Published in final edited form as:

J AAPOS. 2018 February ; 22(1): 32–37. doi:10.1016/j.jaapos.2017.09.005.

Single grading vs double grading with adjudication in the telemedicine approaches to evaluating acute-phase retinopathy of prematurity (e-ROP) study

Ebenezer Daniel, MBBS, PhD^a, Wei Pan, MS^a, Graham E. Quinn, MD, MSCE^{a,b}, Eli Smith, BA^a, Agnieszka Baumritter, MS^b, Gui-Shuang Ying, PhD^a, and the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Cooperative Group*

^aDepartment of Ophthalmology, University of Pennsylvania, Philadelphia

^bDivision of Pediatric Ophthalmology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

Abstract

Purpose—To evaluate the sensitivity and specificity of single, independent, nonphysician trained reader (TR) gradings in the Telemedicine Approaches to Evaluating Acute-phase Retinopathy of Prematurity (e-ROP) study.

Methods—Secondary analyses of image grading results from 1,235 infants of birth weights <1251 g. Two of three TRs independently graded image sets; discrepancies were adjudicated by the reading center director (an ophthalmologist) to reach final grading. Sensitivity and specificity of each TR grading and final grading was calculated by comparing gradings to clinical examination results.

Results—Of 7,808 double graded image sets, TR1 graded 5,165; TR2, 3,787; and TR3, 6,664. Compared to final grading for referral warranted retinopathy of prematurity (RW-ROP), two TRs had relatively lower sensitivity (TR1, 75% vs 79% [$P = 0.03$]; TR2, 73% vs 77% [$P = 0.02$]) and specificity (TR1, 80% vs 83% [$P < 0.001$]; TR2, 82% vs 83% [$P = 0.09$]). TR3 had similar sensitivity (83% vs 83% [$P = 0.78$]) and specificity (83% vs 84% [$P = 0.02$]). Compared to final grading, TR1 had lower sensitivity for zone I ROP (47% vs 56% [$P = 0.04$]) and stage 3 ROP (71% vs 77% [$P = 0.002$]); TR2 had lower sensitivity for stage 3 ROP (69% vs 77% [$P < 0.001$]) and lower specificity for all three components ($P < 0.001$); TR3 had lower sensitivity for detecting plus disease (23% vs 35% [$P < 0.001$]) and similar sensitivity for zone I ROP and stage 3 ROP.

Correspondence: Ebenezer Daniel, MBBS, MS, MPH, PhD, Ophthalmology Reading Center, University of Pennsylvania, 3535 Market Street, Suite 700, Philadelphia, PA 19104 (ebdaniel@mail.med.upenn.edu).

*A listing of the e-ROP Cooperative Group members is provided in eSupplement 1, available at jaapos.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Daniel E, Pan W, Quinn GE, Smith E, Baumritter A, Ying G; the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Cooperative Group. Single grading vs double grading with adjudication in the telemedicine approaches to evaluating acute-phase retinopathy of prematurity (e-ROP) study. J AAPOS online supplement.

Conclusions—There is a small but significant decrease in the sensitivity and specificity for RW-ROP when single-reader grading is compared to double adjudicated grading.

Fewer than 10% of infants examined for possible retinopathy of prematurity (ROP) require treatment¹; nevertheless, the number of infants who must be examined is high compared to the number of ophthalmologists experienced in ROP care. Telemedicine has the potential to provide an efficient approach to detecting the acute phases of severe ROP for timely treatment. Retinal images of infants at risk can be obtained in the neonatal intensive care units and transferred to a remote centralized reading center, where they can be evaluated for the presence of referral-warranted ROP (RW-ROP, defined as plus disease, stage 3, or zone 1 ROP).² If RW-ROP is identified, a report can be sent to the ROP specialist who can then examine the infant to judge whether treatment is necessary. One of the critical components of an effective ROP telemedicine system is a proficient image reader. This role has generally been performed by local ophthalmologists.^{2–6} However, the cost of establishing and running a telemedicine system may be reduced if evaluation of retinal images from at-risk infants is performed by certified nonphysician readers.

