Published in final edited form as:

Can J Ophthalmol. 2013 August ; 48(4): 227–234. doi:10.1016/j.jcjo.2013.02.010.

Fundus Image Diagnostic Agreement Among Multiple Raters in Uveitis Utilizing Free and Open Source Software

Jane Hsieh, BS*, Andrea F. Honda, MD*, Mayte Suárez-Fariñas, PhD⁺, C. Michael Samson, MD, Sanjay Kedhar, MD*, John Mauro, MD, Jasmine Francis, MD*, Jason Badamo, MBA*, Vicente A. Diaz, MD*, John H. Kempen, MD, PhD[†], and Paul A. Lattkany, MD*

*New York Eye and Ear Infirmary, New York, New York USA

⁺Rockefeller University, New York, New York USA

[†]University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania USA

INTRODUCTION

Uveitis is a significant public health problem accounting for approximately 10% of blindness in Western countries with a disproportionately high socioeconomic impact given its earlier onset than most other major blinding conditions.^{1–4} The United States of America has an estimated prevalence of uveitis of 69–115 cases per 100,000,⁵ with an estimated annual incidence of 17–52 cases per 100,000 person-years.⁶ Many patients in rural areas do not access ophthalmic care because of difficulties related to travel distance, time and expense.^{7,8} In the 2007 National Physician Survey in Canada, only 4.0% of ophthalmologists reported that their practice primarily served the needs of residents from rural and remote areas.^{9,10} These results were mirrored in the 2010 Ontario Physicians Human Resources Data Centre (OPHRDC) Report, which indicated that there was 1 ophthalmologist to 11,808 residents in the Toronto Central Local Health Integrated Network (LHIN), compared to 1 ophthalmologist to a range of 22,226 – 68,641 residents in the relatively rural North Simcoe-Muskoka LHIN.^{11,12} In Canadian cities with populations of 10,000–25,000, people have to travel on average 91 km to reach the nearest ophthalmologist.^{13,14} In developing countries, ophthalmologist: resident ratios are less favorable, and specialist care often is unavailable. Therefore reaching underserved areas through alternate modalities, such as telemedicine, should be considered.

We sought to explore whether uveitis also could be diagnosed using fundus photography and fluorescein angiograms alone (helping define images role in a future telemedicine build). We also used free and open source software applications to demonstrate that future image agreement research in ophthalmology could be done with relatively modest resources. In this study, we analyzed the agreement between uveitis specialists who arrived at a diagnosis following a full history and clinical examination supplemented by digital fundus and fluorescein photos compared to ophthalmologists who arrived at a uveitis diagnosis using

Corresponding Author: Paul A. Lattkany MD 225 East 38th Street, New York, NY 10016 p: (212)687-0265 f: (212)687-3463 dvd27email@yahoo.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

FINANCIAL DISCLOSURES: The ancillary study reported herein was unfunded. Dr. Kempen serves or has served as a consultant for Allergan, Alcon, Harbor Biosciences, Lux Biosciences, and Sanofi Pasteur. **OHTER ACKNOWLEDGEMENTS:** None.

digital fundus and fluorescein photos alone. Since all images used in the study were from the MUST trial, all infectious causes of uveitis were excluded from this analysis.

METHODS

The Multi-center Uveitis Steroid Treatment (MUST) Trial ([clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT00132691) identifier NCT00132691) is a 23-site multi-center randomized clinical trial that enrolled 255 patients with active, noninfectious, intermediate, posterior or panuveitis randomized to one of two treatment groups: treatment with a fluocinolone acetonide implant (Retisert™, Bausch & Lomb, Rochester, New York, USA), or with systemic corticosteroids supplemented by immunosuppressive agents when indicated.^{14–16} Photographers trained and certified by the University of Wisconsin Fundus Photograph Reading Center obtained baseline photographs of 247 of the MUST Trial patients (inability to image the fundus was not an exclusion criterion). Certified trial ophthalmologists provided patients' diagnoses and graded clinical status based on ophthalmologic history and examination. At the baseline visit, which is the subject of this analysis, color and red-free fundus photographs, fluorescein angiography images, optical coherence tomography images, lens photos, and Humphrey visual fields were obtained, which were available to the trial ophthalmologists for review as needed. The 30° Zeiss FF4 (or similar models) and FF450-plus fundus cameras as well as the Topcon TRC-50 series (50VT, 50X, 50EX, 50IA, and 50IX or similar models) used at the 35° setting were suitable cameras. Additionally, the Canon UVi (or similar models) used at the 40° setting, and the Kowa, Nikon and Olympus camera models used at the 30° or 35° settings were suitable cameras for this study. There were six color fundus images per eye and 36 angiographic images per eye obtained per the Fundus Photograph Reading Center's (FPRC) Modified 3-Standard Field Color Fundus Photography and Digital Sweep Angiography protocol.

