Impact of number and quality of retinal images in a telemedicine screening program for ROP: results from the e-ROP study

David Morrison, MD[a], **Erick D. Bothun, MD**[b], **Gui-Shuang Ying, PhD**[c], **Ebenezer Daniel, MBBS, MS, PhD**[c], **Agnieshka Baumritter, MS**[d], **Graham Quinn, MD, MSCE**[c,d], and **For the e-ROP Cooperative Group**

[a]Department of Ophthalmology and Visual Sciences, Vanderbilt Eye Institute, Nashville, Tennessee

[b]Departments of Ophthalmology and Visual Neurosciences and Pediatrics, University of Minnesota, Minneapolis

[c]Department of Ophthalmology, Scheie Eye Institute, University of Pennsylvania, Perelman School of Medicine, Philadelphia

[d]Division of Pediatric Ophthalmology, Children's Hospital of Philadelphia, Pennsylvania

Abstract

Background—Telemedicine for the detection of retinopathy of prematurity (ROP) is becoming increasingly common; however, obtaining the required multiple retinal images from an infant can be challenging. This secondary analysis from the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) study evaluated the detection of referral-warranted ROP (RW-ROP) by trained readers when a full set of 5 retinal images could not be obtained.

Methods—A total of 7,905 image sets from 1,257 infants in the study were evaluated. Retinal location of images and image quality were recorded. Sensitivity and specificity of RW-ROP detection by trained readers were calculated by comparing findings in incomplete image sets to the findings on standard eye examination.

Results—The majority of image sets contained all 5 retinal images (92.8%). The disk center view was the image most likely to be present and to be of acceptable image quality (96.8%). The nasal retina was the most difficult to obtain with acceptable image quality (83.4%). Sensitivity of detection of RW-ROP was 82.1% when 5 retinal images of acceptable quality were submitted for grading, 67.2% when 4 acceptable images were submitted, and 66.7% for 3 or fewer acceptable images ($P = 0.02$), with corresponding specificity of 82.2%, 89.0%, and 81.7% respectively ($P < 0.0001$). When images of any quality were evaluated, sensitivity was not increased ($P = 0.74$).

Correspondence: David Morrison, MD, Associate Professor of Ophthalmology and Pediatrics, Vanderbilt Eye Institute, 2311 Pierce Avenue, Nashville, TN 37232 (David.morrison.1@vanderbilt.edu).

Conclusions—The likelihood of detecting RW-ROP by telemedicine screening is decreased when a full set of retinal images is not obtained.

Current guidelines for screening for retinopathy of prematurity (ROP) recommend repeated and carefully timed eye examinations to identify serious disease in at-risk infants. At present in the United States, an infant with birth weight (BW) of <1501 g or gestational age (GA) of 30 weeks requires examinations beginning at 31-32 weeks' postmenstrual age (PMA) and examinations at intervals of 1-2 weeks until ROP treatment is required or the eye is considered at very low risk, that is, mature or regressed ROP.[1] To determine the validity of remote image evaluation in the detection of potentially serious ROP (referral-warranted or RW-ROP defined as presence of zone I ROP, stage 3 or worse ROP, or plus disease),[2] the Telemedicine Approaches to Evaluating Acute-Phase ROP (e-ROP) Study compared the presence of clinical findings consistent with RW-ROP on diagnostic examination performed by an ophthalmologist experienced in performing ROP examinations with the result of image grading by nonphysician readers.[3] The e-ROP protocol required 5 retinal images per eye (disk center and disk nasal, temporal, superior, and inferior) be taken by trained, nonphysician imagers from at-risk infants and uploaded them to a central server for grading by nonphysician trained readers. Among the 855 infants included in the initial publication, the e-ROP study reported a 90% sensitivity and an 87% specificity for detecting RW-ROP on trained reader grading, when both eyes were considered for the presence of RW-ROP in an infant.[4]

An important factor to successful implementation of telemedicine in ROP is the number and quality of the images obtained. In this secondary analysis from the e-ROP study, we report which of the 5 retinal images for an eye were less likely to be obtained, and the resulting sensitivity and specificity of ROP detection using incomplete image sets.

