Title
Variability in Plus Disease Diagnosis using Single and Serial Images

Journal
Ophthalmology Retina, 6(12)

ISSN
2468-7219

Authors
Cole, Emily D
Park, Shin Hae
Kim, Sang Jin
et al.

Publication Date
2022-12-01

DOI
10.1016/j.oret.2022.05.024

Peer reviewed

Published in final edited form as:

Ophthalmol Retina. 2022 December ; 6(12): 1122–1129. doi:10.1016/j.oret.2022.05.024.

Variability in Plus Disease Diagnosis using Single and Serial Images

Emily D. Cole, MD, MPH[1], **Shin Hae Park, MD, PhD**[1,2], **Sang Jin Kim, MD, PhD**[3,4], **Kai B. Kang, MD**[1], **Nita G. Valikodath, MD, MS**[1], **Tala Al-Khaled, MD**[1], **Samir N. Patel, MD**[5], **Karyn E. Jonas, MSN, RN-BC**[1], **Susan Ostmo, MS**[3], **Aaron Coyner, PhD**[3], **Audina Berrocal, MD**[6], **Kimberly A. Drenser, MD**[7], **Aaron Nagiel, MD**[8], **Jason D. Horowitz, MD**[9], **Thomas C. Lee, MD**[10], **Jayashree Kalpathy-Cramer, PhD**[11], **Michael F. Chiang, MD**[12], **J. Peter Campbell, MD, MPH**[3], **R.V. Paul Chan, MD**[1] **Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium**

[1]Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago, Chicago, Illinois.

[2]Department of Ophthalmology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea.

[3]Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, Portland, Oregon.

[4]Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea.

[5]Wills Eye Hospital, Philadelphia, Pennsylvania.

Correspondence: R.V. Paul Chan, MD, MSc, MBA, 1855 W Taylor St, Chicago, IL 60612. rvpchan@uic.edu.

Author Contributions:

Conception and design: Park, Kim Kang, Patel, Jonas, Ostmo, Campbell, Chiang, Chan

Data collection: Cole, Park, Kim, Kang, Valikodath, Al-Khaled, Patel, Jonas, Ostmo, Campbell, Berrocal, Drenser, Nagiel, Lee, Horowitz, Chiang, Chan

Analysis and interpretation: Cole, Park, Kim, Kang, Valikodath, Al-Khaled, Patel, Jonas, Ostmo, Campbell, Berrocal, Drenser, Nagiel, Lee, Horowitz, Chiang, Chan

Obtained funding: Chan, Campbell, Chiang

Overall responsibility: Cole, Park, Kim, Kang, Valikodath, Al-Khaled, Patel, Jonas, Ostmo, Campbell, Berrocal, Drenser, Nagiel, Lee, Horowitz, Chiang, Chan

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form. The author(s) made the following disclosure(s): K.E.J.: Equity owner — Siloam Vision.

A.B.: Consulting fees — Alcon, DORC, Oculus, Zeiss.

J.K.-C.: Support — National Institutes of Health; Grants — Genentech; Royalties — Boston AI.

M.F.C.: Support — National Institutes of Health; Grants — National Institutes of Health, National Science Foundation, Genentech; Consulting fees — Novartis; Equity owner — Inteleretina LLC

J.P.C.: Grants — National Institutes of Health, Research to Prevent Blindness; Consulting fees — Boston AI; Co-founder — Siloam Vision; Equity owner — Siloam Vision; Support — Genentech.

R.V.P.C.: Grants — National Institutes of Health, Research to Prevent Blindness; Leadership role — Vit Buckle Society, Helen Keller International, Prevent Blindness America; Equity owner — Siloam Vision; Scientific Advisory Board — NeoLight LLC; Consulting fees — Alcon; Cofounder — Siloam Vision; Support — Genentech.

HUMAN SUBJECTS: Human subjects were included in this study. This study was approved as a prospective study by the Institutional Review Board at the University of Illinois at Chicago. Written informed consent for the study was obtained from the parents of all infants enrolled in this study, and a waiver of consent was obtained for the use of deidentified retinal images. This study was conducted in accordance with the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

[6]Bascom Palmer Eye Institute, University of Miami, Miami, Florida.

[7]Department of Ophthalmology, Beaumont Eye Institute, Royal Oak, Michigan.

[8]Stein Eye Institute, University of California Los Angeles, Los Angeles, California.

[9]Department of Ophthalmology, Columbia University, New York, New York.

[10]Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, California.

[11]Harvard Medical School, Boston, Massachusetts.

[12]National Eye Institute, National Institutes of Health, Bethesda, Maryland.

Abstract

Purpose: To assess changes in retinopathy of prematurity (ROP) diagnosis in single and serial retinal images.

Design: Cohort study.

Participants: Cases of ROP recruited from the Imaging and Informatics in Retinopathy of Prematurity (i-ROP) consortium evaluated by 7 graders.

