Published in final edited form as:

J Pediatr Ophthalmol Strabismus. 2018 July 01; 55(4): 245–253. doi:10.3928/01913913-20180327-04.

Real-world simulation of an alternative retinopathy of prematurity screening system in Thailand: a pilot study

S. Grace Prakalapakorn, MD, MPH¹, Sharon F. Freedman, MD¹, Amy K. Hutchinson, MD², Piyada Saehout, BA³, Mine Cetinkaya-Rundel, PhD⁴, David K. Wallace, MD, MPH¹, and Kittisak Kulvichit, MD, MSc³

¹Department of Ophthalmology, Duke University, Durham, NC, USA ²Department of Ophthalmology, Emory University, Atlanta, GA, USA ³Department of Ophthalmology, Chulalongkorn University, Bangkok, Thailand ⁴Department of Statistical Science, Duke University, Durham, NC, USA

Abstract

Purpose: The purpose of this study was to evaluate an alternative ROP screening system that identifies infants meriting examination by an ophthalmologist in a middle-income country. We hypothesized that grading posterior pole images for the presence of pre-plus or plus disease has high sensitivity to identify infants with treatment-warranted (type 1) ROP. To test this hypothesis, we evaluated the (1) feasibility of having a non-ophthalmologist health care worker (HCW) obtain retinal images of prematurely-born infants using Pictor™ (a non-contact retinal camera) that were of sufficient quality to grade for pre-plus or plus disease and (2) accuracy of grading these images to identify infants with type 1 ROP.

Methods: We prospectively recruited infants at Chulalongkorn University Hospital (Bangkok, Thailand). On days infants underwent routine ROP screening, a trained HCW imaged their retinas with Pictor™. Two ROP experts graded these serial images from a remote location for image gradability and posterior pole disease.

Results: Fifty-six infants were included. Overall, 69.4% of infant imaging sessions were gradable. Amongst gradable images, the sensitivity of both graders for identifying an infant with type 1 ROP by grading for the presence of pre-plus or plus disease was 1.0 (95% CI: 0.31–1.0) for grader 1 and 1.0 (95% CI: 0.40–1.0) for grader 2 and the specificity was 0.93 (95% CI: 0.76–0.99) for grader 1 and 0.74 (95% CI: 0.53–0.88) for grader 2.

Conclusions: It was feasible for a trained non-ophthalmologist HCW to obtain retinal images of infants using Pictor™ that were of sufficient quality to identify infants with type 1 ROP.

INTRODUCTION

Retinopathy of prematurity (ROP) is a significant cause of blindness in many middle-income countries,^{1,2} where rates of treatment-requiring ROP are high.² While appropriate screening and treatment could reduce blindness due to ROP, there are important barriers to receiving this care, especially in the developing world. Barriers include a shortage of ophthalmologists trained to screen for ROP and poor access to them.^{3,4} While the use of digital retinal images for remote ROP screening has been explored to address these barriers, most studies have focused on using large, expensive wide-field contact retinal cameras (i.e. RetCam (Clarity Medical Systems, Pleasanton, CA) ~\$100K).⁵⁻⁷

In order to increase access to care and sustain ROP screening programs in middle-income countries, alternative approaches should be portable and affordable. Thus, we investigated a Food and Drug Administration approved, commercially available, portable, non-contact digital fundus camera, Pictor ((Volk Optical Inc., Mentor, OH) ~\$10K), for its potential as an ROP screening tool in a developing country. This camera weighs 450g and can be transported in a small bag. A previous study found that a pediatric ophthalmologist could use Pictor to obtain retinal images of prematurely-born infants that were of sufficient quality for ROP experts to accurately grade for pre-plus or plus disease compared to the clinical exam findings.⁸

The main objective of this prospective study was to evaluate an alternative screening system for ROP to identify infants meriting binocular indirect ophthalmoscopy by an ophthalmologist in a middle-income country. We hypothesized that grading posterior pole images for the presence of pre-plus or plus disease has high sensitivity to identify infants with type 1 ROP. To test this hypothesis, we evaluated (1) the feasibility of having a non-ophthalmologist health care worker (HCW) obtain retinal images of prematurely-born infants using a non-contact retinal camera that were of sufficient quality to grade for pre-plus or plus disease and (2) the accuracy of ROP experts to grade these images for pre-plus or plus disease using a real-world simulation to identify infants with type 1 ROP as diagnosed on clinical exam.