The Telemedicine Approaches to Evaluating Acute-phase Retinopathy of Prematurity (e-ROP) Study, designed to evaluate the validity, reliability, feasibility and cost-effectiveness of a telemedicine system in detecting RW-ROP in the eyes of at-risk babies, demonstrated that images obtained can be reliably and accurately graded by certified nonphysician trained readers (TRs).^{7–9} Assessment of a standard set of five retinal images and one pupil image was performed by two TRs independently, and discrepancies were adjudicated by an ophthalmologist reading center director. The system, using double grading with masked TRs and adjudication of discrepancies, yielded high sensitivity (90%) and specificity (87%) for identifying RW-ROP.⁷ The present study investigated whether image grading by a single reader could achieve similar sensitivity and specificity to that of the double grading with adjudication.

Methods

This is a secondary analysis of data from the e-ROP study, conducted from July 1, 2010, to June 30, 2014,^{7–9} which evaluated the validity of a telemedicine system to identify RW-ROP in infants with birth weights of <1251 g. This study was approved by the Institutional Review Board of each of the participating clinical centers, and conformed to the requirements of the US Health Insurance Portability and Accountability Act of 1996. Infants underwent regularly scheduled diagnostic examinations by e-ROP certified ophthalmologists and concurrent digital imaging by trained nonphysician imagers using a widefield RetCam digital camera (Clarity Medical Systems, Pleasanton, CA). Ophthalmologists documented examination findings consistent with RW-ROP (plus disease, stage 3 ROP, zone I ROP), and imagers acquired standard six-image sets from each eye at the bedside. Image sets were uploaded to a secure server and assigned to TRs for masked grading.

Details of TR training and their image grading for e-ROP study have been reported previously.⁹ The training included didactic lectures, interactive sessions and assigned readings that covered classification of ROP, the e-ROP study protocol, telemedicine

principles, the grading protocol, and current ROP treatments. The TRs then independently viewed and graded training image sets with known ROP grading from a previous ROP study database. Their results were compared to a final consensus grading result of a group of experts. Each grading session was followed by review and discussion of the grading results. The TRs graded an additional 100 ROP RetCam image sets using the electronic form and grading protocol. The training was completed by certifying the TRs when they demonstrated good agreement of 85% or more with the consensus grading of 10 image sets from the e-ROP pilot study. The TRs, who had diverse undergraduate academic backgrounds, were trained and certified by experienced ROP specialists to grade ROP morphology in digital retinal images.

Each image set was graded independently by two TRs using customized image-viewing software with functionalities that could enhance the appearance of the images (magnification, gray tone, brightness, and contrast). The e-ROP grading protocol required assessment of both image quality and key morphological features of ROP, including appearance of the posterior pole vessels, the zone of ROP or vascularization, and the stage of ROP.⁹ Agreement on grading between the two TRs was designated as an “agreed” grading. Discrepant results between two readers were adjudicated by an ophthalmologist reading center director and designated as “adjudicated.” Final gradings, therefore consisted on a mixture of agreed and adjudicated gradings. Reproducibility of grading monitored for temporal drift and contemporaneous variability was high.⁹

Statistical Analysis

This analysis used the grading data from all images sets ($n = 7,808$) from e-ROP infants in contrast to grading data from the selected sample of the image sets used in the e-ROP primary paper ($n = 5,520$).⁷ We calculated the sensitivity and specificity of detecting RW-ROP and its components from each TR’s grading by comparing their grading results to the corresponding eye examination findings at the same session. The sensitivity was calculated as the proportion of positive image gradings when examination indicated presence of RW-ROP; specificity was calculated as the proportion of negative image gradings when examination indicated absence of RW-ROP. The 95% confidence intervals for sensitivity and specificity were calculated, with intereye correlation and correlation from multiple image sessions of an eye adjusted by generalized estimating equations using the sandwich robust estimate of variance.¹⁰ As a comparison, we also calculated the sensitivity and specificity from final two-readers grading with adjudication (ie, both adjudicated and agreed gradings combined). All statistical analyses were performed in SAS 9.4 (SAS Institute Inc, Cary, NC).

Results

A total of 7,808 unique image sets from 1235 infants were graded by 2 of the 3 TRs with 5165 sets graded by TR1, 3787 by TR2 and 6,664 sets by TR3.

Sensitivity and Specificity of Single-reader Grading of RW-ROP

The sensitivity and specificity for RW-ROP from single-reader grading and from final two-readers grading with adjudication are provided in Table 1. For the 5,165 image sets graded by TR1, sensitivity for detecting RW-ROP was 0.75 with specificity at 0.80, both of which were lower than the sensitivity (0.79 [$P = 0.03$]) and specificity (0.83 [$P < 0.001$]) from the final grading. TR2 graded 3,787 image sets with sensitivity for RW-ROP of 0.73 and specificity of 0.82, both lower than the sensitivity (0.77 [$P = 0.02$]) and specificity (0.83 [$P = 0.09$]) from the final grading. TR3 graded 5,516 image sets with a sensitivity of 0.83 and specificity of 0.83 for RW-ROP, which did not differ significantly from final grading in sensitivity (0.83 [$P = 0.78$]) but was slightly lower in specificity (0.84 [$P = 0.02$]).

Sensitivity and Specificity of Single-reader Grading for Each RW-ROP Component

The sensitivity and specificity for detecting individual components of RW-ROP are shown in Table 2 for each TR. Compared to the final grading, TR1 had no significant difference in sensitivity (0.40 vs 0.41, [$P = 0.83$]) for detecting plus disease but did have lower sensitivity for zone I ROP (0.47 vs 0.56 [$P = 0.04$]) and stage 3 ROP (0.71 vs 0.77, $P = 0.002$), whereas their specificities were similar. For TR2, only sensitivity for stage 3 ROP was significantly lower (0.69 vs 0.77 [$P < 0.001$]) and the specificity for each component was similar. For TR3, sensitivity was significantly lower for plus disease (0.23 vs 0.35 [$P < 0.001$]), slightly lower for stage 3 ROP (0.79 vs 0.82 [$P = 0.07$]), and somewhat higher for zone I ROP (0.67 vs 0.60 [$P = 0.09$]) while specificities were similar.

Table 3 provides the sensitivity and specificity of RW-ROP for individual TR gradings stratified by the RW-ROP adjudication status for each pair of two readers. For image sets graded by TR1 and TR2 that did not require adjudication, the sensitivity was 0.66; the specificity, 0.88. When their grading did not agree between the two readers and required adjudication, the sensitivity/specificity was 0.71/0.49 for TR1, 0.78/0.35 for TR2, and 0.77/0.48 for the final adjudication grading. Their difference was not significant ($P = 0.48$ for sensitivity; $P = 0.11$ for specificity). For the pair of TR1 and TR3, their agreed gradings yielded high sensitivity (0.81) and specificity (0.91). The adjudicated gradings also had high sensitivity (0.78 for TR1, 0.83 for TR3, and 0.85 for final grading [$P = 0.10$]) but low specificity (0.31 for TR1, 0.48 for both TR3 and final grading [$P < 0.001$]). For the pair of TR2 and TR3, the agreed grading yielded high sensitivity (0.82) and specificity (0.94), and adjudicated gradings had low sensitivity for TR2 (0.66) and high sensitivity for TR3 (0.91), and final adjudication (0.85) had low specificity (0.35 to 0.47, $P < 0.001$).

Discussion

Our results show that the single-reader grading by a TR had lower overall sensitivity and specificity for detecting RW-ROP than the two-readers grading for the e-ROP standard sets of retinal images. However, there were differences between the TRs with regard to the degree to which the sensitivity and specificity were lowered. In the e-ROP study design, two TRs independently graded an image set, and an experienced ophthalmologist adjudicated discrepant results between TRs in identifying RW-ROP. The loss in sensitivity and specificity of a TR to identify RW-ROP has to be weighed against the unavailability of an