After obtaining approval from the Institutional Review Board at the New York Eye and Ear Infirmary (NYEEI) and the Executive Committee for the MUST Trial, three fellowship-trained specialists in uveitis who did not serve as clinicians in the MUST Trial independently analyzed approximately 12,000 de-identified digital images from the 247 patients, completing a short questionnaire after viewing each set of fundus images for each patient. The remaining 8 patients (3.1%) were excluded from our study due to the absence of digital fundus photos. For the purposes of our study, all photographs were de-identified and not associated with any patient history or data. Each patient set consisted solely of images, which included color fundus, fluorescein, and red free images. Each uveitis specialist was provided a Windows 7™ laptop with identical settings appropriate for diagnostic purposes, an Acer Aspire AS5336-2524 15.6" Notebook Computer, to ensure consistent resolution and screen quality across the three reviewers.

We utilized free and open source applications to investigate the feasibility of implementing an inter-observer agreement study with limited resources—our ancillary study assessing the images the MUST Trial provided was unfunded. Epi Info™ is a public domain software program developed by the Centers for Disease Control (CDC; Atlanta, Georgia) used for electronic survey creation and data collection, which is freely available for use, copying, translation and distribution. It is designed for implementation with minimal hardware and software requirements. In addition, we chose to utilize a freely available DICOM viewer, ClearCanvas Workstation 2.0™ (Toronto, Canada) which has available source code under the ClearCanvas License, modeled after the New BSD License.

All photos were accessed on the reviewer's study laptop using ClearCanvas Workstation 2.0™. Reviewer responses for each of the 247 patient-related questionnaires were directly entered into Epi Info™ version 3.5.3. For each patient set, the respondent was asked to

review a series of fundus images and then assess 1) image clarity, 2) presence or absence of retinal vasculitis, 3) ophthalmologic diagnosis, 4) their level of confidence in their diagnosis, 5) whether or not additional information would be needed to determine the diagnosis with greater confidence and/or accuracy, and 6) the presence or absence of 1 chorioretinal scar (full questionnaire shown in Figure 1). To mimic a likely clinical scenario, the questions in the survey were based on the diagnosis of both eyes together.

The responses from each of the three reviewers were compared with each managing clinician's diagnoses as a gold standard, which had been determined previously for the MUST Trial by the patient's ophthalmologist. The two questions in our study that asked for the patients' diagnoses utilized the identical question format used in the baseline MUST data entry forms in order to ensure comparability.

Data Analysis

The focus of our statistical analysis was on the diagnoses made by the reviewers. Our main objective was to measure the level of agreement between the reviewers' diagnoses and the "gold standard" (validity). Our secondary objective was to estimate the measurement of agreement among reviewers (reliability).

