



Published in final edited form as:

J AAPOS. 2008 August ; 12(4): 352–356. doi:10.1016/j.jaapos.2007.11.022.

Agreement Among Pediatric Ophthalmologists in Diagnosing Plus and Pre-plus Disease in Retinopathy of Prematurity

David K. Wallace, MD, MPH¹, Graham E. Quinn, MD, MSCE², Sharon F. Freedman, MD¹, and Michael F. Chiang, MD, MA³

¹*Department of Ophthalmology, Duke University, Durham, North Carolina*

²*Department of Ophthalmology, Children's Hospital of Philadelphia, Pennsylvania*

³*Department of Ophthalmology, Columbia University, New York, New York*

Abstract

Purpose—Plus disease has become the major criterion for laser treatment in infants with retinopathy of prematurity (ROP), but its assessment is subjective. Our purpose was to compare quadrant-level and eye-level assessment of plus disease and pre-plus disease among 3 experienced ROP examiners and to report their rate of agreement.

Methods—One hundred eighty-one high-quality RetCam images from premature infants were graded by 3 of the authors. Dilation and tortuosity were judged separately using a scale of normal or sufficiently abnormal to meet criteria for pre-plus or plus disease.

Results—There was disagreement on the presence of plus disease for 18 images (10%), on tortuosity sufficient for plus disease (plus tortuosity) for 26 images (14%) and on dilation sufficient for plus disease (plus dilation) for 26 images (14%). Of 67 images judged to have pre-plus disease or worse, there was disagreement on the presence of plus disease for 18 images (27%), on plus tortuosity for 25 images (37%), and on plus dilation for 21 images (31%). For distinguishing plus or pre-plus disease from normal, there was disagreement on pre-plus tortuosity for 38 of 181 images (21%) and on pre-plus dilation for 58 of 181 images (32%).

Conclusions—Three experienced ROP examiners disagreed frequently on the diagnosis of plus or pre-plus disease when evaluating cropped clinical photographs of infants, many of which had borderline plus disease. Further study is required to determine the implications of these observations on clinical decision making.

Introduction

Severe retinopathy of prematurity (ROP) can lead to retinal detachment and permanent visual loss. As acute ROP worsens, characteristic changes occur in the blood vessels of the posterior retina. Plus disease is considered to be present when the vascular changes are so marked that the posterior veins are dilated and the arterioles tortuous. (1) A less severe vascular change, pre-plus disease, is defined as abnormalities of the posterior pole that are insufficient for the diagnosis of plus disease but that demonstrate more arterial tortuosity and more venous

Corresponding Author and Reprint Requests: David K. Wallace MD MPH, Duke University Eye Center, DUMC 3802 Durham, NC 27710, email: david.wallace@duke.edu.

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

dilatation than normal. (2) Since results were published from the Early Treatment for Retinopathy of Prematurity (ETROP) study (3), plus disease has become the primary indication for laser treatment. Unfortunately, the assessment of plus disease may be extremely subjective. It is based on comparison to a standard photograph that is more than 20 years old. This photograph is de-centered and blurred, and it shows only a small part of the posterior retina close to the optic nerve. (4,5)

A computer program (“ROptool”) has been developed that traces retinal blood vessels and measures their dilation and tortuosity. (6) Recently, a study was done to evaluate the accuracy of ROptool in assessing tortuosity sufficient for plus disease in comparison to consensus of 3 experts. The accuracy of ROptool was similar and its sensitivity was superior to individual examiners in assessing tortuosity sufficient for plus disease. (7) As part of this study, 190 posterior poles images from 119 different patients were graded by 3 pediatric ophthalmologists. The purpose of the study reported herein was to compare these grades and to report the agreement rate among 3 experienced pediatric ophthalmologists.

Methods

Details of the study that evaluated the accuracy of ROptool have been reported. (7) Three of us (GEQ, SFF, MFC) served as graders for this study. All three graders were fellowship-trained pediatric ophthalmologists with extensive ROP experience at different institutions, and all were certified investigators in the ETROP study.

