

## CLINICAL SCIENCE

## Reliability of expert interpretation of retinal photographs for the diagnosis of toxoplasma retinochoroiditis

M R Stanford, L Gras, A Wade, R E Gilbert

*Br J Ophthalmol* 2002;**86**:636–639

See end of article for authors' affiliations

Correspondence to:  
Dr Ruth Gilbert, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK;  
[r.gilbert@ich.ucl.ac.uk](mailto:r.gilbert@ich.ucl.ac.uk)

Accepted for publication  
29 October 2001

**Aims:** To measure agreement and estimate sensitivity and specificity of uveitis experts' interpretation of retinal photographs for the diagnosis of toxoplasma retinochoroiditis.

**Methods:** The authors collated 96 retinal photographs from patients presenting with symptomatic posterior uveitis to ophthalmology clinics within a defined geographical area. Five uveitis experts independently ranked each photograph as definite, probable, possible, or not toxoplasma retinochoroiditis. They measured interobserver agreement based on an intraclass correlation coefficient and estimated sensitivity and specificity for each observer using a maximum likelihood model

**Results:** The intraclass correlation coefficient for all five observers was 0.62 (95% CI: 0.54 to 0.69). Estimates of sensitivity and specificity for individual observers ranged from 71% to 96% and 58% to 100% respectively when appearances were dichotomised as definite or probable toxoplasma retinochoroiditis, versus possible or not.

**Conclusion:** Agreement between uveitis experts on the interpretation of retinal photographs was moderate to good, but the estimates of sensitivity and specificity of their interpretation for the diagnosis of toxoplasma retinochoroiditis were highly variable. Assuming that interpretation of photographs is similar to real time ophthalmoscopy, clinicians need to beware that treatment decisions made on the basis of a single assessment will include patients without toxoplasma retinochoroiditis and will miss patients with the disease.

The diagnosis of recurrent toxoplasma retinochoroiditis is based on the finding of focal retinitis associated with a retinochoroidal scar. Presentation to ophthalmologists is usually prompted by symptoms of sudden visual impairment or pain.<sup>1</sup> Given these signs and symptoms, the most likely differential diagnoses are ocular tuberculosis, viral retinitis, sarcoidosis, brucellosis, syphilis, serpiginous choroiditis, and fungal infections.<sup>2</sup> Additional information from the clinical history or the results of invasive tests is rarely definitive. For example, a history of similar previous episodes, reported in 50% of patients with acute toxoplasma retinochoroiditis,<sup>1</sup> makes the diagnosis more likely but not certain as recurrent episodes can also occur with sarcoidosis or serpiginous choroiditis.<sup>1</sup> On the other hand, a history of ocular or neurological signs in early childhood associated with congenital toxoplasmosis is rare but makes the diagnosis almost certain.<sup>1</sup> Neither systemic signs of tuberculosis, sarcoidosis, or brucellosis nor serum or skin tests for these diagnoses are highly specific.<sup>3</sup> A positive toxoplasma specific antibody result does not add important information but a negative result rules out toxoplasma retinochoroiditis. Finally, more specific tests such as polymerase chain reaction (PCR) detection of toxoplasma DNA in ocular fluids require invasive sampling and cannot be justified as a routine investigation.

The diagnosis of toxoplasma retinochoroiditis and the decision to treat therefore depend on the clinician's interpretation of the appearance of the retina and on the baseline risk of toxoplasmosis and alternative diagnoses. As there is no reference standard, assessment by uveitis experts is regarded as the best method for diagnosing or excluding toxoplasma retinochoroiditis. We evaluated the performance of assessments made by five uveitis experts by measuring agreement and by modelling the sensitivity and specificity of their categorisation of retinal photographs.

## METHODS

## Study population

We collated 96 fundal photographs from patients presenting to ophthalmology clinics within a defined geographical area

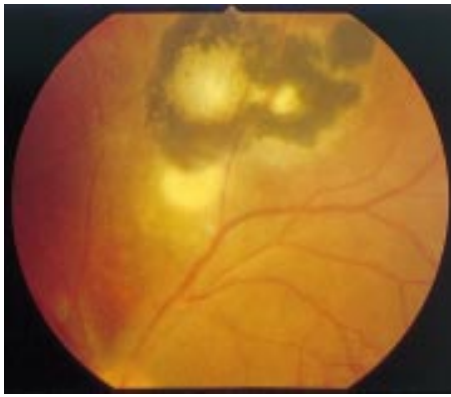
(Birmingham, south London, or north east London). Of these, 66 were from patients with suspected acute symptomatic toxoplasma retinochoroiditis reported to a central surveillance scheme<sup>1</sup> and, in two patients, one photograph was included for each eye. The remaining 28 photographs were from patients with alternative diagnoses based on clinical follow up and laboratory investigations.