Since both the original examiner and our respondents' diagnosis options were not exclusive, such that the number of diagnoses given could differ amongst them, we utilized a rank transformation to adjust for the possible variation in the number of responses per patient.¹⁷ For example, if a rater chose category diagnosis A from the total of 10 categories ($K=10$) as the only diagnosis, then that category received a rank of 1 while each of the remaining 9 categories received the average rank that represented an average of the remaining 9 categories ($(K+2)/2$, in our example $(10+2)/2=6$). If a rater chose diagnosis A and diagnosis B, then both diagnosis A and diagnosis B received a rank of 1.5 while each of the remaining categories received the average rank of the 8 remaining categories, etc. This rank transformation was applied to the gold standard as well as to the responses of the 3 reviewers, resulting in a multivariate composite of 10 ratings per reviewer. This transformation inherently accounts for partial agreements and the top ranks are distributed amongst more categories when many diagnoses are chosen.¹⁸

We measured the level of agreement between the reviewers' diagnoses using an iota statistic.¹⁹ The iota statistic was calculated on the rank-transformed diagnoses. We used this statistic to measure agreement amongst the whole group and between the three pairs of reviewers and between each reviewer and the gold standard. The iota coefficient is an estimator of chance-corrected inter-observer agreement among the multivariate ratings of several raters; it corresponds to the extensions of the kappa coefficient to several raters in the univariate case.²⁰ Because asymptotic distributions are not available for this statistic, we used bootstrapping techniques to estimate the confidence intervals. As with the kappa statistic, iota values of 0.80 and above represent excellent agreement, values between 0.61 and 0.80 represent substantial agreement, 0.41 to 0.61 represent moderate agreements, and values of 0.40 or less suggest fair to poor agreement.²¹ The iota coefficient measures reliability among raters rather than validity against a gold standard. However, the validity of a measure is always less than the reliability.²² For simple two-category agreement evaluations, a standard kappa statistic was used.

All statistical analysis were carried out in R, a free software package.²³ We used the R's package *irr* to calculate the iota coefficient and the package *boot* for bootstrapping simulations. A nonparametric stratified resampling ($N=5000$) strategy was implemented to estimate the iota bias and standard error. The 95% confidence interval was calculated using

bias-corrected and accelerated BCa instead of the percentiles. The (BCa) bootstrap adjusts for both bias and skewness in the bootstrap distribution.

For the questions posed to the reviewer regarding the image quality of patient images and the presence of chorioretinal scar on fundus images, we calculated—as a percentage of the total number of patients—the number of times that each reviewer reported each of the possible answer choices.

RESULTS

Each reviewer completed the study questionnaire for each of the 247 distinct patient image sets. Table 1 displays the distribution of “gold standard” diagnoses as given by the MUST ophthalmologists who examined the patient as well as reviewing all pertinent tests and histories. The most common gold standard diagnoses were intermediate uveitis (38.1%) and multifocal choroiditis (12.6%). Figure 2 demonstrates the proportion of MUST patients per diagnosis for each reviewer and the gold standard. Each of the 3 reviewers and the baseline gold standard had diagnosed very different proportions of each of the diagnostic categories. Per the baseline clinician, 22.0% of the cases did not fit into one of the specific diagnostic categories studied, vs. 90.0%, 76.1% and 21.1% for the three reviewers.

We also analyzed the data reductively by transforming the initial 10-category diagnosis to a dichotomous variable “any of the diagnoses A – I” or “J” (none of the above), and then by measuring agreement between each reviewer and the baseline “gold standard” which is shown in Figure 3. The reliability among the 3 reviewers is $\kappa=0.11$, with 95%-CI $[-0.034, 0.261]$, $p < 0.13$. The reliability between the 3 reviewers and the gold standard, is $\kappa=0.08$ with 95% CI $(-0.24, 0.39)$ and $p=0.63$.

The quality of retinal images as reported by the reviewers is summarized in Table 2. For image clarity per patient image set reviewer 1, 2 and 3 reported 17.0%, 17.8% and 11.3% as inadequate and the remainder as adequate.

In addition to the primary outcome of diagnosis, reviewers were asked to identify the presence of retinal vasculitis and chorioretinal scars. Reviewers 1, 2 and 3 diagnosed 9.3%, 5.3% and 46.2% of MUST patients, respectively, with retinal vasculitis. Table 3 reports reviewers' results regarding the presence or absence of chorioretinal scars. Reviewers 1, 2, and 3 had difficulty being sure whether chorioretinal scars were present (borderline or cannot assess) in 17.4, 19.4, and 34.8% respectively. Among those each reviewer was able to grade, 82.6, 80.6, and 65.2% respectively were graded as having chorioretinal scars, respectively.

