Diagnostic accuracy of microbial keratitis with in vivo scanning laser confocal microscopy
Scott Hau, John Dart, Minna Vesaluoma, Dipak Parmar, Ilse Claerhout, Kanom Bibi, Frank Larkin

To cite this version:
Diagnostic accuracy of microbial keratitis with in vivo scanning laser confocal microscopy

Scott C Hau,1 John K G Dart,1 Minna Vesaluoma,1 Dipak N Parmar,2 Ilse Claerhout,3 Kanom Bibi,1 Daniel F P Larkin1

1 NIHR Biomedical Research Centre in Ophthalmology, Moorfields Eye Hospital and Institute of Ophthalmology, London, UK, EC1V 2PD
2 Dept of Ophthalmology, Whipps Cross Hospital, London, UK, E11 1NR
3 Ghent University Hospital, De Pintelaan 185, 9000 Gent, Belgium

Word count: 2500

Corresponding author: Scott Hau, Research fellow, Moorfields Eye Hospital NHS Foundation Trust, 162 City Road, London, UK, EC1V 2PD
Email: scott.hau@moorfields.nhs.uk
Tel: +447905 295888
Fax: +44207 5662471

Key words: Confocal microscopy, diagnostic accuracy, acanthamoeba keratitis, fungal keratitis, Microsporidia, Nocardia
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in British Journal of Ophthalmology and any other BMJPGL products to exploit all subsidiary rights, as set out in our licence.
ABSTRACT

Aims
To determine the accuracy of diagnosing microbial keratitis by masked medical and non-medical observers using the Heidelberg Retina Tomograph II / Rostock Cornea Module in vivo confocal microscope.

Methods
Confocal images were selected for 62 eyes with culture or biopsy proven infections. The cases comprised of 26 Acanthamoeba, 12 fungus, 3 Microsporidia, 2 Nocardia, and 19 bacterial infections (controls). The reference standard for comparison was a positive tissue diagnosis. These images were assessed on two separate occasions by 4 observers who were masked to the tissue diagnosis. Diagnostic accuracy indices, Kappa (κ) statistic and percentage agreement values were calculated. The Spearman correlation coefficient (r_s) was calculated for the number of correct diagnoses versus duration of disease.

Results
The highest sensitivity and specificity values were 55.8% and 84.2%, and the lowest sensitivity and specificity values were 27.9% and 42.1%. The highest positive and negative likelihood ratios were 2.94 and 0.59, respectively. Agreement values were: fair to moderate (κ, 0.22-0.44) for reference standard versus observer diagnosis, moderate to good in intra-observer variability (repeatability, κ 0.56-0.88), and poor to moderate in inter-observer variability (reproducibility, κ, 0.15-0.47). The correct
diagnosis was associated with duration of disease for *Acanthamoeba* keratitis ($r_s = 0.60$, $p = 0.001$).

**Conclusions**

The diagnostic accuracy of microbial keratitis by confocal microscopy is dependent on observer experience. Intra-observer repeatability was better than inter-observer reproducibility. Difficulty in distinguishing host cells from pathogenic organisms limits the value of confocal microscopy as a stand-alone tool in diagnosing microbial keratitis.
INTRODUCTION

Difficulties in clinical and microbiological diagnosis are one of the major problems in the management of microbial keratitis particularly when caused by protozoa (Acanthamoeba and Microsporidia), fungi or filamentary bacteria. Diagnosis of these pathogens is difficult as they often take days or weeks to grow in culture and, in any case, culture is insensitive with culture positive rates rarely exceeding 60%. [1]

Although culture is still the primary diagnostic tool in tertiary referral centres it is not widely available to many patients because of limited resources.