One-hundred ninety photographs of the posterior retina of premature infants that were obtained using RetCam (Clarity Medical Systems, Inc., Pleasanton, California) were collected. These 190 images came from 119 different patients: both eyes of 70 patients and one eye of 49 patients (one patient had 3 images from 2 eyes – 2 images from the same eye were obtained on different days). None of the infants could be identified from the retinal images, so an Institutional Review Board exemption was granted for this study. The sample was enriched to include a larger proportion of images with plus disease and pre-plus disease than would normally be encountered during routine screening examinations. Included were as many images as possible with “borderline” plus disease as judged by DKW, who was not one of the 3 examiners and did not formally grade images while selecting them. Some images came from photographic databases of previously published studies done elsewhere (8,9) and some images were from teaching files. None of the photographs were taken for purposes of this study. Most images had very high quality and were in sharp focus, and all images included the optic nerve and posterior pole. Images were excluded from analysis if any of the graders could not evaluate all quadrants. It was necessary to include only images with grades for all quadrants because analyses were done per quadrant as well as per eye.

Adobe Photoshop (Adobe, San Jose, California) was used to crop each image in the shape of a circle centered on the optic nerve that approximated the view seen with a 28 diopter lens. This view is the typical one of the posterior retinal vessels that is initially seen on examination and used to assess the presence or absence of plus disease. (Figure) Cropping the RetCam images in this manner insured that the 3 graders used the same view and extent of the retina to judge dilation and tortuosity and that more peripheral findings such as stage 3 disease did not influence this judgment. The cropped images were randomly ordered and distributed to the graders.

All graders were reminded of the definition of plus disease (as detailed in the Introduction), and all were sent a copy of the standard photograph of plus disease as well as examples of plus disease and pre-plus disease from the ICROP Revisited publication. (2) Each grader then independently scored each quadrant of each image by grading tortuosity and dilation separately

(8 total grades per eye) as plus, pre-plus or normal. Graders were not instructed as to whether to consider arterioles, venules, or both. Their scores were used to generate both quadrant-level and eye-level data. Quadrant-level data were based on individual quadrant grades, and eye-level grades were based on a combination of the quadrant-level grades for each image. For example, an eye-level grade of plus disease was present if at least 2 of the 4 quadrants had tortuosity and dilation sufficient for plus disease, in keeping with requirements from recent ROP multi-center clinical trials. (3,10) An eye-level grade of tortuosity sufficient for plus disease was present if at least 2 of the 4 quadrants in a single eye had tortuosity sufficient for plus disease. An eye was considered to have pre-plus disease if at least 2 quadrants had pre-plus dilation and tortuosity, or if one quadrant had plus dilation and tortuosity and one or more quadrants had pre-plus dilation and tortuosity.

For each image and for each grader individually, it was determined whether an eye was judged to have sufficient dilation for plus disease, sufficient tortuosity for plus disease, and sufficient dilation and tortuosity for plus disease. For each of these categories, a distribution of grades was generated (3 Yes / 0 No, 2 Yes / 1 No, 1 Yes / 2 No, and 0 Yes / 3 No). "Percentage disagreement" was calculated by dividing those images for which there was disagreement (2 Yes / 1 No or 1 Yes / 2 No) by the total number of images. This same analysis was repeated at the quadrant level in order to include 4 times as many observations. Next, the same analysis was repeated after excluding those judged by consensus (at least 2 of 3 graders) to be normal images, in order to determine agreement on distinguishing plus disease from pre-plus disease. This analysis was done because there were a large number of "normal" (neither plus nor pre-plus) images in the study, and graders never disagreed on the presence of plus disease in an image judged by at least 2 of 3 to be normal. Two of 3 normal judgments were required to exclude images instead of 3 of 3 in order to exclude images that were likely normal; requiring 3 of 3 would have decreased the number of images excluded and increased the rate of agreement. After excluding "normals", it was possible for the distribution of grades to be 0 Yes (plus) / 3 No (not plus) because those images that were not plus were also not completely normal, as they were judged to be pre-plus by at least 2 of 3 graders. Finally, a similar analysis was done including all images but considering instead whether an eye or quadrant had tortuosity or dilation sufficient for pre-plus or plus disease (any abnormality) or whether it was normal. For this study, no attempt was made to compare the judgment of each grader to a reference standard.

Results

Description of Study Images

One hundred ninety images were graded. Of these, 9 images were excluded because a grader could not evaluate all quadrants (7 images by one grader; 2 images by a different grader). Of the remaining 181 images used for the eye-level analysis, consensus (at least 2 of 3 graders) was that 27 images (15%) had tortuosity sufficient for plus disease, 40 images (22%) had tortuosity sufficient for pre-plus disease but insufficient for plus disease, and 114 images (63%) had no abnormal tortuosity. Consensus of the graders was that 28 images (15%) had dilation sufficient for plus disease, 49 images (27%) had dilation sufficient for pre-plus disease but insufficient for plus disease, and 104 images (57%) had no abnormal dilation.

At the quadrant level, there were 181 images with 4 quadrants each, or 724 total quadrants. Grader consensus was that 90 quadrants (12%) had tortuosity sufficient for plus disease, 144 quadrants (20%) had tortuosity sufficient for pre-plus disease but insufficient for plus disease, and 490 quadrants (68%) had no abnormal tortuosity. For dilation at the quadrant level, consensus was that 88 quadrants (12%) had dilation sufficient for plus disease, 160 quadrants (22%) had dilation sufficient for pre-plus disease but insufficient for plus disease, and 476 quadrants (66%) had no abnormal dilation.

Interobserver Variability of Judgment

Among the 3 graders, there existed differences in judgment of plus disease vs. not plus disease at both the eye level and the quadrant level. When all 181 gradable images were included, there was disagreement among the 3 examiners on plus disease for 18 images (10%), on tortuosity sufficient for plus disease for 26 images (14%), and on dilation sufficient for plus disease for 26 images (14%). (Table) The disagreement rates between examiners 1 and 2, examiners 1 and 3, and examiners 2 and 3 were 9%, 12%, and 7%, respectively. When excluding normal images, there was even greater disagreement in distinguishing plus disease from pre-plus disease. Of 67 abnormal images, there was disagreement among the 3 examiners on plus disease for 18 images (27%), on tortuosity sufficient for plus disease for 25 images (37%), and on dilation sufficient for plus disease for 21 images (31%).

When considering pre-plus disease, there was disagreement among the 3 examiners on tortuosity sufficient for pre-plus disease for 38 of 181 images (21%) and on dilation sufficient for pre-plus disease for 58 of 181 images (32%). Results were similar at the eye level or at the quadrant level. (Table)

Discussion

Accurately diagnosing plus disease is essential to optimize the timing of laser treatment. The ETROP study showed a significant benefit to treatment of eyes with high-risk prethreshold disease. (3) Consequently, the recently revised guidelines endorsed by the American Academy of Ophthalmology, the American Association for Pediatric Ophthalmology and Strabismus, and the American Academy of Pediatrics recommended that laser treatment be performed for eyes with Type 1 ROP, defined as (1) zone I ROP with plus disease, (2) zone II, stage 2 or 3 with plus disease, or (3) zone I, stage 3 without plus disease. (11) Since most eyes with type 1 ROP have plus disease, it is clear that the primary indication for laser treatment is now the presence of plus disease. It is recommended that treatment be performed within 72 hours of determination of treatable disease, because delay in treatment may increase the risk of an unfavorable outcome. (3)

Our results indicate that there is a large amount of interobserver variability in the assessment of plus disease and pre-plus disease. When excluding normal images, there was disagreement among the 3 graders on the presence of plus disease for 18 of 67 images (27%). This observation agrees with the findings of Freedman and associates that 3 CRYO-ROP certified experts disagreed on the presence or absence of plus disease in 29 of 72 cases (40%). (12) Chiang and associates reported that all 22 ROP experts agreed on the presence or absence of plus disease for only 7 of 34 images (21%). Using a 3-level categorization (plus, pre-plus, or neither), they found agreement among all 22 experts in only 4 of 34 images (12%). (13) Our disagreement rate in this study was lower because we used 3 graders (compared to 22 in Chiang's study). When reporting the number of times that at least one grader disagreed, it is expected that a higher rate of disagreement would exist when more graders are used.

