

# Accuracy of ROPtool vs Individual Examiners in Assessing Retinal Vascular Tortuosity

David K. Wallace, MD, MPH; Sharon F. Freedman, MD; Zheen Zhao, PhD; Sin-Ho Jung, PhD

**Objective:** To prospectively determine if tortuosity assessment by a computer program (ROPtool) that traces retinal blood vessels and measures their tortuosity was more accurate than that of individual pediatric ophthalmologists.

**Methods:** One hundred eighty-five high-quality RetCam images from premature infants were circulated to 3 retinopathy of prematurity (ROP) experts and 3 other pediatric ophthalmologists ("examiners") who graded the tortuosity in each quadrant as normal, pre-plus, or plus. These same images were analyzed using ROPtool.

**Results:** Using expert consensus as the standard, ROPtool's overall accuracy of 95% (175 of 185) for identify-

ing tortuosity sufficient for plus disease was similar to that of examiner 1 (93%; 172 of 185;  $P=.50$ ), examiner 2 (93%; 172 of 185;  $P=.50$ ), and examiner 3 (91%; 168 of 185;  $P=.10$ ). ROPtool's sensitivity of 97% (36 of 37) compared favorably with that of examiner 1 (65%; 24 of 37;  $P<.001$ ), examiner 2 (70%; 26 of 37;  $P<.001$ ), and examiner 3 (81%; 30 of 37;  $P=.06$ ).

**Conclusion:** Computer-assisted analysis of retinal images can potentially reduce subjectivity in the diagnosis of plus disease and optimize timing of follow-up and treatment for ROP.

*Arch Ophthalmol.* 2007;125(11):1523-1530

**P**LUS DISEASE IS DEFINED IN THE International Classification of Retinopathy of Prematurity (ICROP) as being present when "the vascular changes are so marked that the posterior veins are enlarged and the arterioles tortuous."<sup>1(p1132)</sup> The Cryotherapy for Retinopathy of Prematurity (CRYO-ROP) study introduced the concept of a standard photograph representing the minimum acceptable abnormality required for plus disease.<sup>2</sup> In the Supplemental Therapeutic Oxygen for Prethreshold Retinopathy of Prematurity (STOP-ROP) study, it was required that at least 2 quadrants of abnormal dilation and tortuosity of the central posterior retinal blood vessels be present to diagnose plus disease.<sup>3</sup> The international committee that revisited ICROP defined pre-plus disease as "vascular abnormalities of the posterior pole that are insufficient for the diagnosis of plus disease but that demonstrate more arterial tortuosity and more venous dilatation than normal."<sup>4(p995)</sup>

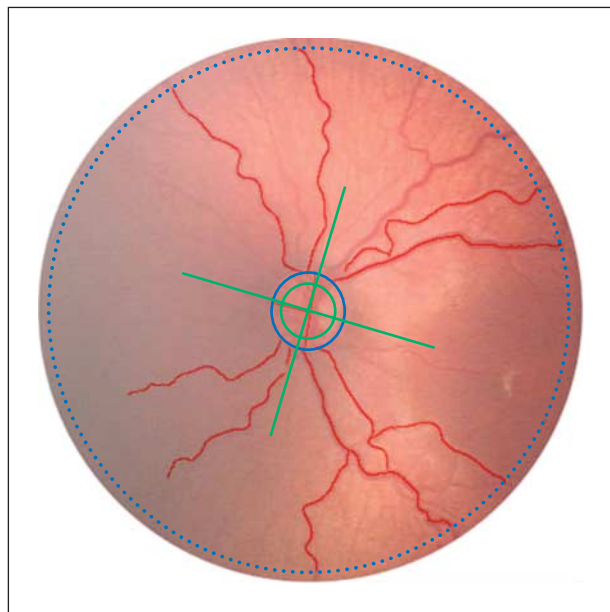
The presence of plus disease is a marker for severe disease, and it is one of the most

important prognostic indicators in retinopathy of prematurity (ROP).<sup>2,4,5</sup> In the CRYO-ROP study, 18 of 29 eyes (62%) with stage 3 and plus disease at 33 to 34 weeks' postmenstrual age had an unfavorable anatomical outcome, compared with only 1 of 31 eyes (3%) with stage 3 but without plus disease at the same age.<sup>2</sup> The

*See also pages  
1531 and 1562*

presence or absence of plus disease is now the primary factor in determining whether an infant with ROP requires laser treatment, since plus disease is the major criterion for type 1 ROP.<sup>6</sup> The international committee that revisited ICROP reaffirmed that "the minimum standard for plus disease is more important than ever in determining which infants should be treated for retinopathy of prematurity."<sup>7(pp1669-1670)</sup> However, it can be very difficult to judge the degree of vascular change present in an infant's posterior pole. Freedman and associates<sup>8</sup> found that 3 principal investigators in the CRYO-ROP study disagreed on the presence or absence of plus

**Author Affiliations:**  
Departments of Ophthalmology (Drs Wallace and Freedman), Pediatrics (Drs Wallace and Freedman), Psychiatry (Dr Zhao), and Biostatistics and Bioinformatics (Dr Jung), Duke University School of Medicine, Durham, North Carolina.



