Evaluation of computer-based retinopathy of prematurity (ROP) education for ophthalmology residents: a randomized, controlled, multicenter study

Ramak Roohipoor, MD,[a,b] Rodrigo Alvarez, MD,[c] Katarzyna Brodowska, MD,[d] Mehdi Yaseri, PhD,[e] Carolyn Kloek, MD,[a] Mohamad Riazi, MD,[b,f] Ramin Nourinia, MD,[g] Homayoun Nikkhah, MD,[g] N. Venkatesh Prajna, MD,[h] Chandrasekharan Krishnan, MD,[i] Sonal Tuli, MD,[j] Laura Green, MD,[k] Divya Srikumaran, MD,[l] Ankoor S. Shah, MD, PhD,[m] Iason S. Mantagos, MD,[m] Michael Chiang, MD,[n,o] R. V. Paul Chan, MD, MSc,[p,q] and John Loewenstein, MD[a]

PURPOSE To evaluate the effect of a computer-based training program—Massachusetts Eye & Ear ROP Trainer—on residents' knowledge of retinopathy of prematurity (ROP) management.

METHODS In this prospective, randomized study, ophthalmology residents from nine different training programs consented to participate. Those who completed the study were randomly assigned to either the Trainer or the control group. The ROP Trainer was created using clinical cases encompassing the stages of ROP in digital pictures and videos. It includes sections on screening decisions, examination techniques, and diagnosis, and a reference section with the expert video clips and a searchable image library. Subjects in the control group were asked to study standard print material on ROP. A pre- and post-test, consisting of theoretical and practical (diagnosis) questions, and a post-intervention satisfaction test were administered. Accuracy of ROP diagnosis was assessed.

RESULTS A total of 180 residents agreed to participate, of whom 60 completed the study. Residents in the Trainer group had statistically significant improvements ($P = 0.003$) in ROP knowledge and diagnostic ability ($P = 0.005$). Residents randomized to the Trainer group were more satisfied with the training materials than were those in the control group. There was no significant difference in improving knowledge by year of training, sex, or country. Considering all training levels, a statistically significant increase was observed in sensitivity for the diagnosis of preplus or worse, zone I or II, ROP stage, category, and aggressive posterior ROP in the Trainer group.

CONCLUSIONS In this study, the Trainer was shown to significantly improve ROP knowledge and diagnostic skills of residents, regardless of sex, year, of training, or country. (J AAPOS 2019;23: 86.e1-7)

Retinopathy of prematurity (ROP) is a leading cause of childhood blindness in the United States,[1] despite advances in prevention and increases in knowledge of its pathophysiology.[2] Timely screening, diagnosis, and treatment can prevent blindness in this population.[3] However, the number of babies at risk for ROP is increasing worldwide as younger gestational-age neonates are surviving, and we are entering a new

Author affiliations: [a]*Massachusetts Eye and Ear, Harvard Medical School, Department of Ophthalmology, Boston, Massachusetts;* [b]*Farabi Eye Hospital, Eye Research Center, Tehran University of Medical Sciences, Tehran, Iran;* [c]*Pontifical Catholic University, Santiago, Chile;* [d]*University Hospitals of Cleveland, Cleveland, Ohio;* [e]*Biostatistics Department, Tehran University of Medical Sciences, Tehran, Iran;* [f]*Department of Ophthalmology, Gavin Herbert Eye Institute, University of California Irvine, Irvine, California;* [g]*Ophthalmic Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran;* [h]*Aravind Eye Hospitals, Madurai, India;* [i]*Tufts University Medical School, Boston, Massachusetts;* [j]*Department of Ophthalmology, University of Florida, Gainesville, Florida;* [k]*Sinai Hospital, Baltimore, Maryland;* [l]*Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland;* [m]*Boston Children's Hospital, Harvard Medical School Departments of Ophthalmology, Boston, Massachusetts;* [n]*Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, Portland, Oregon;* [o]*Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon;* [p]*Department of Ophthalmology and Visual Sciences, Illinois Eye and Ear Infirmary, University of Illinois at Chicago, Chicago, Illinois;* [q]*Center for Global Health, College of Medicine, University of Illinois at Chicago, Chicago, Illinois*

Financial disclosures: Funding from the Mass Eye and Ear Cures for Kids program supported this project. DS is a consultant for Alcon. RVPC is member of the Scientific Advisory Board for Visunex Medical Systems (Fremont, CA), supported by unrestricted departmental funding from Research to Prevent Blindness (New York, NY), and P30 EY001792 from the National Institutes of Health, Bethesda, Maryland. MFC is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA), and a consultant for Novartis (Basel, Switzerland). ST is the recipient of an unrestricted grant from Research to Prevent Blindness.

Submitted May 10, 2018.
Revision accepted November 11, 2018.
Published online March 16, 2019.