Most experienced examiners have presumably been "calibrated" to the degree of abnormality represented in standard photograph of plus disease. So why does so much disagreement exist in the assessment of plus disease? Heretofore, there has been little attention paid to identifying the factors that influence an individual's assessment of plus disease, and a number of unanswered questions remain. For example, is the assessment of tortuosity driven by the most tortuous vessel in each quadrant or by some composite of tortuosity of all of the major vessels in a quadrant? How much weight is given to extremely tortuous short vessel segments ("wiggles") versus more gradual curves in longer vessel segments? (14) How does the amount and degree of dilation influence the assessment of tortuosity? Does the total number or volume of vessels influence this judgment? One might think that status as an experienced ROP

examiner would improve one's ability to accurately assess tortuosity or to discriminate between various degrees of dilation and tortuosity. However, Freedman and associates found that naïve and expert subjects performed indistinguishably when ranking groups of computer-generated retinal vascular tracings in order of increasing mean diameter and tortuosity. (15) The large amount of interobserver variability in the assessment of plus disease suggests that even ROP experts are frequently inaccurate in evaluating plus disease.

These data suggest that there is the need for more objective methods to diagnose plus disease and pre-plus disease, particularly in "borderline" cases. One potentially promising technique for reducing subjectivity of plus disease assessment is computer-assisted analysis of posterior pole images. One such computer program ("ROPTool") demonstrated an overall accuracy similar to 3 individual examiners and better sensitivity than 2 of 3 examiners in diagnosing tortuosity sufficient for plus disease (7) Other investigators have also reported promising results using other computer-assisted methods for quantification of plus disease. (16–19) When confronted with the clinical dilemma of plus versus pre-plus disease, another reasonable alternative is to repeat the examination 2 days later, which still allows for treatment within the recommended window.

This study must be viewed in light of some limitations. First, the sample was enriched to include as many borderline images as possible. Since examiners are more likely to disagree in borderline cases, this method of image selection could increase the rate of disagreement and make it higher than one might typically encounter in clinical practice. A second limitation is that the graders compared cropped RetCam images to the standard photograph of plus disease, but it is unknown whether disagreements regarding RetCam images would translate to disagreements at the bedside where more of the fundus can be viewed using indirect ophthalmoscopy. The cropping of images may have resulted in more disagreement compared to what might have been observed if images were not cropped. However, we felt it was important to crop the images in order to simulate the view of the posterior pole that is typically used when evaluating plus disease." It is possible that contact with the camera altered the appearance of the retinal blood vessels, so the relative proportion of fault between inherent reader ability and any problems with photographs cannot be established. In order to establish the rate of disagreement at bedside, one would need to perform a study in which different clinicians perform indirect ophthalmoscopy of the same infants on the same day. Third, we used individual quadrant grades of dilation and tortuosity and converted them to eye-level grades, based on the requirement that at least two quadrants have dilation and tortuosity sufficient for plus disease. Clinically, examiners may not separate the quadrants in this manner but instead may rely on their overall impression when determining the presence of plus disease. Finally, we used only 3 graders and only images of high enough quality that they could be evaluated by all graders.

In conclusion, we found that experienced ROP examiners frequently disagreed on the presence or absence of plus disease or pre-plus disease when judging high-quality cropped RetCam photographs, many of which were images with borderline plus disease. Further study required to determine the implications of these observations on clinical decision making; however, these data suggest that there is need for more objective and accurate methods for assessing retinal vascular abnormalities in preterm infants.

Acknowledgments

We thank Anna Ells MD of Alberta Children's Hospital, Calgary, Alberta, Canada, Antonio Capone MD of William Beaumont Hospital, Royal Oak, Michigan, and the PHOTO-ROP study group for sharing RetCam photographs.

References

1. The Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Arch Ophthalmol* 1984;102:1130–1134. [PubMed: 6547831]
2. An International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity Revisited. *Arch Ophthalmol* 2005;123:991–999. [PubMed: 16009843]
3. Early Treatment for Retinopathy of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: Results of the early treatment of retinopathy of prematurity randomized trial. *Arch Ophthalmol* 2003;121:1684–1694. [PubMed: 14662586]
4. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity: Preliminary results. *Arch Ophthalmol* 1988;106:471–479. [PubMed: 2895630]
5. Capone A Jr, Ells AL, Fielder AR, Flynn JT, Gole GA, Good WV, et al. Standard image of plus disease in retinopathy of prematurity. *Arch Ophthalmol* 2006;124:1669–1670. [PubMed: 17102031]
6. Wallace DK, Zhao Z, Freedman SF. A pilot study using "ROPTool" to quantify plus disease retinopathy of prematurity. *J AAPOS* 2007;11:381–387. [PubMed: 17532238]
7. Wallace DK, Freedman SF, Zhao Z, Jung SH. Accuracy of "CROPTool" versus Individual Examiners in Assessing Retinal Vascular Tortuosity. *Archives of Ophthalmology* 2007;125:1523–1530. [PubMed: 17998514]
8. Ells AL, Holmes JM, Astle WF, Williams G, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology* 2003;110:2113–2117. [PubMed: 14597517]
9. Photographic Screening for Retinopathy Of Prematurity (PHOTO-ROP) Cooperative Group. The Photographic Screening For Retinopathy Of Prematurity Study (PHOTO-ROP): Study Design and Baseline Characteristics of Enrolled Patients. *Retina* 2006;26(7 Suppl):S4–S10. [PubMed: 16946677]
10. The STOP-ROP Multicenter Study Group. Supplemental Therapeutic Oxygen for Prethreshold Retinopathy of Prematurity (STOP-ROP), A Randomized, Controlled Trial. I: Primary Outcomes. *Pediatrics* 2000;105:295–310. [PubMed: 10654946]
11. 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;117:572–576. [PubMed: 16452383]Erratum published in: *Pediatrics* 2006;118:1324.
12. Freedman SF, Kylstra JA, Hall JG, Capowski JJ. Plus disease in retinopathy of prematurity – photographic evaluation by an expert panel (abstract). *Invest Ophthalmol Vis Sci* 1995;36:76.
13. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *Arch Ophthalmol* 2007;125:875–880. [PubMed: 17620564]
14. Capowski JJ, Kylstra JA, Freedman SF. A numeric index based on spatial frequency for the tortuosity of retinal vessels and its application to plus disease in retinopathy of prematurity. *Retina* 1995;15:490–500. [PubMed: 8747443]
15. Freedman SF, Kylstra JA, Capowski JJ, Realini TD, et al. Observer sensitivity to retinal vessel diameter and tortuosity in retinopathy of prematurity: a model system. *J Pediatr Ophthalmol Strabismus* 1996;33:248–254. [PubMed: 8827562]
16. 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;46:4734–4738. [PubMed: 16303973]
17. 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. *Medical Image Analysis* 2002;6:407–429. [PubMed: 12426111]
18. Swanson CR, Cocker KD, Parker KH, Moseley MJ, Wren SME, Fielder AR. Semi-automated computer analysis of vessel growth in preterm infants without and with ROP. *Br J Ophthalmol* 2003;87:1474–1477. [PubMed: 14660456]
19. 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;143:723–725. [PubMed: 17386296]