The confocal microscope allows detailed in vivo analysis of normal [2] and pathological corneas. In patients with presumed corneal infection, it is used in diagnosis and in examination of the extent of involvement of tissue by infection and associated inflammation. All published studies have been directed at diagnosis and a number have shown both white light and laser confocal microscopy to be effective in diagnosing Acanthamoeba, [3-5] fungal, [6-8] Nocardia [9] and Microsporidia keratitis. [10] However, these studies only present case series or reports and there are limited published data on evaluating the diagnostic accuracy of confocal microscopy. Two recent studies have found high sensitivity and specificity values for diagnosing fungal keratitis (FK) and Acanthamoeba keratitis (AK) with the Confoscan 3.0 (Nidek Technology, Padova, Italy). [11,12] However, factors such as observer or selection bias, the absence of masking the observers from the microbiological diagnosis, and lack of appropriate controls may have resulted in overestimates of the sensitivity and specificity values. Although experience in interpreting confocal keratitis images is essential, the accuracy of diagnosing microbial keratitis by clinicians with differing
levels of confocal microscopy experience and the potential of using trained
technicians in interpreting images have not previously been assessed. These are
important considerations in evaluating this technique. The aim of this study was to
examine the diagnostic accuracy of microbial keratitis with the Heidelberg Retina
Tomograph II / Rostock Cornea Module (HRT II/RCM) in vivo confocal microscope,
as a stand-alone tool, by trained medical and non-medical observers with differing
confocal microscopy experience.

MATERIALS AND METHODS

Patients

This study was approved by the Research & Ethics Committee of Moorfields Eye
Hospital and it adhered to the tenets of Declaration of Helsinki. We retrospectively
reviewed the case notes of a consecutive series of microbial keratitis patients who had
had both corneal cultures or corneal biopsy and confocal microscopy (n=105) from
January 1, 2005 to January 4, 2008. These cases were both those refractory to
conventional treatment and those with unusual clinical features such as perineural
infiltrates and ring infiltrates. Patients were either referred from Moorfields
Emergency Department or from other institutions. Of the 105 cases, 62 culture or
biopsy positive cases (62 eyes) were identified: 26 Acanthamoeba, 11 fungus, 1
fungus and bacteria, 3 Microsporidia, 2 Nocardia, and 19 bacteria. Bacteria were used
as controls because they are normally too small to detect with confocal microscopy,
[9,13] therefore the case which was culture positive for both fungus and bacteria was
classified as a fungal keratitis for the purposes of the study. We did not classify
Nocardia as controls because they are filamentous bacteria and can form filamentous structures that are large enough to be distinguished by confocal microscopy.[9] Empirical treatments started prior to assessment in this study included topical antimicrobial agents and topical steroids for presumed herpes, bacterial or keratitis of unknown cause, respectively. Irrespective of the referring diagnosis, all patients had undergone a full clinical examination by a corneal specialist and repeat corneal scraping for culture and confocal microscopy on the same day. If the scraping was culture negative, and the keratitis progressive, then a corneal biopsy was later performed. Exclusion criteria were culture or biopsy negative keratitis cases, and patients who declined to have confocal microscopy or a corneal culture as part of their clinical investigation. The reference standard for this study was a diagnosis either by isolation on culture of a corneal scraping or histological diagnosis on a corneal biopsy; other ancillary culture sources such as contact lens case and solutions were not used. The clinical outcomes were recorded for all the patients in the study and were consistent with the diagnosis based on culture or histology therefore it is unlikely, but possible, that there was unrecognised polymicrobial infections which may have been identified on confocal but not by culture or biopsy. We followed the Standards for Reporting of Diagnostic accuracy (STARD) initiative in conducting this study.[14]

Culture and biopsy methods

Corneal scrapings for microbial culture were inoculated on the following media: blood agar, Sabouraud’s dextrose agar (fungi), Robertson’s cooked meat (anaerobic bacteria), Escherichia coli-seeded non-nutrient agar (Acanthamoeba), brain heart infusion (fastidious organisms, fungi) and Lowenstein-Jensen (mycobacteria, Nocardia). Scrapings were smeared on sterile glass slides for Gram and Giemsa stains.
All microbiological investigations were undertaken independently in an external laboratory. For biopsy a superficial lamellar disc of the affected cornea was trephined under local anaesthetic to provide a further specimen for microbiology and histopathological staining.

