Published in final edited form as:
J AAPOS. 2011 December ; 15(6): 573–578. doi:10.1016/j.jaapos.2011.06.011.

Accuracy of retinopathy of prematurity image-based diagnosis by pediatric ophthalmology fellows: Implications for training

Jane S. Myung, MD[a], **R.V. Paul Chan, MD**[a], **Michael J. Espiritu, MD**[a], **Steven L. Williams, MD**[b], **David B. Granet, MD**[c], **Thomas C. Lee, MD**[d], **David J. Weissgold, MD**[e], and **Michael F. Chiang, MD**[f]

[a]Department of Ophthalmology, Weill-Cornell Medical Center, New York, New York

[b]New England Eye Center, Tufts Medical Center, Tufts University, Boston, Massachusetts

[c]Department of Ophthalmology, University of California-San Diego, San Diego, California

[d]The Vision Center, Children's Hospital Los Angeles, Los Angeles, California

[e]Retina Center of Vermont and Division of Ophthalmology, College of Medicine, University of Vermont, Burlington, Vermont

[f]Department of Ophthalmology and Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon

Abstract

Purpose—To measure the accuracy of image-based retinopathy of prematurity (ROP) diagnosis by pediatric ophthalmology fellows.

Methods—This was a comparative case series of expert versus nonexpert clinicians in image-based ROP diagnosis. An atlas of 804 retinal images was captured from 248 eyes of 67 premature infants with a wide-angle camera (RetCam-II, Clarity Medical Systems, Pleasanton, CA). Images were uploaded to a study website from which an expert pediatric retinal specialist and five pediatric ophthalmology fellows independently provided a diagnosis (no ROP, mild ROP, type 2 ROP, or treatment-requiring ROP) for each eye. Two different retinal specialists experienced in ROP examination served as additional controls. Primary outcome measures were sensitivity and specificity of image-based ROP diagnosis by fellows compared to a reference standard of image-based interpretation by the expert pediatric retinal specialist. Secondary outcome measure was intraphysician reliability.

Results—For detection of mild or worse ROP, the mean (range) sensitivity among the five fellows was 0.850 (0.670–0.962) and specificity was 0.919 (0.832–0.964). For detection of type 2 or worse ROP by fellows, mean (range) sensitivity was 0.527 (0.356–0.709) and specificity was 0.938 (0.777–1.000). For detection of treatment-requiring ROP, mean (range) sensitivity was 0.515 (0.267–0.765) and specificity was 0.949 (0.805–1.00).

Reprint requests: Michael F. Chiang, MD, MA, Casey Eye Institute, Oregon Health and Science University, 3375 SW Terwilliger Boulevard, Portland, OR, 97239. (chiangm@ohsu.edu).

Presented at the American Academy of Ophthalmology Annual Meeting, Chicago, IL, October 16-19, 2010 and the American Academy of Pediatric Ophthalmology and Strabismus Annual Meeting, San Diego, CA, March 30-April 3, 2011.

Conclusions—Pediatric ophthalmology fellows in this study demonstrated high diagnostic specificity in image-based ROP diagnosis; however, sensitivity was lower, particularly for clinically significant disease.

Retinopathy of prematurity (ROP) management continues to be challenging for clinicians, particularly due to concerns about logistical difficulties and medicolegal liability. A 2006 survey by the American Academy of Ophthalmology (AAO) found that only half of pediatric ophthalmologists and retinal specialists were willing to manage ROP and that over 20% planned to discontinue managing ROP in the near future.[1] Guidelines on ROP screening jointly published by the AAO, the American Academy of Pediatrics, and the American Association for Pediatric Ophthalmology and Strabismus (AAPOS) in 2006 effectively increased the number of infants requiring ROP screening by raising the recommended gestational age cutoff from 28 weeks to 30 weeks and by suggesting shorter follow-up intervals to avoid missing treatment-requiring disease.[2] Finally, the number of infants at risk for ROP is increasing with rising premature birth rates and neonatal survival rates improving during the past several decades.[3,4]