Table 4 summarizes comparisons of the three reviewers with each other, overall against the gold standard, and each reviewer's individual agreement to the gold standard. There were substantial differences found between the gold standard diagnoses and the diagnoses made by each reviewer as well as among the reviewers themselves. There was uniformly poor agreement between the respondents and the gold standard ophthalmologist (iota=0.04, 95% CI, 0.02, 0.06; =0.09 (95% CI, 0.06, 0.13); and =0.10 (95% CI, 0.06, 0.14) respectively). Figure 4.

The reviewers' self-reported overall confidence in their diagnoses was extremely poor with a median of 2, 6 and 1 out of 10, respectively for reviewers 1, 2, and 3. In addition, for every patient throughout the study all reviewers reported a need for more historical patient information to better arrive at a diagnosis.

DISCUSSION

Our well-powered study regarding diagnosis of uveitis based solely on fundus photo and fluorescein angiography images taken under optimal conditions in a clinical trial done under aggressive quality control enforcement by a fundus image reading center demonstrated poor agreement across most diagnostic categories. Agreement was poor between each reviewer and the gold standard diagnosis as well as amongst the three reviewers, suggesting that the validity and reliability of using on image assessment is inadequate. A major reason for disagreement likely is the lack of historical patient information, based on the complaints of the reviewers participating in the study--more complete information including history and radiology and laboratory studies might lead to a more favorable level of agreement. An additional reason for disagreement might reflect discrepancies between the gold standard clinicians and reviewers as to what the precise basis for diagnosis of each of the specific entities studied, as there are no formal diagnostic criteria for most of them. Establishment of precise diagnostic criteria might benefit a telemedicine approach to diagnosis. The duration of disease may have resulted in a common fundus appearance for different diagnoses because the median time from initial diagnosis was 3.8 years (25–75%-tile, 1.2 to 8.0).²⁴

In addition, there was a lack of agreement even regarding the presence of the objective finding of chorioretinal scarring. The observed lack of agreement might relate to difficulties in imaging peripheral scarring using the photographic techniques of the MUST Trial and/or might reflect uncertainty regarding whether abnormalities in the fundus appearance represent an inflammatory scar or not (in the absence of history). Improved photographic technology that better images the peripheral fundus might aid future telemedicine efforts that would require identification of peripheral fundus findings, as is pertinent to the diagnosis of posterior segment uveitis. The statistical analyses are not straightforward due to the characteristics of the data at hand. The variables representing the diagnoses have three complexities: 1) 10 categories, 2) these categories are not exclusive, since several diagnoses can be selected, and 3) there are multiple reviewers. Although a generalization of the Kappa Statistic exists to calculate reliability in the multi-rater/multi-category case, to calculate validity and reliability in the complexities noted above represented a methodological challenge. The estimator of reliability in the multi-rater, multi-category Kappa was an upper boundary for the validity of the construct, which in our case is very low.

A weakness of our study design included not controlling for the exact environment in which the reviewer answered questions. However, since the reviewers were allowed to complete their analysis in what they considered an appropriate setting, the range of circumstances, is a reasonable approximation of what would be encountered in the image analysis segment of a telemedicine program. The agreement observed was sufficiently poor that it is unlikely that the reviewer's equipment was the primary constraint, especially given that the reviewers considered the conditions adequate. It should be noted that some infectious causes of uveitis such as CMV retinitis would lend itself well to an image agreement study but infectious etiology was excluded from MUST.²⁵ Separately, MUST OCT data was analyzed by Domalpally et al who reported excellent intergrader agreement of OCT macular edema.²⁶ Also, the pars plana was not imaged as part of the MUST protocol so its diagnosis would have to be inferred after excluding other characteristic diagnoses such as multifocal choroiditis. The patients were not excluded based on poor image quality to broaden inclusion criteria and help the study results be more generalizable. In addition, part of the interpretation of our results should be based on our utilized photographic equipment, and future work should explore if newer imaging modalities give improved reviewer agreement outcomes. There were no reported limitations by the respondents of utilizing free and open source applications for an inter-observer agreement study, suggesting that studies of this

kind can be done cost-effectively, and further investigation beyond the scope of our study should be analyzed.