**Figure 1.** Appearance of image after analysis by ROPtool. Solid blue circle is optic nerve border; dotted blue circle is 28-diopter circle; green lines are borders of quadrants; and red lines are traced vessels whose tortuosity has been calculated.

disease in 29 of 72 retinal images (40%). Therefore, even for experts, the assessment of the amount of retinal vascular abnormality is often an educated guess. This underscores the potential value of a mechanism to more consistently and accurately quantify retinal vascular dilation and tortuosity in infants with ROP.

In response to this problem, there has been much interest in recent years in measuring as accurately as possible the degree of blood vessel dilation and tortuosity. We have developed a computer program that automatically traces selected retinal blood vessels, measures the tortuosity of each vessel in comparison with the standard photograph of plus disease, and calculates whether an eye has sufficient tortuosity for plus disease. This technology was first applied to the measurement of cerebral vessel dilation and tortuosity in angiograms.<sup>9,10</sup> We recently piloted the newest version of this program (now called ROPtool), and we found that its tortuosity measure was equal to or better than determination of tortuosity sufficient for plus disease by investigator judgment.<sup>11</sup> The measurement of dilation was disappointing in that ROPtool did not consistently discriminate between dilation that was or was not sufficient for plus disease as judged by 2 experts. The purpose of the present study was to assess the validity of ROPtool's measurement of vascular tortuosity by analyzing a much larger image set than in our pilot study. We hypothesized that ROPtool's determination of the presence or absence of tortuosity sufficient for plus disease would be superior to judgment by individual pediatric ophthalmologists.

## METHODS

Six pediatric ophthalmology colleagues of one of us (D.K.W.) were asked to participate in the study. Three of them were certified investigators in the Early Treatment for Retinopathy of Pre-

maturity Study and considered by us to be ROP experts based on their extensive ROP experience. The other 3 pediatric ophthalmologists (the "examiners") had performed approximately 1000, 7000, and 10 000 ROP examinations, respectively. The study was designed to challenge the accuracy of the examiners individually vs ROPtool, and judgment of the experts was considered to be the reference standard for comparison.

RetCam photographs of the posterior retina of 190 different eyes of 117 premature infants were collected. This number was based on a priori sample size calculations. Our institutional review board granted an exemption because none of the infants could be identified from the retinal images. The sample was enhanced to include a larger proportion of images with plus disease and pre-plus disease than would normally be encountered during routine screening examinations, including as many images as possible with "borderline" plus disease. Some images came from photographic databases of previously published studies done elsewhere,<sup>12,13</sup> and some images were from teaching files. Most images were very high quality and in sharp focus, with only a few images slightly blurred or decentered.

Adobe Photoshop (Adobe, San Jose, California) was used to crop each image in the shape of a circle centered on the optic nerve, which approximated the view seen with a 28-diopter lens (**Figure 1**). This is the typical view of the posterior retinal vessels that is initially seen on examination and used to assess the presence or absence of plus disease, although it is a wider field of view than that seen in the standard photograph of plus disease.<sup>2</sup> Cropping the RetCam images in this manner ensured that all ophthalmologists and ROPtool used the same view and extent of the retina to judge tortuosity sufficient for plus disease 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 6 ophthalmologists for grading. The same images were analyzed by ROPtool (**Figure 1**) using a method previously reported.<sup>11</sup> The one of us (D.K.W.) who operated ROPtool was masked to the grades of the other ophthalmologists. He attempted to identify and click on the major arteriole and venule in each of the 4 quadrants, and tortuosity of both arterioles and venules were usually included.

All 6 pediatric ophthalmologists were informed about the general purpose of the study, reminded of the definition of plus disease, and 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.<sup>4</sup> Each ophthalmologist 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. These scores were used to generate both quadrant-level and eye-level data. Eye-level grades were based on a combination of the quadrant-level grades for each image. For example, 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.

The combined grades of the 3 experts were used to establish the reference standard judgment of tortuosity sufficient for plus disease (plus tortuosity) for each image. When there was disagreement as to whether there was eye-level tortuosity sufficient for plus disease, the expert who was the outlier was asked to reconsider the image. Those images for which disagreement still existed were discussed on a conference call among the experts. Expert grades for eye-level determination of tortuosity sufficient for pre-plus disease (pre-plus tortuosity) or for quadrant-level determination of tortuosity sufficient for plus or pre-plus disease were not adjudicated, since they were not primary outcome measures.

The grades assigned to the images by the examiners were used to test the primary hypothesis that the computer program (ROPtool) was more accurate than an individual exam-

iner in determining the presence or absence of tortuosity sufficient for plus disease. No examiner grades were adjudicated, since each examiner's grades were compared separately to expert consensus.

