Remote image based retinopathy of prematurity diagnosis: A receiver operating characteristic analysis of accuracy

10 authors, including:

Michael F Chiang
Oregon Health and Science University
188 PUBLICATIONS **2,918** CITATIONS

SEE PROFILE

William Schiff
Lenox Hill Hospital
47 PUBLICATIONS **1,305** CITATIONS

SEE PROFILE

Ditte Hess
University of Miami
47 PUBLICATIONS **705** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Retinopathy of prematurity View project

EXTENDED REPORT

Remote image based retinopathy of prematurity diagnosis: a receiver operating characteristic analysis of accuracy

M F Chiang, J Starren, Y E Du, J D Keenan, W M Schiff, G R Barile, J Li, R A Johnson, D J Hess, J T Flynn

....

See end of article for authors' affiliations
....

Correspondence to:
Michael F Chiang, MD, Department of Ophthalmology, Columbia University College of Physicians and Surgeons, 635 West 165th Street, Box 92, New York, NY 10032, USA; chiang@dbmi.columbia.edu

Background/aims: Telemedicine offers potential to improve the accessibility and quality of diagnosis of retinopathy of prematurity (ROP). The aim of this study was to measure accuracy of remote image based ROP diagnosis by three readers using receiver operating characteristic (ROC) analysis.

Methods: 64 hospitalised infants who met ROP examination criteria underwent two consecutive bedside procedures: dilated examination by an experienced paediatric ophthalmologist and digital retinal imaging with a commercially available wide angle camera. 410 images from 163 eyes were reviewed independently by three trained ophthalmologist readers, who classified each eye into one of four categories: no ROP, mild ROP, type 2 prethreshold ROP, or ROP requiring treatment. Sensitivity and specificity for detection of mild or worse ROP, type 2 prethreshold or worse ROP, and ROP requiring treatment were determined, compared to a reference standard of dilated ophthalmoscopy. ROC curves were generated by calculating values for each reader at three diagnostic cut-off levels: mild or worse ROP (that is, reader was asked whether image sets represented mild or worse ROP), type 2 prethreshold or worse ROP (that is, reader was asked whether image sets represented type 2 prethreshold or worse ROP), and ROP requiring treatment.

Results: Areas under ROC curves ranged from 0.747–0.896 for detection of mild or worse ROP, 0.905–0.946 for detection of type 2 prethreshold or worse ROP, and 0.941–0.968 for detection of ROP requiring treatment.

Conclusions: Remote interpretation is highly accurate among multiple readers for the detection of ROP requiring treatment, but less so for detection of mild or worse ROP.

Telemedicine is an emerging technology involving computer based transmission of patient data, with subsequent interpretation by a remote expert.[1–3] Retinopathy of prematurity (ROP) is an ideal disease for application of telemedicine systems because existing diagnostic methods require frequent, time consuming, and logistically difficult infant examinations at the neonatal intensive care unit (NICU) bedside. Adequate ophthalmic expertise is often unavailable at the point of care because of training limitations, geographical constraints, and medicolegal concerns. Despite recent treatment advances, ROP continues to be a leading cause of childhood blindness throughout the world.[4] These factors suggest that telemedicine has potential to improve the quality and accessibility of ROP care.

In addition, ROP is a convenient disease for evaluating the efficacy of telemedicine strategies. Established guidelines have defined the population of high risk infants requiring examination,[5] and provided a universal standard for disease description based on retinal appearance.[6] ROP is treatable if diagnosed early, and the multicentre cryotherapy for ROP (CRYO-ROP) and early treatment for ROP (ETROP) trials have established specific criteria to identify disease requiring treatment.[7 8]

Previous studies have determined the sensitivity and specificity of remote image based ROP diagnosis at various diagnostic levels, such as the presence of any ROP, the presence of "referral warranted" ROP, or the presence of ROP requiring treatment.[9–14] However, sensitivity and specificity vary depending on the diagnostic cut-offs used to define "normal" and "abnormal" results. Little published research has examined the relation between referral cut-off and the accuracy of remote ROP diagnosis.[11 12] This is an important gap in knowledge that must be understood if telemedicine is to become an accepted ROP examination strategy.