A recent study showed that many ophthalmologists without pediatric or retinal subspecialty training are performing ROP screening and treatment.[5] Using a Web-based survey, we recently found that up to 25% of ROP examinations are being performed by retina or pediatric ophthalmology fellows without supervision by an attending physician (Wong RK, Ventura CVOC, Espiritu MJ, Lee TC, Chiang MF, Chan RVP. Training fellows for retinopathy of prematurity care: A Web-based survey. J AAPOS 2011;15:e33 [Abstract 122]). These practices raise concerns regarding clinical care and training, particularly since retina fellows do not identify clinically significant ROP as well as expert ophthalmologists.[65] The purpose of this study is to extend research in this area by investigating how well pediatric ophthalmology fellows perform image-based ROP diagnosis.

Methods

This research was conducted under approval of the Columbia University Institutional Review Board and included a waiver of consent for use of de-identified retinal images. Informed consent was obtained from all fellows who participated in the study. All research was performed in compliance with the Health Insurance Portability and Accountability Act of 1996.

Images from consecutive infants whose parents provided informed consent were captured during routine ROP examinations at Columbia University from 2004 to 2005. Retinal images were obtained using a commercially available camera (RetCam-II, Clarity Medical Systems, Pleasanton, CA) by a trained neonatal nurse. Posterior, temporal, and nasal images of each eye were captured along with up to two additional images per eye if they were believed by the photographer to be of diagnostic value.

All study participants viewed wide-angle retinal images on a secure website.[6,7] All images from the right and left eyes were presented simultaneously. The infant's birth weight, gestational age at birth, and postmenstrual age at time of examination were also displayed. A total of 124 image sets (248 eyes) of bilateral retinal examinations from 67 infants were displayed consecutively. Of these 124 examinations, 21 (42 eyes) were randomly selected by the system to be repeated for assessment of intra-physician reliability.

Study subjects were recruited by one author (DBG), who contacted all pediatric ophthalmology fellowship program directors in the United States by email to recruit fellows willing to participate. Fellows were excluded if they did not perform regular ROP

examinations with a faculty member in their training program or if they were fellows visiting from other countries.

Pediatric ophthalmology fellows were oriented to the diagnostic classification of ROP used in this study with a one-page guide developed by the authors. All subjects were asked to diagnose each eye using a four-level system: (1) no ROP; (2) mild ROP, defined as ROP less than type 2 disease; (3) type 2 ROP, defined as (a) zone 1, stage 1 or 2, without plus disease, or (b) as zone 2, stage 3, without plus disease; (4) treatment-requiring ROP, defined as (a) type 1 ROP (zone 1, any stage, with plus disease, or zone 1, stage 3, without plus disease, or zone 2, stage 2 or 3, with plus disease) or (b) threshold ROP (at least 5 contiguous or 8 noncontiguous clock hours of stage 3 in zone 1 or 2, with plus disease). Respondents had the option to provide a diagnosis of "unknown" if they were uncomfortable making a diagnosis based on the data provided. There was no time limit for image interpretation.

As a diagnostic reference standard, the Web-based images evaluated by the five pediatric ophthalmology fellows were also reviewed independently by an expert pediatric retinal specialist (TCL). The expert reference standard physician has over 10 years of experience in ROP examination and treatment at tertiary care centers, has served as a principal investigator in the ETROP study, and has previously coauthored numerous peer-reviewed papers involving ROP. For comparison, as well as validation of the expert reference standard, two additional retina specialists with extensive ROP experience (DJW, RVPC) also reviewed the Web-based images as study experts.

