Published in final edited form as:

Retina. 2012 June ; 32(6): 1148–1155. doi:10.1097/IAE.0b013e31823ac3c3.

Plus disease in retinopathy of prematurity: diagnostic impact of field of view

Rohini Rao, BA¹, Nina J. Jonsson, MD¹, Camila Ventura, MD¹, Rony Gelman, MD¹, Martin A. Lindquist, PhD², Daniel S. Casper, MD, PhD¹, and Michael F. Chiang, MD^{3,4}

¹Department of Ophthalmology, Columbia University College of Physicians and Surgeons (New York, NY)

²Department of Statistics, Columbia University (New York, NY)

³Department of Ophthalmology, Oregon Health & Science University (Portland, OR)

⁴Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University (Portland, OR)

Abstract

Purpose—To examine the impact of retinal field of view and magnification on inter-expert reliability of plus disease diagnosis in retinopathy of prematurity (ROP).

Methods—15 wide-angle images from infants with ROP were cropped and adjusted in magnification to create two additional image categories: medium-angle (40–50°) and narrow-angle (20–30°). These 45 images were uploaded to a web-based system and interpreted independently by 13 ROP experts using a 3-level (plus, pre-plus, neither) and 2-level (plus, not plus) classification. Absolute agreement and kappa statistics were calculated to compare inter-expert reliability.

Results—In the 3-level classification, 70% experts agreed on the same diagnosis in 8/15 (53%) wide-angle images, but only in 3/15 (20%) medium-angle and 3/15 (20%) narrow-angle images. In the 2-level classification, 80% experts agreed on the same diagnosis in 11/15 (73%) wide-angle images, but only in 9/15 (60%) medium-angle and 3/15 (20%) narrow angle images. Mean kappa of each expert compared to all other experts was 0.40–0.59 in 8/13 (62%) experts using wide-angle images, was 0–0.19 in 7/13 (54%) experts using medium-angle images, and was 0.20–0.39 in 9/13 (69%) experts using narrow-angle images.

Conclusions—Inter-expert agreement in plus disease diagnosis in wide-angle images is higher than from medium-angle and narrow-angle images. Plus disease is defined using a narrow-angle standard published photograph, yet this study suggests that peripheral findings also contribute to diagnosis.

Keywords

Retinopathy of prematurity; Retina; Pediatric ophthalmology; Telemedicine; Medical informatics

Correspondence to: Michael F. Chiang, MD, Professor of Ophthalmology & Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, 3375 SW Terwilliger Boulevard, Portland, OR 97239, Tel: 503-418-3087, Fax: 503-494-5748, chiangm@ohsu.edu.

No authors have financial conflicts of interest. MFC is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA).

Presented in part at the 2011 Association of Research in Vision and Ophthalmology conference (Ft. Lauderdale, FL).

INTRODUCTION

Retinopathy of Prematurity (ROP) is a vasoproliferative disease affecting low birth-weight infants, and continues to be a leading cause of childhood blindness worldwide. According to the Cryotherapy for Retinopathy of Prematurity (CRYO-ROP) and Early Treatment for Retinopathy of Prematurity (ETROP) trials, treatment of severe disease with laser photocoagulation or cryotherapy can improve structural and visual outcomes.^{1,2}

Clinical ROP findings are described using an international classification system.^{3,4} A key component of this classification system is “plus disease,” which is defined as arterial tortuosity and venous dilation in the posterior pole that is greater than or equal to that of a standard published photograph selected by expert consensus over 20 years ago.^{1,3,4} More recently, the revised international ROP classification system defined a new “pre-plus” category as abnormal arterial tortuosity or venous dilation within the posterior pole that is less than that of the standard published photograph.⁴ Presence of plus disease has been shown by the CRYO-ROP and ETROP studies to be the most critical factor in identifying cases of severe treatment-requiring ROP.^{1,2} Therefore, it is essential for plus disease to be diagnosed accurately and consistently.