Receiver operating characteristic (ROC) analysis is an ideal method for addressing this question. An ROC curve plots sensitivity against false positive ratio (FPR, or 1 – specificity), in which each point reflects values obtained at a different diagnostic cut-off value.[15 16] ROC methods provide several advantages over traditional sensitivity and specificity measurements. (1) Sensitivity and specificity are characteristics of a diagnostic test at a particular cut-off value, whereas an ROC curve is characteristic of the test itself. (2) The trade-off between sensitivity and specificity can be visualised on an ROC plot as the diagnostic cut-off value is shifted. (3) ROC curves allow quick comparison of the discriminative ability of different diagnostic tests to separate "normal" from "abnormal" results. In particular, area under an ROC curve is an established measure of accuracy for a diagnostic test. The area has a value from 0.0 to 1.0, where 1.0 represents complete ability to discriminate between "normal" and "abnormal" states, 0.0 represents complete inability to discriminate (that is, every patient with an "abnormal" test is actually "normal"), and 0.5 represents pure chance.[17 18]

Our recent studies have examined the remote interpretation by multiple trained readers of ROP images captured by a commercially available wide angle digital fundus contact camera.[11 12] This paper will extend that work by introducing ROC analysis to examine the impact of cut-off level on diagnostic accuracy.

....

Abbreviations: FPR, false positive ratio; NICU, neonatal intensive care unit; ROC, receiver operating characteristic; ROP, retinopathy of prematurity; SE, standard error
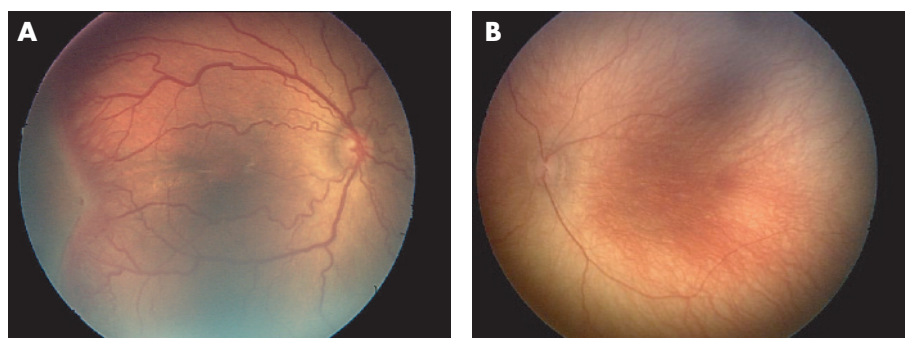
Figure 1 Examples of retinal atlas photographs. (A) Was interpreted by reference standard ophthalmoscopic examination and by all readers as ROP requiring treatment, and (B) was interpreted as no ROP.

METHODS

This study was approved by the institutional review boards at Columbia University Medical Center and the University of Miami School of Medicine.

Examination and development of retinal atlas

An ROP image atlas developed by the authors was used for this study, as described previously.[11] Briefly, infants at Jackson Memorial Hospital who met ROP examination criteria from 1999–2000, and whose parents provided informed consent for imaging, were included. Patients were excluded if they had structural ocular or systemic anomalies, if they had previously received treatment for ROP, or if they were considered by their neonatologist to be systemically unstable for ocular examination.