## DATA ANALYSIS

"Tortuosity index" measured by ROPtool was defined as the vessel tortuosity divided by the average tortuosity of the standard photograph of plus disease<sup>2</sup> multiplied by 10. Therefore, a vessel with a tortuosity index of 15 had 50% more tortuosity than the average tortuosity of the major blood vessels in the standard photograph. Based on results from our pilot study,<sup>11</sup> ROPtool's threshold for tortuosity sufficient for pre-plus disease was determined a priori to be a tortuosity index of 7 or greater. The plus disease threshold was set a priori to a tortuosity index of 9 or greater, because it was more accurate in our pilot study than the default value of 10 that corresponds to the average tortuosity of the vessels in the standard photograph of plus disease.

## PRIMARY ANALYSIS

Sensitivity, specificity, and concordance (overall accuracy) were calculated in comparison with the reference standard for (1) ROPtool's assessment of tortuosity sufficient for plus disease and (2) judgment of each of the examiners individually. Sensitivity was calculated by dividing the number of true positives by the number of positives by the reference standard. Specificity was calculated by dividing the number of true negatives by the number of negatives by the reference standard. Concordance was calculated by dividing the total number of agreements (true positives plus true negatives) by the total number of images. To test the primary hypothesis, the concordance of each examiner was compared with the concordance of ROPtool. This hypothesis was tested 3 times, once for each examiner in comparison with ROPtool. Sensitivities and specificities of ROPtool were also compared with those of the examiners individually.

## SECONDARY ANALYSES

### Concordance, Sensitivity, and Specificity of ROPtool for Detecting Pre-Plus Disease

The methods used for the primary analysis were repeated for pre-plus tortuosity instead of plus tortuosity.

### Comparison of Mean Tortuosities

Using quadrant-level data, the average of the 3 experts' grades of tortuosity were calculated. Since each quadrant was graded as plus (2), pre-plus (1), or normal (0), this calculation resulted in 7 possible mean values: 0.0, 0.3, 0.7, 1.0, 1.3, 1.7, and 2.0. For each of these mean expert values, the average ROPtool tortuosity measurement was calculated with corresponding 95% confidence intervals. These 7 values were also collapsed into 3 categories by rounding to the nearest integer (0, 1, and 2), and the average tortuosities were calculated for each of these 3 categories.

### Receiver Operating Characteristic Curves

Receiver operating characteristic (ROC) curves were constructed to assess the overall value of ROPtool as a diagnostic tool and to determine the optimal thresholds, or cutoff points, for maximizing sensitivity and specificity. The ROC curves plot

sensitivity on the y-axis and 1-specificity (the false-positive rate) on the x-axis. The various thresholds chosen to calculate sensitivity and specificity were all integers between 5 and 15, representing numbers higher and lower than the threshold of 9 used for plus tortuosity in this study. For each threshold value, we determined the sensitivity and specificity in the assessment of tortuosity sufficient for plus disease.

## STATISTICAL ANALYSES

A sample size calculation for the primary outcome was done a priori. It was determined that 190 eyes from 95 infants were necessary to have 80% power to detect a significant difference ( $\alpha = .05$ ) in examiner-expert concordance of 85% vs ROPtool-expert concordance of 92.5%. We estimated that examiners would have 85% concordance (15% discordance), and a 50% relative reduction in discordance to 7.5% (92.5% concordance) was felt to be clinically significant.

Data were collected and analyzed in part using Microsoft Office Excel 2003 (Microsoft, Redmond, Washington), and some statistical analyses were done using SAS version 9.1 (SAS Institute Inc, Cary, North Carolina). Concordance, sensitivity, and specificity were estimated by assigning an equal weight to each eye. To account for possible dependency between eyes of each patient, a generalized estimating equation method with a working independent correlation model was used.<sup>14</sup> Mean tortuosities were compared using a *t* test. All *P* values were 2-sided. Although the study was not powered for a multiple testing adjustment, adjusted *P* values were calculated using the Bonferroni correction. The multiple testing adjustment was performed across 3 examiners for each variable (adjusted  $P = 3 \times P$ ). Adjusted *P* values are reported only for the tests whose marginal *P* values were smaller than .05.

## RESULTS

### DESCRIPTION OF IMAGES

Of the 190 images, 5 (3%) were excluded from the primary eye-level analyses, 3 because of insufficient image quality and 2 because they were from the same eye of the same patient (on different days). Of the remaining 185 images, expert consensus was that 37 images (20%) had tortuosity sufficient for plus disease, 33 images (18%) had tortuosity sufficient for pre-plus disease but insufficient for plus disease, and 115 images (62%) had no abnormal tortuosity. There was unanimous agreement initially on the presence or absence of plus disease for 158 of the 185 images (85%), increasing to 168 (91%) after images were reconsidered by the expert outlier, and 17 (9%) of the images required adjudication. These 185 images came from 117 different patients, both eyes of 68 patients and one eye of 49 patients.