Methods: Seven ophthalmologists reviewed both single and 3 consecutive serial retinal images from 15 cases with ROP, and severity was assigned as plus, preplus, or none. Imaging data were acquired during routine ROP screening from 2011 to 2015, and a reference standard diagnosis was established for each image. A secondary analysis was performed using the i-ROP deep learning system to assign a vascular severity score (VSS) to each image, ranging from 1 to 9, with 9 being the most severe disease. This score has been previously demonstrated to correlate with the International Classification of ROP. Mean plus disease severity was calculated by averaging 14 labels per image in serial and single images to decrease noise.

Main Outcome Measures: Grading severity of ROP as defined by plus, preplus, or no ROP.

Results: Assessment of serial retinal images changed the grading severity for > 50% of the graders, although there was wide variability. Cohen's kappa ranged from 0.29 to 1.0, which showed a wide range of agreement from slight to perfect by each grader. Changes in the grading of serial retinal images were noted more commonly in cases of preplus disease. The mean severity in cases with a diagnosis of plus disease and no disease did not change between single and serial images. The ROP VSS demonstrated good correlation with the range of expert classifications of plus disease and overall agreement with the mode class ($P = 0.001$). The VSS correlated with mean plus disease severity by expert diagnosis (correlation coefficient, 0.89). The more aggressive graders tended to be influenced by serial images to increase the severity of their grading. The VSS also demonstrated agreement with disease progression across serial images, which progressed to preplus and plus disease.

Conclusions: Clinicians demonstrated variability in ROP diagnosis when presented with both single and serial images. The use of deep learning as a quantitative assessment of plus disease has the potential to standardize ROP diagnosis and treatment.

Keywords

Retinopathy of prematurity; ROP imaging; ROP progression; ROP screening; telemedicine

Retinopathy of prematurity (ROP) is a leading cause of childhood blindness worldwide.[1-3] Proper screening and timely treatment of ROP are essential for successful anatomical and functional outcomes. Current screening and treatment guidelines are based on the Cryotherapy for ROP and the Early Treatment for ROP trials.[2-4] Treatment should be considered for any eye with type 1 ROP. For type 2 ROP, continued close retinal examinations should be primarily recommended, and treatment should be performed if progression to type 1 disease is detected. Plus disease is defined as arteriolar tortuosity and venous dilation within the posterior pole greater than that found in a standard published photograph.[5] Although the presence of plus disease is the most critical parameter for determining treatment-requiring ROP, the assessment of plus disease is currently subjective and qualitative with significant variability in diagnosis even among ROP experts.[6-12]

Retinopathy of prematurity is a disease with a continuous spectrum of retinal vascular abnormalities.[11] Early vascular dilation and tortuosity have been associated with worse structural outcomes, with the clinical implication that the identification of preplus disease can identify high-risk infants who will develop severe vision-threatening ROP even in eyes with vascular changes insufficient for a diagnosis of plus disease.[13-20] Many clinicians who manage ROP compare vascular changes through retinal drawings or photographs taken during screening without the systematic evaluation of serial retinal photographs.

Wide-angle digital photography has enabled the objective documentation of disease findings in ROP, creation of digital libraries for education and research, improvement in the accuracy and standardization of ROP diagnosis, and development of telemedicine programs for ROP.[21-23] Digital imaging has allowed the serial comparison of vascular changes of fundus images over time; however, there are little data on how serial retinal images may guide the modern management of ROP. The purpose of this study was to examine how serial color fundus photographs affected the diagnosis of ROP by ophthalmologists.

Methods

This study was approved as a prospective study by the Institutional Review Board at the University of Illinois at Chicago. Clinical data and images for this study were obtained from infants who underwent serial ROP screening examinations in accordance with current, evidence-based guidelines by participating ophthalmologists at 8 participating institutions: (1) Oregon Health & Science University; (2) Weill Cornell Medical College; (3) University of Miami; (4) Columbia University Medical College; (5) Children’s Hospital Los Angeles; (6) Cedars-Sinai Medical Center; (7) William Beaumont Hospital; and (8) Asociacion para evitarla Ceguera en Mexico. Written informed consent for the study was obtained from the parents of all infants enrolled in this study, and a waiver of consent was obtained for the use of deidentified retinal images. This study was conducted in accordance with the tenets of the Declaration of Helsinki.

Image Acquisition and Selection

Wide-angle fundus photographs were acquired during routine ROP screening examinations from 2011 to 2015 using the RetCam (Natus Medical Systems) at 8 participating academic institutions. For each image set, a reference standard diagnosis (RSD) of ROP was established using the previously published methods.[24] The RSD was determined by combining the clinical diagnosis at the time of ROP screening with the masked and independent interpretation of images performed by 3 expert graders (2 ophthalmologists and 1 ROP study coordinator: M.F.C., R.V.P.C., and S.O.). Fifteen eyes were selected for this study, with an RSD of 4 with plus disease, 9 with preplus disease, and 2 normal images. Each eye was required to have ≥ 3 consecutive wide-angle retinal photographs to demonstrate vascular progression over time with the following requirements: (1) serial images were 1—3 weeks apart; (2) sufficient image quality to demonstrate vascular changes; and (3) similarities in the field of view centered on the posterior pole. Each eye had an image set, which consisted of 3 sequential wide-angle retinal images labeled with the date of image acquisition.