Each infant underwent two sequential examinations under topical anaesthesia at the NICU bedside. Firstly, dilated indirect ophthalmoscopy with scleral depression was performed by an experienced paediatric ophthalmologist (JTF).[5] Presence or absence of ROP disease was documented according to the international classification of ROP. Subsequently, wide angle imaging was performed independently by an experienced ophthalmic photographer (DJH) using a digital system (RetCam-120; Clarity Medical Systems, Pleasanton, CA, USA) with a 120° retinal lens. This was done with the aim of imaging as much of the retina as feasible. Because imaging was performed under typical NICU bedside conditions, it was not possible to capture a standard set of photographs for each eye. Images were compiled into an atlas by an ophthalmologist masked to examination findings, and were not annotated with any individually identifiable data or other descriptive information. No patients were excluded because of poor image quality or inability to obtain photographs.

Interpretation of images

Three board certified image readers participated in this study: two retina specialists (readers A and B), and one general ophthalmologist (reader C). Readers were trained to analyse remote images by an experienced paediatric ophthalmologist (JTF). Masked readers independently reviewed each image set in the atlas, and interpretations for each eye were converted to an ordinal scale based on established CRYO-ROP and ETROP criteria[7 8]: (1) no ROP; (2) mild ROP, defined as disease less than type 2 prethreshold; (3) type 2 prethreshold ROP (zone I, stage 1 or 2 ROP without plus disease; or zone II, stage 3 ROP without plus disease); and (4) ROP requiring treatment, defined as type 1 prethreshold disease (zone I, any stage ROP with plus disease; zone I, stage 3 ROP without plus disease; or zone II, stage 2 or 3 ROP with plus disease) or threshold disease (at least five contiguous or eight non-continuous clock hours of stage 3 ROP in zone I or II in the presence of plus disease).

Data analysis

Data were analysed by eye, using statistical software (SPSS 13.0; SPSS Inc, Chicago, IL, USA). Accuracy of image interpretation was determined, compared to reference standard ophthalmoscopy, for three situations: (1) detection of any ROP (that is, mild or worse ROP based on reference standard examination was considered ''abnormal''); (2) detection of type 2 prethreshold or worse ROP (that is, type 2 prethreshold or ROP requiring treatment based on reference standard examination were considered ''abnormal''); and (3) detection of ROP requiring treatment. In each of these situations, the sensitivity and specificity of each reader for every image set was determined at three diagnostic cut-off levels for interpretation: (1) cut-off of mild or worse ROP (that is, reader was asked whether image sets represented mild or worse ROP); (2) cut-off of type 2 prethreshold or worse ROP (that is, reader was asked whether image sets represented type 2 prethreshold or worse ROP); and (3) cut-off of ROP requiring treatment.

McNemar's test was used for comparison of sensitivities and specificities between pairs of raters, for subgroups of positive and negative ROP by the reference standard at each cut-off value. Results were displayed on ROC curves, and areas under the curves were determined.
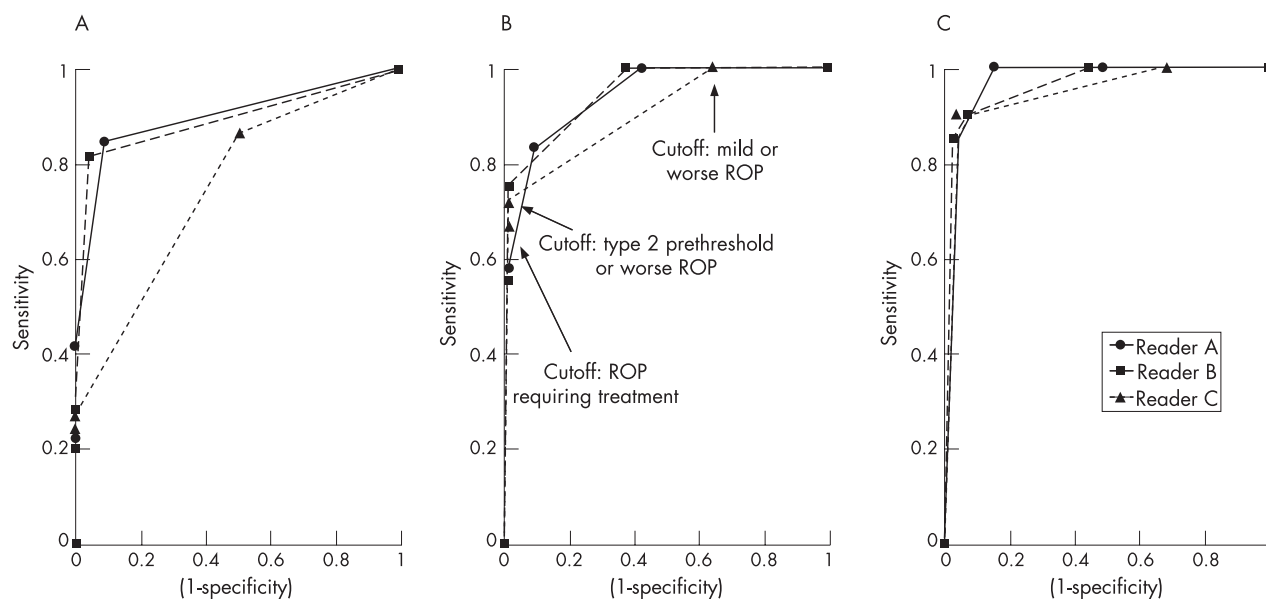
Table 1 Sensitivity and specificity for remote detection of mild or worse ROP, by three trained readers at three diagnostic cut-off levels*