### PRIMARY ANALYSIS

In comparison with expert consensus of eye-level tortuosity sufficient for plus disease, ROPtool had 175 agreements (95%) and 10 disagreements (5%) (**Table 1**). Concordance with expert consensus was similar for ROPtool and for the 3 examiners. However, when ROPtool had a disagreement with expert consensus, it overcalled tortuosity sufficient for plus disease in 9 of these 10 disagreements. Therefore, true tortuosity sufficient for plus

**Table 1. Eye-Level Diagnoses of Plus Disease by ROPtool and 3 Individual Examiners Compared With ROP Expert Consensus<sup>a</sup>**

		Expert Consensus, No. of Diagnoses		Concordance	Sensitivity	Specificity
		Plus	Not Plus			
ROPtool	Plus	36	9	175 of 185 (95%)	36 of 37 (97%)	139 of 148 (94%)
	Not plus	1	139			
Examiner 1	Plus	24	0	172 of 185 (93%); <i>P</i> =.50	24 of 37 (65%); <i>P</i> <.001; <i>P</i> adjusted <.001	148 of 148 (100%); <i>P</i> =.01; <i>P</i> adjusted=.01
	Not plus	13	148			
Examiner 2	Plus	26	2	172 of 185 (93%); <i>P</i> =.50	26 of 37 (70%); <i>P</i> <.001; <i>P</i> adjusted=.002	146 of 148 (99%); <i>P</i> =.03; <i>P</i> adjusted=.10
	Not plus	11	146			
Examiner 3	Plus	30	10	168 of 185 (91%); <i>P</i> =.10	30 of 37 (81%); <i>P</i> =.02; <i>P</i> adjusted=.06	138 of 148 (93%); <i>P</i> =.80
	Not plus	7	138			

<sup>a</sup> *P* values are for the difference between each examiner and ROPtool. The adjusted *P* value was adjusted for multiple comparisons.

disease was missed only once, resulting in a relatively high sensitivity of 97%. This sensitivity compared favorably with the sensitivities of examiner 1 (*P*<.001; *P* adjusted <.001), examiner 2 (*P*<.001; *P* adjusted=.002), and examiner 3 (*P*=.02; *P* adjusted=.06). Specificity was clinically similar among ROPtool and the 3 examiners, although specificity for 1 of the examiners was significantly better than specificity for ROPtool (Table 1).

#### OTHER ANALYSES OF EYE-LEVEL DATA

The numeric threshold for tortuosity sufficient for plus disease of 9 tortuosity units was determined a priori based on pilot data, and it was used to test the primary hypothesis. When we tested different numeric thresholds (or cut-off points), we observed changes in the sensitivity and specificity of ROPtool. Fewer total disagreements with expert consensus occurred when the numeric thresholds were 10 (8 disagreements) or 11 (8 disagreements) instead of 9 tortuosity units (10 disagreements). However, these higher numeric thresholds improved specificity (fewer false positives) at the expense of lower sensitivities (more false negatives). This trade-off between sensitivity and specificity is shown by the ROC curve in **Figure 2A**.

The numeric threshold for tortuosity sufficient for pre-plus disease of 7 tortuosity units was also determined a priori based on pilot data. In comparison with expert consensus of eye-level tortuosity sufficient for plus or pre-plus disease, ROPtool had 87% concordance (161 of 185); concordance for the 3 examiners was 93% (171 of 185), 91% (169 of 185), and 90% (167 of 185), respectively. Figure 2B shows that 7 tortuosity units is a reasonably good threshold for tortuosity sufficient for pre-plus disease, but it also shows that a threshold of 7.5 would have resulted in higher specificity with only a small reduction in sensitivity. In addition, sensitivity increases slightly at values lower than 7 tortuosity units, but specificity drops off sharply (ie, there are many more false positives).

#### ANALYSES OF QUADRANT-LEVEL DATA

**Table 2** shows that the mean ROPtool quadrant tortuosity measurement increased as the mean expert grade increased. All of the differences between categories were statistically significant at the .05 level except for mean

expert grades of 0.7 vs 1.0 and 1.0 vs 1.3. ROPtool had excellent discrimination between the major categories of normal (0.0-0.3), pre-plus (0.7-1.3), and plus disease (1.7-2.0), with mean tortuosities of 4.80 tortuosity units (95% confidence interval [CI], 4.65-4.95), 9.88 tortuosity units (95% CI, 9.17-10.59), and 19.06 tortuosity units (95% CI, 17.36-20.76), respectively.