ROP-specialized ophthalmologist in a particular geographic location and, even if they are available, the utility of using them for interpreting ROP screening images. A recent survey reporting responses from medical directors of 847 level III neonatal intensive care units affirmed the insufficient workforce of ROP-experienced ophthalmologists. In addition, 35% of respondents agreed that a retinal imaging device could replace indirect ophthalmology; 25% disagreed, and 40% were neutral. More respondents agreed than disagreed (30% vs 15%) that telemedicine for ROP screening is safe, but most were neutral on the safety issue (55%).¹² In tele-ophthalmology programs for diabetic retinopathy, TRs have been shown to perform as well as or even better than ophthalmologists performing a clinical examination.¹³ The e-ROP study also demonstrated this; when retrospectively evaluating discrepancies between image grading by TRs and clinician findings, a panel of ROP experts determined the results of the image grading to be accurate 70% of the time.¹⁴

For the RW-ROP, two TRs (TR1 and TR2) had lower sensitivity and specificity, whereas TR3 had similar sensitivity and specificity compared to the final adjudicated grading. This result is somewhat surprising, because all of the three readers went through a similar rigorous training and certification program in the e-ROP study and graded images until the completion of the study.⁹ TRs had similar workstations, and all were masked to the clinical eye examination results, gradings of the fellow eye, and all previous gradings of an eye. The TRs were also monitored to follow the grading protocol consistently. However, there were two major differences between the three TRs; TR1 and TR2 had several years' grading experience (12 years and 8 years, resp.) for other retinal diseases (but not for ROP), whereas TR3, who had better sensitivity and specificity results than TR1 and TR2, had no previous grading experience and was recruited a few months later, after e-ROP image grading had commenced. The more experienced graders (TR1 and TR2) went through a training course that lasted longer due to the concurrent development of the grading form and protocol used in assessing study images. TR3, who joined later, went through a relatively seamless yet similar training before being certified for the study. TR3 also graded more images than did the other two graders. Though these results are contrary to expectations from more experienced graders, they demonstrate that a newly recruited, trained, and certified nonphysician grader with no previous experience in grading can consistently and accurately grade ROP morphology for RW-ROP, equal to, if not even slightly better than, experienced TRs.

Among the individual morphological features (plus disease, stage 3 ROP, zone I ROP) that make up RW-ROP, plus disease had the lowest sensitivity across all readers. This difficulty in identifying plus disease is also evident in the markedly reduced sensitivity results when two independent TRs agreed with each other that there was plus disease, but only one-fourth of these results agreed with those of the clinical examination. This is not surprising, because identifying plus disease both in retinal images as well as in clinical examination has been reported to have wide variability even among ROP specialists.¹⁵ Zone I ROP was correctly identified only half the time by two readers, whereas TR3 had a much higher sensitivity, identifying zone I ROP correctly two-thirds of the time. Locating the foveal center on color images of preterm infants is challenging and extremely variable, even among ROP specialists.^{16,17} In our study, even when two independent grading results agreed on the presence of zone I ROP, only half of such results agreed with the clinical examination.

However, the evidence that one reader could identify zone I ROP three-quarters of the time and, more importantly, had a better sensitivity for zone I ROP than the final grading result that included adjudication implies that the results for zone I ROP could be made more accurate. This might be achieved by having the measurements for zone I ROP performed on a mosaic of the image set or by defining zone I ROP on images using some form of internal standard.^{18,19}

The fluctuations in the sensitivity and specificity seen between TRs and between the agreed gradings and adjudicated gradings are not unexpected. In a recent report by Gschließer and colleagues²⁰ 7 ROP experts individually evaluated 260 high quality wide-field digital photographs of 52 preterm infants with gestational ages of <32 weeks or birth weights of <1500 g. Perfect interexpert agreement was 45% for ROP stage 2 and only 17% for plus disease. The agreed results between two TRs in our study, without excluding poor quality and missing images,²¹ did have lower sensitivity for plus disease but higher sensitivity for ROP stage 3 compared to adjudicated gradings.

Study limitations include the small number of eyes with the presence of individual RW-ROP morphologies when stratifying the results by individual readers. Another limitation was that the clinical examination results that were used as the gold standard sometimes overlooked definite components of RW-ROP pathology that were captured by imaging, and this was later affirmed by a panel of ROP experts.¹⁴ A major strength of the e-ROP study is the use of well-trained and certified nonphysician graders in grading standard image sets from a large population of premature infants with low birth weights of <1251 g.

