Published in final edited form as:

JAMA Ophthalmol. 2016 November 01; 134(11): 1263–1270. doi:10.1001/jamaophthalmol.2016.3502.

Analysis of Discrepancy Between Diagnostic Clinical Examination Findings and Corresponding Evaluation of Digital Images in the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity Study

Graham E. Quinn, MD, MSCE, Anna Ells, MD, Antonio Capone Jr, MD, G. Baker Hubbard, MD, Ebenezer Daniel, MD, PhD, P. Lloyd Hildebrand, MD, and Gui-shuang Ying, PhD for the e-ROP (Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity) Cooperative Group

Department of Ophthalmology, University of Pennsylvania, Philadelphia (Quinn, Daniel, Ying); Division of Ophthalmology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Quinn); Department of Ophthalmology, University of Calgary, Calgary, Alberta, Canada (Ells); Associated Retinal Consultants, Royal Oak, Michigan (Capone); Dean McGee Eye Institute, University of Oklahoma, Oklahoma City (Hildebrand); Department of Ophthalmology, Emory University, Atlanta, Georgia (Hubbard)

Abstract

IMPORTANCE—As effective treatments for potentially blinding retinopathy of prematurity (ROP) have been introduced, the importance of consistency in findings has increased, especially with the shift toward retinal imaging in infants at risk of ROP.

OBJECTIVE—To characterize discrepancies in findings of ROP between digital retinal image grading and examination results from the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity study, conducted from May 2011 to October 2013.

DESIGN, SETTING, AND PARTICIPANTS—A poststudy consensus review of images was conducted by 4 experts, who examined discrepancies in findings between image grades by trained

Corresponding Author: Graham E. Quinn, MD, MSCE, Division of Ophthalmology, The Children's Hospital of Philadelphia, Wood Center, 1st Floor, Philadelphia, PA 19104 (quinn@email.chop.edu).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author Contributions: Drs Quinn and Ying had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Quinn, Ells, Capone, Hildebrand, Ying.

Acquisition, analysis, or interpretation of data: Quinn, Capone, Hubbard, Daniel, Ying.

Drafting of the manuscript: Quinn, Ells, Capone, Hubbard.

Critical revision of the manuscript for important intellectual content: Quinn, Capone, Hubbard, Daniel, Hildebrand, Ying.

Statistical analysis: Quinn, Capone, Ying.

Obtaining funding: Quinn, Ying.

Administrative, technical, or material support: Capone, Daniel, Hildebrand.

Study supervision: Quinn, Ells, Capone, Daniel.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Hildebrand reported receiving support from Inveon Corp and has a patent for Digital Disease Management System, with royalties paid by Inveon to the Board of Regents of the University of Oklahoma. No other disclosures were reported.

nonphysician readers and physician examination results in infants with referral-warranted ROP (RW-ROP). Images were obtained from 13 North American neonatal intensive care units from eyes of infants with birth weights less than 1251 g. For discrepancy categories with more than 100 cases, 40 were randomly selected; in total, 188 image sets were reviewed.

MAIN OUTCOMES AND MEASURES—Consensus evaluation of discrepant image and examination findings for RW-ROP components.

RESULTS—Among 5350 image set pairs, there were 161 instances in which image grading did not detect RW-ROP noted on clinical examination (G−/E+) and 854 instances in which grading noted RW-ROP when the examination did not (G+/E−). Among the sample of G−/E+ cases, 18 of 32 reviews (56.3%) agreed with clinical examination findings that ROP was present in zone I and 18 of 40 (45.0%) agreed stage 3 ROP was present, but only 1 of 20 (5.0%) agreed plus disease was present. Among the sample of G+/E− cases, 36 of 40 reviews (90.0%) agreed with readers that zone I ROP was present, 23 of 40 (57.5%) agreed with readers that stage 3 ROP was present, and 4 of 16 (25.0%) agreed that plus disease was present. Based on the consensus review results of the sampled cases, we estimated that review would agree with clinical examination findings in 46.5% of the 161 G−/E+ cases (95% CI, 41.6–51.6) and agree with trained reader grading in 70.0% of the 854 G+/E− cases (95% CI, 67.3–72.8) for the presence of RW-ROP.

CONCLUSIONS AND RELEVANCE—This report highlights limitations and strengths of both the remote evaluation of fundus images and bedside clinical examination of infants at risk for ROP. These findings highlight the need for standardized approaches as ROP telemedicine becomes more widespread.

Published studies evaluating telemedicine for retinopathy of prematurity (ROP), including the recent Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) study,^{1,2} use ophthalmologists' examination findings as the criterion standard for determining the accuracy of grading corresponding digital retinal images. This approach assumes diagnostic examinations are accurate; however, other studies suggest substantial variability using indirect ophthalmoscopy in a severe ROP diagnosis. In the Cryotherapy for Retinopathy of Prematurity trial,³ second examinations for determining treatment-severity ROP had 12% disagreement. Several studies address the variability of determining ROP status on digital retinal images.^{4–6} One study⁷ reported discrepancies of up to 13% between examination findings and image grading when both were performed by the same physician.