Web-based diagnostic evaluations by an expert pediatric retinal specialist (TCL) were used as the reference standard examination diagnosis. Sensitivity, specificity, and area under the curve (AUC) of receiver operating characteristic (ROC) plots were calculated for detection of mild or worse ROP, type 2 or worse ROP, and treatment-requiring ROP by each fellow.[8] Similar analysis on the two additional expert retinal specialists (DJW, RVPC) was performed for comparison as previously described.[11]

Sensitivity and specificity were analyzed to determine whether the physicians' accuracy improved significantly as they performed more diagnostic examinations. At each cutoff value, logistic regression analysis was used to determine whether physician accuracy improved or worsened as they performed more diagnostic examinations. For cases of no disease, an odds ratio <1 would mean that positive responses were less likely as more diagnoses were performed (ie, improved specificity). For cases of disease, an odds ratio >1 would mean that positive responses were more likely as more diagnoses were performed (ie, improved sensitivity).

Intraphysician reliability was determined using the κ statistic for chance-adjusted agreement in diagnosis. A well-known scale was used for interpretation of results (0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, almost perfect agreement.)[14,15] Unknown diagnoses were considered to be incorrect responses during data analysis. Statistical software was used for analysis of data (Excel 2008, Microsoft, Redmond, WA; SPSS 15.0, SPSS Inc, Chicago, IL). Statistical significance was considered to be a 2-sided *P* value <0.05.

Results

Overall, there were five residency-trained, board-eligible ophthalmologists enrolled in pediatric ophthalmology fellowship programs who met study eligibility criteria and consented to participate. All fellows reported that they had minimal or no ROP screening experience during residency training. All were within the first 6 months (mean, 4.4 months)

of their fellowship training at programs where weekly ROP examinations were performed with a faculty member.

Figure 1 reports the distribution of diagnostic responses of five pediatric ophthalmology fellows. The 143 eyes with a reference standard diagnosis of no ROP were diagnosed as no ROP in 83% to 93% of eyes by all fellows. The 28 eyes with a reference standard diagnosis of type 2 ROP were diagnosed as type 2 ROP in <36% of responses by all fellows. The 15 eyes with a reference standard diagnosis of treatment-requiring ROP were diagnosed as treatment-requiring ROP in 27% to 100% of eyes by all fellows.

Table 1 reports the sensitivity, specificity, and ROC area under the curve (AUC) of five pediatric ophthalmology fellows at 3 diagnostic levels. For detection of type 2 or worse ROP among the fellows, mean (range) sensitivity was 0.527 (0.356–0.709), specificity was 0.938 (0.777–1.000), and AUC was 0.732 (0.678–0.762). For detection of treatment-requiring ROP among the fellows, mean (range) sensitivity was 0.515 (0.267–0.765), specificity was 0.949 (0.805–1.00), and AUC was 0.730 (0.633–0.880). Figure 2 displays examples of images from one eye that was frequently misdiagnosed by fellows.

Logistic regression analysis showed that the overall sensitivities and specificities did not increase as more diagnoses were made by the fellows for all levels of ROP; however, when responses were considered individually for sensitivity of detection, 3 of 5 fellows increased in sensitivity for mild or worse ROP, and 1 of 5 fellows increased in sensitivity for type 2 or worse ROP. When responses were considered individually for specificity of detection, 1 of 5 fellows increased in specificity for mild or worse ROP, 1 of 5 fellows increased in specificity for type 2 or worse ROP, and 2 of 5 fellows increased in specificity for treatment-requiring ROP.

Table 2 shows the intraphysician agreement at three diagnostic levels of each fellow when unknown responses are excluded. Overall, all fellows showed substantial or near-perfect agreement for each level of ROP.

Discussion

The key finding of this study is that pediatric ophthalmology fellows generally demonstrated high specificity for imaged-based detection of mild levels of ROP but showed lower diagnostic sensitivity for detecting clinically significant levels of disease (ie, type 2 and treatment-requiring ROP). Among fellows in this study, mean sensitivities for detecting type 2 and treatment-requiring ROP were approximately 50%. Fellows in this study had variable diagnostic performance and a general tendency to under-call clinically significant levels of disease (ie, low sensitivity with high specificity). A large body of research has shown that infants with type 2 disease must be monitored very closely for progression to treatment-requiring disease, whereas infants with treatment-requiring ROP should receive laser photocoagulation or cryotherapy within 48–72 hours of detection.[2,11]

Based on previously published methods, we compared the diagnostic performance of fellows to two additional expert retinal specialists (DJW, RVPC) using the same expert reference standard.[6] For the detection of type 2 or worse ROP for these two additional expert retinal specialists, mean sensitivity was 0.884 and mean specificity was 0.885. For the detection of treatment-requiring ROP, mean sensitivity was 1.000 and mean specificity was 0.908. Mean sensitivity by these two expert retinal specialists was higher than the pediatric ophthalmology fellows ($P = 0.02$ for type 2 or worse, $P = 0.03$ for treatment-requiring ROP). Mean specificity between these two expert retinal specialists and the pediatric ophthalmology fellows for type 2 or worse and treatment-requiring ROP was not statistically different ($P = 0.51$ for type 2 or worse, $P = 0.53$ for treatment-requiring ROP). This supports

the validity of our study methods using image-based ROP diagnosis by showing that expert retinal specialists were more likely to agree with the expert reference standard diagnosis, as would be expected.

To identify reasons for incorrect responses by fellows, the most commonly misdiagnosed images were reviewed to determine the most likely source of error as judged by author consensus (JSM, RVPC, MFC; Table 3). Among the 248 total image sets, 47 (19%) were diagnosed incorrectly ≥3 of the 5 fellows. Of these image sets, 43 (91.5%) had errors in identification of stage. Nineteen (40.4%) had errors in identification of plus disease, and 16 (34%) had errors in identification of zone. This was similar to findings from a previously published study involving retina fellows in which the majority of errors involved identification of stage.[6] Recognition of these frequent sources of error may help guide education and training programs; however, it is important to note that this study was not specifically designed to elucidate reasons for discrepancies between expert retinal specialists and fellows. Additional research designed to precisely define reasons for error may be warranted.

One published study found that ROP examinations are being performed by an equal number of ophthalmologists who have not completed any fellowship training as who have completed fellowship training in pediatric ophthalmology.[5] With regard to ROP education programs, those authors found that 9% of ophthalmologists who were currently screening for ROP reported that their training did not adequately prepare them to do so.[5] In a different survey on ROP education, pediatric ophthalmology and retinal fellows reported that up to 25% of ROP examinations were performed by the fellow alone, without confirmatory examination by an attending ophthalmologist (Wong RK, et al. Training fellows for retinopathy of prematurity care: A Web-based survey). In fact, pediatric ophthalmology fellows in this study and retinal fellows in a previously published study[11] may have even more training and experience in ROP than many ophthalmologists who currently perform ROP screening without fellowship training. Even among recognized ROP experts specializing in pediatric ophthalmology and retina, several studies have shown that there may be important variability in diagnosis of important parameters such as plus disease and zone I disease.[12–14] Taken together, these factors suggest the need for increased emphasis on ROP education in training programs.

There are a number of limitations in this study. First, The number of subjects was very small (n = 5). While recruiting subjects, we found that there were 35 total fellows being trained at the time.[15] Due to either lack of response or lack of ROP training in the fellowship program, only 11 potential subjects were identified. This study included 45% (5/11) of the potential subjects. However, we note that the purpose of this study is to highlight the importance of ROP training rather than to make generalized conclusions about the strengths and weakness of pediatric ophthalmology fellows.

A second limitation is that diagnosis was assessed by interpretation of a standard set of retinal images. This raises a potential concern that study findings may not accurately translate to diagnostic performance with bedside indirect ophthalmoscopy, which remains the gold standard for ROP diagnosis.[2] Retinal imaging may facilitate visualization of the peripheral retina and ensure that all fellows and experts had an opportunity to review the exact same retinal findings, which might decrease diagnostic variability among graders with less experience in evaluating ROP by indirect ophthalmoscopy. Although a study design involving serial ophthalmoscopic examinations by multiple examiners would be more realistic, that may be impractical because of concerns about infant safety.[16–25] The better performance of the expert retinal specialists also appears to validate the methodology.