However, plus disease diagnosis has been found to be inconsistent, even among experts.^{5–8} The standard published photograph shows numerous vessels of varying degrees of dilation and tortuosity, leaving the true diagnostic cutoff unclear.¹ Furthermore, the standard photograph has a narrower field of view and larger magnification than clinical diagnostic methods such as indirect ophthalmoscopy and wide-angle retinal imaging.^{9–16} Although plus disease is defined only by appearance of the posterior vessels, it is possible that clinicians also consider peripheral or mid-peripheral features while making diagnoses. It is also possible that differences in image magnification or field of view could create biases in diagnosis. Finally, quantitative methods for plus disease detection have been shown to be feasible using computer-based image analysis,^{6,17–23} but the field of view that should be analyzed by these systems is not known. These are important gaps in knowledge with regard to plus disease diagnosis, and with development of quantitative diagnostic methods.

The goal of this study is to examine the impact of retinal field of view and magnification on the consistency of plus disease diagnosis. This is done by examining the diagnoses made by ROP experts using wide-angle images, as well as images cropped to medium-angle and narrow-angle fields of view.

DESIGN AND METHODS

This study was approved by the Institutional Review Board at Columbia University, and was conducted in accordance with Health Insurance Portability and Accountability Act (HIPAA) guidelines. Informed consent was obtained from all expert participants using a click-through web form, and waiver of consent was obtained for use of de-identified retinal images.

Expert Participants

Potential participants were identified by the authors as being ROP experts based on meeting at least one of three criteria: having published at least 5 ROP manuscripts in the peer-reviewed medical literature, having served as a Principal Investigator in the CRYO-ROP or ETROP multi-center trials, or having served as a certified investigator in either trial. Eligible experts were invited to provide informed consent to participate.

Image Selection and Processing

Fifteen wide-angle retinal images were selected for this study. Images were captured by a commercially-available camera (RetCam; Clarity Medical Systems, Pleasanton, CA). Each image was taken from a different infant during routine clinical care. Images that showed a change in vasculature relative to baseline, as perceived by the authors, were used. All wide-angle images were centered on the posterior pole, with no visible peripheral ROP and with no descriptive clinical or demographic information.

Using photo-editing software (Photoshop Elements 8; Adobe Systems, San Jose, CA), each wide-angle image (over 100°) was cropped and adjusted in magnification to create two additional image categories: medium-angle and narrow-angle. Medium-angle images were adjusted to match the field of view (40°–50°) and size (50 mm diameter) of binocular indirect ophthalmoscopy, with the rationale that this is what is seen by examiners during clinical diagnosis. Narrow-angle images were adjusted to match the standard published photograph with regard to field of view (20°–30°), magnification (based on standardizing vertical optic disc height), and size (based on the image in the web-based version of the published article), with the rationale that this is used to train ophthalmologists for clinical diagnosis.^{4,24} In total, three categories of retinal images were created: narrow-angle, medium-angle, and wide-angle. All three fields of view for each of the 15 retinas were presented to expert study participants, for a total of 45 images.

Image Interpretation

All images were uploaded to a web server and stored in a secure database system (SQL 2005; Microsoft, Redmond, WA), with a web interface developed by the authors to display retinal photographs and collect responses. The order of image presentation was standardized for all participants: three sequential narrow-angle images were displayed (e.g. infants #1–3), followed by three sequential medium-angle images (e.g. infants #1–3), followed by three sequential wide-angle images (e.g. infants #1–3). This cycle was repeated (e.g. above with infants #4–6), until all 45 images were presented to expert participants. Narrow-angle and medium-angle images were not identified to expert participants as being cropped and magnified from the same original 15 wide-angle images.

Each expert was given access to the secure web-based review system using an anonymous identification number and password. Informed consent was obtained before any images were displayed. For each of the 45 images, all participants were asked to: (a) categorize the image based on presence of “plus disease,” “pre-plus disease,” or “neither plus nor pre-plus disease,” and (b) rate their confidence in diagnosis (“confident,” “somewhat confident,” or “not confident”). Finally, experts were asked about previous experience interpreting RetCam images (“extensive,” “limited,” or “none”). No reference images or definitions for plus or pre-plus disease were provided, so as to better simulate a real-world diagnostic environment.

Data Analysis

Data were downloaded from the database system onto a spreadsheet (Excel 2008; Microsoft, Redmond, WA). Absolute inter-expert agreement was determined for each image, with the images grouped by angle of view, based on both 3-level (“plus,” “pre-plus,” “neither”) and 2-level (“plus,” “not plus”) categorizations.