Survey Participants

All 7 eligible graders for this study were board-certified practicing retinal specialists who have regularly performed ROP care at their institution and met ≥ one of the following criteria: (1) principal investigator or certified investigator for the Cryotherapy for ROP study or Early Treatment for ROP study; and/or (2) published ≥ 2 peer-reviewed ROP-related articles. These participants are further referred to as experts in this study.

Study Design

Each expert who agreed to provide informed consent was given an individual electronic link to the survey website. Baseline demographic data were initially collected, including the length of time in practice and type of fellowship completed (pediatric ophthalmology, medical retina, or surgical retina). Seven graders were presented with single retinal images from 15 cases with ROP. The single retinal image presented was the most recent image for that case. For each case, experts were asked to choose the ROP diagnosis (plus, preplus, or neither), follow-up and management plan based on vascular appearance (follow-up in 2 weeks, 1 week, < 1 week, treat, or other), and level of confidence in determining the clinical diagnosis based on the image presented (confident, somewhat confident, or not confident). Then, a second set composed of 3 serial retinal images was presented for each of 15 eyes. The series of images from the 15 eyes were displayed in the same sequential order. The graders were not permitted to review the previous images or change their responses. Experts were again asked to choose a diagnosis, management plan, and level of confidence in determining the clinical diagnosis. Additionally, graders were asked if serial images influenced their diagnosis or follow-up assessment.

Statistical Analysis

Data analysis was performed using IBM SPSS statistics, version 24 (IBM). Three-level plus diagnosis (plus, preplus, or neither) was used for statistical analysis. The intragrader agreement between 2 assessments and intergrader agreement of 7 graders when viewing

single versus serial images were analyzed using kappa statistics. Interpretation of the kappa statistic Cohen's and Fleiss' kappa statistic were used to assess agreement between 2 graders and between > 2 graders, respectively. Interpretation of the kappa statistic used a commonly accepted scale as follows: 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1, near-perfect agreement.[25] A *P* value of < 0.05 was considered statistically significant.

Deep Learning Analysis

A secondary analysis was performed using the previously validated Imaging and Informatics in Retinopathy of Prematurity (i-ROP) deep learning system to assign a vascular severity score (VSS) to each image. The i-ROP deep learning system was developed by the i-ROP consortium and has demonstrated expert-level classification of plus disease. In addition to classifying plus disease, the system has introduced a quantitative scale for ROP severity—the VSS—which ranges from 1 to 9, with 9 being the most severe disease. This score has been previously demonstrated to correlate with expert diagnosis and the International Classification of ROP. [26-30] The VSS for each image in this study was determined using the i-ROP deep learning system. The mean plus disease severity was calculated by averaging 14 labels per image in both serial and single images to decrease noise.

Results

All 7 experts had completed a surgical retinal fellowship and had been practicing ophthalmology for a mean of 15 years (range, 2—32 years) since the completion of training. Analysis of intragrader agreement between single and serial image assessment demonstrated a Cohen's kappa with a range of 0.29—1.0, which showed a wide range of agreement from slight to perfect by each grader. There was noted variability in the aggressiveness of the graders, defined as a tendency to assign a more severe ROP diagnosis. This was assessed by comparing serial and single mean severity of all images across 7 graders. There was a range of 4.20—6.07 for serial mean severity and 3.67—5.80 for single mean severity. Four of the 7 graders were significantly more aggressive when assessing both serial ($P = 0.004$) and single ($P = 0.0001$) images. The confusion matrices for the assessment of serial and single images are shown in Figure 1.

There was a nonsignificant difference in the mean grader-assigned severity in cases with no ROP and plus disease ($P = 0.1$) The mean severity in the 10 cases showed an average increase in the classification of disease severity with serial images ($P = 0.08$). Therefore, there was a nonsignificant tendency to increase classification label for preplus images with the use of serial images. The mode diagnosis changed in only 1 case from preplus to plus disease. Figure 2 demonstrates the mean and mode severity in single and serial images. The numerical scale correlates with disease severity, with 1 indicating no disease, 5 indicating preplus disease, and 9 indicating plus disease.

When viewing serial images, a diagnostic change from preplus to plus disease was the most common finding, and 7 eyes were given the diagnosis of treatment-requiring disease. Among the 15 eyes, 12 (80.0%) eyes had a change in the diagnosis or management of ROP after viewing serial images. There were 13 changes from preplus to plus disease, which represents

a change warranting treatment. Notably, a change in preplus to plus disease occurred in > 50% of the graders. Graders with a lower threshold to diagnose the more severe disease (more aggressive graders) tended to be influenced by serial images to make a change in diagnosis; 15 of the 24 total diagnostic changes occurred in the 2 most aggressive graders. Figure 3 displays all changes in diagnosis across all graders after viewing serial images, organized from the least to the most aggressive graders (left to right). Representative cases showing changes toward more and less severe disease when viewing serial retinal images are described in Figure 4.