A literature review on telemedicine in vitreoretinal diseases has shown an increase in cost-effective means in screening for diabetic retinopathy, retinopathy of prematurity, and age-related macular degeneration.^{27,28} However, given that diagnosis based solely on fundus photographs and fluorescein angiogram images performed poorly, our findings support a recommendation against an approach that utilizes fundus images alone to determine diagnosis in uveitis, and support the need for exploration of traditional telemedicine constructs utilizing history and clinician feedback. Our observations are intuitively plausible, because diabetic retinopathy, retinopathy of prematurity, and age-related macular degeneration represent single diagnostic entities in which a limited number of well-defined findings need to be detected, whereas uveitis reflects an array of different conditions mostly without diagnostic criteria and presenting with discrepant findings. Furthermore, prior diabetic image agreement focus on identification of cases that might be abnormal rather than providing support in sorting out a complicated differential diagnosis, as attempted here. Future telemedicine programs aiming to support diagnosis of uveitis involving the posterior segment must develop approaches that recognize the need to establish linkage between additional clinical data and the use of fundus photos in determining a diagnosis, in a manner that nevertheless still would be cost-effective and feasible to scale up if successful.

Acknowledgments

FUNDING/SUPPORT: The contribution of the Multi-center Uveitis Steroid Treatment (MUST) Trial Research Group to the manuscript was supported by cooperative agreements from the National Eye Institute to Mount Sinai School of Medicine (U10 EY 014655), The Johns Hopkins University Bloomberg School of Public Health (U10 EY 014660), and the University of Wisconsin, Madison, School of Medicine (U10 EY 014656). Bausch & Lomb also provided a limited allotment of fluocinolone acetonide 0.59-mg implants for patients randomized to implant treatment in the study who could not afford that treatment. The clinicaltrials.gov identifier number for the MUST Trial is NCT00132691. The MUST Trial and this study were conducted in accordance with the Declaration of Helsinki, with approval by the governing institutional review boards of the participating institutions.

CONTRIBUTION OF AUTHORS: Dr. Kempen's input was partially supported by Research to Prevent Blindness and the Mackall Foundation. Statement about conformity with Author Information: The New York Eye and Infirmary (NYEE) internal review board reviewed and approved the study.

References

1. Nussenblatt RB. The natural history of uveitis. *Int Ophthalmol*. 1990; 14:303–8. [PubMed: 2249907]
2. Suttorp-Schulten MS, Rothova A. The possible impact of uveitis in blindness: a literature survey. *Br J Ophthalmol*. 1996; 80:844–8. [PubMed: 8962842]
3. Smith JA, Mackensen F, Sen HN, et al. Epidemiology and course of disease in childhood uveitis. *Ophthalmology*. 2009; 116:1544–51. [PubMed: 19651312]
4. Jabs DA. Epidemiology of uveitis. *Ophthalmic Epidemiol*. 2008; 15:283–4. [PubMed: 18850463]
5. Darrell RW, Wagener HP, Kurland LR. Epidemiology of uveitis: incidence and prevalence in a small urban community. *Arch Ophthalmol*. 1962; 68:502–14. [PubMed: 13883604]
6. Gritz DC, Wong IG. Incidence and prevalence of uveitis in Northern California; the Northern California Epidemiology of Uveitis Study. *Ophthalmology*. 2004; 111:491–500. [PubMed: 15019324]
7. Suhler EB, Lloyd MJ, Choi D, et al. Incidence and prevalence of uveitis in Veterans Affairs Medical Centers of the Pacific Northwest. *Am J Ophthalmol*. 2008; 146:890–6. [PubMed: 19027424]
8. Reeves SW, Sloan FA, Lee PP, Jaffe GJ. Uveitis in the elderly: epidemiological data from the National Long-term Care Survey Medicare Cohort. *Ophthalmology*. 2006; 113:302–7.