In the e-ROP study,² standard 5-field retinal image sets of eyes of at-risk infants were graded remotely by e-ROP–certified, nonphysician trained readers (TRs),⁸ and results were compared with examination findings of e-ROP–certified ophthalmologists. The accuracy of reader grading was determined by agreement of grades with examination findings for the presence of retinal morphology consistent with referral-warranted ROP (RW-ROP, defined as zone I ROP, stage 3 ROP or worse, or plus disease), indicating the need for evaluation by an ophthalmologist to consider treatment.

When results from paired image grading and diagnostic examination for infants were compared, the TR grading sensitivity for detecting RW-ROP was 90.0%, with a specificity of 87.0%, a negative predictive value of 97.3%, and a positive predictive value of 62.5% at the RW-ROP rate 19.4%.² When considering the known variability in accuracy of clinical

ROP examinations^{3,4,6,7,9} and image grading,¹⁰ we hypothesized that discrepancies between image grading and examination for an eye may result from clinical findings not documented on image grading or from image grading that noted findings not observed on examination. We also acknowledged that current retinal imaging does not provide quality images of the entire retina in premature infants. To better characterize these discrepancies, we analyzed discrepant RW-ROP findings between TR grading and clinical examination using consensus review of images by a panel of ROP experts.

Methods

Details of the e-ROP study have been published^{1,2,8,11,12} and can be found on clinicaltrials.gov (NCT01264276). Only major features pertinent to discrepant ROP findings are described here. The e-ROP protocol was approved by study center institutional review boards. Patients provided written consent at the onset of the study.

The e-ROP study enrolled 1284 infants with birth weights less than 1251 g in neonatal intensive care units in 13 North American centers from May 2011 to October 2013.² In 1257 infants, routine diagnostic examinations were paired with retinal imaging using a wide-field digital camera. A standardized imaging protocol that included 5 defined retinal fields and a morphology-based image grading protocol was used by non-physician TRs supervised by an ophthalmologist.⁸

A consensus panel of 4 ROP experts (A.C., A.L.E., G.B.H., and G.E.Q.) reviewed representative samples of the image sets in which TR grades differed from examination results for RW-ROP morphology. Image sets were presented in groups based on discrepant RW-ROP components (ie, zone IROP, stage 3 ROP or worse, or plus disease), and each component was evaluated independently.

We reviewed image sets with negative image grade findings and positive examination results (G-/E+), including 32 instances of zone I disease and 20 instances of plus disease, as well as image sets with positive image grade findings and negative examination results (G+/E-), including 16 instances of plus disease. For categories with more than 100 cases (ie, 128 G-/E+ image sets for stage 3 ROP as well as 212 and 758 G+/E- image sets for zone I disease and stage 3 ROP, respectively), a random sample of 40 image sets was selected. The panel reviewed a total of 188 sets.

The panel reviewed selected image sets together during sessions in September 2014 using the same grading protocol that TRs used for grading all e-ROP image sets. The panel discussed each set, and consensus was reached on the presence or absence of RW-ROP features on image grading.

To determine the percentage of RW-ROP-positive cases if the panel reviewed all discrepant cases, we first imputed results of discrepant components of RW-ROP for those discrepant cases not selected for review using the panel review results of the selected sample. We then determined RW-ROP status (based on the imputed or observed review results of RW-ROP components) and calculated the percent of RW-ROP-positive cases in all RW-ROP

discrepant cases. To account for the uncertainty of imputation, we calculated the 95% CI by imputing 1000 times.

Results

In the e-ROP study, 5520 image sets were evaluated for the presence of RW-ROP. The status of RW-ROP could not be determined for either grading or the corresponding examination in 170 image sets (3.1%). In the remaining 5350 sets, there was agreement between grading and examination results for the presence of RW-ROP findings in 632 sets (11.8%) and for the absence of RW-ROP findings in 3703 (69.2%) sets. There was disagreement between grading and examination results in 161 G-/E+ cases (3.0%) and 854 G+/E- cases (16.0%).

Analysis of G-/E+ Results

Zone I ROP—For cases in which clinicians diagnosed zone I ROP, 32 image sets were graded by TRs as having no ROP in zone I (8 had no ROP and 24 had ROP in zone II) (Table 1). According to panel review, 18 of 32 image sets (56.3%) had ROP in zone I (consistent with clinicians' diagnoses), 10 (31.3%) had ROP only in zone II, 2 (6.3%) had no ROP visible, and 2 (6.3%) had indeterminant results due to low-quality images. Figure 1A demonstrates review agreement with examination findings, and Figure 1B demonstrates review agreement with TR findings.