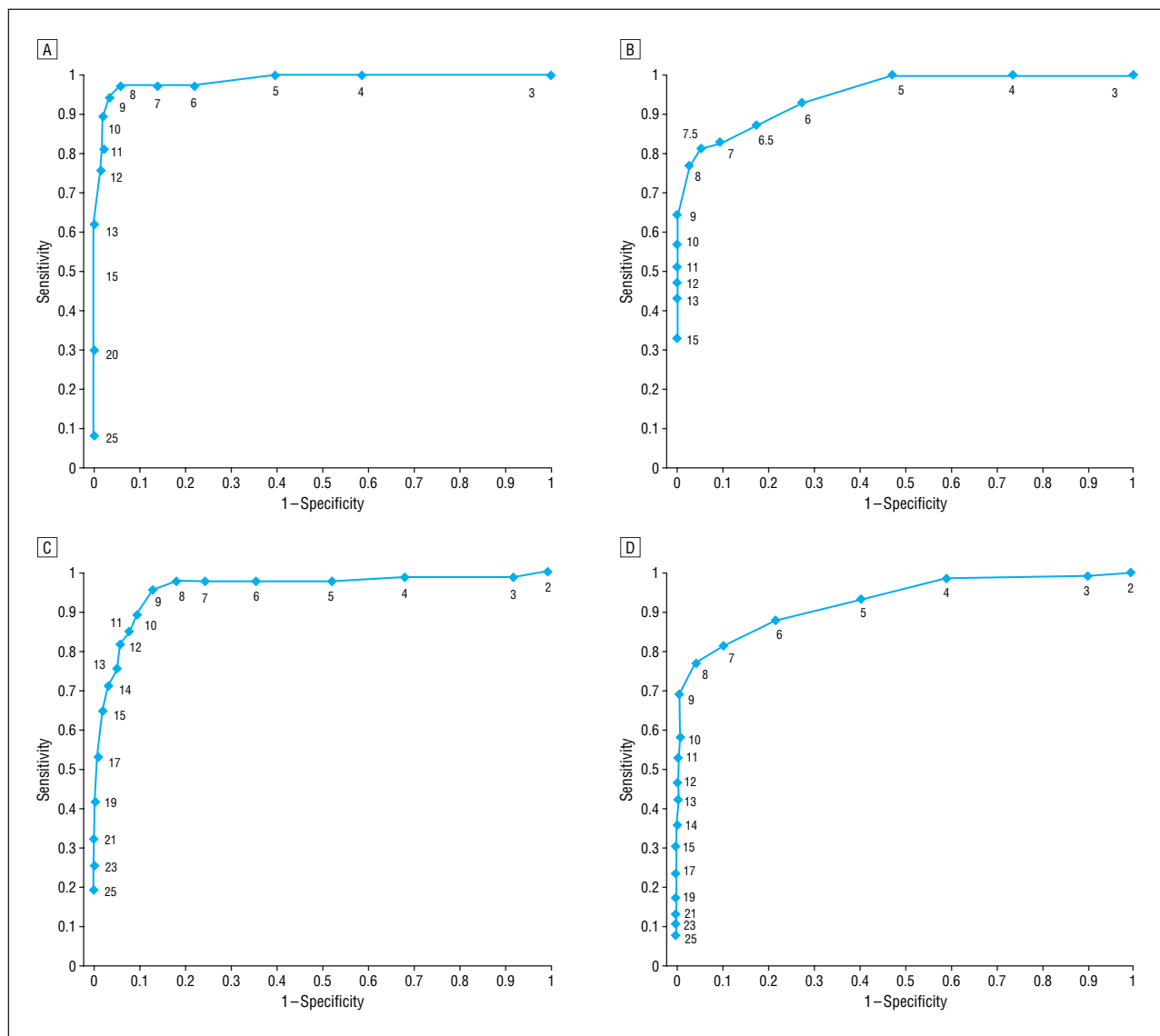
The quadrant-level ROC curve for tortuosity sufficient for plus disease confirmed that a tortuosity index of 9 was a good numeric threshold for plus tortuosity, with a sensitivity of 96% and specificity of 87% (Figure 2C). The quadrant-level ROC curve for tortuosity sufficient for pre-plus disease confirmed that a tortuosity index of 7 was a good numeric threshold for pre-plus tortuosity, with a sensitivity of 96% and specificity of 87% (Figure 2D).

#### COMMENT

In this study, ROPtool showed excellent sensitivity (97%) and specificity (94%) in the detection of tortuosity sufficient for plus disease when consensus of 3 ROP experts was considered to be the reference standard for comparison. Individual examiners also performed well in comparison with expert consensus, so there was not a statistically significant difference in overall concordance with experts between ROPtool (95%) and any of the individual examiners (93%, 93%, and 91%). However, ROPtool's 97% sensitivity in detecting tortuosity sufficient for plus disease compared favorably with that of the individual examiners who had sensitivities of 65%, 70%, and 81%. Specificity was similar between ROPtool (94%) and 2 of the examiners (99% and 93%), and 1 of the examiners had significantly better specificity (100%) than ROPtool.

The finding in this study that sensitivity of ROPtool compared favorably with that of individual examiners is important because high sensitivity is arguably the most desirable aspect of a diagnostic test for plus disease in ROP. High sensitivity means that there are very few cases in which plus disease exists and is not detected. It is important to miss as few cases of plus disease as possible to reduce the risk of delayed treatment and retinal detachment. High specificity, although desirable, is not as





**Figure 2.** Receiver operating characteristic curves for ROPtool's detection of eye-level tortuosity sufficient for plus disease (A), eye-level tortuosity sufficient for pre-plus disease (B), quadrant-level tortuosity sufficient for plus disease (C), and quadrant-level tortuosity sufficient for pre-plus disease (D) in comparison with the consensus of 3 experts. The numbers represent different thresholds for tortuosity index, which are used to calculate ROPtool's sensitivity and specificity relative to expert consensus. For the primary analysis, a tortuosity index of 9 was chosen a priori as the threshold for plus tortuosity.

critical as high sensitivity in diagnosing plus disease. If one assumes that false positives are most likely to occur in cases that are "almost plus disease," many of these eyes will progress to plus disease and need treatment over the next few days or weeks.