9. Chan SM, Hudson M, Weis E. Anterior and Intermediate uveitis cases referred to a tertiary centre in Alberta. *Can J Ophthalmol*. 2007; 42:860–4. [PubMed: 18033328]
10. Campbell RJ, Hatch WV, Bell CM. Canadian health care: a question of access. *Arch Ophthalmol*. 2009; 127:1384–6. [PubMed: 19822859]
11. National Physician Survey. [Accessed November 21, 2010] Results for surgical specialists. Population primarily served in main patient care setting. 2007. Available at: http://www.nationalphysiciansurvey.ca/nps/2007_Survey/2007results-e.asp.
12. Ontario Physicians Human Resources Data Centre. [Accessed November 21, 2010] 2010 Active Physicians in Ontario by Specialty and LHIN. Available at: <https://www.ophrdc.org/Public/Report.aspx?owner=pio>.
13. Rosenthal MB, Zaslavsky A, Newhouse JP. The geographic distribution of physicians revisited. *Health Serv Res*. 2005; 40:1931–52. [PubMed: 16336557]
14. Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group. Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Louis TA, Sugar EA, Thorne JE. Randomized comparison of systemic anti-inflammatory therapy versus fluocinolone acetonide implant for intermediate, posterior, and panuveitis: the multicenter uveitis steroid treatment trial. *Ophthalmology*. Oct; 2011 118(10): 1916–26. [PubMed: 21840602]
15. Multicenter Uveitis Steroid Treatment Trial Research Group. Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. *American Journal of Ophthalmology*. Apr; 2010 149(4):550–561. e10. [PubMed: 20097325]
16. Holbrook JT, Kempen JH, Prusakowski NA, Altaweel MM, Jabs DA, Multicenter Uveitis Steroid Treatment (MUST) Trial Research Group. Challenges in the design and implementation of the Multicenter Uveitis Steroid Treatment (MUST) Trial--lessons for comparative effectiveness trials. *Clin Trials*. Dec; 2011 8(6):736–43. [PubMed: 21994128]
17. Janson H, Olsson U. A measure of agreement for interval or nominal multivariate observations. *Educ Psychol Meas*. 2001; 61:277–89.
18. Gamer M, Lemon J, Fellow I, et al. Various Coefficient of Inter-rater Reliability and Agreement. 2010
19. Kraemer HC. Extension of the kappa coefficient. *Biometrics*. 1980; 36:207–16. [PubMed: 7190852]
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
21. Gwet, KL. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters. Advanced Analytics; Gaithersburg, MD: 2010. p. 11-181.
22. Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond Kappa: a review of inter-rater agreement measures. *Can J Stat*. 1999; 27:3–23.
23. R-project.org
24. Multicenter Uveitis Steroid Treatment Trial Research Group. Kempen JH, Altaweel MM, Holbrook JT, Jabs DA, Sugar EA. The multicenter uveitis steroid treatment trial: rationale, design, and baseline characteristics. *American Journal of Ophthalmology*. Apr; 2010 149(4):550–561. e10. [PubMed: 20097325]
25. Hubbard LD, Ricks MO, Martin BK, Bressler NM, Kempen JH, Dunn JP, Jabs DA, Cytomegalovirus Retinitis and Viral Resistance Study Group. Comparability of two fundus photograph reading centers in grading cytomegalovirus retinitis progression. *Am J Ophthalmol*. Mar; 2004 137(3):426–34. [PubMed: 15013864]
26. Domalpally A, Altaweel MM, Kempen JH, Myers D, Davis JL, Foster CS, Latkany P, Srivastava SK, Stawell RJ, Holbrook JT, MUST Trial Research Group. Optical coherence tomography evaluation in the Multicenter Uveitis Steroid Treatment (MUST) trial. *Ocul Immunol Inflamm*. Dec; 2012 20(6):443–7. [PubMed: 23163490]
27. Ciemins E, Coon P, Peck R, et al. Using telehealth to provide diabetes care to patients in rural Montana: findings from the promoting realistic individual self-management program. *Telemed J E Health*. 2011; 8:596–602. [PubMed: 21859347]