Subjects and Methods

The e-ROP study enrolled 1,284 infants from 11 US and 1 Canadian medical center. All images submitted from these infants were used for this secondary analysis. Inclusion criteria were prematurity with BW <1251 g. Exclusion criteria were PMA of >39 weeks at presentation unless referred for treatment, presence of a structural abnormality of the eye preventing retinal visualization, or previous ROP treatment. The protocol for imaging acquisition, selection for uploading, and image grading has previously been described.[2,5] Briefly, images were obtained using the RetCam Shuttle (Clarity Medical Systems, Pleasanton, CA) by trained and certified nonphysician imagers. The imager sought to obtain the 5 required retinal images for each eye of an infant and to upload them to a central server. Standard indirect ophthalmic examinations were completed in coordination with each imaging session. The ophthalmologist examiners determined the timing of eye examinations based on the infant's clinical needs. Nonphysician trained readers evaluated images in a masked fashion. Image quality was graded as good, fair, poor, or missing. Definitions used for this grading system are outlined in Table 1. For the purposes of this study, good and fair images were combined into a single "acceptable" grade category. The trained readers determined whether the retinal morphology observed was consistent with the presence of ROP, stage of ROP if present, zone of vascularization, and the presence of plus disease. The

result of the grading by the trained readers was then compared to the examination findings recorded by the ophthalmologist. For this secondary analysis, all image sets from the e-ROP database were graded.[4]

Statistical Analysis

We described the frequency distribution for number of images in a submitted image set, image quality among all images submitted and by each image field (disk center, temporal retina, nasal retina, superior retina, inferior retina). We calculated and compared sensitivity and specificity by number of retinal images in an image set and by number of retinal images with acceptable quality using the generalized linear models. Due to the small number of image sets having no more than 3 retinal images, we categorized the number of retinal images into three groups for statistical comparison: 3, 4, and 5 images. The sensitivity was calculated as the proportion of RW-ROP positive image gradings of an eye when indirect ophthalmic examination indicated the presence of RW-ROP in an eye at the same session, and specificity was calculated as the proportion of RW-ROP negative image gradings of an eye when indirect ophthalmic examination indicated the absence of RW-ROP. The 95% confidence intervals for sensitivity and specificity were also calculated. In all these analyses for calculating sensitivity, specificity, and their confidence intervals, and for comparing sensitivity and specificity across image quality groups, the inter-eye correlation and correlation from multiple image sessions were adjusted by generalized estimating equations using the sandwich robust estimate of variance.[6]

Results

Among the 1,284 enrolled infants, 1,257 underwent diagnostic examinations and all but 16 infants had at least one imaging session attempted by the certified retinal imager. Of the 7,905 image sets submitted, 7,332 (92.8%) contained all 5 required retinal images. When image quality was evaluated, more than 90% of image sets had 4 or more acceptable quality images present (Table 2).

Table 3 summarizes the frequency of acceptable, poor, and missing images by retinal view. Disk center was the image most frequently submitted to receive an acceptable image quality grading (96.8%), followed by the temporal retinal view (95.8%) and superior retina (93.2%). The nasal retinal view was the least likely be graded as acceptable (83.4%) and also the most likely to have poor image quality (13.2%) or to not be submitted in an image set (3.4%). When images of any quality are considered, disk center and superior retina were most likely to be submitted, whereas nasal and inferior retinal views were least likely to be submitted in incomplete image sets (Table 4).

When an analysis was performed by the number of acceptable quality images (Table 5), sensitivity increased with a larger number of images of acceptable quality: 66.7% when 3 or fewer acceptable images were submitted, 67.2% for 4 acceptable images, and 82.1% for 5 acceptable images ($P = 0.02$). The corresponding specificity was 81.7%, 89.0% and 82.2%, respectively ($P < 0.0001$). When sets with images of any quality were considered, sensitivity was 76.9% in sessions with 3 or fewer images present, 74.1% when 4 images were present, and 80.2% when 5 images were present ($P = 0.74$). Specificity was 69.6% for 3 or fewer

images present, 85.6% for 4 images present, and 83.5% for all 5 retinal images present (P = 0.0002).