The kappa statistic for inter-expert reliability was calculated for each pair of experts for each image based on 3-level and 2-level categorizations. Kappa values were calculated for each pair of experts, and then used to calculate the mean kappa of each expert compared to all other experts. Weighted intra-expert kappas were also calculated to compare agreement within an expert’s three ratings of different images of the same retina (i.e. low agreement

would imply that field of view is relevant to diagnosis). A published scale was used to interpret kappa values: 0–0.20, slight agreement, 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, near-perfect agreement.²⁵ All statistical calculations were performed using R (R v.2.11.1, GNU General Public License).

RESULTS

Characteristics of Expert Participants

Fifteen experts from around the world were invited for this study, of whom 13 (87%) consented to participate and completed the study. Among these 13 participants, 8 (62%) had served as a Principal Investigator or certified investigator in the CRYO-ROP or ETROP study, and 12 (92%) had coauthored at least 5 published ROP papers. Seven (54%) participants fulfilled more than 1 criterion.

Among the 13 participants, 12 (92%) worked at an institution with a RetCam. 10 (77%) experts described their experience interpreting RetCam images as “extensive”, 3 (23%) described it as “limited,” and no experts described it as “none.” Eight (62%) experts were pediatric ophthalmologists, and 5 (38%) were retina specialists.

Participant Responses

Each expert participant evaluated all 45 study images, resulting in a total of 585 ratings (195 from each of the 3 fields of view). Among the 585 total ratings, a diagnosis of plus disease was given 179 (31%) times, a diagnosis of pre-plus disease was given 291 (50%) times, and a diagnosis of neither plus nor pre-plus disease was given 115 (20%) times. The profile of expert responses is shown in Figure 1. Among all images considered together, confidence in diagnosis was described as “confident” 47% of the time, “somewhat confident” 46% of the time, and “not confident” 7% of the time.

Inter-expert Agreement

Inter-expert agreement was calculated to assess consistency of expert diagnoses based on field of view. As shown in Figure 2, wide-angle images generally had higher levels of inter-expert agreement than medium-angle and narrow-angle images. For example, in the 3-level classification, 70% or more experts agreed on the same diagnosis in 8/15 (53%) wide-angle images, but in only 3/15 (20%) medium-angle and 3/15 (20%) narrow-angle images. In the 2-level classification, 80% or more experts agreed on the same diagnosis in 11/15 (73%) wide-angle images, but only in 9/15 (60%) medium-angle and 3/15 (20%) narrow angle images.

Figure 3 displays the mean kappa statistics for diagnostic agreement of each expert compared to all other experts. In the 3-level classification, the mean kappa of each expert compared to all other experts was 0.21–0.40 in 6/13 (46%) experts and 0.41–0.60 in 4/13 (31%) experts using wide-angle images, was 0–0.20 in 9/13 (69%) experts using medium angle images, and was 0–0.20 in 9/13 (69%) experts using narrow-angle images. In the 2-level classification, the mean kappa of each expert compared to all other experts was 0.41–0.60 in 8/13 (62%) experts using wide-angle images, was 0–0.20 in 7/13 (54%) experts using medium-angle images, and was 0.21–0.40 in 9/13 (69%) experts using narrow-angle images.

Intra-expert Agreement

Experts graded each retina using wide-angle, medium-angle, and narrow-angle views, and intra-expert diagnostic agreement using these three fields of view is displayed in Figure 4.

Six (46%) of the 13 experts had weighted intra-expert kappa from 0.21–0.40 (fair agreement).

When comparing the 195 total sets of 3-level diagnostic responses from all 15 retinas by all 13 experts, the same diagnosis was made by experts in all three fields of view in 77/195 (39%) sets, and the same diagnosis was made in two of the three fields of view in 113/195 (58%) sets. Among the 113 sets when an expert made the same diagnosis in only two of the three fields of view, narrow-angle imaging gave the outlying diagnosis 52/113 (46%) times, medium-angle imaging gave the outlying diagnosis 40/113 (35%) times, and wide-angle imaging gave the outlying diagnosis 21/113 (19%) times. Representative examples are displayed in Figure 5.

