COMPUTER-ASSISTED QUANTIFICATION OF VASCULAR TORTUOSITY IN RETINOPATHY OF PREMATURITY (AN AMERICAN OPHTHALMOLOGICAL SOCIETY THESIS)

BY David K. Wallace MD MPH

ABSTRACT

Purpose: Plus disease is severely abnormal dilation and tortuosity of posterior retinal blood vessels in infants with retinopathy of prematurity (ROP). It has become the major criterion for laser treatment in ROP, but its assessment is subjective and prone to error. ROPtool is a computer program that traces retinal blood vessels, measures their tortuosity, and determines whether there is sufficient tortuosity for plus disease. The purpose of this study was to prospectively determine if assessment of tortuosity by ROPtool is more accurate than by individual pediatric ophthalmologists.

Methods: One hundred eighty-five high-quality RetCam images from premature infants were circulated to 3 ROP experts to develop reference data and to 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. Overall accuracy, sensitivity, and specificity of ROPtool relative to expert consensus were compared to that of the individual examiners.

Results: By expert consensus, 37 of the 185 eyes (20%) had tortuosity sufficient for plus disease. The overall accuracy of ROPtool of 95% (175/185) for identifying tortuosity sufficient for plus disease was similar to that of examiner 1 (93%, 172/185, $P = .5$), examiner 2 (93%, 172/185, $P = .5$), and examiner 3 (91%, 168/185, $P = .1$). Sensitivity of ROPtool of 97% (36/37) was superior to that of examiner 1 (65%, 24/37, $P < .001$), examiner 2 (70%, 26/37, $P < .001$), and examiner 3 (81%, 30/37, $P = .02$). The mean tortuosity of quadrants with plus disease was 19.1 tortuosity units, compared to 9.9 tortuosity units for quadrants with pre-plus ($P < .001$) and 4.8 tortuosity units for normal quadrants ($P < .001$ for pre-plus vs normal).

Conclusions: ROPtool has excellent sensitivity and overall accuracy relative to expert consensus in the detection of tortuosity sufficient for plus disease. Computer-assisted analysis of retinal images has the potential to remove subjectivity from the determination of plus disease and to optimize the timing of follow-up and treatment for ROP.

Trans Am Ophthalmol Soc 2007;105:594-615

INTRODUCTION

PUBLIC HEALTH SIGNIFICANCE OF ROP

Severe retinopathy of prematurity (ROP) can lead to retinal detachment and permanent visual loss. Each year in the United States, approximately 500 infants are blinded by ROP, and 1300 infants develop ROP that is severe enough to require laser treatment.[1] Despite treatment, many eyes develop retinal detachments, which are associated with a poor visual prognosis.[2] Compounding the handicap experienced by the child, there is a very large cost to society based on medical expenses and lost wages over a lifetime for each new case of childhood blindness.[3,4]

INTRODUCTION TO ROP

Normal retinal vascular development begins with the appearance of blood vessels in the optic nerve at 16 weeks gestation. Stimulated by vascular endothelial growth factor and other vasculogenic proteins, these vessels grow progressively outward from the optic disc toward the retinal periphery. In those infants born at full term, the retinal vessels reach the ora serrata nasally by 36 weeks gestational age and temporally by 40 weeks. However, when infants are born prematurely, their retinal blood vessels have not had sufficient time to grow in utero to the far retinal periphery. Exposure to factors in the extrauterine environment, including higher relative oxygen levels, disrupts the normal process of retinal vascular development. As a result, clinically visible signs of ROP develop in the blood vessels in the central, posterior part of the retina (the "posterior pole") and at the border between the vascular and the avascular retina in the peripheral aspect of the retina. These signs are usually present between 31 and 42 weeks postmenstrual age (gestational age at birth + chronological age).

Most cases of ROP are mild to moderate and resolve without serious sequelae. Some cases of ROP are severe, and they are characterized by abnormal dilation and tortuosity of the arterioles and venules (plus disease) and growth of abnormal neovascular tissue on the surface of the retina (stage 3). In the Cryotherapy for Retinopathy of Prematurity (CRYO-ROP) clinical trial, 245 of 4099 infants (6%) developed severe, sight-threatening (threshold) ROP, and 2699 of 4099 (66%) infants developed less severe forms of the disease.[5] In cases of severe ROP, peripheral retinal ablation is done using laser or, less commonly today, using cryotherapy. These treatments reduce the secretion of vasoactive cytokines that stimulate proliferation of neovascular tissue, resulting in regression of disease in most cases. Despite timely treatment, some infants still experience severe disease that progresses to retinal detachment and blindness. Of those eyes treated with cryotherapy and followed for 15 years in the CRYO-ROP study, 85 of 190 (45%) had a visual acuity of 20/200 or worse, and 54 of 180 (30%) had an unfavorable anatomic outcome, defined as a posterior retinal fold or worse.[6]

INTERNATIONAL CLASSIFICATION OF ROP AND ITS REVISION

In 1984, the Committee for the International Classification of Retinopathy of Prematurity (ICROP) published the classification system for ROP that is widely used today.[7] This classification defined key parameters used to describe the disease: zone (location of disease), stage (severity of disease), extent (expressed as number of clock hours of disease), and plus disease (abnormal dilation and tortuosity of the arterioles and venules). Zone I is the most posterior location, zone II represents an intermediate location, and zone III is a crescent of far peripheral, mostly temporal retina (Figure 1). Stage describes the severity of disease at the border of the vascular and avascular retina, and it is graded from 1 (very mild) to 5 (most severe). Stage 1 is a demarcation line, stage 2 is a ridge with height and width, stage 3 is extraretinal fibrovascular proliferation, stage 4A is an extrafoveal retinal detachment, stage 4B is a subtotal retinal detachment involving the fovea, and stage 5 is a total retinal detachment. The circumferential extent of the stage is characterized by describing the number of clock hours (1 through 12), or 30-degree sectors, of retinal involvement.

Plus disease is severely abnormal dilation and tortuosity of posterior retinal blood vessels, and it exists when an eye is judged to have at least as much dilation and tortuosity of its posterior retinal vessels as a standard photograph that was used in the CRYO-ROP study (Figure 2).[5] A revised classification of ROP was published in 2005,[8] and it included an intermediate grade of posterior retinal vascular abnormality known as "pre-plus" disease, which is believed in most cases to be a precursor to plus disease. Whereas plus disease refers to severely abnormal vascular changes, the designation of pre-plus disease applies to mildly to moderately abnormal dilation and tortuosity of the central, posterior retinal blood vessels (Figure 3).
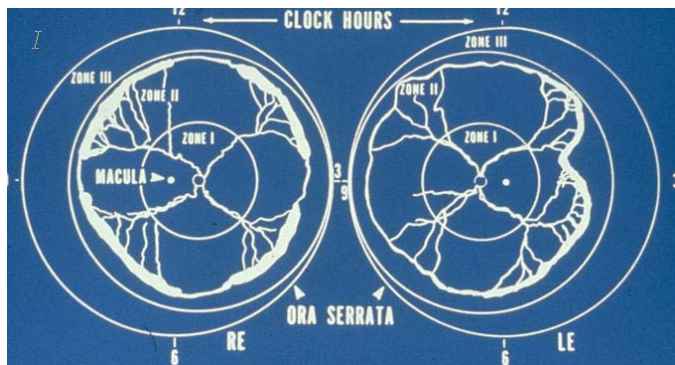

FIGURE 1

Typical appearance of severe (threshold) retinopathy of prematurity, showing plus disease (abnormal dilation and tortuosity of posterior retinal vessels), stage 3 (neovascular tissue) at the border of vascular and avascular retina, and the locations of zones I, II, and III. (Reprinted with permission from *Archives of Ophthalmology.*[5])
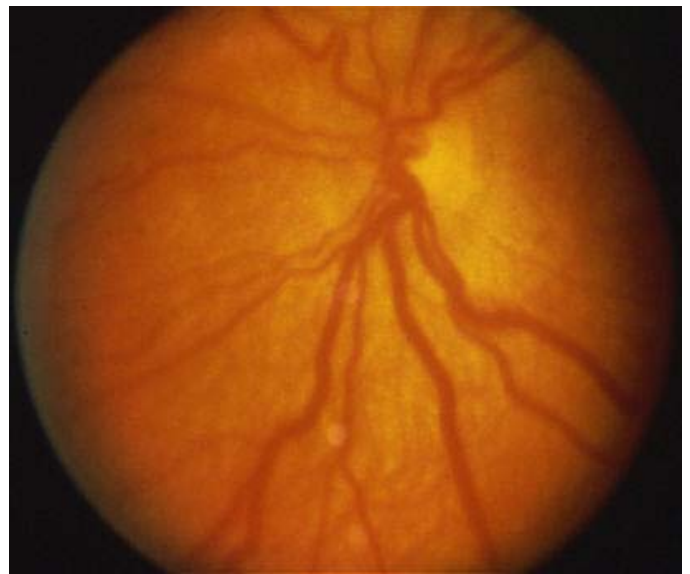

FIGURE 2

Standard photograph of plus disease in retinopathy of prematurity representing the minimum degree of dilation and tortuosity necessary for plus disease. (Reprinted with permission from *Archives of Ophthalmology.*[5])

ROP TREATMENT TRIALS

The CRYO-ROP study was the first multicenter, randomized, controlled clinical trial to demonstrate that ablation of the avascular, peripheral retina reduced the incidence of an unfavorable outcome in cases of severe ROP.[5] Threshold ROP was defined in this study as at least 5 contiguous or 8 total clock hours of extraretinal neovascularization (stage 3) in zone I or zone II with plus disease. Two hundred ninety-one infants developed threshold ROP, and eyes were randomly assigned to treatment with cryotherapy or to observation alone. After 15 years, 254 children were available for follow-up examinations. Unfavorable visual outcome, defined as 20/200 or worse, was present in 45% of treated eyes and in 64% of control eyes ($P < .001$). Unfavorable anatomic outcome, defined as posterior retinal fold, retinal detachment involving the macula, or retrolental mass, was present in 30% of treated eyes and 52% of control eyes ($P < .001$).[6] Compared to eyes with no ROP, a visual field reduction of 30% to 37% existed in treated eyes after 10 years compared to a reduction of 27% to 33% in control eyes. However, when blind eyes were included in this analysis, the visual field area was 25% larger in treated eyes compared to control eyes.[9] There was no significant difference in contrast sensitivity between treated and control eyes.[10] Of concern is that retinal detachments occurred between the 10-year and 15-year examinations in 6 patients in the