Correspondence: John I. Loewenstein, MD, Associate Professor of Ophthalmology, Retina Service, Massachusetts Eye and Ear, Harvard Medical School, Department of Ophthalmology, 243 Charles Street, Boston, MA 02114 (email: John_Loewenstein@meei.harvard.edu).

1091-8531/$36.00
https://doi.org/10.1016/j.jaapos.2018.11.008

epidemic of ROP in middle- and high-income countries.[4-6]

Ophthalmologists are not universally well-trained in the screening, diagnosis, and treatment of ROP; there is a well-recognized lack of standardized educational programs in fellowship[7,8] and residency. Nagiel and colleagues[9] showed that most residents have limited or no ROP patient exposure, and some perform ROP screening examinations without supervision. Given the increasing number of patients at risk for developing ROP and lack of adequate training, we developed a prototype online program for training residents and fellows in the diagnosis and management of ROP.

Methods

Prototype Trainer

We developed the Massachusetts Eye & Ear ROP Trainer, a prototype, computer-based tool that uses an interactive case, a searchable image library allowing for the comparison of cases, and expert feedback via text and videos for teaching, screening, and ROP diagnosis. The stages of ROP are illustrated in digital pictures and videos. In other videos, ROP experts explain the key aspects of the disease as well as controversial issues. The interactive case asks learners to decide whether and when to screen, to make a diagnosis, and to formulate next steps. Learners receive feedback on their actions, and brief scenarios (referred to in the program as "drilldowns") allow the student to practice and refine their understanding of specific aspects of the disease. The Trainer contains screening and staging scenarios. The image library, which can be searched in a variety of ways, contains 140 fundus images. The learner can study all the images of a given location, stage, and type of ROP, or they can compare images of different locations, stages, and types to help define the borders between categories. Feedback is provided for each scenario (eSupplement 1, available at jaapos.org); if a subject answers a question incorrectly, the program provides and explains the correct answer. The Trainer has a help button that guides the participants in each section.

The program gives feedback on decisions and includes specific "Ask a Question" links that direct the learner to relevant questions and answers in the Reference area. Many questions are answered by videos of experts discussing aspects of ROP. The procedure drilldowns can be studied in any sequence. The learner may access the Reference area (eSupplement 2, available at jaapos.org) at any point to explore a topic or learn more about particular issues, regardless of where they are in the program.

Study Design

This study is a randomized, prospective, single-masked study of the effectiveness of the ROP Trainer involving both an intervention group and a control group. A sample size calculation estimated that 60 subjects would be required for a significant study. The calculation was based on the results of a previous randomized trial of a computer-based training tool for teaching principles of phacoemulsification cataract surgery.[10] Ophthalmology residents from 9 programs (Harvard Medical School, Aravind Eye Hospitals, Farabi Eye Hospital, Labbafinejad Eye Hospital, Catholic University Santiago, University of Florida, Tufts University, Johns Hopkins Institutions, and Sinai Hospital Baltimore) were asked to participate. Participants were randomly assigned to an intervention group or a control group. Randomization was performed using a stratified random permuted block design according to center and postgraduate year of training. Stratification by year of training, a characteristic that is likely to influence test scores, would ensure that the teaching assignments would be well balanced. Within each stratum, subjects were assigned at random to one of the groups, to receive either the traditional teaching tools or the ROP Trainer tool. Both groups participated in any ROP training that was conducted as part of their residency.

The intervention group was asked to use the computer-based tool; the control group was asked to study the sections on ROP in the American Academy of Ophthalmology's Basic and Clinical Science Course and key journal articles on ROP: http://pediatrics.aappublications.org/content/131/1/189.full[3] and http://archopht.jamanetwork.com/article.aspx?articleid=417157.[11] The primary outcome measure was the ability to diagnose ROP from fundus photographs. Secondary outcome measures were knowledge about ROP and satisfaction with training tools.

We asked the residents in the ROP Trainer group to explore the Reference area, review videos from ROP experts, read text material, and browse the Image Library. We recommended that subjects in the ROP Trainer group review the following questions and answers in particular:

- What are the guidelines for the timing of the initial screening examination?
- What are the criteria for screening premature infants?
- What is the international classification system for staging ROP?
- What are the different "types" of ROP?
- How is plus disease defined?
- What is aggressive posterior ROP?

We then asked subjects to review the interactive case and proceed to the drilldowns. In the Trainer, we used the term *staging*, although we asked participants for a full diagnosis, including location, stage, and type.

Preliminary work on the ROP Trainer was presented at meetings of the Association for Research in Vision and Ophthalmology (2010), Club Jules Gonin (2010), and Medicine Meets Virtual Reality (2011). The ROP stage in fundus photographs used in the Trainer has been agreed on by at least 4 of 6 experienced ROP practitioners; if at least 4 did not agree on stage, we excluded the photograph.