In conclusion, our investigation into whether just one well-trained nonphysician grader can achieve comparable or better results in identifying RW-ROP than double grading with adjudication by an ophthalmologist is encouraging. One major component of the e-ROP grading protocol that we did not investigate is the necessity of an ophthalmologist adjudicator, because this would have required regrading the entire data set. At present, the option of combining TR assessment with ophthalmologist confirmation of results within an ROP telemedicine system that has robust quality assurance is a reasonable approach.

Acknowledgments

Supported by National Eye Institute of the National Institutes of Health, Department of Health and Human Services U10 EY017014 and R21EY025686.

Members of the e-ROP Study Cooperative Group

Office of Study Chair – The Children’s Hospital of Philadelphia: PI: Graham E. Quinn, MD, MSCE; KellWade, MD, Ph.D; Agnieszka Baumritter, MS; Trang B. Duros; Lisa Erbring

Baltimore, MD - Johns Hopkins University PI: Michael X. Repka, MD; Jennifer A. Shepard, CRNP, David Emmert, B.A; C. Mark Herring

Boston, MA – Boston Children’s Hospital PI: Deborah VanderVeen, MD; Suzanne Johnston, MD; Carolyn Wu, MD; Jason Mantagos, MD; Danielle Ledoux, MD; Tamar Winter RN, BSN, IBCLC; Frank Weng; Theresa Mansfield, RN

Columbus, OH - Nationwide Children's Hospital and Ohio State University Hospital PI: Don L. Bremer, MD; Mary Lou McGregor, MD; Catherine Olson Jordan, MD; David L. Rogers, MD; Rae R. Fellows, M.Ed., CCRC; Suzanne Brandt, RNC, BSN; Brenda Mann, RNC, BSN

Durham, NC - Duke University PI: David Wallace, MD; Sharon Freedman, MD; Sarah K Jones; Du Tran-Viet; Rhonda "Michelle" Young

Louisville, KY – University of Louisville PI: Charles C. Barr, MD; Rahul Bhola, MD; Craig Douglas, MD; Peggy Fishman, MD; Michelle Bottorff; Brandi Hubbuch, RN, MSN, NNP-BC; Rachel Keith, PhD

Minneapolis, MN - University of Minnesota PI: Erick D. Bothun, MD; Inge DeBecker, MD; Jill Anderson, MD; Ann Marie Holleschau, BA, CCRP; Nichole E. Miller, MA, RN, NNP; Darla N. Nyquist, MA, RN, NNP Oklahoma City, OK - University of Oklahoma PI: R. Michael Siatkowski, MD; Lucas Trigler, MD; Marilyn Escobedo, MD; Karen Corff, MS, ARNP, NNP-BC; Michelle Huynh, MS, ARNP; Kelli Satnes, MS, ARNP, NNP-BC

Philadelphia, PA – Children's Hospital of Philadelphia PI: Monte D. Mills, MD; Will Anninger MD; Gil Binenbaum, MD MSCE; Graham Quinn, MD, MSCE; Karen A. Karp, BSN; Denise Pearson, COMT

San Antonio, TX - University of Texas Health Science Center PI: Alice Gong, MD; John Stokes, MD; Clio Armitage Harper, MD; Laurie, Weaver; Carmen McHenry, BSN; Kathryn Conner; Rosalind Heemer; Elnora Cokley, RNC;

Salt Lake City, UT - University of Utah PI: Robert Hoffman, MD; David Dries, MD; Katie Jo Farnsworth; Deborah Harrison, MS; Bonnie Carlstrom; Cyrie Ann Frye, CRA, OCT-C

Daniel E, Pan W, Quinn GE, Smith E, Baumritter A, Ying G; the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity (e-ROP) Cooperative Group. Single grading vs double grading with adjudication in the telemedicine approaches to evaluating acute-phase retinopathy of prematurity (e-ROP) study. J AAPOS online supplement.

Nashville TN - Vanderbilt University PI: David Morrison, MD; Sean Donahue, MD; Nancy Benegas, MD; Sandy Owings, COA, CCRP; Sandra Phillips, COT, CRI; Scott Ruark

Calgary, Alberta, Canada - Hospital of the Foothills Medical Center PI: Anna Ells, MD, FRCS; Patrick Mitchell, MD; April Ingram; Rosie Sorbie, RN

Data Coordinating Center– University of Pennsylvania School of Medicine PI: Gui-shuang Ying, Ph.D; Maureen Maguire, PhD; Mary Brightwell-Arnold, BA, SCP; Max Pistilli, MS; Kathleen McWilliams, CCRP; Sandra Harris; Claressa Whearry

Image Reading Center – University of Pennsylvania School of Medicine PI: Ebenezer Daniel, MBBS, MS, MPH; E. Revell Martin, BA; Candace R. Parker Ostroff; Krista Sepielli; Eli Smith

Expert Readers: Antonio Capone, MD -The Vision Research Foundation, Royal Oak, Michigan; G. Baker Hubbard, MD - Emory University School of Medicine, Atlanta, Georgia; Anna Ells, MD, FRCS – University of Calgary Medical Center, Calgary, Alberta, Canada

Image Data Management Center – Inoveon Corporation PI: Peter Lloyd Hildebrand, MD; Kerry Davis; G. Carl Gibson; Regina Hansen

Cost Effectiveness Component – PI: Alex R. Kemper, MD, MPH, MS; Lisa Prosser, PhD

Data Management and Oversight Committee (DMOC) David C. Musch, PhD, MPH - Chair, Stephen P. Christiansen, MD; Ditte J. Hess, CRA; Steven M. Kymes, PhD; Srinivas R. Sadda, MD; Ryan Spaulding, PhD

National Eye Institute - Eleanor B. Schron, PhD, RN, FAAN

References

1. Early Treatment for Retinopathy of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. *Arch Ophthalmol*. 2003; 121:1684–94. [PubMed: 14662586]
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. Fijalkowski N, Zheng LL, Henderson MT, et al. Stanford University Network for Diagnosis of Retinopathy of Prematurity (SUNDROP): five years of screening with telemedicine. *Ophthalmic Surg Lasers Imaging Retina*. 2014; 45:106–13. [PubMed: 24444469]
4. Murthy KR, Murthy PR, Shah DA, Nandan MR, S NH, Benakappa N. Comparison of profile of retinopathy of prematurity in semiurban/rural and urban NICUs in Karnataka, India. *Br J Ophthalmol*. 2013; 97:687–9. [PubMed: 23603485]
5. 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]
6. Skalet AH, Quinn GE, Ying GS, et al. Telemedicine screening for retinopathy of prematurity in developing countries using digital retinal images: a feasibility project. *J AAPOS*. 2008; 12:252–8. [PubMed: 18289897]
7. 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; 132:1178–84. [PubMed: 24970095]
8. Quinn, Graham E., e-ROP Cooperative Group. Telemedicine approaches to evaluating acute-phase retinopathy of prematurity: study design. *Ophthalmic Epidemiol*. 2014; 21:256–67. [PubMed: 24955738]
9. 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]
10. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
11. Zimmer-Galler IE. Telemedicine for retinopathy of prematurity: an evolving paradigm. *JAMA Ophthalmol*. 2016; 134:1270–1. [PubMed: 27657491]
12. Vartanian RJ, Besirli CG, Barks JD, Andrews CA, Musch DC. Trends in the screening and treatment of retinopathy of prematurity. *Pediatrics*. 2017; :139. pii: e20161978. doi: 10.1542/peds.2016-1978