treated group and 7 patients in the control group. This finding emphasizes the need for long-term monitoring of patients with severe ROP, even after disease regression with or without treatment.[6] Although cryotherapy was the first treatment proven to be effective, it has largely been replaced by laser, and it is now generally used only if there is a poor view of the retina preventing transpupillary laser photocoagulation.
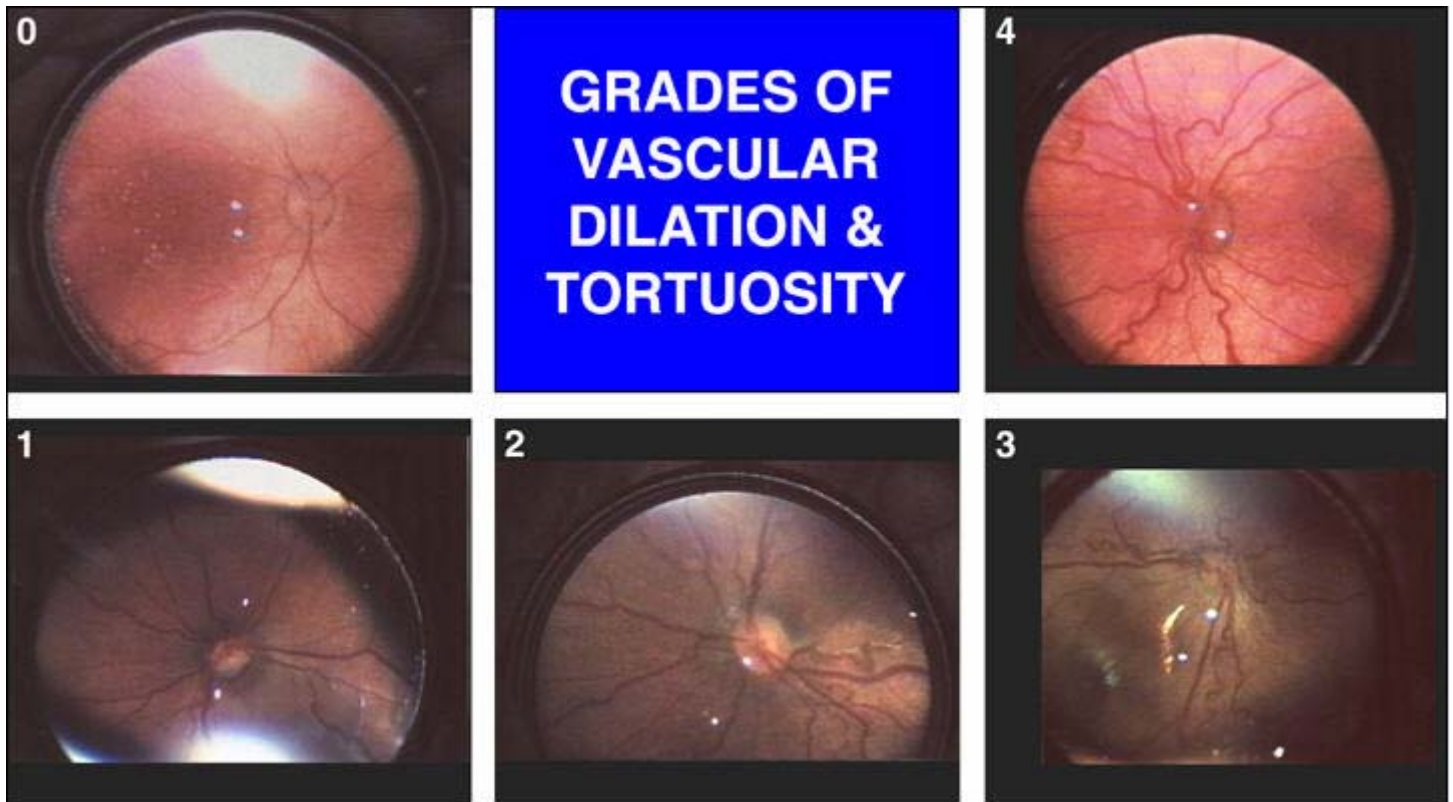

FIGURE 3

Spectrum of posterior pole vascular changes in infants with retinopathy of prematurity. 0 = normal, 1 = mild venous dilation, 2 = mild venous dilation and mild arteriolar tortuosity (pre-plus disease), 3 = moderate vascular dilation and tortuosity (pre-plus, almost plus disease), and 4 = plus disease. (Reprinted with permission from *Journal of AAPOS*.[25])

Investigators in the Early Treatment for ROP (ETROP) study hypothesized that treatment at high-risk prethreshold would reduce the incidence of unfavorable visual outcomes compared to standard treatment at threshold.[11] Prethreshold was defined as any of the following: (1) any ROP in zone I, (2) ROP in zone II with plus disease, (3) zone II with stage 3 ROP and no plus disease, or (4) zone II with plus disease and stage 3 ROP but with less than the 5 contiguous or 8 cumulative clock hours of stage 3 required for the diagnosis of threshold ROP. "High-risk" was determined using a computerized model based on risk factors for progression identified in the CRYO-ROP natural history cohort. Three hundred seventeen bilateral cases and 84 asymmetric cases reached high-risk prethreshold, and eyes were randomly assigned either to laser treatment within 48 hours or to standard treatment at threshold if it developed. After 9 months, an unfavorable visual outcome had occurred in 14.5% of eyes in the early treated group compared to 19.5% of eyes receiving standard treatment. (P = .01) An unfavorable anatomic outcome had occurred in 9.1% of eyes in the early treated group compared to 15.6% of eyes receiving standard treatment. (P < .001). Ocular complication rates were similar in the 2 groups, although systemic complication rates were higher in the early treated group. After 9 months, myopia of 5 D or greater was present in 26% of early treated eyes and in 28% of eyes treated conventionally (P = .2). Based on the results of this study, laser treatment is now strongly considered for any eye with type 1 ROP, defined as (1) zone I, any stage with plus disease, (2) zone I, stage 3 without plus disease, or (3) zone II, stage 2 or 3 with plus disease.[12]

DEFINITION AND ETIOLOGY OF PLUS AND PRE-PLUS DISEASE

Plus disease was originally defined by ICROP as being present when "the vascular changes are so marked that the posterior veins are enlarged and the arterioles tortuous".[7] The CRYO-ROP study introduced the concept of a standard photograph representing the minimum acceptable abnormality required for plus disease[5] (Figure 2). 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.[13] The International Committee that revisited ICROP

reaffirmed the standard photograph from the CRYO-ROP study as the minimum abnormality required and provided additional photographic examples of plus disease. They also 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."[8]

It has been hypothesized that the severely abnormal dilation and tortuosity of blood vessels characteristic of plus disease occur as a result of reduced capillary resistance and increased retinal blood flow through developing arteriovenous shunts.[14-16] This hypothesis is supported by a strong correlation between the appearance of posterior arterioles and venules and the severity of ROP in the retinal periphery.[15,17] However, Holland and associates[18] used color Doppler imaging to show that blood flow (peak systolic velocity) through the optic nerve was not increased in eyes with plus disease. An alternative explanation for the appearance of plus disease is that vascular endothelial growth factor, which is produced in response to retinal ischemia, may be present in sufficient amounts in the vitreous of the eye to act directly on central retinal blood vessels and alter their morphology (Ho M, ARVO Meeting, 2000. Abstract). This proposed mechanism may also help to explain the observation that plus disease occurs before obvious peripheral shunting and neovascularization in some cases of severe, posterior ROP.

PROGNOSTIC SIGNIFICANCE OF PLUS DISEASE

The presence of plus disease is a marker for severe disease, and it is one of the most important prognostic indicators in ROP.[5,7,14] 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 anatomic outcome, compared to only 1 of 31 eyes (3%) with stage 3 but without plus disease at the same age.[5] The presence or absence of plus disease is now the primary factor in determining whether an infant with ROP requires laser treatment or not, since plus disease is one of the criteria for type 1 ROP and for threshold. 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."[19] Timely and accurate assessment of plus disease is critical to optimize the chances of a favorable visual outcome for infants with ROP. Unnecessary treatment may result in loss of peripheral vision due to destruction of peripheral retina, and it can be associated with complications such as cataract[20-22] and glaucoma.[23] Conversely, treatment delay may be associated with a greater risk of retinal detachment and blindness.

PROGNOSTIC SIGNIFICANCE OF PRE-PLUS DISEASE

Although plus disease is graded simply as present or absent, experienced examiners recognize that a spectrum of posterior retinal vascular changes exist in ROP, from very early venous dilation to severe venous and arteriolar dilation and tortuosity. The CRYO-ROP Study Group recognized that "there is a need for a practical way to quantitate plus disease. This would permit further exploration of the effect of plus disease on ROP outcome."[24] In response to this need, a grading system has been developed to describe vascular changes in ROP (Figure 3), which offered a slight modification of one previously developed by Saunders and associates.[15] Use of any grading system was reasoned to be of practical value only if it had prognostic significance, thereby influencing examination intervals or treatment decisions. It was hypothesized, based on clinical experience, that since plus disease had major prognostic significance, pre-plus disease would be predictive of disease progression as well. Two pilot studies have been done that support this hypothesis.

In the first study,[25] video clips of posterior pole images captured at the examination closest to 33 weeks postmenstrual age were analyzed. Images of 64 eyes of 32 infants screened during an 18-month time period were randomly ordered and graded by 2 masked examiners. The examiners graded the images from 0 to 4 in comparison to standard photographs of 5 representative grades of retinal vascular dilation and tortuosity (Figure 3). The standard photographs were digital images captured from video clips taken during indirect ophthalmoscopy. The examiners disagreed on the precise grade in 16 of 64 images (25%), and in these cases, the grades of the 2 examiners were averaged. The incidences of progression to stage 3, development of plus disease, and requirement of laser treatment were compared between infants with normal posterior retinal blood vessels and those judged to have early dilation and tortuosity insufficient for plus disease. This analysis included only one eye per infant, which was the right eye unless the left eye had more severe disease. Of 8 eyes with mild vascular dilation and tortuosity insufficient for plus disease, 5 (63%) progressed to plus disease, 5 (63%) eventually required laser treatment, and 4 (50%) developed stage 3. (These outcomes were not mutually exclusive, and some eyes were treated prior to threshold.) Of the 24 eyes with normal posterior retinal vessels, none required laser treatment, 2 (8%) developed stage 3, and none progressed to plus disease. The group with vascular dilation and tortuosity insufficient for plus disease had a significantly higher incidence of progression to laser treatment ($P = .0003$), stage 3 ($P = .027$), and plus disease ($P = .0003$).[25]

The results of this study suggested that pre-plus disease had prognostic significance at the examination performed closest to 33 weeks postmenstrual age, but whether it was applicable to the entire screening period remained unknown. Therefore, a second pilot study was performed using serial RetCam photographs that had been collected as part of a separate study (Wallace DK, AAPOS Meeting, 2003. Abstract). RetCam photographs were taken at every examination of 44 infants for an entire year, for a total of 371 serial images over the entire screening period.[26] In order to assess the prognostic significance of pre-plus disease at various ages, a masked examiner graded the posterior retinal blood vessels of these images in comparison to the same standard photographs used in the first pilot study. The incidences of progression to stage 3, development of plus disease, and requirement of laser treatment were compared between eyes with normal posterior retinal vessels and those with pre-plus vascular changes at various ages throughout the screening period.