Discussion

In this secondary analysis from the e-ROP study, we highlight some of the challenges that may be encountered in establishing a telemedicine screening program for ROP. Although greater than 90% of imaging attempts resulted in a complete set of images, patterns do emerge regarding the impact of missing images on the detection of ROP. Disk center was the image most likely to be submitted and to be of acceptable quality. When only 1 or 2 images (of any quality) were submitted, superior retina and disk center were most likely to be submitted. These findings are likely explained by imaging technique. RetCam imaging requires the camera contact a coupling gel on the cornea while an eyelid speculum is in place; manipulation of the camera on the eye in a limited space is often difficult. It is more challenging to move the camera to obtain peripheral images, particularly in the setting of smaller palpebral fissures in premature infants[7,8] and/or in infants with various respiratory support devices that may obstruct access to the eye and prevent maneuvering the camera. It is also quite difficult to manipulate the direction of the eye with scleral depression and maintain camera contact with the eye for image acquisition. Thus the disk center image was most often obtained. The superior retina was also often visualized, likely due to the upward eye position associated with the Bell's reflex. The nasal and inferior retinal images were the most difficult to obtain, likely due to eye position in addition to the temporally placed eyelid speculum preventing good camera contact with the eye. Analysis of which images are most likely to demonstrate disease is ongoing, however, and was not evaluated in the present report.

Our data also indicate that when all 5 retinal images were submitted, a greater number of acceptable images provide improved diagnostic sensitivity (82% for 5 acceptable images vs 67% for 3 or fewer acceptable images) and trend in the same direction when images of any quality are examined. This improved sensitivity with image number is not surprising. Features of RW-ROP may be missed if images cannot be obtained or if images are of poor quality and limit visualization of ROP morphology. Specificity often decreases with increased sensitivity. However, in our study specificity does not change dramatically with different numbers of images of acceptable quality (82.2% in 5 images of acceptable quality compared to 81.7% with 3 or fewer images) and in fact increased with 4 images present (89.0%). This is likely a statistical aberration due to the large number of imaging sets without RW-ROP. Specificity does increase when more images of any quality are examined (83.5% in 5 images vs 69.6% in 3 or fewer images). It stands to reason that more images of good quality should improve accurate diagnosis in the presence or absence of ROP despite this statistical finding. It should be noted that all image sets were read even if only one image was submitted. Although images were classified as acceptable or poor quality, the presence of poor image quality did not make it "ungradable."

Several previous studies have evaluated the sensitivity and specificity of retinal imaging for the diagnosis of ROP.[9-17] However, few have reported the quality of images obtained. Ells and colleauges[2] reported that 96% of infants had successful imaging sessions, although

image quality was not specifically reported. Wu and colleauges[18] reported that 21% of retinal images were of poor quality and could not be graded. Chiang and colleagues[19] reported that 93%-100% of imaging sessions provided usable images. In a subsequent study image quality was further evaluated.[20] Among 3 readers, an ungradable image due to poor image quality was reported in 0%-40.6% of images taken at 31-33 weeks' gestational age. The number of ungradable images decreased to 0%-6.7% in infants 35-37 weeks' gestational age, perhaps indicating that older infants are easier to image. Evaluation of e-ROP data has demonstrated that pupil dilation, respiratory support status, and comorbidities were the greatest barriers to successful imaging.[21] Given that better image quality and a larger number of images in an image set were associated with greater sensitivity for detecting RW-ROP diagnosis in e-ROP, real-world implementation of telemedicine ROP screenings may include repeated imaging or referral for standard ROP examination when fewer than 4 acceptable quality images can be obtained.

This study has several limitations. It is inherent difficult to compare image grading results with a criterion standard of known variability.[9,10] While this secondary analysis provides insight into the challenges of retinal imaging for ROP, it was not powered to evaluate sensitivity and specificity of detecting RW-ROP when image sets were not complete. This analysis was performed per eye imaged and not per infant. Thus, sensitivity or specificity could have been altered if both eyes were considered for an infant. The standard of care ophthalmic examination and the imaging sessions occurred at the same time. Thus, if a child was ill and could only tolerate 1 study procedure, only the standard ROP examination was performed. We cannot determine whether these infants could have tolerated the examination if imaging was the only procedure being performed. Finally, timing of examinations was determined by the physician examiner based on clinical need rather than imaging results. Thus, implementation of a real-world telemedicine screening program may require a revision of standard screening intervals for repeat examinations.

In conclusion, the use of a telemedicine system to detect potentially blinding ROP is becoming more widespread, both in the US and abroad.[1,7-20] Obtaining good quality images is an integral part of the implementation of telemedicine programs to maximize the screening sensitivity. Additionally, a full set of retinal images appears to be necessary to provide the best opportunity to detect RW-ROP.