## Performance of expert assessors

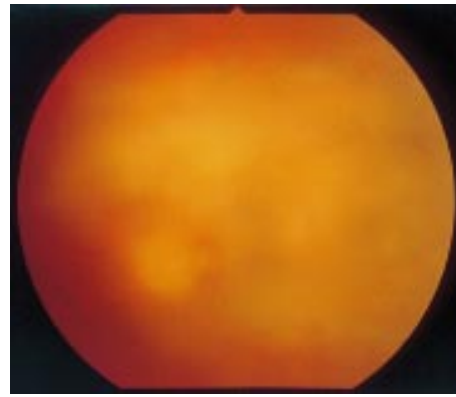
As there is no reference standard for toxoplasma retinochoroiditis, we could not definitively diagnose or exclude the disease. We therefore determined the performance of expert assessments by measuring agreement between assessors. Sensitivity and specificity were estimated for each assessor using a statistical model. Finally, we used the estimated sensitivity and specificity to predict the risk of mislabelling retinal photographs as toxoplasma retinochoroiditis, and of missing the disease in photographs categorised as unaffected.

## Agreement between assessors

Five uveitis experts were asked to independently classify the retinal photographs into four categories without any additional information: definitely, probably, possibly, or not toxoplasma retinochoroiditis. We calculated the level of agreement between five assessors using an intraclass correlation coefficient.<sup>4</sup> This coefficient expresses the variability attributable to true differences between retinal photographs as a proportion of the total variability and gives similar results to a quadratic weighted kappa statistic.<sup>5</sup> The intraclass correlation coefficient ranges from 0 and 1, a value of 1 indicates complete agreement. Residual variability (1 – intraclass correlation coefficient) is due to true variation between assessors and measurement error. Approximate confidence intervals were calculated according to the method described by Fleiss and Shout.<sup>6</sup>



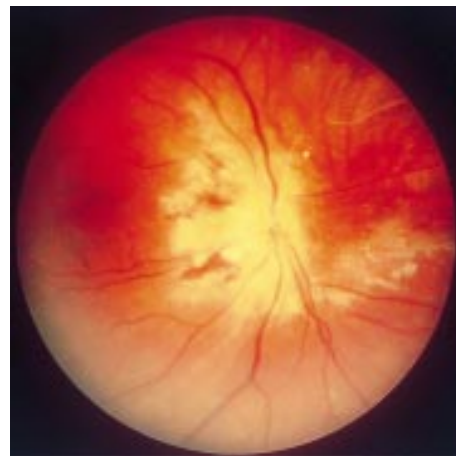
**Figure 1** Retinal photograph used in the study on which all assessors agreed. Definite toxoplasma retinochoroiditis.



**Figure 3** Retinal photograph used in the study on which all assessors agreed. Possible toxoplasma retinochoroiditis.



**Figure 2** Retinal photograph used in the study on which all assessors agreed. Probable toxoplasma retinochoroiditis.



**Figure 4** Retinal photograph used in the study on which all assessors agreed. Not toxoplasma retinochoroiditis.

#### Estimation of sensitivity and specificity

We used a statistical method,<sup>7</sup> which made probabilistic estimates for the true disease prevalence, to estimate sensitivity and specificity for each assessor. As calculation of sensitivity and specificity requires binary data, categories were dichotomised as definite or probable toxoplasma retinochoroiditis versus possible or not. The former group was considered to represent patients likely to be treated without further investigations, while in the latter group, alternative diagnoses would be considered. The model took account of the source of the photographs (central surveillance scheme or alternative diagnoses) and the different prevalences of toxoplasma retinochoroiditis likely in photographs from these two sources. Estimates obtained from the model were based on the assumptions that misclassification errors were independent between assessors and between retinal photographs, conditional on the true disease status.<sup>7</sup> Confidence intervals were calculated using the profile likelihood method.<sup>8</sup>

#### Predicting the risk of mislabelling retinal appearances

We used the estimated sensitivity and specificity for assessor 3 and Bayes's formula<sup>9</sup> to produce illustrative calculations of the risk of mislabelling retinal appearances in: (a) a community setting, where the reported prevalence of toxoplasma retinochoroiditis was 35% (95% CI: 18% to 53%); and (b) a uveitis clinic, reported prevalence 8% (4%–13%).<sup>10</sup> We assumed the spectrum of disease to be similar in both settings. The post-assessment odds of toxoplasma retinochoroiditis in photographs classified as definite or probable toxoplasma retinochoroiditis can be calculated as the pretest odds of toxoplasma

retinochoroiditis  $\times$  (sensitivity/(1 – specificity)). Odds can then be converted to a risk estimate (= odds/1 + odds). The risk of mislabelling retinal appearances as toxoplasma retinochoroiditis is 1 – the post-assessment risk of toxoplasma retinochoroiditis.