Eighteen of the infants had serial photographs taken throughout their entire screening period and were included in the data analysis. In comparison to eyes (one per patient) with normal posterior retinal vessels on every examination, those with pre-plus changes on at least one examination had a higher incidence of stage 3 (5/11, 45% vs 1/4, 25%), plus disease (4/11, 36% vs 0/4, 0%),

and laser treatment (3/11, 27% vs 0/4, 0%), but these differences were not statistically significant. In 7 eyes that required laser treatment, plus disease was the first ominous sign in 3 cases and pre-plus was the first sign (appearing prior to stage 3 or plus disease) in 4 cases. Pre-plus disease was most predictive of progression to stage 3 when it was present at 34 to 39 weeks postmenstrual age. It was predictive of development of plus disease when present at 33 to 34 and 36 weeks, and it predicted need for laser treatment when present at 33 to 38 weeks.

NATURAL HISTORY OF VASCULAR CHANGES

Because infants are typically examined only once every 1 to 2 weeks, the natural history of posterior retinal vascular changes in ROP is not completely understood. In order to ascertain the underlying mechanisms promoting disease, it would be desirable to characterize the timing of various degrees of retinal vascular dilation and tortuosity. Most experts agree that there is a progression of posterior retinal vascular changes that can be observed between normal and plus disease. A grading system was first introduced by Saunders and associates,[15] and it included (0) normal, (1) venous dilation, (2) venous dilation and arteriolar tortuosity, and (3) plus disease. This grading scheme is consistent with the premise that abnormal dilation precedes tortuosity in most cases. The first pilot study on the prognostic significance of pre-plus disease utilized a similar grading system, with the addition of a category for moderate vascular dilation and/or tortuosity, or "almost plus disease."[25]

In the second pilot study, the natural history of posterior retinal blood vessel changes was also studied (Wallace DK, AAPOS Meeting, 2003. Abstract). It was found that grade 2 (mild dilation and tortuosity) was used much more commonly by the masked examiners than either grade 1 (mild dilation) or grade 3 (almost plus disease) to describe the images in the study. In addition, of 8 patients with grade 2 vascular changes, only 2 had prior examinations with grade 1 findings, suggesting that many eyes develop mild vascular dilation and tortuosity simultaneously or so close in time as to escape detection with examinations every 1 to 2 weeks.

Consequently, there remains controversy regarding the usual progression of central retinal blood vessel changes in ROP. Some examiners believe that, contrary to popular opinion, vascular tortuosity precedes dilation in some cases. It may be that the natural history of vascular changes, as well as the timing of events, depends on the location of the disease; that is, more posterior disease tends to progress more rapidly and may develop dilation and tortuosity simultaneously. However, no published study has reported on the natural progression of vascular changes in a large number of infants, and such a study is necessary to verify or refute the accuracy of current classification systems of posterior retinal vascular changes in ROP.

COMPUTER-ASSISTED QUANTIFICATION OF PLUS DISEASE

The accurate and timely determination of plus disease is extremely important to optimize the chance of a favorable visual outcome in infants with ROP. However, even for experienced examiners, it can be very difficult to judge the degree of vascular change present in an infant's posterior pole. Freedman and associates found that 3 principal investigators in the CRYO-ROP study disagreed on the presence or absence of plus disease in 29 of 72 retinal images (40%) (Freedman SF, ARVO Meeting, 1995. Abstract). In addition, there can be significant intraobserver variability due to factors such as the type of ophthalmic lens used, clarity of the image, size of the optic nerve, amount of pigmentation of the underlying retina, and relative amount of dilation and tortuosity compared to other infants examined recently. 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.

An accurate, automated measure of plus disease is needed for several additional reasons. It has become more difficult for many neonatal nurseries in the United States to find ophthalmologists willing to perform ROP examinations. This problem is due in part to fewer graduates from pediatric ophthalmology fellowships in recent years as well as the substantial medicolegal risk of screening for ROP. Also, the number of infants requiring ROP examinations has increased owing to improved survival of very low-birth-weight infants and recent changes in screening guidelines to now include infants born at 29 or 30 weeks gestational age with birth weights over 1500 g.[27] Many developing countries are facing an increasing number of surviving premature infants without an adequate number of ophthalmologists trained to screen and treat ROP. Telemedicine offers a potential solution for these problems, and computer-assisted analysis of images would greatly enhance a telemedicine approach to screening.

In response to these needs, interest has grown in recent years in measuring as accurately as possible the degree of blood vessel dilation and tortuosity. In collaboration with computer scientists, a computer program was developed that automatically traced retinal blood vessels and quantified the degree of tortuosity and dilation. This technology was first applied to the precise measurement of cerebral vessel dilation and tortuosity in angiograms.[28-31] Three pilot studies have been performed to assess the accuracy of this computer program at various stages of its evolution.[32] In the first study, a representative sample of posterior retinal images from 20 premature infants, 10 normal and 10 representing various degrees of dilation and tortuosity, was extracted from a video database and analyzed by both the computer program and by 2 masked examiners. The standard photograph from the CRYO-ROP study (Figure 2) representing the minimum degree of dilation and tortuosity required for plus disease was also digitized, analyzed, and used as a numeric comparison for the automated determination of plus disease. Of the 5 images determined to have plus disease by both examiners, 4 were calculated to have plus disease by the computer program (80% sensitivity). Of the 11 images without plus disease, 10 were calculated to not have plus disease by the computer program (91% specificity).[32] The examiners disagreed on the presence or absence of plus disease for 3 images, and 1 poor-quality image was excluded.

A second study utilized RetCam images from 22 premature infants collected as part of the Photographic Screening for Retinopathy of Prematurity (PHOTO-ROP) study.[33] Of 6 images determined to have plus disease by 2 masked examiners, 5 were calculated to

have plus disease by the computer program (83% sensitivity). Of 13 images without plus disease, 9 were calculated to not have plus disease by the computer program (69% specificity) (Wallace DK, ARVO Meeting, 2003. Abstract). Although promising, these 2 pilot studies had several limitations. First, the sample sizes were small, and larger samples are required to obtain more accurate estimates of sensitivity and specificity. Second, accurate determination of vessel dilation was problematic owing to variables that can result in variations of image size, such as the power of condensing lens used to capture the image. Third, the blood vessels were often not traced by the computer for their entire length. In particular, sharp turns in a vessel were sometimes missed, and this problem affected the measurement of tortuosity. One of the most important variables in success of the tracing program was the quality of the digitized image. Those blood vessels that were out of focus or exhibited poor contrast against the fundus background were more difficult for the computer to identify, trace, and analyze.

Since completion of these pilot studies, the computer program has been refined and named "ROPtool." The program was previously unable to consistently identify on its own the major vessels from each image, so a feature was added that allowed the operator to click on individual vessels. The program now traces the selected vessel from the border of the optic nerve to the extent of the view that an examiner would have with a 28 D lens centered on the optic nerve. This view subtends an angle of approximately 30 degrees, and it includes those segments of arterioles and venules that are generally used to determine the presence or absence of plus disease. Another problem in the pilot studies was that the tortuosity measure was calculated as the ratio of the total vessel length to the distance of a straight line drawn between the beginning and ending points of the vessel. However, this simple formula had the effect of artificially increasing the tortuosity of a nontortuous vessel that simply makes a large, gradual curve (Figure 4). The program was refined to choose points along the vessel that are 200 points (about 40 pixels) apart and to draw a smooth curve connecting these points. The tortuosity measure used by ROPtool is now the ratio of the total length of the vessel to the length of this smooth curve, which likely gives a more accurate measure of tortuosity. Also, features have been added to allow the operator to mark the borders of the optic nerve and the center of the macula. Finally, the interface was modified to make it "user friendly" and to include functions such as cutting, merging, deleting, and undoing vessel tracings. The result is the current working version of ROPtool shown in Figure 5.
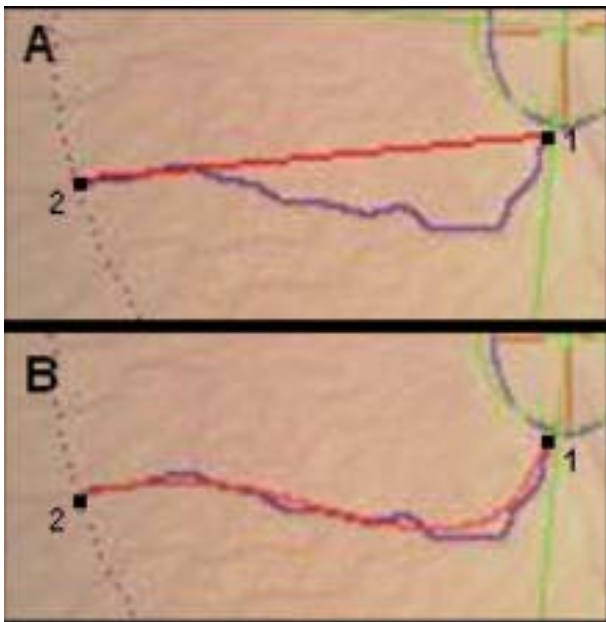

FIGURE 4

Two methods of measuring tortuosity of retinal blood vessels: A, Using the simplest method, tortuosity is calculated as the ratio of the total vessel length (purple line) to a straight line drawn between points 1 and 2 (red line). This method will overestimate tortuosity in a vessel such as this one that curves gradually. B, ROPtool chooses points along the vessel about 40 pixels apart and draws a smooth curve (shown in red) connecting these points. Tortuosity is then calculated as the ratio of the total vessel length (purple line) to the length of the smooth curve (red curve) between points 1 and 2.
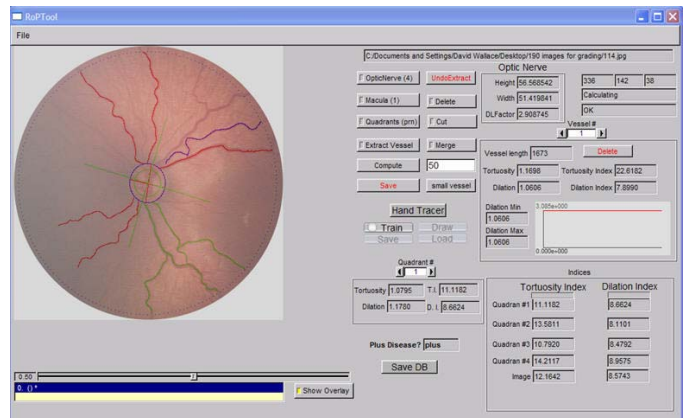

FIGURE 5

User interface for ROPtool, a computer program that traces retinal blood vessels of premature infants and measures their tortuosity. The operator clicks on any of the functions on the right and then executes this function by clicking on the image on the left.

Recently, this new version of ROPtool was piloted in order to troubleshoot for problems, determine its reliability (repeatability), assess its accuracy as a tool to measure dilation and tortuosity, and calculate appropriate numerical thresholds for tortuosity sufficient for plus disease and pre-plus disease.[34] During the pilot study, it was noted that the same vessels were sometimes traced twice and that tracing occasionally continued around the optic nerve border, and ROPtool was modified to correct these problems. Reliability and accuracy of the tortuosity measure of ROPtool were equal to or better than determination of tortuosity sufficient for plus disease by investigator judgment. Tortuosity of both arterioles and venules were included in the overall tortuosity measure, primarily because it is not always possible to accurately differentiate between them, and it is common for arterioles and venules to become tortuous as plus disease develops. 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.

