Is This the Right Reference Standard Diagnosis for Retinopathy of Prematurity?-Reply

Travis K. Redd, MD, MPH,
Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, Portland, Oregon

J. Peter Campbell, MD, MPH,
Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, Portland, Oregon

Michael F. Chiang, MD
Department of Ophthalmology, Casey Eye Institute, Oregon Health and Science University, Portland, Oregon; Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon

In Reply We thank Dr Athikarisamy for his comments. Diagnosis of retinopathy of prematurity, like many ophthalmic conditions, requires interpretation of retinal morphology. In this study,[1] we evaluated 2 methods of visualizing these changes, ophthalmoscopy and telemedicine, and found no significant difference between them for detecting clinically significant retinopathy of prematurity compared with a reference standard diagnosis (RSD). Development of this RSD involved first establishing a composite reference standard using the consensus of 3 independent telemedicine examinations by masked clinicians, combined with a single ophthalmoscopic examination.[2] This was followed by a consensus panel review and adjudication of any cases in which there was a tie between the 4 independent observations. We have previously shown this approach reduces interexaminer variability compared with either telemedicine or ophthalmoscopy.[3]

We will respond to the points raised by Dr Athikarisamy. First, is this RSD vulnerable to incorporation bias? This is a reasonable concern, and we thank Dr Arhikarisamy for raising this issue. Because the telemedicine readers were masked to each other and the results of the ophthalmoscopic examination, incorporation bias could not have played a role in the formation of the reference standard; none of the 4 component diagnoses were dependent on each other. The exception was in cases requiring panel review, during which the panel had access to all of the 4 diagnoses as well as the original image. This raises the possibility that the final RSD was influenced by the index test results, although it is not clear to us how this might influence the RSD or how it could be avoided. Notably, the RSD does have an inherent bias toward ophthalmoscopy in some instances, which is attributable to a different bias: "If no consensus could be obtained owing to lack of confirmatory information in photographs…preference was given to the ophthalmoscopic diagnosis."[1(p500)]

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

Second, does the consensus RSD result in high interexpert variability? It has been demonstrated previously that different experts on retinopathy of prematurity visualizing the same retina often come to different diagnostic conclusions.[4] The fact that this finding is reproduced in our study is not a product of the methods used to establish the reference standard diagnosis but rather is inherent to the current diagnostic paradigm for this condition. Notably, our group is currently evaluating artificial intelligence for diagnosis of retinopathy of prematurity,[5] which may improve diagnostic uniformity.

Finally, does the RSD have poor translational applicability? We agree that it is not feasible to implement the RSD in routine clinical care. It was designed for research purposes, not real-world use. We believe that these results support the conclusion that both ophthalmoscopy and telemedicine can provide accurate diagnosis, although both are limited by the precision and accuracy of the examiner.

We agree that the RSD used in this study is not a perfect gold standard. However, all prior telemedicine studies have usedophthalmoscopy as the gold-standardcomparison, which we propose is inferior to a composite reference standard methodology.[3,6] Thus we encourage the use of similar RSD in studies evaluating diagnostic modalities in retinopathy of prematurity, and we look forward to further discourse to yield even better RSDs in future.

Acknowledgments

References

1. Biten H, Redd TK, Moleta C, et al.; Imaging & Informatics in Retinopathy of Prematurity (ROP) Research Consortium. Diagnostic accuracy of ophthalmoscopy vs telemedicine in examinations for retinopathy of prematurity. JAMA Ophthalmol. 2018;136(5):498–504. doi:10.1001/jamaophthalmol.2018.0649 [PubMed: 29621387]

2. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard: a review of methods. Health Technol Assess. 2007;11(50):iii, ix–51. doi:10.3310/hta11500

3. 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]

4. 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. doi:10.1001/archopht.125.7.875 [PubMed: 17620564]

5. Brown JM, Campbell JP, Beers A, et al.; Imaging and Informatics in Retinopathy of Prematurity (i-ROP) Research Consortium. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA Ophthalmol. 2018;136(7):803–810. doi:10.1001/jamaophthalmol.2018.1934 [PubMed: 29801159]

6. Bertens LCM, Broekhuizen BDL, Naaktgeboren CA, et al. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. PLoS Med. 2013;10(10):e1001531. doi:10.1371/journal.pmed.1001531 [PubMed: 24143138]