



Published in final edited form as:

*Ophthalmology*. 2020 August ; 127(8): 1105–1112. doi:10.1016/j.ophtha.2020.01.052.

## Aggressive Posterior Retinopathy of Prematurity: Clinical and Quantitative Imaging Features in a Large North American Cohort

Kellyn N. Bellsmith<sup>1</sup>, James Brown<sup>2,3</sup>, Sang Jin Kim<sup>1,4</sup>, Isaac H. Goldstein<sup>1</sup>, Aaron Coyner<sup>6</sup>, Susan Ostmo<sup>1</sup>, Kishan Gupta<sup>1</sup>, R.V. Paul Chan<sup>5</sup>, Jayashree Kalpathy-Cramer<sup>3</sup>, Michael F. Chiang<sup>1,6</sup>, J. Peter Campbell<sup>1</sup>

<sup>1</sup>Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, Portland, OR, USA

<sup>2</sup>School of Computer Science, University of Lincoln, Lincoln, United Kingdom

<sup>3</sup>Massachusetts General Hospital & Brigham and Women's Hospital Center for Clinical Data Science, Boston, MA, USA

<sup>4</sup>Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

<sup>5</sup>Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago, Chicago, IL, USA

<sup>6</sup>Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA

### Abstract

**Purpose:** Aggressive posterior retinopathy of prematurity (AP-ROP) is a vision-threatening disease with a significant rate of progression to retinal detachment. The purpose of this study was to quantitatively characterize AP-ROP by demographics, rate of disease progression, and a deep learning based vascular severity score.

**Design:** Retrospective analysis

**Subjects:** The Imaging and Informatics in ROP (i-ROP) cohort from 8 North American centers, consisting of 947 total patients and 5945 clinical eye exams with fundus images, was used. Pre-treatment eyes were categorized by disease severity: None, Mild, Type 2 or Pre-Plus, Treatment-

---

Correspondence to: J. Peter Campbell, MD, MPH, Casey Eye Institute, Oregon Health & Science University, 3375 SW Terwilliger Blvd, Portland OR 97239, Phone : 503-494-7891, campbelp@ohsu.edu.

**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 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.

Meeting Presentation: This data has been presented at Association for Research in Vision and Ophthalmology Annual Meeting 2019 and is under consideration for presentation at the American Academy of Pediatric Ophthalmology and Strabismus Annual Meeting 2020.

Aggressive posterior retinopathy of prematurity is a vision-threatening disease with earlier onset and more rapid progression than less severe forms of ROP. Quantitative analysis using deep learning may enable earlier and more objective diagnosis.

Requiring (TR) without AP-ROP, TR with AP-ROP. Analyses compared TR with AP-ROP to TR without AP-ROP to investigate differences between AP-ROP and other TR disease.

**Methods:** A reference standard diagnosis was generated for each eye exam using previously-published methods combining three independent image-based and one ophthalmoscopic gradings. All fundus images were analyzed using a previously-published deep learning system (i-ROP DL) and assigned a score from 1-9. Demographic data, systemic comorbidities, and post-menstrual age were evaluated for each category.

**Main Outcome Measures:** Birth weight, gestational age, post-menstrual age, vascular severity score

**Results:** Infants who developed AP-ROP were more premature by birth weight (617 g vs 679 g,  $p=0.01$ ) and gestational age (24.3 weeks vs 25.0 weeks,  $p<0.01$ ) and reached peak severity at an earlier post-menstrual age (34.7 weeks vs 36.9 weeks,  $p<0.001$ ) compared to TR without AP-ROP. The mean vascular severity score was greatest in TR with AP-ROP compared to TR without AP-ROP (8.79 vs 7.19,  $p<0.001$ ). Analyzing the severity score over time, the rate of progression was fastest in infants who developed AP-ROP ( $p<0.002$  at 30-32 weeks).

**Conclusion—**Premature infants in North America with AP-ROP are born younger and develop disease earlier than infants with less severe ROP. Disease severity is quantifiable with a deep-learning derived severity score, which correlates with clinically identified categories of disease including AP-ROP. The rate of progression to peak severity of disease is greatest in eyes that develop AP-ROP compared to other treatment-requiring eyes. Analysis of quantitative characteristics of AP-ROP may help improve diagnosis and treatment of an aggressive, vision-threatening form of ROP.

## INTRODUCTION

Aggressive posterior retinopathy of prematurity (AP-ROP) was defined by the International Classification of Retinopathy of Prematurity (ICROP) as an “uncommon, rapidly progressing, severe form of ROP” that “usually progresses to stage 5” ROP if untreated.<sup>1</sup> Characteristic features include “posterior location, prominence of plus disease, and the ill-defined nature of the retinopathy.” Specific features of AP-ROP include shunt vessels, large avascular areas within vascularized retina, and flat extraretinal neovascularization which appears different from typical extraretinal neovascularization that forms a ridge at the vascular-avascular border. Despite the formal recognition of AP-ROP as a diagnostic entity in the 2005 ICROP, there continues to be significant disagreement among ROP experts regarding the precise definition and therapeutic implications of the diagnosis of AP-ROP.<sup>2,3,4,5</sup>

Inter-observer differences in ROP diagnosis lead to differences in care between examiners for all components of the ICROP classification (zone, stage, and plus). This has been shown both in research studies evaluating diagnostic agreement, in clinical trials involving ROP experts,<sup>6,7</sup> and is presumably common in the real-world day-to-day care of ROP. Several prior studies have explored diagnostic disagreements in ROP classification, but few have looked specifically at AP-ROP.<sup>3,8,9</sup> The vascular changes that constitute plus disease present on a spectrum, and one likely cause of diagnostic variability is systematic differences in

interpreting diagnostic cut points (e.g. plus vs. pre-plus vs. normal) among examiners.<sup>10,11</sup> Since AP-ROP also presents along a continuum, several authors have suggested a new classification of “pre-AP-ROP” reflecting this continuous progression of disease, however this entity has yet to be formally recognized in the ROP nomenclature.<sup>2</sup> Others completely differentiate between “staged,” or “classic,” ROP and AP-ROP, suggesting that the pathophysiology between the two may be different.<sup>12,13</sup> Finally, although the definition of AP-ROP implies rapid progression, the ICROP does not formally include rate of progression in the diagnostic criteria and to date there has been no method to evaluate this characteristic as a diagnostic feature of AP-ROP.