PATIENTS AND METHODS

This prospective study was approved by the Institutional Review Boards at both the Faculty of Medicine, Chulalongkorn University and the Duke Health System. It was carried out in compliance with the International guidelines for human research protection as per the Declaration of Helsinki and the International Conference on Harmonization in Good Clinical Practice and conformed to the requirements of the United States Health Insurance Portability and Privacy Act.

Before recruiting study participants, we trained an ophthalmic photographer (PS) without prior neonatal photography experience to use Pictor. Training consisted of: (1) reviewing the manual; (2) obtaining Pictor images of undilated adult volunteers under supervision during two one-hour training sessions; (3) independently practicing imaging of undilated adult

volunteers; and (4) imaging study participants with Pictor for three sessions under supervision.

We recruited participants from all infants undergoing routine ROP screening examinations from July 2013–August 2015 in the Neonatal Intensive Care Unit (NICU) and the outpatient pediatric retina clinic at Chulalongkorn University Hospital, in Bangkok, Thailand. All clinical eye examinations were performed by a fellowship-trained medical and surgical retina specialist (KK) with 25 years experience in ROP screening and treatment. Infants were eligible for this study if they had a birth weight (BW) ≥ 1500 grams or a gestational age (GA) ≥ 31 weeks. After informed consent was obtained, the HCW used Pictor to obtain dilated fundus photographs of participants on the same day they underwent routine ROP screening by the retinal specialist using indirect ophthalmoscopy and a 20-diopter condensing lens. The order of imaging and examination was determined by convenience and was not randomized. Pictor was set to capture and store the color and corresponding red-free image simultaneously when images were taken.

The goal of each imaging session was to obtain and submit focused retinal images of retinal vessels in all four quadrants of each eye of each participant. At each imaging session, the HCW imager held the infant's eyelids open with her fingers with or without the aid of another HCW or study staff member. No topical anesthetic or eyelid speculum was used during imaging. In the NICU, a nurse monitored infants for clinical stability. After each imaging session, the imager chose up to 3 pairs of images (i.e. the color and corresponding red-free image(s)) per eye and uploaded them in random order by infant into a web-based software, FocusROP (Trumbauersville, PA), along with documentation of the BW, GA and post-menstrual age (PMA) at the time of imaging (Figure 1), in order to simulate screening for ROP in a real-world scenario. Only images with a visible optic nerve were included, to provide orientation. Because of Pictor's 45° field of view and the movement of infants' eyes during imaging, the imager was allowed to submit up to 3 images per eye in order to achieve the imaging goal listed above and maximize the length of retinal vessels in all four quadrants captured in the retinal images.

Two pediatric ophthalmologists who have been involved in previous national ROP studies^{5,9} and who have had at least 20 years of experience in screening for and treating ROP (SFF and AKH) served as image graders. Neither grader had examined any of the study participants. Independently, they reviewed the uploaded images from the United States by remotely logging into the web-based software. To simulate a real-world telemedicine screening scenario, graders viewed images for each infant in chronological order (by increasing PMA) and were allowed to review only images submitted for previous weeks. The graders evaluated these images for: (1) Image quality, (2) number of gradable quadrants, and (3) posterior pole disease. Image quality was based on the grader's judgment of how well she could grade the dilation and tortuosity of the retinal vessels in the image(s). Image quality was defined as "good" when there was a clear view of the optic nerve and retinal vessels and the grader could easily determine vessel dilation and tortuosity; "fair" when it was difficult to distinguish either vessel dilation or tortuosity; and "poor" when the grader could determine neither vessel dilation nor tortuosity. The number of gradable quadrants (0–4) was determined by the number of quadrants where the grader could satisfactorily visualize 1 disc

diameter length of a major vessel. The posterior pole was graded as normal or having pre-plus or plus disease, based on ICROP definitions.^{10,11}

SAS 9.3 (SAS Institute Inc, Cary, NC) and R v.3.3.1 (<https://www.R-project.org>) were used for all statistical analysis.