Nine residency programs participated in this study. Institutional review board approval was sought at all sites, and the study was determined to be exempt at all sites. The residency program directors at each institution were asked to provide us with the email addresses of their residents. We contacted the residents to solicit participation in the study. The activities of the program directors with regard to this protocol were limited to providing

approval and access to recruiting the potential subjects at their institutions as well as review of aggregate data and manuscripts.

Subjects' email addresses were converted to a unique study number. After each resident agreed to participate, pre-test multiple-choice questionnaires (27 questions) and then pre-test staging questions (23 questions) for grading fundus images were sent to them to determine a baseline level of competency in ROP diagnosis. Prior to developing the Trainer, we asked 5 experienced ROP clinicians to list the most important things a trainee should know about ROP screening and diagnosis. These teaching points were used to develop the content of the Trainer. We composed the pre-test from these teaching points and then asked 6 experienced ROP clinicians to review the questions for face and content validity. The questions were revised, as necessary, according to comments from the reviewers (JM, PC, MCh, RR, RA, ASH). After subjects completed the pre-tests, they were randomly assigned to one of the groups and sent a link to either the ROP Trainer or the print materials. After studying the assigned material, they were sent the post-test multiple-choice questionnaires and post-test staging questions, and finally they were asked to answer the satisfaction questions. Each satisfaction question had a 5-point Likert scale. The photographs in the staging (severity, zone, presence of plus) post-test were not those used in the teaching drill-downs.

No immediate feedback was provided after completion of the pre- or post-test, and the trainees were not made aware of correct or incorrect responses after the post-test. Subjects were allowed to complete the study materials at their own pace and were allowed to save their work and reenter the program at a later time.

Statistical Methods

To present data, we used mean, standard deviation, median, and range. Normal distribution of data was assessed using Kolmogorov-Smirnov test and Q-Q plot. To compare the baseline values between groups, we used χ^2, the Fisher exact test, the *t* test, and the Mann-Whitney test. A paired *t* test was used to evaluate the change in each group. To assess the difference between groups considering their baseline values, we used interaction analysis of group and time within a linear mixed model. Accuracy of ROP diagnosis (eg, plus disease, zone, stage, category) was determined using sensitivity and specificity. We assessed the difference between groups regarding the sensitivity and specificity in pre- and post-test scores (correct answers) with a χ^2 test. The evaluation of the changes in sensitivity and specificity in each group was performed by MacNemar test. Also, the changes in these values were compared between the groups by the time and group interaction within a random intercept logistic regression. In all analyses, *P* values of <0.05 were considered statistically significant. All the analyses were performed using SPSS software, version 22.0 for Windows (IBM Corp., Armonk, NY).

Results

A total of 180 residents agreed to participate in the study over a 2-year period, and all completed the two pre-tests. Since the progress of many subjects was slow after the pre-tests, the data analysis was performed once 60 (the number needed according to the power calculation) completed the study. Mean and standard deviation of baseline theoretical question (26 questions) scores of participants who completed all the steps was 72 ± 14.2, which was statistically comparable to 72 ± 18.8 in participants who did not complete all the steps (difference = 0.0; 95% CI, −6.1 to 5.8; $P = 0.987$). On the other hand, the mean score on the staging pre-test (21 questions) was significantly different between groups (55.4 ± 13.1 vs 50.5 ± 13.8 in participants who completed all materials vs those who did not; difference = 4.9, 95% CI, 0.9-8.9; $P = 0.023$).

Table 1 shows the participants' sex and residency year and US versus non-US distribution in the Trainer group versus the control group. There was no significant difference in these parameters between groups. In the control group, 31% of subject spent less than 1 hour, 60% spent 1-3 hours, and 9% spent 3-5 hours on the study materials, while in the Trainer group 9% spent less than 1 hour, 50% spent 1-3 hours, 27% spent 3-5 hours, and 13% spent more than 5 hours using the Trainer.

There was significant improvement in the control and training groups in theoretical and staging questions. Although there was no significant difference between the Trainer and control group in improving theoretical knowledge of ROP ($P = 0.094$), there was significant improvement in staging skills in the Trainer group ($P = 0.005$). Combining the results on both theoretical and staging tests, the Trainer group improved significantly compared with the control group ($P = 0.003$; Table 2).

Our mixed model analysis showed that there was no statistically significant effect of sex on the improvement in theoretical ($P = 0.745$), staging ($P = 0.716$), and total scores ($P = 0.925$). The effect of participant's grade level on the improvement of theoretical ($P = 0.669$), staging ($P = 0.908$), and total scores ($P = 0.493$) was also not statistically significant.

The Trainer and control groups both had statistically significant improvement in identifying pre-plus or worse ($P = 0.002$ vs $P = 0.032$). When all training years are considered together, the ROP Trainer group had a statistically significant improvement in the accuracy of ROP diagnosis for pre-plus, zone, stage, category of ROP, and aggressive posterior ROP (APROP). See Table 3.

Participants assigned to the Trainer group were more satisfied (Table 4) with the training material than the control group.

There was no statistically significant correlation between self-reported rate of knowledge in Table 3 and objective theoretical ($r = 0.245$; $P = 0.061$) and staging ($r = 0.049$; $P = 0.707$) evaluations and total score ($r = 0.140$; $P = 0.287$) before assigning the materials.

Discussion

In this randomized, prospective, single-masked study, we found significant effectiveness of the ROP Trainer

Table 1. Participants' sex, training year, and country of residence in Trainer group versus control group

	Total score			Routine group			Trainer group		
	Test			Test			Test		
	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change
Theoretical									
Country									
USA	72.4 ± 11.6	89.1 ± 9.4	16.7 ± 16.3	73.1 ± 13	88 ± 12.2	14.9 ± 20.1	71.8 ± 11	90.2 ± 6.7	18.4 ± 13.2
Others	71.8 ± 15.3	83.8 ± 9.1	12.1 ± 17.1	72.2 ± 16	81.5 ± 9.1	9.3 ± 18.3	71.4 ± 14.9	86 ± 8.7	14.8 ± 15.7
P[a]	0.884	0.047	0.441	0.886	0.13	0.482	0.948	0.209	0.559
Grade									
2	70.9 ± 16.4	85.3 ± 11.7	14.5 ± 19.8	72.4 ± 18.7	82.9 ± 13.5	10.5 ± 24.5	69.2 ± 14.4	88.1 ± 9.3	18.8 ± 12.6
3	72.7 ± 13.1	84.1 ± 8.1	11.5 ± 14.8	72.6 ± 13.2	82.3 ± 8.1	9.7 ± 14.3	72.8 ± 13.4	85.7 ± 8	13.2 ± 15.6
4	73.8 ± 11	90.8 ± 7	16.9 ± 15.5	65.4 ± 10.9	90.4 ± 8.2	25 ± 19	79.5 ± 8	91 ± 8	11.5 ± 13.9
P[b]	0.921	0.438	0.73	0.796	0.734	0.738	0.276	0.677	0.233
Sex									
Male	73.4 ± 15.3	85.8 ± 9.6	12.4 ± 18.4	74.6 ± 17.3	83.8 ± 10.7	9.2 ± 20.7	72.3 ± 13.6	87.7 ± 8.3	15.4 ± 16
Female	69.6 ± 12.1	84.6 ± 9.3	15.2 ± 14	68.9 ± 10.1	82.5 ± 9.9	13.6 ± 15.1	70.3 ± 14.3	86.5 ± 8.6	16.8 ± 13.4
P[a]	0.324	0.651	0.536	0.331	0.757	0.542	0.708	0.722	0.809
Total	72 ± 14.2	85.3 ± 9.4	13.4 ± 16.9	72.4 ± 15	83.3 ± 10.2	10.9 ± 18.6	71.5 ± 13.6	87.2 ± 8.3	15.9 ± 14.9
Staging									
Country									
USA	57.1 ± 10.5	66.1 ± 15.2	9 ± 18.4	56 ± 8.7	58.9 ± 10.2	3 ± 13.2	58.2 ± 12.3	72.5 ± 16.6	14.3 ± 21.4
Others	54.7 ± 14	65.6 ± 17.3	10.9 ± 21	52.8 ± 14.7	59.6 ± 10.1	6.8 ± 14.2	56.5 ± 13.5	71.2 ± 20.8	14.7 ± 25.7
P[a]	0.521	0.909	0.747	0.58	0.867	0.514	0.746	0.871	0.965
Grade									
2	54.9 ± 14.2	67.3 ± 17.7	12.5 ± 21.7	54.1 ± 11.9	54.5 ± 11.1	0.4 ± 11	55.7 ± 17	81.4 ± 11.8	25.7 ± 23.4
3	56.7 ± 11.7	64 ± 15.9	7.3 ± 19.4	55.6 ± 13.3	63.2 ± 8.3	7.6 ± 13.9	57.7 ± 10.5	64.7 ± 20.7	7 ± 23.6
4	53.3 ± 19.5	68.6 ± 18.3	15.2 ± 18.3	40.5 ± 23.6	61.9 ± 0	21.4 ± 23.6	61.9 ± 14.3	73 ± 24.4	11.1 ± 18
P[b]	0.653	0.861	0.552	0.493	0.138	0.215	0.645	0.151	0.153
Sex									
Male	55.1 ± 14.3	65.8 ± 15.9	10.7 ± 19.9	55.6 ± 14.1	59.5 ± 9.1	4 ± 12.5	54.6 ± 14.9	71.7 ± 18.7	17 ± 23.6
Female	55.9 ± 11.1	65.6 ± 18	9.7 ± 21	50.6 ± 11.7	59.3 ± 11.5	8.7 ± 15.9	60.7 ± 8.4	71.4 ± 21.2	10.7 ± 25.5
P[a]	0.816	0.976	0.861	0.341	0.956	0.384	0.21	0.973	0.487
Total	55.4 ± 13.1	65.7 ± 16.6	10.3 ± 20.2	53.7 ± 13.2	59.4 ± 9.9	5.7 ± 13.8	57 ± 13	71.6 ± 19.4	14.6 ± 24.2
Total score									
Country									
USA	65.6 ± 9.1	78.8 ± 10	13.3 ± 14.1	65.4 ± 9.4	75 ± 8.3	9.6 ± 13.7	65.7 ± 9.4	82.3 ± 10.6	16.5 ± 14.4
Others	63.2 ± 12.7	75.7 ± 10.4	12.4 ± 14.9	63.5 ± 12.3	71.7 ± 7.9	8.2 ± 13	63 ± 13.3	79.4 ± 11.3	16.4 ± 15.8
P[a]	0.49	0.284	0.841	0.698	0.336	0.805	0.575	0.519	0.986
Grade									
2	63.7 ± 14.1	77.3 ± 11.2	13.6 ± 17.2	64.2 ± 14.9	70.2 ± 10.3	6 ± 16.2	63.2 ± 14	85.1 ± 5.7	21.9 ± 14.8
3	64.3 ± 10.4	75.1 ± 9.7	10.8 ± 12.5	65 ± 9.3	73.8 ± 6.4	8.8 ± 10.6	63.7 ± 11.5	76.3 ± 12	12.6 ± 14
4	64.7 ± 10.9	80.9 ± 11.2	16.2 ± 12.5	54.3 ± 4.5	77.7 ± 4.5	23.4 ± 0	71.6 ± 6.8	83 ± 14.9	11.3 ± 14.9
P[a]	0.887	0.598	0.591	0.665	0.578	0.384	0.373	0.173	0.122
Sex									
Male	65.2 ± 12.6	76.8 ± 10.3	11.6 ± 15.8	66.1 ± 12.7	72.9 ± 8.5	6.9 ± 13.6	64.4 ± 12.8	80.5 ± 10.7	16.1 ± 16.8
Female	61.8 ± 10.1	76.1 ± 10.6	14.3 ± 12.4	60.7 ± 8.8	72.1 ± 7.6	11.4 ± 12	62.8 ± 11.5	79.8 ± 11.8	17 ± 12.7
P[a]	0.277	0.803	0.486	0.23	0.803	0.396	0.724	0.861	0.876
Total	63.9 ± 11.7	76.6 ± 10.3	12.7 ± 14.6	64 ± 11.5	72.6 ± 8	8.6 ± 13	63.8 ± 12.2	80.2 ± 11	16.5 ± 15.1

[a]Mann-Whitney test.
[b]Kruskall-Wallis test.

prototype compared with mandated use of standard teaching materials about ROP in training residents. We found that knowledge of ROP and ability to stage the disease in pre-tests was not optimal, echoing the results of previous studies.[8,12-14]

The ROP Trainer could improve the diagnostic accuracy of ROP by residents specifically in diagnosis of pre-plus or worse, staging, ROP requiring treatment, and APROP. Of these, diagnosis of treatment-requiring ROP is particularly important. Our results are comparable to the i-ROP and GEN-ROP studies.[14,15]

Chan and colleagues[16] developed a tele-education system with the i-ROP and GEN-ROP study groups.[15] In an initial trial of the effectiveness of this tool, 31 of 55 ophthalmology resident volunteers completed the study.[16] The residents were from programs in the United States and Canada. There was no separate control group, because resident performance was compared to baseline competence. The residents improved significantly in sensitivity for the diagnosis of plus disease, zone, stage, category, and presence of APROP. A subsequent study of their system was performed with a group

Table 2. Participants' pre- and post-test scores within and between groups (control and Trainer)

Variable	Total		Routine		Train		*P* value
	Mean ± SD	Median (IQR)	Mean ± SD	Median (IQR)	Mean ± SD	Median (IQR)	
Theoretical items							
Pre-test	72 ± 14.2	73.1 (30.8 to 96.2)	72.4 ± 15	73.1 (30.8 to 96.2)	71.5 ± 13.6	73.1 (46.2 to 96.2)	0.815[a]
Post-test	85.3 ± 9.4	84.6 (57.7 to 100)	83.3 ± 10.2	84.6 (57.7 to 100)	87.2 ± 8.3	88.5 (65.4 to 100)	0.094[c]
Change	13.4 ± 16.9	11.5 (−30.8 to 46.2)	10.9 ± 18.6	7.7 (−30.8 to 46.2)	15.9 ± 14.9	11.5 (−15.4 to 46.2)	
P-within[b]			0.004		<0.001		
Practical items (staging)							
Pre-test	55.4 ± 13.1	52.4 (23.8 to 85.7)	53.7 ± 13.2	52.4 (23.8 to 81)	57 ± 13	57.1 (23.8 to 85.7)	0.334[a]
Post-test	65.7 ± 16.6	66.7 (33.3 to 100)	59.4 ± 9.9	61.9 (38.1 to 76.2)	71.6 ± 19.4	71.4 (33.3 to 100)	0.005[c]
Change	10.3 ± 20.2	7.1 (−38.1 to 76.2)	5.7 ± 13.8	4.8 (−23.8 to 38.1)	14.6 ± 24.2	14.3 (−38.1 to 76.2)	
P-within[b]			0.033		0.002		
Total							
Pre-test	63.9 ± 11.7	64.9 (31.9 to 87.2)	64 ± 11.5	68.1 (31.9 to 83)	63.8 ± 12.2	63.8 (31.9 to 87.2)	0.925[a]
Post-test	76.6 ± 10.3	76.6 (48.9 to 100)	72.6 ± 8	74.5 (48.9 to 80.9)	80.2 ± 11	78.7 (59.6 to 100)	0.003[c]
Change	12.7 ± 14.6	10.6 (−23.4 to 51.1)	8.6 ± 13	6.4 (−23.4 to 31.9)	16.5 ± 15.1	17 (−12.8 to 51.1)	
P-within[b]			0.001		<0.001		

IQR, interquartile range.
[a]Independent sample *t* test.
[b]Paired *t* test.
[c]Interaction analysis within a linear mixed model.

Table 3. Accuracy of ROP diagnosis in the pre- and post-test by all trainees who participated in the Trainer group versus the routine group

Diagnosis	Sensitivity, %								Specificity, %							
	Routine group			Trainer group			Group comparison[b]		Routine group			Trainer group			Group comparison[b]	
	Test			Test					Test			Test				
	Pre	Post	*P* value[a]	Pre	Post	*P* value[a]	Pre	Post	Pre	Post	*P* value[a]	Pre	Post	*P* value[a]	Pre	Post
Plus																
Plus	82.8	96.6	0.103	74.2	90.3	0.096	0.430	0.342	56.3	78.2	0.000	69.9	83.9	0.009	0.060	0.331
Pre or worse	58.6	79.3	0.002	59.7	75.8	0.032	0.907	0.649	67.2	86.2	0.010	82.3	95.2	0.020	0.058	0.091
Stage																
1 or worse	53.2	52.6	0.273	49.3	68.5	0.000	0.426	0.000	69.0	79.3	0.264	93.5	96.8	0.325	0.013	0.036
2 or worse	53.2	52.6	0.273	49.3	68.5	0.000	0.426	0.000	69.0	79.3	0.264	93.5	96.8	0.325	0.013	0.036
3 or worse	50.6	50.6	1.000	51.6	66.1	0.001	0.844	0.003	69.0	58.6	0.204	64.5	75.8	0.051	0.609	0.045
4 or worse	81.0	91.4	0.109	77.4	90.3	0.031	0.629	0.843	46.6	39.7	0.122	47.3	61.3	0.001	0.886	0.000
5	93.1	96.6	0.573	90.3	96.8	0.325	0.703	0.963	49.8	46.3	0.407	49.8	64.5	0.000	0.997	0.000
Type																
Mild or worse	49.4	58.6	0.008	56.3	68.8	0.000	0.112	0.014	69.0	79.3	0.264	93.5	96.8	0.325	0.013	0.036
Type 2 or worse	51.3	58.2	0.059	57.7	69.8	0.001	0.162	0.008	51.7	70.7	0.007	69.4	79.0	0.109	0.049	0.296
Type 1	34.5	44.8	0.106	37.6	54.8	0.009	0.662	0.181	58.6	67.5	0.018	69.6	78.8	0.008	0.019	0.009
Type AP	43.1	53.4	0.224	33.9	62.9	0.000	0.303	0.298	53.4	62.5	0.009	66.5	73.8	0.029	0.003	0.008
ROP	49.4	58.6	0.008	56.3	68.8	0.000	0.112	0.014	67.8	83.9	0.005	86.0	95.7	0.012	0.003	0.008
Zone																
I	44.8	62.1	0.023	58.1	74.2	0.169	0.313	0.321	59.5	62.9	0.495	63.7	74.2	0.047	0.503	0.060
I & II[c]	56.6	62.8	0.150	62.6	74.2	0.016	0.503	0.060	—	—	—	—	—	—	—	—

AP-ROP, aggressive posterior retinopathy of prematurity; *ROP*, retinopathy of prematurity.
[a]Statistically significant ($P < 0.05$); Based on Wilcoxon signed-rank test.
[b]Mann-Whitney test.
[c]Specificity is undefined as all eyes contained disease in zone I and/or zone II.