Analysis of the VSSs demonstrated good overall agreement with the mode class as assigned to no ROP, preplus, and plus disease ($P = 0.001$), as demonstrated in Figure 5A. The VSS also correlated with the range of disease severity, as determined by averaging clinician responses of disease severity (no plus disease =1, preplus =5, and plus disease = 9) with a correlation coefficient of 0.8 (Fig 5B).

When assessing single images, in 81 (77.1%) of the images, graders reported that they were confident and for 24 (22.9%) images the graders reported that they were somewhat confident in their clinical diagnosis. With serial retinal images, in 85 (81.0%) of the images, graders reported that they were confident and for 20 images (19.0%), the graders reported that they were somewhat confident. There was no statistically significant difference in confidence in diagnosis between single and serial images ($P = 0.435$). Graders were asked whether the serial images influenced their diagnosis or follow-up. For thirty-five (33.3%) images, graders felt that their clinical decision-making was affected. In a total of 25 (23.8%) images, graders felt that follow-up was the most affected, in 6 images (5.7%), graders felt that both follow-up and diagnosis were affected, and in 4 images (3.8%), graders felt that only diagnosis was affected.

Discussion

The key findings from this study are as follows: (1) exposure to serial images at the time of plus disease classification changed the diagnosis of borderline cases from preplus to plus for more than half of the graders in ≥ 1 case. This may have implications for the treatment and application of evidence-based medicine; (2) graders who tended to have a lower threshold for the diagnosis of preplus and plus were more likely to be influenced by serial progression images; and (3) the use of a deep learning—derived quantitative VSS may standardize the assessment of plus disease among experts and may facilitate a better understanding of the role that disease progression plays in treatment decisions, which has not been evaluated in clinical trials.

The first key finding of this study is that exposure to serial images at the time of plus disease classification changed the diagnosis of borderline cases from preplus to plus for more than half of the graders in ≥ 1 case. Overall, most changes in diagnosis in our study (18 of 24 [75%]) were toward more severe disease, although clinicians were inconsistent in whether serial versus single image review changed their diagnosis. It has been well known that interexpert agreement in ROP classification varies with a single image because of differences among cut points of vascular abnormality required for plus disease, differences

in the field of view considered, identification of different vascular parameters by different clinicians, and differences in training and education.[6-12] Serial images have been previously used as a tool in telemedicine to assess follow-up after intravitreal anti-VEGF injection.[31] A single retinal image may be insufficient to understand the patient's individual disease progression, especially with regard to the tempo of disease progression.[32] There has been a large amount of interobserver variability in the assessment of plus and preplus disease with single retinal photographs, which was also noted in the current study.[6,7] A previous study by our group demonstrated that ROP experts correctly identified the vascular progression of ROP with an accuracy of approximately 76% with static side-by-side presentation of image pairs.[33] These findings imply that serial images might provide more contextualized information to decide the need for treatment compared with single images alone.[9]

The second key finding is that the graders who tended to have a lower threshold for diagnosis of preplus and plus were more likely to be influenced by serial progression images. Our previous study revealed that ROP experts occasionally recommended ROP treatment in eyes with disease milder than type 1 mainly for structural changes, prompting concern about future anatomic complications, active ROP disease at an advanced postmenstrual age, and logistical, systemic, and other diagnostic concerns.[32] Although graders were not provided with additional demographic information when viewing the images, experts are more likely to diagnosis plus disease if demographic characteristics suggest a higher pretest probability for more severe disease.[34] Further studies are required to better understand variability in graders' threshold to diagnose plus disease.

The third key finding is that deep learning can measure disease severity that corresponds with expert agreement and disease progression. The use of a deep learning—derived quantitative VSS may standardize the assessment of plus disease among experts and may facilitate a better understanding of the role disease progression ought to play in treatment decisions, which has not been evaluated in clinical trials. This represents an area of opportunity for the use of artificial intelligence as an adjunct to clinical decision-making when viewing serial images in ROP. Numerous computer-based image analysis programs have been developed for the quantification of vascular features related to plus disease using different algorithms for measuring vascular dilation and tortuosity.[19,20,35,36] ROPTool and Retinal Image multiScale Analysis have provided objective measurement of the changes in retinal vascular dilation and tortuosity in individual eyes over time.[35,36,37] The i-ROP consortium has demonstrated expert-level classification of plus disease. The i-ROP deep learning system has introduced a quantitative scale for ROP severity (1—9) that has been shown to correlate with the International Classification of ROP. The integration of a quantitative severity score may assist in providing a more consistent and precise identification of retinal vascular progression in conjunction with qualitative assessment by readers.