Recent advances in automated diagnosis of plus disease using deep learning enable objective evaluation of vascular severity along a continuum. We previously demonstrated that a fully-automated deep learning classifier (i-ROP DL) can diagnose plus disease with comparable or better accuracy to experts,<sup>14,15</sup> that a quantitative vascular severity score (1-9 scale) derived using this classifier can identify clinically-significant ROP with high accuracy,<sup>16</sup> and that this quantitative severity score can monitor ROP progression and regression after treatment.<sup>17,18</sup> However, the potential for quantitative diagnosis of AP-ROP with deep learning has not been evaluated. This paper addresses this gap in knowledge by characterizing the clinical features of AP-ROP in a large North American patient cohort, and by applying a deep learning-based ROP severity score to eyes with all stages of ICROP disease severity. Improved methods for identifying and characterizing AP-ROP may result in earlier diagnosis, more consistent diagnosis, and better understanding of disease pathophysiology.

## METHODS

This project was conducted as part of the multicenter Imaging and Informatics in ROP (i-ROP) consortium. This study was approved by the Institutional Review Board at the coordinating center (Oregon Health & Science University), and by each of the 8 participating institutions (Columbia University, Cornell University, University of Illinois at Chicago, William Beaumont Hospital, Children’s Hospital Los Angeles, Cedars-Sinai Medical Center, University of Miami, Asociación para Evitar la Ceguera en México). All institutions abided by the tenets of the Declaration of Helsinki, and written informed consent was obtained from parents of all infants enrolled.

### Dataset and reference standard diagnosis

All neonates who underwent dilated ophthalmoscopic examination per standard ROP screening guidelines at each institution between July 2011 and December 2016 were included in this study. The following demographic data were collected from the study database for analysis: birth weight, gestational age, post-menstrual age (PMA) at time of peak disease severity (if any ROP). In addition, clinical comorbidities were collected for each baby, including history of: any grade of intraventricular hemorrhage (IVH), chronic lung disease (defined as requiring oxygen at 36 weeks PMA), sepsis, necrotizing enterocolitis (NEC), blood transfusion, or peri-ventricular leukomalacia (PVL). We compared the distributions of demographic features for each category using a Tukey test,

and adjusted using the Holm-Bonferroni method. Distribution of race, ethnicity, and treatment type (anti-VEGF, laser, surgery) between categories was also included in the Holm-Bonferroni adjustment.

Each examination included capture of standard five-field retina images using a wide-angle camera (RetCam; Natus Medical Incorporated, Pleasanton, CA). Images were de-identified and each eye examination was given a reference standard diagnosis using previously-published methods,<sup>19</sup> based on combined findings from ophthalmoscopic examination by the examining clinician and image-based examinations by 3 image readers. The reference standard diagnosis was established for all ICROP components: zone (I-III), stage (1-5), plus (plus, pre-plus, no plus), and AP-ROP (yes, no) based on majority vote and/or after group discussion in the absence of a majority. In addition, an overall disease category was assigned for each eye examination: No ROP, Mild ROP (less than type 2 disease and normal vessels), Type 2 or pre-plus disease, treatment-requiring (TR) ROP without AP-ROP, or TR-ROP with AP-ROP. Images were excluded if 2 out of 3 image graders labeled the quality as “unacceptable for diagnosis,” if the presenting clinical diagnosis was stage 4 or 5 ROP with retinal detachment, or if the eye had received treatment previously. Eyes were excluded from the analysis if there was no pre-treatment photo after the above exclusions. Of all the available images of each eye, the image that represented the peak disease severity (pre-treatment if applicable) was selected for quantitative analysis. Statistical analysis was performed using Excel 2016 (Microsoft, Redmond, WA), Stata v.11.0 (College Station, TX), and R version 3.2.2.<sup>20</sup>

### **Cross-sectional and longitudinal analysis of vascular severity in AP-ROP**

The i-ROP DL system was used to analyze each posterior pole image in the dataset and assign an ROP vascular severity score from 1 (normal retinal vasculature) to 9 (severe plus disease) using methods previously described: (1 x probability of no disease + 5 x probability of pre-plus disease + 9 x probability of plus disease).<sup>15</sup> Images were then grouped into the reference standard diagnosis categories described above for cross-sectional analysis. We compared the distributions of vascular severity for each category using a Tukey test, and adjusted using the Holm-Bonferroni method. We then compared longitudinal change in vascular severity in 3 groups: eyes that did not require treatment (No Treatment Received), eyes that required treatment but were not labeled AP-ROP (TR without AP-ROP), and eyes with AP-ROP (TR with AP-ROP). Comparisons between groups were performed using the Wilcoxon rank-sum test.

