



HHS Public Access

Author manuscript

JAMA Ophthalmol. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

JAMA Ophthalmol. 2018 May 01; 136(5): 498–504. doi:10.1001/jamaophthalmol.2018.0649.

Diagnostic Accuracy of Ophthalmoscopy vs Telemedicine in Examinations for Retinopathy of Prematurity

Hilal Biten, MD, Travis K. Redd, MD, MPH, Chace Moleta, MD, MS, J. Peter Campbell, MD, MPH, Susan Ostmo, MS, Karyn Jonas, MS, RN, R. V. Paul Chan, MD, Michael F. Chiang, MD, and for the Imaging & Informatics in Retinopathy of Prematurity (ROP) Research Consortium

Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, Portland (Biten, Redd, Moleta, Campbell, Ostmo, Chiang); Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago (Jonas, Chan); Center for Global Health, College of Medicine, University of Illinois at Chicago (Chan); Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland (Chiang)

Abstract

IMPORTANCE—Examinations for retinopathy of prematurity (ROP) are typically performed using binocular indirect ophthalmoscopy. Telemedicine studies have traditionally assessed the accuracy of telemedicine compared with ophthalmoscopy as a criterion standard. However, it is not known whether ophthalmoscopy is truly more accurate than telemedicine.

OBJECTIVE—To directly compare the accuracy and sensitivity of ophthalmoscopy vs telemedicine in diagnosing ROP using a consensus reference standard.

DESIGN, SETTING, AND PARTICIPANTS—This multicenter prospective study conducted between July 1, 2011, and November 30, 2014, at 7 neonatal intensive care units and academic ophthalmology departments in the United States and Mexico included 281 premature infants who met the screening criteria for ROP.

Corresponding Author: Michael F. Chiang, MD, Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, 3375 SW Terwilliger Blvd, Portland, OR 97239 (chiangm@ohsu.edu).

Group Information: The Imaging & Informatics in Retinopathy of Prematurity (ROP) Research Consortium members are listed at the end of this article.

Author Contributions: Drs Biten and Redd contributed equally to this article. Dr Chiang had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Biten, Campbell, Jonas, Chan, Chiang.

Acquisition, analysis, or interpretation of data: Biten, Redd, Moleta, Ostmo, Jonas, Chan, Chiang.

Drafting of the manuscript: Biten, Redd, Moleta, Campbell, Ostmo, Chan.

Critical revision of the manuscript for important intellectual content: Biten, Redd, Moleta, Campbell, Jonas, Chan, Chiang.

Obtained funding: Chiang.

Administrative, technical, or material support: Redd, Moleta, Ostmo, Jonas, Chiang.

Study supervision: Campbell, Chan, Chiang.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Chan reported serving as a consultant for Visunex Medical Systems. Dr Chiang reported serving as an unpaid member of the Scientific Advisory Board for Clarity Medical Systems and as a consultant for Novartis. No other disclosures were reported.

EXPOSURES—Each examination consisted of 1 eye undergoing binocular indirect ophthalmoscopy by an experienced clinician followed by remote image review of wide-angle fundus photographs by 3 independent telemedicine graders.

MAIN OUTCOMES AND MEASURES—Results of both examination methods were combined into a consensus reference standard diagnosis. The agreement of both ophthalmoscopy and telemedicine was compared with this standard, using percentage agreement and weighted κ statistics.

RESULTS—Among the 281 infants in the study (127 girls and 154 boys; mean [SD] gestational age, 27.1 [2.4] weeks), a total of 1553 eye examinations were classified using both ophthalmoscopy and telemedicine. Ophthalmoscopy and telemedicine each had similar sensitivity for zone I disease (78% [95% CI, 71%–84%] vs 78% [95% CI, 73%–83%]; $P > .99$ [n = 165]), plus disease (74% [95% CI, 61%–87%] vs 79% [95% CI, 72%–86%]; $P = .41$ [n = 50]), and type 2 ROP (stage 3, zone I, or plus disease: 86% [95% CI, 80%–92%] vs 79% [95% CI, 75%–83%]; $P = .10$ [n = 251]), but ophthalmoscopy was slightly more sensitive in identifying stage 3 disease (85% [95% CI, 79%–91%] vs 73% [95% CI, 67%–78%]; $P = .004$ [n = 136]).

CONCLUSIONS AND RELEVANCE—No difference was found in overall accuracy between ophthalmoscopy and telemedicine for the detection of clinically significant ROP, although, on average, ophthalmoscopy had slightly higher accuracy for the diagnosis of zone III and stage 3 ROP. With the caveat that there was variable accuracy between examiners using both modalities, these results support the use of telemedicine for the diagnosis of clinically significant ROP.

Retinopathy of prematurity (ROP) is a leading cause of childhood blindness worldwide,^{1–4} and its effect on public health continues to grow as advances in perinatal medicine allow for improved survival of premature infants.^{5–7} Retinopathy of prematurity is amenable to screening interventions, as it is detectable before it causes loss of vision and prompt recognition and treatment can delay or reverse adverse outcomes.^{8–11} As a result, the American Academy of Pediatrics, American Academy of Ophthalmology, American Association for Pediatric Ophthalmology and Strabismus, and American Association of Certified Orthoptists have issued a joint policy statement detailing guidelines for ROP screening.¹² The consensus statement specifies that all infants who meet the screening criteria should undergo dilated retinal examination using binocular indirect ophthalmoscopy.