13. Hutchinson A, McIntosh A, Peters J, et al. Effectiveness of screening and monitoring tests for diabetic retinopathy: a systematic review. *Diabet Med*. 2000; 17:495–506. [PubMed: 10972578]
14. Quinn GE, Ells A, Capone A Jr, et al. e-ROP (Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity) Cooperative Group. 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. *JAMA Ophthalmol*. 2016; 134:1263–70. [PubMed: 27657673]
15. Campbell JP, Kalpathy-Cramer J, Erdogmus D, et al. Imaging and Informatics in Retinopathy of Prematurity Research Consortium. Plus disease in retinopathy of prematurity: a continuous spectrum of vascular abnormality as a basis of diagnostic variability. *Ophthalmology*. 2016; 123:2338–44. [PubMed: 27591053]
16. Patel SN, Klufas MA, Ryan MC, et al. Color fundus photography versus fluorescein angiography in identification of the macular center and zone in retinopathy of prematurity. *Am J Ophthalmol*. 2015; 159:950–57. [PubMed: 25637180]
17. Chiang MF, Thyparampil PJ, Rabinowitz D. Interexpert agreement in the identification of macular location in infants at risk for retinopathy of prematurity. *Arch Ophthalmol*. 2010; 128:1153–9. [PubMed: 20837799]
18. Slidsborg C, Forman JL, Fielder AR, et al. Experts do not agree when to treat retinopathy of prematurity based on plus disease. *Br J Ophthalmol*. 2012; 96:549–53. [PubMed: 22174097]
19. Rao R, Jonsson NJ, Ventura C, et al. Plus disease in retinopathy of prematurity: diagnostic impact of field of view. *Retina*. 2012; 32:1148–55. [PubMed: 22466473]
20. 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; 160:553–60. [PubMed: 26004406]
21. Morrison D, Bothun ED, Ying GS, Daniel E, Baumritter A, Quinn G, e-ROP Cooperative Group. Impact of number and quality of retinal images in a telemedicine screening program for ROP: results from the e-ROP study. *J AAPOS*. 2016; 20:481–5. [PubMed: 27702612]

Table 1

Comparison of single grading to final adjudicated grading for referral warranted—retinopathy of prematurity (RW-ROP)

Images grading ^a	Clinical exam positive for RW-ROP			Clinical exam negative for RW-ROP		
	No. image sets	Grading positive	Sensitivity (95% CI)	No. image sets	Grading positive	Sensitivity (95% CI)
All images graded by reader 1						
Reader 1 grading	582	438	0.75 (0.70–0.80)	4583	3670	0.80 (0.78–0.82)
Final grading	582	457	0.79 (0.73–0.83)	4583	3781	0.83 (0.80–0.84)
<i>P</i> value			0.03			<0.001
All images graded by reader 2						
Reader 2 grading	434	316	0.73 (0.6–0.78)	3353	2755	0.82 (0.80, 0.84)
Final grading	434	336	0.77 (0.72–0.82)	3353	2792	0.83 (0.81–0.85)
<i>P</i> value			0.02			0.09
All images graded by reader 3						
Reader 3 grading	612	511	0.83 (0.79–0.87)	6052	5005	0.83 (0.81–0.84)
Final grading	612	509	0.83 (0.79–0.87)	6052	5059	0.84 (0.82–0.85)
<i>P</i> value			0.78			0.02

CI, confidence interval.^aFinal grading = Trained reader agreed gradings that required no adjudication plus discrepant gradings between readers that required adjudication.

Table 2

Comparison of single grading to final grading for individual components of referral-warranted retinopathy of prematurity (RW-ROP)