When comparing the 195 total sets of 2-level diagnostic responses from all 15 retinas by 13 experts, the same diagnosis was made by experts in all three fields of view in 114/195 (58%) sets, and the same diagnosis was made in two of the three fields of view in 81/195 (42%) sets. Among the 81 times when an expert made the same diagnosis in only two of the three fields of view, narrow-angle imaging gave the outlying diagnosis 46/81 (57%) times, medium-angle imaging gave the outlying diagnosis 23/81 (28%) times, and wide-angle imaging gave the outlying diagnosis 12/81 (15%) times.

DISCUSSION

This study was designed to examine and compare the inter-expert and intra-expert reliability of plus disease diagnosis in wide-angle, medium-angle, and narrow-angle images from the same retinas. To our knowledge, this is the first study to investigate plus disease diagnosis as a function of field of view. The key findings are that: (1) agreement in plus disease diagnosis from wide-angle images is higher than from medium-angle and narrow-angle images, suggesting that peripheral retinal findings contribute diagnostic information; and (2) inter-expert agreement in plus disease diagnosis is imperfect.

In this study, diagnostic agreement using wide-angle images was significantly higher than agreement from smaller fields of view. Using a 3-level categorization, 70% experts agreed on the same diagnosis in 8/15 (53%) wide-angle images, but in only 3/15 (20%) medium-angle and 3/15 (20%) narrow-angle images (Figure 2). The mean kappa for each expert compared to all others ranged from 0.08–0.52 in wide-angle images, –0.10–0.23 in medium-angle images, and –0.08–0.29 in narrow-angle images (Figure 3). Analysis using a 2-level categorization showed consistent findings. From a different perspective, the intra-grader diagnostic agreement among wide-angle, medium-angle, and narrow-angle images ranged from –0.01–0.66. The largest number of experts had intra-grader kappa from 0.21–0.40 (fair agreement) (Figure 5). This level of disagreement with medium-angle and narrow-angle images suggests that peripheral retinal features, or perhaps magnification, are considered by experts during clinical plus disease diagnosis.

The landmark multi-center CRYO-ROP and ETROP studies showed that presence of plus disease is the most critical parameter for identifying treatment-requiring ROP.^{1,2} Therefore, appropriate clinical management requires accurate and consistent identification of plus disease. However, findings from this study support previous work showing that there may be significant diagnostic inconsistencies.^{5–8} Possible underlying reasons for discrepancy include differing subjective perceptions of vascular tortuosity and dilation during examination, variable interpretations of the amount and nature of abnormality required for plus disease, confusion from differences in magnification,²⁶ and differing interpretations of the significance of peripheral and mid-peripheral retinal features. Further studies to clarify the exact retinal features used by experts to characterize plus disease will be informative.

Plus disease is defined by a standard published narrow-angle photograph, which was selected by expert consensus during the 1980s. In this study, the consistency of plus disease diagnosis was weakest with narrow-angle images, which had similar magnification and field of view compared to the standard photograph. Furthermore, previous studies have suggested that the arterial tortuosity and venous dilation in the standard photograph are actually much less than what is being used by experts during clinical diagnosis.²⁶ This potential tendency to over-estimate the severity of the standard photograph may occur because the magnification of the standard photograph is much larger than what is used during routine clinical examination. Taken together, these results suggest that there are benefits of identifying newer reference photographs for plus disease with larger fields of view and with image magnifications that are closer to what is seen during clinical examination, as was done in the revised international classification of ROP.⁴ This may provide additional clinically-relevant features to train ophthalmologists toward more accurate and consistent diagnosis.

Computer-based image analysis has been proposed as a strategy for improving the objectivity and reproducibility of plus disease diagnosis.^{6,17–23} Pilot studies suggest that semi-automated analysis of retinal vascular tortuosity and dilation may identify plus disease with comparable or better accuracy than experts.^{6,17,19,20} However, the retinal field of view that should be processed by these computer-based systems is not known. For example, some studies have performed image analysis within cropped narrow-angle fields that are comparable to the standard published photograph, some have performed image analysis within cropped medium-angle images, some have used wide-angle fields of view, and others have used inconsistent fields of view.^{6,17–23,27–29} In the longer-term, a generalizable method for quantitative plus disease diagnosis will likely require analysis of standardized retinal vessels within a standard field of view. This study has implications for the development of computer-based analysis methods for plus disease, because of the finding that larger fields of view may be required to capture the required diagnostic information. Additional research will be required to create optimal vascular analysis algorithms, to identify the retinal features that are most relevant for clinical diagnosis, and to validate system performance compared to experts.