Unfortunately, a lack of access to trained ophthalmologists with experience diagnosing ROP via ophthalmoscopy prevents many premature infants from receiving adequate screening, both in developed and underdeveloped countries.^{13–17} Telemedical screening via remote review of dilated ophthalmoscopic images has been proposed as a substitute for ophthalmoscopy to address this gap, and the use of telemedicine as a substitute for bedside ophthalmoscopy in real-world diagnosis of ROP is increasing.^{18–21}

Prior studies have demonstrated that telemedicine is highly accurate in identifying clinically significant (type 2 or worse) ROP.^{22–31} These studies have established the accuracy of telemedicine as a screening tool using ophthalmoscopy as the reference standard. However, prior work has suggested that there may be significant variability in ROP categorization via ophthalmoscopy, even among experts who are highly experienced in the disease.³² Furthermore, numerous studies have suggested that critical aspects of ROP diagnosis, such

as identification of plus disease and zone I disease, have significant variability among different experts.^{33–39} By definition, a criterion standard must have complete accuracy and consensus.⁴⁰ This definition raises questions about the design of prior studies that examined the accuracy of telemedicine for ROP examination. To our knowledge, little published literature has directly compared the accuracy of ophthalmoscopy with that of telemedicine for ROP diagnosis, without assuming that ophthalmoscopy is the criterion standard.^{31,35} This fact is important not only to better understand the accuracy of ROP diagnosis but also to improve the design of future studies involving emerging diagnostic technologies across other ophthalmic diseases.

The purpose of this study is to directly compare the accuracy of ophthalmoscopy with that of telemedicine for ROP diagnosis in a large data set. To our knowledge, this is the first study to have examined this question in patients with ROP. This comparison is done by developing a consensus reference standard diagnosis (RSD) and by comparing both telemedicine and ophthalmoscopy with this new reference standard to determine the relative accuracy and sensitivity of each for diagnosing ROP.^{40–42}

Methods

Study Population

The study was conducted as part of the multicenter prospective Imaging and Informatics in ROP study. All data were collected prospectively from 7 participating institutions: Oregon Health & Science University, Weill Cornell Medical College, University of Miami, Columbia University Medical Center, Children's Hospital Los Angeles, Cedars-Sinai Medical Center, and Asociación para Evitar la Ceguera en México. Inclusion criteria were infants who were either admitted to a participating neonatal intensive care unit or were transferred to a participating center for specialized ophthalmic care between July 1, 2011, and November 30, 2014; met published criteria for a screening examination for ROP; and had parents who provided written informed consent for data collection. Images were deidentified for analysis and were labeled only with birth weight, gestational age, and postmenstrual age at the time of examination. This study was conducted in accordance with Health Insurance Portability and Accountability Act guidelines; prospectively obtained institutional review board approval from Oregon Health & Science University, Weill Cornell Medical College, University of Miami, Columbia University Medical Center, Children's Hospital Los Angeles, Cedars-Sinai Medical Center, and Asociación para Evitar la Ceguera en México; and adhered to the tenets of the Declaration of Helsinki.⁴³

Clinical Grading and Image Acquisition

In accordance with current guidelines for ROP screening at each institution, all infants underwent serial dilated ophthalmoscopic examinations by a participating ophthalmologist (R.V.P.C. and M.F.C.). The clinical examination findings were obtained using ophthalmoscopy and documented according to the international classification of ROP.⁴⁴ Findings at each examination were incorporated into an overall disease category, based on specifications from the Multicenter Trial of Cryotherapy for Retinopathy of Prematurity Study¹⁰ and from the Early Treatment for Retinopathy of Prematurity (ETROP) study⁹: (1)

no ROP; (2) mild ROP, defined as ROP less than type 2 disease; (3) type 2 ROP (zone I, stage 1 or 2, without plus disease; or zone II, stage 3, without plus disease; or any ROP less than type 1 but with preplus disease); and (4) type 1 ROP or ROP requiring treatment (zone I, any stage, with plus disease; zone I, stage 3, without plus disease; or zone II, stage 2 or 3, with plus disease). Examinations were performed by experienced, board-certified ophthalmologists (R.V.P.C. and M.F.C.) who had undergone specialty training in either pediatric ophthalmology or vitreoretinal surgery, and all were either principal investigators or certified investigators in the ETROP study or had published more than 10 peer-reviewed articles on ROP.

Retinal images were captured by an ophthalmologist or trained photographer after the clinical examination using a wide-angle camera (RetCam; Clarity Medical Systems). De-identified clinical and image data were uploaded to a secure web-based database system developed by us. Cases with clinical diagnoses of stage 4 or 5 ROP were excluded to focus on identification of the onset of clinically significant disease. For analysis of the diagnostic accuracy of ophthalmoscopy, owing to small numbers we excluded 2 participating examiners with less than 50 examinations. Thus, the accuracy of ophthalmoscopy was evaluated for 5 participating clinicians.

Telemedical Image Reading

Three experts independently conducted remote, dilated ophthalmoscopic image review and interpretation of all images via a secure socket layer–encrypted web-based module developed by us. Two of the three experts (R.V.P.C. and M.F.C.) had more than 10 years of clinical ROP experience and more than 50 ROP-related publications. The third expert (S.O.) was a non-physician ROP study coordinator who had previously helped validate a published computer-based ROP severity scale with very high accuracy and intraexpert reliability.⁴⁵ Images were graded according to the same criteria as ophthalmoscopic examinations.