### **Inter-reader agreement on AP-ROP diagnosis**

A total of 11 image readers participated in the image gradings for the development of the reference standard diagnosis during this time period. Because the prevalence of AP-ROP was low, readers with fewer than 2000 total readings (averaging fewer than 10 diagnoses of AP-ROP) were excluded from analysis for inter-reader agreement. Five readers (3 physicians and 2 non-physician study coordinators) with over 2000 readings each were included in the final inter-reader analysis by Cohen kappa score. The Cohen kappa score was calculated for each pair of graders using online tools (StatsToDo: Kappa (Cohen and Fleiss); [www.statstodo.com](http://www.statstodo.com)) for ordinal data. Cohen kappa values were interpreted using a

commonly-accepted scale: 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1.00, near-perfect agreement.<sup>21,22</sup>

## RESULTS

### Clinical and demographic features of study cohort

Table 1 displays demographic features of the study cohort based on category of peak disease severity and associated comorbidities. A total of 5945 eye examinations from 947 infants (1889 eyes) were analyzed. Sixty-two (3%) of 1889 eyes developed a peak disease severity of TR with AP-ROP. Infants in the TR with AP-ROP category had a lower mean gestational age ( $24.3 \pm \text{SD } 0.9$  weeks,  $p < 0.01$ ) and birth weight ( $617 \pm 119$  grams,  $p < 0.01$ ) than all other disease categories and were treated on average at an earlier PMA ( $34.7 \pm \text{SD } 1.3$  weeks) than eyes with TR without AP-ROP ( $36.9 \pm \text{SD } 2.8$  weeks,  $p < 0.001$ ). There were no infants who developed TR-ROP in the cohort who were born after 26 weeks GA (range 23-26 weeks). Adjusting for gestational age, infants developed TR with AP-ROP at a slightly earlier mean post-natal age ( $8.3 \pm \text{SD } 1.5$  weeks) compared to eyes with TR without AP-ROP ( $11.5 \pm \text{SD } 2.8$  weeks). In general, infants with more severe ROP tended to have more comorbidities. Table 2 demonstrates that type 2 or worse disease was more likely in patients with red blood cell transfusions ( $p < 0.01$ ), TR-ROP was more likely in patients with history of sepsis ( $p < 0.05$ ), and TR-ROP was more common in patients with necrotizing enterocolitis ( $p < 0.05$ ). Compared to all other groups, infants with AP-ROP had higher rates of chronic lung disease ( $p < 0.05$  compared with TR without AP-ROP).

### Cross-sectional analysis of quantitative vascular severity score

Of the 1889 eyes meeting inclusion criteria, 1507 had corresponding images to use for the analysis of vascular severity score. Figure 1 displays a box plot of the median quantitative vascular severity scores for each of the 5 clinical severity groups shown in Table 1. For each ICROP category, a higher ROP vascular severity score was associated with more severe disease, including AP-ROP ( $p < 0.001$ ). The median vascular severity score for the TR with AP-ROP group was 8.8 (interquartile range (IQR) 8.2-9.0), 7.2 (IQR 5.3-8.7) for TR without AP-ROP, 4.3 (IQR 2.2-5.1) for type 2 or pre-plus disease, 1.2 (IQR 1.0-1.8) for mild ROP, and 1.0 (IQR 1.0-1.3) for babies with no ROP ( $p < 0.001$  for all comparisons between all groups).

### Longitudinal analysis of vascular severity score over time

We compared the distribution of ROP vascular severity score between the three cohorts (no treatment received, TR without AP-ROP, and TR with AP-ROP) in Figure 2. As seen in Table 1, eyes that developed TR with AP-ROP developed peak disease earlier than those without AP-ROP (top cohort in black), and were characterized both earlier onset and more rapid progression of disease prior to treatment. The distribution of scores for eyes with TR with AP-ROP were higher than eyes without AP-ROP at all time points ( $p < 0.002$  at 30-32 weeks, and  $p < 0.001$  after 32 weeks). Eyes with TR without AP-ROP had higher vascular severity scores than non-treatment-requiring eyes at all time points after 32 weeks ( $p < 0.002$  at 32-34 weeks and  $p < 0.001$  after 34 weeks).

### Analysis of inter-reader agreement on AP-ROP diagnosis

Table 3 demonstrates the inter-observer agreement for clinician image readers in this study. Reader 1 and 3 showed moderate agreement, however all other pairs of readers showed no or fair agreement (kappa  $-0.01$  to  $0.52$ ). Three of the 5 readers showed substantial agreement with the reference standard diagnosis, which would be expected as it is a composite of 3 individual readers' diagnoses and the clinical exam. However, the mean kappa for each individual reader, representing the overall agreement of a particular reader with other readers and the reference standard diagnosis did not exceed fair agreement for any individual reader (kappa  $0.12$  to  $0.52$ ). Figure 3 displays example images of babies with 3 different categories of disease by reference standard diagnosis: type 2 or pre-plus ROP (3A), TR without AP-ROP (3B), and TR disease with AP-ROP (3C).

## DISCUSSION

In this study, we analyzed a cohort of premature infants in North America to determine the demographics, clinical comorbidities, quantitative vascular severity, and inter-expert diagnostic agreement for patients with AP-ROP. There were several key findings: (1) Premature infants in North America with AP-ROP are born younger, develop disease earlier, and have more chronic lung disease than infants with other categories of ROP. (2) Quantitative evaluation of vascular severity using a deep-learning derived vascular severity score correlates with all ICROP categories of ROP including AP-ROP. (3) Diagnostic agreement on AP-ROP is variable, even among experienced image graders, suggesting the need for more quantitative diagnosis.

The first key finding is that infants in North America with AP-ROP tend to be more premature, develop disease earlier, and have more chronic lung disease than treatment-requiring infants without AP-ROP (Tables 1 and 2). In this cohort of patients from 2011–2016, we did not observe any AP-ROP in babies born after 26 weeks of gestation, whereas TR without AP-ROP babies varied more significantly on birth weight and gestational age. The finding that 80% of the patients with AP-ROP have chronic lung disease may suggest that AP-ROP is at least partially related to a higher total exposure to oxygen than eyes without AP-ROP, since these babies have higher oxygen requirements for a longer time. This is not to suggest that oxygen exposure is unrelated to less severe forms of ROP, but only that the most severe phenotypes may be more common in eyes with the highest exposure and the least developed retinas at birth. It is also possible that the relationship is merely an association of two conditions that increase with extremes of prematurity, rather than a causal relationship between oxygen and AP-ROP.

Another potential explanation for earlier onset of AP-ROP is that post-natal age may be a better predictor of treatment timing than post-menstrual age, which was suggested in a recent paper from Sweden that found peak risk at 12 weeks post-birth, independent of gestational age at birth.<sup>23</sup> It is worth noting here that the epidemiology of AP-ROP varies between regions. In high-income countries, AP-ROP is rare (<5% of ROP cases) and typically only affects the lowest birth weight infants as we found here,<sup>1,24</sup> which is not the case in many low- and middle-income countries (LMIC), where there is high incidence of AP-ROP in older and larger birth weight babies.<sup>25,26</sup> The phenotype of AP-ROP in bigger



babies can, in certain cases, be directly attributable to exposure to unblended (100%) oxygen. In both cases, oxygen concentration and/or total exposure (concentration over time) may be the most important risk factor for development of the phenotype of AP-ROP, which presents earlier and more aggressively than TR without AP-ROP in both populations.<sup>26</sup>

The second key finding is that the full spectrum of ICROP ROP disease severity including AP-ROP is quantifiable with a deep-learning derived automated vascular severity score (Figure 1). This has several potential implications: (1) Incorporating a vascular severity score into the diagnostic criteria for ROP may bring a level of objectivity to the clinical determination of all levels of disease severity. Kalpathy-Cramer et al previously demonstrated that while graders would often disagree on ICROP classification, they tended to agree on relative disease severity.<sup>10</sup> A continuous severity score would allow objective measurement of disease severity that could be put into appropriate clinical context for treatment decisions, as well as studied prospectively to better understand optimal timing of interventions in different phenotypes. (2) By analyzing rate of change of the vascular severity score, we may better understand the kinetics of various ROP phenotypes including AP-ROP. This latter finding also raises the question of whether pace of disease ought to be incorporated into the diagnostic criteria for AP-ROP since both the term “aggressive” as well as the older term “rush” disease have kinetic implications that are not encoded in the current ICROP criteria. Taylor et al and Gupta et al have shown retrospectively how the vascular severity score can differentiate progression to treatment-requiring disease and regression after treatment.<sup>17,18</sup> Data from this study suggest that many of the eyes that will eventually progress to TR with AP-ROP may be identified as early as 2–4 weeks prior to treatment (Figure 2), which may have implications for risk modeling, disease screening, and enable future prospective evaluation of earlier (and more phenotype-specific) treatment thresholds.

The third key finding is that we found significant levels of inter-reader disagreement in diagnosis of AP-ROP (Table 3). On the one hand, it is surprising that colleagues in the i-ROP research consortium with more than 2000 gradings demonstrated only fair agreement for such an important phenotype. On the other hand, since the diagnostic description of AP-ROP is nonspecific (“ill-defined nature of the retinopathy”), plus disease is a necessary (but subjective) component, and both plus disease and AP-ROP evolve along a continuum, the lack of perfect agreement is consistent with what has been reported previously in ROP.<sup>12,25–27</sup> Further complicating precise and accurate phenotypic classification of AP-ROP, Flynn and Chan-Ling described a hybrid form of ROP, with a mixture of more typical ridge pathology (e.g. Figure 3B) and ill-defined flat neovascularization (e.g. Figure 3C).<sup>13,25</sup> The terminology in the literature is also conflicted as to whether the severe posterior disease seen in less premature babies in LMIC (e.g. infants > 30 weeks gestation and > 2000 grams) ought to be called AP-ROP versus “oxygen-induced” ROP, the latter of which is not a formal disease classification in the ICROP, but is a well-recognized entity in LMIC populations, and primarily related to suboptimal oxygen monitoring and excessive oxygen exposure.<sup>28,29</sup> It is worth noting however, that since the ICROP was first established in the 1980s, 30 years after the end of the first epidemic of ROP in the US and Europe,<sup>30</sup> phenotypic variations of ROP related to oxygen may not be reflected within the current ICROP classification system.

There are several limitations to this study. The first is the acknowledgment that there is no gold standard for the diagnosis of AP-ROP. In this study, we utilized a reference standard diagnosis, combining multiple independent, image-based expert readings and the clinical diagnosis by binocular indirect ophthalmoscopy. Given that readers within our study often disagreed on the diagnosis of AP-ROP individually, it is likely that the same would be true with some of the readers of this paper, for many of the reasons mentioned above. We believe that this suggests the need for more objective metrics of disease severity that are more directly tied to anatomic and visual outcomes. Second, all post-treatment exams were excluded from the study given that there are no formal classifications for post-treatment ROP. Therefore, this analysis does not address the issue of disease recurrence post AP-ROP treatment, which is known to be higher than TR without AP-ROP.<sup>31,32</sup> Third, we excluded images of poor quality and therefore the effect of quality on the vascular severity score was not evaluated. Fourth, as alluded to above, there are known differences between AP-ROP seen in North America, and the much more prevalent severe ROP seen in LMIC worldwide and therefore the generalizability of these findings to those populations is unknown.

## CONCLUSION

Visual outcomes from untreated AP-ROP are poor, which make accurate and timely diagnosis critical. The use of quantitative disease metrics, such as the ROP vascular severity scale, may represent a way to improve diagnostic agreement and enable earlier recognition of AP-ROP in the future.<sup>33</sup> In this North American population, we found that the disease tends to occur only in the most premature babies, who have multiple comorbidities, and tends to present earlier and more aggressively than severe treatment-requiring that was not diagnosed as AP-ROP. Combining demographic risk with kinetic monitoring of a vascular severity score may lead to earlier recognition of babies progressing towards AP-ROP in the future. Globally, AP-ROP is both more prevalent and more aggressive in regions of the world with less well-regulated oxygen monitoring, and the greatest potential application of quantitative disease monitoring and early detection may be in this population. Finally, these results demonstrate the potential applicability of deep learning not only for image-based disease diagnosis (i.e. referable diabetic retinopathy) but for quantitative diagnosis of disease, which may be broadly relevant to other imaging technologies within ophthalmology and medicine.

## Acknowledgments

Financial Support: This project was supported by grants R01EY19474, K12EY027720, T15LM007088, and P30EY10572 from the National Institutes of Health (Bethesda, MD), by grants 1622542 and SCH-1622679 from the National Science Foundation (Arlington, VA), and by unrestricted departmental funding and a Career Development Award (JPC) from Research to Prevent Blindness (New York, NY).

Disclosures: Sang Jin Kim is a Consultant for Novartis (Basel, Switzerland), Curacle (Seongnam, Korea), Hanmi Pharmaceutical (Seoul, Korea), and Reyon Pharmaceutical Co., Ltd. (Seoul, Korea). R.V. Paul Chan is on the Scientific Advisory Board for Phoenix Technology Group (Pleasanton, CA), a Consultant for Novartis (Basel, Switzerland), and a Consultant for Alcon (Ft. Worth, TX). Michael F. Chiang is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA), a Consultant for Novartis (Basel, Switzerland), and an initial member of Intelere retina (Honolulu, HI). Michael F. Chiang, J. Peter Campbell, R.V. Paul Chan, and Jayashree Kalpathy-Cramer receive research support from Genentech. R.V. Paul Chan receives research support from Regeneron. J. Peter Campbell, James M. Brown, Susan Ostmo, Aaron Coyner, R.V. Paul Chan, Jayashree Kalpathy-Cramer, and Michael F. Chiang have a preliminary patent application submitted on the i-ROP DL system.

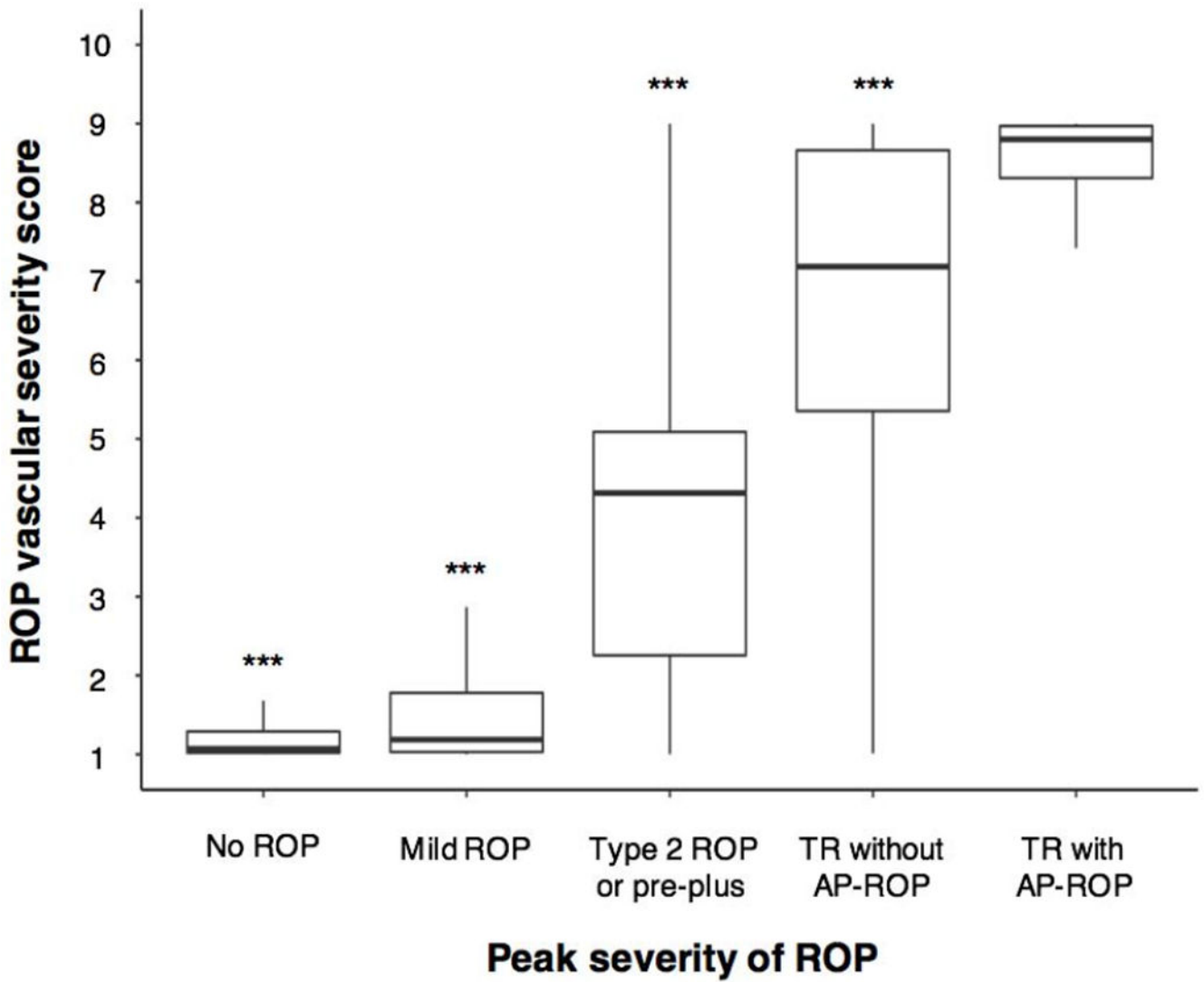


None of the funding agencies had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

## BIBLIOGRAPHY

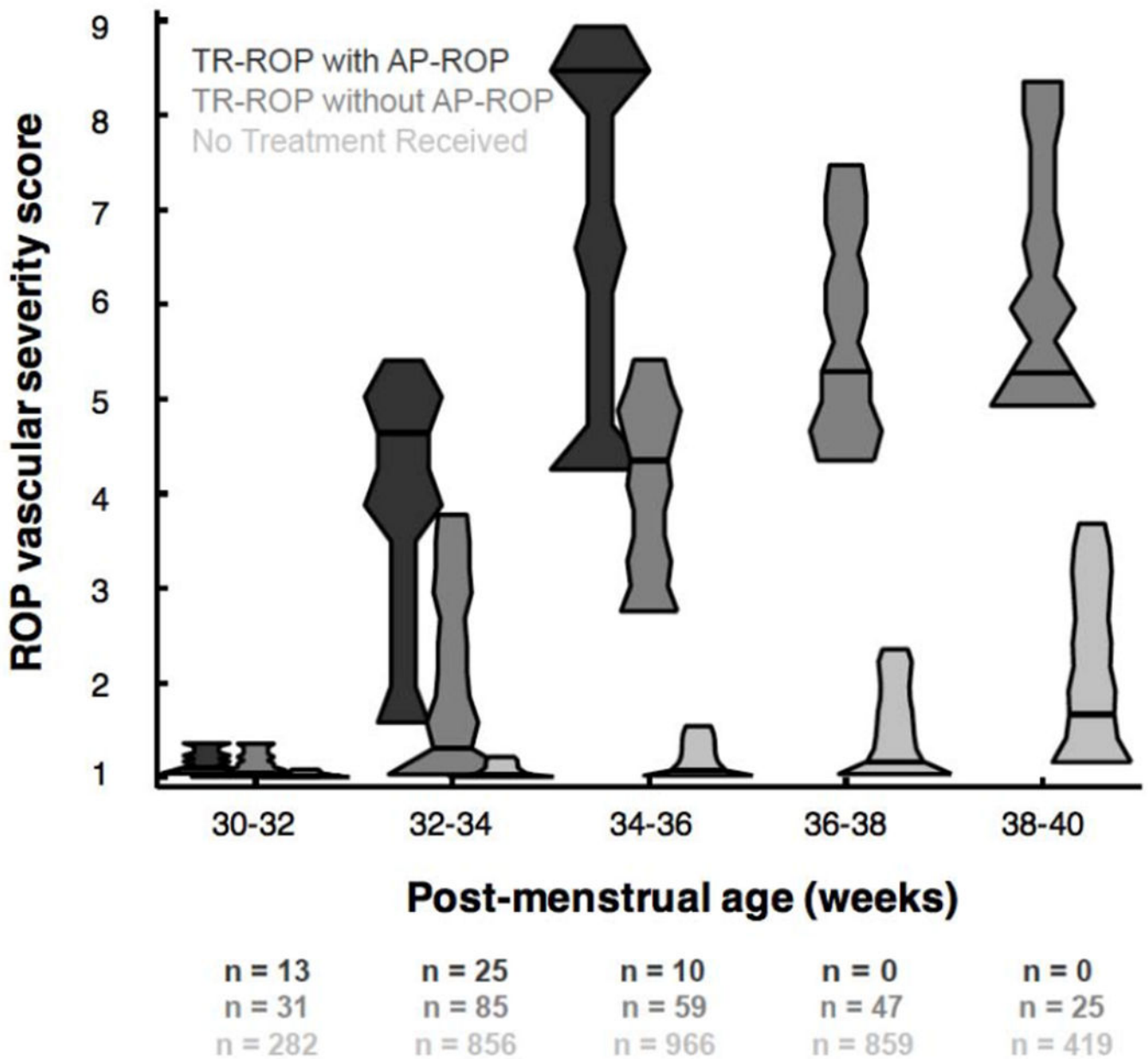
1. International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. *Arch Ophthalmol* 2005;123:991–999. [PubMed: 16009843]
2. Shapiro MJ, Blair MP, Garcia-Gonzalez JM. Experts contradict established classification. *Graefes Arch Clin Exp Ophthalmol* 2016;254:199. [PubMed: 26194405]
3. Chiang MF, Chan RVP, Vinekar A, Woo R. Science and art in retinopathy of prematurity diagnosis. *Graefes Arch Clin Exp Ophthalmol* 2016;254:201–202. [PubMed: 26198712]
4. Patel SN, Singh Ranjodh, Jonas Karen E, et al. Inconsistencies in the Diagnosis of Aggressive Posterior Retinopathy of Prematurity. *Journal of Vitreo Retinal Diseases* 2017;1:181–186.
5. Fielder AR, Wallace DK, Stahl A, et al. Describing Retinopathy of Prematurity: Current Limitations and New Challenges. *Ophthalmology* 2019;126:652–654. [PubMed: 31005186]
6. 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;120:1470–1476. [PubMed: 12427059]
7. Fleck BW, Williams C, Juszczak E, et al. An international comparison of retinopathy of prematurity grading performance within the Benefits of Oxygen Saturation Targeting II trials. *Eye (Lond)* 2018;32:74–80. [PubMed: 28752837]
8. Patel SN, Singh R, Jonas KE, et al. Telemedical Diagnosis of Stage 4 and Stage 5 Retinopathy of Prematurity. *Ophthalmol Retina* 2018;2:59–64. [PubMed: 31047304]
9. Campbell JP, Ryan MC, Lore E, et al. Diagnostic Discrepancies in Retinopathy of Prematurity Classification. *Ophthalmology* 2016;123:1795–1801. [PubMed: 27238376]
10. Kalpathy-Cramer J, Campbell JP, Erdogmus D, et al. Plus Disease in Retinopathy of Prematurity: Improving Diagnosis by Ranking Disease Severity and Using Quantitative Image Analysis. *Ophthalmology* 2016;123:2345–2351. [PubMed: 27566853]
11. Campbell JP, Kalpathy-Cramer J, Erdogmus D, et al. Plus Disease in Retinopathy of Prematurity: A Continuous Spectrum of Vascular Abnormality as a Basis of Diagnostic Variability. *Ophthalmology* 2016;123:2338–2344. [PubMed: 27591053]
12. Agarwal K, Jalali S. Classification of retinopathy of prematurity: from then till now. *Community Eye Health* 2018;31: S4–S7. [PubMed: 30275659]
13. Flynn JT, Chan-Ling T. Retinopathy of prematurity: two distinct mechanisms that underlie zone 1 and zone 2 disease. *Am J Ophthalmol* 2006;142:46–59. [PubMed: 16815250]
14. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Expert Diagnosis of Plus Disease in Retinopathy of Prematurity From Computer-Based Image Analysis. *JAMA Ophthalmol* 2016;134:651–657. [PubMed: 27077667]
15. Brown JM, Campbell JP, Beers A, et al. Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks. *JAMA Ophthalmol* 2018;136:803–810. [PubMed: 29801159]
16. Redd TK, Campbell JP, Brown JM, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol* 2018;103:580–584.
17. Taylor S, Brown JM, Gupta K, et al. Monitoring Disease Progression With a Quantitative Severity Scale for Retinopathy of Prematurity Using Deep Learning. *JAMA Ophthalmol* 2019;137:1022–1028.
18. Gupta K, Campbell JP, Taylor S, et al. A Quantitative Severity Scale for Retinopathy of Prematurity Using Deep Learning to Monitor Disease Regression After Treatment. *JAMA Ophthalmol* 2019;137:1029–1036.
19. 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]

20. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013 Available at: <http://www.R-project.org/>.
21. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–363. [PubMed: 15883903]
22. Hripcsak G, Heitjan DF. Measuring agreement in medical informatics reliability studies. *J Biomed Inform* 2002;35:99–110. [PubMed: 12474424]
23. Pivodic A, Hard A- L, Lofqvist C, et al. Individual Risk Prediction for Sight-Threatening Retinopathy of Prematurity Using Birth Characteristics. *JAMA Ophthalmol* 2019:1–9.
24. Ahn YJ, Hong KE, Yum HR, et al. Characteristic clinical features associated with aggressive posterior retinopathy of prematurity. *Eye (Lond)* 2017;31:924–930. [PubMed: 28234354]
25. Sanghi G, Dogra MR, Dogra M, et al. A hybrid form of retinopathy of prematurity. *Br J Ophthalmol* 2012;96:519–522. [PubMed: 22096143]
26. Shah PK, Subramanian P, Venkatapathy N, et al. Aggressive posterior retinopathy of prematurity in two cohorts of patients in South India: implications for primary, secondary, and tertiary prevention. *J AAPOS* 2019;23:264.e1–264.e4. [PubMed: 31521847]
27. Hewing NJ, Kaufman DR, Chan RVP, Chiang MF. Plus disease in retinopathy of prematurity: qualitative analysis of diagnostic process by experts. *JAMA Ophthalmol* 2013;131:1026–1032. [PubMed: 23702696]
28. Shah PK, Narendran V, Kalpana N. Aggressive posterior retinopathy of prematurity in large preterm babies in South India. *Arch Dis Child Fetal Neonatal Ed* 2012;97:F371–375. [PubMed: 22611114]
29. Martinez-Castellanos MA, Velez-Montoya R, Price K, et al. Vascular changes on fluorescein angiography of premature infants with low risk of retinopathy of prematurity after high oxygen exposure. *Int J Retina Vitreous* 2017;3:2. [PubMed: 28105373]
30. Gilbert C Retinopathy of prematurity: a global perspective of the epidemics, population of babies at risk and implications for control. *Early Hum Dev* 2008;84:77–82. [PubMed: 18234457]
31. Tong Q, Yin H, Zhao M, et al. Outcomes and prognostic factors for aggressive posterior retinopathy of prematurity following initial treatment with intravitreal ranibizumab. *BMC Ophthalmol* 2018;18:150. [PubMed: 29940900]
32. Mintz-Hittner HA, Geloneck MM, Chuang AZ. Clinical Management of Recurrent Retinopathy of Prematurity after Intravitreal Bevacizumab Monotherapy. *Ophthalmology* 2016;123:1845–1855. [PubMed: 27241619]
33. Smith LEH, Hellström A, Stahl A, et al. Development of a Retinopathy of Prematurity Activity Scale and Clinical Outcome Measures for Use in Clinical Trials. *JAMA Ophthalmol* 2019;137:305–311. [PubMed: 30543348]

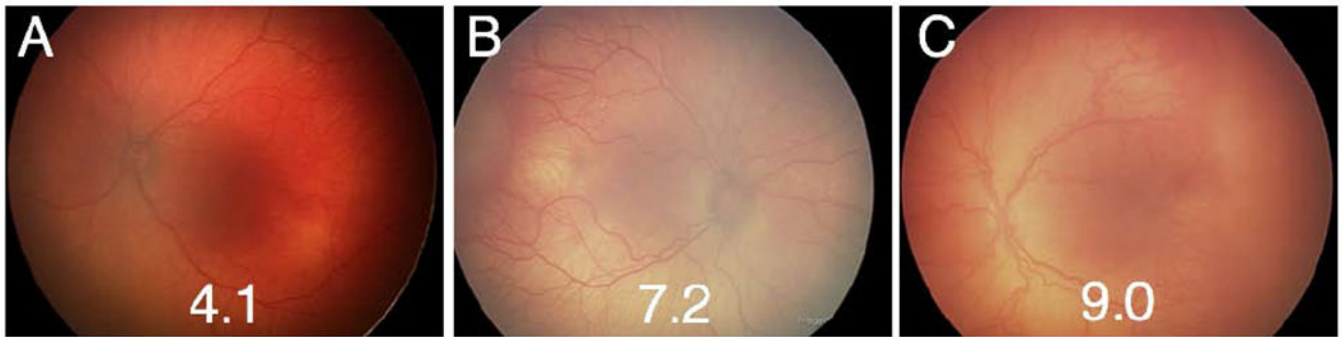


	No ROP	Mild ROP	Type 2 ROP or pre-plus	TR without AP-ROP	TR with AP-ROP
Median	1.07	1.19	4.32	7.19	8.79
Interquartile range	1.02 - 1.29	1.03 - 1.78	2.24 - 5.10	5.34 - 8.68	8.23 - 8.97

**Figure 1. Box-plot distribution of vascular severity scores by peak disease severity.** The vascular severity score at the time of peak disease severity for eyes in three cohorts in the Imaging and Informatics in retinopathy of prematurity (ROP) cohort study was significantly different for each disease category including treatment-requiring (TR) disease with aggressive posterior retinopathy of prematurity (AP-ROP,  $p < 0.001$  for all comparisons).



**Figure 2. Longitudinal evaluation of vascular severity score.** Violin plot distribution of vascular severity score for eyes in three cohorts in the Imaging and Informatics in retinopathy of prematurity (ROP) cohort study. Number (n) in each cohort as well as median and interquartile-range (IQR) of vascular severity score at each time point displayed below. Eyes with aggressive posterior retinopathy of prematurity (AP-ROP) demonstrated earlier and more rapid progression of disease severity compared to treatment-requiring (TR) eyes without AP-ROP and those eyes with non-TR ROP ( $p < 0.002$  for 30-32 weeks, and  $p < 0.001$  for all other time points).



**Figure 3. Representative fundus photos and vascular severity scores of eyes with and without aggressive posterior retinopathy of prematurity (AP-ROP).**

Three example images from eyes of patients with and without AP-ROP. A) Left eye of an infant with type 2 ROP. On ophthalmoscopy, this patient had zone II, stage 2 disease with pre-plus and a severity score of 4.1. B) Right eye of an infant with treatment-requiring (TR) ROP. On ophthalmoscopy, this infant was diagnosed with zone 2, stage 3 disease with plus disease. C) Left eye of an infant diagnosed with TR-ROP with AP-ROP. On ophthalmoscopy, this patient was diagnosed with zone 1, stage 3 with plus disease. Compared to the baby in B, the baby in C demonstrated extensive flat stage 3 disease rather than the more typical stage configuration seen in B.

**Table 1.**  
**Clinical and demographic features of study cohort displayed by peak disease severity.**

Features of infants who developed treatment-requiring (TR) with aggressive posterior retinopathy of prematurity (AP-ROP) are significantly different compared to infants who developed other disease categories.

Peak severity	n (eyes)	Mean birth weight, g (SD)	Mean gestational age, wk (SD)	Mean PMA at peak, wk (SD)
No ROP	841	1142 (309) ***	28.8 (2.0) ***	Not applicable
Mild ROP	584	921 (264) ***	27.0 (2.0) ***	37.0 (3.8) ***
Type 2 ROP or pre-plus	282	742 (210) *	25.7 (1.8) ***	38.2 (4.4) ***
TR without AP-ROP	120	679 (179) **	25.0 (1.5) **	36.9 (2.8) ***
TR with AP-ROP	62	617 (119)	24.3 (0.9)	34.7 (1.3)

\* Statistically significant difference compared to TR with AP-ROP,  $p < 0.05$

\*\* Statistically significant difference compared to TR with AP-ROP,  $p < 0.01$

\*\*\* Statistically significant difference compared to TR with AP-ROP,  $p < 0.001$

g = grams; wk = weeks; SD = standard deviation; PMA = post-menstrual age; ROP = retinopathy of prematurity; TR = treatment-requiring; AP-ROP = aggressive posterior retinopathy of prematurity



**Table 2.**  
**Comorbidities and use of targeted oxygen saturations of study cohort displayed by peak disease severity.**

Infants who developed treatment-requiring (TR) disease with aggressive posterior retinopathy of prematurity (AP-ROP) had higher rates of chronic lung disease than other infants with ROP.

Peak severity	n (eyes)	RBC transfusion	Chronic lung disease	Sepsis	IVH	NEC	PVL <sup>†</sup>
No ROP	841	50% <sup>***</sup>	17% <sup>***</sup>	19% <sup>***</sup>	18% <sup>***</sup>	7.0% <sup>***</sup>	3.1%
Mild ROP	584	75% <sup>**</sup>	34% <sup>***</sup>	31% <sup>***</sup>	24% <sup>*</sup>	21% <sup>**</sup>	3.5%
Type 2 or Pre-Plus	282	91%	58% <sup>**</sup>	36% <sup>*</sup>	33%	21% <sup>*</sup>	9%
TR without AP-ROP	120	93%	58% <sup>*</sup>	51%	41%	29%	7%
TR with AP-ROP	62	93%	80%	59%	40%	36%	7%

\* Statistically significant difference compared to TR with AP-ROP category,  $p < 0.05$

\*\* Statistically significant difference compared to TR with AP-ROP,  $p < 0.01$

\*\*\* Statistically significant difference compared to TR with AP-ROP,  $p < 0.001$

<sup>†</sup> Chi-squared approximation may be invalid due to few subjects with PVL overall, including limited number within TR with AP-ROP group

ROP= retinopathy of prematurity; TR= treatment-requiring; AP-ROP= aggressive posterior retinopathy of prematurity; RBC= red blood cell; IVH= intraventricular hemorrhage; NEC= necrotizing enterocolitis; PVL= periventricular leukomalacia

**Table 3.**  
**Inter-reader agreement on the diagnosis of aggressive posterior retinopathy of prematurity (AP-ROP).**

Agreement between individual readers, as well as between readers and the reference standard diagnosis (RSD), was variable.

Total readings (n)	RSD 5958	Reader 1 5756	Reader 2 3533	Reader 3 3352	Reader 4 2575	Reader 5 2277
RSD						
Reader 1	0.72					
Reader 2	0.52	0.45				
Reader 3	0.74	0.69	0.50			
Reader 4	0.74	0.37	-0.012	0.52		
Reader 5	0.16	0.16	0.15	0.16	-0.0024	
Mean kappa score	0.54	0.50	0.32	0.52	0.32	0.12

RSD= reference standard diagnosis

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript