



Published in final edited form as:

JAMA Ophthalmol. 2013 August ; 131(8): 1026–1032. doi:10.1001/jamaophthalmol.2013.135.

Plus disease in retinopathy of prematurity: qualitative analysis of diagnostic process by experts

Nina J. Hewing, MD^{1,2}, David R. Kaufman, PhD³, RV. Paul Chan, MD¹, and Michael F. Chiang, MD⁴

¹Department of Ophthalmology Weill Cornell Medical College New York, New York

²Department of Ophthalmology Campus Benjamin Franklin Charité - Universitaetsmedizin Berlin Berlin, Germany

³Department of Biomedical Informatics Columbia University College of Physicians and Surgeons New York, New York

⁴Departments of Ophthalmology & Medical Informatics and Clinical Epidemiology Casey Eye Institute Oregon Health & Science University Portland, Oregon

Abstract

Purpose—To examine the diagnostic reasoning process of experts for plus disease in retinopathy of prematurity (ROP) using qualitative research techniques.

Methods—Six experts were video-recorded while independently reviewing seven wide-angle retinal images from infants with ROP. Experts were asked to explain their diagnostic process in detail (think-aloud protocol), mark findings relevant to their reasoning, and diagnose each image (plus vs. pre-plus vs. neither). Subsequently, each expert viewed the images again while being asked to examine arteries and veins in isolation, and answer specific questions. Video-recordings were transcribed and reviewed. Diagnostic process of experts was analyzed using a published cognitive model.

Results—Based on the think-aloud protocol, 5/6 experts agreed on the same diagnosis in three study images, and 3/6 experts agreed in three images. When experts were asked to rank images in order of severity, the mean correlation coefficient between pairs of experts was 0.33 (range - 0.04–0.75). All experts considered arterial tortuosity and venous dilation while reviewing each image. Some considered venous tortuosity, arterial dilation, peripheral retinal features, and other factors. When experts were asked to re-review images to diagnose plus disease based strictly on definitions of sufficient arterial tortuosity and venous dilation, all but 1 expert changed their diagnosis compared to the think-aloud protocol.

Conclusions—Diagnostic consistency in plus disease is imperfect. Experts differ in their reasoning process, retinal features that they focus on, and interpretations of the same features. Understanding these factors may improve diagnosis and education. Future research defining more precise diagnostic criteria may be warranted.

Corresponding author and reprints: Michael F. Chiang, MD Oregon Health & Science University 3375 SW Terwilliger Boulevard Portland, OR 97239 Tel: 503-418-3087 | Fax: 503-494-5347 | chiangm@ohsu.edu.

The authors have no commercial, proprietary, or financial interest in any of the products or companies described in this article. MFC is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA). MFC had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

INTRODUCTION

Retinopathy of prematurity (ROP) is a vasoproliferative disease affecting low birth-weight infants. An international classification for ROP (ICROP) has been developed to standardize clinical diagnosis.¹ Multi-center randomized trials such as the Cryotherapy for ROP (CRYO-ROP) and Early Treatment for ROP (ETROP) studies have found that severe ROP may be treated successfully by laser photocoagulation or cryotherapy.^{2,3} Pharmacological treatments are being studied.⁴ Despite these advances, ROP continues to be a leading cause of childhood blindness throughout the world.

“Plus disease” is a critical component of the ICROP system, and is defined as arterial tortuosity and venous dilation in the posterior pole greater than or equal to that of a standard photograph selected by expert consensus during the 1980s.^{1,2} More recently, the revised ICROP system defines “pre-plus disease” as vascular abnormalities insufficient for plus disease but with more arterial tortuosity and venous dilation than normal.¹ Presence of plus disease is a necessary feature for threshold disease and a sufficient feature for type 1 ROP, both of which have been shown to warrant prompt treatment. Therefore, accurate diagnosis of plus disease is essential.

However, there are limitations regarding the definition of plus disease. Studies have found diagnostic inconsistency, even among experts.⁵⁻⁷ The standard published photograph has a larger magnification and narrower field of view than clinical evaluation tools such as indirect ophthalmoscopy and wide-angle retinal imaging, and this difference in perspective may cause difficulty for ophthalmologists.^{8,9} Vessels in the standard published photograph have varying degrees of tortuosity and dilation, creating uncertainty regarding which vessels to focus on during examination. Finally, although plus disease is defined solely from arteriolar tortuosity and venous dilation within the posterior pole, it is possible that other vascular features or the rate of vascular change are relevant for diagnosis.^{9,10} Better understanding of the examination features characterizing plus disease may improve diagnostic accuracy and consistency.

It has been our observation that many ophthalmologists who trained within the past 25 years, after dissemination of ICROP and the CRYO-ROP study findings,^{1,2} perform ROP examination predominantly by classifying the zone, stage, and presence of plus disease based on “venous dilation and arteriolar tortuosity of the posterior vessels.” The premise of this study is that reliance only on this classification system, without attention to description of rich underlying retinal features, may over-simplify the characterization of clinically-significant findings. This study is designed to encode detailed qualitative thoughts of experts during plus disease diagnosis, using research methods from cognitive informatics.¹¹ The overall goals are to ascertain levels of agreement as well as to better understand underlying reasons for diagnostic discrepancy among experts, and to obtain more precise information about specific retinal features of plus disease.

METHODS

This study was approved by the Institutional Review Boards at Columbia University (New York, NY) and Oregon Health & Science University (Portland, OR). Informed consent was obtained from all expert participants, and waiver of consent was obtained for use of de-identified retinal images.

Expert Participants

The authors assembled a panel of international ROP experts who had the experience of using qualitative retinal features as their primary basis for clinical diagnosis. In our view,

this could be accomplished by identifying experts who had practiced ophthalmology before publication of the CRYO-ROP findings,² who participated as CRYO-ROP principal investigators, or who participated on national ROP standards committees. The rationale was that this would identify a small number of experts with the background and perspective to articulate their underlying qualitative reasons for diagnosis.

Image Selection

A set of 7 wide-angle retinal images (RetCam; Clarity Medical Systems, Pleasanton, CA) was captured from premature infants during routine clinical ROP care. Each image showed the posterior retina, and reflected some degree of vascular abnormality in the opinion of the authors. Images were printed on high-resolution photo paper (Kodak; Rochester, NY) in a 5" × 7" format.

No additional information such as birth-weight, systemic findings, or post-menstrual age was provided. This was to ensure that experts focused only on retinal features, without potential confounding factors. Neither the standard photograph nor any definitions of plus or pre-plus disease were provided to experts. This was to simulate a real-world examination scenario and avoid biasing expert opinions. We believed that study experts would be intimately familiar with these definitions through previous experiences, and through contributing to the creation of those definitions in many cases.

Think-aloud protocol and specific image questions

This study was conducted in two rounds, in which each study expert was asked a series of scripted questions (**Appendix**) by a co-author (NJH): (1) *Round 1* ("think-aloud protocol"). The 7 retinal images were shown individually and in the same order to each expert, who was asked to diagnose each image as either "plus disease," "pre-plus disease," or "neither plus nor pre-plus." Each expert was asked to verbalize thoughts while reviewing the image, to explain the process that led to the final diagnosis, and to annotate the most important findings on the printed image using a marking pen. Finally, each expert was asked to rate the degree of confidence in the diagnosis for each image (certain, somewhat certain, or uncertain). Experts were encouraged by the observer (NJH) to verbalize all of their thoughts, but were not otherwise interrupted or coached during the think-aloud protocol. (2) *Round 2* ("specific questions"). The 7 study images were displayed again in the same order, and each expert was asked a series of specific questions about each image. For each image, experts were asked whether the arteriolar tortuosity was sufficient for plus disease, whether the venous dilation was sufficient for plus disease, and whether the overall image reflected plus disease, pre-plus disease, or neither. Experts were asked to rank the 7 images in order of increasing arteriolar tortuosity, increasing venous dilation, and increasing overall severity of vascular abnormality. Additional specific questions were custom-tailored to each image regarding features used by experts to identify severe ROP, perceptions about the nature and location of vascular abnormalities, and other diagnostic heuristics (**Appendix**).

