The CHOP Postnatal Weight Gain, Birth Weight, and Gestational Age Retinopathy of Prematurity Risk Model

Gil Binenbaum, MD, MSCE; Gui-shuang Ying, PhD; Graham E. Quinn, MD, MSCE; Jiayan Huang, MS; Stephan Dreisieitl, PhD; Jules Antigua, MD; Negar Foroughi, MD; Soraya Abbasi, MD

Objective: To develop a birth weight (BW), gestational age (GA), and postnatal–weight gain retinopathy of prematurity (ROP) prediction model in a cohort of infants meeting current screening guidelines.

Methods: Multivariate logistic regression was applied retrospectively to data from infants born with BW less than 1501 g or GA of 30 weeks or less at a single Philadelphia hospital between January 1, 2004, and December 31, 2009. In the model, BW, GA, and daily weight gain rate were used repeatedly each week to predict risk of Early Treatment of Retinopathy of Prematurity type 1 or 2 ROP. If risk was above a cut-point level, examinations would be indicated.

Results: Of 524 infants, 20 (4%) had type 1 ROP and received laser treatment; 28 (5%) had type 2 ROP. The model (Children’s Hospital of Philadelphia [CHOP]) accurately predicted all infants with type 1 ROP; missed 1 infant with type 2 ROP, who did not require laser treatment; and would have reduced the number of infants requiring examinations by 49%. Raising the cut point to miss one type 1 ROP case would have reduced the need for examinations by 79%. Using daily weight measurements to calculate weight gain rate resulted in slightly higher examination reduction than weekly measurements.

Conclusions: The BW-GA-weight gain CHOP ROP model demonstrated accurate ROP risk assessment and a large reduction in the number of ROP examinations compared with current screening guidelines. As a simple logistic equation, it can be calculated by hand or represented as a nomogram for easy clinical use. However, larger studies are needed to achieve a highly precise estimate of sensitivity prior to clinical application.

Arch Ophthalmol. 2012;130(12):1560-1565
prospective cohort of 367 infants with BWs less than 1000 g in which the model was developed, the PINT ROP model accurately identified all 33 infants requiring laser treatment, while reducing the number of infants requiring examinations by 30%.12 As infants with higher BW are at lower risk for developing treatment-requiring ROP, application of the same approach to a cohort more representative of current screening guidelines may result in a greater reduction in the number of infants requiring diagnostic examinations.

We sought to develop a predictive model applying the same modeling approach as the PINT ROP model in a cohort of infants meeting current US screening guidelines. We hypothesized that the model would accurately predict the risk of type 1 or 2 ROP while demonstrating a significant reduction in the need for examinations. We also sought to determine the use of daily vs weekly weight measurements in the calculation of the rate of postnatal weight gain and of adding a BW and GA interaction term to the model.

METHODS

The study was approved by the joint Institutional Review Board of the Hospital of the University of Pennsylvania and Pennsylvania Hospital (Philadelphia) and was carried out in compliance with the principles of the Declaration of Helsinki and the US Health Insurance Portability and Accountability Act.

SUBJECTS

Eligible subjects were infants born between January 1, 2004, and December 31, 2009, at Pennsylvania Hospital with a BW less than 1301 g or GA of 30 weeks or less and with a known ROP outcome, defined as diagnosis in either eye of type 1 or 2 ROP (defined below) or diagnosis in each eye of 1 of the following: regressing or regressed ROP that had not met type 1 or 2 ROP criteria, immature retinal vasculature in zone III without prior ROP in zones I or II, or mature retinal vasculature. There were no further medical or surgical exclusion criteria. Type 1 ROP, which was an indication for treatment with retinal laser photocoagulation, was defined according to the Early Treatment of Retinopathy of Prematurity (ETROP) Study type 1 high-risk prethreshold ROP criteria: stage 2 or 3 ROP in zone II with plus disease, stage 3 disease in zone I without or without plus disease, or stage 1 or 2 disease in zone 1 with plus disease. Type 2 ROP, for which observation was indicated, was also defined according to ETROP criteria: stage 1 or 2 ROP in zone I without plus disease or stage 3 ROP in zone II without plus disease.

DATA COLLECTION

Data were retrospectively collected from the medical records of infants meeting the inclusion criteria. Medical data included BW, GA, and all available weight measurements, which were typically taken daily. ROP data included stage of disease, zone of disease, and presence or absence of plus disease for each eye at each examination and any laser or other treatments performed. In the PINT ROP model development study, numerous additional covariates were considered, including maternal race, sex, use of medications (systemic corticosteroids, erythropoietin, methyloxanthines, doxapram, and indomethacin), red blood cell transfusion volume, patent ductus arteriosus diagnoses or surgery, necrotizing enterocolitis diagnosis or surgery, other surgical procedures (hernia repair and others), intraventricular hemorrhage, ventriculomegaly, abnormal head ultrasound, perinatal infection, postnatal sepsis, and cerebrospinal fluid infection. However, none of these factors were significant predictors of ROP in the model with BW, GA, and weight gain. Therefore, such data were not considered for the current study.

STATISTICAL ANALYSIS

Analyses were done using SAS statistical software (version 9.1; SAS, Inc.). Multivariate logistic regression was used to develop a model containing terms for BW, GA, and weight gain rate to calculate the probability of type 1 or 2 ROP. The structure of the PINT ROP model containing these terms was used as a starting point for the new model.13 However, when a predictive model is applied to a new cohort with significant differences in the distribution of the predictor variables, as with BW and GA in this study, it is often necessary to update the model.19 As the development cohort of the PINT ROP model was at much higher risk for ROP than the current study cohort, it was anticipated that updating would be necessary, including recalibration of the coefficients and the alarm threshold.19 The addition of an interaction term for BW and GA was evaluated as well because small-for-GA birth weight has been reported to be a risk factor for ROP.20,21 Daily weight gain rate was calculated alternatively using weekly measurements (the difference between the current and prior weeks’ weight, divided by 7), as it had been in the PINT ROP study, and using daily measurements (the difference between the average of all available daily weights during the prior 7 days and the average of the penultimate week’s daily weights, divided by 7). Weight measurements after severe ROP developed were excluded. Very low-birth-weight infants show an initial reduction in weight through the first week of life,22 so weight gain measurements before age 7 days were also excluded.

A need for eye examinations was indicated if the predicted risk of type 1 or 2 ROP was greater than an alarm cut-point level. The cut point was set empirically to evaluate the performance of the model at multiple levels, including a level to ensure that all type 1 ROP cases were captured and a level at which one type 1 ROP case was missed. In this manner, the trade-off between identifying severe ROP and reducing the number of examinations was explored. The predicted risk was reassessed on a weekly basis, beginning with the second week of life, to determine whether examinations would be indicated. If at any week the predicted risk was greater than the cut-point level, then the child would receive eye examinations from that point going forward according to the routine schedule used clinically, and further weekly evaluation of risk for that infant using the model would be unnecessary.

The performance of the model was assessed by calculating the sensitivity and specificity for detecting type 1 or 2 ROP combined, the sensitivity for detecting type 1 ROP alone, the reduction in the number of infants requiring eye examinations, and the interval between first alarm and diagnosis of type 1 or 2 ROP. Because model development and performance assessment were completed on the same data set, we evaluated optimistic bias in the performance of the model by completing an internal validation using the bootstrap methods of Harrell et al.23 The validation is based on 1000 bootstrap replicates, each consisting of 524 subjects sampled with replacement. With each replicate, a prediction model was developed, and its performance was evaluated on both the replicate and the original data by calculating the sensitivity and specificity for predicting type 1 or 2 ROP using the same cut point of predicted probability.
as chosen for the original data set. The “optimism” in sensitivity or specificity was the difference between that from the bootstrap replicate and that from the original data set using the replicate prediction model. The mean optimisms and 95% CIs were calculated based on 1000 replicates.

To create a sample clinical tool to determine ROP risk, the final logistic model was converted into graphic form as a nomogram with Mathematica software (version 7; Wolfram Research) using previously described methods.\(^{24}\)

### RESULTS

Five hundred twenty-four infants met the inclusion criteria (Table 1). Twenty infants (4%) developed type 1 ROP and received laser retinal photocoagulation. An additional 28 infants (5%) reached type 2 ROP but regressed spontaneously and did not require treatment. No infants developed zone 1 ROP.

The base model contained terms for BW, GA, and daily weight gain rate calculated using daily weight measurements (Table 2). The probability of severe ROP was calculated on a weekly basis for each child, and when the calculated risk was greater than 0.014, the child was flagged as needing examinations. In this manner, the model accurately predicted all the infants with type 1 ROP, missed 1 infant with type 2 ROP, and would have resulted in 255 fewer infants (49%) undergoing examinations (Table 3). The sensitivity for predicting type 1 or 2 ROP was 98% (95% CI, 89% to 100%), specificity 53% (49% to 58%), positive predictive value 17% (13% to 22%), and negative predictive value 100% (98% to 100%). The sensitivity for predicting type 1 ROP was 100% (95% CI, 84% to 100%). The median time between alarm and diagnosis of type 1 ROP was 9.5 weeks (range, 4.0-15.3 weeks). When the cut point was raised to miss just one case of type 1 ROP (>0.159), 6 cases of type 2 ROP were missed, and the model would have resulted in 416 fewer infants (79%) undergoing examinations. When weekly weight measurements were used to calculate the daily weight gain rate, and the risk cut point set at greater than 0.010, the reduction in examinations was slightly less (46% with weekly vs 49% with daily).

An interaction term between BW and GA was added to the base predictive model and the cut-point level set at 0.055. Applied in this fashion, the model that included an interaction term identified all infants with type 1 ROP, missed 4 infants with type 2 ROP, and would have resulted in 379 fewer infants (72%) requiring examinations (Table 3). However, the \(P\) value for the interaction term was .50, and the model coefficients and relative risks were impractically large (data not shown), suggesting that the model was overfitted.

From 1000 bootstrap replicates, the average optimism for sensitivity of type 1 or 2 ROP for the base model was 1.6% (95% CI, −5.1% to 8.4%) and for specificity 0.09% (95% CI, −3.3% to 3.4%), providing, based on this internal validation technique, estimates of sensitivity of
96% (88% to 98%) and specificity of 53% (49% to 58%). A pilot nomogram was created based on the regression equation for the base model (Figure).

Building on the PINT ROP model, we developed an updated model in a cohort that included larger, more mature infants. The model uses BW, GA, and daily weight gain rate to predict the risk that an infant will develop severe ROP and does so early enough to ensure ROP examinations are scheduled. The model contains revised equation coefficients, intercept, and alarm cut points. Therefore, to distinguish between the models, we will refer to the new model as the CHOP ROP model. The model accurately predicted all the cases of type 1 ROP while reducing the number of infants requiring examinations by 49%. As anticipated, this reduction was greater than the 30% reduction observed in the PINT ROP study because of the difference in risk profiles between the 2 cohorts. The high-risk PINT ROP infants all had a BW less than 1000 g, while the low-risk CHOP ROP cohort was more representative of current US screening guidelines. The 49% reduction is less than the 76% reduction seen when the Swedish WINROP model was applied to a similar low-risk US cohort, but WINROP involves a more complex, multistep statistical algorithm.

The use of daily rather than weekly weight measurements resulted in a slight improvement in the specificity of the model. The addition of an interaction term also seemed to improve model performance. This finding would be consistent with the reported association between small-for-GA birth weights and severe ROP. Accounting for this effect by including an interaction term in the model resulted in a greater reduction (72% vs 49%) in the number of infants requiring examinations. However, we are concerned that the significance level for the interaction term and unrealistically large relative risks for the GA levels suggest overfitting to the data. Overfitting can occur when the number of outcome events is small in comparison to the complexity of the model, and the model is, in fact, describing random error rather than a true underlying association. In this study, the higher GA strata contained as expected very few cases of type 1 or 2 ROP, and the high performance of the interaction model may be due to capturing a single case or two of severe

### Table 2. Multivariate Logistic Regression Analysis of the Association of Candidate Predictors With the Development of Type 1 or 2 ROP

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Regression Coefficient (SE)</th>
<th>Relative Risk (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base modela</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−1.50 (1.54)</td>
<td></td>
<td>.008</td>
</tr>
<tr>
<td>Gestational age, wk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>4.24 (1.24)</td>
<td>69.7 (6.2 to 791)</td>
<td>.0006</td>
</tr>
<tr>
<td>24</td>
<td>3.49 (1.06)</td>
<td>32.8 (4.1 to 263)</td>
<td>.001</td>
</tr>
<tr>
<td>25</td>
<td>3.60 (1.05)</td>
<td>36.7 (4.7 to 288)</td>
<td>.0006</td>
</tr>
<tr>
<td>26</td>
<td>2.33 (1.09)</td>
<td>10.3 (1.2 to 86.5)</td>
<td>.03</td>
</tr>
<tr>
<td>27</td>
<td>2.48 (1.17)</td>
<td>11.9 (1.2 to 118)</td>
<td>.03</td>
</tr>
<tr>
<td>28</td>
<td>2.05 (1.25)</td>
<td>7.7 (0.7 to 90.5)</td>
<td>.10</td>
</tr>
<tr>
<td>29-33</td>
<td>1.0 [Reference]</td>
<td>1.0 [Reference]</td>
<td></td>
</tr>
<tr>
<td>Birth weight, g</td>
<td>−0.0037 (0.0015)</td>
<td>0.996 (0.993 to 0.999)</td>
<td>.02</td>
</tr>
<tr>
<td>Weight gain rate calculated using weight measurements, g/d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>−0.0186 (0.0058)</td>
<td>0.982 (0.970-0.993)</td>
<td>.001</td>
</tr>
<tr>
<td>Weekly</td>
<td>−0.0140 (0.0039)</td>
<td>0.986 (0.978-0.994)</td>
<td>.0005</td>
</tr>
</tbody>
</table>

**Abbreviations:** BW, birth weight; GA, gestational age; ROP, retinopathy of prematurity; SE, standard error.

*a The equation for calculating the probability of type 1 or 2 ROP was as follows: Probability of type 1 or 2 ROP = 1/[1 + exp (−risk score)], where risk score = (−1.50) + (4.24 if GA = 23) + (3.49 if GA = 24) + (3.60 if GA = 25) + (2.33 if GA = 26) + (2.48 if GA = 27) + (2.05 if GA = 28) + [−(0.0037) × (BW)] + [(−0.0186) × (weight gain rate)], where weight gain rate was calculated as [(mean of daily weights of the preceding week − mean of daily weights of the penultimate week) / 7].

### Table 3. The Prediction of Type 1 ROP (n = 20) and Type 2 ROP (n = 28) Based on Birth Weight, Gestational Age, and Daily Weight Gain Rate in 524 Premature Infants at Risk for ROP

<table>
<thead>
<tr>
<th>Weight</th>
<th>Alarm Cut Point</th>
<th>ROP</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Reduction in Infants Requiring Examinations</th>
<th>Time From Alarm to Diagnosis, Median (Range), wk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>0.014</td>
<td>Type 1</td>
<td>20 (100)</td>
<td>255 (50.6)</td>
<td>255 (48.7)</td>
<td>9.5 (4.0-15.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type 1 or 2</td>
<td>47 (97.9)</td>
<td>254 (53.4)</td>
<td>255 (48.7)</td>
<td>10.0 (5.7-17.3)</td>
</tr>
<tr>
<td>Weekly</td>
<td>0.010</td>
<td>Type 1</td>
<td>20 (100)</td>
<td>239 (47.4)</td>
<td>239 (45.6)</td>
<td>9.8 (5.4-15.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Type 1 or 2</td>
<td>47 (97.9)</td>
<td>238 (50.0)</td>
<td>239 (45.6)</td>
<td>10.0 (5.7-17.3)</td>
</tr>
</tbody>
</table>

**Abbreviation:** ROP: retinopathy of prematurity.

©2012 American Medical Association. All rights reserved.
and GA once an alarm has sounded.13 Finally, with regard primarily tracks growth and only secondarily considers BW can arise with the use of a multistep algorithm that pri-

25-week GA infant) are consistently flagged to undergo ex-

required. Another advantage of this single-equation regres-

tion-based model include that it allows risk to be cal-

Table 1. Sample nomogram to predict the risk of severe retinopathy of prematurity (ROP) based on the CHOP (Children’s Hospital of Philadelphia) ROP model. A straight line is drawn between the values for birth weight (BW) and daily weight gain rate. The intersection of this line with the gray auxiliary axis is then connected to the value for gestational age (GA). The intersection of this second line with the probability line provides the predicted probability of severe ROP. If the risk is greater than .014, eye examinations are indicated. Note: This nomogram requires further validation and is not intended for clinical use at this time.

ROP. Such a model may not perform well when applied to new data. Therefore, we are cautious about accepting the performance of the interaction model and believe that assessment in a new, larger data set is necessary to evaluate further the validity of the BW-GA interaction.

Transparency, ease of use, and physician acceptance are important factors for the successful clinical implementation of a prediction model.19 Advantages of a simple regression-based model include that it allows risk to be calculated directly and it may be represented as a nomogram for easy clinical use (Figure).12,24 With further development and validation, a printed nomogram could be used to plot BW, GA, and weight gain on a weekly basis to determine the risk for ROP examinations. Alternatively, a simple risk calculator could be incorporated into the electronic medical record. In either case, once a need for an examination is signaled, future assessments would not be required. Another advantage of this single-equation regression model is that infants with very low BW or GA (eg, a 25-week GA infant) are consistently flagged to undergo examinations because these factors are weighed heavily in the model. Neonatologists are unlikely to forgo ROP examinations for such infants, despite a statistical model predicting that examinations are unnecessary. This situation can arise with the use of a multistep algorithm that primarily tracks growth and only secondarily considers BW and GA once an alarm has sounded.13 Finally, with regard to workload, BW, GA, and serial weight measurements are already routinely recorded by the neonatal team and with a reduced number of infants requiring examinations, the burden of coordinating ROP examinations would likely be decreased by the use of this risk stratification process.

The accurate prediction of type 1 ROP with only BW, GA, and serial assessments of postnatal weight gain supports the hypothesis that many risk factors for ROP may act via a common pathway that is captured by consider-

Lower risk

Higher risk

Alarm level

Probability line

Daily weight gain, g/d

Auxiliary axis

BW, g

GA, wk

23
24-25
26-27
28
29-34

400 600 800 1000 1200 1400 1600 1800

0– 40
20
40
60
80

©2012 American Medical Association. All rights reserved.

ing weight gain, a surrogate measure for serum IGF-1 levels. A low IGF-1 level limits vascular endothelial growth factor activity and retinal blood vessel growth,17,18 and factors like sepsis and acidosis have been shown to lower serum IGF-1 levels.3,25 Numerous factors were considered in the PINT ROP study and were not significant when weight gain was introduced into the model12; they are listed in the “Methods” section.

It is reassuring that the bootstrap internal validation suggested minimal optimistic bias in the estimates of sensitivity and specificity. However, the study had important limitations. The sample size was too small to obtain a highly precise estimate of the model’s sensitivity for detecting severe ROP. While the point estimate of sensitivity for detecting type 1 ROP was 100%, the lower boundary of the 95% CI was only 84% because, although 524 infants were studied, the number of infants with type 1 ROP (n = 20) was still small, making the CI wide. The same is true of the PINT ROP study12 and the WINROP studies in Sweden13 and Boston.14 Studies an order of magnitude larger, with hundreds of outcome events,26 will be required to raise the lower CI boundary high enough (>99%) to provide high confidence in the sensitivity of the model. Therefore, we believe it is important to characterize these models as still being in the development stage. External validity will be determined in subsequent studies after model development is complete, and only then should clinical use be considered. Larger studies may also uncover false-negative signals from conditions that cause weight gain but are not associated with increased IGF-1 levels. In countries with developing neonatal care systems, higher BW and GA infants may develop type 1 ROP.1,27 Recalibrated or even restructured models will need to be developed separately for those populations. For example, the sensitivity of WINROP was only 90% in Brazil28 and 55% in Mexico (85% for GA <32 weeks, 5% for GA ≥32 weeks).29 Finally, the performance of the model may be altered by the selection of the alarm cut point level, with a trade-off between sensitivity and reduced examinations. Raising the risk cutpoint to a level where one type 1 ROP case was missed resulted in a nearly 80% reduction in the number of infants undergoing examinations. A consensus will need to be reached to determine which trade-offs are most acceptable to physicians, and those decisions may require further consideration as more treatment modalities are developed.

In this study, the CHOP ROP model provided accurate risk assessment and a large reduction in ROP examinations. Consideration of postnatal weight gain in determining which infants should undergo ROP examinations and treating BW and GA as continuous rather than dichotomized risk factors could represent important advances beyond the current guidelines. Models, such as CHOP ROP and WINROP, have the potential to greatly reduce the screening burden, identify early those infants who might benefit from preventive interventions, and possibly be used alongside other innovations, such as fundus imaging, in a tiered approach to ROP screening. However, larger studies are needed to achieve high confidence in the accuracy of a model before it can be applied safely to clinical practice.
Submitted for Publication: April 8, 2012; final revision received June 21, 2012; accepted June 24, 2012.

Correspondence: Gil Bienenbaum, MD, MSCE, The Children’s Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104 (bienenbaum@email.chop.edu).

Author Contributions: Dr Bienenbaum had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by grants K12 EY015398 and P30 EY01583-26 from the National Institutes of Health.

Role of the Sponsors: The funding organization has played no role in the study design, conduct, analysis, or manuscript preparation.

Previous Presentations: Presented in part at the 2012 Annual Meeting of the American Association of Pediatric Ophthalmology and Strabismus; March 27, 2012; San Antonio, Texas, and the 2012 Annual Meeting of the Pediatric Academic Societies; April 29, 2012; Boston, Massachusetts.

Additional Contributions: Haresh Kirpalani, MD, MSc, reviewed the manuscript, and Karen Karp, BSN, Emidio Sivieri, MSE, and Toni Mancini, BSN, helped with data collection.

REFERENCES