Several study limitations should be noted: (1) This study involved the presentation of images, rather than actual ophthalmoscopic examinations. Therefore, the narrow-angle and medium-angle study images may not represent true diagnostic situations, where examiners can visualize the peripheral retinal features using scleral depression. However, we emphasize that the purpose of this study is to investigate the anatomic location of retinal features that are used by examiners for plus disease diagnosis, rather than to simulate real-world clinical diagnosis. (2) Similarly, this study did not include features such as patient demographics, systemic factors, and rate of vascular change that may be considered by experts during real-world diagnosis.²³ In particular, progression toward plus disease is a continuous process rather than a point in time, and factors beyond vascular dilation and tortuosity may be important for diagnosis. In the future, comparison of serial photographs may be used to identify progression toward plus disease. However, we note that plus disease is defined based solely on retinal vascular abnormality at a static point in time. Furthermore, neither the standard published photograph nor other representative published photographs include any clinical features.^{1,4} (3) This study measured reliability of plus disease diagnosis, and did not include a reference standard for the “correct” diagnosis. We feel that a true gold standard for plus disease would be extremely difficult to define, and note that the CRYO-ROP and ETROP studies were not designed to assess presence of plus disease independent of other clinical features in ROP.^{1,2} However, to explore this factor, clinical diagnoses made during standard ophthalmoscopic examinations were compared with the 3-level and 2-level diagnoses made by the majority of study experts. In the 3-level classification, agreement of

image-based diagnosis with standard ophthalmoscopic diagnosis occurred in 14/15 (93%) wide-angle images, 10/15 (67%) medium-angle images, and 8/15 (53%) narrow-angle images. In the 2-level classification, agreement of image-based diagnosis with standard ophthalmoscopic diagnosis occurred in 14/15 (93%) wide-angle images, 11/15 (73%) medium-angle images, and 11/15 (73%) narrow-angle images. This further supports the notion that peripheral retinal findings contribute to plus disease diagnosis, and that there is variability in plus disease diagnosis among experts. (4) Eligibility criteria were based on an academic definition of expertise, in which participants were required to have assumed leadership roles in multi-center clinical trials or to have published clinical papers involving ROP. Therefore, the pattern of responses from study experts may not be representative of responses from the larger group of ophthalmologists with clinical expertise in ROP. We hypothesize that diagnostic consistency among the academic experts in this study would be higher than what would be seen among the overall larger group of clinical experts, because academic experts may have had additional training and calibration during multi-center clinical ROP studies. (5) This study was unable to distinguish between the diagnostic impact of differences in field of view from the impact of differences in image magnification. For example, narrow-angle images may have been judged differently by experts because they were unable to visualize important peripheral retinal features, or because the larger image magnification created misinterpretations of disease severity. Among cases in which narrow-angle imaging gave a different diagnosis from that of medium-angle and wide-angle images, it resulted in a more severe diagnosis 46/52 (88%) times and in a less severe diagnosis 6/52 (12%) times. Additional research may help clarify this issue.

In summary, our study findings suggest that mid-peripheral and peripheral retinal features are used by clinical experts during plus disease diagnosis. This raises concerns about the narrow-angle standard photograph of plus disease, which has been used to train and calibrate clinicians since the 1980s. These findings also have implications for identifying the retinal features that are used to define plus disease, which is important from a clinical perspective to improve diagnostic consistency and a scientific perspective to identify mechanisms underlying abnormal retinal vascular growth.

Acknowledgments

MFC is supported by grant EY19474 from the National Institutes of Health (Bethesda, MD). RR is supported by grant T35 HL007616 from the National Institutes of Health (Bethesda, MD). NJJ funded/supported by the Dr. Werner Jackstaedt Foundation (Wuppertal, Germany). RR, NJJ, CV, RG, DSC, and MFC are supported by unrestricted departmental funding from Research to Prevent Blindness (New York, NY). The authors would like to thank the 13 expert subjects who participated in this study.

References

1. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity. Preliminary results. Arch Ophthalmol. 1988; 106:471–9. [PubMed: 2895630]
2. Early Treatment For Retinopathy Of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. Arch Ophthalmol. 2003; 121:1684–94. [PubMed: 14662586]
3. The Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. Arch Ophthalmol. 1984; 102:1130–4. [PubMed: 6547831]
4. The International Classification of Retinopathy of Prematurity revisited. Arch Ophthalmol. 2005; 123:991–9. [PubMed: 16009843]
5. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. Arch Ophthalmol. 2007; 125:875–80. [PubMed: 17620564]

6. Gelman R, Jiang L, Du YE, Martinez-Perez ME, Flynn JT, Chiang MF. Plus disease in retinopathy of prematurity: pilot study of computer-based and expert diagnosis. *J AAPOS*. 2007; 11:532–40. [PubMed: 18029210]
7. Reynolds JD, Dobson V, Quinn GE, et al. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. *Arch Ophthalmol*. 2002; 120:1470–6. [PubMed: 12427059]
8. Wallace DK, Quinn GE, Freedman SF, Chiang MF. Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity. *J AAPOS*. 2008; 12:352–6. [PubMed: 18329925]
9. Roth DB, Morales D, Feuer WJ, Hess D, Johnson RA, Flynn JT. Screening for retinopathy of prematurity employing the RetCam 120: sensitivity and specificity. *Arch Ophthalmol*. 2001; 119:268–72. [PubMed: 11176990]
10. Ells AL, Holmes JM, Astle WF, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology*. 2003; 110:2113–7. [PubMed: 14597517]
11. Chiang MF, Keenan JD, Starren J, et al. Accuracy and reliability of remote retinopathy of prematurity diagnosis. *Arch Ophthalmol*. 2006; 124:322–7. [PubMed: 16534051]
12. Wu C, Petersen RA, VanderVeen DK. RetCam imaging for retinopathy of prematurity screening. *J AAPOS*. 2006; 10:107–11. [PubMed: 16678743]
13. Chiang MF, Wang L, Busuioc M, et al. Telemedical retinopathy of prematurity diagnosis: accuracy, reliability, and image quality. *Arch Ophthalmol*. 2007; 125:1531–8. [PubMed: 17998515]
14. Scott KE, Kim DY, Wang L, et al. Telemedical diagnosis of retinopathy of prematurity intraphysician agreement between ophthalmoscopic examination and image-based interpretation. *Ophthalmology*. 2008; 115:1222–8 e3. [PubMed: 18456337]
15. Photographic Screening for Retinopathy of Prematurity (PHOTO-ROP) Cooperative Group. The photographic screening for retinopathy of prematurity study (photo-ROP). Primary outcomes Retina. 2008; 28 (3 Suppl):S45–54.
16. Silva RA, Murakami Y, Lad EM, Moshfeghi DM. Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDROP): 36-Month Experience with Telemedicine Screening. *Ophthalmic Surg Lasers Imaging*. 2011; 42:12–9.
17. Gelman R, Martinez-Perez ME, Vanderveen DK, Moskowitz A, Fulton AB. Diagnosis of plus disease in retinopathy of prematurity using Retinal Image multiScale Analysis. *Invest Ophthalmol Vis Sci*. 2005; 46:4734–8. [PubMed: 16303973]
18. Johnson KS, Mills MD, Karp KA, Grunwald JE. Semiautomated analysis of retinal vessel diameter in retinopathy of prematurity patients with and without plus disease. *Am J Ophthalmol*. 2007; 143:723–5. [PubMed: 17386296]
19. Koreen S, Gelman R, Martinez-Perez ME, et al. Evaluation of a computer-based system for plus disease diagnosis in retinopathy of prematurity. *Ophthalmology*. 2007; 114:e59–67. [PubMed: 18054630]
20. Wallace DK, Freedman SF, Zhao Z, Jung SH. Accuracy of ROPtool vs. individual examiners in assessing retinal vascular tortuosity. *Arch Ophthalmol*. 2007; 125:1523–30.
21. Wilson CM, Cocker KD, Moseley MJ, et al. Computerized analysis of retinal vessel width and tortuosity in premature infants. *Invest Ophthalmol Vis Sci*. 2008; 49:3577–85. [PubMed: 18408177]
22. Wallace DK, Freedman SF, Zhao Z. Evolution of plus disease in retinopathy of prematurity: quantification by ROPtool. *Trans Am Ophthalmol Soc*. 2009; 107:47–52. [PubMed: 20126481]
23. Thyparampil PJ, Park Y, Martinez-Perez ME, et al. Plus disease in retinopathy of prematurity: quantitative analysis of vascular change. *Am J Ophthalmol*. 2010; 150:468–75 e2. [PubMed: 20643397]
24. Capone A Jr, Ells AL, Fielder AR, et al. Standard image of plus disease in retinopathy of prematurity. *Arch Ophthalmol*. 2006; 124:1669–70. [PubMed: 17102031]
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]