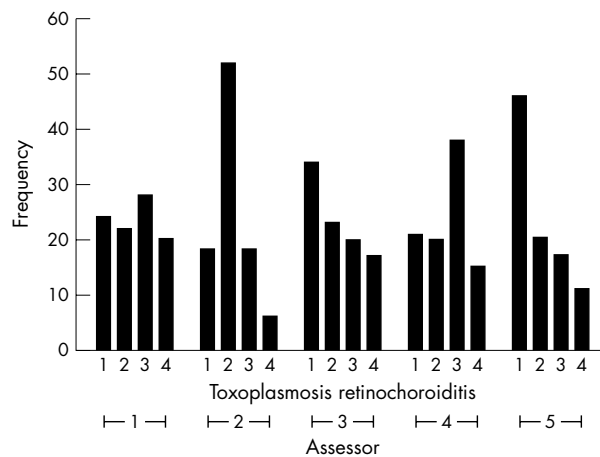
## RESULTS

### Agreement between experts

Figures 1–4 provide examples of photographs that were classified as definite, probable, possible, and not toxoplasma retinochoroiditis respectively by all five assessors. Of the 96 photographs examined, two were excluded from the analysis owing to lack of categorisation by one or more assessor. The frequency of categorisation as definitely, probably, or possibly toxoplasma retinochoroiditis varied between assessors, but all assessors categorised fewest photographs as not toxoplasma retinochoroiditis (Fig 5). The intraclass correlation coefficient for all assessors across the four categories was 0.62 (95% CI: 0.54 to 0.69). The intraclass correlation coefficients for pairs of assessors across the four categories varied between 0.53 and 0.73 (Table 1). When categories were dichotomised as definitely/probably versus possibly/not toxoplasma retinochoroiditis the intraclass correlation coefficient for all five assessors was 0.46 (95% CI: 0.37 to 0.54).

### Estimation of sensitivity and specificity

Point estimates for the sensitivity for toxoplasma retinochoroiditis, based on categories of definitely or probably toxoplasma retinochoroiditis versus probably or not, ranged from



**Figure 5** Histogram showing frequency of categorisation of photographs by the five assessors as definite (1), probable (2), possible (3), or not toxoplasma retinochoroiditis (4).

**Table 1** Intraclass correlation coefficients for all four categories for each pair of assessors

Assessors	Assessors				
	1	2	3	4	5
1		0.58	0.69	0.70	0.69
2	0.58		0.54	0.53	0.63
3	0.69	0.54		0.68	0.73
4	0.70	0.53	0.68		0.66
5	0.69	0.63	0.73	0.66	

71% to 96% (confidence intervals shown in Table 2). Point estimates for specificity ranged from 58% to 100%.

### Predicting the risk of mislabelling retinal appearances

In a community setting (prevalence of toxoplasma uveitis 34%), the risk of mislabelling retinal appearances as toxoplasma retinochoroiditis in photographs from patients without the disease would be 28%. Conversely, the diagnosis would be missed in 7% of photographs labelled as possible or not toxoplasma retinochoroiditis.

In a uveitis referral clinic (reported prevalence of toxoplasma retinochoroiditis 8%<sup>10</sup>), the estimated risk of mislabelling retinal appearances as toxoplasma retinochoroiditis would be 70%. The risk of missing the diagnosis in photographs labelled as possible or not toxoplasma retinochoroiditis would be 1%. These risks should be regarded as illustrative as the confidence intervals around the sensitivity and specificity are wide and require several assumptions.

## DISCUSSION

We found moderate agreement between uveitis experts' attribution of retinal appearances to toxoplasma retinochoroiditis

or not which resulted in substantial interobserver variation in the estimated sensitivity and specificity. Our calculations of the predicted risks of mislabelling retinal appearances indicate that, in low prevalence settings, uveitis experts are more likely to diagnose toxoplasma retinochoroiditis and possibly treat patients who do not have the disease than to miss patients who have the disease. The consequences of these errors depend on the harm in delayed or no treatment in patients with toxoplasma retinochoroiditis and inappropriate treatment in patients with alternative diagnoses.