Each of the expert sessions was recorded using a video camera (Handycam; Sony, Tokyo, Japan). A digital recorder (Garage Band; Apple, Cupertino, CA) was used as a backup. The video camera was directed to record the retinal images and hands of each expert. Personal features of experts were not recorded, and experts were identified only by a study number.

Data analysis

In Round 1 (think-aloud protocol): Digital files were processed using video editing software (iMovie; Apple, Cupertino, CA). All video and audio files were manually transcribed for analysis. A modified protocol of the Hassebrock coding-scheme was used to analyze the transcribed files.¹² The scheme was designed to analyze medical reasoning and coding of

verbal think-aloud protocols. Inter-expert agreement was examined based on overall diagnosis provided after the think-aloud protocol. Specific examples were identified to represent differences in underlying qualitative diagnostic rationale among experts.

In Round 2 (specific questions): inter-expert agreement was examined by calculating correlation coefficients among each pair of experts who were asked to rank the 7 retinal images from least to most severe based on arterial tortuosity alone, venous dilation alone, and overall severity of vascular abnormalities related to plus disease. A published scale was used to interpret correlation coefficients: 0–0.30, small correlation; 0.31–0.50, medium correlation; 0.51–1.00, strong correlation.¹³

Intra-expert agreement in plus disease diagnosis was calculated. As described above, each expert initially provided a diagnosis (plus, pre-plus, or neither) in Round 1 while “thinking aloud” to explain their rationale. Each expert then provided a diagnosis in Round 2 after responding to a series of questions about specific image features. Absolute intra-expert agreement and weighted kappa (k) statistic were calculated for each expert using these diagnoses. A published scale was used to interpret k values: 0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, near-perfect agreement.¹⁴

Finally, data from Rounds 1 and 2 were analyzed to identify qualitative retinal features contributing to the ROP diagnostic process by experts, and to identify the relationship among individual retinal features and overall diagnosis.

RESULTS

Characteristics of Study Experts

Six ROP experts participated: 5/6 (83%) were principal investigators in the CRYO-ROP and/or ETROP studies, 5/6 (83%) published 5 peer-reviewed ROP papers, 5/6 (83%) practiced ophthalmology before publication of initial CRYO-ROP findings, and 5/6 (83%) contributed to expert consensus activities such as selection of the standard published photograph, development of ICROP, or creation of screening guidelines.^{15,16} All experts met 2 of these criteria.

Inter-Expert Agreement

Table 1 summarizes inter-expert agreement in plus disease diagnosis among the 6 experts based on the Round 1 think-aloud protocol. In particular, 5/6 (83%) experts agreed on the same diagnosis in three images, 4/6 (67%) experts agreed in one image, and 3/6 (50%) agreed in three images. Several images were diagnosed differently by experts. For example, image #2 was diagnosed as plus disease by 3/6 (50%) experts, as pre-plus disease by 1/6 (17%) experts, and as neither by 2/6 (33%) experts.

In Round 2, experts were asked to rank all 7 images in order of increasing overall severity of vascular abnormality related to plus disease, in order of increasing arterial tortuosity alone, and in order of increasing venous dilation alone. The correlation in ordering of arterial tortuosity among pairs of experts was strong (mean [range] correlation coefficient 0.89 [0.80–1.00]), whereas there was only small correlation in ordering of venous dilation (mean [range] correlation coefficient 0.27 [–0.04–1.00]) (Table 2).

Example of Qualitative Analysis of Diagnostic Discrepancy

To examine differences in underlying diagnostic process that may have led to discrepancies among experts, transcripts of think-aloud protocols were examined and compared. For

example, image #5 was diagnosed as plus disease by 2/6 (33%) experts, pre-plus by 3/6 (50%) experts, and neither by 1/6 (17%) expert. The Figure displays examples of differences in key retinal features that were discussed and annotated in that image by three different experts.

Intra-expert Agreement

Table 3 summarizes intra-expert agreement between plus disease diagnosis provided in the think-aloud protocol from Round 1 and the diagnosis provided after responding to a series of specific questions about image features in Round 2. Absolute intra-expert agreement ranged from 4/7 (57%, 1 expert) to 7/7 (100%, 1 expert), and weighted k ranged from 0.30 (fair agreement, 1 expert) to 1.00 (perfect agreement, 2 experts).

Retinal Features of Plus Disease

In specific questions of Round 2, experts were asked to characterize arterial tortuosity sufficient or insufficient for plus disease, characterize venous dilation sufficient or insufficient for plus disease, and provide an overall diagnosis (Table 4). Five individual ratings were excluded because an expert provided no response about arterial tortuosity or venous dilation; therefore, there were 37 total ratings by the 6 experts for the 7 images. In 5/37 (14%) ratings, there was inconsistency between the expert diagnostic process and published definitions of plus disease (which requires both sufficient arterial tortuosity and venous dilation).¹ In another 5/37 (14%) ratings, there was inconsistency with the published definition of pre-plus disease (which requires arterial tortuosity and venous dilation which is insufficient for plus disease).¹

Table 5 displays retinal features considered by experts in plus disease diagnosis during the think-aloud protocol of Round 1. In addition to retinal features mentioned in the published definition of plus disease (sufficient arterial tortuosity and venous dilation within 2 quadrants of the central retina),¹ experts cited many different features such as venous tortuosity, arterial dilation, peripheral retinal features, and vascular branching.

DISCUSSION

To our knowledge, this is the first study using qualitative research methods to examine the process of plus disease diagnosis by ROP experts. Key findings are that: (1) There are inconsistencies in plus disease diagnosis among experts, (2) Some diagnostic discrepancies may occur because experts are considering different retinal features, and (3) The current concept of plus disease as arteriolar tortuosity and venous dilation within the posterior pole appears over-simplified based on expert behavior.

Our results regarding inter-expert disagreement in plus disease diagnosis support findings from previous studies involving image-based diagnosis,^{5,6} and from a previous study showing that certified CRYO-ROP experts performing unmasked ophthalmoscopic examinations to confirm presence of threshold disease disagreed with the first expert diagnosis in 12% of cases.⁷ One demonstration of inter-expert inconsistency in the current study is shown in Table 1. Another demonstration is summarized in Table 2, showing that correlation among experts for ranking severity of arterial tortuosity (mean correlation coefficient 0.89) was much higher than correlation for ranking severity of venous dilation (mean correlation coefficient 0.27) or overall vascular severity (mean correlation coefficient 0.33). This suggests that arterial tortuosity is easier for experts to recognize and order visually. Conversely, venous dilation may be more difficult to identify visually, more subjective, or perhaps more difficult to represent using wide-angle images. There are several possible reasons explaining the low inter-expert correlation for overall severity, including

the fact that there are differences in retinal features considered by different experts. Other possibilities are that there are differences in the interpretation of the same retinal features by different experts, or that the significance of particular features is weighted differently among experts (Figure). A final demonstration of variability is summarized in Table 3, showing intra-expert differences in plus disease diagnosis using different methods. Taken together, these findings suggest that there are significant inconsistencies, and that experts appear to consider different retinal features and interpret the same features differently.

The traditional definition of plus disease was created by expert consensus during the 1980s, and has been used for major multi-center trials.^{2,3} However, another key finding from the current study is that experts consider retinal features beyond arterial tortuosity and venous dilation within the posterior pole when diagnosing plus disease. As shown in Table 5, experts explicitly mentioned many additional factors while explaining their diagnostic rationale during the think-aloud protocol. These included retinal features such as venous tortuosity and vascular branching, as well as anatomic factors such as peripheral retinal appearance and macular features, none of which are described in the published definition of plus disease.^{1,2} Furthermore, as shown in Table 4, there were 10/37 study ratings in which expert diagnoses of plus or pre-plus disease were inconsistent with published definitions.¹ Overall, these findings suggest that plus disease diagnosis is considerably more complex than current rules which combine arterial tortuosity and venous dilation in the posterior pole, and that experts do not appear to consider the same retinal features even when examining the same images.

Qualitative cognitive research methods have been used to characterize complex processes pertaining to visual diagnosis in fields such as dermatology, pathology, and radiology.¹⁷⁻¹⁹ The premise of this study is that current ROP management strategies are based on an international classification system,¹ along with diagnosis and treatment guidelines resulting from groundbreaking multi-center trials.^{2,3} By nature, this translates the qualitative nuances of retinal examination into discrete evidence-based rules. Findings from the current study support the notion that plus disease diagnosis is over-simplified by these rules involving only central arterial tortuosity and venous dilation. In particular, most study experts had practice experience before publication of these definitions and rules, and might therefore have greater insight about ROP diagnosis based on qualitative retinal characteristics. Follow-up research to encode the diagnostic methods and heuristics used by these experts may improve standardization and education in ROP care.^{20,21} Evidence-based protocols provide enormous benefits through guidelines to improve clinical management, and methods from this study can complement these protocols by providing additional information about subtle diagnostic factors.

Computer-based image analysis is an emerging method for improving accuracy and reproducibility of plus disease diagnosis using quantitative retinal vascular parameters.^{8,10,22-35} Development of these systems requires identification of the relevant vascular features to analyze (e.g. arterial tortuosity); definition of algorithms for quantifying these features; selection of the appropriate vessels for analysis (e.g. all vessels, worst vessels); and combination of individual feature values into an overall diagnosis.³⁴ Currently, there are no standard methods for performing most of these tasks. This study provides information about the diagnostic process used by experts, which may eventually provide a scientific basis for developing computer algorithms that better mimic expert diagnosis.

Several additional study limitations should be noted: (1) Wide-angle retinal images were used for expert review, rather than ophthalmoscopic exams. This may have biased findings to the extent that examiners were less familiar with image-based diagnosis. However, we note that all study experts had experience with ROP imaging, and that multiple studies have

shown that image-based ROP diagnosis agrees closely with ophthalmoscopic diagnosis.³⁶⁻⁴⁴ We felt that image review was the best study design to allow multiple experts to analyze the exact same retinal features. (2) Retinal images were reviewed by experts with no clinical information. This may have affected study findings to the extent that experts interpret retinal findings in the context of clinical data. However, the purpose of this study was to understand the significance of qualitative retinal features and the expert diagnostic process, not to simulate the process of ophthalmoscopic examination. (3) The number of study experts was limited. This may affect the generalizability of study findings to the extent that these 6 academic experts may not be representative of the larger group of clinical ROP specialists. However, we note that the foundation of qualitative research is to collect detailed verbal descriptions to portray varying perspectives about complex phenomena.⁴⁵ (4) This study focused only on identifying factors relevant to plus disease diagnosis. Other potentially-relevant factors, such as location of retinal disease, were not explicitly asked about but were noted if mentioned by experts (e.g. Table 5). New research, such as studies relating vascular appearance with zone, may be useful.

In summary, this study suggests that agreement in plus disease diagnosis among experts is imperfect, and that there are differences in the underlying diagnostic reasoning process and the retinal features examined. This study provides evidence that plus disease diagnosis is based on multiple factors, which may depend on the specific examiner. Updated definitions based on detailed analysis of expert behavior, using qualitative research methods such as those used in this study, may lead to improved diagnostic accuracy and standardization. This may have implications for future definitions of plus disease, for education and consistency of care, and for development of computer-based diagnostic tools.

Acknowledgments

Supported by grant EY19474 from the National Institutes of Health (Bethesda, MD) (MFC), the Dr. Werner Jackstaedt Foundation, Wuppertal, Germany (NJH), the Friends of Doernbecher foundation (Portland, OR) (MFC), unrestricted departmental funding from Research to Prevent Blindness (New York, NY) (RVPC, MFC), the St. Giles Foundation (New York, NY) (RVPC). The authors are very grateful to the 6 experts who generously agreed to participate in this study. Portions of this study were presented at the 2012 ARVO Annual Meeting (Ft. Lauderdale, FL).

REFERENCES

1. International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol*. 2005; 123:991-9. Comment in: *Arch Ophthalmol* 2006; 124:1669-70. [PubMed: 16009843]
2. 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]
3. 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]
4. Mintz-Hittner HA, Kennedy KA, Chuang AZ, BEAT-ROP Cooperative Group. Efficacy of intravitreal bevacizumab for stage 3+ retinopathy of prematurity. *N Engl J Med*. 2011; 364:603-15. [PubMed: 21323540]
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. 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]

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. 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]
9. Rao R, Jonsson NJ, Ventura C, et al. Plus disease in retinopathy of prematurity: diagnostic impact of field of view. *Retina.* 2012; 32:1148–55. [PubMed: 22466473]
10. Thyparampil PJ, Park Y, Martinez-Perez ME, et al. Plus Disease in Retinopathy of Prematurity: Quantitative Analysis of Vascular Change. *Am J Ophthalmol.* 2010
11. Ericsson, KS. H. Protocol Analysis: Verbal Reports as Data. 2nd ed.. MIT Press; Boston: 1993.
12. Hassebrock F, Prietula M. A protocol-based scheme for the analysis of medical reasoning. *Int J Man-Machine Stud.* 1992; 37:613–52.
13. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull.* 1968; 70:213. [PubMed: 19673146]
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–74. [PubMed: 843571]
15. Section on Ophthalmology, American Academy of Pediatrics; American Academy of Ophthalmology; American Association for Pediatric Ophthalmology and Strabismus. Screening examination of premature infants for retinopathy of prematurity. *Pediatrics.* 2006; 117:572–6. Erratum in: *Pediatrics* 2006; 118:1324. [PubMed: 16452383]
16. Wilkinson AR, Haines L, Head K, Fielder AR. UK retinopathy of prematurity guideline. *Eye.* 2009; 23:2137–9. [PubMed: 18836408]
17. Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Dermatol.* 1989; 125:1063–8. [PubMed: 2757402]
18. Crowley RS, Naus GJ, Stewart J 3rd, Friedman CP. Development of visual diagnostic expertise in pathology -- an information-processing study. *J Am Med Inform Assoc.* 2003; 10:39–51. [PubMed: 12509356]
19. Azevedo, R.; Faremo, S.; Lajoie, SP. Expert-novice differences in mammogram interpretation. In: Trafton, DSMJG., editor. *Proceedings of the 29th Annual Cognitive Science Society*; Austin, Tx. 2007.
20. Paul Chan RV, Williams SL, Yonekawa Y, Weissgold DJ, Lee TC, Chiang MF. Accuracy of retinopathy of prematurity diagnosis by retinal fellows. *Retina.* 2010; 30:958–65. [PubMed: 20168274]
21. Myung JS, Paul Chan RV, Espiritu MJ, et al. Accuracy of retinopathy of prematurity image-based diagnosis by pediatric ophthalmology fellows: implications for training. *J AAPOS.* 2011; 15:573–8. [PubMed: 22153403]
22. 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]
23. 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. [PubMed: 17998514]
24. Wallace DK, Jomier J, Aylward SR, Landers MB. Computer-automated quantification of plus disease in retinopathy of prematurity. *J AAPOS.* 2003; 7:126–30. [PubMed: 12736626]
25. Wallace DK, Freedman SF, Zhao Z. A Pilot Study Using Roptool to Measure Retinal Vascular Dilation. *Retina.* 2009; 29:1182–7. [PubMed: 19491720]
26. Swanson C, Cocker KD, Parker KH, Moseley MJ, Fielder AR. Semiautomated computer analysis of vessel growth in preterm infants without and with ROP. *Br J Ophthalmol.* 2003; 87:1474–7. [PubMed: 14660456]
27. 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]

28. 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]
29. Rabinowitz MP, Grunwald JE, Karp KA, Quinn GE, Ying GS, Mills MD. Progression to severe retinopathy predicted by retinal vessel diameter between 31 and 34 weeks of postconception age. *Arch Ophthalmol*. 2007; 125:1495–500. [PubMed: 17998510]
30. 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]
31. 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]
32. Shah DN, Karp KA, Ying GS, Mills MD, Quinn GE. Image analysis of posterior pole vessels identifies type 1 retinopathy of prematurity. *J AAPOS*. 2009; 13:507–8. [PubMed: 19840733]
33. Wallace DK, Zhao Z, Freedman SF. A pilot study using “ROPtool” to quantify plus disease in retinopathy of prematurity. *J AAPOS*. 2007; 11:381–7. [PubMed: 17532238]
34. Chiang MF, Gelman R, Martinez-Perez ME, et al. Image analysis for retinopathy of prematurity diagnosis. *Journal of Aapos*. 2009; 13:438–45. [PubMed: 19840720]
35. Chiang MF, Gelman R, Williams SL, et al. Plus disease in retinopathy of prematurity: Development of composite images by quantification of expert opinion. *Invest Ophth Vis Sci*. 2008; 49:4064–70.
36. 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]
37. 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]
38. Wu C, Petersen RA, VanderVeen DK. RetCam imaging for retinopathy of prematurity screening. *J AAPOS*. 2006; 10:107–11. [PubMed: 16678743]
39. 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]
40. 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]
41. Balasubramanian M, Capone A, Hartnett ME, et al. The photographic screening for retinopathy of prematurity study (Photo-ROP): primary outcomes. *Retina*. 2008; 28:S47–54. [PubMed: 18317345]
42. Lorenz B, Spasovska K, Elflein H, Schneider N. Wide-field digital imaging based telemedicine for screening for acute retinopathy of prematurity (ROP). Six-year results of a multicentre field study. *Graefes Arch Clin Exp Ophthalmol*. 2009; 247:1251–62. [PubMed: 19462177]
43. 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. [PubMed: 20954641]
44. Dai S, Chow K, Vincent A. Efficacy of wide-field digital retinal imaging for retinopathy of prematurity screening. *Clin Experiment Ophthalmol*. 2011; 39:23–9. [PubMed: 20796264]
45. Li AC, Kannry JL, Kushniruk A, et al. Integrating usability testing and think-aloud protocol analysis with “near-live” clinical simulations in evaluating clinical decision support. *Int J Med Inform*. 2012

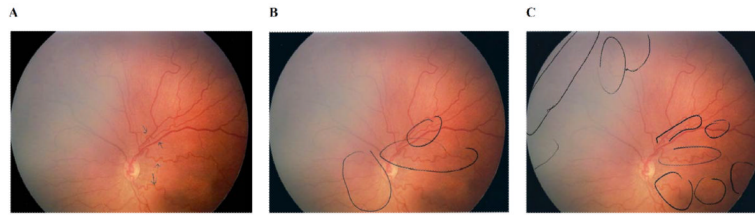


Figure. Example of differences in diagnostic process among experts reviewing the same image independently

Experts were asked to provide a diagnosis (plus, pre-plus, or neither) and annotate key findings, while being videotaped “thinking aloud” to describe their reasoning. Videotapes were transcribed, coded, and analyzed to examine qualitative diagnostic process. (A) was diagnosed as plus disease by Expert #1: “...looks like a very low gestational age baby, it's taken quite a long time to get to this stage. There is a lot of arterial tortuosity [annotated], there is a little bit of venous congestion in the superior temporal and superior nasal quadrant, more in the superior half of the retina [annotated]. By definition I think this has to be plus, because it's two quadrants at least, and even the other quadrants aren't normal...” (B) was diagnosed as pre-plus disease by Expert #2: “...there is a lot of tortuosity of the arteries, the veins are about 2 to 1. This could either be a pre-plus eye or a normal variant, depending on a quick look to the periphery. Curiously, there is a lot of tortuosity down here [annotated], it looks like there is disease up here [annotated].” (C) was diagnosed as neither by Expert #4: “...vessels seem to be branching excessively in that region [superonasal area annotated] and some increased tortuosity [superotemporal area annotated] as well, and this vein looks too fat [superotemporal area annotated]. If all the quadrants were like this quadrant [superotemporal] then it would be at least pre-plus and verging on plus, but since it's only one quadrant that's highly questionable. Would not classify it as plus, I could see why some would call it pre-plus ... I would call it no plus.”

Table 1
Inter-expert agreement in plus disease diagnosis by 6 retinopathy of prematurity (ROP)
experts from reviewing 7 wide-angle retinal images

Experts were asked to provide a diagnosis (plus, pre-plus, or neither) while “thinking aloud” to describe their underlying reasoning process.

Image	Diagnosis		
	Plus	Pre-Plus	Neither
1	0/6 (0%)	4/6 (67%)	2/6 (33%)
2	3/6 (50%)	1/6 (17%)	2/6 (33%)
3	3/6 (50%)	2/6 (33%)	1/6 (17%)
4	5/6 (83%)	1/6 (17%)	0/6 (0%)
5	2/6 (33%)	3/6 (50%)	1/6 (17%)
6	5/6 (83%)	1/6 (17%)	0/6 (0%)
7	5/6 (83%)	1/6 (17%)	0/6 (0%)

Table 2
Inter-expert agreement among experts ranking severity of overall vascular abnormality, arterial tortuosity alone, and venous dilation alone in retinopathy of prematurity (ROP)

Six experts were asked to rank 7 wide-angle retinal images from least to most severe, and correlation coefficient for rank ordering was compared among each pair of experts for each parameter (overall vascular abnormality, arterial tortuosity alone, venous dilation alone). Mean, highest, and lowest correlation coefficients among pairs of experts are displayed.

Parameter	Correlation coefficient among pairs of experts		
	Lowest	Mean	Highest
Overall vascular abnormality	-0.04	0.33	0.75
Arterial tortuosity alone	0.80	0.89	1.00
Venous dilation alone	-0.04	0.27	1.00

Table 3
Intra-expert agreement in plus disease diagnosis by 6 retinopathy of prematurity (ROP) experts from reviewing 7 wide-angle retinal images

Each expert was initially asked to provide a diagnosis (plus, pre-plus, or neither) while “thinking aloud” to explain their underlying rationale. Experts were subsequently asked to provide a diagnosis after responding to a series of questions about specific features of each image. Intra-expert agreement is represented as absolute agreement and weighted kappa (κ) statistic.

Expert	<u>Intra-expert agreement metric</u>	
	Absolute	Weighted κ
1	6/7 (86%)	0.78
2	7/7 (100%)	1.00
3	5/7 (71%)	0.36
4	4/7 (57%)	0.36
5	5/7 (71%)	0.30
6	6/7 (86%)	0.78

Table 4
Relationship between perceived vascular abnormality and overall plus disease diagnosis

Experts were asked to review wide-angle retinal images in detail, provide an overall diagnosis (plus, pre-plus, or neither), characterize arterial tortuosity (AT, sufficient or insufficient for plus disease), and characterize venous dilation (VD, sufficient or insufficient for plus disease). Shaded cells represent inconsistencies between expert diagnostic process and published definitions. Six experts reviewed 7 wide-angle images, for a total of 42 ratings. Five ratings were excluded because experts provided no response for AT or VD.

Diagnosis	Sufficient AT or VD for plus disease?			
	Both AT and VD	Only AT	Only VD	Neither AT nor VD
Plus	15/37	5/37	0/37	0/37
Pre-Plus	5/37	3/37	3/37	0/37
Neither	0/37	3/37	1/37	2/37

Table 5
Retinal features considered by experts during plus disease diagnosis

Experts were asked to review wide-angle retinal images in detail, while “thinking aloud” to describe their underlying reasoning process. Six experts reviewed 7 wide-angle images, for a total of 42 ratings. Shaded cells represent features explicitly mentioned in published definitions.

Retinal feature	Mentioned in image ratings
Arterial tortuosity	42/42 (100%)
Arterial dilation	8/42 (19%)
Venous tortuosity	10/42 (24%)
Venous dilation	42/42 (100%)
Central vessels	8/42 (19%)
Peripheral vessels	14/42 (33%)
Number of quadrants of abnormality	23/42 (55%)
Vascular branching	8/42 (19%)
Macular features	3/42 (7%)
Other retinal vascular features	7/42 (17%)