Acknowledgments

References

1. Fierson WM. American Academy of Pediatrics Section on Ophthalmology, American Academy of Ophthalmology, American Association for Pediatric Ophthalmology and Strabismus, American association of Certified Orthoptists. Screening examination of premature infants for retinopathy of prematurity. Pediatrics. 2013; 131:189–95. [PubMed: 23277315]

2. 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. [PubMed: 14597517]

3. Quinn GE, Ying GS, Daniel E, et al. e-ROP Cooperative Group. Telemedicine approaches to evaluating acute-phase retinopathy of prematurity: study design. Ophthalmic Epidemiol. 2014; 21:256–67. [PubMed: 24955738]

4. Quinn GE, Ying GS, Daniel E, et al. e-ROP Cooperative Group. Validity of a telemedicine system for evaluation of acute-phase retinopathy of prematurity. JAMA Ophthalmol. 2014; 132:1178–84. [PubMed: 24970095]

5. Daniel E, Quinn GE, Hildebrand PL, et al. e-ROP Cooperative Group. validated system for centralized grading of retinopathy of prematurity: Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Study. JAMA Ophthalmol. 2015; 133:675–82. [PubMed: 25811772]

6. Smith PJ, Hadgu A. Sensitivity and specificity for correlated observations. Stat Med. 1992; 11:1503–9. [PubMed: 1410962]

7. Yen KG, Hess D, Burke B, Johnson RA, Feuer WJ, Flynn JT. 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. [PubMed: 11997800]

8. 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. [PubMed: 11176990]

9. Trese MT. What is the real gold standard for ROP screening? Retina. 2008; 28(3 Suppl):S1–2. [PubMed: 18317337]

10. Phelps DL. It's plus disease, isn't it? Arch Opthalmol. 2007; 125:963–64.

11. Schwartz SD, Harrison SA, Ferrone PJ, Trese MT. Telemedical evaluation and management of retinopathy of prematurity using a fibe-ROPtic digital fundus camera. Ophthalmology. 2000; 107:25–8. [PubMed: 10647714]

12. Photographic Screening for Retinopathy of Prematurity (Photo-ROP) Cooperative Group. The Photographic Screening for Retinopathy of Prematurity Study (Photo-ROP): primary outcomes. Retina. 2008; 28(3 Suppl):S47–54. [PubMed: 18317345]

13. Dhaliwa C, Wright E, Graham C, McIntosh N, Fleck BW. Wide-field digital retinal imaging versus binocular indirect ophthalmoscopy for retinopathy of prematurity screening: a two-observer prospective, randomised comparison. Br J Ophthalmol. 2009; 93:355–9. [PubMed: 19028742]

14. Lorenz B, Spasovska K, Elflein H, Schneider N. Wide-field digital imaging based telemedicine for screening for acute retinopathy of prematurity (ROP): six-year results of a multicentre field study. Graefes Arch Clin Exp Ophthalmol. 2009; 247:1251–62. [PubMed: 19462177]

15. Silva RA, Murakami Y, Lad EM, Moshfeghi DM. Stanford University network for diagnosis of retinopathy of prematurity (SUNDROP): 36-month experience with telemedicine screening. Ophthalmic Surg Lasers Imaging. 2011; 42:12–19. [PubMed: 20954641]

16. Dai S, Chow K, Vincent A. Efficacy of wide-field digital retinal imaging for retinopathy of prematurity screening. Clin Experiment Ophthalmol. 2011; 39:23–9. [PubMed: 20796264]

17. Weaver DT, Murdock TJ. Telemedicine detection of type 1 ROP in a distant neonatal intensive care unit. J AAPOS. 2012; 16:229–33. [PubMed: 22681938]

18. Wu C, Petersen RA, VanderVeen DK. RetCam imaging for retinopathy of prematurity screening. J AAPOS. 2006; 10:107–11. [PubMed: 16678743]

19. Chiang MF, Keenan JD, Starren J, et al. Accuracy and reliability of remote retinopathy of prematurity diagnosis. Arch Ophthalmol. 2006; 124:322–7. [PubMed: 16534051]

20. Chiang MF, Wang L, Busuioc M, et al. Telemedical retinopathy of prematurity diagnosis: accuracy, reliability, and image quality. Arch Ophthalmol. 2007; 125:1531–8. [PubMed: 17998515]

