Clinical Models and Algorithms for the Prediction of Retinopathy of Prematurity

A Report by the American Academy of Ophthalmology

Amy K. Hutchinson, MD,1 Michele Melia, ScM,2 Michael B. Yang, MD,3 Deborah K. VanderVeen, MD,4 Lorri B. Wilson, MD,5 Scott R. Lambert, MD1

Objective: To assess the accuracy with which available retinopathy of prematurity (ROP) predictive models detect clinically significant ROP and to what extent and at what risk these models allow for the reduction of screening examinations for ROP.

Methods: A literature search of the PubMed and Cochrane Library databases was conducted last on May 1, 2015, and yielded 305 citations. After screening the abstracts of all 305 citations and reviewing the full text of 30 potentially eligible articles, the panel members determined that 22 met the inclusion criteria. One article included 2 studies, for a total of 23 studies reviewed. The panel extracted information about study design, study population, the screening algorithm tested, interventions, outcomes, and study quality. The methodologist divided the studies into 2 categories—model development and model validation—and assigned a level of evidence rating to each study. One study was rated level I evidence, 3 studies were rated level II evidence, and 19 studies were rated level III evidence.

Results: In some cohorts, some models would have allowed reductions in the number of infants screened for ROP without failing to identify infants requiring treatment. However, the small sample size and limited generalizability of the ROP predictive models included in this review preclude their widespread use to make all-or-none decisions about whether to screen individual infants for ROP. As an alternative, some studies proposed approaches to apply the models to reduce the number of examinations performed in low-risk infants.

Conclusions: Additional research is needed to optimize ROP predictive model development, validation, and application before such models can be used widely to reduce the burdensome number of ROP screening examinations. Ophthalmology 2016;123:804-816 © 2016 by the American Academy of Ophthalmology.

The American Academy of Ophthalmology prepares Ophthalmic Technology Assessments to evaluate new and existing procedures, drugs, and diagnostic and screening tests. The goal of an Ophthalmic Technology Assessment is to review systematically the available research for clinical efficacy and safety. After review by members of the Ophthalmic Technology Assessment Committee, relevant subspecialty societies, and legal counsel, assessments are submitted to the Academy’s Board of Trustees for consideration as official Academy statements. The purpose of this assessment is to assess the ability of available retinopathy of prematurity (ROP) predictive models to detect clinically significant ROP and to what extent and at what risk these models allow for the reduction of screening examinations for ROP.

Background

Retinopathy of prematurity is a vasoproliferative retinal disorder that affects premature infants and is the leading cause of preventable childhood blindness in high- and middle-income countries.1 Preventing blindness caused by ROP requires timely treatment, which depends on appropriate screening of infants at risk. Current United States screening guidelines recommend at a minimum examining all infants with birth weights (BW) of 1500 g or less or estimated gestational age (GA) at birth of 30 weeks or less.2 Although the screening criteria have high sensitivity to detect infants in need of treatment, implementation of these guidelines results in many unnecessary examinations, because only a small percentage of infants screened will meet criteria for treatment.3,4

A number of risk factors for ROP have been described, including BW, GA,5 and oxygen exposure. Oxygen-dependent growth factors, such as vascular endothelial growth factor, play a primary role in the pathophysiology of ROP.7 Deficiencies of non–oxygen-dependent growth factors, such as insulin-like growth factor-1 (IGF-1), normally passed to the developing fetus through the placenta, also play a key role in the pathophysiology of ROP.6,7

804 © 2016 by the American Academy of Ophthalmology
Published by Elsevier Inc. http://dx.doi.org/10.1016/j.ophtha.2015.11.003 ISSN 0161-6420/15
Table 1. Levels of Evidence for Retinopathy of Prematurity Screening Predictive Model Studies

**Predictive Model Development Studies**
- Level I: good-quality model development study including multiple (>1) cohorts
- Study cohorts are representative of population at risk of ROP*
- Study cohorts have differing risks or prevalences of disease, are from different institutions in varying geographic locations, or have differing racial or ethnic composition
- Screening was conducted using indirect ophthalmoscopy by an examiner with ROP experience, with appropriate frequency and duration
- Internal validation
  - Adequate sample size
  - Model discrimination (sensitivity and specificity, AUC, or C index) was reported or could be calculated based on data provided
- Level II: good-quality predictive model development study including single cohort or >1 cohort but with similar ROP risk, geographic location, and racial or ethnic characteristics
- Study cohort is representative of population at risk of ROP*
- Screening was conducted using indirect ophthalmoscopy by an examiner with ROP experience, with appropriate frequency and duration
- Internal validation
  - Adequate sample size
  - Model discrimination (sensitivity and specificity, AUC, or C index) was reported or could be calculated based on data provided
- Level III: low-quality predictive model development study
  - At least 1 of the following applies:
    - Study cohort(s) are not representative of population at risk of ROP*
    - Screening was conducted not using indirect ophthalmoscopy, or not by an examiner with ROP experience, or not of appropriate frequency or duration
    - No validation
  - Inadequate sample size
  - Model discrimination (sensitivity and specificity, AUC, or C index) was not reported or could not be calculated based on data provided

**Model Validation Studies**
- Level I: good-quality predictive model validation study including multiple (>1) cohorts
  - Study cohort(s) are independent of the cohort(s) used to develop the model (i.e., no overlap in infants included in the model development cohort[s] and the model validation cohort[s])
  - Study cohorts were representative of population at risk of ROP*
  - Study cohorts expand on the population(s) used for model development (e.g., to populations with varying risks or prevalences of disease, different institutions in varying geographic locations, or differing racial or ethnic composition, or same institutions but different periods with varying ROP risks or prevalences)
  - Screening was conducted using indirect ophthalmoscopy by an examiner with ROP experience, with appropriate frequency and duration
  - ROP outcome assessor was masked to model determination
  - Adequate sample size
  - Model discrimination (sensitivity and specificity, AUC, or C index) was reported or could be calculated based on data provided
- Level II: good-quality predictive model validation study including single cohort or >1 cohort but with similar ROP risk, demographic, and racial or ethnic characteristics
  - Study cohort is independent of model development cohort (i.e., no overlap in infants was included in the model development cohort[s] and the model validation cohort[s])
  - Study cohort is representative of population at risk of ROP*
  - Screening was conducted using indirect ophthalmoscopy by an examiner with ROP experience, with appropriate frequency and duration
  - ROP outcome assessor was masked to model determination
  - Adequate sample size
  - Model discrimination (sensitivity and specificity, AUC, or C index) was reported or could be calculated based on data provided
- Level III: poor-quality predictive model validation study
  - At least 1 of the following applies:
    - Study cohort was not representative of population at risk of ROP
    - Screening was conducted not using indirect ophthalmoscopy, or not by an examiner with ROP experience, or was not of appropriate frequency or duration
    - ROP outcome assessor was not masked to model determination
  - Inadequate sample size
  - Model discrimination (sensitivity and specificity, AUC, or C index) was not reported or could not be calculated based on data provided

AUC = area under the receiver operating characteristic curve; C index = concordance index; ROP = retinopathy of prematurity.