Toxoplasma treatment consists of antibiotics and high dose systemic corticosteroids for at least 3–4 weeks.<sup>11</sup> Treatment aims to reduce the duration of acute symptoms and reduce the risk of permanent visual impairment.<sup>2</sup> However, there is no evidence from randomised controlled trials that antibiotic or steroid treatment of toxoplasma retinochoroiditis, compared with no treatment, improves visual symptoms or acuity<sup>12–13</sup> and both treatments are associated with major adverse effects. In addition, steroid treatment may exacerbate tuberculosis and the rarer infectious causes of retinitis such as viruses, brucellosis, or fungi.<sup>14</sup> In Europe, the most common differential diagnosis is sarcoidosis,<sup>15–16</sup> which is usually treated with corticosteroids, but treatment may nevertheless delay a definitive diagnosis.

One previous study reported clinician agreement on the interpretation of quiescent toxoplasma retinal lesions,<sup>17</sup> but we know of no previous reports on the performance of retinal examination for toxoplasma retinochoroiditis in clinical practice. Three limitations of our study mean that our estimate of the performance of clinical interpretation of retinal photographs is likely to overestimate agreement, sensitivity, and specificity. Firstly, the assessors were all uveitis experts. Non-experts might be expected to show more disagreement.<sup>18</sup> Secondly, the model used to estimate sensitivity and specificity required the assumption that misclassification between assessors was unrelated, conditional on the true disease status. In practice, classification is likely to be part of a continuum and may be correlated due to qualitatively similar mistakes made by assessors with similar background and training.<sup>7</sup> Thirdly, although photographs from patients with suspected acute toxoplasma retinochoroiditis were derived from a population based study, the remaining photographs were from patients with retinochoroiditis because of alternative diagnoses. Selection of the latter photographs may have favoured those with unambiguous appearances of an alternative diagnosis, thereby introducing spectrum bias.<sup>19</sup>

Finally, generalisation of our results, based on retinal photographs, to real time, biomicroscopic ophthalmoscopy, is a matter of judgment. Studies comparing ophthalmoscopy with assessments of photographs of diabetic retinopathy reported moderate to good agreement<sup>18</sup> but no such comparisons have been published for toxoplasma retinochoroiditis. Agreement may be increased given the clinical information available in practice, but this has not been assessed. A further limitation of our results is that they reflect interpretation of retinal photographs taken from patients in the United Kingdom. Further studies are required to evaluate clinician assessments in populations with a different spectrum of

**Table 2** Estimated sensitivity and specificity (%) for each assessor comparing categories of definite/probable versus possible/not toxoplasma retinochoroiditis. (95% confidence interval)

	Assessor				
	1	2	3	4	5
Sensitivity	75 (62 to 86)	95 (87 to 99)	88 (77 to 95)	71 (58 to 83)	96 (89 to 99)
Specificity	91 (79 to 98)	58 (42 to 92)	82 (66 to 93)	100 (94 to 100)	71 (53 to 86)

differential diagnoses and, possibly, a different spectrum of toxoplasma retinochoroiditis, as may be the case in endemic areas such as south Brazil.<sup>17</sup> Finally, the criteria used by each of the assessors to classify toxoplasma retinochoroiditis were not investigated and may have differed. Agreement would be improved if standardised criteria could be developed and adopted by ophthalmologists.

Unless a readily usable reference test for toxoplasma retinochoroiditis becomes available, future evaluations of clinician assessments should involve comparison with a clinical reference standard which measures the evolution of retinochoroiditis in the long term.<sup>1</sup> In addition, the methods need to take account of the fact that retinal assessment would be part of the test and reference standard.<sup>20</sup> Such long term follow up studies would provide more valid estimates of the performance of clinician assessments together with important information on the baseline risk of toxoplasma retinochoroiditis and the differential diagnoses.

## ACKNOWLEDGEMENTS

The authors thank Professor A Bird, Dr A Rothova, Professor G Dutton, Dr A Brezin, and Professor G Holland for acting as independent assessors in this study. Professors P Murray and S Lightman provided the retinal photographs of alternative diagnoses. The work was supported by the Iris Fund for Prevention of Blindness.

