Title: Postnatal Growth and Retinopathy of Prematurity Model Development Study

Study Key Name G-ROP-1
Protocol No: 12-009727
Protocol Date: November 13, 2012

Amendment 1 Date: Amendment 4 Date:
Amendment 2 Date: Amendment 5 Date:
Amendment 3 Date: Amendment 6 Date:

Funding Source: National Institute of Health (NIH)/National Eye Institute (NEI)
Grant number 1R01EY021137-01A1

Study Principal Investigator:

Gil Binenbaum MD MSCE
The Children’s Hospital of Philadelphia
Ophthalmology, 9-MAIN
34th Street and Civic Center Blvd
Philadelphia, PA, 19104
Phone 215-590-4594
binnenbaum@email.chop.edu
# TABLE OF CONTENTS

Table of Contents ............................................................................................................. 2  
Abbreviations and Definitions of Terms ........................................................................... 4  
Abstract ............................................................................................................................... 6  

1 BACKGROUND INFORMATION AND RATIONALE ......................................................... 8  
   1.1 INTRODUCTION .................................................................................................. 8  
   1.2 COMPLIANCE STATEMENT ............................................................................. 8  
   1.3 RELEVANT LITERATURE AND DATA .............................................................. 8  

2 STUDY OBJECTIVES ........................................................................................................ 12  
   2.1 PRIMARY OBJECTIVE (OR AIM) ...................................................................... 13  
   2.2 SECONDARY OBJECTIVES (OR AIM) ............................................................. 13  

3 INVESTIGATIONAL PLAN ................................................................................................. 13  
   3.1 GENERAL SCHEMA OF STUDY DESIGN ....................................................... 13  
   3.2 STUDY DURATION, ENROLLMENT AND NUMBER OF SITES ....................... 13  
      3.2.1 Duration of Study for Subject ................................................................ 13  
      3.2.2 Total Number of Study Sites/Total Number of Subjects Projected .......... 13  
   3.3 STUDY POPULATION ....................................................................................... 13  
      3.3.1 Inclusion Criteria ..................................................................................... 13  
      3.3.2 Exclusion Criteria ................................................................................... 14  
      3.3.3 Case ascertainment ............................................................................... 14  
      3.3.4 Data sources .......................................................................................... 14  
   3.4 BIAS AND BLINDING ....................................................................................... 15  

4 STUDY PROCEDURES .................................................................................................... 15  
MEDICAL RECORDS DATA COLLECTION. SEE SECTION 6.1 MEDICAL RECORDS DATA ABSTRACTION BELOW FOR DETAILED INFORMATION. .................................................. 15  

5 STUDY ENDPOINTS AND EVALUATIONS .................................................................. 15  
   5.1 PRIMARY ENDPOINT ...................................................................................... 15  
   5.2 SECONDARY ENDPOINTS .............................................................................. 15  

6 MEASUREMENTS AND EVALUATIONS ........................................................................ 16  
   6.1 MEDICAL RECORDS DATA ABSTRACTION ................................................... 16  
      6.1.1 ROP Data collection ............................................................................... 16  
      6.1.2 Candidate predictors data collection ...................................................... 16  
      6.1.3 Costs data collection ............................................................................. 16  
      6.1.4 Protected Health Information (PHI) ....................................................... 17  

7 STATISTICAL CONSIDERATIONS ................................................................................. 17  
   7.1 PRIMARY AND SECONDARY ENDPOINTS .................................................... 17  
   7.2 CLINICAL ENDPOINT ...................................................................................... 18  
   7.3 STATISTICAL METHODS .................................................................................. 18  
      7.3.1 Model development .............................................................................. 18  
      7.3.2 Model performance assessment ........................................................... 19  
      7.3.3 Alternative strategies for model development ....................................... 19  
      7.3.4 Cost-effectiveness analysis .................................................................. 20  
   7.4 SAMPLE SIZE AND POWER ........................................................................... 21  