Development and Rationale for Reference Standards

The overall RSD was developed by integration of the telemedicine diagnoses of all 3 image readers with the ophthalmoscopic diagnosis of the examining ophthalmologist using previously published methods.⁴⁶ In cases of discrepancy between image-based and clinical diagnoses, the 3 image readers and a moderator (K.J.) reviewed all medical records to reach a consensus for the overall reference standard. If no consensus could be obtained owing to lack of confirmatory information in photographs (such as ophthalmoscopic diagnosis of zone 3 not clearly seen in photographs), preference was given to the ophthalmoscopic diagnosis. The rationale for this reference standard is that a more accurate diagnosis may be possible by combining ophthalmoscopic and telemedicine findings, and that such a standard may be feasible and applicable in rigorous clinical research settings. Previous work has shown that this approach for developing an RSD can provide higher accuracy and intergrader agreement than either telemedicine or ophthalmoscopy alone.⁴⁶

Statistical Analysis

We compared ophthalmoscopic and telemedicine diagnoses against the RSD as the criterion standard for each ordinal subcategory of zone (I–III), stage (0–3), plus (none, preplus, or

plus), and overall disease category (no ROP, mild ROP, type 2 ROP, or type 1 ROP). This agreement was also calculated for clinically significant binary classifications of zone (zone I vs not), stage (stage 3 vs not), vascular morphologic characteristics (plus disease vs not), and presence of ROP warranting referral to an ophthalmologist (type 2 or worse ROP). All agreements were reported as absolute agreement (compared with a *t* test) and as κ statistic for chance-adjusted agreement. Interpretation of the κ statistic used a commonly accepted scale (0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, near perfect agreement). To analyze the homogeneity of the distributions of categorical data, descriptive statistics and χ^2 testing was used. No adjustment was made for including both eyes of an individual in some cases. We used Excel 2011 (Microsoft Corp) for data management and Stata/SE, version 11 (Stata Corp) for all statistical analysis. $P < .05$ was considered significant. All reported P values were 2-sided.

Results

A total of 281 infants met the eligibility criteria for this study and underwent serial examinations in accordance with current guidelines for ROP screening (mean, 3.7 examinations per infant; range, 1–14).^{12,32} The mean (SD) gestational age was 27.1 (2.4) weeks, and 127 infants (45.2%) were female. A total of 127 infants (45.2%) were white, 74 (26.3%) were Asian, 73 (26.0%) were African American, 4 (1.4%) were Hispanic, and 3 (1.1%) were of undisclosed ethnicity. All screening sessions included an evaluation of each eye, for a total of 1576 study eye examinations. Twenty-three eye examinations were excluded owing to a clinical diagnosis of stage 4 or 5 ROP, yielding 1553 total study eye examinations for comparison.

The distribution of examination findings and disease severity for both telemedicine and ophthalmoscopy is described in Table 1. The RSD identified the presence of ROP in 913 examinations (58.8%); this included mild ROP in 512 examinations (33.0%), type 2 or preplus disease in 313 examinations (20.2%), and type 1 ROP in 88 examinations (5.7%). All telemedicine graders appeared to identify preplus disease more frequently than ophthalmoscopic examiners. In addition, zone III disease was more frequently diagnosed by ophthalmoscopy than telemedicine.

Table 2 describes the accuracy of individual telemedicine graders and ophthalmoscopic examiners compared with the RSD for each ordinal subcategory of zone, stage, plus disease, and overall disease category. There was statistically significant intergrader variability in diagnostic accuracy, regardless of examination method (Figure). This variability among graders was statistically significant for all categories of examination findings and disease classification except for zone among telemedicine graders. As a group, examiners using telemedicine were slightly more accurate than those using ophthalmoscopy in identifying normal vs preplus vs plus disease (92% vs 88%; $P < .001$) compared with the RSD. However, ophthalmoscopy was more accurate in identifying zone (91% [95% CI, 89%–92%] vs 88% [95% CI, 87%–89%]; $P = .009$), stage (88% [95% CI, 86%–89%] vs 75% [95% CI, 74%–77%]; $P < .001$), and category (84% [95% CI, 82%–85%] vs 77% [95% CI, 76%–79%]; $P < .001$).

Table 3 displays sensitivity for the detection of clinically significant disease (ie, type 2 or worse ROP). There were no statistically significant differences between telemedicine and ophthalmoscopy in detecting zone I disease (78% [95% CI, 73%–83%] vs 78% [95% CI, 71%–84%]; $P > .99$), plus disease (79% [95% CI, 72%–86%] vs 74% [95% CI, 61%–87%]; $P = .41$), or type 2 or worse disease (79% [95% CI, 75%–83%] vs 86% [95% CI, 80%–92%]; $P = .10$). However, ophthalmoscopy did have a statistically significantly higher sensitivity for the detection of stage 3 disease (85% [95% CI, 79%–91%]) vs telemedicine (73% [95% CI, 67%–78%]; $P = .004$).

Discussion

This study analyzed the accuracy and sensitivity of telemedicine grading of dilated fundus imaging vs binocular indirect ophthalmoscopy for ROP examination, compared with a consensus reference diagnosis. Key findings were: (1) there was no statistically significant difference in the sensitivity of ophthalmoscopy vs telemedicine to detect clinically significant (type 2 or worse) ROP; (2) ophthalmoscopy had slightly higher accuracy than telemedicine for detecting zone III and stage 3 ROP; and (3) there was statistically significant interobserver variability in the accuracy of ROP classification, regardless of examination method.