For the feasibility portion of this study, we evaluated the number of imaging sessions in which submitted images were “gradable” for screening, defined as receiving a grade of fair or good image quality and having 3 or 4 gradable quadrants (Figure 2). Analysis was performed both by eye and by infant. If no images could be captured for an eye, that eye was considered “not gradable”. Images for both eyes had to be considered “gradable” for an infant’s images to be considered “gradable”.

For the accuracy portion of our analysis, we compared each expert grader’s image interpretations to the retina specialist’s clinical exam findings. For an imaging session to be included in this portion of the analysis, the infant had to have clinical examination results available (Figure 2). In order to include enough images with abnormal posterior pole findings for our analysis, we created an enhanced sample of images by choosing one imaging session per infant as follows: if an infant was treated for ROP, we chose the last imaging session prior to treatment. If an infant was not treated for ROP, we chose the date the infant had the most severe posterior pole disease according to the retina specialist’s clinical exam findings (plus>pre-plus>normal posterior pole). If there was more than one exam date with the most severe posterior pole disease, we chose the exam performed closest to PMA of 36 weeks.

Overall accuracy was calculated for each grader as the number of times the image grader’s determination of normal, pre-plus, or plus disease agreed with the retina specialist’s clinical examination findings of normal, pre-plus, or plus disease, respectively, divided by the total number of images graded. Analysis was performed both by eye and by infant. The infant was categorized according to the disease in the worse eye as determined by the retina specialist. To test intra-grader (test-retest) reliability for each grader, 20% of the study populations’ (n=10 infants) complete set of images (total of 40 imaging sessions) were selected and uploaded twice in random order into the online system for each grader to evaluate.

To simulate a real-world scenario, analysis of sensitivity and specificity was performed only by infant. A positive screening test was defined as an imaging session where an infant was graded as having pre-plus or plus disease in either eye by the telemedicine grader. Because it is critical in screening programs to identify those infants requiring treatment, i.e. infants with type 1 ROP, we selected the diagnosis of type 1 ROP (zone I with stage 3 or plus disease; or zone II, stage 2 or 3 with plus disease) on clinical examination by the retina specialist to be the “reference standard”. Two separate analyses of the data were performed as follows: (1) including “gradable” images only (i.e. images with fair or good image quality and 3 or 4 gradable quadrants) and (2) including all images with any “ungradable” image considered a positive screening test. We calculated confidence intervals using a modified version of Wilson’s score method by employing a continuity correction in order to align the minimum

coverage probability (rather than the average) with the nominal value.^{12–14} To test inter-grader reliability, Cohen's unweighted Kappa statistic (κ) was used calculate how well the graders agreed with each other's evaluation of normal, pre-plus, or plus disease by eye.

RESULTS:

Fifty-six infants were included in this study (56 for the feasibility portion and 50 for the accuracy portion). Average GA was 28.1 weeks (range: 23–33) and BW was 1156 grams (range: 585–2000).

Analysis of feasibility

For the feasibility analysis, all 56 infants were included. A total of 215 clinical examinations were performed. On the same day as the clinical examinations, imaging was attempted for all but two infants: one was not attempted due to infant instability and one was discontinued due to episodes of oxygen desaturation. Submitted images were considered gradable by both graders for 86% of all eye imaging sessions and 79% of all infant imaging sessions (Table 1, Part 1: Feasibility Analysis). The HCW was unable to obtain an image of an eye in 4.7% of eye imaging sessions (Table 1, Part 1: Feasibility Analysis).

The imager reported no difficulty in imaging eyes 72% of the time and infants 67% of the time. Barriers to imaging eyes or infants included the following: movement of the eye or child (23% eyes, 28% infants); presence of a respiratory assist device including nasal cannula, continuous positive airway pressure mask, or intubation tube (5% eyes, 6% infants); difficulty opening eyelids (2% eyes or infants); and pupil size (2% eyes or infants).

Taking the average gradability of all imaging sessions by infant (i.e. the contribution of each infant was weighted equally), 78% of all eye imaging sessions and 69% of all infant imaging sessions were considered gradable (table 2). Gradability of images submitted ranged between 0–100% by PMA, with 1–21 infants being imaged at each specified PMA (Table 2).