28. Polisena J, Tran K, Cimon K, et al. Home telehealth for diabetes management: a systemic review and meta-analysis. *Diabetes Obes Metab*. 2009; 11:913–30. [PubMed: 19531058]

1. Patient Number (1-247)
2. Is the image quality high enough to make a diagnostic evaluation?
 - a. Clarity is high and there are no significant problems caused by image quality
 - b. Clarity is high, but still unable to make diagnosis
 - c. Clarity is adequate but suboptimal image quality noticeably interferes with diagnosis evaluation
 - d. Clarity is inadequate to determine major disease parameters
3. Would you diagnose the patient with retinal vasculitis?
4. Would you diagnose this patient with any of the following? Select all that apply?
 - a. Serpiginous choroiditis
 - b. Multifocal choroiditis with panuveitis
 - c. Punctate inner choroidopathy
 - d. Birdshot retinochoroidopathy
 - e. Pars Planitis
 - f. Intermediate uveitis other than pars planitis
 - g. Intermediate uveitis with anterior uveitis
 - h. Vogt-Koyanagi Harada disease
 - i. Sympathetic Ophthalmia
 - j. None of the Above
5. On a scale of 1-10 where 1 is very unsure of the diagnosis and 10 is as confident as you can be in making the diagnosis, what number would you give to the most confident diagnosis you selected in question 4
6. As compared to interpreting images for presence of diabetic retinopathy, would having more patient historical information be helpful to making a diagnosis of the type of uveitis?
7. Is there a chorioretinal scar present?

Figure 1.
Image of FIDAU survey provided to the reviewers

Figure 2.
Proportion of MUST patients with each diagnosis categories for each reviewer and the gold standard.

Figure 3. Agreement between reviewers 1, 2, and 3 with the gold standard using grouping A-I vs. J as binary variables. (see text)

Figure 4.
A. Red-free fundus photograph of a patient with a baseline diagnosis of serpiginous choroidopathy and B fluorescein angiogram of a patient with VKH syndrome.

Table 1

Diagnoses of study participants (n=247), as given by MUST Tria ophthalmologists (gold standard).

Diagnosis	N	%
Intermediate Uveitis	93	38.0
Multifocal Choroiditis	30	12.2
Birdshot Retinochoroidopathy	21	8.6
Vogt-Koyangi-Harada Disease	15	6.1
Panuveitis	15	6.1
Pars Planitis	12	4.9
Punctate Inner Choroiditis	1	0.4
Sympathetic Ophthalmia	3	1.2
Serpiginous Choroiditis	1	0.4
Other	54	22.0

Table 2

Quality of retinal image reported by reviewers (n=247)

Image Quality	Reviewer 1		Reviewer 2		Reviewer 3	
	N	%	N	%	N	%
high clarity, no significant problems	5	2.0	44	17.8	16	6.5
high clarity, unable to diagnose	118	47.8	80	32.4	12	4.9
adequate clarity	82	33.2	79	32.0	191	77.3
inadequate clarity	42	17.0	44	17.8	28	11.3

Table 3

Presence of one or more chorioretinal scars on retinal images, as reported by reviewers (n=247).

Presence of 1 Chorioretinal Scar	Reviewer 1		Reviewer 2		Reviewer 3	
	N	%	N	%	N	%
Yes	58	23.5	40	16.2	69	27.9
No	146	59.1	159	64.4	92	36.2
Borderline	6	2.4	14	5.7	52	21.1
Cannot Assess	37	15.0	34	13.8	34	13.8

Table 4

Reviewer vs Gold Standard Diagnosis Agreement, Iota (95% confidence interval)

	κ	95% CI
Overall agreement among the 3 reviewers	0.11	0.08, 0.13
Agreement between the 3 reviewers in aggregate and the gold standard	0.09	0.07, 0.11
Agreement between Reviewer 1 and the gold standard	0.04	0.02, 0.06
Agreement between Reviewer 2 and the gold standard	0.09	0.06, 0.13
Agreement between Reviewer 3 and the gold standard	0.10	0.06, 0.14

An iota coefficient was used for categorical analysis between the gold standard and each of the 3 clinical investigators.