Confocal microscopy measurement protocol

In vivo confocal microscopy was performed on all 62 eyes by a single experienced observer (SH) with the HRT II / RCM (Heidelberg Engineering GmbH, Dossenheim, Germany) confocal microscope following a Standard Operating Procedure as follows. A sterile Tomocap (Heidelberg Engineering GmbH, Dossenheim, Germany) was mounted over the objective of the microscope (Zeiss, x 63), and Polyacrylic acid 0.2% (Viscottears, Novartis) was used as a coupling agent between the cap and the lens objective. Topical anaesthetic (Proxymetacaine hydrochloride 0.5%, Chauvin) and Carmellose sodium 1% (Celluvisc, Allergan) was instilled into both eyes to provide comfort and act as a coupling fluid between the front of the Tomocap and the cornea. Options for image acquisition include section (a single image at a particular depth), volume (a series of images over 60µm depth) and sequence scans (a video sequence at a particular depth). The volume scan option was selected for image acquisition because it allowed the capturing of large number of images over a short space of time. The central region of the corneal ulcer or corneal infiltrate was scanned first followed by the top, left, bottom and right margin of the lesion. At each point, the epithelial layer of the affected area was scanned first and the focal plane of the microscope adjusted until the whole depth of the ulcer or infiltrate had been scanned. When there was more than one infiltrate, the same scanning sequence was repeated for each infiltrate. The wavelength of the laser employed in the HRT II / RCM is 670 nm and
each standard 2 dimensional image consists of 384 x 384 pixels covering an area of 400 μm x 400 μm. The axial resolution is 7.6μm; compared to other instruments such as the Tandem scanning microscope (9μm) and ConfoScan 4 (29μm).[15]

**Image selection**

The confocal images of all the scans were reviewed by two experienced confocal microscopist (SH and JD). In diagnosing keratitis, a considerable amount of time is often needed to find an image that would yield sufficient information to be able to identify the organism. This is due to masking of the organisms by the cellular inflammatory response and that they seldom distribute evenly within the cornea during active infection. Therefore, to ensure all our observers had the maximum likelihood in diagnosing the type of keratitis, the best quality 384 x 384 pixel resolution digital image indicating clearly the culture proven pathogen from the corneal ulcer or infiltrate was selected and exported onto Microsoft Power Point® (Microsoft Corp., Redmond, WA, USA). These included those of Acanthamoeba - round single or double walled hyper-reflective objects (~10-20 μm) consistent with *Acanthamoeba* cysts,[4,5] fungus - linear irregular branching hyper-reflective objects consistent with fungal hyphae,[6,7] *Microsporidia* - small round hyper-reflective deposits (~ 2 μm) located in between keratocytes,[10] *Nocardia* - small branching filamentous structures within the corneal stroma,[9] and bacteria (control) – a mixture of inflammatory cells.
Intra- and inter-observer agreement

All digital images were assessed prospectively in the same standard fashion in the Reading Centre at Moorfields Eye Hospital by 4 observers (3 ophthalmologists and 1 medical technician) with differing levels of experience in assessing keratitis on confocal microscopy as follows. Of the 3 ophthalmologists, observer A had 6 years of experience in assessing microbial keratitis with confocal microscopy, observer B, 10 years of experience in confocal microscopy but not keratitis, and observer C, 6 months of experience in assessing keratitis with confocal microscopy. Observer D was a medical technician who had 2 years of experience in performing confocal microscopy using the HRT II / RCM and analysing keratitis images but with no experience in the clinical appearance and treatment of different types of keratitis. To ensure each observer was familiar with the image appearance of different cell types obtained from the HRT II / RCM confocal microscope, examples of both normal cellular morphology and the standard images of different pathogens were shown in a presentation before their assessment. In addition, a series of five recent articles on diagnosing keratitis with the HRT II / RCM [4,5,7,9,10] were given to each observer to read 2 weeks prior to their scheduled assessment date.

The confocal images were viewed in random order and assigned an identification number from 1 to 62. To ensure that there was masking between observers, the order of viewing the images were randomised by computer before being assessed by the next observer on a different day. No clinical details regarding each case were made available to the observers. Each observer assessed the series of images in a masked fashion on slide show in Microsoft Powerpoint® and recorded the diagnosis corresponding to one of the following categories: AK, FK, Microsporidia (MK),
Nocardia (NK) or bacterial keratitis (BK). A reference sheet showing the range of sizes of resident and inflammatory cells including epithelium and macrophages, and pathogenic cells e.g. diameter of Acanthamoeba cysts was given to each observer. Intra-observer variability (repeatability) was evaluated by asking each observer to reassess the images, randomised in a different order, three weeks later in the same standard fashion. Inter-observer variability (reproducibility) was assessed by determining the level of agreement in diagnosis between observers. Readings of all the digital images were collected on a standard pro-forma and analysed.

Data analysis

Data analysis was performed with SPSS V14.0 (SPSS Inc, Chicago, USA). We calculated sensitivity, specificity, positive and negative likelihood ratios (LR) for both image set for each observer. Positive LR predicts the probability of a positive test result in patients with disease compared to those who do not have the disease. Negative LR predicts the probability of a negative test in those who have the disease compared to those who do not. The level of agreement between the reference standard and different observers, and both intra and inter-observer variability were determined using Kappa (κ) statistic. The interpretation of κ statistic is as follows: ‘poor’ if κ ≤ 0.20, ‘fair’ if κ 0.21 – 0.40, ‘moderate’ if κ 0.41 – 0.60, ‘substantial’ if κ 0.61 – 0.80 and ‘good’ if κ > 0.80.[16] In addition, we also calculated percentage agreement values between reference standard and observers, within-observers, and between different observers. Spearman’s rank correlation coefficient (r_s) was used to determine the relationship between the number of correct diagnoses and the duration of disease for AK, FK and BK respectively. The duration of disease was defined as the time from symptom onset to presentation to the Corneal and External Disease Service at
Moorfields. A value of $P < 0.05$ was deemed statistically significant. MK and NK were excluded from this analysis because the numbers were too small.

RESULTS

The reference standard consisted of 52 culture positive cases from corneal scrapings and 10 histopathologically confirmed cases on corneal biopsy. Sensitivity, specificity and likelihood ratio values for each observer are shown in Table 1.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive LR (95% CI)</th>
<th>Negative LR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st assessment</td>
<td>55.8 (47.7-63.2)</td>
<td>57.9 (39.5-74.7)</td>
<td>1.32 (0.79-2.49)</td>
<td>0.76 (0.49-1.33)</td>
</tr>
<tr>
<td>2nd assessment</td>
<td>55.8 (47.6-61.8)</td>
<td>73.7 (55.0-87.3)</td>
<td>2.12 (1.06-4.87)</td>
<td>0.60 (0.44-0.95)</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st assessment</td>
<td>53.4 (45.4-58.8)</td>
<td>78.9 (60.5-91.0)</td>
<td>2.53 (1.15-6.51)</td>
<td>0.59 (0.45-0.90)</td>
</tr>
<tr>
<td>2nd assessment</td>
<td>46.5 (38.6-50.9)</td>
<td>84.2 (66.4-94.2)</td>
<td>2.94 (1.15-8.82)</td>
<td>0.64 (0.52-0.93)</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st assessment</td>
<td>27.9 (20.5-32.3)</td>
<td>84.2 (67.5-94.2)</td>
<td>1.77 (0.63-5.56)</td>
<td>0.86 (0.72-1.18)</td>
</tr>
<tr>
<td>2nd assessment</td>
<td>27.9 (20.4-33.2)</td>
<td>78.9 (61.9-90.9)</td>
<td>1.32 (0.54-3.66)</td>
<td>0.91 (0.74-1.29)</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st assessment</td>
<td>44.2 (36.8-52.3)</td>
<td>42.1 (25.3-60.5)</td>
<td>0.76 (0.49-1.33)</td>
<td>1.33 (0.79-2.49)</td>
</tr>
<tr>
<td>2nd assessment</td>
<td>48.8 (40.7-56.3)</td>
<td>57.9 (39.6-74.8)</td>
<td>1.16 (0.67-2.23)</td>
<td>0.88 (0.59-1.50)</td>
</tr>
</tbody>
</table>

95% CI = 95% confidence intervals; LR = likelihood ratio
Positive LR = the probability of a positive test result in patients with disease compared to those who do not have the disease.
Negative LR = the probability of a negative test result in patients with disease compared to those who do not have the disease.