8 SAFETY MANAGEMENT ................................................................................................. 21  

9 STUDY ADMINISTRATION ............................................................................................. 22  
   9.1 DATA COLLECTION AND MANAGEMENT ..................................................... 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2</td>
<td>CONFIDENTIALITY</td>
<td>24</td>
</tr>
<tr>
<td>9.3</td>
<td>REGULATORY AND ETHICAL CONSIDERATIONS</td>
<td>25</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Data and Safety Monitoring Plan</td>
<td>25</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Risk Assessment</td>
<td>25</td>
</tr>
<tr>
<td>9.3.3</td>
<td>Potential Benefits of Trial Participation</td>
<td>25</td>
</tr>
<tr>
<td>9.3.4</td>
<td>Risk-Benefit Assessment</td>
<td>25</td>
</tr>
<tr>
<td>9.4</td>
<td>RECRUITMENT STRATEGY (OR CASE ASCERTAINMENT)</td>
<td>25</td>
</tr>
<tr>
<td>9.5</td>
<td>INFORMED CONSENT/ASSENT</td>
<td>25</td>
</tr>
<tr>
<td>9.5.1</td>
<td>Waiver of Consent</td>
<td>25</td>
</tr>
<tr>
<td>9.5.2</td>
<td>Waiver of HIPAA authorization</td>
<td>26</td>
</tr>
<tr>
<td>9.6</td>
<td>PAYMENT TO SUBJECTS/FAMILIES</td>
<td>26</td>
</tr>
<tr>
<td>9.7</td>
<td>G-ROP STUDY GROUP</td>
<td>26</td>
</tr>
<tr>
<td>9.7.1</td>
<td>Steering Committee</td>
<td>26</td>
</tr>
<tr>
<td>9.7.2</td>
<td>Study Headquarters</td>
<td>27</td>
</tr>
<tr>
<td>9.7.3</td>
<td>Clinical Centers</td>
<td>27</td>
</tr>
<tr>
<td>9.7.4</td>
<td>Data Coordinating Center</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>PUBLICATION</td>
<td>27</td>
</tr>
<tr>
<td>11</td>
<td>REFERENCES</td>
<td>27</td>
</tr>
</tbody>
</table>
### ABBREVIATIONS AND DEFINITIONS OF TERMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Adverse event</td>
</tr>
<tr>
<td>BW</td>
<td>Birth weight</td>
</tr>
<tr>
<td>CE</td>
<td>Cost effectiveness</td>
</tr>
<tr>
<td>CHOP</td>
<td>Children’s Hospital of Philadelphia</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CPOB</td>
<td>Center for Preventive Ophthalmology and Biostatistics at the University of Pennsylvania</td>
</tr>
<tr>
<td>CRCU</td>
<td>Clinical Research Computing Unit at the University of Pennsylvania</td>
</tr>
<tr>
<td>CRF</td>
<td>Case report form</td>
</tr>
<tr>
<td>CRYO-ROP</td>
<td>Cryotherapy for Retinopathy of Prematurity (Study)</td>
</tr>
<tr>
<td>DBA</td>
<td>Oracle Database Administrator</td>
</tr>
<tr>
<td>DCC</td>
<td>Data Coordinating Center</td>
</tr>
<tr>
<td>DMS</td>
<td>Data Management System</td>
</tr>
<tr>
<td>ETROP</td>
<td>Early Treatment of Retinopathy of Prematurity (Study)</td>
</tr>
<tr>
<td>GA</td>
<td>Gestational age</td>
</tr>
<tr>
<td>HIPAA</td>
<td>Health Insurance Portability and Accountability Act</td>
</tr>
<tr>
<td>HUP</td>
<td>Hospital of the University of Pennsylvania</td>
</tr>
<tr>
<td>ICROP</td>
<td>International Classification of ROP</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Insulin-Like Growth Factor 1</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>IV</td>
<td>Intravenous</td>
</tr>
<tr>
<td>IVH</td>
<td>Intraventricular hemorrhage</td>
</tr>
<tr>
<td>MRN</td>
<td>Medical record number</td>
</tr>
<tr>
<td>NEC</td>
<td>Necrotizing enterocolitis</td>
</tr>
<tr>
<td>NEI</td>
<td>National Eye Institute</td>
</tr>
<tr>
<td>NICU</td>
<td>Neonatal intensive care unit</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institute of Health</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative predicted value</td>
</tr>
<tr>
<td>PENN</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>PHI</td>
<td>Protected health information</td>
</tr>
<tr>
<td>PINT</td>
<td>Premature Infants in Need of Transfusion (a research study)</td>
</tr>
<tr>
<td>PMA</td>
<td>Postmenstrual age</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive predicted value</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality adjusted life years</td>
</tr>
<tr>
<td>RDC</td>
<td>Oracle Clinical Remote Data Capture</td>
</tr>
<tr>
<td>ROP</td>
<td>Retinopathy of Prematurity</td>
</tr>
<tr>
<td>RW-ROP</td>
<td>Referral Warranted ROP (any zone I disease, stage 3 disease, or plus disease)</td>
</tr>
<tr>
<td>SAE</td>
<td>Severe adverse event</td>
</tr>
<tr>
<td>Se</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Sp</td>
<td>Specificity</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
</tbody>
</table>
ABSTRACT

Context:
Retinopathy of prematurity (ROP) is a leading cause of preventable childhood blindness. Diagnosis involves repeated, stressful, resource-intensive eye examinations. Under current screening guidelines, less than 5% of infants require laser surgery. Slow postnatal growth, a surrogate measure for low serum insulin-like growth factor 1 (IGF-1), is associated with subsequent development of severe ROP and can be used to predict which infants are at risk for severe ROP and therefore require eye exams. However, studies to date have been limited by sample size.

Objectives:
1. Develop a prognostic model that includes postnatal weight gain, birth weight, and gestational age at birth to predict infants who are likely to develop severe ROP
2. Evaluate the relative cost-effectiveness of the prognostic model versus conventional ROP screening guidelines

Study Design:
Multi-center retrospective cohort study

Setting/Participants:
- Setting:
  Study headquarters: CHOP
  Clinical centers: 19 (18 US, 1 Canada), including CHOP/HUP as 1 center
  Data coordinating centers: CHOP and PENN
- Participants, primary inclusion criteria:
  1. Born between January 1 2006 and December 31 2011
  2. Birth weight (BW)<1501 g, or gestational age (GA)<32 weeks, or had ROP exams
  3. ROP outcome known
- Chart review period: January 1, 2006 to June 30, 2012
- Sample size: 9,850 infants overall, 649 infants at CHOP/HUP clinical center

Study Interventions and Measures:
- Procedures: medical records data abstraction
- Main study outcome measures:
1. Sensitivity and specificity for predicting severe ROP
2. Reduction in the number of infants requiring eye examinations compared to current screening guidelines
3. Incremental cost-effectiveness ratios (cost per severe ROP case detected, cost per quality-adjusted life years)
1 BACKGROUND INFORMATION AND RATIONALE

1.1 Introduction

Retinopathy of prematurity (ROP) is a leading cause of preventable blindness in children.\textsuperscript{1,2} ROP can be effectively treated with laser retinal ablative surgery if diagnosed in time,\textsuperscript{3-5} but detection involves subjecting at-risk infants to physically stressful, resource-intensive and costly, serial diagnostic eye exams. Risk of ROP and the need for eye examinations is currently assessed using birth weight and gestational age dichotomized cut off levels, but these criteria have poor specificity for identifying severe ROP. In the US, of an estimated 65,000 babies per year undergoing exams,\textsuperscript{6} less than 5% of infants examined require laser surgery.\textsuperscript{5,7-12} Nevertheless, because of the serious consequences of missing a potential case of blindness, the protocol must maintain high sensitivity, even at the cost of repeatedly examining children who never require treatment, many of whom never develop any retinopathy. Following insights into the role of serum insulin-like growth factor 1 (IGF-1) in the pathogenesis of ROP, recent studies have demonstrated that slow postnatal weight gain, a surrogate measure for low IGF-1, is a predictor of the subsequent development of severe ROP.\textsuperscript{13-19} If a predictive model could be developed and validated to more accurately predict ROP risk, changes to current screening guidelines could be proposed in order to reduce the number of infants requiring examinations. However, the studies done to date have been limited by inadequate sample size and in some cases model complexity.