In this study, both examination methods had greater accuracy and sensitivity relative to the other for particular aspects of the ROP examination. Specifically, ophthalmoscopy was slightly more accurate in identifying zone, stage, and category of ROP. With respect to disease stage in particular, on average, ophthalmoscopy had greater accuracy for identifying stage 3 disease. It may be that the stereopsis afforded by ophthalmoscopic examination allows better visualization of the 3-dimensional nature of fibrovascular proliferation into the vitreous that occurs in this stage. However, image grader 3 demonstrated comparable accuracy using telemedicine. Image grader 3 is a nonophthalmologist who has been trained exclusively on fundus images, which suggests that there may be non-stereoscopic cues in the images that a trained grader can use to achieve similar accuracy. In addition, ophthalmoscopy was much more accurate for detecting zone III disease in this study. In fact, telemedicine graders almost never identified zone III disease, which is likely owing to the inability to visualize the far temporal retina via wide-field fundus photography in infants. This finding indicates that in some cases ophthalmoscopy may be required to detect pathologic characteristics, although in the absence of plus disease, zone III disease is unlikely to be clinically significant.⁹

There was no difference in sensitivity between ophthalmoscopy and telemedicine for detecting type 2 or worse ROP, which would typically require referral to an ophthalmologist. This finding is consistent with similar studies in which ophthalmoscopy was used as the criterion standard for telemedicine comparison.^{22–31} Based on these findings, the above-mentioned differences between ophthalmoscopy and telemedicine should not preclude implementation of screening programs using telemedicine in situations in which ophthalmoscopy is not readily available, particularly considering the unmet need for screening of at-risk infants.

As in this study, others have previously reported on the marked intergrader variability in diagnosing plus disease, even among highly experienced readers.^{33–36} Diagnosis of plus disease is based on interpretation of venous dilation and arteriolar tortuosity. These are both continuous variables, which are then transformed into a categorical outcome (no plus disease, preplus disease, or plus disease) based on comparison with a single reference standard photograph.¹⁰ This makes diagnosis of plus disease inherently subjective.⁴⁷ Furthermore, it has been shown that in classifying plus disease, examiners focus on different pathologic features and have different interpretations of the same features.³⁷ Consideration of plus disease as a spectrum of disease and quantification using a continuous scoring system rather than a categorical outcome may allow for more accurate diagnosis in the future.³⁸ Limiting the subjective component of diagnosis of plus disease using computer-based image assessment may also allow greater diagnostic uniformity.⁴⁸ Several computer-based image assessment tools have shown efficacy in diagnosing plus disease in ROP.^{49–51} We suggest that future validation of these computerized image assessment programs should use a consensus reference diagnosis as the criterion standard comparison to establish validity.

Limitations

This study has several limitations. Remote image interpretation was performed by only 3 image readers, which may limit generalizability. However, this group was composed of 2 ophthalmologist experts in ROP and 1 nonophthalmologist trained to recognize features of ROP. In some cases, the telemedicine grader and ophthalmoscopic examiner for an examination were the same ophthalmologist. To minimize recall bias, images were generally reviewed several months after acquisition and no patient data, except demographic information such as gestational age, postmenstrual age, and birth weight, were visible during image reading. Because the total of 1553 examinations includes serial assessments of the same infant and the 2 eyes of each infant were regarded as separate study participants, these observations are not truly independent and there was no statistical adjustment to account for this. However, all identifying data were removed and each eye was separately assessed on telemedical review, which should not bias the results to favor ophthalmoscopy or telemedicine preferentially. Finally, a single ophthalmoscopic examination by an expert was performed for each participant. Therefore, the reference standard was based on only 1 ophthalmoscopic examination combined with 3 telemedicine examinations, and the accuracy and reliability of ophthalmoscopic examination is difficult to analyze in this study. Ophthalmoscopic examination has long been considered the criterion standard for ROP diagnosis, and we did not think there was any practical way for infants to undergo multiple masked sequential ophthalmoscopic examinations for this study because of concerns about infant safety.⁵²

Conclusions

Overall, this study demonstrates that binocular indirect ophthalmoscopy and telemedicine using remote review of dilated ophthalmoscopic imaging possess similar sensitivity for detecting type 2 or worse ROP, and that both are limited by interobserver variability. Future

investigations of screening and diagnostic modalities in ROP should use a consensus reference diagnosis as the criterion standard to improve validity.

Acknowledgments

Funding/Support: This work was supported by grants R01 EY19474 and P30 EY010572 from the National Institutes of Health; grants SCH-1622679, SCH-1622542, and SCH-1622536 from the National Science Foundation; and unrestricted departmental funding from Research to Prevent Blindness.

Role of the Funder/Sponsor: The funding sources 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.

Appendix

Group Information: The members of the Imaging & Informatics in Retinopathy of Prematurity (ROP) Research Consortium include Michael F. Chiang, MD, Susan Ostmo, MS, Sang Jin Kim, MD, PhD, Kemal Sonmez, PhD, and J. Peter Campbell, MD, MPH (Oregon Health & Science University, Portland); R.V. Paul Chan, MD, and Karyn Jonas, MS, RN (University of Illinois at Chicago); Jason Horowitz, MD, Osode Coki, RN, Cheryl-Ann Eccles, RN, and Leora Sarna, RN (Columbia University, New York, New York); Anton Orlin, MD (Weill Cornell Medical College, New York, New York); Audina Berrocal, MD, and Catherin Negron, BA (Bascom Palmer Eye Institute, Miami, Florida); Kimberly Denser, MD, Kristi Cumming, RN, Tammy Osentoski, RN, Tammy Check, RN, and Mary Zajechowski, RN (William Beaumont Hospital, Royal Oak, Michigan); Thomas Lee, MD, Evan Kruger, BA, and Kathryn McGovern, MPH (Children's Hospital Los Angeles, Los Angeles, California); Charles Simmons, MD, Raghu Murthy, MD, and Sharon Galvis, NNP (Cedars-Sinai Hospital, Los Angeles, California); Jerome Rotter, MD, Ida Chen, PhD, Xiaohui Li, MD, Kent Taylor, PhD, and Kaye Roll, RN (LA Biomedical Research Institute, Los Angeles, California); Jayashree Kalpathy-Cramer, PhD, Ken Chang, BS, and Andrew Beers, BS (Massachusetts General Hospital, Boston); Deniz Erdogmus, PhD, and Stratis Ioannidis, PhD (Northeastern University, Boston, Massachusetts); and Maria Ana Martinez-Castellanos, MD, Samantha Salinas-Longoria, MD, Rafael Romero, MD, Andrea Arriola, MD, Francisco Olguin-Manriquez, MD, Miroslava Meraz-Gutierrez, MD, Carlos M. Dulanto-Reinoso, MD, and Cristina Montero-Mendoza, MD (Asociación para Evitar la Ceguera en México, Mexico City).

References

1. Muñoz B, West SK. Blindness and visual impairment in the Americas and the Caribbean. *Br J Ophthalmol.* 2002; 86(5):498–504. [PubMed: 11973241]
2. Steinkuller PG, Du L, Gilbert C, Foster A, Collins ML, Coats DK. Childhood blindness. *J AAPOS.* 1999; 3(1):26–32. [PubMed: 10071898]
3. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020—the right to sight. *Bull World Health Organ.* 2001; 79(3):227–232. [PubMed: 11285667]
4. Sommer A, Taylor HR, Ravilla TD, et al. Council of the American Ophthalmological Society. Challenges of ophthalmic care in the developing world. *JAMA Ophthalmol.* 2014; 132(5):640–644. [PubMed: 24604415]

5. Finnström O, Olausson PO, Sedin G, et al. The Swedish national prospective study on extremely low birthweight (ELBW) infants: incidence, mortality, morbidity and survival in relation to level of care. *Acta Paediatr.* 1997; 86(5):503–511. [PubMed: 9183490]
6. National NeoKnowledge Network. Multi-Institutional Comparative Analysis for Births in 1996: Based on 1810 Liveborn Infants <1000 g Birth Weight Wayne, PA: MDS; 1997
7. Strebler R, Bucher HU. Improved chance of survival for very small premature infants in Switzerland [in German]. *Schweiz Med Wochenschr.* 1994; 124(38):1653–1659. [PubMed: 7939532]
8. Palmer EA. Optimal timing of examination for acute retrolental fibroplasia. *Ophthalmology.* 1981; 88(7):662–668. [PubMed: 6894977]
9. 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]
10. Cryotherapy for Retinopathy of Prematurity Cooperative Group. Multicenter trial of cryotherapy for retinopathy of prematurity: preliminary results. *Arch Ophthalmol.* 1988; 106(4):471–479. [PubMed: 2895630]
11. Landers MB III, Semple HC, Ruben JB, Serdahl C. Argon laser photocoagulation for advanced retinopathy of prematurity. *Am J Ophthalmol.* 1990; 110(4):429–431. [PubMed: 2220983]
12. 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(1):189–195. [PubMed: 23277315]
13. Chen Y, Li X. Characteristics of severe retinopathy of prematurity patients in China: a repeat of the first epidemic? *Br J Ophthalmol.* 2006; 90(3):268–271. [PubMed: 16488941]
14. Chen Y, Li XX, Yin H, et al. Beijing ROP Survey Group. Risk factors for retinopathy of prematurity in six neonatal intensive care units in Beijing. *Br J Ophthalmol.* 2008; 92(3):326–330. China [retracted in: *Br J Ophthalmol.* 2008;92(8):1159]. [PubMed: 18303154]
15. Gilbert C. Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early Hum Dev.* 2008; 84(2):77–82. [PubMed: 18234457]
16. Gilbert C, Fielder A, Gordillo L, et al. International NO-ROP Group. Characteristics of infants with severe retinopathy of prematurity in countries with low, moderate, and high levels of development: implications for screening programs. *Pediatrics.* 2005; 115(5):e518–e525. [PubMed: 15805336]
17. Vinekar A, Dogra MR, Sangtam T, Narang A, Gupta A. Retinopathy of prematurity in Asian Indian babies weighing greater than 1250 grams at birth: ten year data from a tertiary care center in a developing country. *Indian J Ophthalmol.* 2007; 55(5):331–336. [PubMed: 17699940]
18. Wang SK, Callaway NF, Wallenstein MB, Henderson MT, Leng T, Moshfeghi DM. SUNDROP: six years of screening for retinopathy of prematurity with telemedicine. *Can J Ophthalmol.* 2015; 50(2):101–106. [PubMed: 25863848]
19. Weaver DT. Telemedicine for retinopathy of prematurity. *Curr Opin Ophthalmol.* 2013; 24(5):425–431. [PubMed: 23872819]
20. Dai S, Chow K, Vincent A. Efficacy of wide-field digital retinal imaging for retinopathy of prematurity screening. *Clin Exp Ophthalmol.* 2011; 39(1):23–29. [PubMed: 20796264]
21. Vartanian RJ, Besirli CG, Barks JD, Andrews CA, Musch DC. Trends in the screening and treatment of retinopathy of prematurity. *Pediatrics.* 2017; 139(1):e20161978. [published online December 14, 2016]. doi: 10.1542/peds.2016-1978 [PubMed: 27974588]
22. Chiang MF, Melia M, Buffenn AN, et al. Detection of clinically significant retinopathy of prematurity using wide-angle digital retinal photography: a report by the American Academy of Ophthalmology. *Ophthalmology.* 2012; 119(6):1272–1280. [PubMed: 22541632]
23. Roth DB, Morales D, Feuer WJ, Hess D, Johnson RA, Flynn JT. Screening for retinopathy of prematurity employing the RetCam 120: sensitivity and specificity. *Arch Ophthalmol.* 2001; 119(2):268–272. [PubMed: 11176990]
24. 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(2):64–70. [PubMed: 11997800]

25. 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]
26. Chiang MF, Keenan JD, Starren J, et al. Accuracy and reliability of remote retinopathy of prematurity diagnosis. *Arch Ophthalmol*. 2006; 124(3):322–327. [PubMed: 16534051]
27. Chiang MF, Keenan JD, Du YE, et al. Assessment of image-based technology: impact of referral cutoff on accuracy and reliability of remote retinopathy of prematurity diagnosis. *AMIA Annu Symp Proc*. 2005; 2005:126–130.
28. Chiang MF, Starren J, Du YE, et al. Remote image based retinopathy of prematurity diagnosis: a receiver operating characteristic analysis of accuracy. *Br J Ophthalmol*. 2006; 90(10):1292–1296. [PubMed: 16613919]
29. Williams SL, Wang L, Kane SA, et al. Telemedical diagnosis of retinopathy of prematurity: accuracy of expert versus non-expert graders. *Br J Ophthalmol*. 2010; 94(3):351–356. [PubMed: 19955195]
30. 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]
31. 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]
32. 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]
33. 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]
34. 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]
35. Campbell JP, Ryan MC, Lore E, et al. Imaging & Informatics in Retinopathy of Prematurity Research Consortium. Diagnostic discrepancies in retinopathy of prematurity classification. *Ophthalmology*. 2016; 123(8):1795–1801. [PubMed: 27238376]
36. 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(9):1153–1159. [PubMed: 20837799]
37. Hewing NJ, Kaufman DR, Chan RVP, Chiang MF. Plus disease in retinopathy of prematurity: qualitative analysis of diagnostic process by experts. *JAMA Ophthalmol*. 2013; 131(8):1026–1032. [PubMed: 23702696]
38. 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(11):2338–2344. [PubMed: 27591053]
39. Kalpathy-Cramer J, Campbell JP, Erdogmus D, et al. Imaging and Informatics in Retinopathy of Prematurity Research Consortium. Plus disease in retinopathy of prematurity: improving diagnosis by ranking disease severity and using quantitative image analysis. *Ophthalmology*. 2016; 123(11): 2345–2351. [PubMed: 27566853]
40. Hripsak G, Heitjan DF. Measuring agreement in medical informatics reliability studies. *J Biomed Inform*. 2002; 35(2):99–110. [PubMed: 12474424]
41. Hripsak G, Wilcox A. Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance. *J Am Med Inform Assoc*. 2002; 9(1):1–15. [PubMed: 11751799]
42. Friedman, CP., Wyatt, JC. *Evaluation Methods in Medical Informatics* New York, NY: Springer-Verlag; 1997
43. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013; 310(20):2191–2194. [PubMed: 24141714]

44. 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]
45. Abbey AM, Besirli CG, Musch DC, et al. Evaluation of screening for retinopathy of prematurity by ROPTool or a lay reader. *Ophthalmology*. 2016; 123(2):385–390. [PubMed: 26681393]
46. Ryan MC, Ostmo S, Jonas K, et al. Development and evaluation of reference standards for image-based telemedicine diagnosis and clinical research studies in ophthalmology. *AMIA Annu Symp Proc*. 2014; 2014:1902–1910. [PubMed: 25954463]
47. Quinn GE. The dilemma of digital imaging in retinopathy of prematurity. *J AAPOS*. 2007; 11(6):529–530. [PubMed: 18086430]
48. Davitt BV, Wallace DK. Plus disease. *Surv Ophthalmol*. 2009; 54(6):663–670. [PubMed: 19665743]
49. Wittenberg LA, Jonsson NJ, Chan RVP, Chiang MF. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus*. 2012; 49(1):11–19. [PubMed: 21366159]
50. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, et al. i-ROP Research Consortium. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the “i-ROP” system and image features associated with expert diagnosis. *Transl Vis Sci Technol*. 2015; 4(6):5.
51. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Imaging and Informatics in ROP (i-ROP) Research Consortium. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. *JAMA Ophthalmol*. 2016; 134(6):651–657. [PubMed: 27077667]
52. 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. [PubMed: 26299381]

Key Points

Question

Is ophthalmoscopy or telemedicine more accurate in diagnosing clinically significant retinopathy of prematurity (ROP) when compared with a reference standard diagnosis?

Findings

In this multicenter study, each method was slightly more accurate for different components of retinopathy of prematurity (zone, stage, and plus disease), but there was no statistically significant difference in their ability to detect clinically significant retinopathy of prematurity. Both methods demonstrate high interexaminer variability.

Meaning

Telemedicine is as effective as ophthalmoscopy in identifying clinically significant retinopathy of prematurity, but both methods demonstrate high interexaminer variability; future studies should use a consensus reference rather than ophthalmoscopy as the criterion standard.

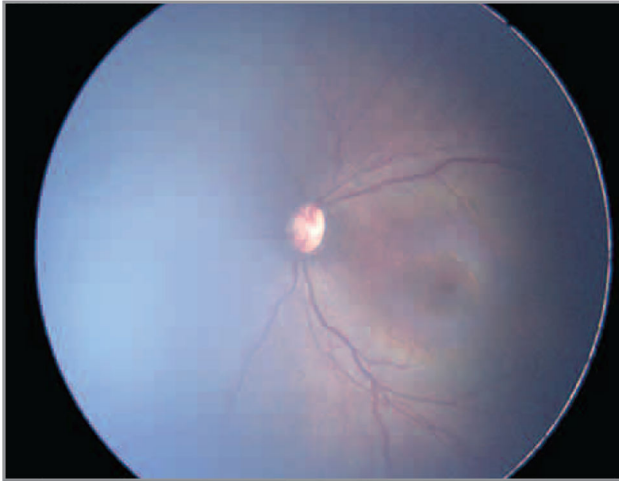
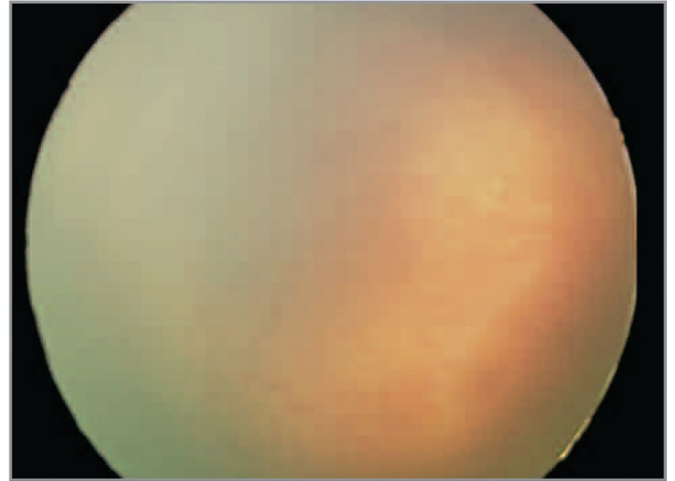
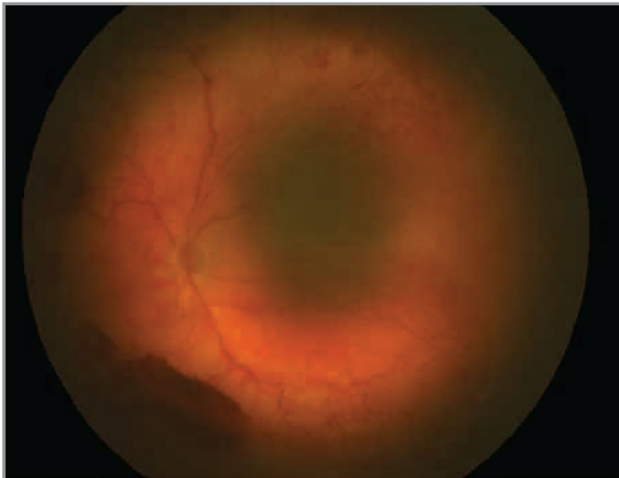
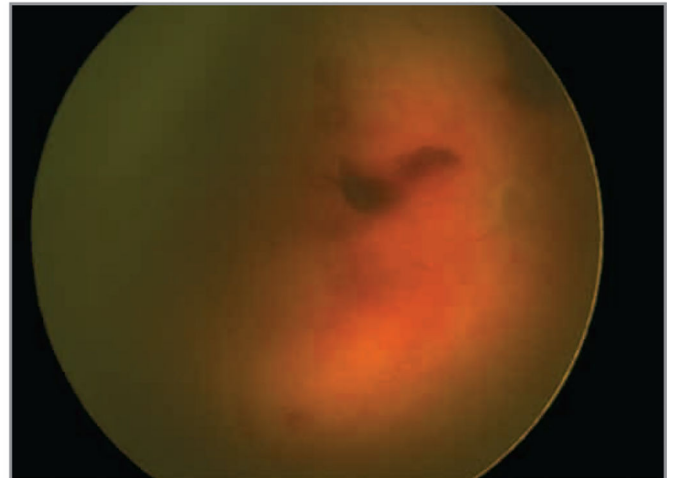
A Type 1 ROP vs mild retinopathy**B** No retinopathy vs mild retinopathy**C** Type 2 ROP vs type 1 ROP**D** Type 1 ROP vs type 2 ROP

Figure. Examples of Errors in Ophthalmoscopic vs Telemedicine Diagnoses in Retinopathy of Prematurity (ROP) Examination

A, Classified as type 1 ROP by ophthalmoscopy (zone I, stage 3, with plus disease), whereas all telemedicine graders classified it as mild retinopathy (zone II, stage 1–2, without plus disease). B, Classified as no retinopathy by ophthalmoscopy (zone III, stage 0, no plus disease), but classified as mild retinopathy according to the reference standard diagnosis (zone II, stage 2, no plus disease). C, Classified as type 2 ROP by 1 telemedicine grader (zone I, stage 2, preplus disease), whereas the ophthalmoscopic and reference standard diagnoses classified it as type 1 ROP (zone I, stage 3, with plus disease). D, Classified as type 1 ROP by 1 telemedicine grader (zone I, stage 3, no plus disease), compared with the reference standard classification of type 2 ROP (zone I, stage 2, no plus disease).