of residents and fellows from a program in Mexico. These participants were randomized into a control group (n = 29) and an intervention group (n = 29). The control subjects took a pre-test followed by a post-test (without participating in the tele-education system in between tests), whereas the study subjects participated in the tele-education system between the pre- and post-test. Subjects who participated in the tele-education system performed better in grading certain categories of ROP compared with the control group.[6]

The ROP Trainer and GEN-ROP are both Internet-based learning tools that can be used for ROP tele-education. Wide-angle photographs taken by Ret-Cam have are used in both. In both programs, the ROP findings in the photographs were validated by multiple experienced

Table 4. Satisfaction items. difference between control and Trainer group when applicable for both groups

Satisfaction item[a]	Total, mean ± SD	Control group, mean ± SD	Trainer group, mean ± SD	Difference	95% CI		*P* value[b]
					Lower	Upper	
Q1	3.73 ± 1.18	3.38 ± 1.35	4.06 ± 0.89	−0.69	−1.27	−0.10	0.049
Q2	3.75 ± 1.14	3.34 ± 1.17	4.13 ± 0.99	−0.78	−1.34	−0.22	0.009
Q3	2.75 ± 1.2	2.14 ± 1.06	3.32 ± 1.05	−1.18	−1.73	−0.64	0.000
Q4	3.83 ± 1.39	3.38 ± 1.52	4.26 ± 1.12	−0.88	−1.57	−0.19	0.017
Q5	4.15 ± 0.93	3.86 ± 0.94	4.35 ± 0.88	−0.49	−1.00	0.01	0.043
Q6	4.24 ± 1.08	3.74 ± 1.1	4.55 ± 0.96	−0.81	−1.41	−0.22	0.002
Q7	4.18 ± 1.13	3.67 ± 1.14	4.48 ± 1.03	−0.82	−1.45	−0.18	0.004
Q8	3.22 ± 1.26	2.68 ± 1.25	3.71 ± 1.07	−1.03	−1.64	−0.43	0.002
Q9	4.07 ± 1.07	3.56 ± 1.26	4.38 ± 0.8	−0.82	−1.47	−0.18	0.028
Q10	3.87 ± 1.05	3.38 ± 1.02	4.13 ± 0.97	−0.76	−1.38	−0.14	0.023
Q11	3.89 ± 1.23	3.12 ± 1.31	4.3 ± 0.99	−1.18	−1.87	−0.48	0.003

[a]Satisfaction items. Q1. How well do you think the study materials helped you to understand ROP in general? Q2. Did you find the study material to be easy to use? Q3. Did you find that the study material was engaging/fun to use? Q4. Would you like to use this type of study material again? Q5. Photographs of fundus/ROP. Q6. Immediate feedback on your answers. Q7. Case presentation format. Q8. Ability to check/ask questions. Q9. Feedback in the form of outlining of disease features on RetCAM photographs. Q10. Selecting the evidence for your findings. Q11. Feedback in text form.
[b]Based on Mann–Whitney test.

clinicians, although there were discrete differences in the validation methods. Both systems ask the trainee to grade ROP and identify zone, stage, plus, and category, and both provide feedback with correct answers. The ROP Trainer includes content on screening criteria and timing of follow-up examinations with interactive cases and explanation. The tutorial in the GEN-ROP/i-ROP tele-education system also provides content on screening criteria and follow-up. There are a few pedagogical differences between the trainers. The Trainer has a searchable photographic library that allows the trainee to see many examples of each finding and stage and allows for image comparison (eg stage 1 to 2, etc). It also uses video clips of experts explaining aspects of ROP, giving learners some didactic instruction. The GEN-ROP tele-education system provides immediate feedback for the trainee in the form of annotated images of the correct and incorrect responses. Also, the GEN-ROP system has a tutorial including text, graphics, and photographs that the student may review prior to beginning. Despite differences, both ROP training programs were shown to be effective in improving the ability of subjects to diagnose ROP and to identify key features from fundus photographs.

The participants in the GEN-ROP studies were from the United States, Canada, and Mexico, whereas the residents in our study were from the United States, Iran, India, and Chile. Campbell and colleagues[16] evaluated the ROP tele-education system for the diagnosis of ophthalmic disease by international trainees in Brazil, Mexico, and the Philippines. They showed that a Web-based tele-education system can be effective at improving the diagnostic accuracy for ROP for ophthalmology trainees. The value of online ROP education tools therefore appears valid across diverse training programs internationally.

