitative PCR. *Lancet* 362: 198–204.
- Bailey R, Osmond C, Mabey DC, Whittle HC, Ward ME, 1989. Analysis of the household distribution of trachoma in a Gambian village using a Monte Carlo simulation procedure. *Int J Epidemiol* 18: 944–951.
- Solomon AW, Holland MJ, Alexander ND, Massae PA, Aguirre A, Natividad-Sancho A, Molina S, Safari S, Shao JF, Courtright P, Peeling RW, West SK, Bailey RL, Foster A, Mabey DC, 2004. Mass treatment with single-dose azithromycin for trachoma. *N Engl J Med* 351: 1962–1971.
- Freudenstein U, Verne J, 2001. A national screening programme for diabetic retinopathy. Needs to learn the lessons of existing screening programmes. *BMJ* 323: 4–5.
- Olson JA, Strachan FM, Hipwell JH, Goatman KA, McHardy KC, Forrester JV, Sharp PF, 2003. A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy. *Diabet Med* 20: 528–534.