Analysis of accuracy

For the accuracy portion of the study, 50 of the 56 infants (100 eyes) were included. Six infants were excluded for the following reasons: infant judged too unstable to attempt imaging (N=1), imaging did not commence until after treatment for ROP (N=3), or clinical examination findings were unavailable because the infants' clinical charts could not be located (N=2). A total of 191 infant imaging sessions (382 eye imaging sessions) qualified for inclusion in this portion of the study. These infant imaging sessions were categorized by the status of the posterior pole vessels in the eye with the worse disease as determined by the retina specialist's clinical examination: 7 had plus disease, 10 pre-plus, and 174 normal posterior pole vessels. Of the 382 eye imaging sessions, 13 had plus, 19 pre-plus and 350 normal posterior pole vessels.

Overall accuracy of each image grader's determination of normal, pre-plus, or plus disease in images compared to the retina specialist's respective clinical examination findings by eye was 90.9% for grader 1 and 71.2% for grader 2, and by infant was 90.9 % for grader 1 and 67.6% for grader 2 (Table 3). Among the 10 infants with a total of 40 imaging sessions

uploaded twice to assess intra-grader (test-retest) reliability, grader 1 was consistent with her own grading of the posterior pole vessels 99% of the time, and grader 2 was consistent 75% of the time.

For the analysis of sensitivity and specificity, we selected one imaging session for each infant. Average PMA at time of imaging was 37.5 weeks (range: 33–47). The graders felt that 64% of the selected infant imaging sessions were gradable (Table 1, Part 2: Accuracy Analysis). The imager was not able to obtain images of one or both eyes of an infant 10% of the time (Table 1, Part 2: Accuracy Analysis). Of these infant imaging sessions, categorized by the disease in the worse eye as determined by clinical examination, 7 had plus disease, 1 pre-plus, 42 normal posterior pole vessels, and 7 had type 1 ROP. Looking at each grader's ability to identify an infant with type 1 ROP by grading for the presence of pre-plus or plus disease, (1) amongst gradable images: sensitivity was 1.0 (95% CI: 0.31–1.0) for grader 1 and 1.0 (95% CI: 0.40–1.0) for grader 2 and the specificity was 0.93 (95% CI: 0.76–0.99) for grader 1 and 0.74 (95% CI: 0.53–0.88) for grader 2 and (2) including all images: sensitivity was 1.0 (95% CI: 0.56–1.0) for grader 1 and 1.0 (95% CI: 0.56–1.0) for grader 2 and the specificity was 0.79 (95% CI: 0.64–0.89) for grader 1 and 0.65 (95% CI: 0.49–0.79) for grader 2 (Table 4). Among all 100 eyes evaluated, the graders agreed with each other's evaluation of normal, pre-plus, or plus disease by eye in 80% of the uploaded images (unweighted $\kappa = 0.6$).

DISCUSSION:

In a middle-income country, it was feasible for a trained non-ophthalmologist HCW to obtain retinal images of infants using a non-contact retinal camera that were of sufficient quality to identify infants with type 1 ROP with high sensitivity using a real-world telemedicine screening simulation. While this proposed ROP screening system shows promise as a way to identify infants in need of an examination using indirect ophthalmoscopy by an ophthalmologist, there is a need to improve the ability of the HCW to consistently obtain gradable images.

While the trained HCW could obtain images with Pictor that experts felt were gradable to identify pre-plus and plus disease, there is opportunity for improvement. The greatest obstacle to obtaining images with Pictor was movement of the eye or infant. Swaddling the infant or using a pacifier, bottle, or oral sucrose to calm the infant could make imaging easier and more efficient. While the presence of a respiratory assist device made imaging difficult about 5% of the time, we did not find an association with any particular device, nor did we find a clear association between ability to image and PMA at time of imaging. The imager reported that opening the eyelids and small pupil size each interfered with imaging about 2% of the time. Examining dilation protocols could also improve pupil dilation to aid image capture using Pictor. In a previous study, one of the authors (SGP) trained herself to image prematurely-born infants using Pictor by reading the user's manual and practicing on undilated adults.⁸ While we trained our imager similarly and also provided supervised training sessions imaging adults and infants, increasing the amount of supervised training could increase an imager's ability to obtain images using Pictor.

Using the identification of pre-plus or plus disease on image grading as a screening test to identify infants who developed type 1 ROP, we found high sensitivity, and, as expected, lower specificity. When screening for ROP, it is important to detect every case of type 1 ROP (i.e. high sensitivity), even at the expense of a lower specificity (identifying an infant for further evaluation who turns out not to have type 1 ROP). While neither grader missed any infants who developed type 1 ROP, by including pre-plus as a positive screening test, we inherently increased the number of false positives and expected a lower specificity. When we included all images compared to just those considered gradable, false positives increased and specificity decreased since ungradable images were considered positive screening tests. Thus, in our screening paradigm, focusing on improving image gradability is paramount to a more efficient screening system.

Our results should be considered in light of several study limitations. First, our results may not be generalizable to other countries since ROP screening guidelines are region-specific.¹⁶ Second, there was a single imager in this study, with a background in ophthalmic photography. A study evaluating the ability to train HCWs without experience in ophthalmic imaging is currently underway. Third, while both graders have extensive experience screening and treating ROP, they were not given any specific training before grading these images and the variability of their grading underscores the need for standardized training for image grading, which has been shown to be effective in other ROP screening studies.^{15,17} Fourth, while using Pictor to image prematurely-born infants appeared safe in this preliminary study (i.e. imaging was stopped in <0.5% of attempted imaging sessions because of infant instability), a formal comparison of the safety of the standard indirect ophthalmoscopy exam to Pictor imaging was not performed. It has been reported that the binocular indirect ophthalmoscopy examination can cause pain and physiologic instability in prematurely-born infants. In particular, the use of a lid speculum during binocular indirect ophthalmoscopy can cause physiologic stress to the infant.¹⁵ An advantage of a non-contact retinal camera is that it is non-contact and neither a topical anesthetic nor a lid speculum is needed for imaging.

This pilot study showed that a trained HCW could capture gradable retinal images with a non-contact retinal camera and that these narrow-field images could be evaluated with high sensitivity to detect infants with type 1 ROP. While not ready to be implemented at this time, when image capture is improved, this screening strategy may serve as a safe and viable alternative to expand access to ROP screening in geographic locations underserved by ophthalmologists and thereby may bridge a gap in patient care (lack of manpower¹⁶ and accessibility) in developing middle-income countries.

Acknowledgments

Financial Conflict of interest: Dr. Prakalapakorn is supported by NIH K23EY024268. Drs. Freedman and Wallace have developed a technology (ROPtool) which is not being used in this study, but ROPtool has been purchased by FocusROP (which is being used in this study). The developers and Duke University may benefit financially from sales of FocusROP. This project was partially funded by an unrestricted grant from Research to Prevent Blindness to the Duke Eye Center. The funding organizations had no role in the design or conduct of this research or the decision to submit this report for publication.

REFERENCES:

1. Gilbert C, Rahi J, Eckstein M, O'Sullivan J, Foster A. Retinopathy of prematurity in middle-income countries. *Lancet* 1997;350:12–4. [PubMed: 9217713]
2. Gilbert C Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early human development* 2008;84:77–82. [PubMed: 18234457]
3. Kemper AR, Wallace DK. Neonatologists' practices and experiences in arranging retinopathy of prematurity screening services. *Pediatrics* 2007;120:527–31. [PubMed: 17766525]
4. Roach L, Francis B. ROP crisis near, survey says. *EyeNet* 2006.
5. Quinn GE, Ying GS, Daniel E, et al. Validity of a Telemedicine System for the Evaluation of Acute-Phase Retinopathy of Prematurity. *JAMA Ophthalmol* 2014.
6. Fijalkowski N, Zheng LL, Henderson MT, Wallenstein MB, Leng T, Moshfeghi DM. Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDRP): Four-years of Screening with Telemedicine. *Curr Eye Res* 2013.
7. Vinekar A, Jayadev C, Mangalesh S, Shetty B, Vidyasagar D. Role of tele-medicine in retinopathy of prematurity screening in rural outreach centers in India - a report of 20,214 imaging sessions in the KIDROP program. *Seminars in fetal & neonatal medicine* 2015;20:335–45. [PubMed: 26092301]
8. Prakalapakorn SG, Wallace DK, Freedman SF. Retinal imaging in premature infants using the Pictor noncontact digital camera. *J Aapos* 2014;18:321–6. [PubMed: 25173892]
9. Early Treatment For Retinopathy Of Prematurity Cooperative G. 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. [PubMed: 14662586]
10. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol* 2005;123:991–9. [PubMed: 16009843]
11. Multicenter trial of cryotherapy for retinopathy of prematurity. Preliminary results. Cryotherapy for Retinopathy of Prematurity Cooperative Group. *Arch Ophthalmol* 1988;106:471–9. [PubMed: 2895630]
12. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857–72. [PubMed: 9595616]
13. Blyth CR, Still HA. Binomial Confidence Intervals. *Journal of the American Statistical Association* 1983;78:108–16.
14. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions: John Wiley and Sons; 2013.
15. Mehta M, Adams GG, Bunce C, Xing W, Hill M. Pilot study of the systemic effects of three different screening methods used for retinopathy of prematurity. *Early human development* 2005;81:355–60. [PubMed: 15814220]
16. Trinavarat A, Atchaneeyasakul LO, Udompunturak S. Applicability of American and British criteria for screening of the retinopathy of prematurity in Thailand. *Japanese journal of ophthalmology* 2004;48:50–3. [PubMed: 14767651]
17. Vinekar A, Gilbert C, Dogra M, et al. The KIDROP model of combining strategies for providing retinopathy of prematurity screening in underserved areas in India using wide-field imaging, tele-medicine, non-physician graders and smart phone reporting. *Indian journal of ophthalmology* 2014;62:41–9. [PubMed: 24492500]

Figure 1.

Example of grader interface with web-based software, FocusROP, when grading images for a real-world telemedicine retinopathy of prematurity screening scenario. The birth weight (BW), gestational age (GA) and post-menstrual age (PMA) at time of imaging were displayed with the images. Up to 3 image sets (i.e. color and corresponding red-free image(s)) could be uploaded per eye. Color images were displayed above and red-free images were displayed below.

Figure 2.

Flow diagram showing which infants and images were included in the feasibility and accuracy portions of the study. Images submitted were considered “gradable” for screening if they received a grade of fair or good image quality and were felt to have 3 or 4 gradable quadrants by the grader. If no images could be captured for an eye, that eye was considered “not gradable”. Images for both eyes had to be considered “gradable” for an infant to be considered “gradable”.

Table 1.

Gradable images among eyes and infants included in the feasibility and accuracy portions of the study:

	Part 1: Feasibility Analysis N=215 infant (430 eye) imaging sessions		Part 2: Accuracy Analysis N=50 infant (100 eye) imaging sessions	
	Grader #1	Grader #2	Grader #1	Grader #2
By eye				
Gradable images	371 (86.3%)	370 (86.0%)	77 (77.0%)	76 (76.0%)
Not gradable images	40 (9.3%)	41 (9.5%)	18 (18.0%)	19 (19.0%)
Reasons for failure [*]				
-poor image quality	13 (3.0%)	15 (3.5%)	6 (6.0%)	6 (6.0%)
-0–2 gradable quadrants	36 (8.4%)	41 (9.5%)	16 (16.0%)	19 (19.0%)
-no images	20 (4.7%)		5 (5%)	
By infant				
Gradable images	171 (79.5%)	169 (78.6%)	33 (66.0%)	31 (62.0%)
Not gradable images	29 (13.5%)	31 (14.4%)	12 (24.0%)	14 (28.0%)
Reasons for failure [*]				
-poor image quality	11 (5.1%)	13 (6.0%)	4 (8.0%)	5 (10%)
-0–2 gradable quadrants	30 (14.0%)	34 (15.8%)	12 (24.0%)	15 (30.0%)
-no images in 1 or both eyes	15 (7.0%)		5 (10%)	

^{*} A given eye or infant could have more than one reason for failure (of gradability).

Table 2.

Gradability of images submitted for grading by post-menstrual age (PMA)

By PMA (weeks)	Number of infants imaged at specified PMA	Percentage of images gradable ¹			
		By eye		By infant ²	
		Grader 1	Grader 2	Grader 1	Grader 2
31	3	83.3	100.0	66.7	100.0
32	2	100.0	100.0	100.0	100.0
33	5	80.0	80.0	60.0	60.0
34	9	77.8	72.2	77.8	66.7
35	11	72.7	68.2	72.7	63.6
36	21	73.8	78.6	61.9	66.7
37	21	81.0	88.1	66.7	81.0
38	20	97.5	97.5	95.0	95.0
39	21	90.5	85.7	85.7	71.4
40	18	97.2	100.0	94.4	100.0
41	15	80.0	90.0	66.7	80.0
42	11	90.9	86.4	81.8	72.7
43	11	100.0	95.5	100.0	90.9
44	7	78.6	71.4	71.4	71.4
45	16	90.6	84.4	87.5	81.3
46	8	68.8	68.8	62.5	62.5
47	6	100.0	91.7	100.0	83.3
48	5	100.0	100.0	100.0	100.0
49	1	100.0	100.0	100.0	100.0
50	1	50.0	50.0	0.0	0.0
51	3	83.3	66.7	66.7	33.3
Average gradability of all imaging sessions ³		78.2	77.2	69.8	68.9

¹ Images submitted were considered “gradable” for screening if they received a grade of fair or good image quality and judged to have 3 or 4 gradable quadrants by the grader. If no images could be captured for an eye, that eye was considered “not gradable”. Images for both eyes had to be considered “gradable” for an infant to be considered “gradable”.

² If either eye was considered not gradable or did not have images captured, then the infant was considered not gradable for that imaging session.

³ Contribution of each infant was weighted equally

Table 3:

Each grader's evaluation of images for plus, pre-plus disease, or normal vessels compared the retina specialist's respective clinical examination findings by eye and infant imaging session

	Clinical examination findings (reference standard)						
	Plus	Pre-plus	Normal vessels		Plus	Pre-plus	Normal vessels
By eye imaging session							
Grader #1				Grader #2			
Plus	6	0	0	Plus	6	0	2
pre plus	6	0	8	pre plus	6	0	78
Normal	0	19	323	Normal	0	19	254
unable to grade due to quality or number of visible quadrants	1	0	6	unable to grade due to quality or number of visible quadrants	1	0	3
no images	0	0	13	no images	0	0	13
Total number of eye imaging sessions	13	19	350	Total number of eye imaging sessions	13	19	350
By infant imaging session ¹							
Grader #1				Grader #2			
Plus	4	0	0	Plus	4	0	1
pre plus	2	0	4	pre plus	2	0	43
Normal	0	10	155	Normal	0	10	117
not gradable due to:	1	0	15	not gradable due to:	1	0	14
-quality or number of visible quadrants ²	1	0	5	-quality or number of visible quadrants ²	1	0	2
-no image in one eye ²	0	0	11	-no image in one eye ²	0	0	11
-no images both eyes	0	0	1	-no images both eyes	0	0	1
Total number of infant imaging sessions	7	10	174	Total number of infant imaging sessions	7	10	174

¹The infant was categorized according to the disease in the worse eye.

²not mutually exclusive, if either eye was considered not gradable or did not have images captured, then the infant was considered not gradable for that imaging session.

Table 4:

2 × 2 tables for sensitivity and specificity analysis of a screening test for identifying infants¹ who developed type 1 ROP for each telemedicine image grader

	Clinical examination findings (reference standard) ²				
	Type 1 ROP	No Type 1 ROP		Type 1 ROP	No Type 1 ROP
Grader #1			Grader #2		
Including only gradable images ³					
Positive screening test ⁴	3	2	Positive screening test ⁴	4	7
Negative screening test ⁵	0	28	Negative screening test ⁵	0	20
Including all images ⁶					
Positive screening test ⁴	7	9	Positive screening test ⁴	7	15
Negative screening test ⁵	0	34	Negative screening test ⁵	0	28

¹The infant was categorized according to the disease in the worse eye.

²The “reference standard” was the diagnosis of type 1 ROP (zone I with stage 3 or plus disease; or zone II, stage 2 or 3 with plus disease) on clinical examination by a retina specialist.

³Analysis was performed including “gradable” images only (i.e. images with fair or good image quality and 3 or 4 gradable quadrants by that grader). If no images could be captured for an eye, that eye was considered “not gradable”. Images for both eyes had to be considered “gradable” for an infant’s images to be considered “gradable”.

⁴A positive screening test was defined as an imaging session where an infant was graded as having pre-plus or plus disease in either eye by the telemedicine grader.

⁵A negative screening test was defined as an imaging session where an infant was graded as having normal vessels in both eyes by the telemedicine grader.

⁶Analysis was performed including all images with any “ungradable” image considered a positive screening test.