In summary, a computer program (ROPtool) has been developed that automatically traces retinal blood vessels, measures the tortuosity of each vessel in comparison to the standard photograph of plus disease, and calculates whether an eye has sufficient tortuosity for plus disease or not. Several pilot studies have been completed, and results have been used to refine the computer program. The purpose of this study was to assess the accuracy of measurement of vascular tortuosity by ROPtool. It was hypothesized that determination of the presence or absence of tortuosity sufficient for plus disease by ROPtool would be superior to judgment by individual pediatric ophthalmologists.

HYPOTHESIS

Computer-assisted quantification of posterior retinal vascular tortuosity in premature infants is more accurate than assessment of tortuosity by experienced pediatric ophthalmologists.

METHODS

Six pediatric ophthalmology colleagues of the author were identified and asked to participate in the study. Three of them were considered by the author to be ROP experts, based on their extensive ROP experience and certification as investigators in the ETROP study. The other 3 ophthalmologists (the "examiners") had different amounts of experience in ROP diagnosis and treatment. 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 (Clarity Medical Systems, Pleasanton, California) photographs of the posterior retina of 190 different eyes of premature infants were collected. This number was based on a priori sample size calculations. 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 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. The ages of the infants at the time of the photographs were not known, so age was not considered when selecting images. Images from these 190 eyes were not included in any of the pilot studies that were used to train ROPtool. Some images came from photographic databases of previously published studies done elsewhere,[26,33] 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 that approximated the view seen with a 28 D lens (Figure 6). 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. Cropping the RetCam images in this manner assured that all ophthalmologists and ROPtool used the same view and extent of the retina to judge tortuosity sufficient for plus disease and those 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 using ROPtool by the author (the "operator"), who was masked to the grades of the other ophthalmologists.

All 6 pediatric ophthalmologists were informed about the general purpose of the study and sent a copy of the standard photograph of plus disease (Figure 2) as well as examples of plus disease and pre-plus disease from the ICROP Revisited publication.[8] They were also reminded of the definition of plus disease ("the standard photograph defines the minimum amount of vascular dilation and tortuosity required to make the diagnosis of plus disease") and pre-plus disease ("vascular abnormalities of the posterior pole that are insufficient for the diagnosis of plus disease but that demonstrate more arterial tortuosity and more venous dilation than normal"). 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. 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 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 three experts were used to establish the reference judgment of tortuosity sufficient for plus disease ("plus tortuosity") for each image. Expert consensus of eye-level tortuosity sufficient for plus disease versus not plus disease (the reference standard) was determined as follows. If at least 2 quadrants had tortuosity sufficient for plus disease, then the eye was determined to have tortuosity sufficient for plus disease. When there was disagreement as to whether there was or was not 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 between the experts. If consensus was still not reached, then each expert

recast a vote, and consensus was determined by majority. 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. An eye was considered to have tortuosity sufficient for pre-plus disease if at least 2 quadrants had pre-plus tortuosity, or if 1 quadrant had plus tortuosity and 1 or more quadrants had pre-plus tortuosity.

The 3 examiners had different amounts of experience in ROP diagnosis and treatment in order to represent pediatric ophthalmologists in practice. The total number of ROP examinations performed by the each of them was approximately 1000, 7000, and 10,000, respectively. Their grades for the images were used to test the primary hypothesis that the computer program (ROPtool) is more accurate than an individual examiner 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.

HOW DOES ROPtool WORK?

The interface of ROPtool is shown in Figure 5. Images are analyzed using the following technique. First, the operator clicks (using the computer's mouse) on the superior, inferior, left, and right borders of the optic nerve, and ROPtool generates an oval that approximates the optic nerve border (Figure 7). Segments of vessels on the optic nerve that are within this oval are ignored. Second, the operator clicks on the center of the macula, and ROPtool draws a circle with the optic nerve in the center and with a radius equal to the distance between the center of the optic nerve and the center of the macula. The image within this circle (the "28 D circle") approximates the view with a 28 D lens, and vessels outside of this circle are excluded from the analysis. ROPtool also uses the position of the center of the macula to generate 4 green lines that separate the quadrants from each other (Figure 7). One line connects the center of the optic nerve to the center of the macula, and the other lines are 90, 180, and 270 degrees from the first one. Only blood vessels or segments of blood vessels in a quadrant defined by the boundaries of 2 green lines are included in the analysis of that particular quadrant. The operator can adjust the position of these quadrant boundaries by clicking on the "quadrant" tab and dragging one of the lines in either direction. This adjustment can be useful, for example, to include the entire length of a vessel that starts nasally before curving temporally.
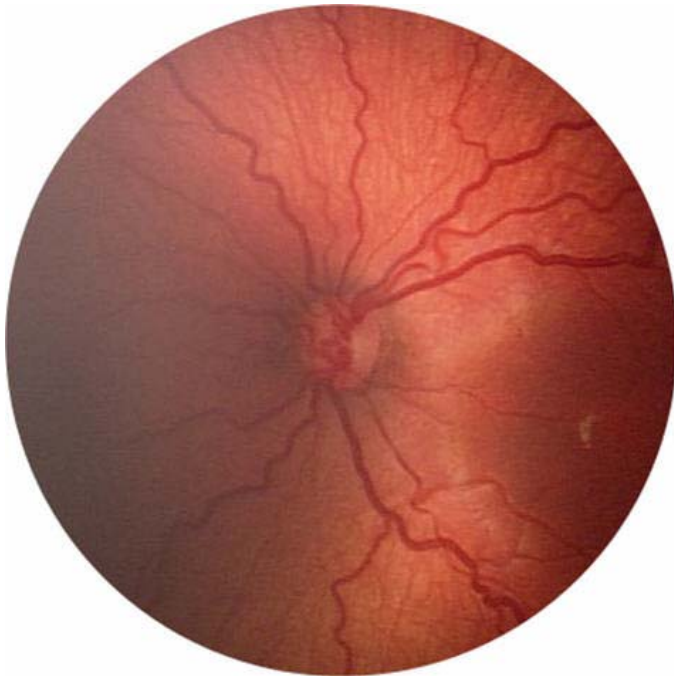

FIGURE 6

Cropped version of RetCam image of a premature infant's retina approximating the view obtained by a 28 D lens with the optic nerve centered.
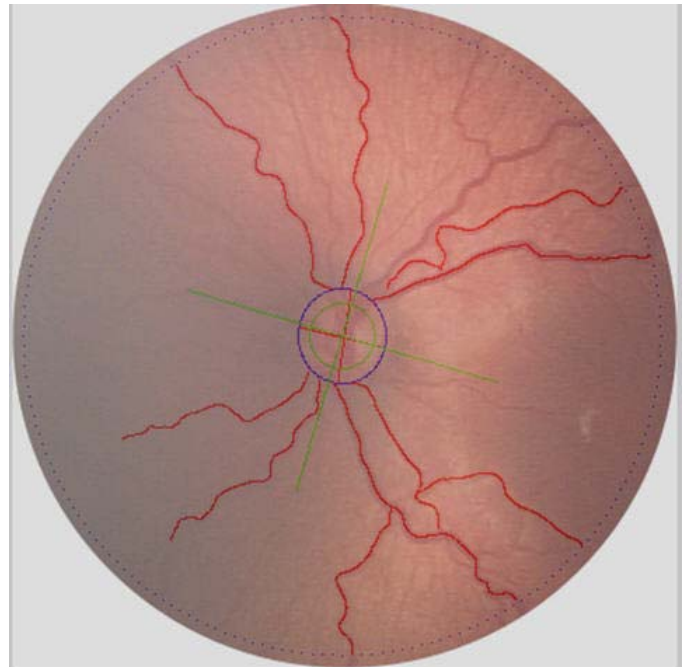

FIGURE 7

Appearance of a retinal image after analysis by the computer program ROPtool. Solid blue circle is optic nerve border; dotted blue circle is 28 D circle; green lines are borders of quadrants; red lines are traced vessels whose tortuosity has been calculated.

Next, the operator clicks on the major blood vessels (arterioles or venules) one at a time, and ROPtool automatically traces the selected blood vessel in 2 directions. The tracing of the vessel ends when (1) it encounters the optic nerve border or the 28D circle, (2) it begins to follow the course of a vessel that has already been traced, or (3) it no longer finds a vessel to trace. The vessel extraction will continue through an intersection with another traced vessel. As the vessel is traced, ROPtool automatically calculates vessel

tortuosity, which is a ratio of the total length of the vessel divided by a smooth curve generated from points spaced 200 points (about 40 pixels) apart on the vessel (Figure 4). *Tortuosity index* is defined as the vessel tortuosity divided by the average tortuosity of the standard photograph of plus disease[5] multiplied by 10. Therefore, a vessel with a tortuosity index of 15 has 50% more tortuosity that the average tortuosity of the major blood vessels in the standard photograph. The program also calculates dilation and dilation index, but this feature of the program has not been fully developed and is not the subject of this investigation. If ROPtool traces the wrong vessel, then the operator can click on "Undo Extract" to cancel the most recent tracing. Sometimes, 2 segments of the same vessel are traced separately, because ROPtool does not recognize them as part of the same vessel. In these cases, the "Merge" function can be used to fuse the 2 vessels into one, generating a new tortuosity value for the merged vessel that is an average weighted by the length of each segment. If ROPtool partially traces a vessel but then inaccurately continues to trace away from it, then the "Cut" function can be used to splice a vessel into 2 parts, allowing the inaccurate segment to be deleted. (In this study, the operator attempted to identify and click on the major arteriole and venule in each of the 4 quadrants, tracing a total of 8 vessels for each image. Tortuosity of both arterioles and venules was usually included, primarily because it was not always possible to accurately differentiate between arterioles and venules. Sometimes, fewer or more vessels were included, although at least 1 major vessel was always traced in each quadrant.)

Once all of the major vessels are identified, the operator clicks on "Compute," and ROPtool calculates the mean tortuosity of each quadrant and determines whether eye-level tortuosity is sufficient for plus disease, based on two quadrants meeting or exceeding a numeric threshold for tortuosity sufficient for plus disease. (For this study, the plus disease threshold was set a priori at a tortuosity index of 9, because it was more accurate in pilot studies than the default value of 10 that would correspond with the average tortuosity of the vessels in the standard photograph of plus disease.) Values for each quadrant and for each vessel can then be reviewed individually. Finally, all of the values can be transferred to a database such as Microsoft Excel by clicking on "Save DB".

PRIMARY ANALYSIS

The primary hypothesis was that computer-assisted quantification (by ROPtool) of posterior pole vascular tortuosity in premature infants was more accurate than assessment of tortuosity by an individual examiner. Accuracy was determined by measuring the concordance of ROPtool and the concordance of individual examiners to a reference standard. The reference standard was based on expert consensus of whether there was eye-level tortuosity sufficient for plus disease or not.