Kirkpatrick[17] describes four levels of evaluation of training programs: reaction, learning, behavior, and results. Reaction, the first level, is similar to "residents' satisfaction." This helps us understand how well the training program was received by our residents. Level 2, learning, generally refers to acquisition of knowledge, skills, or attitude. Level 3 is change in behavior, which is more challenging to measure. Level 4 measures outcomes. In our study, the higher resident satisfaction scores in the Trainer group indicate a Kirkpatrick level 1. Both the control and Trainer groups had significant improvements in ROP knowledge (level 2). The Trainer was not better than standard text materials in this regard. Both groups improved in staging ability, which could be considered a surrogate for behavioral change (level 3), although the Trainer group showed significantly better than the control group. The GEN-ROP/i-ROP tele-education system also had positive survey results, indicating that today's residents respond well to online education.

There are several limitations to our study. First, residents who were approached but did not choose to participate at all could differ from those that agreed to participate. We know from pre-test results that participants who failed to complete all the study materials did not do as well in the staging pre-test. Second, although this study demonstrated better knowledge acquisition and better performance in staging in the ROP Trainer group, we cannot be sure that this advantage will cause better performance in staging in the neonatal intensive care unit or better patient outcomes, because this study was based on image-based diagnosis, which may differ from diagnosis by indirect ophthalmoscopy. Third, a randomized educational trial in ophthalmology residency often requires multiple programs to provide enough study subjects. Each program, however, has a different curriculum and may use different materials and strategies. This makes the selection of a control teaching tool problematic. We chose standard print materials that are typically available to residents. The residents in the Trainer group spent more time using this teaching tool than residents in the control group spent

on their study materials. One could argue that increased time alone could account for the study results. If so, commitment to using the Trainer could still be said to motivate residents to spend more time learning about ROP than they would have otherwise. (We deliberately did not specify how much time residents were expected to work with the materials in either group.) Lastly, there is significant regional variability in screening criteria and management of ROP. Our Trainer used criteria applicable in the US; thus, we cannot fully judge the utility of the program for other regions. We used images captured by RetCam rather than diagnosis by indirect ophthalmoscopy. The latter is considered the "gold standard" by many ROP experts, although studies have shown advantages of both image-based diagnosis and diagnosis by indirect ophthalmoscopy.[18]

References

1. National Eye Institute. Retinopathy of Prematurity. Available at: http://www.nei.nih.gov/health/rop. Accessed May 18, 2010.
2. Hartnett ME. Pathophysiology and mechanisms of severe retinopathy of prematurity. Ophthalmology 2015;122:200-210.
3. 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;131:189-95.
4. 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:e518-25.
5. Chattopadhyay MP, Pradhan A, Singh R, Datta S. Incidence and risk factors for retinopathy of prematurity in neonates. Indian Pediatr 2015;52:157-8.
6. Roohipoor R, Karkhaneh R, Farahani A, et al. Retinopathy of prematurity screening criteria in Iran: new screening guidelines. Arch Dis Child Fetal Neonatal Ed 2016;101:F288-93.
7. Kemper AR, Freedman SF, Wallace DK. Retinopathy of prematurity care: patterns of care and workforce analysis. J AAPOS 2008;12:344-8.
8. Wong RK, Ventura CV, Espiritu MJ, et al. Training fellows for retinopathy of prematurity care: a Web-based survey. J AAPOS 2012;16:177-81.
9. Nagiel A, Espiritu MJ, Wong RK, et al. Retinopathy of prematurity residency training. Ophthalmology 2012;119:2644-5.e1-2.
10. Henderson BA, Kim JY, Golnik KC, et al. Evaluation of the virtual mentor cataract training program. Ophthalmology 2010;117:253-8.
11. International Committee for the Classification of Retinopathy of Prematurity. The International Classification of Retinopathy of Prematurity revisited. Arch Ophthalmol 2005;123:991-9.
12. Roohipoor R, Loewenstein JI. Need for refinement of international retinopathy of prematurity guidelines and classifications. J Ophthalmic Vis Res 2015;10:355-7.
13. Wallace DK. Fellowship training in retinopathy of prematurity. J AAPOS 2012;16:1.
14. Patel SN, Martinez-Castellanos MA, Berrones-Medina D, et al, GEN-ROP, i-ROP Research Consortium. Assessment of a teleeducation system to enhance retinopathy of prematurity training by international ophthalmologists-in-training in Mexico. Ophthalmology 2017;124:953-61.
15. Chan RV, Patel SN, Ryan MC, et al. The Global Education Network for Retinopathy of Prematurity (Gen-Rop): development, implementation, and evaluation of a novel tele-education system (an American Ophthalmological Society thesis). Trans Am Ophthalmol Soc 2015;113:T2.
16. Campbell JP, Swan R, Jonas K, et al. Implementation and evaluation of a tele-education system for the diagnosis of ophthalmic disease by international trainees. AMIA Annu Symp Proc 2015;2015:366-75.
17. Kirkpatrick D. A Practical Guide for Supervisory Training and Development. Reading, MA: Addison-Wesley; 1971.
18. 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;115:1222-1228.e3.