The relative value of sensitivity vs specificity has implications for choosing the appropriate numeric thresholds (cutoff points) for tortuosity sufficient for plus disease and for pre-plus disease. The ROC curves are particularly valuable for visualizing this trade-off between sensitivity and specificity (Figure 2). When ROPtool was first developed, the threshold for tortuosity sufficient for plus disease was 10 tortuosity units, because it was the average amount of tortuosity present in the standard photograph of plus disease.<sup>2</sup> After our pilot study, the threshold was adjusted to 9 tortuosity units and a pre-plus threshold of 7 tortuosity units was established.<sup>11</sup> The results presented herein confirm that 9 tortuosity units

is an appropriate threshold for tortuosity sufficient for plus disease. Although using a threshold of 10 or 11 tortuosity units would have improved the overall concordance, it would have done so at the expense of decreasing sensitivity to an unacceptable level (ie, <90%). An important part of the continuing "training" of ROPtool will be to adjust these numeric thresholds as results are obtained from analyses of additional image sets.

It is initially counterintuitive that the mean tortuosities of quadrants with pre-plus and plus tortuosity were 9.9 and 19.1 tortuosity units, respectively (Table 2), even though the numeric thresholds were only 7 for pre-plus tortuosity and 9 for plus tortuosity. This occurred because the range of tortuosity measurements in the study was 1.7 to 45.3 tortuosity units. Therefore, a few very high tortuosity measurements increased the mean value. In addition, 2 quadrants measuring 7 tortuosity units or greater were required for pre-plus tortuosity at the eye

**Table 2. Mean Quadrant Tortuosity Measured by Computer Program (ROPtool) for Each Level of Mean Expert Grade**

Mean Expert Grade <sup>a</sup>	Mean Quadrant Tortuosity (95% Confidence Interval)	P Value <sup>b</sup>
0.0	4.49 (4.34-4.64)	<.001
0.3	6.20 (5.84-6.56)	<.001
0.7	8.56 (7.59-9.52)	.20
1.0	9.67 (8.44-10.90)	.05
1.3	11.49 (10.16-12.83)	<.001
1.7	16.03 (14.01-18.05)	.002
2.0	21.30 (18.91-23.69)	
Normal (0.0-0.3)	4.80 (4.65-4.95)	<.001
Pre-plus (0.7-1.3)	9.88 (9.17-10.59)	<.001
Plus (1.7-2.0)	19.06 (17.36-20.76)	

<sup>a</sup>Three experts graded the tortuosity of each quadrant as 0=normal, 1=pre-plus, and 2=plus.

<sup>b</sup>P values are for differences between adjacent categories.

level (and 9 units for plus tortuosity). As a result, 1 quadrant with even a very high tortuosity was insufficient for eye-level preplus or plus tortuosity.

In addition to the primary analysis involving plus disease, this study looked at ROPtool's ability to distinguish eye-level tortuosity sufficient for pre-plus disease from normal. ROPtool did not perform quite as well when assessing tortuosity sufficient for pre-plus disease, as there were 24 disagreements with expert consensus (compared with only 10 disagreements for plus tortuosity). One possible reason for this observation is that expert consensus for tortuosity sufficient for pre-plus was not defined as rigorously as for plus disease because it was not the primary analysis. Consensus for pre-plus was based on majority opinion, and disagreements were not reconsidered by experts or adjudicated on a conference call. In addition, there is no standard photograph defining the minimum amount of tortuosity for pre-plus disease, so experts may not be well calibrated for diagnosing pre-plus disease. For these reasons, expert consensus for tortuosity sufficient for pre-plus may not have been as accurate as that for plus disease.

Other investigators have reported early success with computer-assisted or computer-automated quantification of dilation and tortuosity of posterior retinal vessels in infants with ROP.<sup>15-20</sup> Heneghan and associates<sup>18</sup> used a general technique for segmenting out vascular structures in retinal images and characterizing the segmented blood vessels. Using images from 23 subjects, they reported a sensitivity of 82% and a specificity of 75% in the detection of plus disease. Swanson and associates<sup>19</sup> used Retinal Image multiScale Analysis (RISA), a semi-automated image analysis software package, to quantify the size and tortuosity of retinal blood vessels. They found that arteriolar tortuosity was associated with ROP severity, but no significant association was observed between ROP severity and venular and arteriolar diameters. Gelman and associates<sup>16</sup> applied an enhanced version of RISA that used their concept of integrated curvature to 32 eyes of 16 infants. They found that eyes with plus disease had vessels with significantly higher integrated curvature, di-

ameter, and tortuosity than those without plus disease. The best diagnostic accuracy was observed with integrated curvature, as evidenced by curvature's area under the ROC curve that was greater than both dilation and tortuosity for both arterioles and venules. We do not directly compare our results with those of these investigators because they used different image sets and addressed different questions.