1.2 Compliance Statement

This study will be conducted in full accordance with all applicable Children’s Hospital of Philadelphia Research Policies and Procedures and all applicable Federal and state laws and regulations including 45 CFR 46, and the HIPAA Privacy Rule. Any episode of noncompliance will be documented.

The investigators will perform the study in accordance with this protocol, will obtain consent and assent (unless a waiver is granted), and will report unexpected problems in accordance with The Children’s Hospital of Philadelphia IRB Policies and Procedures and all federal requirements. Collection, recording, and reporting of data will be accurate and will ensure the privacy, health, and welfare of research subjects during and after the study.

1.3 Relevant Literature and Data

Current Screening Criteria

The current US ROP screening guidelines represent a simple prediction model with two dichotomized predictors: birth weight (BW) and gestational age (GA) at birth. The model has high sensitivity (catches almost all cases of severe ROP) but is not specific (most examined infants do not develop severe ROP), as a priority is placed upon avoiding blindness in even a single child. A need for exams is determined by the degree of prematurity at birth by using two cut-off levels, one for BW (<1501 grams) and one for GA (<30, though <32 weeks is commonly observed due to a publication error).\textsuperscript{20,21} Babies with either BW or GA beneath these levels receive exams. BW and GA are still the most significant risk factors for ROP in countries with highly developed Neonatal Intensive Care Unit (NICU) systems.\textsuperscript{5,7,8,10-12,22-25} The smaller and younger an infant was at birth, the
greater the risk of ROP and of severe ROP. However, this predictive information is lost by using only BW and GA cut-off levels. The development of a model that treats BW and GA as continuous or ordinal variables will address this issue. Likely pathophysiologic correlates represented by low BW and GA include retinal vascular immaturity and low endogenous production of serum IGF-1 (discussed below).

A more accurate prognostic model is likely to reduce exams primarily in larger BW and GA infants, as the majority of sight-threatening ROP occurs in infants with lower BW and GA. US 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.8, 23, 26-32 Based upon preliminary studies described 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.

Few infants screened in the US need treatment. We estimate the proportion to be 5% or less, based upon multiple studies multiple large US,5, 7, 8, 11, Canadian10, and UK8, 12 studies (all 3 have similar ROP guidelines12, 33). Conversely, treatable disease can develop in infants falling outside current guidelines. Raising the GA cut-off to <33 weeks would capture these infants, but the cost in terms of additional unnecessary negative screening exams would be high, as the incidence of such cases is very small.26, 27 A model that treats GA and BW as continuous variables will not identify these infants either, as the risk decreases as GA and BW increase. Therefore, another method is needed to identify such infants, such as tracking postnatal growth—a system that could be integrated simply, as weights are collected as part of daily neonatal care. In addition, since current guidelines are not 100% sensitive, a new model could improve sensitivity by identifying infants with higher BW or GA.

ROP Pathogenesis

ROP develops in two phases, a hypoxic preclinical phase, during which we believe slow postnatal growth can be used to predict risk, and a subsequent proliferative clinical phase. These phases result from alterations in serum IGF-1, a somatic growth factor, and retinal vascular endothelial growth factor (VEGF, a hypoxia-induced vasoproliferative factor necessary for normal retinal vascular development).34 Serum IGF-1 falls with premature birth from loss of maternal sources and poor endogenous production.35-38 IGF-1 plays a permissive role in VEGF-induced retinal vascular growth.39, 40 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. Extensive laboratory work, primarily by Smith et al., supports this model.34, 39-45 Clinically, both prolonged early IGF-1 deficits and slow postnatal weight gain are associated with a higher risk of subsequent severe ROP.14, 15, 17, 19, 35, 46-50 Serum IGF1 levels correlate highly with fetal and postnatal growth, so postnatal growth is a good surrogate measure for serum IGF-1.36, 38, 51-53, and weight measurements are simple, quick, cheap, and routinely collected, while IGF-1 assays are costly and require blood, a laboratory, and processing time. Note that this pathogenic model does not apply to higher GA infants born in countries with developing neonatal care systems.
Growth-based ROP prediction models

**WINROP:** Based upon their groundbreaking work on IGF-1 and ROP, Lofqvist, Smith, Hellstrom 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.\(^{46,54-56}\) WINROP uses a cumulative-deviations statistical approach: 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, risk for 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 need exams by 76% \(^{48}\); 100% in a Boston cohort 318 infants, reducing infants who need exams by 75% \(^{15}\); and most recently fell slightly to 98.6% in a larger, multi-center US and Canadian cohort of 1,706 infants\(^{16}\). When WINROP was studied in countries with developing NICU systems, however, the sensitivity fell further: 91% in a Brazilian cohort of 366 infants\(^{57}\), and 55% (85% if GA<32 weeks, 5% if GA≥32 weeks) in a Mexican cohort of 352 infants\(^{58}\).