The highest sensitivity value obtained was 55.8% and the highest specificity value 84.2%. We found fair to moderate agreement between observers and reference standard ($\kappa$, 0.22-0.44), moderate to good agreement in intra-observer variability ($\kappa$, 0.56-0.88), and poor to moderate agreement in inter-observer variability ($\kappa$, 0.15-0.47),
Table 2. One observer (observer B) obtained the highest positive and lowest negative LR for diagnosing microbial keratitis. This observer also achieved the best overall kappa and percentage agreement values in diagnoses compared to reference standard, Table 3. The best inter-observer agreement (percentage agreement, 61.3-66.1%; \( \kappa \), 0.43-0.47) was between observer A and B, the two most experienced observers in the study. Observer C was the most repeatable (percentage agreement, 93.5%; \( \kappa \), 0.88) despite having the lowest kappa and percentage agreement values compared to reference standard, Tables 2 & 3.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Reference standard</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.36 (0.42)</td>
<td>0.56</td>
<td>0.43 (0.47)</td>
<td>0.23 (0.15)</td>
<td>0.28 (0.29)</td>
</tr>
<tr>
<td>B</td>
<td>0.44 (0.40)</td>
<td>0.43 (0.47)</td>
<td>0.76</td>
<td>0.40 (0.32)</td>
<td>0.18 (0.32)</td>
</tr>
<tr>
<td>C</td>
<td>0.24 (0.22)</td>
<td>0.23 (0.15)</td>
<td>0.40 (0.32)</td>
<td>0.88</td>
<td>0.24 (0.33)</td>
</tr>
<tr>
<td>D</td>
<td>0.27 (0.36)</td>
<td>0.28 (0.29)</td>
<td>0.18 (0.32)</td>
<td>0.24 (0.33)</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Reference standard = culture or biopsy proven cases
Values in parentheses = observers’ second assessment

<table>
<thead>
<tr>
<th>Observer</th>
<th>Reference standard</th>
<th>A (%)</th>
<th>B (%)</th>
<th>C (%)</th>
<th>D (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>56.5 (61.3)</td>
<td>70.9</td>
<td>61.3 (66.1)</td>
<td>45.2 (43.5)</td>
<td>45.2 (50)</td>
</tr>
<tr>
<td>B</td>
<td>61.3 (58.1)</td>
<td>61.3 (66.1)</td>
<td>85.5</td>
<td>64.5 (61.3)</td>
<td>37.1 (53.2)</td>
</tr>
<tr>
<td>C</td>
<td>45.2 (43.5)</td>
<td>45.2 (43.5)</td>
<td>64.5 (61.3)</td>
<td>93.5</td>
<td>40.3 (53.2)</td>
</tr>
<tr>
<td>D</td>
<td>43.5 (51.6)</td>
<td>45.2 (50)</td>
<td>37.1 (53.2)</td>
<td>40.3 (53.2)</td>
<td>72.6</td>
</tr>
</tbody>
</table>

Reference standard = culture or biopsy proven cases
Values in parentheses = observers’ second assessment
Complete agreement in diagnosis between all the observers and reference standard for both assessments were found in 3/26 (11.5%) cases of AK, 8/19 (42.1%) cases of BK, and 1/12 (8.3%) case of FK. In contrast, none of the observers identified *Acanthamoeba* in 5/26 (19.2%) cases, fungus in 4/12 (33.3%), and confused BK with other diagnoses in 2/19 (10.5%) cases. Observer B was the only one who managed to diagnose NK correctly in one case. The percentage correct diagnosis for the different types of keratitis is shown in Table 4. A breakdown of all the diagnoses for each observer for the different keratitis category is shown in the appendix.

<table>
<thead>
<tr>
<th>Reference standard Diagnosis</th>
<th>Observer A</th>
<th>Observer B</th>
<th>Observer C</th>
<th>Observer D</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acanthamoeba</em> (n=26)</td>
<td>19 (73.1)</td>
<td>11 (42.3)</td>
<td>4 (15.4)</td>
<td>12 (46.2)</td>
</tr>
<tr>
<td><em>Bacteria</em> (n=19)</td>
<td>10 (52.6)</td>
<td>15 (78.9)</td>
<td>15 (78.9)</td>
<td>8 (42.1)</td>
</tr>
<tr>
<td><em>Fungus</em>† (n=12)</td>
<td>1 (8.3)</td>
<td>5 (41.2)</td>
<td>5 (41.2)</td>
<td>5 (41.2)</td>
</tr>
<tr>
<td><em>Microsporidia</em> (n=3)</td>
<td>1 (33.3)</td>
<td>2 (66.7)</td>
<td>2 (66.7)</td>
<td>1 (33.3)</td>
</tr>
<tr>
<td><em>Nocardia</em> (n=2)</td>
<td>0</td>
<td>1 (50)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Percentage correct diagnosis was calculated by taking into account only those cases where each observer had made the correct diagnosis on both assessments.

† 1 *Alternaria*, 1 *Candida*, 10 *Fusarium*.

Figure 1 shows a series of images demonstrating the difference in appearance between correctly diagnosed versus incorrectly diagnosed cases. Figures 1 A and F show a case of late diagnosed AK versus early diagnosed AK; note the presence of inflammatory cells in the epithelium in early AK making distinction between host cells and *Acanthamoeba* cysts and trophozoites difficult, whereas in delayed diagnosed AK, single or clusters of cysts were seen in the stroma with minimal host immune and resident cells seen. The incorrectly diagnosed cases demonstrate the
difficulties in distinguishing host cells from pathogenic organisms, and *Nocardia* (Fig 217) from FK because of their similarity in appearance on confocal microscopy.

Figure 2 shows a plot between the number of correct diagnoses for Acanthamoeba, bacteria and fungal keratitis versus the duration of disease (days). The graph shows a moderate correlation between the number of correctly diagnosed cases and the duration of disease for AK ($r_s = 0.60$, $p = 0.001$), but not for BK ($r_s = 0.17$, $P = 0.49$) or FK ($r_s = -0.19$, $p = 0.57$), respectively. Therefore, the longer the duration of AK, the higher the likelihood that a correct diagnosis was made by the observers in grading the confocal images.

**DISCUSSION**

Acanthamoeba and fungus are uncommon causes of corneal infection for which early diagnosis is paramount because it yields better prognosis and reduces ocular morbidity.[17,18] Although the current reference standard for diagnosing microbial keratitis is corneal culture, the sensitivity varies because of numerous factors.[19]

The HRT II / RCM *in vivo* confocal microscope has been shown to be useful in diagnosing a range of pathogens but validation studies of this new technology are few. A recent review has reported the efficacy of diagnosing infections keratitis with confocal microscopy to be inconclusive, with the possible exception of AK.[20] Our results show moderate sensitivity and moderate to high specificity values in diagnosing microbial keratitis with the HRT II / RCM confocal microscope, whereas
both Kanavi et al[11] and Tu et al[12] found very high sensitivity (>90%) in
diagnosing AK and FK respectively with the Confoscan 3. Tu et al,[12] using multi-
test referencing standards, reported that when there are both clinical characteristics
and objective evidence of AK, the adjunctive usage of confocal microscopy exhibited
a sensitivity of 90.6% and specificity of 100%. In our study, we set out to evaluate the
diagnostic accuracy of confocal microscopy as a stand alone tool rather than a
supportive investigative technique, without the bias and influence of clinical findings.
Although assessing confocal images in the absence of clinical data does not reflect the
use of confocal microscopy in clinical settings, it is the only way to avoid bias when
analysing the images. Our inclusion criteria were based on culture positive cases
irrespective of confocal classification. Although we chose only one representative
image from each case this was the best available image for the organism that was
cultured from each case giving the observers the best opportunity to make a correct
confocal diagnosis; we believe that reviewing a series of images from each case
would either have made a correct confocal diagnosis more difficult or have had no
effect on the outcome. In addition, it allowed standardisation when viewing the
images so that all observers assessed the same number of images consecutively. The
absence of controls in the previous studies and the use of confocal ‘positive’ without
culture confirmation as a reference standard, or for the case definition[11,12], could
lead to selection bias and misdiagnosis resulting in an overestimation of sensitivity
values.[12,21,22] This is evident from our controls in which immune cells can often
be confused with AK cysts and vice versa leading to erroneous diagnosis.
Furthermore, ‘good’ confocal images have been illustrated in most published studies
to present findings without discussion of difficulties in analysing equivocal images.
We found fair to moderate agreement between reference standard and observer
diagnosis when a case mix of equivocal and unequivocal images were analysed by our observers. The rigorous criteria in our study design in regard to the use of masked observers and controls could explain why sensitivity values, even for the most experienced observer, were lower.

Another explanation for the very high sensitivity values reported in one previous study was the use of only one ophthalmology trained observer who, in addition to being unmasked to the clinical findings, was experienced in the use of confocal microscopy for keratitis diagnosis: this makes it difficult to extrapolate the results to what might be expected outside their centres.[12] To evaluate the potential of using this technology in clinics where an ophthalmologist with experience in confocal microscopy may not be available, our graders included 2 experienced ophthalmologists, an inexperienced ophthalmologist and an experienced technician. We found a two-fold difference in sensitivity between the most experienced and the least experienced observer indicating higher diagnostic accuracy with clinicians experienced in confocal microscopy. Our results indicate the sensitivity value with a trained technician, with no experience in the clinical appearance of different types of microbial keratitis, was better than an inexperienced medical observer but with a lower specificity value and positive LR. This raises the possibility of training non-medical personnel, in performing and analysing keratitis images. The highest positive LR and lowest negative LR was achieved by observer B who was experienced in confocal imaging of normal corneal anatomy and various pathological conditions other than microbial keratitis, indicating experience gained in other aspects of confocal microscopy improves the diagnostic outcome.
Intra-observer agreement (repeatability) was found to be moderate to good, indicative of good observer repeatability in grading the images irrespective of the accuracy of their diagnoses. Observer experience did not appear to improve intra-observer repeatability as the observer with the lowest sensitivity had the highest repeatability and vice versa. Inter-observer agreement (reproducibility) was poor to moderate between different observers because of factors such as observer experience and differences in techniques of classifying images by different observers. The two observers who had the highest sensitivity values also had the best inter-observer reproducibility, indicating experienced observers achieved a higher diagnostic accuracy and reproducibility than less experienced observers. Therefore, to improve reliability the same experienced operator should be employed if sequential imaging of a patient is required.

Our observers were able to diagnose AK more accurately than any other type of keratitis. The unique appearance of *Acanthamoeba* cysts on confocal microscopy and the higher number of AK compared to other conditions in our study might explain this outcome. However, AK was commonly confused with controls and vice versa because of the diagnostic difficulty with some of the equivocal images. There was a marked association between the accuracy of diagnosing AK and the duration of disease. Previous case reports have mainly described the morphological features of cysts and trophozoites in the epithelium and stroma during active infection,[4,5] but have not related the number of cysts seen and the way they distribute with the different stages of the disease process. In early disease, where the organism is mainly confined to the epithelium, the presence of large numbers of inflammatory cells made diagnosing AK more difficult because of the difficulty in distinguishing AK cysts and particularly
trophozoites from inflammatory cells.[5] Late presentation was associated with either a greater number of *Acanthamoeba* cysts seen in the images or the fact that they were easier to identify because of a reduction in the type and number of host cells seen. Our experience, therefore, suggests that AK is easier to identify with confocal microscopy in the later stages of infection.

The use of confocal microscopy in diagnosing FK has been widely reported in the literature.[6-8] Filamentous fungal hyphae have characteristic linear hyper-reflective lesions branching at 45 or 90 degrees angle,[7] whereas candida infection produces pseudofilaments.[7] Despite these well described confocal appearances of FK in the literature, the percentage of correct diagnosis in our series was low possibly due to difficulties in differentiating other linear images from fungal hyphae.[23]

*Nocardia* and *Microsporidia* species are rare causes of microbial keratitis.[24] Clinically, *Nocardia* may be misdiagnosed as mycotic or mycobacterial keratitis,[9,25] whilst *Microsporidia* can be misdiagnosed as AK or herpes simplex keratitis. Despite the rarity of these organisms, because of the unique appearance on confocal microscopy with *Microsporidia*,[10] two observers managed to identify this organism correctly in both of their assessments. Only observer B managed to obtain the correct diagnosis in both assessments for diagnosing one case of Nocardia keratitis; the unfamiliarity in interpreting confocal images of *Nocardia*, the similarity in appearance of fungal hyphae and *Nocardia* filaments, and the small number of cases in our study made diagnosing this organism difficult. The inclusion of both *Nocardia* and *Microsporidia* cases might have reduced the overall sensitivity and specificity values
but as confocal findings of both organisms have been reported, we believe it was appropriate to include them in the study.

In summary, to the best of our knowledge, this is the first study evaluating the diagnostic accuracy of microbial keratitis using a single reference standard for different masked observers with the HRT II / RCM confocal microscope. Although confocal microscopy is non-invasive and can provide a rapid diagnosis for microbial keratitis, (i) similarities between inflammatory and pathogenic cells, and (ii) difficulty in interpreting equivocal images, limits its usefulness as a stand-alone tool in diagnosing keratitis. Confocal microscopy is a useful adjunct in managing refractory cases and we have shown that the diagnostic accuracy improves with clinician experience. However, the diagnostic accuracy of confocal microscopy used in isolation from the clinical assessment is still too low to be a substitute for tissue diagnosis, particularly in patients with progressive disease. Improvement in clinician training and experience, greater standardization of image interpretation, and the development of new software in tandem with higher resolution imaging is likely to improve the diagnostic accuracy of this technology in diagnosing microbial keratitis in the future.

Acknowledgements
The authors would like to thank Dr Catey Bunce for her statistical advice and support.

Competing Interests
None declared
REFERENCES


FIGURE LEGENDS

Figure 1

Confocal scans of correctly diagnosed versus incorrectly diagnosed cases.

Figures A to D demonstrate the characteristic features of inflammatory cells and pathogenic organisms on confocal microscopy in which all the observers had made the correct diagnoses. A, *Acanthamoeba* cysts (white arrow), some with double-walled appearance (dotted white arrow). B, Inflammatory cells (black arrows). C, Fungal hyphae (black arrows). D, *Microsporidia* organisms (white arrows).

Figures E to J show a series of images of incorrectly diagnosed cases demonstrating the difficulty in distinguishing host cells from pathogenic organisms and *Nocardia* from fungal keratitis. E, Shows *Nocardia* filaments (white arrows) - only observer B identified this correctly with all the other observers graded it as fungal hyphae. F, Cultured *Acanthamoeba* - misdiagnosed as bacterial keratitis by observers B, C and D; possible *Acanthamoeba* cysts (white arrows), and possible inflammatory cells (black arrows). G, Cultured bacteria (*Staphylococcus aureus*) but was misdiagnosed as *Acanthamoeba* by all the observers; multiple round lesions that could be identified as inflammatory or *Acanthamoeba* cysts (white arrows). H, Cultured bacteria - diagnosed as fungal keratitis by observers C & D; linear hyphae-like opacities that were confused with fungal hyphae (white arrows). I, Cultured bacteria - diagnosed as *Microsporidia* by all the observers; small hyper-reflective granules that appear similar to *Microsporidia* organisms (white arrows). J, Cultured positive for *Alternaria* and
Staphylococcus aureus but diagnosed as Nocardia by observers A, B and D; hyphae type lesions that appear similar to Nocardia filaments (black arrows).

Figure 2

Scattered plot showing the relationship between number of correctly diagnosed cases and duration of disease (days) for Acanthamoeba, bacteria and fungal keratitis.