Stage 3 ROP or Worse—For cases in which clinicians diagnosed stage 3 ROP or worse, 128 image sets were graded by TRs as not having stage 3 ROP or worse (Table 1). Among a random sample of 40 image sets, 18 (45.0%) had stage 3 ROP or worse according to the panel (consistent with clinicians' diagnoses), 9 (22.5%) had no ROP, and 13 (32.5%) had stage 2 ROP. Figure 1C demonstrates an example of review agreement with TR findings and an example of review agreement with examination findings.

Plus Disease—For cases in which clinicians diagnosed plus disease, 20 image sets were graded by TRs as not having plus disease (Table 1). All 20 sets were reviewed. Review determined that 1 set (5.0%) had plus disease, 6 (30.0%) had normal posterior pole vessels, and 13 (65.0%) had 1 to 4 quadrants of preplus disease. Figure 1D demonstrates an example of review agreement with TR findings and an example of review agreement with examination findings.

Overall, if all 161 G-/E+ discrepant cases were reviewed, review would agree with examination findings on the presence of RW-ROP in 46.5% (95% CI, 41.6–51.6) of these cases.

Analysis of G+/E- Results

Zone I ROP—For cases in which clinicians did not diagnose zone I ROP, 212 image sets were graded by TRs as having zone I ROP (Table 2). A random sample of 40 sets was reviewed, with 36 (90.0%) found to have ROP in zone I (consistent with TR grading) and 4 (10.0%) with zone II ROP. Figure 2A demonstrates review agreement with TR findings, and Figure 2B demonstrates review agreement with examination findings.

Stage 3 ROP or Worse—For cases in which clinicians did not diagnose stage 3 ROP or worse, 758 image sets were graded by TRs as having stage 3 ROP or worse (Table 2). A random sample of 40 sets was reviewed, and 23 (57.5%) had stage 3 ROP or worse (consistent with grading by TRs), 14 (35.0%) had stage 2 ROP, and 3 (7.5%) had indeterminant results. Figure 2C demonstrates an example of review agreement with TR findings and an example of review agreement with examination findings.

Plus Disease—For cases in which clinicians did not diagnose plus disease, TRs graded 16 image sets as having plus disease (Table 2). Among all 16 sets reviewed, 4 (25.0%) had plus disease (consistent with grading) and 12 (75.0%) had 1 to 4 quadrants of preplus disease. Figure 2D demonstrates an example of review agreement with TR findings and an example of review agreement with examination findings.

Overall, if all 854 G+/E– discrepant cases were reviewed, review would agree with TR findings on the presence of RW-ROP in 70.0% (95% CI, 67.3–72.8) of these cases.

Discussion

In the e-ROP study, results of diagnostic examinations performed by e-ROP–certified ophthalmologists of infants with birth weights less than 1251 g were used as the criterion standard for the accuracy (ie, sensitivity, specificity, and negative and positive predictive values) of grading of corresponding image sets obtained by e-ROP–certified imagers. We demonstrated that combining results from each eye of an infant improved sensitivity, an important approach for clinical care.¹ However, as there were a number of disagreements in individual eyes between the clinical examination and the corresponding grading of an image set, we convened a panel of ROP experts to examine these disagreements to determine whether G–/E+ and G+/E– results were indeed discrepancies and why. Using infants rather than the eye as the unit of analysis for discrepant analysis would likely obscure some eye-level discrepancies.

In analyzing the G–/E+ cases reviewed by the consensus panel, the greatest difference was determined to be an image grading error for zone I ROP, with the consensus panel agreeing with examination findings in 56.3% of cases, similar to the 61.5% (8 of 13 cases) noted by Scott et al.⁷ Several potential explanations may account for the G–/E+ findings. First, there is an inherent imprecision in the outer boundary of zone I. Because the outer extent of zone I is defined as twice the disc-fovea distance, any change in the estimated location of the fovea can change the boundary of zone I substantially. Trained readers may have more variability in estimating the location of the fovea than clinicians, perhaps explaining why the consensus panel agreed with the bedside examination more frequently. Second, identifying the clinical features of ROP in zone I are often subtle, in particular flat stage 3 ROP, and challenge clinicians who manage ROP.^{13–15} The retinopathy may have been apparent on the image but incorrectly interpreted by TRs. Finally, at least 2 of the G–/E+ cases could have been avoided if the threshold for rating an image set as “not gradable” was lower, which would have designated those image sets as having RW-ROP.

Most of the G-/E+ image sets adjudicated for stage 3 ROP (22 of 40 [55.0%]) and plus disease (19 in 20 [95.0%]) were found to be accurately assessed by TRs, ie, the consensus panel was more likely to agree with TRs than with the examining ophthalmologist. Interestingly, G-/E+ discrepancies on the presence of plus disease weighed strongly in favor of TRs; all but 1 of the image sets reviewed by the consensus panel were determined to have no plus disease present. The threshold for determining plus disease by experienced clinicians is highly variable, as reported in both clinical trials¹⁶ and in studies using digital image analysis.^{4,5} These studies underscore the challenges of clinically diagnosing ROP requiring treatment and highlight the inherent variability in the clinical diagnosis of critical ROP features.

There was a considerable number of G+/E- cases (854 [16.0%]). When reviewing a random sample of zone I G+/E- cases, the panel determined that zone IROP was indeed present 90.0% of the time. This finding is particularly noteworthy because there were 212 G+/E- cases for zone I in the e-ROP study, ie, there were many instances in which the bedside examination did not find ROP in zone I and yet there was good photographic evidence for the presence of zone IROP among the random sample reviewed 90% of the time. For G+/E- cases of stage 3 ROP, the panel determined the TRs correctly documented its presence in 57.5% of the image sets, ie, the images showed findings inconsistent with the clinical examination. In the remaining 42.5% of sets, only stage 2 or no ROP was noted by the consensus panel. For plus disease, the consensus panel determined that 25.0% of image grades in the G+/E- sample did indeed show plus disease. In the remaining 75.0%, morphology consistent with preplus disease was noted.

In the e-ROP study, the number of G-/E+ and G+/E- cases varied considerably among the 3 RW-ROP components (Table 1 and Table 2). Given the high number of disagreements for stage 3 ROP (128 for G-/E+ and 758 for G+/E-), this morphologic indicator of severe ROP may be the most difficult RW component to document accurately, both on clinical examination and on image grading. For clinicians, mild to moderate stage 3 ROP may be interpreted as stage 2 ROP. For readers of digital images, stage 3 ROP may not have been included in the image set, and the extraretinal neovascularization in stage 3 ROP may be challenging to differentiate from stage 2 ROP on 2-dimensional images. Further, image clarity and focus are essential for detection of stage 3 ROP outside of zone I, and image quality is also affected by poor dilation, darkly pigmented fundus, and vitreous haze. In our analysis of discrepant cases, the panel agreed with TRs in most cases for the other components easily captured by the camera (ie, zone IROP and presence of plus disease). Intuitively, it seems more likely that image grading would be less effective at ruling out stage 3 ROP than detecting zone I or plus disease because of limitations in obtaining quality images of more peripheral retinal pathology. This may explain why no ROP was detected on consensus grading in 22.5% of G-/E+ cases in which the bedside examination documented stage 3 ROP.

This report has several limitations. The panel did not review all discrepant cases, and the panel was aware that the image sets represented discrepant cases. In addition, a consensus panel is not an ideal format for determining the accuracy of differences or agreements, but given the retrospective data available for understanding the discrepancies, we consider it a

reasonable approach to analyze the disagreements between examination results and TR grading.

In telemedicine studies of diabetic retinopathy and age-related macular degeneration, remote capture of standardized image sets with centralized image set grading by TRs has shown higher accuracy than indirect ophthalmoscopy performed by clinicians.^{17–21} In digital imaging for both diabetic retinopathy and age-related macular degeneration, imaging protocols, grading algorithms, disease morphologies, and classification and staging of disease have been extensively examined and rigorously evaluated. The current reference standard for detecting, staging, and monitoring the progression of diabetic retinopathy is a protocol determined by the Early Treatment Diabetic Retinopathy Study that requires the use of a set of 7 different 30° retinal fields taken in stereo. This approach has been broadly applied both in research and teleretinal care programs and is acknowledged as the criterion standard in practice guidelines published by the American Diabetes Association and the American Academy of Ophthalmology Preferred Practice Patterns.²²

The evolution of our understanding of the usefulness of telemedicine in ROP is still under way. This review found that in more than half of G–/E+ cases, review of the images documented plus disease, stage 3 ROP, or zone I ROP. Further, in a review of the G+/E– discrepant cases, almost 3 of 4 cases were estimated to be consistent with RW-ROP. These findings argue for fewer qualitative assessments of at-risk eyes and for the development of more quantitative standards. They also argue that a composite of image grading and clinical examination results are a reasonable choice for determining treatment effects in clinical trials. Other modalities, such as fluorescein angiography²³ and optical coherence tomography,^{24–26} may enhance detection of serious ROP.

Conclusions

At present, decisions about whether to implement ROP telemedicine must be based on available evidence. A major concern has been whether telemedicine will detect potentially significant morphological features of ROP that would influence clinical care, and this study offers insight into the limitations of telemedicine. Importantly, these data also highlight that bedside examination has limitations in identifying potentially important features of ROP.

The findings of this study suggest that the region of the retina in which most severe disease occurs (ie, zone I) may be best delimited by assessment of retinal images; however, the subtleties of the clinical findings of stage 3 ROP in zone I still require insight/oversight by experienced clinicians, possibly at the bedside. There are also, at present, technical limitations in retinal imaging in the neonatal intensive care units, as current cameras have low resolution, and there are difficulties in field definition with imaging. However, present indications for treatment of ROP²⁷ require the presence of plus disease except when stage 3 ROP is noted in zone I, and it appears that image grading is more specific for the determination of plus disease. Further, the definition of RW-ROP is set sufficiently low to capture eyes requiring treatment prior to the presence of type 1 ROP. The joint technical report on Telemedicine for Evaluation of Retinopathy of Prematurity recently published by the American Academy of Pediatrics Section on Ophthalmology, American Academy of

Ophthalmology, and American Association of Certified Orthoptists²⁸ is a step toward developing a standard approach to ROP telemedicine. Further refinements in ROP image grading are needed. It may improve sensitivity to train nonphysician readers to consider birth weight, gestational age, and postmenstrual age when grading images. Further studies should examine the effect of grading both eyes at the same session and comparing previous imaging sessions. In addition, developing improved grading protocols to determine location of the retinopathy and to describe clearly which image sets are ungradable are essential for establishing performance standards for telemedicine-based ROP evaluation programs.

Acknowledgments

Funding/Support: This study was funded by cooperative agreement grant U10 EY017014 from the National Eye Institute of the National Institutes of Health.

Group Information

The Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity study investigators included the following: Office of Study Chair: *The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania*: Graham E. Quinn, MD, MSCE (principal investigator [PI]); Kelly Wade, MD, PhD, MSCE; Agnieszka Baumritter, MS; Trang B. Duros, BA; and Lisa Erbring. Study centers: *Johns Hopkins University, Baltimore, Maryland*: Michael X. Repka, MD (PI); Jennifer A. Shepard, CRNP; Pamela Donohue, ScD; David Emmert, BA; and C. Mark Herring, CRA. *Boston Children's Hospital, Boston, Massachusetts*: Deborah VanderVeen, MD; Suzanne Johnston, MD; Carolyn Wu, MD; Jason Mantagos, MD; Danielle Ledoux, MD; Tamar Winter, RN, BSN, IBCLC; Frank Weng, BS; and Theresa Mansfield, RN, BSN. *Nationwide Children's Hospital, Columbus, Ohio* and *Ohio State University Hospital, Columbus, Ohio*: Don L. Bremer, MD (PI); Richard Golden, MD; Mary Lou McGregor, MD; Catherine Olson Jordan, MD; David L. Rogers, MD; Rae R. Fellows, MEd, CCRC; Suzanne Brandt, RNC, BSN; and Brenda Mann, RNC, BSN. *Duke University, Durham, North Carolina*: David Wallace, MD (PI); Sharon Freedman, MD; Sarah K. Jones, BS; Du Tran-Viet, BS; and Rhonda "Michelle" Young, RN. *University of Louisville, Louisville, Kentucky*: Charles C. Barr, MD (PI); Rahul Bhola, MD; Craig Douglas, MD; Peggy Fishman, MD; Michelle Bottorff, BS; Brandi Hubbuch, RN, MSN, NNP-BC; and Rachel Keith, PhD. *University of Minnesota, Minneapolis*: Erick D. Bothun, MD (PI); Inge DeBecker, MD; Jill Anderson, MD; Ann Marie Holleschau, BA, CCRP; Nichole E. Miller, MA, RN, NNP; and Darla N. Nyquist, MA, RN, NNP. *University of Oklahoma, Oklahoma City*: R. Michael Siatkowski, MD (PI); Lucas Trigler, MD; Marilyn Escobedo, MD; Karen Corff, MS, ARNP, NNP-BC; Michelle Huynh-Blunt, MS, ARNP; and Kelli Satnes, MS, ARNP, NNP-BC; *The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania*: Monte D. Mills, MD; Will Anninger, MD; Gil Binenbaum, MD, MSCE; Graham E. Quinn, MD, MSCE; Karen A. Karp, BSN; and Denise Pearson, COMT. *University of Texas Health Science Center at San Antonio*: Alice Gong, MD (PI); John Stokes, MD; Clio Armitage Harper, MD; Laurie Weaver, RNC, BSN; Carmen McHenry, BSN; Kathryn Conner, RN, BSN; Rosalind Heemer, BSN; Elnora Cokley, RNC; and Robin Tragus, MSN, RN, CCRC. *University of Utah, Salt Lake City*: Robert Hoffman, MD (PI); David Dries, MD; Katie Jo Farnsworth, BS; Deborah Harrison, MS; Bonnie Carlstrom,

COA; and Cyrie Ann Frye, CRA, OCT-C. *Vanderbilt University, Nashville, Tennessee*: David Morrison, MD (PI); Sean Donahue, MD; Nancy Benegas, MD; Sandy Owings, COA, CCRP; Sandra Phillips, COT, CRI; and Scott Ruark, DO. *Foothills Medical Center, Calgary, Alberta, Canada*: Anna Ells, MD, FRCS (PI); Patrick Mitchell, MD; April Ingram, BS; and Rosie Sorbie, RN. Data Coordinating Center: *Perelman School of Medicine, University of Pennsylvania, Philadelphia*: Gui-shuang Ying, PhD (PI); Maureen Maguire, PhD; Mary Brightwell-Arnold, BA, SCP; Max Pistilli, MS; Kathleen McWilliams, CCRP; Sandra Harkins; and Claressa Whearry. Image Reading Center: *Perelman School of Medicine, University of Pennsylvania*: Ebenezer Daniel, MBBS, MS, MPH (PI); E. Revell Martin, BA; Candace R. Parker Ostroff, BA; Krista Sepielli, BFA; and Eli Smith, BA. Expert Readers: *The Vision Research ROPARD Foundation, Novi, Michigan*: Antonio Capone, MD. *Emory University School of Medicine, Atlanta, Georgia*: G. Baker Hubbard, MD. *Foothills Medical Center, Calgary, Alberta, Canada*: Anna Ells, MD, FRCS. Image Data Management Center: *Inoveon Corporation, Oklahoma City, Oklahoma*: P. Lloyd Hildebrand, MD (PI); Kerry Davis, BA; G. Carl Gibson, BBA, CPA; and Regina Hansen, COT. Cost-Effectiveness Component: *Duke University, Durham, North Carolina*: Alex R. Kemper, MD, MPH, MS (PI). *University of Michigan, Ann Arbor, Michigan*: Lisa Prosser, PhD. Data Management and Oversight Committee: *Kellogg Eye Center, University of Michigan, Ann Arbor*: David C. Musch, PhD, MPH (chair); *Department of Ophthalmology, Boston Medical Center, Boston University School of Medicine, Boston, Massachusetts*: Stephen P. Christiansen, MD; *Bascom Palmer Eye Institute, University of Miami Leonard M. Miller School of Medicine, Miami, Florida*: Ditte J. Hess, CRA; *Department of Ophthalmology and Visual Sciences, Center for Health Policy, Washington University School of Medicine, St Louis, Missouri*: Steven M. Kymes, PhD; *Department of Ophthalmology, Doheny Eye Center UCLA, University of California at Los Angeles, Arcadia*: Srinivas R. Sadda, MD; *University of Kansas Center for Telemedicine and Telehealth, Kansas City*: Ryan Spaulding, PhD. *National Eye Institute, Bethesda, Maryland*: Eleanor B. Schron, PhD, RN.

References

1. Quinn GE. e-ROP Cooperative Group. Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity: study design. *Ophthalmic Epidemiol.* 2014; 21(4):256–267. [PubMed: 24955738]
2. Quinn GE, Ying GS, Daniel E, et al. e-ROP Cooperative Group. Validity of a telemedicine system for the evaluation of acute-phase retinopathy of prematurity. *JAMA Ophthalmol.* 2014; 132(10): 1178–1184. [PubMed: 24970095]
3. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity: three-month outcome. *Arch Ophthalmol.* 1990; 108(2):195–204. [PubMed: 2405827]
4. Gschließer A, Stifter E, Neumayer T, et al. Inter-expert and intra-expert agreement on the diagnosis and treatment of retinopathy of prematurity. *Am J Ophthalmol.* 2015; 160(3):553–560. e3. [PubMed: 26004406]
5. Chiang MF, Wang L, Busuioc M, et al. Telemedical retinopathy of prematurity diagnosis: accuracy, reliability, and image quality. *Arch Ophthalmol.* 2007; 125(11):1531–1538. [PubMed: 17998515]
6. Wallace DK, Quinn GE, Freedman SF, Chiang MF. Agreement among pediatric ophthalmologists in diagnosing plus and pre-plus disease in retinopathy of prematurity. *J AAPOS.* 2008; 12(4):352–356. [PubMed: 18329925]

7. Scott KE, Kim DY, Wang L, et al. Telemedical diagnosis of retinopathy of prematurity intraphysician agreement between ophthalmoscopic examination and image-based interpretation. *Ophthalmology*. 2008; 115(7):1222–1228. e3. [PubMed: 18456337]
8. 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(6):675–682. [PubMed: 25811772]
9. Chiang MF, Jiang L, Gelman R, Du YE, Flynn JT. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. *Arch Ophthalmol*. 2007; 125(7):875–880. [PubMed: 17620564]
10. Ells AL, Holmes JM, Astle WF, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology*. 2003; 110(11):2113–2117. [PubMed: 14597517]
11. Kemper AR, Wade KC, Hornik CP, Ying GS, Baumritter A, Quinn GE. Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Study Cooperative Group. Retinopathy of prematurity risk prediction for infants with birth weight less than 1251 grams. *J Pediatr*. 2015; 166(2):257–61. e2. [PubMed: 25449219]
12. Wade KC, Pistilli M, Baumritter A, et al. e-Retinopathy of Prematurity Study Cooperative Group. Safety of retinopathy of prematurity examination and imaging in premature infants. *J Pediatr*. 2015; 167(5):994–1000. e2. [PubMed: 26299381]
13. International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol*. 2005; 123(7):991–999. [PubMed: 16009843]
14. Shaikh S, Capone A Jr, Schwartz SD, Gonzales C, Trese MT. ROP Photographic Screening Trial (Photo-ROP) Study Group. Inadvertent skip areas in treatment of zone 1 retinopathy of prematurity. *Retina*. 2003; 23(1):128–131. [PubMed: 12652254]
15. Koreen S, Lopez R, Jokl DH, Flynn JT, Chiang MF. Variation in appearance of severe zone 1 retinopathy of prematurity during wide-angle contact photography. *Arch Ophthalmol*. 2008; 126(5):736–737. [PubMed: 18474793]
16. Reynolds JD, Dobson V, Quinn GE, et al. CRYO-ROP and LIGHT-ROP Cooperative Study Groups. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. *Arch Ophthalmol*. 2002; 120(11):1470–1476. [PubMed: 12427059]
17. Mansberger SL, Shepler C, Barker G, et al. Long-term comparative effectiveness of telemedicine in providing diabetic retinopathy screening examinations: a randomized clinical trial. *JAMA Ophthalmol*. 2015; 133(5):518–525. [PubMed: 25741666]
18. Shi L, Wu H, Dong J, Jiang K, Lu X, Shi J. Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis. *Br J Ophthalmol*. 2015; 99(6):823–831. [PubMed: 25563767]
19. Silva PS, Aiello LP. Telemedicine and eye examinations for diabetic retinopathy: a time to maximize real-world outcomes. *JAMA Ophthalmol*. 2015; 133(5):525–526. [PubMed: 25742322]
20. Vaziri K, Moshfeghi DM, Moshfeghi AA. Feasibility of telemedicine in detecting diabetic retinopathy and age-related macular degeneration. *Semin Ophthalmol*. 2015; 30(2):81–95. [PubMed: 24171781]
21. Zimmer-Galler IE, Kimura AE, Gupta S. Diabetic retinopathy screening and the use of telemedicine. *Curr Opin Ophthalmol*. 2015; 26(3):167–172. [PubMed: 25759962]
22. American Academy of Ophthalmology Preferred Practice Pattern Retina/Vitreous Panel. [Accessed August 25, 2016] Screening for diabetic retinopathy: 2014. <http://www.aao.org/clinical-statement/screening-diabetic-retinopathy--june-2012>
23. Klufas MA, Patel SN, Ryan MC, et al. Influence of fluorescein angiography on the diagnosis and management of retinopathy of prematurity. *Ophthalmology*. 2015; 122(8):1601–1608. [PubMed: 26028345]
24. Vinekar A, Mangalesh S, Jayadev C, Maldonado RS, Bauer N, Toth CA. Retinal imaging of infants on spectral domain optical coherence tomography. *Biomed Res Int*. 2015; 2015(2015):782420. [PubMed: 26221606]

25. Maldonado RS, Izatt JA, Sarin N, et al. Optimizing hand-held spectral domain optical coherence tomography imaging for neonates, infants, and children. *Invest Ophthalmol Vis Sci*. 2010; 51(5): 2678–2685. [PubMed: 20071674]
26. Maldonado RS, Toth CA. Optical coherence tomography in retinopathy of prematurity: looking beyond the vessels. *Clin Perinatol*. 2013; 40(2):271–296. [PubMed: 23719310]
27. 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(12):1684–1694. [PubMed: 14662586]
28. Fierson WM, Capone A Jr. American Academy of Pediatrics Section on Ophthalmology; American Academy of Ophthalmology, American Association of Certified Orthoptists. Telemedicine for evaluation of retinopathy of prematurity. *Pediatrics*. 2015; 135(1):e238–e254. [PubMed: 25548330]

Key Points

Question

What are the discrepancies between image grading and clinical examination findings in infants at risk for retinopathy of prematurity (ROP) in the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity study?

Findings

Consensus review of discrepant referral-warranted ROP findings (ie, eyes with zone I ROP, stage 3 ROP, or plus disease) by a panel of ROP experts found that nonphysician readers were less likely to detect stage 3 ROP than clinicians. However, trained readers were generally better than physicians at identifying the zone of ROP and presence of plus disease.

Meaning

This report highlights different strengths and weaknesses of both the grading of retinal images and the bedside examination of infants at risk for ROP.

Figure 1. Images From Eyes in Which Grading Did Not Detect Findings Consistent With Referral-Warranted Retinopathy of Prematurity (ROP) Noted on Examination

A, Example of ROP identified as being in zone I on clinical examination but graded as not being in zone I by trained readers. Left, The radius of zone I was determined as twice the distance between the disc and fovea. Right, The radius was applied to the same eye, demonstrating that the ROP lies inside zone I. B, Example of ROP identified as being in zone I on clinical examination but graded as not being in zone I by trained readers. Left, The radius of zone I was determined as twice the distance between the disc and fovea. Right, The radius was applied to the same eye, demonstrating that the ROP lies outside zone I. C, Examples of discrepant image grading in which trained readers did not identify stage 3 ROP that was noted on clinical examination. Left, Review agreed with grading by trained readers. Right, Review agreed with clinical examination findings. D, Examples of discrepant image grading in which trained readers did not identify plus disease that was noted on clinical examination. Left, Review agreed with grading by trained readers. Right, Review agreed with clinical examination findings. G+/E- indicates positive findings on image grading with negative findings on clinical examination (or a false positive with positive findings on image grading not noted on clinical examination).

Figure 2. Images From Eyes in Which Grading Detected Findings Consistent With Referral-Warranted Retinopathy of Prematurity (ROP) Not Noted on Examination

A, Example of ROP identified as not being in zone I on clinical examination but graded as being in zone I by trained readers. Left, The radius of zone I was determined as twice the distance between the disc and fovea. Right, The radius was applied to the same eye, demonstrating that the ROP lies within zone I. B, Example of ROP identified as not being in zone I on clinical examination but graded as being in zone I by trained readers. Left, The radius of zone I was determined as twice the distance between the disc and fovea. Right, The radius was applied to the disc right image of the same eye, demonstrating that the ROP lies within zone I. C, Examples of discrepant image grading in which trained readers identified stage 3 ROP that was not noted on clinical examination. Left, Review agreed with grading by trained readers. Right, Review agreed with clinical examination findings. D, Examples of discrepant image grading in which trained readers identified plus disease that was not noted on clinical examination. Left, Review agreed with grading by trained readers. Right, Review agreed with clinical examination findings. G+/E+ indicates negative findings on image grading with positive findings on clinical examination (or a false negative with positive findings on clinical examination not noted on image grading).

Table 1

Discrepant Cases With Negative Grades and Positive Examination Findings

	No. of Image Sets			
Diagnosis	Clinical Examination	Trained Readers	Trained Reader Finding in Sample Reviewed	Consensus Review Results, No. (%)
Zone I ROP ^a				
No ROP	NA	8	8	2 (6.3)
Zone II	NA	24	24	10 (31.3)
Zone I	32	NA	NA	18 (56.3)
Poor image quality (RW-ROP)	NA	NA	NA	2 (6.3)
Stage 3 ROP ^b				
No ROP	NA	33	11	9 (22.5)
Stage 1	NA	2	NA	0
Stage 2	NA	93	29	13 (32.5)
Stage 3	128	NA	NA	18 (45.0)
Plus disease ^c				
Normal vessels	NA	5	5	6 (30.0)
1–4 Quadrants of preplus disease	NA	15	15	13 (65.0)
Plus disease	20	NA	NA	1 (5.0)

Abbreviations: NA, not applicable; ROP, retinopathy of prematurity; RW-ROP, referral-warranted retinopathy of prematurity.

^aAll 32 discrepant cases were reviewed.^bA random sample of 40 of the 128 discrepant cases was reviewed.^cAll 20 discrepant cases were reviewed.

Table 2

Discrepant Cases With Positive Grades and Negative Examination Findings

Diagnosis	No. of Image Sets		Examination Finding in Sample Reviewed	Consensus Review Results, No. (%)
	Clinical Examination	Trained Readers		
Zone I ROP ^a				
No ROP	17	NA	2	0
Zone II	195	NA	38	4 (10.0)
Zone I	NA	212	NA	36 (90.0)
Stage 3 ROP ^b				
No ROP	36	NA	1	0
Stage 1	130	NA	13	0
Stage 2	592	NA	26	14 (35.0)
Stage 3	NA	758	NA	23 (57.5)
Indeterminant	NA	NA	NA	3 (7.5)
Plus disease ^c				
Normal vessels	4	NA	4	0
1–4 Quadrants of preplus disease	12	NA	12	12 (75.0)
Plus disease	NA	16	NA	4 (25.0)

Abbreviations: NA, not applicable; ROP, retinopathy of prematurity.

^a A random sample of 40 of the 212 discrepant cases was reviewed.^b A random sample of 40 of the 128 discrepant cases was reviewed.^c All 16 discrepant cases were reviewed.