21. Karp KA, Baumritter A, Pearson DJ, et al. e-ROP Cooperative Group. Training retinal imagers for retinopathy of prematurity (ROP) screening. J AAPOS. 2016; 20:214–19. [PubMed: 27164425]

Table 1

The retinal image quality classification system[a]

Retinal image	Good (acceptable)	Fair (acceptable)	Poor
Posterior pole	Good focus, illumination, and exposure of disk center image; optic disk and the vessels in circular area up to at least 3 disk diameters from edge of disk in all 4 quadrants are clearly observed	Reasonable focus, illumination, and exposure of disk center image; apresence of artifacts may obscure parts of disk center image so that vessels within circular region of 3 disk diameters surrounding disk cannot be visualized clearly in one or more quadrants	Unable to visualize vessels within circular region of 3 disk diameters surrounding disk in all 4 quadrants
Temporal retina Nasal retina Superior retina Inferior retina	Good focus, illumination, and exposure; vascular components, presence or absence of pathology, and avascularity can be ascertained with certainty in peripheral retina of the image	Reasonable focus, illumination and exposure; presence of artifacts may obscure some parts of peripheral retina; not all peripheral vascular components, pathology, and avascularity can be determined with certainty	Unable to visualize vessels, pathology, or avascular areas in retinal periphery

[a]A single category of acceptable consisting of "good" and "fair" images was used in this analysis.

Table 2

The frequency distribution for number of retinal images in an image set (N = 7905 image sets)

No. retinal images of *any* quality in image set	No. image sets (%)
1	34 (0.43)
2	48 (0.61)
3	152 (1.92)
4	339 (4.29)
5	7332 (92.8)

No. retinal images with *acceptable* quality among sets with at least 1 image	No. image sets (%)
0	57 (0.72)
1	104 (1.32)
2	167 (2.11)
3	408 (5.16)
4	1126 (14.2)
5	6043 (76.5)

Table 3

The frequency distribution for image quality for each of 5 retinal image field based on retinal view[a]

Retinal image field	Acceptable n (%)	Poor n (%)	Missing n (%)
Disk center	7652 (96.8)	180 (2.3)	73 (0.9)
Temporal retina	7578 (95.8)	236 (3.0)	91 (1.2)
Nasal Retina	6593 (83.4)	1046 (13.2)	266 (3.4)
Superior retina	7366 (93.2)	393 (5.0)	146 (1.9)
Inferior retina	7192 (91.0)	366 (4.6)	347 (4.4)

[a]7905 possible images.

Table 4

The frequency distribution for image type by number of retinal images of any image quality

No. retinal images of any image quality in image set	No. image sets	Disk center n (%)	Temporal retina n (%)	Nasal retina n (%)	Superior retina n (%)	Inferior retina n (%)
1	34	12 (35.3)	6 (17.7)	0 (0)	15 (44.1)	1 (2.94)
2	48	30 (62.5)	23 (47.9)	7 (14.6)	31 (64.6)	5 (10.4)
3	152	138 (90.8)	130 (85.5)	43 (28.3)	106 (69.7)	39 (25.7)
4	339	320 (94.4)	323 (95.3)	257 (75.8)	275 (81.1)	181 (53.4)
5	7332	7332 (100)	7332 (100)	7332 (100)	7332 (100)	7332 (100)

Table 5

Sensitivity and specificity for the detection of referral-warranted retinopathy of prematurity (RW-ROP) by number of retinal images of acceptable image quality

No. retinal images of acceptable image quality in image set	Exams with RW-ROP			Exams without RW-ROP		
	Number of image sets/exams analyzed[a]	Positive from image grading	Negative from image grading	Sensitivity % (95% CI)	Positive from image grading	Negative from image grading	Specificity % (95% CI)
3	725	32	16	66.7 (48.7-80.8)	124	553	81.7 (78.1-84.7)
4	1118	43	21	67.2 (55.1-77.4)	116	938	89.0 (86.5-91.1)
5	5976	576	126	82.1 (77.8-85.6)	938	4336	82.2 (80.1-84.1)
P value				0.02			<0.0001

[a] Image sets excluded if there no corresponding diagnostic eye exam or if diagnostic exam could not determine features of RW-ROP.