**ROPScore:** Eckert et al. developed a somewhat less complex model named ROPScore in a Brazilian cohort of 474 infants.\(^{18}\) Their 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. The model is only calculated once per child using an Excel spreadsheet. The model had a sensitivity of 98% and specificity of 56% in the Brazilian development cohort.

**PINT ROP:** Binenbaum et al. developed an even simpler logistic regression based model, named PINT ROP after the dataset used to develop the model.\(^{13}\) Prospectively collected data from 367 infants with BW<1000 g in the Premature Infants in Need of Transfusion (PINT) Study was used to develop a logistic regression model containing terms for BW, GA, and daily weight gain rate, calculated from the current and prior week’s weight measurements. Numerous additional candidate predictors were evaluated, including ethnicity, perinatal and postnatal comorbidities, and medical and surgical interventions, but all these factors fell out in multivariate analyses. Therefore, we hypothesize that many previously described systemic risk factors for ROP act through a common pathway, i.e., by affecting IGF1 levels, and are therefore “captured” through weight measurements. The PINT ROP equation was calculated on a weekly basis, and if the predicted risk of ROP is greater than a cut point level, exams will be indicated and the equation does not need to be re-calculated again. In this manner, the PINT ROP correctly predicted in advance all 33 infants who required laser surgery (100% sensitivity for treated ROP) and 66 of 67 infants who developed stage 3 ROP, while reducing the number of infants requiring exams by 30%.

**CHOP-ROP:** The PINT ROP cohort was at high risk for ROP. The investigators subsequently applied the same modeling approach to a low risk cohort (BW<1500 g) more representative of current US ROP screening criteria, in order to develop an updated model called CHOP ROP.\(^{59}\) Among the 524 infants, the model accurately predicted all 20 infants receiving laser while reducing the number of infants requiring exams by 49%. If the cut off was raised to miss 1 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.
Model complexity

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. Transparency and ease of use relate to model complexity, while “confidence” relates to the precision of the point estimate of sensitivity of the model for predicting severe ROP, which is addressed below under “Sample size.” A cumulative-deviations model such as WINROP is relatively complex and lacks transparency. Clinical data are entered into a computer, which performs multi-stage calculations, and it is not possible to calculate ROP risk by hand. 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, for even simpler clinical application. Sample nomograms were created for the PINT ROP and CHOP ROP models (Figure).

Finally, because the logistic based models treat BW and GA as continuous rather than dichotomous variables, all very low BW and GA infants are flagged to received examinations, because the models weigh these factors heavily. In contrast, the cumulative deviations based models may indicate that an infant born at 26 weeks GA does not require exams, a decision with which a neonatologist would likely feel uncomfortable.

Limitations and the need for additional studies

The studies described above have important limitations, which must be addressed with additional and much larger studies before changes to ROP screening guidelines can be considered. Two primary limitations are sample size and generalizability.

Sample size: The models were developed using cohorts that were too small. Development of a prognostic model ideally occurs with a much larger dataset, including several hundred outcome events (i.e., cases of severe ROP). While a point estimate of sensitivity of 100% is reported, the confidence interval (CI) around that point estimate is too wide for clinicians to have confidence that infants with severe ROP would not be missed. For example, the
CHOP ROP study involved 524 infants, but only 20 had severe ROP, so while the sensitivity was 100% the 95% confidence interval 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%. Arguably, in order to have sufficient confidence in a model to use it alone for screening decisions, the lower boundary of that CI should be much higher, above 99%, which would require a dataset with hundreds of cases of ROP. Therefore, it is best to conservatively consider these studies as still in “development,” and only once that high degree of precision of sensitivity (very narrow CI) is obtained would subsequent validation studies be undertaken. The primary purpose of the G-ROP model development study in this protocol is to obtain that desired sample size, and a second G-ROP model validation study will be undertaken under a separate protocol.

**Generalizability:** 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 these models to countries where higher BW and GA infants develop severe ROP (countries with developing neonatal care systems) will be limited. When the WINROP model was applied to a Brazilian cohort of infants, it demonstrated less than 90% sensitivity, and when applied to a Mexican cohort, sensitivity was only 55%. Additional model development studies need to be carried out in such populations. The ROPScore study was carried out in Brazil and with additional terms (blood transfusion, oxygen), model performance was higher. A likely explanation for why separate models are required for older GA infants is that at older post menstrual ages, endogenous production of IGF-1 has already increased, so that low IGF-1 plays less of a role in the pathogenesis and tracking growth will not provide adequate prediction of severe ROP. Rather, ROP is driven primarily by high oxygen exposure, which cause inhibition of VEGF and vessel destruction, as it does in oxygen-induced animal models of ROP. This hypothesis is supported by the dramatic difference in performance of WINROP among Mexican infants with GA<32 weeks (85% sensitivity) and GA≥32 weeks (5% sensitivity).

**Economics**

The Institute of Medicine estimates that prematurity costs the US more than $26 billion a year, a significant proportion of acute pediatric hospital care. More accurate ROP risk assessment may help to reduce this enormous sum. However, severe visual impairment, such as from high-stage ROP, is associated with a lifetime economic burden of more than $615,000 per affected individual (2003 USD). If improved specificity is associated with even a slight decrement in sensitivity, missed cases resulting in blindness would have a net negative economic impact, in addition to tragic health consequences. In an increasingly constrained health care system, such economic trade-offs must be examined rigorously, so health policy decisions optimize use of scarce resources to generate the most health. Although changes in screening frequency, birth weight thresholds, and risk factors have been shown to affect the cost-effectiveness of ROP management, no analyses have been performed of a formal prognostic model such as the one to be developed in this protocol.

2  STUDY OBJECTIVES

The purpose of the study is to develop a prognostic model using postnatal weight gain to identify infants who are likely to develop severe ROP among a large, diverse retrospective
cohort of premature infants. The cohort must be large enough to obtain a highly precise estimate of the sensitivity of the model for predicting severe ROP.

2.1 Primary Objective (or Aim)

The primary objective of this study is to develop a prognostic model using postnatal weight gain to identify premature infants who are likely to develop severe ROP.

2.2 Secondary Objectives (or Aim)

The secondary objective of the study is to evaluate the relative cost-effectiveness of the prognostic model versus conventional ROP screening guidelines.

3 INVESTIGATIONAL PLAN

3.1 General Schema of Study Design

Multi-center retrospective cohort study

3.2 Study Duration, Enrollment and Number of Sites

3.2.1 Duration of Study for Subject

Infants born between January 1, 2006, and December 31, 2011, will be included in the study, with follow-up data collected through June 30, 2012, as follows. Eye exam data will be collected until (1) retinal vascular maturity, (2) disease regression, or (3) retinal vascularization into zone III without prior disease in zones I or II. Typically, ROP examinations begin at 31 weeks post menstrual age (PMA) or 4 weeks chronological age, whichever occurs later, and go until 40 to 44 weeks PMA, including outpatient exams, but may extend further, particularly if treatment involves intraocular injection of bevacizumab. Inpatient medical data will be collected from birth until discharge from the hospital or 40 weeks PMA, whichever occurs first.

3.2.2 Total Number of Study Sites/Total Number of Subjects Projected

Study data will be collected at 18 investigative centers in the United States and 1 center in Toronto, Canada. Some academic centers care for children at more than one hospital in their health system, so approximately 34 NICU’s are expected to be represented in the dataset. The study headquarter is CHOP.

Study-wide subjects: 9,850 ENROLLED subjects are projected in order to obtain 8,422 EVALUABLE infants (see section 7.4 Sample Size below for further details).

3.3 Study Population

3.3.1 Inclusion Criteria

1. Males or female born between January 1, 2006, and December 31, 2011
2. BW<1501 grams or GA<32 weeks or had ROP exams
3. ROP outcome known, for which the infant must meet either A or B below:
   A. EITHER eye had one of the following:
a. Severe ROP (Any stage disease in Zone I, or any stage 3 disease, or plus disease)

b. Treated ROP (laser, cryo, bevacizumab injections, retinal detachment surgery)

B. EACH eye had one of the following:

a. Regressing stage 1 or 2 disease in Zone II

b. Regressed stage 1 or 2 disease in Zone II

c. Mature retinal vasculature

d. Immature retinal vasculature in Zone III without prior disease in Zone I or II

3.3.2 Exclusion Criteria
Death or loss to follow-up before ROP outcome determined.

3.3.3 Case ascertainment

Enrollable subjects will be identified by querying neonatology and ophthalmology billing records and clinical databases at each study center for infants born with BW<1501 g or born with GA<32 weeks or who had ROP exams during the study period. These queries will include searches for premature infants who received ROP exams based upon ophthalmology billing charge diagnoses (ICD codes for ROP and prematurity), ophthalmology and neonatology clinical databases of infants receiving exams (often weekly lists are maintained to track which infants require exams), and neonatology clinical (e.g., hospital census) and billing databases for infants born within the BW and GA criteria. In this manner, the sample of enrolled infants will be representative of the population of infants that would meet current ROP screening guidelines (BW<1501 g, GA<32 weeks, or larger infants with a poor postnatal course, examined at the discretion and request of the neonatologist).

Evaluable subjects must have sufficient ROP and medical data. First, medical records of enrolled infants will be reviewed to identify a known ROP outcome. An “Eligibility” case report form will include a list of each inclusion criterion and check boxes in which to indicate that the infant does or does not meet the inclusion criteria. To be included in the regression modeling, an infant must have a known outcome, so if there is insufficient ROP follow-up (death, transfer, discharge without follow up, etc.), it will be unnecessary to collect growth and medical data, which will only be collected on infants with a known ROP outcome. Based upon the preliminary studies, it is estimated that approximately 85.5% of enrolled infants will have both a known ROP outcome and sufficient growth and medical data to be evaluable and included in the regression modeling.

3.3.4 Data sources
Infant demographic, medical, and ophthalmological data will be retrospectively abstracted from written and electronic inpatient and outpatient medical records at each of the study centers.
3.4 Bias and Blinding

**Selection Bias:** The prognostic model will be developed on as diverse a sample as possible, particularly with regards to risk profile. The study centers were chosen to capture a geographically and racially diverse cohort of infants at hospitals representing a wide case-mix spectrum. Sites with lower risk populations were deliberately included to accompany the larger academic centers. In addition, the inclusion criteria are designed to capture all infants who may receive ROP exams under current recommended guidelines. This group includes infants born less than 32 weeks GA or 1501 g BW, even though the guidelines used at many of the centers earlier during the study period (2006) may have had a lower GA cutoff. The group also includes infants with larger BW and GA who received ROP exams at the request of the neonatologist, which is a third criterion under current guidelines.

**Information Bias:** All data will be collected directly from the written or electronic medical record. ROP exam findings during the study period were reported using standardized diagnostic criteria (International Classification of ROP, ICROP) across all centers and were based upon direct observation of clinical findings on dilated fundus examination. Treatment decisions were made using guidelines from the Early Treatment of ROP (ETROP) study, but some variability may exist across centers. Candidate predictor data, such as weight measurements, are routinely charted. Center effects on model performance will be assessed by stratified analysis to identify systematic differences between centers in data recording or collection, or in ROP treatment decisions. If necessary, specific data elements could be re-abstracted, but systematic differences in the way ROP or medical data were recorded, if identifiable, may be non-remediable.

4 STUDY PROCEDURES

Medical records data collection. See Section 6.1 Medical records data abstraction below for detailed information.

5 STUDY ENDPOINTS AND EVALUATIONS

5.1 Primary Endpoint

The primary endpoints will be (1) sensitivity of the model for predicting severe ROP and (2) reduction in the number of infants requiring exams.