A third limitation may be that some images may have been difficult to interpret due to relatively low quality (Table 3). Although expert pediatric retinal specialists in this study had higher performance while reviewing the same images, it is not clear whether this difference was because of better understanding of ROP diagnosis or more experience with review of wide-angle RetCam images.

A fourth limitation may be that subjects may have had differences in exposure to ROP examinations and diagnosis during their residency and early fellowship training. To minimize this bias, all pediatric ophthalmology fellows completed this study within the first 6 months of their training.

Our results show that there are important subtleties to the diagnosis of clinically significant ROP that may not be recognized by all trainees. This has important implications for the delivery of ROP care and for the enhancement of ROP training programs.

Acknowledgments

This study was conducted at Weill Cornell Medical College, New York, New York and Columbia University College of Physicians and Surgeons, New York, New York

This study was supported by grant EY19474 from the National Institutes of Health (MFC), by the St. Giles Foundation (RVPC), and by the Research to Prevent Blindness (JSM, RVPC, DBG, MFC). The sponsors of funding organizations had no role in the design or conduct of this research. The authors have no commercial, proprietary, or financial interest in any of the products or companies described in this article. MFC is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA).

References

1. Ocular Surgery News U.S. Edition. [Accessed August 14, 2010] Survey: Physicians being driven away from ROP treatment. 2006. http://www.revophth.com/content/d/retinal_insider/i/1287/c/24797/
2. Section on Ophthalmology, American Academy of Pediatrics, American Academy of Ophthalmology, American Association for Pediatric Ophthalmology and Strabismus. Screening examination of premature infants for retinopathy of prematurity. Pediatrics. 2006; 117:572–6. Erratum in: Pediatrics 2006;118:1324. [PubMed: 16452383]
3. Matthews TJ, Minino AM, Osterman MJ, Strobino DM, Guyer B. Annual summary of vital statistics: 2008. Pediatrics. 2011; 127:146–57. [PubMed: 21173001]
4. Lad EM, Hernandez-Boussard T, Morton JM, Moshfeghi DM. Incidence of retinopathy of prematurity in the United States: 1997 through 2005. Am J Ophthalmol. 2009; 148:451–8. [PubMed: 19541285]
5. Kemper AR, Freedman SF, Wallace DK. Retinopathy of prematurity care: Patterns of care and workforce analysis. J AAPOS. 2008; 12:344–8. [PubMed: 18440256]
6. Chan RVP, Williams SL, Yonekawa Y, et al. Accuracy of retinopathy of prematurity diagnosis by retinal fellows. Retina. 2010; 30:958–65. [PubMed: 20168274]
7. Chiang MF, Wang L, Busuioc M, et al. Telemedical retinopathy of prematurity diagnosis: Accuracy, reliability, and image quality. Arch Ophthalmol. 2007; 125:1531–8. [PubMed: 17998515]
8. Chiang MG, Starren J, Du YE, et al. Remote image based retinopathy of prematurity diagnosis: A receiver operating characteristic analysis of accuracy. Br J Ophthalmol. 2006; 90:1292–6. [PubMed: 16613919]
9. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37–46.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–74. [PubMed: 843571]
11. 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 Opthalmol. 2003; 121:1684–94.