This study should be viewed in the light of some limitations: (1) The sample size for both image analysis and number of graders was relatively small, with 7 graders and 15 cases. Future research with additional standardized cases and participants may provide more clear and generalizable answers on the role of serial retinal images in ROP surveillance; (2) the RSD was determined by a single "latest" image, not by serial images. This makes it

difficult to directly compare the accuracy of ROP classification using single versus serial images; (3) readers were not informed of the infants' birth weight, gestational age at birth, postmenstrual age, and findings from previous examinations, which does not represent current clinical practice in which these factors are considered together with retinal findings when determining management;[21,34] (4) the image-based diagnosis of ROP in this study limits its generalizability primarily to a telemedicine setting and may not reflect grading practices via binocular indirect ophthalmoscopy. However, if an image-based diagnosis is routinely performed, examiners may be able to use serial images to assess progression; (5) the images in this study were collected in a standardized fashion at regular time points. However, in clinical practice, there may be variability in the frequency of imaging and follow-up examination based on screening criteria and other considerations; and (6) experts who participated in the study were also part of the i-ROP consortium. This could have potentially caused bias by the experts and therefore influenced the results of the study. The images, however, were presented in a deidentified and randomized fashion so that the graders would be less likely to recognize cases. Despite the limitations we present here, our study provides some evidence to support the use of incorporating the concept of dynamic disease progression into ROP assessments.

Currently, there has been no consensus on how to integrate information regarding retinal vascular progression into the diagnosis and management of ROP and how it can best be used in clinical practice. Future studies are needed to assess how a quantitative VSS can play an adjunctive role in the diagnosis and management of ROP, how this may change the decision to treat or not, and how clinicians interact with this quantitative diagnosis in the clinical setting. Recognizing vascular progression toward more or less severe disease with serial images affected classification of ROP severity as well as their follow-up and management decisions. These findings suggest the potential role of using serial retinal images in the current management of ROP, as well as the promising role of a quantitative VSS in standardizing ROP diagnosis.

Acknowledgments

Supported by grants R01EY19474, K12EY027720, P30 EY001792, and P30EY10572 from the National Institutes of Health (Bethesda, Maryland), grants SCH-1622679, SCH-1622542, and SCH-1622536 from the National Science Foundation (Arlington, Virginia), unrestricted departmental funding, and a Career Development Award (J.P.C.) from Research to Prevent Blindness (New York, New York).

Abbreviations and Acronyms:

i-ROP	Imaging and Informatics in Retinopathy of Prematurity
ROP	retinopathy of prematurity
RSD	reference standard diagnosis
VSS	vascular severity score

References

1. Gilbert C Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. Early Hum Dev. 2008;84:77–82. [PubMed: 18234457]
2. Multicenter trial of cryotherapy for retinopathy of prematurity. Preliminary results. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Arch Ophthalmol. 1988;106:471–479. [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;121:1684–1694. [PubMed: 14662586]
4. Christiansen SP, Dobson V, Quinn GE, et al. Progression of type 2 to type 1 retinopathy of prematurity in the Early Treatment for Retinopathy of Prematurity Study. Arch Ophthalmol. 2010;128:461–465. [PubMed: 20385942]
5. International Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. Pediatrics. 1984;74:127–133. [PubMed: 6547526]
6. Chiang MF, Jiang L, Gelman R, et al. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. Arch Ophthalmol. 2007;125:875–880. [PubMed: 17620564]
7. 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;12:352–356. [PubMed: 18329925]
8. Gschließer A, Stifter E, Neumayer T, et al. Inter-expert and intra-expert agreement on the diagnosis and treatment of retinopathy of prematurity. Am J Ophthalmol. 2015;160:553–560e3. [PubMed: 26004406]
9. Slidsborg C, Forman JL, Fielder AR, et al. Experts do not agree when to treat retinopathy of prematurity based on plus disease. Br J Ophthalmol. 2012;96:549–553. [PubMed: 22174097]
10. Hewing NJ, Kaufman DR, Chan RV, Chiang MF. Plus disease in retinopathy of prematurity: qualitative analysis of diagnostic process by experts. JAMA Ophthalmol. 2013;131:1026–1032. [PubMed: 23702696]
11. Campbell JP, Kalpathy-Cramer J, Erdogmus D, et al. Plus disease in retinopathy of prematurity: a continuous spectrum of vascular abnormality as a basis of diagnostic variability. Ophthalmology. 2016;123:2338–2344. [PubMed: 27591053]
12. Campbell JP, Ryan MC, Lore E, et al. Diagnostic discrepancies in retinopathy of prematurity classification. Ophthalmology. 2016;123:1795–1801. [PubMed: 27238376]
13. Schaffer DB, Palmer EA, Plotsky DF, et al. Prognostic factors in the natural course of retinopathy of prematurity. The Cryotherapy for Retinopathy of Prematurity Cooperative Group. Ophthalmology. 1993;100:230–237. [PubMed: 8437832]
14. Wallace DK, Kylstra JA, Chesnutt DA. Prognostic significance of vascular dilation and tortuosity insufficient for plus disease in retinopathy of prematurity. J AAPOS. 2000;4:224–229. [PubMed: 10951298]
15. Heneghan C, Flynn J, O'Keefe M, Cahill M. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. Med Image Anal. 2002;6:407–429. [PubMed: 12426111]
16. 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;147:1086–1091. [PubMed: 19286152]
17. Wallace DK, Freedman SF, Hartnett ME, Quinn GE. Predictive value of pre-plus disease in retinopathy of prematurity. Arch Ophthalmol. 2011;129:591–596. [PubMed: 21555612]
18. Ghodasra DH, Karp KA, Ying GS, et al. Risk stratification of preplus retinopathy of prematurity by semiautomated analysis of digital images. Arch Ophthalmol. 2010;128:719–723. [PubMed: 20547949]
19. Thyparampil PJ, Park Y, Martinez-Perez ME, et al. Plus disease in retinopathy of prematurity: quantitative analysis of vascular change. Am J Ophthalmol. 2010;150:468–475e2. [PubMed: 20643397]
20. 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]

