Online Resource 1

Non-ophthalmologist agreement with consultant AS-OCT anterior chamber cell counts in paediatric uveitis: an exploratory masked study

International Ophthalmology. Ali S. Ibrahim, Ekin Kahraman, Sean Zhou, Narman Puvanachandra, Saad Khan, Anas Injarie. Corresponding author: Ali S. Ibrahim, Norwich Medical School, University of East Anglia (Ali.ibrahim8@nhs.net)

Table S1 Batch-level agreement with the reference consultant (descriptive)

Batch (grading order)	n	SUN mix[a]	Grader	ICC(A,1) (F-based 95% CI)	MAD, cells	Bias, cells
1 (1–19)	19	6/5/3/4/1	Optometrist	0.899 (0.761–0.960)	1.26	+0.61
			Medical student	0.948 (0.871–0.979)	0.68	−0.29
			Nurse	0.947 (0.870–0.979)	0.98	+0.21
2 (20–38)	19	12/0/2/4/1	Optometrist	0.959 (0.859–0.986)	0.71	+0.71
			Medical student	0.904 (0.771–0.961)	0.79	+0.65
			Nurse	0.912 (0.790–0.965)	0.72	+0.41
3 (39–57)	19	12/4/0/0/3	Optometrist	0.930 (0.829–0.972)	0.62	+0.50
			Medical student	0.955 (0.889–0.982)	0.46	+0.22
			Nurse	0.882 (0.723–0.953)	0.69	+0.39
4 (58–76)	19	6/7/3/2/1	Optometrist	0.928 (0.824–0.972)	0.35	+0.24
			Medical student	0.934 (0.833–0.974)	0.43	+0.25
			Nurse	0.882 (0.585–0.960)	0.62	+0.47
5 (77–95)	19	7/4/4/2/2	Optometrist	0.916 (0.797–0.966)	0.89	+0.47
			Medical student	0.967 (0.915–0.987)	0.35	−0.27
			Nurse	0.956 (0.861–0.984)	0.70	+0.49
6 (96–113)	18	6/2/5/3/2	Optometrist	0.917 (0.793–0.968)	0.68	−0.05
			Medical student	0.880 (0.711–0.953)	0.77	+0.36
			Nurse	0.943 (0.789–0.981)	0.83	+0.61

[a] Number of episodes at SUN grade 0/0.5+/1+/2+/3+. Batches followed the randomised grading order, so case mix varied between batches. ICC, intraclass correlation coefficient; MAD, mean absolute difference; SUN, Standardization of Uveitis Nomenclature

Material S2 AS-OCT cell counting framework

Training protocol. Duration approximately 10 minutes, delivered once before grading commenced. Format: verbal explanation of AS-OCT image features accompanied by the written rule-based framework below. No practice cases were provided.

Definition of a countable hyperreflective focus. For counting purposes, a countable focus (referred to as a 'cell') is a hyperreflective focus that meets all of the following criteria. This is an operational image definition; the biological nature of the foci was not confirmed. An annotated example is shown in Fig. S1.

- Small and discrete: a distinct dot, not a diffuse signal
- Well-circumscribed: clear boundaries distinguishable from background
- Located within the anterior chamber: between the corneal endothelium and the anterior iris surface
- Not contiguous with a tissue surface: not touching or extending from the corneal endothelium, iris surface or lens capsule

Exclusion criteria. The following should not be counted:

- Edge reflections: bright signals at the boundary of the scan field
- Linear signals: elongated reflections inconsistent with a discrete cell
- Background speckle: diffuse low-level noise throughout the image
- Surface-contiguous reflections: bright signals arising from or continuous with the corneal endothelium, iris pigment epithelium or lens surface
- Poorly defined faint signals: indistinct reflections that do not meet the well-circumscribed criterion, likely representing noise or artefact

Grading procedure.

- Open the seven central B-scans for the episode
- For each B-scan, count the number of hyperreflective foci meeting the above criteria
- Record the count for each of the seven B-scans
- The episode-level cell count is the mean across the seven B-scans

Guidance notes.

