Algorithms for the Prediction of Retinopathy of Prematurity based upon Postnatal Weight Gain

Gil Binenbaum, MD, MSCE
Gil Binenbaum: binenbaum@email.chop.edu

Synopsis
Current ROP screening guidelines represent a simple risk model, consisting of two dichotomized factors, birth weight (BW) and gestational age at birth (GA). The pioneering work of Smith, Hellstrom, Lofqvist, and colleagues has shown that while many risk factors have been associated with severe ROP, low insulin-like growth factor 1 (IGF-1) may represent a common pathway through which many of these risk factors act to influence ROP risk, and that this influence may be captured by tracking postnatal weight gain, a surrogate measure for IGF-1. In countries with highly developed neonatal care systems, predictive models including weight gain have demonstrated accurate ROP risk assessment and a large reduction in the number of ROP examinations, compared to current guidelines based upon BW and GA alone. These models include WINROP, a computer-based algorithm that identifies cumulative deviations from expected weight gain; ROPSscore, a spreadsheet-based equation; and CHOP ROP, a nomogram-based model that consists of a single equation with continuous terms for BW, GA, and weight gain. However, clinicians should appreciate important limitations of the models, including the need for larger studies to achieve a highly precise estimate of sensitivity and the poor generalizability of the models to countries where higher BW and GA infants require treatment, as low IGF-1 may play less of a role in the pathogenesis of ROP in these larger and older babies.

Keywords
Infant; nomogram; prediction model; retinopathy of prematurity; risk model

Introduction
Prognostic statistical models allow clinicians to predict a patient’s risk of developing a specific medical outcome. Although not typically described in such terms, our current retinopathy of prematurity (ROP) screening criteria are nothing more than a simple prediction model with two dichotomized (yes or no) predictors, birth weight (BW) and gestational age at birth (GA). A need for exams is determined by the degree of prematurity at birth by using two cut-off levels, and other risk factors are not considered in a systematic fashion. For example, in the United States, babies with either BW<1501 grams or GA ≤30 weeks receive exams. A priority is placed upon avoiding blindness in even a single child,
and this screening “model” has high sensitivity (it catches almost all cases of severe ROP), but it is not specific (most examined infants do not develop severe ROP). In fact, less than five per cent of infants examined require laser surgery, based upon multiple large US\textsuperscript{3–6}, Canadian\textsuperscript{7}, and UK\textsuperscript{8, 9} studies.

**Growth, IGF-1, and ROP**

In an effort to improve this specificity, postnatal growth-based ROP risk models have been developed. The models are based upon ground-breaking work by Smith, Hellstrom, Lofqvist, and colleagues, who have provided our current understanding of the pathophysiology underlying ROP.

ROP develops in two phases, a hypoxic preclinical phase, during which slow postnatal growth can be used to predict risk, and a subsequent proliferative clinical phase. These phases result from alterations in serum insulin-like growth factor 1 (IGF-1), a somatic growth factor, and retinal vascular endothelial growth factor (VEGF), a hypoxia-induced vasoproliferative factor necessary for normal retinal vascular development.\textsuperscript{10} Serum IGF-1 falls with premature birth from loss of maternal sources and poor endogenous production.\textsuperscript{11–14} Importantly, IGF-1 plays a permissive role in VEGF-induced retinal vascular growth.\textsuperscript{15, 16} Therefore, low serum IGF-1 hinders retinal vessel development, with localized hypoxia and VEGF accumulation, as metabolic demands increase within the developing retina. With increasing age and size, endogenous production of IGF-1 rises, permitting VEGF activity, and proliferative retinopathy develops.

Much laboratory work supports this model.\textsuperscript{10, 15–21} Clinically, multiple investigators have demonstrated that both prolonged early IGF-1 deficits and slow postnatal weight gain are associated with a higher risk of subsequent severe ROP.\textsuperscript{11, 22–30} Serum IGF1 levels correlate with fetal and postnatal growth, so postnatal growth is a good surrogate measure for serum IGF-1.\textsuperscript{12, 14, 31–33} In addition, weight measurements are simple, quick, cheap, and routinely collected, while IGF-1 assays are costly and require blood, a laboratory, and processing time.

**Postnatal Growth ROP Models**

Specific postnatal growth based ROP predictive models are discussed below. In evaluating these models, it is most useful to consider their performance with regards to sensitivity for detecting severe ROP and the reduction in children requiring eye exams that would have resulted from their use, which is a more clinically intuitive measure than specificity. Severe ROP has been variably defined as Early Treatment of ROP Study type 1 ROP, treated ROP, and or stage 3 ROP in the studies.

All three models use weight gain as a measure of postnatal growth and consider BW and GA in the determination of risk of ROP along with growth. BW and GA are still the most significant risk factors for ROP in countries with highly developed Neonatal Intensive Care Unit (NICU) systems.\textsuperscript{3–7, 9, 34–37} The smaller and younger an infant was at birth, the greater the risk of ROP and of severe ROP. Likely pathophysiologic correlates include the degree of retinal vascular immaturity and low postnatal endogenous production of serum IGF-1. However, this predictive information is lost by using only BW and GA cut-off levels. Treating BW and GA as continuous or ordinal variables addresses this issue, but then another factor, such as slow postnatal growth, must be introduced to identify higher BW and GA infants at risk for severe ROP.

With regards to additional risk factors, current screening guidelines do include an “unstable clinical course” as judged by the neonatologist as a criteria by which to examine larger
infants. However, in practice, postnatal factors are not considered in a systematic fashion, even though multiple other risk factors for ROP have been described, such as excessive supplemental oxygen, necrotizing enterocolitis (NEC), intraventricular hemorrhage (IVH), anemia, apnea, sepsis, and blood transfusions. Based upon the studies below, it is hypothesized that most such factors may act via a common pathway, by lowering serum IGF-1 levels, and may be “captured” in a predictive model simply by considering postnatal weight gain. Supplemental oxygen is an important exception, particularly in countries with developing neonatal care systems, where higher BW and GA infants may develop severe ROP via a pathophysiologic mechanism unrelated to low serum IGF-1. The limited potential for weight gain ROP models to be used in such infants is a critical issue discussed below.

**WINROP**

Based upon their work on IGF-1 and ROP, Lofqvist, Hellstrom, Smith, et al. developed a computer-based ROP risk algorithm named WINROP to detect slowdowns in postnatal weight gain, predict severe ROP, and greatly reduce the number of infants requiring exams. WINROP uses a cumulative-deviations statistical approach in multiple steps: each week the infant’s actual weight is compared to an expected growth curve of infants who developed no or mild ROP; the differences or deviations between the expected weight and the actual weight are accumulated from week to week; and when these cumulative deviations surpass a threshold alarm level, the risk of severe ROP is categorized using dichotomized cut off levels for BW, GA, and alarm timing to determine a need for eye exams.

WINROP demonstrated very high sensitivity for detecting severe ROP in retrospective studies: 100% in a Swedish cohort of 353 infants, reducing infants who would have received exams by 76%; 100% in a Boston cohort of 318 infants, reducing infants who need exams by 75%; and most recently fell slightly to 98.6% in a larger, multi-center US and Canadian cohort of 1,706 infants. When WINROP was studied in countries with developing NICU systems, however, the sensitivity fell further: 91% in a Brazilian cohort of 366 infants, and 55% (85% if GA<32 weeks, 5% if GA ≥32 weeks) in a Mexican cohort of 352 infants.

The WINROP algorithm involves complex calculations, which have limited transparency for the user. However, the developers have created a web-based application (www.winrop.com) that permits the user to enter BW, GA, and weight measurements, allows tracking of multiple infants, and provides a simple indication of low or high risk status for developing site-threatening ROP. The authors suggest that WINROP be used as “an adjunct to and not a replacement for standard ophthalmological screening.” They describe abbreviated exam schedules for low-risk infants, as opposed to no examinations at all.

**ROPScore**

Eckert, Filho, et al. recently developed a less complex model named ROPScore. The model consists of a logistic regression equation, which is used to calculate risk only once per child, using an Excel spreadsheet. The model includes continuous rather than dichotomized terms for BW and GA, weight gain at a single time point (6 weeks postnatal age) as a proportion of BW, and dichotomous terms for blood transfusion and use of oxygen in mechanical ventilation during the first six weeks of life. Assuming a specific cut off level for low or high risk, ROPScore had a sensitivity of 98% and specificity of 56% for treatment-requiring ROP in a development cohort of 474 Brazilian infants. Additional studies are underway. As with WINROP, the authors suggest that ROPScore be used not to determine overall screening criteria but rather to reduce the frequency of exams in low risk infants.
PINT ROP and CHOP ROP

Binenbaum et al. developed a simpler logistic regression based model named PINT ROP.\textsuperscript{30} Prospectively collected data from 367 infants with BW<1000 g in the Premature Infants in Need of Transfusion (PINT) randomized controlled trial were used to develop a model containing terms only for BW, GA, and daily weight gain rate, which was calculated from the current and prior week’s weight measurements. The equation is calculated on a weekly basis, and if the predicted risk of ROP is greater than a cut point level, exams are indicated and the equation does not need to be re-calculated again. In this manner, PINT ROP had 100% sensitivity for treated ROP, while reducing the number of infants requiring exams by 30% in the high risk cohort.\textsuperscript{30}

Of note, numerous additional candidate predictors were evaluated, including ethnicity, perinatal and postnatal comorbidities, and medical and surgical interventions. All these factors fell out in multivariate analyses\textsuperscript{30}, supporting the hypothesis that many previously described risk factors for ROP act through a common pathway, by affecting IGF1 levels, and are therefore “captured” through weight measurements.

The PINT ROP cohort was at high risk for ROP. Therefore, the investigators applied the same modeling approach to a low risk cohort more representative of current US ROP screening criteria (BW<1501 g), to develop an updated model called CHOP ROP.\textsuperscript{52} Among 524 infants, the model had 100% sensitivity for type 1 ROP, while reducing the number of infants requiring exams by 49%. If the risk cut off was raised to miss 1 infant requiring laser, the reduction in exams was 79%, suggesting a trade off that might be explored further. There was a small (3%) advantage to using daily versus weekly weight measurements. Paper nomograms to simplify use of the models were created based upon the PINT ROP model and updated with the CHOP ROP model (Figure 1), but the authors stressed that further studies are needed prior to clinical use.\textsuperscript{30, 52}

Model Development and Complexity

The creation of a usable clinical prediction model is a stepwise process.\textsuperscript{1} First, a development study is undertaken, containing as large a sample size as possible in order to avoid over fitting of the model.\textsuperscript{1, 53} Over fitting may occur when the complexity of the model is high relative to the number of outcome events (cases of severe ROP) in the cohort, and the model effectively describes random error rather than a true association. Such a model will not perform well in a new groups of patients. Second, the model must be validated in new patients and if necessary updated, preferably using both the original and new datasets.\textsuperscript{1, 54, 55} Model updating may involve adjusting the model coefficients or even adding new variables if necessary.\textsuperscript{55} In the case of the CHOP ROP model, the study cohort had a significantly different risk profile than the PINT ROP cohort, so the model coefficients and alarm cut point needed to be adjusted.

The final step is to evaluate the impact of the model in clinical use.\textsuperscript{156} Successful implementation of a clinical predictive model requires physician acceptance, which depends upon transparency, ease of use, and in this case confidence that no cases of severe ROP will be missed.\textsuperscript{56} Transparency and ease of use relate to model complexity, while “confidence” relates most directly to the precision of the point estimate of sensitivity of the model for predicting severe ROP. A cumulative-deviations model is relatively more complex and to some degree lacks transparency with regards to the calculations being done to determine risk. The process becomes more user-friendly with the use of a web-based application.\textsuperscript{22, 23, 25} A logistic regression equation based model provides a simpler calculation of risk, and it can be represented as a paper nomogram, which does not require any calculation to be done (Figure 1).
Sample Size: An important limitation

Despite studies in cohorts ranging from 300 to 1700 infants, sample size remains a limiting factor. While a point estimate of sensitivity of 100% has been reported, the confidence interval (CI) around that point estimate is likely too wide for clinicians to have confidence that infants with severe ROP would not be missed. The reason is that the width of the CI is driven by the number of cases of severe ROP, not the overall number of infants. For example, the CHOP ROP study involved 524 infants, but only 20 infants had severe ROP, so while the sensitivity was 100%, the 95% CI was 84% to 100%. Even in the most recent, multi-center WINROP study, which involved 1706 infants and reported a sensitivity of 98.6%, the CI was only 96.7% to 100%.

To have sufficient confidence in a model to justify its use for primary “all or none” screening decisions, the lower boundary of the CI arguably should be much higher, perhaps above 99%, which would require a dataset with hundreds of cases of ROP. Therefore, it may be best to conservatively consider these studies as still in the “development” stage, and only once that high degree of precision of sensitivity (very narrow CI) is obtained would subsequent validation and updating studies be undertaken. Larger studies might also identify confounding medical factors, the presence of which would default an infant to receive eye examinations. However, alternative perspectives and approaches may be considered. Rather than replacing current screening guidelines, a model may be used alongside current guidelines, changing exam frequency or timing based upon predicted risk. WINROP has been used in Sweden in this fashion, based upon repeatedly high sensitivity demonstrated across multiple WINROP studies. Alternatively, a sensitivity of less than 100% might be acceptable under some circumstances, such as when ophthalmologic resources are limited and must be rationed. Finally, a model’s alarm level could be lowered to more confidently ensure 100% sensitivity, while accepting a decrease in specificity. This approach could work well in a multi-tiered screening approach together with telemedicine, once both modalities have been sufficiently validated: the risk model determines which infants require fundus photographs, and photographic grading determines which infants require eye examination by an ophthalmologist.

Generalizability: A second important limitation

There are several reasons why a predictive model may behave poorly in new patients, including the methods used to design the model and differences in health care systems and patient characteristics. There is good evidence that the generalizability of models such as WINROP and CHOP ROP to countries where higher BW and GA infants develop severe ROP will be limited, and that separate model development studies need to be carried out in such populations. When WINROP was applied to a Brazilian cohort, it demonstrated less than 90% sensitivity, and when it was applied to a Mexican cohort, sensitivity was only 55%. ROPScore was developed in Brazil, and model performance was higher. However, additional terms for blood transfusion and oxygen supplementation are included in the model, and performance across multiple settings still must be assessed, as there is high variability in ROP risk profile across neonatal centers in the country, even within the same city.

The poor performance of postnatal weight gain ROP models in countries with developing neonatal care system may be related to differences in ROP pathophysiology, particularly in older GA infants. At older post menstrual ages, endogenous production of IGF-1 has already increased, so that low IGF-1 may play less of a role in the pathogenesis of severe ROP. Rather, ROP in such infants might be driven primarily by high oxygen exposure, which causes inhibition of VEGF and retinal blood vessel destruction, as it does in oxygen-induced
animal models of ROP. This hypothesis is supported by the dramatic difference in the performance of WINROP among Mexican infants with GA<32 weeks (85% sensitivity) and GA ≥32 weeks (5% sensitivity).

Benefits and Future Studies

Improved ROP risk assessment could have broad-reaching benefits in neonatology, ophthalmology, and public health. Revised ROP screening guidelines might reduce both the number of children requiring stressful diagnostic eye examinations and the frequency of exams for lower-risk infants. Professional and infrastructure resources may be better allocated to high risk infants, particularly in areas and countries with limited resources, where better targeted resources could lower the burden of blindness due to ROP. Tiered approaches, combining a growth-based risk model with telemedicine assessments, might further improve the efficiency of screening. In general, the cost effectiveness of ROP diagnostic examinations would be increased.

Postnatal weight gain based ROP prediction models have great potential to influence the clinical management of ROP. Further development, validation, and clinical impact studies across multiple international settings can and should be pursued, but there are no set epidemiological rules by which to determine when sufficient validation has been achieved to safely use a clinical prediction model; criteria depend upon the clinical context. ROP screening involves high stakes decisions, and perhaps model validation and updating will best be viewed as an ongoing process. The studies discussed above have established “proof of concept.” And if even larger development and validation studies continue to demonstrate very high sensitivity for predicting severe ROP, then the research objective will shift towards simply maximizing the predictive value of the data upon which ROP screening guidelines are based. Retrospective pooling of datasets by investigators and an ongoing prospective data registry would permit a model to be updated in an ongoing fashion and maximize the use of available data. This type of collaborative approach would best ensure that all infants at risk for site threatening ROP undergo diagnostic eye examinations, while the potential benefits of postnatal growth based ROP models are realized.

References


Key points

1. Slow postnatal growth is a surrogate measure for low serum insulin-like growth factor 1, which is an important risk factor for severe ROP.

2. Risk models that consider postnatal weight gain, along with birth weight and gestational age, include WINROP, ROPScore, and CHOP ROP.

3. These models predict severe ROP with much greater specificity than current ROP screening guidelines.

4. Two important limitations are study sample size and poor generalizability in countries with developing neonatal care systems.

5. Postnatal growth based models have great potential to reduce the number of diagnostic ROP examinations being performed.
Figure 1.
Sample nomogram to predict the risk of severe retinopathy of prematurity based upon the CHOP ROP model. A straight line is drawn between the values for birth weight and daily weight gain rate. The intersection of this line with the grey auxiliary axis is then connected to the value for GA. The intersection of this second line with the probability line provides the predicted probability of severe ROP. If the risk is greater than 0.014, eye exams are indicated. NOTE: This nomogram requires further validation and is not intended for clinical use at this time. From 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(12):1560–1565; with permission.