Sensitivity, specificity, and concordance (overall accuracy) were calculated in comparison to the reference standard (1) for the assessment of tortuosity sufficient for plus disease by ROPtool and (2) for 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 to the concordance of ROPtool. This hypothesis was tested 3 times, once for each examiner in comparison to ROPtool. Sensitivities and specificities of ROPtool were also compared to that 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. In these analyses, the accuracy of ROPtool in determining whether tortuosity was sufficient for *pre-plus* disease was assessed. An eye was considered to have tortuosity sufficient for pre-plus disease if at least 2 quadrants had pre-plus tortuosity or if 1 quadrant had plus tortuosity and 1 or more quadrants had pre-plus tortuosity. Pre-plus disease was not adjudicated; instead, expert consensus was considered to be the majority opinion of the 3 experts. The threshold for tortuosity sufficient for pre-plus disease used by ROPtool was determined a priori to be a tortuosity index of 7 or greater.

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 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 values 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 in order 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. 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 above and below the threshold of 9 used for plus tortuosity for this study. For each threshold value, sensitivity and specificity in the assessment of tortuosity sufficient for plus disease was determined. This analysis was first done for eye-level data for tortuosity sufficient for plus disease, and it was repeated for eye-level data for pre-plus disease, for quadrant-level data for plus disease, and for quadrant-level data for pre-plus disease.

Interobserver Variability of Experts

Eye-level data was used to calculate disagreement between each of the 3 possible pairs of experts (expert 1 vs expert 2, expert 1 vs expert 3, and expert 2 vs expert 3) with regard to the presence of tortuosity sufficient for plus disease and tortuosity sufficient for pre-plus disease. The number of images for which any of the experts differed from the other 2 experts was also calculated for plus vs not plus and for pre-plus vs normal. Finally, the plus vs. not plus analysis was repeated using only those eyes judged to have plus or pre-plus by expert consensus.

STATISTICAL ANALYSES

A sample size calculation for the primary outcome was done a priori. It was determined that 190 eyes were necessary in order to have 80% power to detect a significant difference (alpha = .05) in examiner-expert consensus concordance of 85% vs ROPtool-expert consensus concordance of 92.5%. It was necessary for a statistician to derive a formula for sample size, in part because of the correlation resulting from ROPtool and the examiners grading the same set of images.

Data were collected and analyzed in part using Microsoft Office Excel 2003, and some statistical analyses were done using SAS 9.1 (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.[35] Mean tortuosities were compared using a *t* test. All *P* values were 2-sided.

RESULTS

DESCRIPTION OF IMAGES

One hundred ninety images were analyzed by the computer program (ROPtool) and graded by 3 experts and 3 examiners. Of these, 5 images (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. These 185 images came from 117 different patients—both eyes of 68 patients and 1 eye of 49 patients.

Initially, the 3 experts disagreed on the presence or absence of tortuosity sufficient for plus disease for 27 of the 185 analyzable images. After the outlier expert for each image reconsidered it, the number of disagreements decreased to 17. On the conference call, agreement of all 3 experts was ultimately reached on all images except one. For this single image, the judgment of tortuosity sufficient for plus disease was determined by majority opinion (2 of 3).

At the quadrant level, there were 190 images with 4 quadrants each, or 760 total quadrants. Of these, 17 quadrants (2%) were excluded because of insufficient image quality, leaving a total of 743 units for the quadrant-level analyses. Three experts independently graded a total of 2259 quadrants, and of these, 313 quadrants (14%) were graded as tortuous enough for plus disease, 474 quadrants (21%) were graded as tortuous enough for pre-plus disease but not tortuous enough for plus disease, and 1472 quadrants (65%) were graded as normal.

The operator was able to successfully trace all of the images except one, and this image was one of the three excluded from the primary analysis because of insufficient image quality. The analysis using ROPtool took approximately 5 minutes per image. The better the image quality, the simpler and faster the analysis could be performed. Images of poorer quality due to image blur and/or those with less contrast between vessels and retina were more difficult and time-consuming to analyze because the "cut" and "delete" edit functions had to be used more frequently.

PRIMARY ANALYSIS

In comparison to 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 (Table 2). 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 disease was missed only once, resulting in a relatively high sensitivity (proportion of those with tortuosity sufficient for plus disease with a positive test) of 97%. This sensitivity compared favorably to the sensitivities of examiner 1 (*P* < .001), examiner 2 (*P* < .001), and examiner 3 (*P* = .02). Specificity (proportion of those without tortuosity sufficient for plus disease with a negative test) was clinically similar among ROPtool and the 3 examiners, although there was a statistically significant difference between ROPtool and two of the examiners (Table 2).

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. Table 3 shows the effect of using different numeric thresholds (or cutoff points) on 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 detailed in the ROC curve in Figure 8.

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 to expert consensus of eye-level tortuosity sufficient for plus or pre-plus disease, ROPtool had 161 agreements (87%) and 24 disagreements (13%) (Table 4). The concordance with expert consensus of tortuosity sufficient for plus or pre-plus disease for the 3 examiners was 93%, 91%, and 90%, respectively. Figure 9 is an ROC curve showing the effect of changing the numeric threshold for tortuosity sufficient for pre-plus disease on the sensitivity and specificity of ROPtool. It shows that 7 tortuosity units is a reasonably good threshold for tortuosity sufficient for pre-plus disease. Although sensitivity increases slightly at values below 7 tortuosity units, specificity drops off sharply (ie, there are many more false-positive results).

TABLE 1. NUMBER OF AGREEMENTS AND DISAGREEMENTS WITH REGARD TO THE PRESENCE OR ABSENCE OF PLUS DISEASE BETWEEN CONSENSUS OF 3 RETINOPATHY OF PREMATURITY EXPERTS AND THE COMPUTER PROGRAM (ROPtool) AND 3 PEDIATRIC OPHTHALMOLOGISTS

		EXPERT CONSENSUS	
		Plus	**Not Plus**
ROPtool	Plus	36	9
	Not Plus	1	139
Examiner 1	Plus	24	0
	Not Plus	13	148
Examiner 2	Plus	26	2
	Not Plus	11	146
Examiner 3	Plus	30	10
	Not Plus	7	138

TABLE 2. CONCORDANCE, SENSITIVITY, AND SPECIFICITY OF THE COMPUTER PROGRAM (ROPtool) AND 3 INDIVIDUAL EXAMINERS IN DETERMINING THE PRESENCE OR ABSENCE OF PLUS DISEASE IN COMPARISON TO CONSENSUS OF 3 EXPERTS*

	CONCORDANCE	SENSITIVITY	SPECIFICITY
ROPtool	175/185 (95%)	36/37 (97%)	139/148 (94%)
Examiner 1	172/185 (93%)	24/37 (65%)	148/148 (100%)
	$P = .5$	$P < .001$	$P = .01$
Examiner 2	172/185 (93%)	26/37 (70%)	146/148 (99%)
	$P = .5$	$P < .001$	$P = .03$
Examiner 3	168/185 (91%)	30/37 (81%)	138/148 (93%)
	$P = .1$	$P = .02$	$P = .8$

*P values are for the difference between each examiner and the computer program.

ANALYSES OF QUADRANT-LEVEL DATA

Since 3 experts graded each quadrant as 0 (normal), 1 (pre-plus), or 2 (plus), the means of these expert grades generated 7 categories (0.0, 0.3, 0.7, 1.0, 1.3, 1.7, and 2.0). These means were also grouped together based on the nearest integer into 3 categories (0.0-0.3, 0.7-1.3, and 1.7-2.0). Table 5 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 0.05 level except for 0.7 vs 1.0 and 1.0 vs 1.3. Figure 10 shows that 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. Median tortuosity values were 4.48 tortuosity units for normal quadrants, 9.17 tortuosity units for pre-plus tortuosity, and 17.45 tortuosity units for plus tortuosity.

Quadrant-level data were also used to generate ROC curves showing the effect of changing the numeric threshold for tortuosity sufficient for plus disease and for pre-plus disease on the sensitivity and specificity of ROPtool. The quadrant-level ROC curve for tortuosity sufficient for plus disease confirms that a tortuosity index of 9 is a good numeric threshold for plus tortuosity, with a sensitivity of 96% and specificity of 87% (Figure 11). The quadrant-level ROC curve for tortuosity sufficient for pre-plus disease

confirms that a tortuosity index of 7 is a good numeric threshold for pre-plus tortuosity, with a sensitivity of 96% and specificity of 87% (Figure 12).

TABLE 3. EFFECT OF USING DIFFERENT NUMERIC THRESHOLDS ("CUTOFF POINTS") ON THE AGREEMENT OR DISAGREEMENT OF THE COMPUTER PROGRAM (ROPtool) WITH EXPERT CONSENSUS*

		EXPERT CONSENSUS	
		Plus	**Not Plus**
ROPtool with cutoff of 8 tortuosity units	Plus	36	21
	Not Plus	1	127
ROPtool with cutoff of 9 tortuosity units (a priori)	Plus	36	9
	Not Plus	1	139
ROPtool with cutoff of 10 tortuosity units	Plus	35	6
	Not Plus	2	142
ROPtool with cutoff of 11 tortuosity units	Plus	33	4
	Not Plus	4	144
ROPtool with cutoff of 12 tortuosity units	Plus	30	4
	Not Plus	7	144
ROPtool with cutoff of 13 tortuosity units	Plus	28	2
	Not Plus	9	146
ROPtool with cutoff of 15 tortuosity units	Plus	30	0
	Not Plus	14	148

*Note that as the cutoff point increases, sensitivity decreases (more false-negatives) and specificity increases (fewer-false positives).
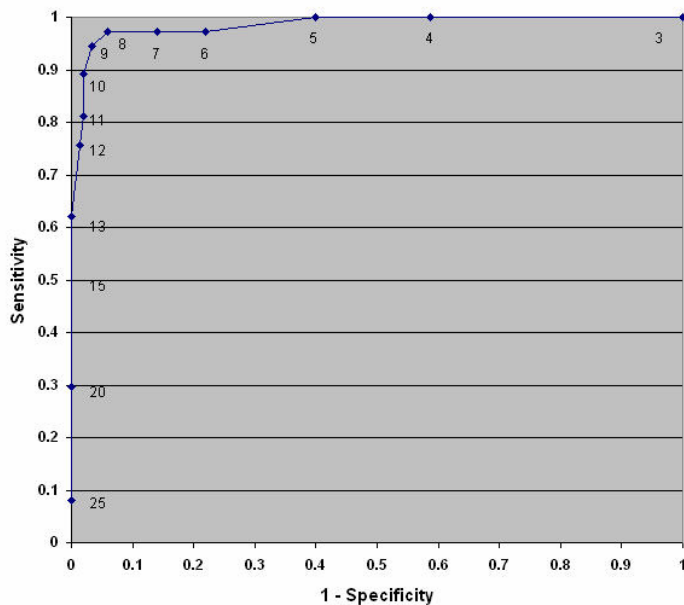

FIGURE 8

Receiver operating characteristic curve showing sensitivity vs 1-specificity (false-positive rate) for the detection of tortuosity sufficient for *plus* disease in retinopathy of prematurity at the *eye* level (unit = eye) by the computer program (ROPtool) in comparison to consensus of 3 experts.

TABLE 4. NUMBER OF AGREEMENTS AND DISAGREEMENTS WITH REGARD TO THE PRESENCE OR ABSENCE OF PLUS DISEASE OR PRE-PLUS DISEASE BETWEEN CONSENSUS OF 3 EXPERTS AND THE COMPUTER PROGRAM (ROPtool) AND 3 PEDIATRIC OPHTHALMOLOGISTS

		EXPERT CONSENSUS	
		PLUS OR PRE-PLUS	NORMAL
ROPtool	Plus or pre-plus	58	12
	Normal	12	103
Examiner 1	Plus or pre-plus	62	6
	Normal	8	109
Examiner 2	Plus or pre-plus	67	13
	Normal	3	102
Examiner 3	Plus or pre-plus	61	9
	Normal	9	106

INTEROBSERVER VARIABILITY OF EXPERT JUDGMENT

Even among experts, there existed differences in judgment of tortuosity sufficient for plus disease vs not plus disease at both the eye level and the quadrant level. When all 187 analyzable images were included (excluding only the 3 poor-quality images), there was disagreement among the 3 examiners for 27 or these images (14%) (Table 6). When normal images were excluded, leaving only those images judged by the experts to have pre-plus tortuosity or worse, the 3 experts disagreed on the presence of tortuosity sufficient for plus disease for 26 of 70 images (37%) (Table 7). For the assessment of tortuosity sufficient for plus or pre-plus disease vs normal, there was disagreement among the experts for 41 of 187 images (22%) (Table 8).
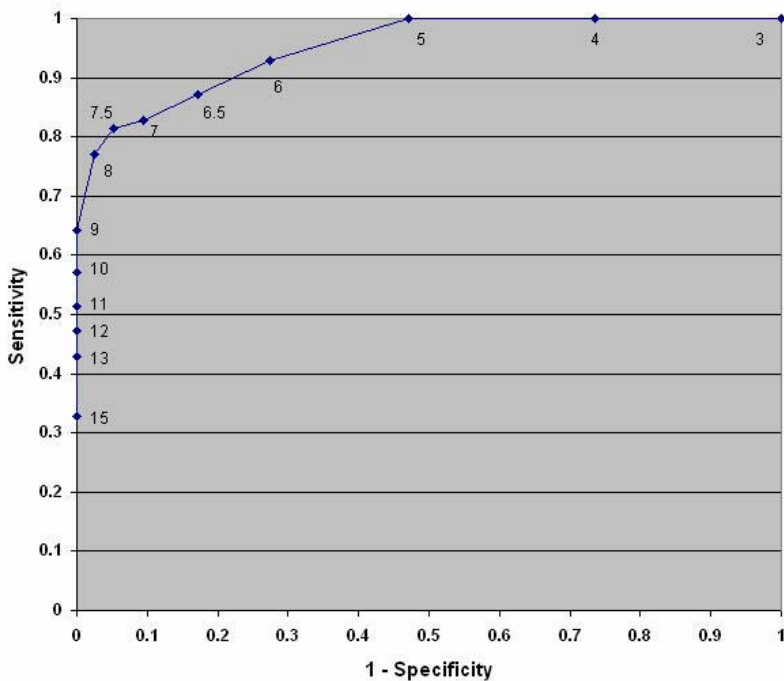

FIGURE 9

Receiver operating characteristic curve showing sensitivity vs 1-specificity (false-positive rate) for the detection of tortuosity sufficient for *plus* or *pre-plus* disease in retinopathy of prematurity at the *eye* level (unit = eye) by the computer program (ROPtool) in comparison to consensus of 3 experts.

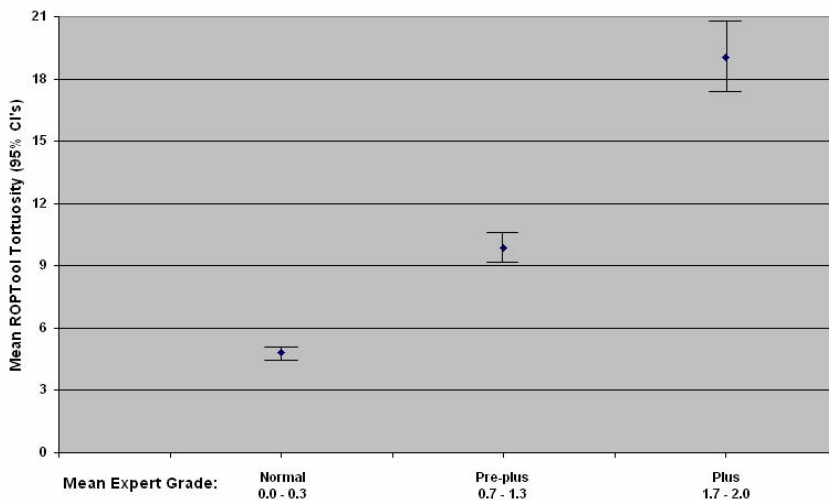| TABLE 5. MEAN QUADRANT TORTUOSITY MEASURED BY COMPUTER PROGRAM (ROPtool) FOR EACH LEVEL OF MEAN EXPERT GRADE		
MEAN EXPERT GRADE*	**MEAN QUADRANT TORTUOSITY (95% CI)**	***P* VALUE**
0.0	4.49 (4.34-4.64)	
		<.001
0.3	6.20 (5.84-6.56)	
		<.001
0.7	8.56 (7.59-9.52)	
		.2
1.0	9.67 (8.44-10.90)	
		.05
1.3	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)	
CI, confidence interval. *Three experts graded the tortuosity of each quadrant as 2 = plus disease, 1 = pre-plus disease, and 0 = normal.		

FIGURE 10

Mean quadrant tortuosity (and 95% confidence intervals [CI]) measured by the computer program (ROPtool) for 3 levels of mean expert grade. Three experts graded the tortuosity of each quadrant of images of premature infants' retinal blood vessels as 2 = plus disease, 1 = pre-plus disease, and 0 = normal.
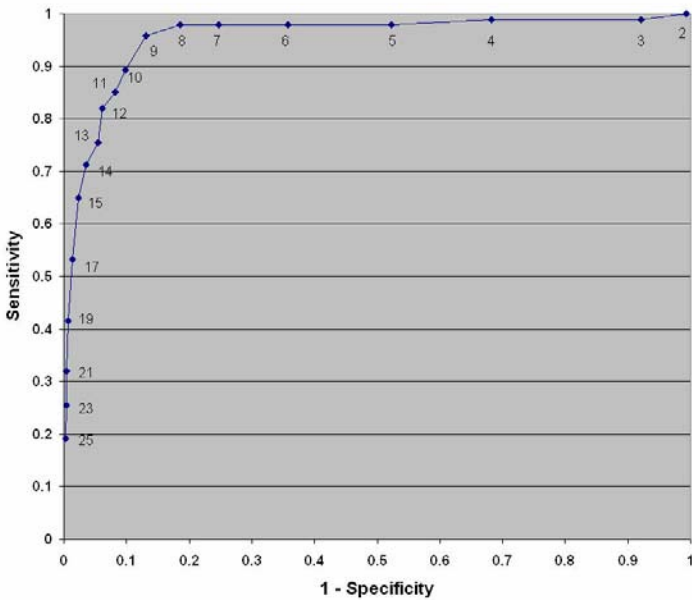
FIGURE 11

Receiver operating characteristic (ROC) curve showing sensitivity vs 1-specificity (false-positive rate) for the detection of tortuosity sufficient for *plus* disease in retinopathy of prematurity at the *quadrant* level (unit = quadrant) by the computer program (ROPtool) in comparison to consensus of 3 experts.
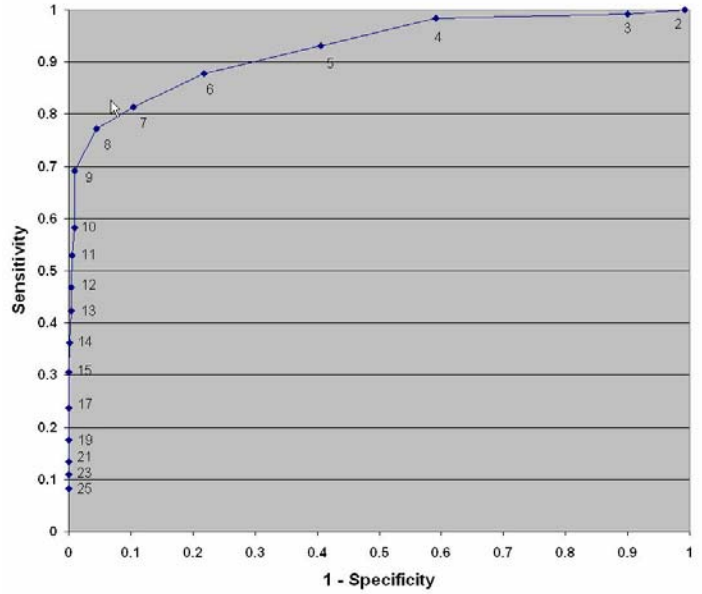

FIGURE 12

Receiver operating characteristic (ROC) curve showing sensitivity versus 1-specificity (false-positive rate) for the detection of tortuosity sufficient for *plus* or *pre-plus* disease in retinopathy of prematurity at the *quadrant* level (unit = quadrant) by the computer program (ROPtool) in comparison to consensus of 3 experts.

TABLE 6. AGREEMENT AND DISAGREEMENT BETWEEN 3 EXPERTS WITH REGARD TO EYE-LEVEL TORTUOSITY SUFFICIENT FOR PLUS DISEASE

		EXPERT 1		EXPERT 3	
		PLUS	NOT PLUS	PLUS	NOT PLUS
EXPERT 2	**PLUS**	25	0	15	10
	NOT PLUS	17	145	4	158
		Discordance = 17/187 (9%)		Discordance = 14/187 (7%)	
EXPERT 3	**PLUS**	19	0		
	NOT PLUS	23	145	Any disagreement among 3 experts: 27/187 (14%)	
		Discordance = 23/187 (12%)			

TABLE 7. AGREEMENT AND DISAGREEMENT BETWEEN 3 EXPERTS WITH REGARD TO EYE-LEVEL TORTUOSITY SUFFICIENT FOR PLUS DISEASE IN THOSE EYES WHERE EXPERT CONSENSUS WAS PLUS DISEASE OR PRE-PLUS DISEASE (EXCLUDING NORMAL EYES)

		EXPERT 1		EXPERT 3	
		PLUS	NOT PLUS	PLUS	NOT PLUS
EXPERT 2	PLUS	25	0	15	10
	NOT PLUS	16	29	4	41
		Discordance = 16/70 (23%)		Discordance = 14/70 (20%)	
EXPERT 3	PLUS	19	0		
	NOT PLUS	22	29	Any disagreement among 3 experts: 26/70 (37%)	
		Discordance = 22/70 (31%)			

TABLE 8. AGREEMENT AND DISAGREEMENT BETWEEN 3 EXPERTS WITH REGARD TO EYE-LEVEL TORTUOSITY SUFFICIENT FOR PRE-PLUS DISEASE

		EXPERT 1		EXPERT 3	
		PLUS OR PRE-PLUS	NORMAL	PLUS OR PRE-PLUS	NORMAL
EXPERT 2	PLUS OR PRE-PLUS	67	0	54	13
	NORMAL	26	94	5	116
		Discordance = 26/187 (14%)		Discordance = 18/187 (10%)	
EXPERT 3	PLUS OR PRE-PLUS	57	1		
	NORMAL	36	93	Any disagreement among 3 experts: 41/187 (22%)	
		Discordance = 37/187 (20%)			

DISCUSSION

ROPtool is a computer program that measures retinal vascular tortuosity and calculates whether this tortuosity is sufficient for plus disease or pre-plus disease. 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 to 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%). Therefore, the null hypothesis that ROPtool is no more accurate than individual examiners in assessing the presence or absence of tortuosity sufficient for plus disease could not be rejected. However, 97% sensitivity by ROPtool in detecting tortuosity sufficient for plus disease was superior to that of the individual examiners, who had lower sensitivities of 65%, 70%, and 81%. Specificities were clinically similar among ROPtool (94%) and the individual examiners (100%, 99%, and 93%), although there was a statistically significant difference between specificity of ROPtool and the 2 examiners who had specificities of 100% and 99%, respectively.

ROPtool also demonstrated the ability to discriminate between plus tortuosity, pre-plus tortuosity, and normal as shown by the significant difference in mean tortuosities between each of these categories (4.8, 9.9, and 19.1 tortuosity units, respectively).

The timely and accurate diagnosis of plus disease is extremely important. The ETROP study demonstrated a significant benefit to treatment of eyes with high-risk prethreshold disease.[12] 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.[27] Since the third criterion (zone I, stage 3 without plus disease) is relatively uncommon, it is clear that the primary indication for ROP 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. Therefore, failure to accurately detect plus disease may place an infant at increased risk for retinal detachment and blindness.

The finding in this study that the sensitivity of ROPtool was superior to 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 false-negative results, or very few cases in which plus disease exists and the diagnostic test fails to detect it. Plus disease has become the primary criterion for laser treatment, so it is important to miss as few cases of plus disease as possible in order to reduce the risk of delayed treatment and retinal detachment. High specificity, although desirable, is not as critical as high sensitivity in diagnosing plus disease. It is true that false-positive results could lead to unnecessary laser treatment. However, if one assumes that false-positive results 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. Although unnecessary treatment is not desirable and can result in reduced peripheral vision, the incidence of severe side effects from laser treatment is low when it is done by an experienced ophthalmologist. In the ETROP study, serious side effects of laser treatment, such as closure of the central retinal artery or inadvertent burns outside of the target zone, occurred in only 3 of 597 infants (0.5%).[12]

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. ROC curves are particularly valuable for visualizing this trade-off between sensitivity and specificity (Figures 8, 9, 11 and 12). 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.[5] After the pilot study, the threshold was adjusted to 9 tortuosity units and a pre-plus threshold of 7 tortuosity units was established. A threshold of 9 tortuosity units for tortuosity sufficient for plus disease was chosen because it had the best sensitivity without significantly compromising specificity. 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. 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.

At first glance, it is counterintuitive that the mean tortuosity of quadrants with pre-plus tortuosity was 9.9 tortuosity units, even though the numeric thresholds were 7 for pre-plus tortuosity and 9 for plus tortuosity. This occurred because the range of tortuosity measurements in the study was between 1.7 and 45.3 tortuosity units. Therefore, very low tortuosity measurements could be no more than 5.3 (7 minus 1.7) tortuosity units below the threshold for pre-plus tortuosity of 7 tortuosity units, whereas some very high tortuosity measurements were more than 15 tortuosity units above the threshold, skewing the mean value upward. In addition, 2 quadrants measuring 7 tortuosity units or greater were required for pre-plus tortuosity at the eye level. As a result, 1 quadrant with a very high tortuosity was insufficient for eye-level pre-plus tortuosity. For example, an eye with tortuosities in the 4 quadrants of 7.5, 8, 8.5, and 16 tortuosity units would have pre-plus tortuosity, because only 1 quadrant exceeds the plus tortuosity cutoff of 9 tortuosity units, even though the overall mean tortuosity of 10 tortuosity units exceeds this cutoff point.

ROPtool performed well in this study; however, the individual examiners also performed quite well, so the null hypothesis that the overall accuracy of ROPtool is no better than that of an examiner could not be rejected. Overall accuracy, or concordance, is greatly influenced by the prevalence of disease in the study population, whereas sensitivity and specificity depend much less on disease prevalence. This fact explains why sensitivity of ROPtool was superior to examiner's sensitivity, but concordance was so similar. Sixty-two percent of the images in the study were judged by experts to be normal, 18% had pre-plus tortuosity, and only 20% had plus tortuosity. If the study were to be repeated using a sample enhanced with a greater proportion of images with plus disease, and if similar sensitivities and specificities were found, then the overall concordance of ROPtool would likely be better than that of individual examiners. In addition, the proportion of disagreements in a study such as this one depends heavily on the number of "borderline" images included in the sample. If one wanted to make a new diagnostic test appear to be very accurate, using a sample enhanced with normals and obvious abnormals would likely give the best results. This is one reason why it was important to include comparison of the accuracy of ROPtool with that of individual examiners. In this study, the sample consisted of as many borderline images as possible. However, it is difficult to obtain a very large number of high-quality RetCam photos from infants with ROP, especially with this study's rigid criterion of using only one image from each eye.

For most calculations, the unit of analysis was a single image or eye (eye-level data). For these analyses, the overall grade of the eye was determined by the second most tortuous quadrant, because 2 quadrants of sufficient tortuosity were necessary in order for an eye to be given a grade of plus or pre-plus tortuosity. For example, if the grades of quadrants of an eye were 0, 0, 1, and 2, then the eye-level grade was 1 (pre-plus tortuosity), because 2 quadrants had a grade of 1 or more (and the second most tortuous quadrant had a

grade of 1). Additional analyses were done using each quadrant as the unit of analysis, or 4 units from each image. This quadrant-level approach could be used because the experts also graded each quadrant individually. This method has the advantage of increasing the number of observations by a factor of 4. It is particularly useful for analysis techniques such as ROC curves when establishment of the best numeric threshold for plus disease or for pre-plus disease is limited by the relatively small number of individual observations with values between 2 consecutive cutoff points. One limitation to this approach is that there is correlation between tortuosity of vessels in the same eye or image, so quadrant-level observations are not completely independent of each other.

In addition to the primary analysis involving plus disease, this study looked at the ability of ROPtool 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 to only 10 disagreements for plus tortuosity). There are several possible reasons for this observation. First, 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. Second, there is no "standard photo" defining the minimum amount of tortuosity for pre-plus disease, so experts may not be well calibrated for diagnosing pre-plus disease. For the above reasons, expert consensus for tortuosity sufficient for pre-plus may not have been as accurate as that for plus disease. Third, a few images were slightly blurry or relatively magnified, which can lead to an abnormally high tortuosity measure by ROPtool as its vessel-tracing program follows a choppy course through large or slightly blurry blood vessels, leading to overestimation of tortuosity and false-positive assessments of tortuosity sufficient for pre-plus disease.

In addition to the advantages of quantifying plus disease, it is desirable to accurately measure pre-plus disease for several reasons. First, the presence of pre-plus disease could influence the timing of the next examination. Current ROP screening guidelines state that infants without pre-threshold disease can be examined at 2-week intervals.[27] Type 2 ROP is defined as any zone I, stage 1 or 2 without plus disease, or zone II, stage 3 without plus disease, and it is generally an indication for close follow-up.[12] However, there are some cases in which pre-plus is present without type 2 ROP, and in these cases, pre-plus may be the first sign of impending disease progression. If pre-plus disease is present but not noted by the examiner, then a routine follow-up examination of 2 weeks would likely be recommended. If the eye progresses rapidly, then an opportunity to treat at the optimal time could be missed, resulting in an increased risk of an unfavorable outcome. Second, ophthalmologists are often asked by neonatologists to predict the likelihood of progression to severe ROP before an infant is transferred to another hospital or discharged home. The presence of pre-plus could alert physicians and other caretakers that an infant should not be transferred to another hospital unless the receiving hospital staff is able to provide follow-up and treatment. Third, the presence or absence of pre-plus disease may aid in counseling parents of premature infants regarding the risk of disease progression. Frequently, parents are understandably anxious to know whether their infants' eyes are "getting worse" and whether laser treatment will be necessary.

Even among experts, there is a tremendous amount of interobserver variability in the assessment of tortuosity sufficient for plus disease. When excluding normal images, there was disagreement among the 3 experts in this study for 26 (37%) of the 70 images. This observation agrees with the findings of Freedman and associates[36] that 3 CRYO-ROP certified experts disagreed on the presence or absence of plus disease in 29 of 72 cases (40%). Why does so much variability exist? Heretofore, there has been little attention paid to identifying the factors that influence an individual's assessment of plus disease, so there are numerous unanswered questions. 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") vs more gradual curves in longer vessel segments? How does the amount and degree of dilation influence the assessment of tortuosity? How does the total number of vessels influence this judgment? Interestingly, status as an ROP expert or as an experienced examiner of infants with ROP does not necessarily improve one's ability to accurately assess tortuosity or to discriminate between various degrees of dilation and tortuosity. Freedman and associates[36] found in a separate study that naïve and expert subjects performed indistinguishably when ranking groups of computer-generated retinal vascular tracings in order of increasing mean diameter and tortuosity. The large amount of interobserver variability in the assessment of plus disease suggests that even ROP experts are frequently inaccurate in their assessment of plus disease at the bedside. Therefore, a computer-automated system such as ROPtool has the potential to improve diagnostic accuracy, particularly in cases of "borderline" plus disease.

Various methods have been used to measure the tortuosity of retinal blood vessels. The original version of ROPtool used the simplest approach, dividing the total length of the vessels by the straight-line distance between its starting and ending points. A major problem with this approach was that a nontortuous vessel that gradually curved in one direction was falsely measured as tortuous. In order to avoid this problem, ROPtool generates a smooth curve using several points along the vessel, and it compares the total length of the vessel to the length of this smooth curve (Figure 4). Capowski and associates[37] developed a novel numeric index based on spatial frequency, and they compared its accuracy in assessing tortuosity of retinal blood vessels to the standard ratio of vessel length to straight-line distance. Their numeric index was calculated by selecting all possible segments of a vessel with lengths that varied from 780 to 1200 micrometers. The tortuosity of each segment was determined using the standard ratio of vessel length to straight-line distance. The overall vessel tortuosity was then calculated as the average of the 5 highest values whose segments did not overlap. They found that this numeric index was a more sensitive indicator of tortuosity than the standard ratio of vessel length to straight-line distance. Gelman and associates[38] quantified tortuosity using a different method, which incorporated the concept of "integrated curvature." They converted segmented vessels into a skeleton of 1-pixel-width trees, and they calculated integrated curvature as the sum of the angles along the "skeleton," normalized by the length of the vessel.

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[36-40] (Capowski, ARVO Meeting, 1994, Abstract; Kylstra JA, ARVO Meeting, 1995, Abstract). Freedman, Kylstra, and Capowski[36,37] were the first to use computer technology to quantify vascular dilation and tortuosity in ROP. Heneghan and associates[39] utilized 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[40] used Retinal Image multiScale Analysis (RISA), a semiautomated 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[38] applied an enhanced version of RISA that utilized 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, diameter, and tortuosity than those without plus disease. The best diagnostic accuracy was observed with integrated curvature, as curvature had an area under the ROC curve that was greater than both dilation and tortuosity for both arterioles and venules.

In this study, the ability of ROPtool 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. The dilation measurement used by ROPtool is still in development, and dilation is more difficult than tortuosity to accurately quantify. This is primarily because the degree of change from a normal to an abnormally dilated vessel is relatively small, as opposed to the relatively large change from a straight to a tortuous vessel. In addition, detection of 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 automated measure of dilation is perfected, it is possible that it will add very little to the overall assessment of plus disease. Kylstra and associates 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 to evaluation by an expert (Kylstra JA, ARVO Meeting, 1995. Abstract). Their computer-assisted model differed from ROPtool in that they used the computer's mouse to mark a series of points along each vessel, and they calculated vessel tortuosity as the length of the vessel divided by the straight-line distance from its origin to its termination. Another reason why dilation may not add much to the assessment of plus disease is that abnormal dilation probably precedes tortuosity in most cases; therefore, it is the development of a threshold level of tortuosity that usually triggers the diagnosis of plus disease.[15,25] Once tortuosity is known, inclusion of 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 the author's 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 (Figure 2) that was first used to calibrate ROPtool, both the arterioles and the venules are tortuous. Third, both arteriolar and venular tortuosities were quantified in the pilot study of ROPtool, 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 or not, but it seems likely that both dilation and tortuosity of all vessels are considered. In future studies, it would be interesting to measure examiners' abilities to differentiate arterioles from venules. If these blood vessels can be differentiated with reasonable accuracy, then new cutoff points could be calculated and the accuracy of ROPtool in determining tortuosity sufficient for plus disease when considering arterioles only could be determined.

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 plus tortuosity. 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 an accurate assessment of tortuosity does not necessarily translate into an accurate assessment of the presence or absence of plus disease. More work needs to be done to develop the dilation measurement of ROPtool or to show that dilation is unimportant or less meaningful than tortuosity in the overall assessment of plus disease. Third, 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. Fourth, a few images were difficult for the graders and/or ROPtool to assess owing to decentering or slight blurring of the image. Although this may have affected the results, it may make them more generalizable, since any future clinical application of ROPtool will likely include excellent as well as suboptimal images.

In addition to improving the assessment of dilation, much work remains to improve and refine ROPtool before it can augment or replace human judgment of plus disease. Analysis of a large number of images with different vessel shapes and patterns will allow the program to "learn" and will allow adjustment of the appropriate numeric thresholds for tortuosity sufficient for plus disease and pre-plus disease. Other techniques for measuring dilation or tortuosity, such as the "spatial frequency" method of Capowski and associates[37] or the "integrated curvature" measurement of Gelman and associates,[38] may be shown to improve accuracy, and if so, they could be incorporated into future versions of ROPtool. Also, ROPtool can be used to analyze high-quality RetCam images, but it is unknown how it will perform with other types of images that may be easier and less expensive to capture, such as those from a Nidek camera or still images captured during video indirect ophthalmoscopy.

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 probably 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, utilizing the internet to transmit images to ROP reading centers. Several studies have established the feasibility of remote screening for ROP.[26,41-45] Analysis of transmitted images could include a computer-assisted calculation of its tortuosity with or without dilation. One approach could be to use expert opinion and ROPtool, 2 diagnostic tests with different degrees of sensitivity and specificity, as either simultaneous or sequential tests. The precise degree of tortuosity with or without dilation 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,[25] and the ability to quantify the precise amount of vascular abnormality could influence follow-up intervals. For example, assuming a threshold for tortuosity sufficient for plus disease of 9 tortuosity units, an examiner would likely follow very closely an infant whose tortuosity increases from 5 to 7 tortuosity units in 1 week. 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. One can envision a clinical trial in which an image of an infant with severe ROP is transmitted to a reading center where experts confirm that treatment is needed, assisted by computerized assessment of plus disease. Finally, computer-assisted analysis of plus disease may eventually have a role at the bedside. Still images captured from video indirect ophthalmoscopy could potentially 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 remove subjectivity in the assessment of tortuosity sufficient for plus disease, thereby increasing accuracy and optimizing the timing of laser treatment. Its greatest potential application in the near future is in research and telemedicine. 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.

ACKNOWLEDGMENTS

REFERENCES

1. Good WV, Gendron RL. Retinopathy of prematurity. *Ophthalmol Clin North Am* 2001;14:513-519.
2. Quinn GE, Dobson V, Barr CC, et al. Visual acuity of eyes after vitrectomy for retinopathy of prematurity: follow-up at 5 1/2 years. The Cryotherapy for Retinopathy of Prematurity Cooperative Group. *Ophthalmology* 1996;103:595-600.
3. Javitt J, Dei Cas R, Chiang YP. Cost-effectiveness of screening and cryotherapy for threshold retinopathy of prematurity. *Pediatrics* 1993;91:859-866.
4. Brown GC, Brown MM, Sharma S, et al. Cost-effectiveness of therapy for threshold ROP. *Pediatrics* 1999;104:e47.
5. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity: preliminary results. *Arch Ophthalmol* 1988;106:471-479.
6. Cryotherapy for Retinopathy of Prematurity Cooperative Group. 15-year outcomes following threshold retinopathy of prematurity. *Arch Ophthalmol* 2005;123:311-318.
7. The Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Arch Ophthalmol* 1984;102:1130-1134.
8. An International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity Revisited. *Arch Ophthalmol* 2005;123:991-999.
9. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Effect of retinal ablative therapy for threshold retinopathy of prematurity: results of Goldmann perimetry at the age of 10 years. *Arch Ophthalmol* 2001;119:1120-1125.
10. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Contrast sensitivity at age 10 years in children who had threshold retinopathy of prematurity. *Arch Ophthalmol* 2001;119:1129-1133.

11. Good WV, Hardy RJ, ETROP Multicenter Study Group. The multicenter study of early treatment for retinopathy of prematurity (ETROP) [editorial]. *Ophthalmology* 2001;108:1013-1014.
12. 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.
13. 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.
14. Schaffer DB, Palmer EA, Plotsky DF, et al, for the Cryotherapy for Retinopathy of Prematurity Cooperative Group. Prognostic factors in the natural course of retinopathy of prematurity. *Ophthalmology* 1993;100:230-237.
15. Saunders RA, Bluestein EC, Sinatra RB, et al. The predictive value of posterior pole vessels in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus* 1995;32:82-85.
16. Flynn JT, O'Grady GE, Herrera J, et al. Retrolental fibroplasia I. Clinical observations. *Arch Ophthalmol* 1977;95:217-223.
17. Fielder AR, Shaw DE, Robinson J, Ng YK. Natural history of retinopathy of prematurity: a prospective study. *Eye* 1992;6:233-242.
18. Holland DR, Saunders RA, Kagemann LE, et al. Color doppler imaging of the central retinal artery in premature infants undergoing examination for retinopathy of prematurity. *J AAPOS* 1999;3:194-198.
19. Capone A Jr, Ells AL, Fielder AR, Flynn JT, et al. Standard image of plus disease in retinopathy of prematurity. *Arch Ophthalmol* 2006;124:1669-1670.
20. Christiansen SP, Bradford JD. Cataract in infants treated with argon laser photocoagulation for threshold retinopathy of prematurity. *Am J Ophthalmol* 1995;119:175-180.
21. Paysse EA, Miller A, Brady McCreery KM, Coats DK. Acquired cataracts after diode laser photocoagulation for threshold retinopathy of prematurity. *Ophthalmology* 2002;109:1662-1665.
22. Simons BD, Wilson MC, Hertle RW, Schaefer DB. Bilateral hyphemas and cataracts after diode laser retinal photoablation for retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus* 1998;35:185-187.
23. Trigler L, Weaver RG Jr, O'Neil JW, Barondes MJ, Freedman SF. Case series of angle-closure glaucoma after laser treatment for retinopathy of prematurity. *J AAPOS* 2005;9:17-21.
24. Cryotherapy for Retinopathy of Prematurity Cooperative Group. The natural ocular outcome of premature birth and retinopathy. *Arch Ophthalmol* 1994;112:903-912.
25. 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.
26. 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.
27. 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 [erratum in *Pediatrics* 2006;118:1324].
28. Aylward SR, Bullitt E. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Trans Med Imaging* 2002;21:61-75.
29. Bullitt E, Aylward SR. Volume rendering of segmented image objects. *IEEE Trans Med Imaging* 2002;21:998-1002.
30. Bullitt E, Gerig G, Pizer SM, Lin W, Aylward SR. Measuring tortuosity of the intracerebral vasculature from MRA images. *IEEE Trans Med Imaging* 2003;22:1163-1171.
31. Bullitt E, Muller KE, Jung I, Lin W, Aylward S. Analyzing attributes of vessel populations. *Med Image Anal* 2005;9:39-49.
32. Wallace DK, Jomier J, Aylward SR, Landers III MB. Computer automated quantification of plus disease in retinopathy of prematurity. *J AAPOS* 2003;7:126-130.
33. 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.
34. Wallace DK, Zhao Z, Freedman SF. A pilot study using "ROPtool" to quantify plus disease in retinopathy of prematurity. *J AAPOS* epub May 25, 2007.
35. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13-22.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.

41. Schwartz SD, Harrison SA, Ferrone PJ, Trese MT. Telemedical evaluation and management of retinopathy of prematurity using a fiberoptic digital fundus camera. *Ophthalmology* 2000;107:25-28.
42. Yen KG, Hess D, Burke B, Johnson RA, et al. The optimum time to employ telephotoscreening to detect retinopathy of prematurity. *Trans Am Ophthalmol Soc* 2000;98:145-150.
43. Roth DB, Morales D, Feuer WJ, et al. Screening for retinopathy of prematurity employing the RetCam 120: sensitivity and specificity. *Arch Ophthalmol* 2001;119:268-272.
44. Chiang MF, Keenan JD, Starren J, et al. Accuracy and reliability of remote retinopathy of prematurity diagnosis. *Arch Ophthalmol* 2006;124:322-327.
45. Chiang MF, Starren J, Du YE, et al. Remote image based retinopathy of prematurity diagnosis: a receiver operating characteristic analysis of accuracy. *Br J Ophthalmol* 2006;90:1292-1296