*Clinical recommendations for the population requiring screening for ROP varied by calendar time and country for the studies included in this review. A study was downgraded if the study cohort was not representative of the screening population based on screening guidelines in place at the time of the study.

1Internal validation uses the same study cohort combined with a statistical method, such as split-sample, cross-validation, or bootstrapping, to adjust for overfitting or optimism. With external validation, the predictive model is applied in a study population that is independent of and differs from the model development population by geographic location. In general, external validation is superior to internal validation and was accepted in place of internal validation if reported as part of a model development study.

1For predictive model development, sample size was considered adequate if there were at least 10 occurrences of the outcome of interest (e.g., type 1 ROP) per predictor variable in the model. For example, for a model with 3 predictor variables, an adequate sample size included at least 30 occurrences of the outcome. For model validation, sample size was considered adequate if the width of the 95% confidence interval on sensitivity and negative predictive value was not wider than 10%.

1Ideally, model calibration also was performed, but studies were not rated based on model calibration. In model calibration, the study cohort is divided into risk groups (approximately 10 is recommended) based on predicted probability of the outcome, and the observed versus predicted probability in the risk groups are plotted and inspected for deviation from the 45° line of equality. Close adherence of all points to the 45° line indicates the model is well calibrated (i.e., it performs equally well across groups with differing risk).
The number of infants requiring screening for ROP is increasing in both high- and middle-income countries as evolving neonatal intensive-care practices improve the survival rate of infants at risk for ROP. However, there is a declining availability of physicians who are willing and able to screen them. Telemedicine has been considered an approach to address this shortfall. Predictive models also could be used to identify and screen only infants who are at highest risk for ROP that requires treatment, to reduce the number of screening examinations performed in low-risk infants, or both. Current screening criteria for ROP are based primarily on 2 predictive factors: BW and GA. These criteria are designed to have very high sensitivity with attendant low specificity; they are based largely on data from the Multicenter Trial of Cryotherapy for Retinopathy of Prematurity and the Light Reduction in Retinopathy of Prematurity studies. In recent years, predictive models have shown improved specificity to detect high-risk patients by incorporating additional factors, including IGF-1 levels or rate of weight gain, which is a surrogate for IGF-1 levels. Existing models perform differently depending on the characteristics of the population of infants being screened and the level of neonatal services provided in their home country.

Questions for Assessment

This assessment addressed the following questions: What is the accuracy with which available ROP predictive models detect clinically significant ROP, and to what extent and at what risk do these models allow for the reduction of screening examinations for ROP?

Description of Evidence

A literature search of the PubMed and Cochrane Library databases was conducted last on May 1, 2015. The search strategy used the following MeSH and text terms: (retinopathy of prematurity [MeSH] OR retinopathy of prematurity AND winrop OR ropscore OR chop OR screening algorithm OR screening algorithms OR prediction model OR prediction models OR screening strategy OR screening strategies OR (screening AND (risk factor OR risk factors OR risk model OR inclusion criteria OR reduction)) AND (costs and cost analysis [Mesh] OR cost-benefit analysis [Mesh] OR cost [tiab] OR ppp OR positive predictive value OR npp OR negative predictive value OR sensitivity OR specificity OR diagnostic accuracy OR models, statistical [Mesh] OR prevention and control [Subheading] OR predictor* OR reduction OR reduced OR prevention OR preventive OR innovation* OR prognosis OR safety index). The searches retrieved 305 citations that had an English abstract. The panel members assessed the abstracts of these citations and identified 30 articles for full-text review. They determined that 22 of the 30 articles reviewed met the following inclusion criteria: they consisted of original research, they involved a clinical prediction model to identify infants at high risk for clinically significant ROP, and they included a prospective or retrospective cohort consisting of premature infants at risk for all stages of ROP. One of the articles included both a model development and a model validation study, for a total of 23 studies. From these, the authors abstracted information about study design, study population, the screening algorithm tested, interventions, ROP outcomes evaluated, and metrics used to evaluate the model.

The methodologist (M.M.) assigned levels of evidence ratings to the studies using a rating scale developed specifically for this assessment, based on published guidelines for the development and validation of a prognostic model. Separate criteria were used for model development and model validation studies. A level I rating was assigned to high-quality studies that included multiple (>1) screening cohorts, a level II rating was assigned to high-quality studies that included a single screening cohort, and a level III rating was assigned to low-quality studies.

None of the model development studies reviewed met the criteria for level I evidence. One model validation study met the criteria for level I evidence. One model development study, and 2 model validation studies met the criteria for level II evidence. The remaining 19 studies were categorized as level III. The major quality deficiencies in the model development studies were that there was no validation (5 studies), sample size was small (5 studies), and the study population was not representative of the population at risk (3 studies). Some studies had more than 1 deficiency. The major deficiency in model validation studies was small sample size, particularly with respect to the number of infants with positive results for the ROP outcome being evaluated, resulting in a wide confidence interval (CI) on algorithm sensitivity (10 studies). Table 1 describes the levels of evidence for predictive model development and validation studies.

Published Studies

All studies included in this assessment evaluated the ability of a predictive model to identify infants at high risk for clinically significant ROP and the extent to which application of the model could reduce the need for ROP screening examinations for the population being studied. The studies were divided into 2 types: model development studies and model validation studies.

In a model development study, a prognostic model is developed using data from a defined cohort of infants for whom the ROP outcome was known and potential risk factors for ROP were assessed. Internal validation is an important component of prognostic model development. During model development, covariates retained in the final predictive model typically are identified from a longer list of potential predictive factors using statistical significance testing. Covariates that by chance seem to be more predictive than they are in truth may be statistically significant and may be included in the final model. Hence, the operating characteristics (sensitivity, specificity) of the final model tend to be overestimated (e.g., the model is overfitted or optimistic). Ideally, the optimism is corrected using statistical techniques for internal validation (i.e., validation such as split-sample, cross-validation, or bootstrapping. Based on the same study cohort).
However, even a model developed using good methodology may not perform well in practice as a result of too much variation in outcome that is unexplained by the model. The role of an external model validation study is to test the model in a cohort drawn from a new population to determine whether the model performs well in practice. This step is critical before using the model in clinical practice, particularly if it will be used in populations that differ from the model development population.

Studies also differed in the following other respects: (1) the definition of the clinically significant ROP outcome that the model was designed to detect (severe ROP, type 1 or 2 ROP; treatment-requiring or treatment-warranting ROP, or both; prethreshold or threshold ROP; stage 3 or higher ROP; any ROP; referral-warranted ROP); (2) the ROP screening criteria used for the population studied; (3) the predictive variables included in the model (GA, BW, IGF-1 level, weekly or daily weight gain, etc.); (4) the metrics used to evaluate the model, which include sensitivity (probability that a patient with clinically significant ROP is identified by the model as being in need of screening examinations), specificity (probability that a patient without clinically significant ROP is identified by the model as not being in need of screening), positive predictive value (probability that a patient identified by the model as being in need of screening demonstrates clinically significant ROP), and negative predictive value (probability that a patient identified by the model as not being in need of screening does not demonstrate clinically significant ROP); and (5) the method by which the authors quantified reduction in screening examinations afforded by the model (number or percent of infants, or both; number or percent of eyes, or both; or number or percent of examinations, or both). In this regard, some authors propose the use of the model to replace existing screening criteria, whereas others recommend the use of the model to reduce the number of examinations without changing existing screening criteria.

Published studies have used several measures of accuracy. For purposes of cross-study comparison, the sensitivity, specificity, positive predictive value, negative predictive value, and corresponding 95% confidence intervals (CI) were abstracted directly from each article or were calculated by the methodologist based on data provided in the article. Confidence intervals were calculated using the binomial exact method.

**Review of Clinical Models for the Prediction of Clinically Significant Retinopathy of Prematurity**

This review is presented in chronological order of model development. Table 2 includes additional details about all model development studies, and Table 3 includes additional details about all model validation studies.

**Safety-Index Model.** In 1996, Schalij-Delfos et al. published a level III retrospective external validation study performed in The Netherlands on the safety-index (S-index) model developed by Meier-Gibbons et al. in 1991. The S-index is calculated based on the following equation: \( S = \log(BW(kg)) + \log(GA) - \log(1 + n) \) of days on oxygen. Development of any ROP was the binary-dependent outcome variable (BDOV) for this validation study, and 33% of 312 infants included in the study showed positive results for the outcome. The sensitivity and specificity of a negative S-index (<0) at the chronologic age of 35 days to detect any ROP in this cohort were 77.7% (95% CI, 68%–85%) and 56.5% (95% CI, 49%–63%), respectively. Application of this model in this cohort of patients would have eliminated the need for screening in 45% of the infants, but it would have missed ROP in 23 infants, including 1 with stage 3 ROP. The authors considered a variety of ways to apply the model and concluded that the best approach would be to stratify the infants into a high-risk group (negative S-index at day 35 or positive S-index but still requiring oxygen at day 35) and low-risk group (positive S-index and not requiring oxygen at day 35), and then perform traditional screening on the high-risk patients and performing a single screening examination on the low-risk patients. Applying the model in this way would have reduced the number of examinations by 10% without missing any cases of stage 3 or higher ROP.

**Termote Model.** In 2005, Termote et al. published a level III model development study performed in The Netherlands describing a multivariate risk model developed and tested retrospectively on 275 patients. Stepwise logistic regression identified 3 significant risk factors for demonstrating any stage of ROP. Internal validation with adjustment for overfit was performed. A diagnostic guideline was developed: (if \( BW + 2 \times \) (gestational age − 20) − 6 × erythrocyte transfusion value within the first 4 weeks of life ≥34, then no ROP screening was necessary). Using the diagnostic guideline, sensitivity and specificity to detect severe ROP was 100% (95% CI, 77%–100%) and 24.5% (95% CI, 19%–30%), respectively. Sixty-four of the 275 infants (23%) in the study group could have been excluded from screening without missing severe ROP.

**Weight, Insulin-like Growth Factor-1, Neonatal, Retinopathy of Prematurity Model.** In 2006, Löfqvist et al. published a level III model development study that introduced a proprietary 2-part ROP surveillance algorithm (Weight, Insulin-like Growth Factor-1, Neonatal, Retinopathy of Prematurity [WINROP]) developed prospectively from 79 patients in Sweden. The sensitivity and specificity of the 2-step screening method in predicting treatment-requiring ROP were 100% (95% CI, 54%–100%) and 83.6% (95% CI, 73%–91%), respectively. The article describes a proposed screening method for ROP using the WINROP algorithm and illustrates the reduction in screening examinations that would have occurred in this cohort. Using the screening protocol, 20% of the infants would not have required any screening examinations, although surveillance of weight and IGF-1 levels would have continued. Fifty-seven percent of infants would have required significantly less screening, but also with continued surveillance, and 23% would have required a traditional number of screening examinations.

In 2009, Löfqvist et al. published a level III prospective external validation study on the WINROP algorithm on 50 patients in Sweden. The sensitivity and specificity of both parts of the WINROP algorithm combined to detect...
sight-threatening ROP were 100% (95% CI, 66%–100%) and 68% (95% CI, 52%–82%), respectively. Application of part 1 of the algorithm (designed to predict the risk for proliferative ROP) would have eliminated the need for screening in 26% of the infants. After application of part 2 of the algorithm, an additional 30% of infants were stratified into a low risk for sight-threatening ROP group and potentially could have undergone a reduced number of examinations. The remaining 44%, considered to be at high risk, would have received the standard number of examinations.

Table 2. Predictive Model

<table>
<thead>
<tr>
<th>Author(s), Year</th>
<th>Institution, Country, Period</th>
<th>Level of Evidence</th>
<th>Predictive Model Studied (Covariates Used)</th>
<th>No. of Infants</th>
<th>Inclusion Criteria</th>
<th>Gestational Age (wks)</th>
<th>Birth Weight (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Termote et al, 2005</td>
<td>Wilhemina Children’s Hospital, The Netherlands, 1996–2000</td>
<td>III</td>
<td>Utrecht model (GA, BW, no. of erythrocyte transfusions during first 4 wks of life)</td>
<td>275</td>
<td>BW &lt;1500 g or GA &lt;32 wks or &gt;3 days FiO2 ≥0.4</td>
<td>Mean, 29; range, 25–34</td>
<td>Mean, 1115; range, 450–2080</td>
</tr>
<tr>
<td>Löfqvist et al, 2006</td>
<td>Queen Silvia Hospital &amp; Uppsala University Hospital, Sweden, 1999–2002</td>
<td>III (prospective study)</td>
<td>WINROP (version 1; PMA, BW, weight gain, IGF-I, IGFBP-3)</td>
<td>79</td>
<td>GA &lt;32 wks</td>
<td>Median, 28; range, 24–32</td>
<td>Median: 1014; range, 530–2015</td>
</tr>
<tr>
<td>Yang and Donovan, 2009</td>
<td>UHC, United States, 1996–2003</td>
<td>III</td>
<td>UHC model (GA, race, gender, multiple births)</td>
<td>357</td>
<td>BW 401–1250 g</td>
<td>Mean, 27; range, NR</td>
<td>Mean, 933; range, NR</td>
</tr>
<tr>
<td>Schildberg et al, 2011</td>
<td>Denmark National Birth Register, Denmark, 2002–2006</td>
<td>III</td>
<td>Denmark model (GA, BW)</td>
<td>4182</td>
<td>BW &lt;1750 g or GA &lt;32 wks</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Binenbaum et al, 2011</td>
<td>PINT (multicenter RCT), United States, 2000–2003</td>
<td>III (prospective study)</td>
<td>PINT (GA, BW, weight gain)</td>
<td>367</td>
<td>Enrolled in PINT study (BW &lt;1000 g)</td>
<td>Median, 26; range, 22–34</td>
<td>Median: 800; range, 445–995</td>
</tr>
<tr>
<td>Binenbaum et al, 2012</td>
<td>CHOP, United States, 2004–2009</td>
<td>II</td>
<td>CHOP (GA, BW, weight gain)</td>
<td>524</td>
<td>BW &lt;1501 g or GA ≤30 wks</td>
<td>Mean, 28; range, 23–33</td>
<td>Mean, 1031; range, 400–1671</td>
</tr>
<tr>
<td>Eckert et al, 2012</td>
<td>Hospital de Clínicas de Porto Alegre, Brazil, 2002–2009</td>
<td>III (prospective study)</td>
<td>ROPScore model (GA, BW, weight gain, oxygen, blood transfusion)</td>
<td>474</td>
<td>BW &lt;1500 g and/or GA ≤32 wks</td>
<td>Mean, 30; range, NR</td>
<td>Mean, 1217; range, NR</td>
</tr>
<tr>
<td>Van Sorge et al, 2013</td>
<td>NEDROP cohort, The Netherlands, 2009</td>
<td>III</td>
<td>NEDROP (GA, BW, 1 or more of AV, NEC, sepsis, postnatal glucocorticoids, or cardiotoracic)</td>
<td>1380</td>
<td>BW &lt;1500 g and/or GA &lt;32 wks or ≥3 days with ≥40% oxygen</td>
<td>Median, 30; range, 28–31</td>
<td>Median, 1260; range, 1020–1520</td>
</tr>
<tr>
<td>Ying et al, 2015</td>
<td>e-ROP (multicenter RCT), North America, 2011–2013</td>
<td>III</td>
<td>e-ROP (gender, race/ethnicity, GA, BW, preplus, ROP stage, retinal hemorrhage, respiratory support, weight gain)</td>
<td>979</td>
<td>BW &lt;1251 g</td>
<td>Mean, 27; range, 23–34</td>
<td>Mean, 860; range, 330–1250</td>
</tr>
</tbody>
</table>

AUC = area under the receiver operating characteristic curve; AV = artificial ventilation duration; BW = birth weight; CHOP = Children’s Hospital factor-1; IGFBP-3 = insulin-like growth factor-1 binding protein 3; NEC = necrotizing enterocolitis; NPV = negative predictive value; NR = not reported; retinopathy of prematurity; RW-ROP = referral-warranted retinopathy of prematurity; UHC = University Hospital of Cincinnati; WINROP = Weight, Insulin-like Growth Factor-1, Neonatal, Retinopathy of Prematurity Model (Weight Gain Only Version of Weight, Insulin-like Growth Factor-1, Neonatal, Retinopathy of Prematurity Model [WINROP 2]). In 2009, Hellström et al published a retrospective level II external validation study from Sweden on a modified (weight gain only) version of the WINROP algorithm. Whereas the original WINROP algorithm was an online surveillance system that was based on weekly postnatal recordings of weight and serum IGF-1 levels, the
modified version (WINROP 2) eliminated serum IGF-1 levels from analysis. The simplified screening procedure was expected to reduce costs and stress on infants. Presence or absence of stage 3 ROP was the BDOV for this validation study, and 9.9% of the 353 patients demonstrated stage 3 ROP (or showed positive results for the outcome). On the basis of a WINROP 2 alarm occurring before 32 weeks postmenstrual age constituting positive test results, the sensitivity and specificity of the model were 100% (95% CI, 90%–100%) and 84.3% (95% CI, 80%–88%), respectively. Application of this model in this cohort of patients would have eliminated the need for screening in 76% of the infants.

Since 2009, a number of other retrospective external validation studies based on the WINROP 2 model have been performed. Each study evaluated the validity of the WINROP 2 model in a cohort of preterm infants using a BDOV, which was the presence or absence of some form of severe ROP. The studies are discussed here in chronological order.

In the Wu et al27 (level III) study performed in the United States in which the BDOV was severe ROP (any prethreshold, stage 3, or threshold ROP), 8.8% of the 318

<table>
<thead>
<tr>
<th>Development Studies</th>
<th>Outcome/Prevalence (%)/No. of Outcomes</th>
<th>Sensitivity (95% Confidence Interval)</th>
<th>Specificity (95% Confidence Interval)</th>
<th>Positive Predictive Value (95% Confidence Interval)</th>
<th>Negative Predictive Value (95% Confidence Interval)</th>
<th>Missed Diagnoses (%)</th>
<th>Savings (%)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe ROP/3.1/14</td>
<td>100% (77%–100%)</td>
<td>24.5% (19%–30%)</td>
<td>6.7% (3.7%–11%)</td>
<td>100% (94%–100%)</td>
<td>0</td>
<td>23</td>
<td></td>
<td>Small sample size</td>
</tr>
<tr>
<td>Treatment required/7.6/6</td>
<td>100% (54%–100%)</td>
<td>83.6% (73%–91%)</td>
<td>33.3% (13%–59%)</td>
<td>100% (94%–100%)</td>
<td>0</td>
<td>20</td>
<td></td>
<td>No validation, small sample size</td>
</tr>
<tr>
<td>Prethreshold or threshold ROP/21/75</td>
<td>90.3% (84%–95%)</td>
<td>72.7% (69%–76%)</td>
<td>45.8% (40%–52%)</td>
<td>96.7% (95%–98%)</td>
<td>9.7</td>
<td>60</td>
<td>(13% reduction in mean no. of examinations)</td>
<td></td>
</tr>
<tr>
<td>Treatment required/2.8/116</td>
<td>99.99% (99.93%–99.99%)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>0.01</td>
<td>17</td>
<td></td>
<td>Not all children screened with indirect ophthalmoscopy</td>
</tr>
<tr>
<td>Severe ROP/18.3/67</td>
<td>97% (92%–100%)</td>
<td>36% (32%–40%)</td>
<td>26% (20%–31%)</td>
<td>99% (95%–99%)</td>
<td>3</td>
<td>30</td>
<td></td>
<td>Study population not representative of population at risk</td>
</tr>
<tr>
<td>Type 1 or 2 ROP/9.2/48</td>
<td>96% (88%–98%)</td>
<td>53% (49%–58%)</td>
<td>18% (13%–21%)</td>
<td>99.6% (98%–99%)</td>
<td>4</td>
<td>49</td>
<td></td>
<td>Study population not representative of population at risk</td>
</tr>
<tr>
<td>Severe ROP**/5.1/24</td>
<td>96% (79%–99.9%)</td>
<td>56% (51%–61%)</td>
<td>10.4% (7%–15%)</td>
<td>99.6% (98%–99%)</td>
<td>4.2</td>
<td>53</td>
<td></td>
<td>Study population not representative of population at risk</td>
</tr>
<tr>
<td>Severe ROP/2.1/29</td>
<td>100% (88%–100%)</td>
<td>20% (18%–22%)</td>
<td>2.6% (1.8%–3.7%)</td>
<td>100% (99%–100%)</td>
<td>0</td>
<td>20</td>
<td></td>
<td>No validation, small sample size</td>
</tr>
<tr>
<td>RW-ROP/15.2/149</td>
<td>96% (91%–99%)</td>
<td>53% (49%–56%)</td>
<td>27% (23%–31%)</td>
<td>99% (97%–99.5%)</td>
<td>4</td>
<td>~45</td>
<td></td>
<td>Study population not representative of population at risk, no validation</td>
</tr>
</tbody>
</table>
patients showed positive results for the outcome. The sensitivity and specificity of a high-risk alarm to predict severe ROP in this cohort were 100% (95% CI, 88%–100%) and 81.7% (95% CI, 77%–86%), respectively, with 75% of the infants eliminated from the need for screening without missing any infants with severe ROP.

In the Hård et al.28 (level III) study performed in Brazil in which the BDOV was proliferative ROP, 5.7% of the 336 patients showed positive results for the outcome. The sensitivity and specificity of a high-risk alarm, a low-risk alarm, or both that sounded before 32 weeks to detect proliferative ROP in this cohort were 90.5% (95% CI,
In the Wu et al. (level I) study performed in the United States in which the BDOV was type 1 ROP, 8.6% of 1706 patients showed positive results for the outcome. The sensitivity and specificity of the WINROP 2 algorithm in this cohort to detect type 1 ROP were 98.6% (95% CI, 95%–99.8%) and 38.7% (95% CI, 36.2%–41.1%), respectively. This study proposed that application of the WINROP 2 algorithm in this cohort of patients could have

<table>
<thead>
<tr>
<th>Outcome/Prevalence (%)/No. of Outcomes</th>
<th>Sensitivity (95% Confidence Interval)</th>
<th>Specificity (95% Confidence Interval)</th>
<th>Positive Predictive Value (95% Confidence Interval)</th>
<th>Negative Predictive Value (95% Confidence Interval)</th>
<th>Missed Outcomes (%)</th>
<th>Savings (%)*</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 3 or higher/4.5/14</td>
<td>100% (77%–100%)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>0</td>
<td>10%</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Sight-threatening ROP/18/9</td>
<td>100% (66%–100%)</td>
<td>68% (52%–82%)</td>
<td>41% (21%–64%)</td>
<td>100% (88%–100%)</td>
<td>0</td>
<td>26</td>
<td>Small sample size; no infants with zone 1 ROP included</td>
</tr>
<tr>
<td>Stage 3 or higher ROP/9.9/35</td>
<td>100% (90%–100%)</td>
<td>84% (80%–88%)</td>
<td>41% (31%–52%)</td>
<td>100% (99%–100%)</td>
<td>0</td>
<td>76</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Prethreshold or threshold ROP/18/180</td>
<td>89.4% (84%–94%)</td>
<td>68% (65%–71%)</td>
<td>39% (34%–44%)</td>
<td>97% (95%–98%)</td>
<td>10.6</td>
<td>58</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Severe ROP/8.8/28</td>
<td>100% (88%–100%)</td>
<td>81.7% (77%–86%)</td>
<td>34.6% (24%–46%)</td>
<td>100% (98%–100%)</td>
<td>0</td>
<td>75</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Stage 3 or higher ROP/5.7/21</td>
<td>90.5% (70%–99%)</td>
<td>55.1% (50%–60%)</td>
<td>11% (7%–17%)</td>
<td>99% (96%–&gt;99%)</td>
<td>9.5</td>
<td>52</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Type 1 ROP/8.6/146</td>
<td>98.6% (95%–99.8%)</td>
<td>39% (36%–41%)</td>
<td>13% (11%–15%)</td>
<td>99.7% (98.8%–&gt;99.9%)</td>
<td>1.4</td>
<td>35</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Type 1 ROP/51/98</td>
<td>85% (76%–91%)</td>
<td>27% (18%–37%)</td>
<td>55% (46%–63%)</td>
<td>63% (46%–77%)</td>
<td>15.3</td>
<td>21</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Severe ROP/9.5/56</td>
<td>89% (78%–96%)</td>
<td>89% (86%–91%)</td>
<td>46% (36%–57%)</td>
<td>99% (97%–&gt;99%)</td>
<td>10.7</td>
<td>82</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Type 1 ROP/12.7/40</td>
<td>90% (76%–97%)</td>
<td>53% (46%–59%)</td>
<td>22% (16%–29%)</td>
<td>97% (93%–&gt;99%)</td>
<td>10</td>
<td>47</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Type 1 ROP/11.5/47</td>
<td>96% (85%–99%)</td>
<td>24% (22%–29%)</td>
<td>14% (10%–18%)</td>
<td>98% (92%–&gt;99%)</td>
<td>4</td>
<td>~53</td>
<td>Study population not representative of population at risk</td>
</tr>
<tr>
<td>Treated ROP/16.2/66</td>
<td>94% (85%–98%)</td>
<td>25% (20%–30%)</td>
<td>19% (15%–24%)</td>
<td>95% (89%–99%)</td>
<td>6</td>
<td>~6</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Severe ROP/3.9/16</td>
<td>88% (62%–98%)</td>
<td>63% (58%–68%)</td>
<td>9% (5%–14%)</td>
<td>99% (97%–&gt;99%)</td>
<td>12.5</td>
<td>62</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Treatment required/4.8/5</td>
<td>100% (48%–100%)</td>
<td>59% (48%–68%)</td>
<td>11% (3.6%–24%)</td>
<td>100% (94%–100%)</td>
<td>0</td>
<td>NR</td>
<td>Small sample size</td>
</tr>
<tr>
<td>Treatment required/11.5/17</td>
<td>65% (38%–86%)</td>
<td>55% (46%–64%)</td>
<td>16% (8%–26%)</td>
<td>92% (84%–97%)</td>
<td>35</td>
<td>NR</td>
<td>Small sample size</td>
</tr>
</tbody>
</table>

Notes:
- NPV = negative predictive value
- PPV = positive predictive value
- RCT = randomized clinical trial
- Retinopathy of Prematurity.

The GSH metrics are reported in this table. Prevalence and model evaluation metrics were reported at the eye level. Although the study population was graded down for this restriction; the development study was graded down for it because the restricted BW criterion did not meet screening guidelines.

70%–99%) and 55% (95% CI, 50%–60%), respectively, with 52% of the infants eliminated from the need for screening at the cost of missing 2 infants who demonstrated proliferative ROP.

In the Wu et al. (level I) study performed in the United States in which the BDOV was type 1 ROP, 8.6% of 1706 patients showed positive results for the outcome. The sensitivity and specificity of the WINROP 2 algorithm in this cohort to detect type 1 ROP were 98.6% (95% CI, 95%–99.8%) and 38.7% (95% CI, 36.2%–41.1%), respectively. This study proposed that application of the WINROP 2 algorithm in this cohort of patients could have
resulted in “reduced ophthalmologic examinations for almost 30% of infants and still detected 100% of type 1 ROP” if it were used in a manner similar to what is being done in several Swedish intensive care units. In these intensive care units, infants born at GA more than 29 weeks who do not receive an alarm are screened once at 5 weeks chronologic age, and if no ROP is detected, ROP screening examinations are discontinued. For infants born at GA 29 weeks or sooner, clinical judgment is used to determine whether additional examinations should be performed if no ROP is present at 5 weeks chronologic age.

The Zepeda-Romero model (level III) study was performed in Mexico, and the BDOV was type 1 ROP. Because the WINROP model was developed from a cohort of 192 infants with GA of less than 32 weeks, separate analyses were performed for infants with GA of less than 32 weeks (very preterm) and for those with GA of 32 weeks or more (moderately preterm). For the group of infants with GA of less than 32 weeks, 98 patients (51%) showed positive results for the outcome. The sensitivity and specificity of an alarm before 33 weeks to detect type 1 ROP in this cohort were 84.7% (95% CI, 76%–91%) and 26.6% (95% CI, 18%–37%), respectively, with 21% of the infants eliminated from the need for screening at the cost of missing 15 infants who demonstrated type 1 ROP. For the group of infants with GA of more than 32 weeks, the WINROP algorithm performed poorly, with a sensitivity of 5.3% and specificity of 88.3%.

In the Sun et al study (level III) study performed in China in which the BDOV was severe ROP (defined as any prethreshold, stage 3, or threshold ROP), 9.5% of 590 patients showed positive results for the outcome. The sensitivity and specificity of a WINROP 2 alarm to detect severe ROP in this cohort were 89.3% (95% CI, 78%–96%) and 89% (95% CI, 86%–91%), respectively, with 82% of the infants eliminated from the need for screening at the cost of missing 6 infants with severe ROP.

In the Choi et al study (level III) study performed in Korea in which the BDOV was type 1 ROP, 12.7% of 314 patients showed positive results for the outcome. The sensitivity and specificity of a high-risk WINROP alarm to detect type 1 ROP in this cohort were 90.0% (95% CI, 76%–97%) and 52.6% (95% CI, 46%–59%), respectively, with 47% of the infants eliminated from the need for screening at the cost of missing 4 infants with type 1 ROP.

In the Lundgren et al study (level III) study performed in Sweden in which the BDOV was type 1 ROP, 11.5% of 407 patients showed positive results for the outcome. The sensitivity and specificity of a WINROP alarm to detect severe ROP in this cohort were 95.7% (95% CI, 85%–99%) and 23.9% (95% CI, 20%–29%), respectively. An alarm did not occur in 2 infants diagnosed and treated for type 1 ROP. The authors state that a 53% reduction in the number of examinations performed on infants with no alarm and no ROP could have been achieved if these infants had received routine eye examinations at 3 time points (31, 33, and 36 weeks GA) instead of according to their current protocol.

In the Piyasena et al study (level III) study performed in Scotland in which the BDOV was any ROP in zone 1, stage 2 ROP in zone 2 with plus disease, or any stage 3 ROP, 3.9% of 410 patients showed positive results for this outcome. The sensitivity and specificity of a high-risk WINROP alarm to detect severe ROP in this cohort were 87.5% (95% CI, 62%–98%) and 63.4% (95% CI, 58%–68%), respectively, with 62% of the infants eliminated from the need for screening at the cost of missing 2 infants with severe ROP.

In the Eriksson et al study performed in Sweden in which the BDOV was severe ROP (defined as stage 3 or treatment-requiring ROP), 4.8% of 104 patients showed positive results for the outcome. The sensitivity and specificity of a WINROP alarm to detect severe ROP in this cohort were 100% (95% CI, 48%–100%) and 58.6% (95% CI, 48%–68%), respectively.

In the Ko et al study performed in Taiwan in which the BDOV was treatment-demanding ROP (TD-ROP), 11.5% of 148 patients showed positive results for the outcome. The sensitivity and specificity of a WINROP alarm to detect TD-ROP in this cohort were 64.7% (95% CI, 38%–86%) and 55% (95% CI, 46%–64%), respectively. An alarm did not occur in 6 infants who were treated for ROP, all of whom had BW of more than 1250 g, GA of more than 30 weeks, or both.

Yang Model. In 2009, Yang and Donovan published a report of a level III study describing a multivariate risk model on 357 patients from the United States. They then performed a level II validation study of the model by applying it retrospectively to an independent validation cohort of 491 patients from the United States. The predictive variables used in the final model were BW, GA, multiple birth, race, and gender. Presence or absence of prethreshold or threshold ROP was the BDOV. One hundred forty-five eyes (20%) in 75 infants (21%) of the University Hospital in Cincinnati cohort and 180 eyes (18%) of the Good Samaritan Hospital cohort showed positive results for this outcome. The area under the receiver operating characteristic curve was used to describe the model’s predictive capacity, and a probability of 0.15 or more of prethreshold or threshold ROP was selected as the cutoff for high risk, because it resulted in the best combination of sensitivity and specificity. The model then was applied to infants from both cohorts, and the eyes were assigned to a high-risk group or a low-risk group. Infants in the high-risk group were screened conventionally (initial screening at 32 weeks GA or 5 to 6 weeks chronological age, whichever was later, and the screening sequence was modified retrospectively to conform to current screening recommendations). Infants in the low-risk group were screened according to an alternative 35q3 protocol (initial screening at 35 weeks GA or 5 to 6 weeks chronological age, whichever is later, and subsequent screenings every 3 weeks, adjusted based on severity of ROP findings). This allowed for a reduction in the mean number of screening examinations by 13.4% per infant with no more than 1 week’s delay in detecting prethreshold or threshold ROP. If 2013 screening guidelines had been applied and infants with BW between 1250 and 1500 g had been included in the cohort, the reduction in screening examinations likely would have been greater. The sensitivity, specificity, positive
predictive value, and negative predictive value of the model to detect prethreshold or threshold ROP for each cohort are shown in Table 1. If the model had been applied in such a way that all low-risk eyes were not screened at all, 60% of eyes in the University Hospital in Cincinnati cohort and 58% of eyes in the Good Samaritan Hospital cohort would not have been screened, missing 9.7% and 10.6% of pre-threshold or worse ROP in each cohort, respectively. Of note, none of these prethreshold ROP eyes went on to require treatment.

Danish Retinopathy of Prematurity Model. In 2011, Slidsborg et al17 published a level III study describing a nonlinear logistic regression model developed and tested retrospectively on 4182 patients in Denmark. The predictive variables used in the model were BW and GA. Presence or absence of TD-ROP was the BDOV and 116 (2.8%) showed positive results for this outcome. Some cases of ROP were treated at threshold and others at prethreshold. Internal validation using the bootstrapping method was performed. Risk-based isolines were constructed (various combinations of GA and BW with identical risk of TD-ROP), and theoretical application of the 0.13% risk isoline resulted in the best outcome, with no TD-ROP patients missed and 17.4% fewer infants requiring screening than with the current guidelines. A concern of this study is that TD-ROP was threshold ROP in many cases, and current treatment strategy involves detection of type 1 ROP.

Premature Infants in Need of Transfusion Retinopathy of Prematurity Model. In 2011, Binenbaum et al20 published a level III study describing a multivariate risk model (Premature Infants in Need of Transfusion ROP model) developed from 367 patients from the United States. The predictive variables used in the final model were GA, BW, and daily weight gain rate calculated from the current and previous weeks’ weights. Presence or absence of severe ROP (defined as stage 3 or greater ROP or treated ROP) was the BDOV for the study, and 67 patients (18.3%) showed positive results for the outcome. Internal validation using the bootstrap methods of Harrell et al20 was performed, and adjustments were made for overfit. To create a pilot clinical tool to determine ROP risk, the final logistic model was converted into graphical form as nomograms. When the cut point was raised to miss just 1 case of type 1 ROP (>0.159), 6 cases of type 2 ROP were missed and 25 (5.1%) showed positive results for the outcome. The best cut point for sensitivity and specificity was established as 14.5.

At this cut point, the sensitivity and specificity of ROPSscore in predicting severe ROP were 95.8% (95% CI, 79%–99.9%) and 56.0% (95% CI, 51%–61%), respectively. Application of the model accurately predicted 96% of infants with severe ROP, missed 4% of infants with severe ROP, and would have resulted in 53% fewer infants undergoing examinations.

Netherlands Retinopathy of Prematurity Study Model [NEDROP]. In 2013, van Sorge et al40 published a level III study describing this multivariate risk model developed prospectively from 474 patients from Brazil. The predictive variables used in the model were GA, BW, history of blood transfusion, use of oxygen in mechanical ventilation up to the sixth week of life, and proportional weight gain at 6 weeks of life. Presence or absence of severe ROP (as defined by the Early Treatment for Retinopathy of Prematurity Study [ETROP]) was the BDOV for the study, and 25 patients (5.1%) showed positive results for the outcome. The best cut point for sensitivity and specificity was established as 14.5.

At this cut point, the sensitivity and specificity of ROPSscore in predicting severe ROP were 95.8% (95% CI, 79%–99.9%) and 56.0% (95% CI, 51%–61%), respectively. Application of the model accurately predicted 96% of infants with severe ROP, missed 4% of infants with severe ROP, and would have resulted in 53% fewer infants undergoing examinations.

Evaluating Acute-Phase Retinopathy of Prematurity Model. In 2015, the Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity Cooperative Group (Ying et al)10 published a level III study describing a multivariate risk model (Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity model) developed from a cohort of 979 infants from the United States enrolled in the prospective Telemedicine Approaches to Evaluating Acute-Phase Retinopathy of Prematurity study. The predictive variables used in the final
model were gender, race or ethnicity, BW, GA, quadrants of plus, stage of ROP, presence of retinal hemorrhage, degree of respiratory support, and weight gain. Presence or absence of referral-warranted ROP (plus disease, zone 1 ROP, or stage 3 or greater ROP) was the BDOV for the study and was present in 149 infants (15.2%); the area under the receiver operating characteristic curve was 0.88 (95% CI, 0.85–0.91). When a predicted probability of referral-warranted ROP of 0.05 was used as a cut point, the model had 96% (95% CI, 91%–99%) sensitivity and 53% (95% CI, 49%–56%) specificity. No internal or external validation of this model was performed.

**Future Research**

Clinical prognostic models have great potential to improve ROP management by reducing the burden of unnecessary screenings and identifying infants who may benefit from preventive measures. However, models must be developed and validated rigorously, and they must be applied carefully. Binzenbaum identified small sample size and limited generalizability as the primary factors that limit the widespread clinical application of existing ROP screening models. The CIs for estimates of sensitivity are too wide for clinicians to apply the models clinically, although in selected cohorts, sensitivity has been reported to be as high as 100%. To narrow the CIs, it is necessary for studies to include a large number of patients with the outcome variable that the model is designed to predict. Binzenbaum argues that for clinicians to trust a model enough to justify its use to make all-or-none screening decisions, the lower boundary of the CI for sensitivity should be very high, perhaps even greater than 99%. This is because any cost savings afforded by applying these types of models could be negated by even a handful of cases of missed ROP that result in lifelong blindness. Even a strategy whereby examinations are reduced in low-risk infants rather than eliminated altogether requires a very high sensitivity and a protocol that does not result in excessive delay in detecting severe ROP.

Limited generalizability also affects the usefulness of existing models. Binzenbaum points out that predictive models may perform poorly because of disparities in neonatal practices and patient characteristics across patient populations. Especially in regions where higher BW and GA patients demonstrate clinically significant ROP, separate model development and validation studies will need to be performed. They may include predictive variables that are additional to or different from the ones in the models used in populations where only very low BW and GA infants develop ROP.

Postnatal weight gain increasingly is recognized as an extremely useful predictive variable as a surrogate for IGF-1, especially in identifying larger infants who are at risk for clinically significant ROP in high-income countries with advanced neonatal practices. Insulin-like growth factor-1 recently was shown to play a primary role in the pathophysiology of ROP and likely acts as a common pathway for many other previously described risk factors. Other factors, such as supplemental oxygen exposure, may prove to be important predictors for models developed for use in regions with less advanced neonatal practices, where larger infants more frequently develop clinically significant ROP.

Experts in the care of infants with ROP should come to a consensus about the definition of clinically significant ROP and strive to conduct studies that include large numbers of patients who meet those criteria. Existing models use a variety of definitions (e.g., type 1 ROP, stage 3 ROP, treated ROP). Clinically significant ROP may be defined differently in some models, such as in areas of the world that have varying levels of development where the biological activity of disease may differ from that in high-income countries (e.g., in India and Mexico, where a high number of patients with stage 2, zone 2 ROP receive treatment). Data harvesting from electronic health records specifically designed to capture ROP-related data could prove invaluable, especially if obtained from multiple centers in a prospective manner. In turn, the same electronic health record could be programmed to apply the very model it helps to develop and refine.

Development, validation, and application of predictive models to aid in ROP screening and treatment are still in their early stages, and no existing predictive model is ready for widespread application. Future research in this area should adhere to published guidelines for prognostic research and should strive for level I evidence, as described in Table 1. Models should be validated externally before clinical use, and they should be as accessible and simple to use as possible. If CIs for sensitivity cannot be narrowed enough for clinicians to apply the models in an all-or-none fashion to eliminate screenings completely, models may be used to reduce the number of screening examinations performed in infants deemed at low risk. Assessment of the ability of the model to alter physicians’ behaviors, to enhance patient outcomes, and to increase cost effectiveness is the final and possibly the most important phase of prognostic research and should be conducted to determine the impact of the models. Current research in this area includes the National Institutes of Health-funded Postnatal Growth and Retinopathy of Prematurity studies. This work will develop a prognostic model based on retrospective data from more than 7500 premature infants that subsequently will undergo prospective validation on a cohort of 4000 infants, and it will assess the cost effectiveness of the overall program.

**References**


Footnotes and Financial Disclosures

Originally received: November 3, 2015.
Final revision: November 4, 2015.
Accepted: November 4, 2015.

1 Department of Ophthalmology, Emory University School of Medicine, Atlanta, Georgia.
2 Jaeb Center for Health Research, Tampa, Florida.
3 Department of Ophthalmology, Abrahamson Pediatric Eye Institute, Cincinnati Children’s Hospital Medical Center, University of Cincinnati, College of Medicine, Cincinnati, Ohio.
4 Department of Ophthalmology, Boston Children’s Hospital, Harvard Medical School, Boston, Massachusetts.
5 Casey Eye Institute, Oregon Health & Science University, Portland, Oregon.

Prepared by the Ophthalmic Technology Assessment Committee Pediatric Ophthalmology/Strabismus Panel and approved by the American Academy of Ophthalmology’s Board of Trustees September 18, 2015.

Financial Disclosure(s):
The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Funded without commercial support by the American Academy of Ophthalmology.

Abbreviations and Acronyms:

BDOV = binary-dependent outcome variable; BW = birth weight; CI = confidence interval; GA = gestational age; IGF-1 = insulin-like growth factor-1; ROP = retinopathy of prematurity; S-index = safety index; TD-ROP = treatment-demanding retinopathy of prematurity; WINROP = Weight, Insulin-like Growth Factor-1, Neonatal, Retinopathy of Prematurity.

Correspondence:
Jennifer Harris, MS, American Academy of Ophthalmology, Quality and Data Science, P.O. Box 7424, San Francisco, CA 94120-7424. E-mail: jharris@aoa.org.