## Authors' affiliations

**M R Stanford**, Medical Eye Unit, St Thomas's Hospital, London, UK  
**L Gras, A Wade, R E Gilbert**, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, London, UK

## REFERENCES

- Gilbert RE**, Dunn D, Lightman S, *et al.* Incidence of symptomatic toxoplasma eye disease: aetiology and public health implications. *Epidemiol Infect* 1999;**123**:283–9.
- Holland GN**, O'Connor GR, Belfort-Junior R, *et al.* Toxoplasmosis. In: Pepose JS, Holland GN, Wilhelmus KR, eds. *Ocular infection and immunity*. St Louis: Mosby, 1996:1183–223.
- Power W**, Rodriguez A, Pedroze-Seres M. The value of combined serum angiotensin converting enzyme and gallium scan in diagnosing ocular sarcoidosis. *Ophthalmology* 1995;**102**:2007–11.
- Streiner DL**, Norman GR, eds. Reliability. In: *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford Medical Publications, 1995:104–27.
- Fleiss JL**, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurements* 1973;**33**:613–19.
- Fleiss J L**, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika* 1978;**43**:259–62.
- Walter SD**. Measuring the reliability of clinical data: the case for using three observers. *Rev Epidemiol Sante Publique* 1984;**32**:206–11.
- Cox DR**, Oakes D, eds. *Analysis of survival data*. London: Chapman and Hall, 1984:51–2.
- Jaeschke R**, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994;**271**:703–7.
- McCannel CA**, Holland GN, Helm CJ, *et al.* Causes of uveitis in the general practice of ophthalmology. *Am J Ophthalmol* 1996;**121**:35–46.
- Engstrom RE Jr**, Holland GN, Nussenblatt RB, *et al.* Current practices in the management of ocular toxoplasmosis. *Am J Ophthalmol* 1991;**111**:601–10.
- Acers TE**. Toxoplasmic retinochoroiditis: a double blind therapeutic study. *Arch Ophthalmol* 1964;**71**:58–62.
- Rothova A**, Meenken C, Buitenhuis HJ, *et al.* Therapy for ocular toxoplasmosis. *Am J Ophthalmol* 1993;**115**:517–23.
- Stanbury RM**, Graham EM. Systemic corticosteroid therapy—side effects and their management. *Br J Ophthalmol* 1998;**82**:704–8.
- Paivonsalo Hietanen T**, Tuominen J, Vaahoranta Lehtonen H, *et al.* Incidence and prevalence of different uveitis entities in Finland. *Acta Ophthalmol Scand* 1997;**75**:76–81.
- Pezzi PP**, Accarinti M, Carva M, *et al.* Endogenous uveitis: an analysis of 1,417 cases. *Ophthalmologica* 1996;**210**:234–8.
- Glasner PD**, Silveira C, Kruszon Moran D, *et al.* An unusually high prevalence of ocular toxoplasmosis in southern Brazil. *Am J Ophthalmol* 1992;**114**:136–44.
- Cook HL**, Stanford MR, Marshall J. Reference standards in diabetic retinopathy screening. *Lasers and Light in Ophthalmology* 1999;**9**:15–27.
- Ransohoff DF**, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;**299**:926–30.
- Joseph L**, Gyorkos TW. Inferences for likelihood ratios in the absence of a gold standard. *Medical Decision Making* 1996;**16**:412–17.

## Contributors please note:

Communications from **all countries except the UK and Republic of Ireland** should be sent to Professor C Hoyt, Editor, *British Journal of Ophthalmology*, University of California, Department of Ophthalmology, 10 Kirkham Street, K 301, San Francisco, CA 94143-0730, USA (tel: 001 415 502 6871; fax: 001 415 514 1521).

Manuscripts from the **UK and the Republic of Ireland** should be sent to Professor Andrew Dick, UK Editor, *British Journal of Ophthalmology*, Division of Ophthalmology, University of Bristol, Lower Maudlin Street, Bristol BS1 2LX (tel: +44 (0) 117 929 4496; fax: +44 (0) 117 929 4607).



## Reliability of expert interpretation of retinal photographs for the diagnosis of toxoplasma retinochoroiditis

M R Stanford, L Gras, A Wade, et al.

*Br J Ophthalmol* 2002 86: 636-639

doi: 10.1136/bjo.86.6.636

---

Updated information and services can be found at:

<http://bjo.bmj.com/content/86/6/636.full.html>

---

*These include:*

### References

This article cites 15 articles, 3 of which can be accessed free at:

<http://bjo.bmj.com/content/86/6/636.full.html#ref-list-1>

Article cited in:

<http://bjo.bmj.com/content/86/6/636.full.html#related-urls>

### Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

---

### Topic Collections

Articles on similar topics can be found in the following collections

[Choroid](#) (434 articles)

---

### Notes

---

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>