Image grading ^a	Clinical exam positive for a specific component of RW-ROP			Clinical exam negative for a specific component of RW-ROP		
	No.	Grading positive	Sensitivity (95% CI)	No.	Grading negative	Specificity (95% CI)
Plus disease						
Reader 1 grading	157	63	0.40 (0.31–0.49)	5052	4954	0.98 (0.97–0.99)
Final grading ^a		64	0.41 (0.32–0.50)	5052	4974	0.98 (0.98–0.99)
<i>P</i> value		0.83				0.0499
Reader 2 grading	111	44	0.40 (0.29–0.51)	3713	3617	0.97 (0.97–0.98)
Final grading	111	43	0.39 (0.28–0.50)	3713	3647	0.98 (0.98–0.99)
<i>P</i> value		0.78				<0.001
Reader 3 grading	186	42	0.23 (0.16–0.31)	6537	6408	0.98 (0.97–0.98)
Final grading	186	65	0.35 (0.27–0.43)	6537	6437	0.98 (0.98–0.99)
<i>P</i> value		<0.001				0.04
Zone I ROP						
Reader 1 grading	100	47	0.47 (0.30–0.64)	5106	4829	0.95 (0.94–0.95)
Final grading	100	56	0.56 (0.40–0.71)	5106	4832	0.95 (0.93–0.96)
<i>P</i> value		0.04				0.89
Reader 2 grading	54	26	0.48 (0.33–0.63)	3766	3451	0.92 (0.90–0.93)
Final grading	54	26	0.48 (0.33–0.64)	3766	3517	0.93 (0.92–0.95)
<i>P</i> value		1.00				0.001
Reader 3 grading	114	76	0.67 (0.53–0.78)	6606	6083	0.92 (0.91–0.93)
Final grading	114	68	0.60 (0.47–0.71)	6606	6239	0.94 (0.93–0.95)
<i>P</i> value		0.09				<0.001
Stage 3 or above						
Reader 1 grading	498	356	0.71 (0.66–0.77)	4667	3765	0.81 (0.79–0.83)
Final grading	498	381	0.77 (0.71–0.81)	4667	3907	0.84 (0.82–0.85)
<i>P</i> value		0.002				<0.001
Reader 2 grading	383	266	0.69 (0.63–0.75)	3404	2814	0.83 (0.80–0.85)
Final grading	383	295	0.77 (0.71–0.82)	3404	2890	0.85 (0.83–0.87)
<i>P</i> value		<0.001				<0.001

Image grading ^a	Clinical exam positive for a specific component of RW-ROP			Clinical exam negative for a specific component of RW-ROP		
	No.	Grading positive	Sensitivity (95% CI)	No.	Grading negative	Specificity (95% CI)
Reader 3 grading	515	407	0.79 (0.74–0.83)	6149	5312	0.86 (0.85–0.88)
Final grading	515	420	0.82 (0.77–0.86)	6149	5225	0.85 (0.83–0.87)
<i>P</i> value			0.07			<0.001

CI, confidence interval.

^aFinal grading = Trained Reader agreed gradings that required no adjudication plus discrepant gradings between readers that required adjudication

Table 3

Comparison of grading results from paired gradings stratified by adjudication status for referral-warranted retinopathy of prematurity (RW-ROP)

RW-ROP grading stratified by adjudication status ^a		Clinical exam positive for RW-ROP			Clinical exam negative for RW-ROP		
		No.	Grading positive	Sensitivity (95% CI)	No.	Grading negative	Specificity (95% CI)
Images graded by readers 1 and 2							
Not adjudicated	Agreed results	119	78	0.66 (0.5–0.76)	750	660	0.88 (0.8–0.91)
	Results from reader 1	83	59	0.71 (0.5–0.81)	181	89	0.49 (0.4–0.58)
	Results from reader 2	83	65	0.78 (0.6–0.86)	181	63	0.35 (0.2–0.44)
	Results from final grading	83	64	0.77 (0.6–0.85)	181	86	0.48 (0.3–0.57)
	<i>P</i> value			0.48			0.11
Images graded by readers 1 and 3							
Not adjudicated	Agreed results	183	148	0.81 (0.7–0.87)	2969	2705	0.91 (0.8–0.93)
	Results from reader 1	197	153	0.78 (0.6–0.85)	673	206	0.31 (0.2–0.35)
	Results from reader 3	197	163	0.83 (0.7–0.88)	673	323	0.48 (0.4–0.53)
	Results from final grading	197	167	0.85 (0.7–0.90)	673	320	0.48 (0.4–0.52)
	<i>P</i> value			0.10			<0.001
Images graded by readers 2 and 3							
Not adjudicated	Agreed results	122	100	0.82 (0.7–0.89)	1931	1811	0.94 (0.9–0.95)
	Results from reader 2	110	73	0.66 (0.5–0.75)	479	209	0.44 (0.3–0.49)
	Results from reader 3	110	100	0.91 (0.8–0.95)	479	166	0.35 (0.3–0.40)
	Results from final grading	110	94	0.85 (0.7–0.91)	479	223	0.47 (0.4–0.52)
	<i>P</i> value			0.003			<0.001

^aNot adjudicated = image sets requiring no adjudication because two readers agreed on their grading results; adjudicated = image sets required adjudication for discrepant gradings between two readers.