12. Reynolds JD, Dobson V, Quinn GE, et al. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. Arch Ophthalmol. 2002; 120:1470–6. [PubMed: 12427059]

13. Chiang MF, Jiang L, Gelman R, et al. Interexpert agreement of plus disease diagnosis in retinopathy of prematurity. Arch Ophthalmol. 2007; 125:875–80. [PubMed: 17620564]

14. 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; 12:352–6. [PubMed: 18329925]

15. San Francisco Ophthalmology Fellowship Match. [Accessed July 26, 2010] Ophthalmology Fellowship Match Report. http://www.sfmatch.org/fellowship/f_ophthalmology/links.htm

16. Belda S, Pallás CR, De la Cruz J, Tejada P. Screening for retinopathy of prematurity: Is it painful? Biol Neonate. 2004; 86:195–200. [PubMed: 15240989]

17. Kumar H, Nainiwal S, Singha U, et al. Stress induced by screening for retinopathy of prematurity. J Pediatr Ophthalmol Strabismus. 2002; 39:349–50. [PubMed: 12458847]

18. Mukherjee AN, Watts P, Al-Madfai H, et al. Impact of retinopathy of prematurity screening examination on cardiorespiratory indices: A comparison of indirect ophthalmoscopy and retcam imaging. Ophthalmology. 2006; 113:1547–52. [PubMed: 16828505]

19. Moral MT, Caserio S, Pallas C, et al. Pain and stress assessment after retinopathy of prematurity screening examination: A comparison study between indirect ophthalmoscopy and digital fundus imaging. Early Hum Dev. 2008; 84(suppl):S19.

20. Kirchner L, Jeitler V, Pollak A, et al. Must screening examinations for retinopathy of prematurity necessarily be painful? Retina. 2009; 29:586–91. [PubMed: 19262437]

21. Laws DE, Morton C, Weindling M, Clark D. Systemic effects of screening for retinopathy of prematurity. Br J Ophthalmol. 1996; 80:425–8. [PubMed: 8695564]

22. Sun X, Lemyre B, Barrowman N, O'Connor M. Pain management during eye examinations for retinopathy of prematurity in preterm infants: A systematic review. Acta Paediatr. 2010; 99:329–34. [PubMed: 19958293]

23. Kleberg A, Warren I, Norman E, et al. Lower stress responses after Newborn Individualized Developmental Care and Assessment Program care during eye screening examinations for retinopathy of prematurity: A randomized study. Pediatrics. 2008; 121:e1267–78. [PubMed: 18450869]

24. O'Sullivan A, O'Connor M, Brosnahan D, McCreery K, Dempsey EM. Sweeten, soother, and swaddle for retinopathy of prematurity screening: A randomised placebo controlled trial. Arch Dis Child Fetal Neonatal Ed. 2010; 95:F419–22. [PubMed: 20876596]

25. Dhaliwal CA, Wright E, McIntosh N, Dhaliwal K, Fleck BW. Pain in neonates during screening for retinopathy of prematurity using binocular indirect ophthalmoscopy and wide-field digital retinal imaging: A randomised comparison. Arch Dis Child Fetal Neonatal Ed. 2010; 95:F146–8. [PubMed: 19815939]

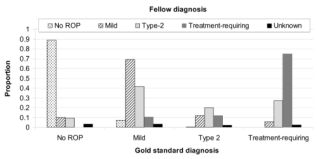

FIG 1.

Mean distribution of ROP (retinopathy of prematurity) diagnoses by pediatric ophthalmology fellows. Of 143 eyes with no ROP, fellows agreed in 127 (89%). Of 62 eyes with mild ROP, fellows agreed in 43 (69%). Of 28 eyes with type 2 ROP, fellows agreed in 6 (20%). Of 15 eyes with treatment-requiring ROP, fellows agreed in 11 (75%).

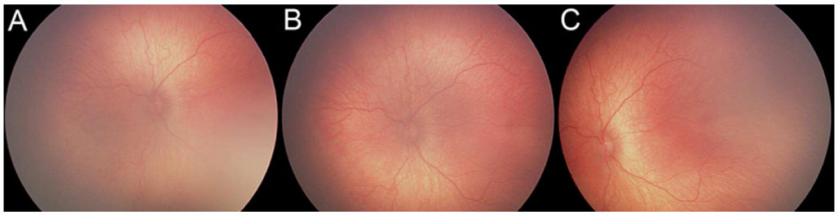

FIG 2.