- When uncertain whether a signal is a countable focus or artefact, do not count it
- Count each discrete dot once, even if it appears bright
- Do not attempt to judge clinical significance; count all qualifying dots regardless of number
- If no qualifying dots are visible in a B-scan, record zero

Table S3 Inter-consultant agreement on the second-consultant subsample

A second consultant paediatric ophthalmologist regraded 75 B-scans (50 episodes, 20 patients) drawn from the 791 study B-scans by random sampling stratified by SUN grade, using the identical written counting framework and masked to the reference consultant's counts and the clinical SUN grade. Because of the stratification, SUN grade 2+ is over-represented relative to the cohort.

Metric (n = 75 B-scans)	Value
ICC(A,1) (patient-level bootstrap 95% CI)	0.901 (0.794–0.950)
Pearson r	0.902
Quadratic weighted kappa	0.899
Mean signed difference (second − reference consultant), cells	−0.31
Mean absolute difference, cells	0.97
SD of differences, cells	1.76
95% limits of agreement, cells	−3.75 to +3.13

SUN grade	n	Reference consultant mean count	Second consultant mean count	Mean difference	MAD
0	33	0.82	0.76	−0.06	0.30
0.5+	11	1.64	1.73	+0.09	0.64
1+	8	2.12	2.12	0.00	1.25
2+	17	4.65	4.12	−0.53	1.71
3+	6	8.67	6.50	−2.17	2.83

Counts are per B-scan. The mean absolute difference rose with inflammation grade, reflecting greater counting difficulty at higher cell densities. B-scan-level agreement would be expected to be lower than episode-level agreement, because averaging across seven B-scans reduces measurement error, so these values are not directly comparable with the episode-

level ICCs in the main analysis. Agreement of each non-specialist with each consultant on these B-scans is reported in Table 3 of the main text. MAD, mean absolute difference; SUN, Standardization of Uveitis Nomenclature

Table S4 Counting error by clinical SUN grade (episode level)

SUN grade	Episodes	Optometrist MAD / bias	Medical student MAD / bias	Nurse MAD / bias
0	49	0.12 / +0.09	0.06 / +0.01	0.14 / +0.14
0.5+	22	0.29 / +0.18	0.25 / +0.03	0.85 / +0.77
1+	17	0.77 / +0.66	0.52 / +0.24	0.89 / +0.59
2+	15	1.74 / +1.67	0.47 / +0.07	1.11 / +0.28
3+	10	3.34 / +0.26	4.10 / +1.13	2.80 / +1.06

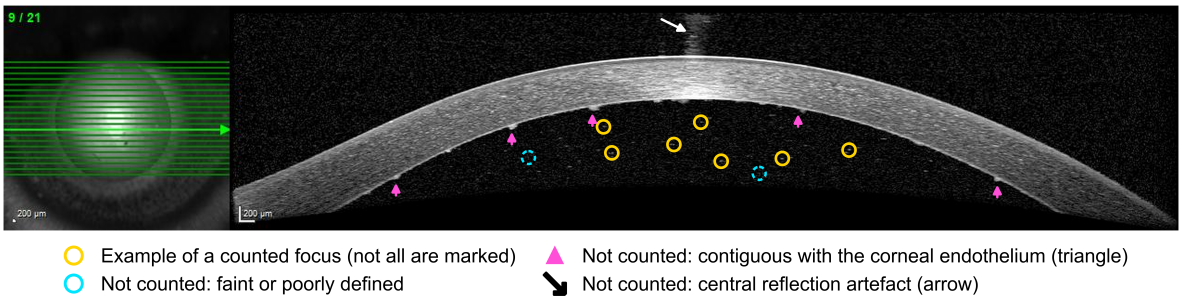
Values are in cells (episode-level mean of seven B-scans), grader minus consultant. MAD, mean absolute difference; SUN, Standardization of Uveitis Nomenclature

Table S5 Secondary and sensitivity analyses

Metric	Optometrist	Medical student	Nurse
Agreement with the reference consultant: additional summaries			
ICC(A,1), F-based 95% CI (pre-specified; assumes independent episodes)	0.884–0.947	0.890–0.946	0.878–0.944
Pearson r	0.931	0.932	0.928
Repeated-measures 95% limits of agreement, cells[a]	−2.47 to +3.29	−2.88 to +3.18	−2.57 to +3.41
Counting error across the grading sequence			
Batch ICC range (six batches; Table S1, Fig. S2)	0.899–0.959	0.880–0.967	0.882–0.956
Change in absolute error per 10 episodes, adjusted[b] (95% CI)	−0.049 (−0.101 to +0.002)	−0.015 (−0.062 to +0.033)	−0.008 (−0.070 to +0.054)
Grading time (30 episodes, 13 patients)[c]			
Median (IQR), s	90.5 (82.8–95.8)	100.0 (89.0–110.0)	90.0 (79.8–98.0)
Change per batch, adjusted[d], s (95% CI)	−4.1 (−5.8 to −2.5)	−4.2 (−6.6 to −1.7)	−4.1 (−5.9 to −2.4)
Within-patient clustering			
Design effect[a]	1.07	1.00	1.17
One episode per patient (27 patients): median ICC (2.5th–97.5th percentile)[e]	0.905 (0.840–0.958)	0.942 (0.884–0.970)	0.933 (0.847–0.974)
Consultant-positive episodes, one per patient (18 patients)[e]	0.882 (0.805–0.946)	0.930 (0.868–0.962)	0.907 (0.830–0.957)

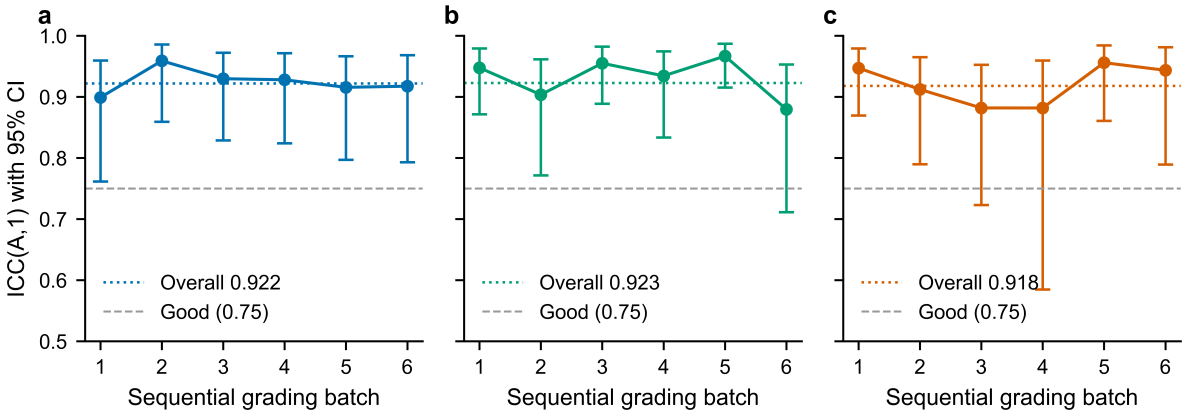
Agreement across all four graders: ICC(A,1) 0.905 (F-based 95% CI 0.875–0.929). [a] Linear mixed model of grader − consultant differences with patient as a random intercept. [b] Linear mixed model of episode-level absolute difference on grading order, adjusted for consultant count, patient random intercept. [c] Consultant: median 77.5 s (IQR 67.5–81.0); adjusted change per batch −4.1 s (95% CI −6.4 to −1.8). [d] Regression on batch number adjusted for consultant count, patient-clustered standard errors. [e] Variation from episode selection, not confidence intervals; 1000 iterations (seeds 0–999). The pre-specified threshold for a learning effect was a difference in batch ICC > 0.10; no grader exceeded it. ICC, intraclass correlation coefficient; IQR, interquartile range

Fig. S1 Annotated example of the counting framework



De-identified central B-scan (frame 9 of 21) from the study cohort, annotated to illustrate the counting rules in Material S2. Yellow circles, examples of foci meeting the counting definition (not every counted focus is marked); dashed cyan circles, faint or poorly defined signals that were not counted; magenta triangles, bright foci contiguous with the corneal endothelium that were not counted; arrow, central reflection artefact. Background speckle in the anterior chamber was not counted. The near-infrared image (left) shows the scan position

Fig. S2 Batch-level agreement with the reference consultant



ICC(A,1) with F-based 95% confidence intervals across six sequential grading batches for (a) the optometrist, (b) the medical student and (c) the nurse versus the reference consultant. Dotted line, overall ICC; dashed line, threshold for good agreement (0.75). n = 18–19 episodes per batch; batch case mix varied (Table S1)