Trachoma: evaluation of a new grading scheme in the United Republic of Tanzania

H. R. Taylor, S. K. West, S. Kata, & A. Foster

A new simplified grading system for trachoma, which is based on the presence or absence of five selected key signs, has been assessed. The level of inter-observer variation and of variation for individual observers (intra-observer variation) showed that the system had good reproducibility following a training period that included interactive clinical teaching. The grading scheme was quickly learned by experienced ophthalmologists and auxiliary health personnel (ophthalmic nurses). The scheme should therefore be suitable for widespread application in field surveys of trachoma.

Recently, a new simplified scheme for grading trachoma has been developed (1), principally to provide paramedical staff with a simple method for use in the field. The scheme needed not only to have an acceptable degree of reproducibility but also to be sufficiently straightforward so that it could be taught, within a reasonable training period, to non-specialist health personnel.

A field study of trachoma was recently undertaken in the United Republic of Tanzania. Prior to adopting the simplified grading for this study, a series of observer variation trials was conducted to assess the reproducibility of the method and the amount of training required to instruct non-specialists in its use. Here, we report the findings of the trials.

MATERIALS AND METHODS

Trachoma grading

Essentially, the recently simplified trachoma grading scheme (1) consists of noting the presence of five signs of the disease, each of which is assessed independently. It can be summarized as described below.

Trachomatous inflammation—follicular (TF). The presence of five or more follicles in the upper tarsal conjunctiva (follicles must be at least 0.5 mm in diameter).

Trachomatous inflammation—intense (TI). Pronounced inflammatory thickening of the tarsal conjunctiva that obscures more than half of the normal deep tarsal vessels.

Trachomatous conjunctival scarring (TS). The presence of easily visible scarring in the tarsal conjunctiva.

Trachomatous trichiasis (TT). At least one eyelash rubs on the eyeball. Evidence of recent removal of in-turned lashes was also graded as trichiasis.

Corneal opacity (CO). Easily visible corneal opacity over the pupil, so dense that at least part of the pupil margin is blurred when viewed through the opacity.

Examinations were performed using binocular loupes (×2.5) and a torch. Each eye was examined first for in-turned lashes (TT), and the cornea was then inspected for corneal opacities (CO). The upper eyelid was subsequently examined for inflammation (TF and TI) and scarring (TS). Eyes were viewed and assessed separately. Signs had to be clearly visible in order to be considered present.
Training

A set of specially selected slides, which included examples of the presence and absence of each sign and of the various combinations of signs, was used for the initial training session. Also included were slides showing examples of commonly confused conditions, such as concretions. Two ophthalmic nurses and an ophthalmologist took part in the training session, which consisted of a 2-hour interactive session during which the grading scheme was reviewed, followed by a detailed discussion of the features shown on the slides. There followed a 1-hour clinical training session during which patients were examined and their conditions discussed. Subsequently, the first observer variation trial was conducted in which 25 patients were examined independently by each grader to assess inter-observer agreement. In order to assess the consistency of grading for individual observers (intra-observer variation), the patients were re-examined at least 1 hour later, in random order, by each grader. The results of these examinations are shown as the “first trial” in Tables 2 and 3.

On the following day, further clinical training sessions lasting several hours were held. On this occasion, groups of about 10 patients were examined independently by each grader. At the end of the examination, any patients for whom there were grading discrepancies were reviewed by all the graders together and then used as teaching examples. A final trial to assess agreement was held in which 20 patients were examined independently by the instructor and one of the ophthalmic nurses. The same patients were re-examined by the ophthalmic nurse in order to provide an estimate of the variation for individual observers.

For each of the trial and clinical teaching sessions, both adults and children were examined so that each set of patients contained examples of each of the signs of trachoma that were being graded.

Analysis of data

The presence or absence of each sign was recorded on a data collection form. All data were then analysed on a microcomputer using cross-tabulation and Kappa (k)-statistical programs. The average of the k-statistics between graders provided a summary measure of overall agreement. The k-statistic, rather than simple proportion, was used to assess inter-observer agreement and consistency of individual observers since it corrects for the amount of agreement that might be expected to occur by chance alone (2). Values of the k-statistic can range from +1 (complete agreement) to −1 (complete disagreement). Values greater than 0 indicate agreement greater than that expected by chance alone, while those above 0.7 indicate good agreement (3).

RESULTS

Overall, each of the four observers (three trainees and the instructor) showed good grading consistency for each sign in the first trial (Table 1). In most cases, agreement for individual observers was excellent, although one grader showed only fair reproducibility

---

Table 1. Values of the k-statistic for consistency of individual observers among four graders using the simple scheme of grading signs of trachoma in the first trial

<table>
<thead>
<tr>
<th>Grader</th>
<th>TT^a</th>
<th>CO^a</th>
<th>TF^a</th>
<th>TI^a</th>
<th>TS^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.0</td>
<td>0.90</td>
<td>0.90</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>1.0</td>
<td>0.81</td>
<td>0.83</td>
<td>0.96</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>1.0</td>
<td>0.68</td>
<td>0.48</td>
<td>0.71</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>0.83</td>
<td>0.57</td>
<td>0.86</td>
<td>0.37</td>
</tr>
</tbody>
</table>

^a No of patients examined = 25, right eyes only presented
^TT = Trachomatous trichiasis
^CO = Corneal opacity
^TF = Trachomatous inflammation – follicular
^TI = Trachomatous inflammation – intense
^TS = Trachomatous conjunctival scarring

Table 2. Values of the k-statistic for inter-observer agreement for three graders (compared to grader 1) using the simple scheme of grading signs of trachoma in the first trial

<table>
<thead>
<tr>
<th>Grader</th>
<th>TT^a</th>
<th>CO^a</th>
<th>TF^a</th>
<th>TI^a</th>
<th>TS^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.0</td>
<td>1.0</td>
<td>0.80</td>
<td>0.66</td>
<td>0.73</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>0.83</td>
<td>0.91</td>
<td>0.71</td>
<td>0.39</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>0.83</td>
<td>0.60</td>
<td>0.88</td>
<td>0.27</td>
</tr>
</tbody>
</table>

^a No of patients = 25, right eyes only presented
^TT = Trachomatous trichiasis
^CO = Corneal opacity
^TF = Trachomatous inflammation – follicular
^TI = Trachomatous inflammation – intense
^TS = Trachomatous conjunctival scarring
Table 3. Values of the \( \kappa \)-statistic for the grading consistency of an individual observer and inter-observer agreement for three signs of trachoma following further training

<table>
<thead>
<tr>
<th>Agreement</th>
<th>TF(^b)</th>
<th>TI(^c)</th>
<th>TS(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual observer</td>
<td>0.79</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>Inter-observer</td>
<td>0.79</td>
<td>0.95</td>
<td>0.87</td>
</tr>
</tbody>
</table>

\(^a\) No. of patients = 20, right eyes only presented
\(^b\) TF = Trachomatous inflammation -- follicular
\(^c\) TI = Trachomatous inflammation -- intense
\(^d\) TS = Trachomatous conjunctival scarring.

for the TI sign, while another had some difficulty in consistently grading the TS sign.

Inter-observer agreement was also very good for four signs (Table 2), although two graders (3 and 4) showed poorer levels of agreement for TS. Further analysis showed that these two observers identified fewer patients with conjunctival scarring (TS) than graders 1 or 2. When re-tested for three of the signs after further training, there was a marked improvement in both the grading consistency of observer 3 and in inter-observer agreement when compared with that of the instructor (Table 3).

**DISCUSSION**

Inter-observer agreement and consistency of individual observers were good with the new grading scheme. Grading skills were readily assimilated by an ophthalmologist and by two ophthalmic nurses. The initial training period included review of photographs with minimal instruction using patients. At the completion of training, each grader showed acceptable-to-excellent individual consistency. Although inter-observer agreement was also very good for TT, CO, TF and TI, inter-observer variation for TS after the initial training scheme was less satisfactory. One observer (grader 3), who was re-tested after further interactive clinical instruction, exhibited a marked improvement in individual consistency for TS, while the inter-observer variation for this sign also increased. The teaching method used, whereby the instructor and pupils graded cases independently and then reviewed cases where there was disagreement, was particularly useful for establishing the threshold for grading the signs, especially TS. This finding emphasizes the need for a period of close interactive clinical instruction as well as the use of demonstration slides.

The results obtained compare favourably with those reported previously for other schemes by Tielsch et al., who re-evaluated the results and calculated the \( \kappa \)-statistics for a number of studies that used various grading schemes for trachoma (4). For studies using the McCallan classification of trachoma, the \( \kappa \)-statistic for inter-observer variation ranged from 0.37 to 0.47 and those for individual observers from 0.48 to 0.63. Also, for the WHO grading scheme (5), or its modified versions, the \( \kappa \)-statistic for inter-observer variation ranged from 0.28 to 0.80, while the variation for individual observers was 0.50 to 0.94.

Initial studies using the grading system were conducted by ophthalmologists who had extensive field experience dealing with trachoma (1). Analysis of the results showed that the \( \kappa \)-statistics for both inter-observer and individual observer variation were greater than 0.68 for all signs, with the exception of TI, which was subsequently redefined. The results we have reported here not only confirm the high reproducibility of the new grading scheme, including that of the previously problematic sign TI, but also demonstrate that it can be readily learned by non-specialist health workers. We believe the scheme has much to recommend it and that it will find wide use in field studies.

**ACKNOWLEDGEMENTS**

The authors thank the Tanzanian National Prevention of Blindness Committee for their support, as well as all members of the Kongwa Eye Project in particular Sater N. N. Lyimo; Helen Keller International, New York; and the Christophel Blindenmission, Federal Republic of Germany.
RÉSUMÉ

TRACHEOMÉ: ÉVALUATION D'UN NOUVEAU SYSTÈME DE COTATION EN RÉPUBLIQUE-UNIE DE TANZANIE

Un nouveau système simplifié de cotation du trachome, basé sur la présence ou l'absence de cinq signes clés de la maladie, a été évalué en République-Unie de Tanzanie. Deux ophtalmologistes confirmés et une infirmière en ophtalmologie ont participé aux essais destinés à évaluer la reproductibilité de la méthode et la formation nécessaire. Les premières séances de formation comprenaient l'observation de diapositives et un examen croisé des malades. Cet essai a révélé une bonne uniformité de la cotation pour chaque observateur (le kappa (κ) moyen allant de 0,74 à 1,0) mais une assez grande variation d'un observateur à l'autre. D'autres séances de formation clinique avec observations croisées plus intensives ont beaucoup amélioré la variation entre observateurs (κ de 0,79 à 0,95). Les ophtalmologistes confirmés comme les infirmières ont rapidement enregistré ce système de cotation, qui devrait trouver de nombreuses applications dans les enquêtes sur le terrain.

REFERENCES