Example of study images frequently misdiagnosed by fellows. A, B, C, Nasal, posterior, and temporal images diagnosed as type 2 ROP by the expert reference standard and no ROP by 2 of 5 fellows (40%), mild ROP by 1 of 5 fellows (20%), type 2 by 1 of 5 fellows (20%), and treatment-requiring by 1 of 5 of fellows (20%).

Table 1

Sensitivity, specificity, and area under the receiver operating curve (AUC) for retinopathy of prematurity (ROP) diagnosis by five pediatric ophthalmology fellows[a]

	Mild or worse ROP			Type 2 or worse ROP			Treatment-requiring ROP		
Physician	**Sensitivity**	**Specificity**	**AUC**	**Sensitivity**	**Specificity**	**AUC**	**Sensitivity**	**Specificity**	**AUC**
Fellow 1	0.841	0.925	0.871	0.709	0.777	0.743	0.556	0.805	0.680
Fellow 2	0.944	0.915	0.933	0.545	0.990	0.762	0.764	0.996	0.880
Fellow 3	0.670	0.957	0.812	0.356	1.000	0.678	0.324	0.972	0.648
Fellow 4	0.962	0.832	0.933	0.465	0.985	0.725	0.267	1.000	0.633
Fellow 5	0.833	0.964	0.900	0.560	0.939	0.751	0.667	0.974	0.810
Mean	0.850	0.919	0.890	0.527	0.938	0.732	0.515	0.949	0.730

[a]Results are calculated for diagnosis of mild or worse ROP, type-2 or worse ROP, and treatment-requiring ROP compared with the expert reference standard of diagnosis. Results are displayed as proportions, and unknown diagnoses by graders are considered incorrect responses. Unknown diagnoses were provided for 12 (5%) by fellow 1, 2 (1%) by fellow 2, 21 (8%) by fellow 3, 0 (0%) by fellow 4, and 7 (3%) by fellow 5.

Table 2

Intraphysician reliability for ROP diagnosis by the expert reference standard and five pediatric ophthalmology fellows[a]

Physician	Mild or worse ROP	Type 2 or worse ROP	Treatment-requiring ROP
Reference standard	0.904 (0.066)	1.000 (−)	0.730 (0.145)
Fellow 1	0.8835 (0.080)	0.727 (0.113)	1.000 (−)
Fellow 2	1.000 (−)	0.659 (0.128)	0.828 (0.119)
Fellow 3	0.936 (0.063)	0.762 (0.131)	0.742 (0.142)
Fellow 4	0.843 (0.087)	1.000 (−)	1.000 (−)
Fellow 5	0.845 (0.086)	0.750 (0.119)	0.857 (0.099)

[a]Results are displayed as kappa statistic (standard error) for ability to detect mild or worse ROP, type-2 or worse ROP, and treatment-requiring ROP.

J AAPOS. Author manuscript; available in PMC 2012 December 1.

Table 3

Reasons for discrepancy by ROP diagnoses provided by five pediatric ophthalmology fellows, compared to diagnosis provided by expert reference standard. All image sets that were diagnosed incorrectly by over half of the fellows were reviewed to identify the most likely source of error as judged by consensus of study authors.

Reasons for discrepancy	Number of images[a]	Percentage
Identification of stage	43	91.5
Identification of plus disease	19	40.4
Identification of zone	16	34.0
Poor image quality	2	4.3
Total number of images	47[b]	-

[a] Number of images that were diagnosed incorrectly by over half of the fellows. The sum is larger than the total number of images because some images were diagnosed incorrectly by fellows for different reasons.

[b] Total number of images that were diagnosed incorrectly by over half of the fellows

J AAPOS. Author manuscript; available in PMC 2012 December 1.