21. Wu C, Petersen RA, VanderVeen DK. RetCam imaging for retinopathy of prematurity screening. J AAPOS. 2006;10:107–111. [PubMed: 16678743]
22. Hussein MA, Coats DK, Paysse EA. Use of the RetCam 120 for fundus evaluation in uncooperative children. Am J Ophthalmol. 2004;137:354–355. [PubMed: 14962432]
23. Chiang MF, Melia M, Buffenn AN, et al. Detection of clinically significant retinopathy of prematurity using wide-angle digital retinal photography: a report by the American Academy of Ophthalmology. Ophthalmology. 2012;119:1272–1280. [PubMed: 22541632]
24. Ryan MC, Ostmo S, Jonas K, et al. Developmen and evaluation of reference standards for image-based telemedicine diagnosis and clinical research studies in ophthalmology. AMIA Annu Symp Proc. 2014;2014:1902–1910. [PubMed: 25954463]
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174. [PubMed: 843571]
26. Kalpathy-Cramer J, Campbell JP, Erdogmus D, et al. Plus disease in retinopathy of prematurity: improving diagnosis by ranking disease severity and using quantitative image analysis. Ophthalmology. 2016;123:2345–2351. [PubMed: 27566853]
27. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. Br J Ophthalmol. 2019;103:580–584.
28. Coyner AS, Swan R, Brown JM, et al. Deep learning for image quality assessment of fundus images in retinopathy of prematurity. AMIA Annu Symp Proc. 2018;2018:1224–1232. [PubMed: 30815164]
29. Coyner AS, Swan R, Campbell JP, et al. Automated fundus image quality assessment in retinopathy of prematurity using deep convolutional neural networks. Ophthalmol Retina. 2019;3:444–450. [PubMed: 31044738]
30. Brown JM, Campbell JP, Beers A, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. 2018;136:803–810. [PubMed: 29801159]
31. Ji MH, Callaway NF, Greven MA, et al. Telemedicine follow-up for intravitreal bevacizumab injection in the Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDROP) cohort. Clin Ophthalmol. 2020;14:1161–1163. [PubMed: 32431486]
32. Gupta MP, Chan RVP, Anzures R, et al. Practice patterns in retinopathy of prematurity treatment for disease milder than recommended by guidelines. Am J Ophthalmol. 2016;163:1–10. [PubMed: 26705094]
33. Myung JS, Gelman R, Aaker GD, et al. Evaluation of vascular disease progression in retinopathy of prematurity using static and dynamic retinal images. Am J Ophthalmol. 2012;153:544–551e2. [PubMed: 22019222]
34. Gschließer A, Stifter E, Neumayer T, et al. Effect of patients' clinical information on the diagnosis of and decision to treat retinopathy of prematurity. Retina. 2018;38:2253–2259. Published correction appears in *Retina*. 2019;39:e20. [PubMed: 28984736]
35. Cabrera MT, Freedman SF, Kiely AE, et al. Combining ROPtool measurements of vascular tortuosity and width to quantify plus disease in retinopathy of prematurity. J AAPOS. 2011;15:40–44. [PubMed: 21397804]
36. Wittenberg LA, Jonsson NJ, Chan RV, Chiang MF. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity. J Pediatr Ophthalmol Strabismus. 2012;49:11–19. [PubMed: 21366159]
37. Taylor S, Brown JM, Gupta K, et al. Monitoring disease progression with a quantitative severity scale for retinopathy of prematurity using deep learning. JAMA Ophthalmol. 2019;137:1022–1028. [PubMed: 31268518]

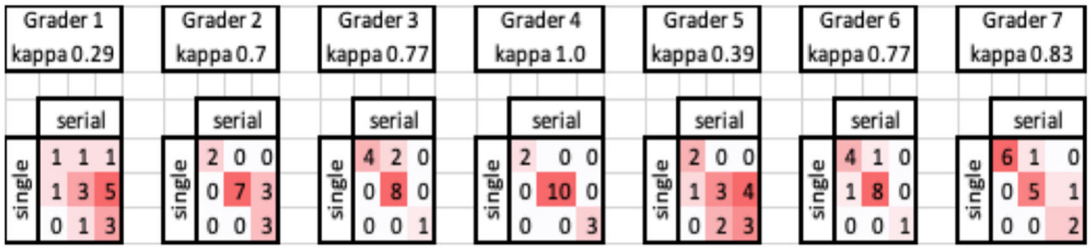

Figure 1.