26. Gelman SK, Gelman R, Callahan AB, et al. Plus disease in retinopathy of prematurity: quantitative analysis of standard published photograph. *Arch Ophthalmol.* 2010; 128:1217–20. [PubMed: 20837812]
27. Wallace DK. Computer-assisted quantification of vascular tortuosity in retinopathy of prematurity (an American Ophthalmological Society thesis). *Trans Am Ophthalmol Soc.* 2007; 105:594–615. [PubMed: 18427631]
28. Grunwald L, Mills MD, Johnson KS, et al. The rate of retinal vessel dilation in severe retinopathy of prematurity requiring treatment. *Am J Ophthalmol.* 2009; 147:1086–91. 91 e1–2. [PubMed: 19286152]
29. Chiang MF, Gelman R, Martinez-Perez ME, et al. Image analysis for retinopathy of prematurity diagnosis. *J AAPOS.* 2009; 13:438–45. [PubMed: 19840720]

Figure 1. Diagnostic responses of expert participants based on field of view

Thirteen experts were asked to diagnose presence of plus disease in 15 retinal images cropped to wide-angle, medium-angle, and narrow-angle fields of view. Responses are displayed using (A) 3-level (plus, pre-plus, neither) and (B) 2-level (plus, not plus) categorizations. The number of experts making each diagnosis is shown by field of view (out of 195 total responses for each angle).

Figure 2. Absolute inter-expert agreement based on field of view

Thirteen experts were asked to diagnose presence of plus disease in 15 retinal images cropped to wide-angle, medium-angle, and narrow-angle fields of view (45 images total). Percentage of experts in agreement of an image diagnosis is shown by field of view and for an increasing number of images. The number of images having a particular percentage of expert agreement are displayed using (A) 3-level (plus, pre-plus, neither) and (B) 2-level (plus, not plus) categorizations.

Figure 3. Inter-expert kappa statistic based on field of view

Thirteen experts were asked to diagnose presence of plus disease in 15 retinal images cropped to wide-angle, medium-angle, and narrow-angle fields of view (45 images total). The mean kappa of each expert compared to all other experts was calculated for each field of view. The strength of agreement by field of view is shown, as indicated by the number of kappas in certain numerical ranges: a higher kappa implies greater strength. (A) 3-level (plus, pre-plus, neither) and (B) 2-level (plus, not plus) categorization.

Figure 4. Intra-expert kappa statistic based on field of view

Thirteen experts were asked to diagnose presence of plus disease in 15 retinal images cropped to wide-angle, medium-angle, and narrow-angle fields of view (45 images total). Intra-expert kappa measures agreement of an expert's ratings of the three images of the same retina, one image from each angle. Results are displayed using a 3-level (plus, pre-plus, neither) categorization.

Figure 5. Representative examples of intra-expert agreement

Thirteen experts were asked to diagnose presence of plus disease in 15 retinal images cropped to wide-angle, medium-angle, and narrow angle fields of view. (A), (B), and (C) display wide-angle, medium-angle, and narrow-angle views from one retina. (A) was diagnosed as pre-plus by 9/13 experts, (B) was diagnosed as pre-plus by 8/13 experts, and (C) was diagnosed as pre-plus by 1/13 experts and plus by 12/13 experts. (D), (E), and (F) display wide-angle, medium-angle, and narrow-angle views from a second retina. (D) was diagnosed as pre-plus by 10/13 experts, (E) was diagnosed as pre-plus by 10/13 experts, and (F) was diagnosed as pre-plus by 3/13 experts and plus by 10/13 experts.