In this study, ROPtool's ability to measure tortuosity sufficient for plus disease was assessed, even though the diagnosis of plus disease depends on a minimum amount of tortuosity and dilation. ROPtool's dilation measurement is still in development, and it has been our experience that dilation is more difficult than tortuosity to accurately quantify.<sup>11</sup> This is primarily because the degree of change from normal to abnormally dilated is relatively small, as opposed to the relatively large change from a straight to a tortuous vessel. Detecting a small change in vessel width is limited by the number of pixels in a digital image. Even if these challenges are overcome and an accurate automated measure of dilation is developed, it is possible that it will add very little to the overall assessment of plus disease. Kylstra and associates<sup>21</sup> noted that computer-assisted determination of plus disease based on a minimum degree of tortuosity alone showed both excellent sensitivity (85%) and specificity (91%) when compared with evaluation by an expert. Also, if one assumes that abnormal dilation precedes tortuosity in most cases, then it is the development of a threshold level of tortuosity that usually triggers the diagnosis of plus disease.<sup>22,23</sup> Once tortuosity is known, including dilation will change the overall assessment of plus disease only in cases when the vessels are tortuous enough for plus disease but not dilated enough for plus disease. In our experience, this scenario is relatively uncommon.

Plus disease was defined by ICROP as being present when the vascular changes are so marked that the posterior veins are enlarged and the arterioles tortuous. In this study, ROPtool was used to measure the tortuosity of both arterioles and venules for several reasons. First, it is not always possible to reliably differentiate between arterioles and venules in an image of a premature infant's retina. Second, as plus disease progresses from mild to severe, both the arterioles and the venules become dilated and tortuous. In the standard photograph of plus disease that was first used to calibrate ROPtool, both the arterioles and the venules are tortuous. Third, both arteriolar and venular tortuosities were quantified in ROPtool's pilot study, and these data were used to determine cutoff points for tortuosity sufficient for plus disease and for pre-plus disease for this study. Even when both arterioles and venules are tortuous, the arterioles are usually more tortuous than the venules, so using arterioles only would increase the tortuosity of each quadrant and would change these cutoff points. Fourth, it is not yet understood precisely which vessels are used by examiners when determining if plus disease is present, but it seems likely that dilation and tortuosity of all vessels are considered.

This study must be viewed in light of some limitations. First, there is no ideal method for determining the "truth" or "reference standard" with regard to the presence or absence of tortuosity sufficient for plus disease.

This aspect of the study was designed as rigorously as possible by including an initial assessment by experts, reconsideration by the expert outlier when disagreement existed, and a conference call to reach consensus when disagreements persisted. However, it remains possible that in cases where ROPtool or the individual examiners were considered to be incorrect, they were actually correct and expert consensus was wrong. A second limitation of this study is that it would have been preferable to include a larger number of "borderline" images that were not obviously plus disease or normal. This would have made it more difficult for ROPtool and for the examiners, possibly amplifying differences between them. Third, examiners and experts compared cropped RetCam images with the standard photograph of plus disease, and the standard photograph gives a more magnified view of the fundus than that seen on examination or in the cropped images used for this study.

How might computer-assisted or computer-automated analysis be used in the management of infants with ROP in the future? Its greatest potential impact in the near future is in the areas of telemedicine and research. It has become increasingly difficult to provide ROP screening for infants at risk because of the lack of availability of qualified examiners, particularly in developing countries. One possible solution is telemedicine screening, using the Internet to transmit images to ROP reading centers. Several studies have established the feasibility of remote screening for ROP.<sup>12,24-27</sup> Analysis of transmitted images could include a computer-assisted calculation of tortuosity with or without dilation. This information could aid in decisions to treat with laser or to transfer to a center with an ophthalmologist for closer follow-up. Even if plus disease is not present, pre-plus vascular changes may be associated with the eventual need for laser treatment,<sup>23</sup> and the ability to quantify the precise amount of vascular abnormality could influence follow-up intervals. Computer-assisted quantification of plus disease also offers potential benefits to research in ROP. Greater accuracy in measuring the components of plus disease is likely to lead to better understanding of its evolution. This knowledge may facilitate studies of causes of plus disease and the effect of new or existing treatments for ROP. Finally, computer-assisted analysis of plus disease may eventually have a role at the bedside. Retinal images could be analyzed in real time, and a measurement of posterior retinal vascular tortuosity with or without dilation could be included in the record of every examination.

In conclusion, ROPtool is a promising technology that has the potential to reduce subjectivity in the assessment of tortuosity sufficient for plus disease, thereby increasing accuracy and optimizing the timing of laser treatment. Results from this study demonstrate that ROPtool can accurately discriminate between tortuosity sufficient or insufficient for plus disease, but more work remains before it will augment or replace the human observer.

**Submitted for Publication:** May 6, 2007; final revision received July 3, 2007; accepted July 9, 2007.

**Correspondence:** David K. Wallace, MD, MPH, Duke University Eye Center, DUMC 3802, Durham, NC 27710 (david.wallace@duke.edu).

**Author Contributions:** Dr Wallace 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.  
**Financial Disclosure:** None reported.

**Funding/Support:** This project was supported by grant K23 EY015806 from the National Eye Institute.

**Additional Information:** This article is an abridgement of a thesis submitted in partial fulfillment of requirements for membership in the American Ophthalmological Society, May 2007.

**Additional Contributions:** Michael Chiang, MD, MA, Columbia University, New York, New York, David Coats, MD, Baylor College of Medicine, Houston, Texas, Graham Quinn, MD, MSCE, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, and Laura Enyedi, MD, and Terri Young, MD, Duke University, Durham, North Carolina, served as experts and examiners. Sarah Jones, BS, Duke University, assisted with data analysis. Dr Chiang, Anna Ells, MD, Alberta Children's Hospital, Calgary, Alberta, Canada, Antonio Capone, MD, William Beaumont Hospital, Royal Oak, Michigan, and the PHOTO-ROP study group shared RetCam photographs.

## REFERENCES

1. The Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Arch Ophthalmol.* 1984;102(8):1130-1134.
2. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity: preliminary results. *Arch Ophthalmol.* 1988;106(4):471-479.
3. 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(2):295-310.
4. An International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol.* 2005;123(7):991-999.
5. Schaffer DB, Palmer EA, Plotsky DF, et al. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Prognostic factors in the natural course of retinopathy of prematurity. *Ophthalmology.* 1993;100(2):230-237.
6. 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(12):1684-1694.
7. Capone A Jr, Ells AL, Fielder AR, et al. Standard image of plus disease in retinopathy of prematurity. *Arch Ophthalmol.* 2006;124(11):1669-1670.
8. Freedman SF, Kylstra JA, Hall JG, Capowski JJ. Plus disease in retinopathy of prematurity—photographic evaluation by an expert panel [abstract]. Paper presented at: Annual Meeting of the Association for Research in Vision and Ophthalmology; May 14, 1995; Fort Lauderdale, FL.
9. Aylward SR, Bullitt E. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Trans Med Imaging.* 2002;21(2):61-75.
10. Bullitt E, Gerig G, Pizer SM, Lin W, Aylward SR. Measuring tortuosity of the intracerebral vasculature from MRA images. *IEEE Trans Med Imaging.* 2003;22(9):1163-1171.
11. Wallace DK, Zhao Z, Freedman SF. A pilot study using "ROPtool" to quantify plus disease in retinopathy of prematurity [published online ahead of print May 29, 2007]. *J AAPOS.* 2007;11(4):381-387.
12. Ells AL, Holmes JM, Astle WF, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology.* 2003;110(11):2113-2117.
13. Photographic Screening for Retinopathy of Prematurity (PHOTO-ROP) Cooperative Group; Balasubramanian M, Capone A Jr, Hartnett ME, Pignatto S, Trese MT. 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.

14. Jung SH, Ahn C, Donner A. A study of an adjusted chi-square statistic as applied to observational studies involving clustered binary data *Stat Med*. 2001;20(14):2149-2161.
15. 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(6):490-500.
16. Gelman R, Martinez-Perez ME, Vanderveen DK, Moskowitz A, Fulton AB. Diagnosis of plus disease in retinopathy of prematurity using retinal image multi-scale analysis. *Invest Ophthalmol Vis Sci*. 2005;46(12):4734-4738.
17. Freedman SF, Kylstra JA, Capowski JJ, et al. Observer sensitivity to retinal vessel diameter and tortuosity in retinopathy of prematurity: a model system. *J Pediatr Ophthalmol Strabismus*. 1996;33(4):248-254.
18. 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(4):407-429.
19. Swanson C, 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(12):1474-1477.
20. 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(4):723-725.
21. Kylstra JA, Freedman FS, Capowski JJ, Hall JG. Measurement of retinal vessel tortuosity as a means of diagnosing plus disease in ROP. Paper presented at: Annual Meeting of the Association for Research in Vision and Ophthalmology; May 14, 1995; Fort Lauderdale, FL.
22. Saunders RA, Bluestein EC, Sinatra RB, et al. The predictive value of posterior pole vessels in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus*. 1995;32(2):82-85.
23. 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(4):224-229.
24. Schwartz SD, Harrison SA, Ferrone PJ, Trese MT. Telemedical evaluation and management of retinopathy of prematurity using a fiberoptic digital fundus camera. *Ophthalmology*. 2000;107(1):25-28.
25. Yen KG, Hess D, Burke B, et al. The optimum time to employ telephotoscreening to detect retinopathy of prematurity. *Trans Am Ophthalmol Soc*. 2000;98:145-150.
26. Roth DB, Morales D, Feuer WJ, et al. Screening for retinopathy of prematurity employing the RetCam 120: sensitivity and specificity. *Arch Ophthalmol*. 2001;119(2):268-272.
27. Chiang MF, Keenan JD, Starren J, et al. Accuracy and reliability of remote retinopathy of prematurity diagnosis. *Arch Ophthalmol*. 2006;124(3):322-327.

#### From the Archives of the Archives

The author regards the tumors of the optic nerve as belonging to the group of elephantiasis neuromatodes. He believes that almost always there are, besides the evident tumor of the nerve, other tumor nodules within the cranium, and thus it happens that months or years after the operation the patient dies with cerebral symptoms.

**Reference:** Emanuel (Leipsic). The relation of tumors of the optic nerve to elephantiasis neuromatodes. *Arch Ophthalmol*. 1902;31.