Reader	Mild or worse ROP		Type 2 prethreshold or worse		ROP requiring treatment	
	Sensitivity	Specificity†	Sensitivity‡	Specificity	Sensitivity	Specificity
A	0.845 (0.036)	0.910 (0.035)	0.417 (0.074)	1.000 (−)	0.223 (0.086)	1.000 (−)
B	0.816 (0.038)	0.955 (0.025)	0.282 (0.083)	1.000 (−)	0.204 (0.086)	1.000 (−)
C	0.864 (0.034)	0.493 (0.061)	0.272 (0.083)	1.000 (−)	0.243 (0.085)	1.000 (−)

*Results are shown at diagnostic cut-offs of mild or worse ROP (that is, readers were asked whether image sets showed mild or worse ROP), type 2 prethreshold or worse ROP (that is, readers were asked whether image sets showed type 2 prethreshold or worse ROP), and ROP requiring treatment (that is, readers were asked whether image sets showed ROP requiring treatment). Results are reported as proportion (standard error, SE).
†Differences in specificity at diagnostic cut-off of mild or worse ROP are statistically significant for reader pairs A v C and B v C (p<0.001).
‡Differences in sensitivity at diagnostic cut-off of type 2 prethreshold or worse ROP are statistically significant for reader pairs A v B and A v C (p<0.001).

Figure 2 Receiver operating characteristic (ROC) curves for (A) detection of mild ROP or worse ROP, (B) detection of type 2 prethreshold or worse ROP, and (C) detection of ROP requiring treatment from remote image interpretation by three readers. Points on ROC curves display sensitivity and specificity for each reader at three diagnostic cut-offs: mild or worse ROP, type 2 prethrehsold or worse ROP, and ROP requiring treatment (shown in (B) for interpretation by reader C).

RESULTS

Study population

Results of reference standard examination classifications, along with study population characteristics, have been reported previously.[11] Mean birth weight of infants was 812 g (range 480–1440 g), and mean gestational age was 26 weeks (range 23–32 weeks). The retinal image atlas consisted of 163 unique sets of digital images (81 right eyes, 82 left eyes) from 64 consecutive infants. Sixty four (39.3%) image sets had reference standard evaluations showing no ROP, 65 (39.9%) had mild ROP, 16 (9.8%) had type 2 prethreshold ROP, and 18 (11.0%) had ROP requiring treatment. Each image set consisted of one to seven photographs from a single eye, taken at 32–45 weeks' postmenstrual age. Example images are shown in figure 1. No corneal injuries, infections, or other ocular complications occurred during imaging.

Detection of mild ROP

Table 1 shows the sensitivity and specificity for each image reader, based on a reference standard classification in which "abnormal" is considered to be presence of mild or worse ROP. When image readers used a diagnostic cut-off of mild ROP (that is, readers were asked whether image sets represented mild or worse ROP), sensitivity of remote diagnosis for each reader was at least 81%, and specificity for readers A and B was at least 91%. However, when using a diagnostic cutoff of mild ROP, the specificity of remote diagnosis by reader C was significantly worse (49.3%, p<0.001).

As the diagnostic cut-off level for image interpretation was shifted to type 2 prethreshold ROP (that is, readers were asked whether image sets represented type 2 prethreshold or worse ROP) and ROP requiring treatment, the sensitivity for each reader decreased (that is, more false negative referrals) and specificity increased to 100% (that is, no false positive referrals). Sensitivity for detection of mild or worse ROP, using a diagnostic cut-off of type 2 prethreshold or worse ROP, was significantly higher for reader A than for readers B or C (p<0.001). Figure 2A displays these results as ROC curves. Areas under these curves ranged from 0.747 (reader C) to 0.896 (reader A) (table 2).

Detection of type 2 prethreshold ROP

Table 3 shows sensitivity and specificity based on a reference standard classification in which "abnormal" is considered to be presence of type 2 prethreshold or worse ROP. When image readers used a diagnostic cut-off of type 2 prethreshold ROP (that is, readers were asked whether image sets represented type 2 prethreshold or worse ROP), sensitivity of remote diagnosis for each reader was at least 72% and specificity was at least 90%. The specificity for detection of type 2 prethreshold or worse ROP, using a diagnostic cut-off of type 2 prethreshold or worse ROP, was significantly lower for reader A than for reader B (p = 0.003) or reader C (p = 0.001).

As this diagnostic cut-off level for image interpretation was shifted to mild ROP (that is, readers were asked whether image sets represented mild or worse ROP), the sensitivity increased to 100% but specificity decreased. The specificity for

Table 2 Areas under receiver operating characteristic (ROC) curves for detection of mild or worse ROP, type 2 prethreshold or worse ROP, and ROP requiring treatment by three readers*

Reader	Mild or worse ROP	Type 2 prethreshold or worse ROP	ROP requiring treatment
A	0.896 (0.025)	0.938 (0.019)	0.968 (0.013)
B	0.892 (0.026)	0.946 (0.019)	0.960 (0.020)
C	0.747 (0.037)	0.905 (0.031)	0.941 (0.029)

*Results are reported as proportion (SE).

Table 3 Sensitivity and specificity for remote detection of type 2 prethreshold or worse ROP, by three trained readers at three diagnostic cut-off levels*

Reader	Mild or worse ROP		Type 2 prethreshold or worse		ROP requiring treatment	
	Sensitivity	Specificity†	Sensitivity	Specificity‡	Sensitivity	Specificity
A	1.000 (−)	0.575 (0.055)	0.833 (0.062)	0.903 (0.026)	0.583 (0.106)	0.985 (0.011)
B	1.000 (−)	0.619 (0.033)	0.750 (0.072)	0.985 (0.022)	0.556 (0.109)	0.993 (0.007)
C	1.000 (−)	0.351 (0.069)	0.722 (0.075)	0.985 (0.011)	0.667 (0.094)	0.993 (0.007)

*Results are shown at diagnostic cut-offs of mild or worse ROP (that is, readers were asked whether image sets showed mild or worse ROP), type 2 prethreshold or worse ROP (that is, readers were asked whether image sets showed type 2 prethreshold or worse ROP), and ROP requiring treatment (that is, readers were asked whether image sets showed ROP requiring treatment). Results are reported as proportion (SE).
†Differences in specificity at diagnostic cut-off of mild or worse ROP are statistically significant for reader pairs A v C and B v C (p<0.001).
‡Difference in specificity at diagnostic cut-off of type 2 prethreshold or worse ROP are statistically significant for reader pairs A v B (p = 0.003) and A v C (p = 0.001).

detection of type 2 prethreshold or worse ROP, using a diagnostic cut-off of mild or worse ROP, was significantly lower for reader C than for readers A or B (p<0.001). In contrast, when the diagnostic cut-off level of image interpretation was shifted to ROP requiring treatment, the sensitivity decreased but specificity increased. Figure 2B displays these results as ROC curves. Areas under these curves ranged from 0.905 (reader C) to 0.946 (reader B) (table 2).

Detection of ROP requiring treatment
Table 4 shows sensitivity and specificity based on a reference standard classification in which "abnormal" is considered to be presence of ROP requiring treatment. When readers used a diagnostic cut-off of ROP requiring treatment (that is, readers were asked whether image sets represented ROP requiring treatment), sensitivity of remote diagnosis for each reader was at least 85% and specificity was at least 95%.

As the diagnostic cut-off level for image interpretation was shifted to type 2 prethreshold ROP (that is, readers were asked whether image sets represented type 2 prethreshold or worse ROP) and mild ROP, the sensitivity of each reader increased to 100% (that is, no false negative referrals) but specificity decreased (that is, more false positive referrals). Specificity for detection of ROP requiring treatment, using a diagnostic cut-off of type 2 prethreshold or worse ROP, was significantly lower for reader A than for reader B (p = 0.002) or reader C (p<0.001). Figure 2C displays these results as ROC curves. Areas under these curves ranged from 0.941 (reader C) to 0.968 (reader A) (table 2).

DISCUSSION
This study evaluates the accuracy of remote image based ROP diagnosis by three masked readers, compared to a reference standard of dilated ophthalmoscopy. ROC analysis was used to examine the ability of each reader to detect any ROP, type

2 prethreshold ROP, and ROP requiring treatment as a function of diagnostic cut-off level. ROC curves display values for all sensitivity/specificity pairs from continuously varying the decision cut-off level used by image readers, and are therefore a convenient tool for performing detailed analysis of readers' ability to discriminate between "normal" and "abnormal" images.

The accuracy of remote diagnosis varied based on two primary factors. Firstly, readers A and B had higher specificity than reader C in several situations (for example, table 1). However, consistency among the three readers was higher for detection of ROP requiring treatment than for detection of mild ROP (fig 2). Secondly, the accuracy by each reader was highest for the detection of ROP requiring treatment, and lowest for the detection of mild or worse ROP. This may be best demonstrated by comparing the three sets of ROC curves (fig 2 and table 2). The area under an ROC curve has several equivalent interpretations: the average value of specificity at all possible values of sensitivity, the average value of sensitivity at all possible values of specificity, or the probability that a randomly chosen subject from the "abnormal" group has a test value of higher severity than a randomly chosen subject from the "normal" group.[17] [18] Areas under ROC curves ranged from 0.747–0.896 for detection of mild or worse ROP to 0.941–0.968 for detection of ROP requiring treatment, indicating near perfect discriminative ability for the latter. Taken together, these findings suggest that remote image interpretation using commercially available devices performs best for the discrimination between ROP that does and does not require treatment, and least well for the discrimination between presence and absence of ROP (table 2).

The finding that accuracy of remote ROP detection is dependent on diagnostic cut-off value is consistent with results from previous studies.[10–14] This discrepancy is not surprising, given that mild ROP is often subtle in appearance,

Table 4 Sensitivity and specificity for remote detection of ROP requiring treatment, by three trained readers at three diagnostic cut-off levels*

Reader	Mild or worse ROP		Type 2 prethreshold or worse		ROP requiring treatment	
	Sensitivity	Specificity†	Sensitivity	Specificity‡	Sensitivity	Specificity
A	1.000 (−)	0.513 (0.056)	1.000 (−)	0.846 (0.032)	0.850 (0.080)	0.960 (0.016)
B	1.000 (−)	0.553 (0.054)	0.900 (0.070)	0.926 (0.022)	0.850 (0.080)	0.973 (0.013)
C	1.000 (−)	0.313 (0.067)	0.900 (0.070)	0.933 (0.021)	0.900 (0.022)	0.953 (0.017)

*Results are shown at diagnostic cut-offs of mild or worse ROP (that is, readers were asked whether image sets showed mild or worse ROP), type 2 prethreshold or worse ROP (that is, readers were asked whether image sets showed type 2 prethreshold or worse ROP), and ROP requiring treatment (that is, readers were asked whether image sets showed ROP requiring treatment). Results are reported as proportion (SE).
†Differences in specificity at diagnostic cutoff of mild or worse ROP are statically significant for reader pairs A v C and B v C (p<0.001).
‡Difference in specificity at diagnostic cut-off of type 2 prethreshold or worse ROP are statistically significant for reader pairs A v B (p = 0.002) and A v C (p<0.001).

located in the peripheral retina, and therefore more difficult to capture photographically. In addition, newer ETROP guidelines for ROP requiring treatment are often based on presence of plus disease, or presence of stage 3 disease in zone I, without the need for precisely measuring the extent of peripheral disease.[8] It could be argued that higher accuracy and reliability for detection of ROP requiring treatment might be expected because it is technically easier to obtain posterior pole images than multiple peripheral images. However, additional studies are needed to determine whether posterior pole images are sufficient for the diagnosis of severe disease. In particular, increased ocular pressure from a contact camera, such as the one used in this study, could reduce blood flow and thereby mask the appearance of plus disease. For example, the image in figure 1A taken alone might be considered by some observers to represent ''pre-plus'' disease, although the reference standard examiner in this study diagnosed plus disease and ROP requiring treatment.

In general, ROC curves are a simple method for displaying the accuracy of a diagnostic test. These curves are independent of disease prevalence and permit quick comparison among multiple tests on a normalised axis. Use of ROC analysis has been widespread in other image based specialties such as radiology.[18] Although application of these methodologies in ophthalmology has been relatively limited, it may be particularly useful for evaluation of new technologies such as imaging devices and other diagnostic modalities.[19 20]

This study design has several limitations. (1) Retinal photographs were not annotated with any clinical data. Lack of access to these data may have biased against remote interpretation by readers. (2) The number of readers in this study was small. Readers did not have extensive clinical experience in ROP examination or treatment, but two of the three were retina specialists. Although our results suggest higher accuracy and consistency among readers for detection of ROP requiring treatment compared to detection of mild or worse ROP, further studies may be required to examine the generalisability and extrapolation to other readers such as ROP experts or trained technicians working under ophthalmologist supervision. (3) All images were obtained by a single, experienced ophthalmic photographer. Future studies to determine whether successful image capture is generalisable among photographers would be useful.

This study demonstrates that remote image based ROP screening is technically feasible using existing technologies, based on the extremely high ability to detect type 2 prethreshold and ROP requiring treatment. However, it is not clear whether the most appropriate cut-off defining clinically relevant disease requiring ophthalmic referral and patient transfer in a real world setting should be ROP requiring treatment, type 2 prethreshold ROP, or some other level. Before large scale implementation of telemedicine strategies, studies must be performed to standardise the image capture process, and examine cost-benefit trade-offs of unnecessary over-referrals or failure to detect clinically significant disease. If these issues are resolved, telemedicine will offer the potential to improve workflow for ophthalmologists and neonatologists, and to provide accessibility to the highest standard of care for patients.

. . . .

Authors' affiliations
M F Chiang, J D Keenan, W M Schiff, G R Barile, J Li, J T Flynn, Department of Ophthalmology, Columbia University College of Physicians and Surgeons, New York, NY, USA

M F Chiang, J Starren, Department of Biomedical Informatics, Columbia University College of Physicians and Surgeons, New York, NY, USA
J Starren, Department of Radiology, Columbia University College of Physicians and Surgeons, New York, NY, USA
Y E Du, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA
R A Johnson, Department of Pediatric Nursing, Jackson Memorial Hospital, Miami, FL, USA
D J Hess, Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA

REFERENCES

1 **Grigsby J**, Sanders JH. Telemedicine: where it is and where it's going. *Ann Intern Med* 1998;**129**:123–7.
2 **Bashshur RL**, Reardon TF, Shannon GW. Telemedicine: a new health care delivery system. *Annu Rev Public Health* 2000;**21**:613–37.
3 **Field MJ**, ed. *Telemedicine: a guide to assessing telecommunications in health care.* Washington, DC: National Academy Press, 1996.
4 **Gilbert C**, Foster A. Childhood blindness in the context of Vision 2020: the right to sight. *Bull World Health Organ* 2001;**79**:227–32.
5 **Fierson WM**, Palmer EA, Petersen RA, *et al.* Screening examination of premature infants for retinopathy of prematurity. *Pediatrics* 2001;**108**:809–11.
6 **Committee for the Classification of Retinopathy Of Prematurity**. An international classification of retinopathy of prematurity. *Arch Ophthalmol* 1984;**102**:1130–4.
7 **Cryotherapy for Retinopathy of Prematurity Cooperative Group**. Multicenter trial of cryotherapy for retinopathy of prematurity: preliminary results. *Arch Ophthalmol* 1988;**106**:471–9.
8 **Early Treatment for Retinopathy of Prematurity Cooperative Group**. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. *Arch Ophthalmol* 2003;**121**:1684–94.
9 **Schwartz SD**, Harrison SA, Ferrone PJ, *et al.* Telemedical evaluation and management of retinopathy of prematurity using a fiberoptic digital fundus camera. *Ophthalmology* 2000;**107**:25–8.
10 **Ells AL**, Holmes JM, Astle WF, *et al.* Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology* 2003;**110**:2113–17.
11 **Chiang MF**, Keenan JD, Starren J, *et al.* Accuracy and reliability of remote retinopathy of prematurity diagnosis. *Arch Ophthalmol* 2006;**124**:322–7.
12 **Chiang MF**, Keenan JD, Du YE, *et al.* Assessment of image-based technology: impact of referral cutoff on accuracy and reliability of remote retinopathy of prematurity diagnosis. *Proc AMIA Symp* 2005:126–30.
13 **Roth DB**, Morales D, Feuer WJ, *et al.* Screening for retinopathy of prematurity employing the RetCam 120: sensitivity and specificity. *Arch Ophthalmol* 2001;**119**:268–72.
14 **Yen KG**, Hess D, Burke B, *et al.* Telephotoscreening to detect retinopathy of prematurity: preliminary study of the optimum time to employ digital fundus camera imaging to detect ROP. *J AAPOS* 2002;**6**:64–70.
15 **Lusted LB**. Signal detectability and medical decision-making. *Science* 1971;**171**:1217–9.
16 **Swets JA**. Measuring the accuracy of diagnostic systems. *Science* 1988;**240**:1285–93.
17 **Zweig MH**, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;**39**:561–77.
18 **Obuchowski NA**. Receiver operating characteristic curves and their use in radiology. *Radiology* 2003;**229**:3–8.
19 **Lachkar Y**, Cohn H. Sensitivity and specificity of optic disc variables and analysis of a new variable (MP/D) for glaucoma diagnosis with the Glaucoma-Scope. *Br J Ophthalmol* 1997;**81**:846–51.
20 **Wadood AC**, Azuara-Blanco A, Aspinall P, *et al.* Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, and Humphrey Swedish Interactive Threshold Algorithm-fast perimetry in a glaucoma practice. *Am J Ophthalmol* 2002;**133**:327–32.