Table 1

Categorization of ROP Examination Findings Using Telemedicine vs Ophthalmoscopic Examination

Categorization	Telemedicine, No. (%) (N = 1553)			Ophthalmoscopy, No. (%) (N = 1553) ^a	RSD, No. (%) (N = 1553)
	Grader 1	Grader 2	Grader 3		
Zone					
I	170 (10.9)	146 (9.4)	171 (11.0)	132 (8.5)	165 (10.6)
II	1382 (89.0)	1403 (90.3)	1375 (88.5)	1204 (77.5)	1272 (81.9)
III	1 (0.06)	4 (0.3)	7 (0.5)	217 (14.0)	116 (7.5)
Stage					
0	533 (34.3)	891 (57.4)	752 (48.4)	713 (46.0)	643 (41.4)
1	551 (35.5)	317 (20.4)	330 (21.2)	385 (24.8)	431 (27.8)
2	368 (23.7)	224 (14.4)	353 (22.7)	321 (20.7)	341 (22.0)
3	98 (6.3)	120 (7.7)	116 (7.5)	134 (8.6)	136 (8.8)
Plus disease					
None	1259 (81.1)	1161 (74.8)	1243 (80.0)	1390 (89.5)	1238 (79.7)
Preplus	229 (14.7)	351 (22.6)	273 (17.6)	104 (6.7)	265 (17.1)
Plus	65 (4.2)	41 (2.6)	37 (2.4)	59 (3.8)	50 (3.2)
Disease category					
No ROP	531 (34.2)	849 (54.7)	737 (47.5)	714 (46.0)	640 (41.2)
Mild ROP	642 (41.3)	297 (19.1)	428 (27.6)	573 (36.9)	512 (33.0)
Type 2 ROP	296 (19.1)	350 (22.5)	327 (21.1)	181 (11.7)	313 (20.2)
Type 1 ROP	84 (5.4)	57 (3.7)	61 (3.9)	85 (5.5)	88 (5.7)

Abbreviations: ROP, retinopathy of prematurity; RSD, reference standard diagnosis.

^a Aggregated results from 5 different expert clinical examiners

Table 2
Comparison of Accuracy in ROP Diagnosis Between Telemedicine and Ophthalmoscopy

Examination	Zone			Stage			Plus Disease			Category		
	No.	RSD Agreement, % (95%CI)	Weighted κ	RSD Agreement, % (95%CI)	Weighted κ	RSD Agreement, % (95%CI)	Weighted κ	RSD Agreement, % (95%CI)	Weighted κ	RSD Agreement, % (95%CI)	Weighted κ	
Telemedicine												
Grader 1	1553	88 (87–90)	0.56	70 (68–73)	0.70	93 (92–94)	0.82	72 (70–74)	0.68			
Grader 2	1553	87 (85–88)	0.48	71 (69–73)	0.69	87 (85–89)	0.68	73 (71–76)	0.70			
Grader 3	1553	89 (88–91)	0.60	85 (84–87)	0.85	97 (96–98)	0.91	86 (85–88)	0.85			
<i>P</i> value ^a	NA	.06	NA	<.001	NA	<.001	NA	<.001	NA	<.001	NA	
Ophthalmoscopy												
Clinician 1	616	95 (93–97)	0.83	94 (92–96)	0.91	95 (94–97)	0.86	94 (92–96)	0.91			
Clinician 2	162	96 (93–99)	0.91	88 (83–94)	0.86	78 (71–84)	0.57	81 (75–87)	0.75			
Clinician 3	82	94 (89–99)	0.89	77 (68–86)	0.69	72 (62–82)	0.13	77 (68–86)	0.68			
Clinician 4	172	97 (95–99)	0.83	89 (85–94)	0.76	97 (95–99)	0.70	92 (88–96)	0.82			
Clinician 5	446	86 (83–90)	0.63	82 (78–86)	0.70	82 (78–85)	0.17	71 (66–75)	0.51			
<i>P</i> value ^a	NA	<.001	NA	<.001	NA	<.001	NA	<.001	NA	<.001	NA	

Abbreviations: NA, not applicable; ROP, retinopathy of prematurity; RSD, reference standard diagnosis.

^aBased on Pearson χ^2 test.

Table 3

Comparison of Diagnostic Sensitivity Between Telemedicine and Ophthalmoscopy

Characteristic	% (95% CI)			
	Zone I Disease	Plus Disease	Stage 3 Disease	Type 2 or Worse
Telemedicine	78 (73–83)	79 (72–86)	73 (67–78)	79 (75–83)
Ophthalmoscopy	78 (71–84)	74 (61–87)	85 (79–91)	86 (80–92)
<i>P</i> value	>.99	.41	.004	.10

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript