AMERICAN ACADEMY™
OF OPHTHALMOLOGY

Ophthalmic Technology Assessment

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, MD[1]

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 (BWs) 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,[4] and oxygen exposure. Oxygen-dependent growth factors, such as vascular endothelial growth factor, play a primary role in the pathophysiology of ROP.[5] 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]

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 cohorts 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.

†Internal 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.

‡For 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%.

§Ideally, 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.[8] However, there is a declining availability of physicians who are willing and able to screen them.[9] Telemedicine has been considered an approach to address this shortfall.[10] 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.[11] 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.[12] 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.[13] 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 *ppv* OR *positive predictive value* OR *npv* OR *negative predictive value* OR *sensitivity* OR *specificity* OR *diagnostic accuracy* OR *models*, *statistical* [Mesh] OR *prevention and control* [Subheading] OR *predict** OR *reduction* OR *reduced* OR *prevention* OR *preventive* OR *innovation** OR *prognosis* OR *safety index*).

The searches retrieved 305 citations that had an English language 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.[14] 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.[15,16] 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[17] met the criteria for level I evidence. One model development study,[18] and 2 model validation studies[14,19] 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.[15,16] 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,[15,20,21] 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[22] 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[23] in 1991. The S-index is calculated based on the following equation: $S = \log BW(kg) + \log GA - \log(1 + \text{no. of days on oxygen})$. 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[24] 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 $BW + 2 \times (\text{gestational age} - 20) - 6 \times$ 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[25] 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[26] 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

Table 2. Predictive Model

Author(s), Year	Institution, Country, Period	Level of Evidence	Predictive Model Studied (Covariates Used)	No. of Infants	Inclusion Criteria	Gestational Age (wks)	Birth Weight (g)
Termote et al, 2005[24]	Wilhemina Children's Hospital, The Netherlands, 1996–2000	III	Utrecht model (GA, BW, no. of erythrocyte transfusions during first 4 wks of life)	275	BW <1500 g or GA <32 wks or >3 days FiO_2 ≥0.4	Mean, 29; range, 25–34	Mean, 1115; range, 450–2080
Löfqvist et al, 2006[25]	Queen Silvia Hospital & Uppsala University Hospital, Sweden, 1999–2002	III (prospective study)	WINROP (version 1; PMA, BW, weight gain, IGF-I, IGFBP-3)	79	GA <32 wks	Median: ~28[§]; range, 24–32	Median: 1014; range, 530–2015
Yang and Donovan, 2009[14,¶]	UHC, United States, 1998–2003	III	UHC model (BW, GA, race, gender, multiple births)	357	BW 401–1250 g	Mean, 27; range, NR	Mean, 933; range, NR
Slidsborg et al, 2011[37]	Denmark National Birth Register, Denmark, 2002–2006	III	Denmark model (GA, BW)	4182	BW <1750 g or GA <32 wks	NR	NR
Binenbaum et al, 2011[38]	PINT (multicenter RCT), United States, 2000–2003	III (prospective study)	PINT (GA, BW, weight gain)	367	Enrolled in PINT study (BW <1000 g)	Median, 26; range, 22–34	Median: 800; range, 445–995
Binenbaum et al, 2012[18]	CHOP, United States, 2004–2009	II	CHOP (GA, BW, weight gain)	524	BW <1501 g or GA ≤30 wks	Mean, 28[§]; range, 23–33	Mean, 1031[§]; range, 400–1671
Eckert et al, 2012[39]	Hospital de Clinicas de Porto Alegre, Brazil, 2002–2009	III (prospective study)	ROPScore model (GA, BW, weight gain, oxygen, blood transfusion)	474	BW ≤1500 g and/or GA ≤32 wks	Mean, 30; range, NR	Mean, 1217; range, NR
Van Sorge et al, 2013[40]	NEDROP cohort, The Netherlands, 2009[††]	III	NEDROP (GA, BW, 1 or more of AV, NEC, sepsis, postnatal glucocorticoids, or cardiotonica)	1380	BW <1500 g and/or GA <32 wks or ≥3 days with ≥40% oxygen	Median, 30; range, 28–31	Median, 1260; range, 1020–1500
Ying et al, 2015[10]	e-ROP (multicenter RCT), North America, 2011–2013	III	e-ROP (gender, race/ethnicity, GA, BW, preplus, ROP stage, retinal hemorrhage, respiratory support, weight gain)	979	BW <1251 g	Mean, 27; range, 23–34	Mean, 860; range, 330–1250

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,
*Percent of infants for whom examinations can be reduced or eliminated (unless otherwise noted).
†Stage 3, 4, or 5 (van Sorge et al) or treated ROP (Binenbaum et al).
‡Study included internal validation, but measure corrected for overfit was not reported.
§A weighted mean was calculated from data in Table 1.
‖Sensitivity, specificity, PPV, NPV, and percent saved from examination are likely to be overestimated and percent of missed diagnoses underestimated
¶This study included 2 cohorts, the UHC cohort for model development and the Good Samaritan Hospital (GSH) cohort for model validation. The UHC in Table 3). Unlike other studies, the model evaluation metrics were reported at the eye level rather than the infant level.
#Not all infants were screened for ROP by indirect ophthalmoscopy. Determination of treatment-requiring ROP was based on inclusion in national
**Retinopathy of prematurity severe enough to require treatment according to the ETROP trial.
††The NEDROP cohort included all infants eligible for ROP screening in The Netherlands in 2009.

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.

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[19] 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

Development Studies

Outcome/ Prevalence (%)/No. of Outcomes	Sensitivity (95% Confidence Interval)	Specificity (95% Confidence Interval)	Positive Predictive Value (95% Confidence Interval)	Negative Predictive Value (95% Confidence Interval)	Missed Diagnoses (%)	Savings (%)*	Comments
Severe ROP†/5.1/14	100% (77%–100%)‡	24.5% (19%–30%)‡	6.7% (3.7%–11%)‡	100% (94%–100%)‡	0	23	Small sample size
Treatment required/7.6/6	100% (54%–100%)‖	83.6% (73%–91%)‖	33.3% (13%–59%)‖	100% (94%–100%)‖	0	20	No validation, small sample size
Prethreshold or threshold ROP/21/75	90.3% (84%–95%)	72.7% (69%–76%)	45.8% (40%–52%)	96.7% (95%–98%)	9.7	60 (13% reduction in mean no. of examinations)	Study population not representative of population at risk
Treatment required#/2.8/116	99.99% (99.93%–>99.99%)	NR	NR	NR	0.01	17	Not all children screened with indirect ophthalmoscopy
Severe ROP†/18.3/67	97% (92%–100%)	36% (32%–40%)	26% (20%–31%)‡	99% (95%–>99%)‡	3	30‡	Study population not representative of population at risk
Type 1 or 2 ROP/9.2/48	96% (88%–98%)	53% (49%–58%)	18% (13%–23%)‡	99.6% (98%–>99%)‡	4	49	
Severe ROP**/5.1/24	96% (79%–99.9%)‖	56% (51%–61%)‖	10.4% (7%–15%)‖	99.6% (98%–>99%)‖	4.2	53	No validation, small sample size
Severe ROP†/2.1/29	100% (88%–100%)	20% (18%–22%)	2.6% (1.8%–3.7%)‖	100% (99%–100%)‖	0	20	No validation, small sample size
RW-ROP/15.2/149	96% (91%–99%)	53% (49%–56%)	27% (23%–31%)‖	99% (97%–99.5%)‖	4	~45	Study population not representative of population at risk, no validation

of Philadelphia; CI = confidence interval; e-ROP = Evaluating Acute-Phase Retinopathy of Prematurity; GA = gestational age; IGF = insulin-like growth
PINT = Premature Infants in Need of Transfusion; PMA = postmenstrual age; PPV = positive predictive value; RCT = randomized clinical trial; ROP =
IGF-1, Neonatal, Retinopathy of Prematurity.

because of lack of validation or correction for overfit.
metrics are reported in this table. These are not corrected for overfit, but the study is credited for performing validation by use of the UHC cohort (reported

registries that were used to identify children treated for ROP or blind due to ROP.

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 al[27] (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 3. Predictive Model

Author(s), Year	Institution, Country, Period	Level of Evidence	Predictive Model Studied (Covariates Used)	No. of Infants	Inclusion Criteria	Gestational Age (wks)	Birth Weight (g)
Schalij-Delfos et al, 1996[22]	Wilhemina Children's Hospital, The Netherlands, 1987–1992	III	Modified S-index (GA, BW, oxygen)	312	BW <1500 g or GA <32 wks or high risk	NR	NR
Löfqvist et al, 2009[26]	Lund NICU, Sweden, 2005–2007	III (prospective study)	WINROP1 (GA, BW, weight gain, IGF-I, IGFBP-3)	50	GA <31 wks	Mean, 26; range, 23–31	Mean, 891; range, 460–1716
Hellstrom et al, 2009[19]	Sahlgrenska University Hospital, Sweden, 2004–2008	II	WINROP2 (GA, BW, weight gain)	353	GA <32 wks	Median, 29; range, 23–31	Median, 1290; range, 425–2210
Yang and Donovan, 2009[14,‡]	Good Samaritan Hospital, United States, 1998–2003	II	UHC model (BW, GA, race, gender, multiple births)	491	BW 401–1250 g	NR	NR
Wu et al, 2010[27]	Brigham & Women's Hospital, United States, 2005–2008	III	WINROP2 (GA, BW, weight gain)	318	GA <32 wks	Median, 29; range, 23–32	Median, 1050; range, 450–2400
Hård et al, 2010[28]	Hospital de Clinicas de Porto Alegre, Brazil, 2002–2008	III	WINROP2 (GA, BW, weight gain)	336	BW ≤1500 g or GA ≤32 wks or high risk	Median, 30; range, 24–33	Median, 1215; range, 505–2000
Wu et al, 2012[17]	10 level-3 NICUs, United States & Canada, 2006–2009	I	WINROP2 (GA, BW, weight gain)	1706	BW and GA met current criteria for screening	Median, 28; range, 22–31	Median, 1016; range, 378–2240
Zepeda-Romero et al, 2012[29]	Hospital Civil de Guadalajara, Mexico, 2005–2010	III	WINROP2 (GW, BW, weight gain)	192	GA <32 wks	Median, 30; range, 25–31	NR
Sun et al, 2013[30]	Zhengzhou Children's Hospital & Shanghai Children's Hospital, China, 2008–2011	III	WINROP2 (GA, BW, weight gain)	590	GA <32 wks	Median, 30; range, 26–32	Median, 1416; range, 800–2000
Choi et al, 2013[31]	Chonnam National University Hospital, South Korea, 2006–2008	III	WINROP2 (GA, BW, weight gain)	314	GA <32 wks	Mean, 29; range, 25–32	Mean, 1264; range, 505–2260
Lundgren et al, 2013[32]	EXPRESS cohort, Sweden, 2004–2007	III	WINROP2 (GA, BW, weight gain)	407	GA <27 wks	Median, 25 + 4 (wks+days); range, 23 + 0–26 + 6	Median, 784; range, 348–1315
Piyasena et al, 2014[33]	Simpson Center for Reproductive Health, Scotland, 1999–2009	III	WINROP2 (GA, BW weight gain)	410[¶]	GA <32 wks	Mean, 30; range, NR	Mean, 1340; range, NR
Eriksson et al, 2014[34]	Eskilstuna & Vasteras Hospital, Sweden, 2009–2011	III	WINROP2 (GA, BW, weight gain)	104	GA <32 wks	Mean, 29; range, 24–32	Mean, 1208; range, 477–2340
Ko et al, 2015[35]	Chang Gun Memorial Hospital, Taiwan, 2008–2010	III	WINROP2 (GA, BW, weight gain)	148	GA <32 wks	Median, 29; range, 23–32	Median, 1272; range, 565–1950

BW = birth weight; GA = gestational age; IGF = insulin-like growth factor-1; IGFBP-3 = insulin-like growth factor-1 binding protein 3; NICU = neonatal PMA = postmenstrual age; ROP = retinopathy of prematurity; UHC = University of Cincinnati; WINROP = Weight, IGF-1, Neonatal,
[*]Percent of infants for whom examinations can be reduced or eliminated (unless otherwise noted).
[†]Percent reduction in examinations by using a protocol to stratify the infants into high and low risk based on safety index and performing a single
[‡]This study included 2 cohorts, the UHC cohort for model development and the Good Samaritan Hospital (GSH) cohort for external model validation. limited to infants with BW <1251 g, this is the population in whom the algorithm was developed and intended to be used, so the validation study was not
[§]Any prethreshold ROP.
[‖]Any prethreshold, stage 3, or threshold ROP.
[¶]For complete weight data cohort.
[#]Any ROP in zone 1, stage 2 ROP in zone 2 with plus disease, or any stage 3 ROP (basically, any prethreshold ROP).

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,

Validation Studies

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

intensive care unit; NPV = negative predictive value; NR = not reported; PPV = positive predictive value; RCT = randomized clinical trial; Retinopathy of Prematurity.

screening examination on the low-risk patients.
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[17] (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 et al[29] (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[30] (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[31] (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[32] (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[33] (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[34] (level III) 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[35] (level III) 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[14] 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[36]). 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 prethreshold or worse ROP in each cohort, respectively. Of note, none of these prethreshold ROP eyes went on to require treatment.

Denmark Retinopathy of Prematurity Model. In 2011, Slidsborg et al[37] 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 al[38] 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 al[20] 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. The model was run at weekly intervals and an alarm was triggered when the risk of ROP was more than 0.085. The sensitivity and specificity of the final model in predicting severe ROP were 97% (95% CI, 92%−100%) and 36.1% (95% CI, 32%−40%), respectively. Application of the base model (not corrected for overfit) accurately predicted all infants with severe ROP except one who ultimately did not require treatment and would have resulted in 30% fewer infants undergoing examinations.

Children's Hospital of Philadelphia Retinopathy of Prematurity Model. In 2012, Binenbaum et al[18] published a level II study describing a multivariate risk model (the Children's Hospital of Philadelphia ROP model) developed from 524 patients from the United States. The predictive variables used in the model were GA, BW, and daily weight gain rate. Presence or absence of type 1 or 2 ROP was the BDOV for the study, and 48 patients (9.2%) showed positive results for the outcome. If at any point the predicted risk was greater than the cut point (set at 0.014), the child would undergo eye examinations from that point on. Internal validation with adjustment for overfit was performed, and with these applied, the sensitivity and specificity of the model in predicting type 1 or 2 ROP were 96% (95% CI, 88%−98%) and 53% (95% CI, 49%−58%), respectively. Application of the model accurately predicted all infants with type 1 ROP, missed 1 infant with type 2 ROP, and would have resulted in 255 fewer infants (49%) undergoing examinations. 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 416 (79%) fewer examinations would have been performed than if standard screening criteria had been applied. To create a simple clinical tool, the final model was converted into graphic form as nomograms.

ROPScore Model. In 2012, Eckert et al[39] 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 ROPScore 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 al[40] published a level III study describing a multivariate risk model developed and tested retrospectively on 1380 patients from The Netherlands. Five models were created, and the final model included the predictive variables of infants with GA of less than 30 weeks, BW of less than 1250 g, or both; or GA of 30 to 32 weeks, BW of 1250 to 1500 g, or both, with 1 or more of the following risk factors: artificial ventilation, necrotizing enterocolitis, sepsis, postnatal glucocorticoids, or cardiotonica. Sensitivity and specificity of the final model to detect severe ROP (which occurred in 2.1%) were 100% (95% CI, 88%−100%) and 20% (95% CI, 18%−22%), respectively. This model would have reduced the number of infants who required eye examinations by 20%.

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 preplus, 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.

Binenbaum[11] 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. Binenbaum 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. Binenbaum[11] 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[8] and likely acts as a common pathway for many other previously described risk factors.[38] 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).[29] 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 so no existing predictive model is ready for widespread application. Future research in this area should adhere to published guidelines for prognostic research[15,16,41,42] 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.[41] Current research in this area includes the National Institutes of Health-funded Postnatal Growth and Retinopathy of Prematurity studies.[43] 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

1. Kong L, Fry M, Al-Samarraie M, et al. An update on progress and the changing epidemiology of causes of childhood blindness worldwide. J AAPOS 2012;16:501–7.
2. 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.
3. Early Treatment for Retinopathy of Prematurity Cooperative Group. Revised indications for the treatment of retinopathy of

prematurity: results of the early treatment for retinopathy of prematurity randomized trial. Arch Ophthalmol 2003;121:1684–94.
4. Palmer EA, Flynn JT, Hardy RJ, et al; Cryotherapy for Retinopathy of Prematurity Cooperative Group. Incidence and early course of retinopathy of prematurity. Ophthalmology 1991;98:1628–40.
5. Pierce EA, Foley ED, Smith LE. Regulation of vascular endothelial growth factor by oxygen in a model of retinopathy of prematurity. Arch Ophthalmol 1996;114:1219–28.
6. Smith LE, Shen W, Perruzzi C, et al. Regulation of vascular endothelial growth factor-dependent retinal neovascularization by insulin-like growth factor-1 receptor. Nat Med 1999;5:1390–5.
7. Hellström A, Perruzzi C, Ju M, et al. Low IGF-I suppresses VEGF-survival signaling in retinal endothelial cells: direct correlation with clinical retinopathy of prematurity. Proc Natl Acad Sci U S A 2001;98:5804–8.
8. Quinn GE, Gilbert C, Darlow BA, Zin A. Retinopathy of prematurity: an epidemic in the making. Chin Med J (Engl) 2010;123:2929–37.
9. Kemper AR, Wallace DK. Neonatologists' practices and experiences in arranging retinopathy of prematurity screening services. Pediatrics 2007;120:527–31.
10. Ying GS, Quinn GE, Wade KC, et al; e-ROP Cooperative Group. Predictors for the development of referral-warranted retinopathy of prematurity in the telemedicine approaches to evaluating acute-phase retinopathy of prematurity (e-ROP) study. JAMA Ophthalmol 2015;133:304–11.
11. Binenbaum G. Algorithms for the prediction of retinopathy of prematurity based on postnatal weight gain. Clin Perinatol 2013;40:261–70.
12. Reynolds JD, Dobson V, Quinn GE, et al; CRYO-ROP and LIGHT-ROP Cooperative Study Groups. Evidence-based screening criteria for retinopathy of prematurity: natural history data from the CRYO-ROP and LIGHT-ROP studies. Arch Ophthalmol 2002;120:1470–6.
13. Hikino S, Ihara K, Yamamoto J, et al. Physical growth and retinopathy in preterm infants: involvement of IGF-I and GH. Pediatr Res 2001;50:732–6.
14. Yang MB, Donovan EF. Risk analysis and an alternative protocol for reduction of screening for retinopathy of prematurity. J AAPOS 2009;13:539–45.
15. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ 2009;338:b604.
16. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.
17. Wu C, Löfqvist C, Smith LE, et al. Importance of early postnatal weight gain for normal retinal angiogenesis in very preterm infants: a multicenter study analyzing weight velocity deviations for the prediction of retinopathy of prematurity. Arch Ophthalmol 2012;130:992–9.
18. Binenbaum G, Ying GS, Quinn GE, et al. The CHOP postnatal weight gain, birth weight, and gestational age retinopathy of prematurity risk model. Arch Ophthalmol 2012;130:1560–5.
19. Hellström A, Hård AL, Engström E, et al. Early weight gain predicts retinopathy in preterm infants: new, simple, efficient approach to screening. Pediatrics 2009;123:e638–45.
20. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
21. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 1983;78:316–31.
22. Schalij-Delfos NE, Zijllmans BL, Cats BP. Towards a universal approach for screening of retinopathy of prematurity (ROP). Doc Ophthalmol 1996;92:137–44.
23. Meier-Gibbons F, Korner F, Bossi E, Fauchere JC. Value of the RPM safety index in retinopathy of prematurity [in German]. Klin Monbl Augenheilkd 1991;198:487–8.
24. Termote JU, Donders AR, Schalij-Delfos NE, et al. Can screening for retinopathy of prematurity be reduced? Biol Neonate 2005;88:92–7.
25. Löfqvist C, Andersson E, Sigurdsson J, et al. Longitudinal postnatal weight and insulin-like growth factor I measurements in the prediction of retinopathy of prematurity. Arch Ophthalmol 2006;124:1711–8.
26. Löfqvist C, Hansen-Pupp I, Andersson E, et al. Validation of a new retinopathy of prematurity screening method monitoring longitudinal postnatal weight and insulinlike growth factor I. Arch Ophthalmol 2009;127:622–7.
27. Wu C, Vanderveen DK, Hellström A, et al. Longitudinal postnatal weight measurements for the prediction of retinopathy of prematurity. Arch Ophthalmol 2010;128:443–7.
28. Hård AL, Löfqvist C, Fortes Filho JB, et al. Predicting proliferative retinopathy in a Brazilian population of preterm infants with the screening algorithm WINROP. Arch Ophthalmol 2010;128:1432–6.
29. Zepeda-Romero LC, Hård AL, Gomez-Ruiz LM, et al. Prediction of retinopathy of prematurity using the screening algorithm WINROP in a Mexican population of preterm infants. Arch Ophthalmol 2012;130:720–3.
30. Sun H, Kang W, Cheng X, et al. The use of the WINROP screening algorithm for the prediction of retinopathy of prematurity in a Chinese population. Neonatology 2013;104:127–32.
31. Choi JH, Löfqvist C, Hellström A, Heo H. Efficacy of the screening algorithm WINROP in a Korean population of preterm infants. JAMA Ophthalmol 2013;131:62–6.
32. Lundgren P, Stoltz Sjöström E, Domellöf M, et al. WINROP identifies severe retinopathy of prematurity at an early stage in a nation-based cohort of extremely preterm infants. PLoS One 2013;8:e73256.
33. Piyasena C, Dhaliwal C, Russell H, et al. Prediction of severe retinopathy of prematurity using the WINROP algorithm in a birth cohort in South East Scotland. Arch Dis Child Fetal Neonatal Ed 2014;99:F29–33.
34. Eriksson L, Lidén U, Löfqvist C, Hellström A. WINROP can modify ROP screening praxis: a validation of WINROP in populations in Sörmland and Västmanland. Br J Ophthalmol 2014;98:964–6.
35. Ko CH, Kuo HK, Chen CC, et al. Using WINROP as an adjuvant screening tool for retinopathy of prematurity in southern Taiwan. Am J Perinatol 2015;30:149–54.
36. Section on Ophthalmology American Academy of Pediatrics, American Academy of Ophthalmology, American Association for Pediatric Ophthalmology and Strabismus. Screening examination of premature infants for retinopathy of prematurity. Pediatrics 2006;117:572–6.
37. Slidsborg C, Forman JL, Rasmussen S, et al. A new risk-based screening criterion for treatment-demanding retinopathy of prematurity in Denmark. Pediatrics 2011;127:e598–606.
38. Binenbaum G, Ying GS, Quinn GE, et al. A clinical prediction model to stratify retinopathy of prematurity risk using postnatal weight gain. Pediatrics 2011;127:e607–14.
39. Eckert GU, Fortes Filho JB, Maia M, Procianoy RS. A predictive score for retinopathy of prematurity in very low birth weight preterm infants. Eye (Lond) 2012;26:400–6.

40. van Sorge AJ, Schalij-Delfos NE, Kerkhoff FT, et al. Reduction in screening for retinopathy of prematurity through risk factor adjusted inclusion criteria. Br J Ophthalmol 2013;97:1143–7.
41. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.
42. Moons KG, Royston P, Vergouwe Y, et al. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.
43. National Institutes of Health. Research Portfolio Online Reporting Tools (RePORT). Postnatal growth and retinopathy of prematurity (G-ROP) studies. Project number 5R01EY021137−03. Available at: http://projectreporter.nih.gov/project_info_description.cfm?aid=8712493&icde=25399089. Accessed August 12, 2015.

Footnotes and Financial Disclosures

Originally received: November 3, 2015.
Final revision: November 4, 2015.
Accepted: November 4, 2015.
Available online: January 28, 2016. Manuscript no. 2015-1943.

[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@aao.org.