The confusion matrices and weighted kappa values for the assessment of single images versus the same image presented at the end of a series of images (serial), for each grader.

mean severity		mode severity	
single	serial	single	serial
1.0	1.6	1	1
1.0	1.0	1	1
3.3	3.3	5	5
3.9	3.9	5	5
3.9	5.6	5	5
3.9	2.1	5	1
4.4	6.7	5	5
4.4	5.6	5	5
5.0	5.6	5	5
5.0	6.1	5	5
5.0	6.1	5	5
6.1	6.7	5	5
7.3	7.9	9	9
7.9	7.3	9	9
9.0	8.4	9	9

Figure 2.

The mean and mode severity of single and serial images. Using a numeric scale which corresponds to the color spectrum (1 = no plus = green, 5 = preplus, and 9 = plus = red), the above table demonstrates the mean (left) and mode (right) values for each image for the group of experts presented with the single image and serial image set. Qualitatively, the mean severity was higher when the images were presented in series.

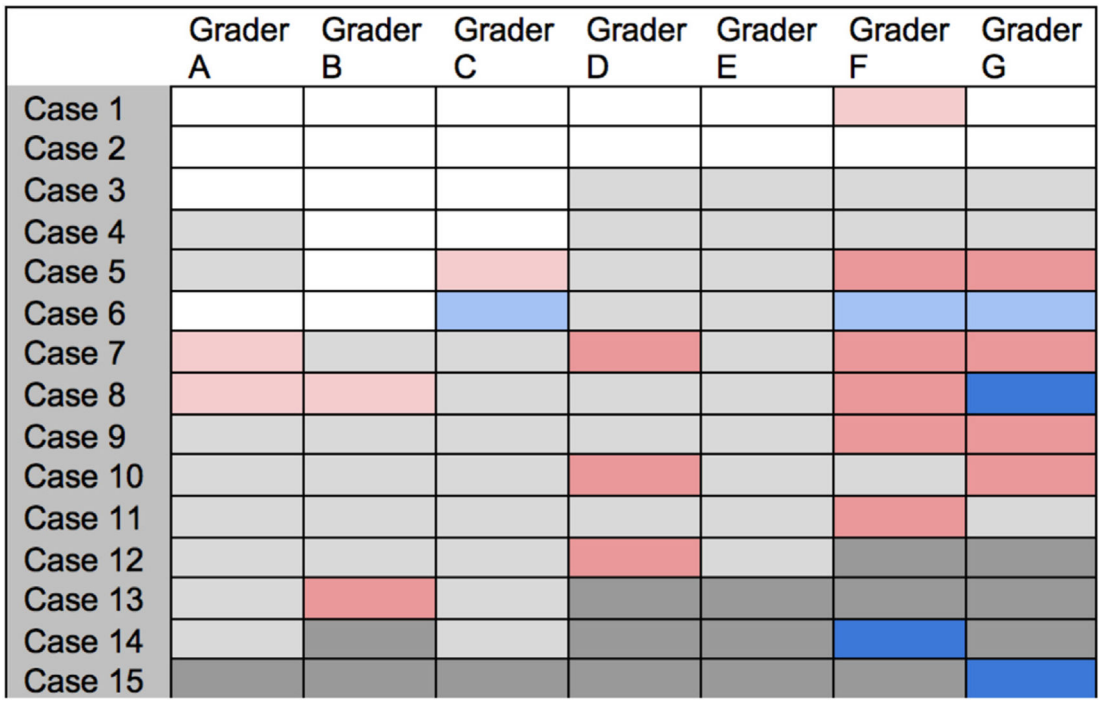

Figure 3.

Compared with the standard single image grading, this figure presents the changes in diagnosis across all graders after viewing serial images, organized from least to most aggressive graders (left to right). Graders are presented in the order of least (A) to most aggressive (G) grading. Grayscale boxes indicate images with no change in single versus serial images (white = no retinopathy of prematurity [ROP], light gray = preplus, dark gray = plus). The red boxes indicate a change in diagnosis toward more severe disease with serial image viewing (light red = no ROP to preplus, dark red = preplus to plus). The blue boxes indicate a change in diagnosis toward less severe disease (dark blue = plus to preplus, light blue = preplus to no disease).

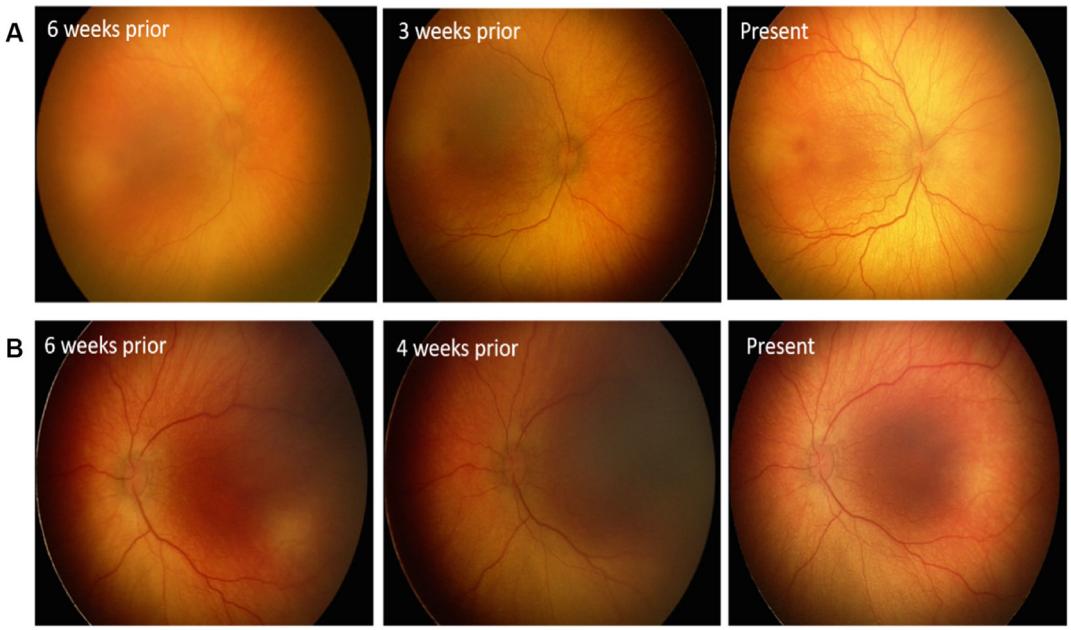

Figure 4.

Two sets of serial images of cases with a reference standard diagnosis of preplus disease. For single images, graders viewed only the last image in the series. In set (**A**), the graders showed diagnostic changes toward more severe disease with serial images. In set (**B**), the graders showed clinical diagnostic changes toward less severe disease.

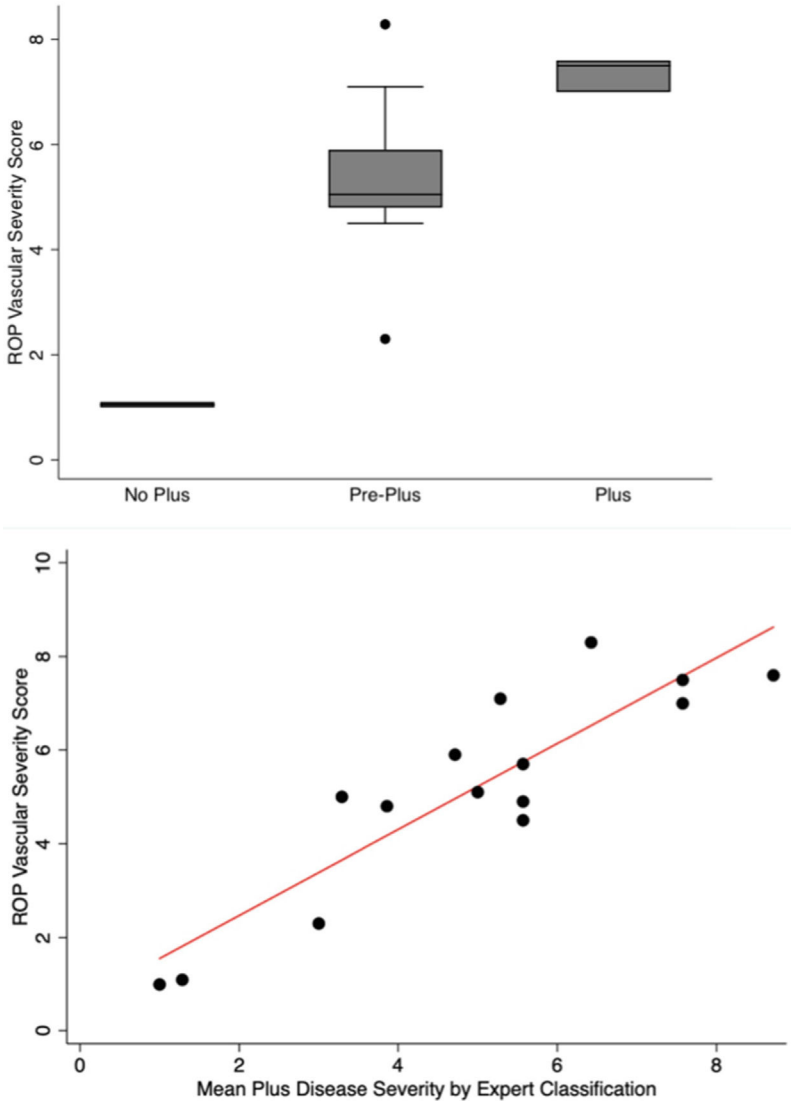

Figure 5.

A, Vascular severity scores demonstrated good overall agreement with the mode class as assigned to no plus, preplus, and plus disease ($P = 0.001$). **B,** The retinopathy of prematurity (ROP) vascular severity score correlated with the range of disease severity, as determined by averaging the clinician responses of disease severity (no plus disease = 1, preplus = 5, plus disease = 9) with a correlation coefficient of 0.8.