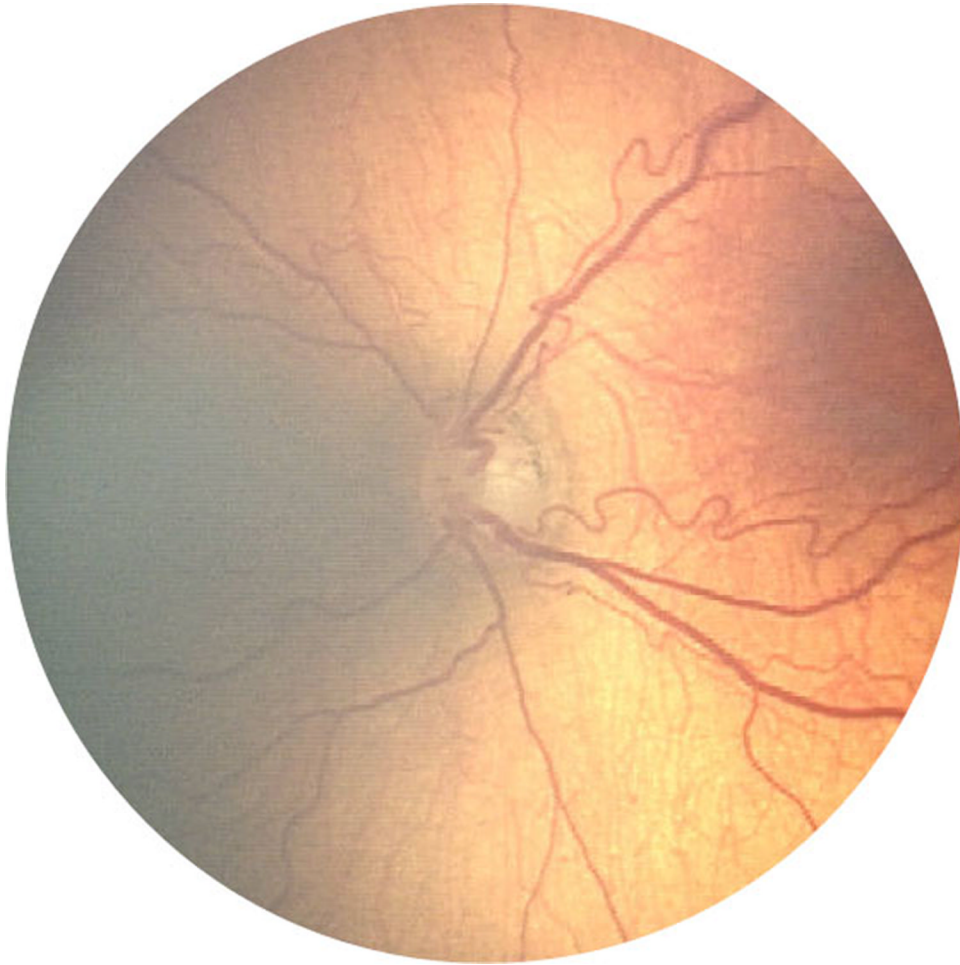


Figure 1. Example of an image for which there was substantial disagreement. One grader scored both tortuosity and dilation as sufficient for plus disease, one graded both tortuosity and dilation as sufficient for pre-plus disease, and one graded tortuosity as sufficient for pre-plus disease and dilation as normal. Therefore, overall eye grades were plus disease for one grader, pre-plus disease for one grader, and neither plus nor pre-plus for one grader. (Image courtesy of the PHOTO-ROP Study Group)

Table 1
 Distribution of grades by 3 ROP experts for tortuosity, dilation, or tortuosity and dilation sufficient for (1) plus disease vs. not plus disease or for (2) pre-plus disease vs. normal at both the eye level and the quadrant level.

	Distribution of Grades (Y es/No)			Percentage Disagreement
	3/0 [*]	2/1 ^{**}	1/2 ^{**}	
Sufficient for plus disease				
All images				
Eye-level				
Tortuosity	14	13	13	14%
Dilation	14	14	12	14%
Tortuosity and dilation	10	6	12	10%
Quadrant-level				
Tortuosity	51	39	59	14%
Dilation	44	44	53	13%
Excluding normal images				
Eye-level				
Tortuosity	14	13	12	37%
Dilation	14	13	8	31%
Tortuosity and dilation	10	6	12	27%
Quadrant-level				
Tortuosity	51	39	55	35%
Dilation	44	41	42	31%
Tortuosity and dilation	32	26	50	28%
Sufficient for pre-plus disease				
All images				
Eye-level				
Tortuosity	52	15	23	21%
Dilation	44	33	25	32%
Quadrant-level				
Tortuosity	177	57	91	20%
Dilation	139	109	118	31%

* Number of eyes (or quadrants) for which 3 examiners judged that tortuosity (or dilation) was sufficient for plus disease (or pre-plus disease) and for which 0 examiners judged that tortuosity (or dilation) was sufficient for plus disease (or pre-plus disease).

** Number of eyes (or quadrants) for which there was disagreement among the 3 examiners