5.2 Secondary Endpoints

Secondary endpoints will include the following:

Incremental cost-effectiveness ratios comparing current screening guidelines with applications of the predictive model using different risk cut-point levels:

- Primary, short-run, analysis: cost per severe ROP case detected
- Long-run analysis: cost per quality-adjusted life years (QALYs) inferred from visual acuity using previously published equations
6 MEASUREMENTS AND EVALUATIONS

6.1 Medical records data abstraction

6.1.1 ROP Data collection

ROP examination results will be abstracted from inpatient NICU and inpatient and outpatient ophthalmology medical records for all recorded exams until (1) retinal vascular maturity, (2) disease regression, or (3) retinal vascularization into zone III without prior disease in zones I or II. These exams typically begin at approximately 31 weeks PMA or 4 weeks chronological age, whichever occurs later; continue to between 40 and 44 weeks PMA; occur every 1 or 2 weeks, depending upon disease severity; and may occur more frequently or continue past 44 weeks PMA.

The following information will be collected for each eye at each exam: stage (stages 1-5, as well as regressing, regressed, or mature vasculature), zone, presence of pre-plus or plus disease, type and date of any treatment, including laser, cryotherapy, retinal detachment surgery (e.g., scleral buckle or vitrectomy), and intravitreal injection of an anti-VEGF agent, such as bevacizumab (Avastin).

ROP data will be collected first on all enrolled infants to determine if the criteria for “known ROP outcome” as defined in section 3.3.1 above.

6.1.2 Candidate predictors data collection

Candidate predictors data will be abstracted from NICU medical records. These data will include the following:

- Baseline measures: birth weight, estimated gestational age at birth, birth length, birth head circumference, gender, maternal age, gravida, para, type of birth (vaginal, cesarean section), multiple birth, race, ethnicity, birth location (inborn or outborn), maternal diabetes, APGAR scores, perinatal steroids, chorioamnionitis, resuscitation at birth
- Daily measures: weight in grams
- Weekly measurements until 40 weeks PMA or hospital discharge, whichever occurs first: length, head circumference, medications (caffeine, steroids, Epogen, diuretics, vitamins), nutritional status, respiratory status
- Other events: intraventricular hemorrhage, NEC, hydrocephalus, sepsis, cerebrospinal fluid infection, surgical procedures, blood product transfusions

Abbreviated baseline data (e.g., BW, GA, gender, race, ethnicity, multiple birth, maternal age, birth location) will be collected on all enrolled infants and will be used to characterize and compare the populations of enrolled infants who are and who are not evaluable. The remaining baseline, daily, and weekly medical data described above will be collected only on those infants who have a known ROP outcome.

6.1.3 Costs data collection

NICU resource utilization data from the NICU medical record will be collected for a subset (<5%) of evaluable subjects on the days prior to, of, and following ROP exams and laser,
using a validated checklist. The checklist includes the following items (recorded as yes/no):

- Type of respiratory support (Oxygen only, Continuous positive airway pressure, intermittent mandatory ventilation, high frequency oscillatory ventilation)
- Intravenous access and nutrition (intravenous (IV) maintenance fluid (PN), IV maintenance fluid (non PN), saline bolus, albumin bolus)
- Radiographic imaging (chest radiograph, abdomen radiograph)
- Blood bank (red blood cell transfusion, platelet transfusion)
- Laboratory (blood gas, complete blood count, blood culture)
- Procedures (central intravenous line placement, peripheral intravenous line placement, endotracheal intubation, lumbar puncture, chest tube placement, surgery (with type))
- Pharmacy (narcotic (fentanyl, morphine sulfate, other), muscle relaxant (pancuronium, vecuronium, Other), IV antibiotic (vancomycin, gentamicin, ampicillin, Other))

6.1.4 Protected Health Information (PHI)

PHI to be collected at each local site includes patient name, mother’s name, medical record number (MRN), date of birth, and dates associated with medical records data, for example, dates of weight measurements, ROP exams, ROP laser treatment, etc. (see list above). In order to best protect PHI, names and MRN will remain only at the study centers and kept separately from other study data on a study enrollment form. These forms will be kept in locked filed cabinets in the offices of the local investigators. It is critical to collect and maintain this PHI in order to facilitate data quality checks and return to the medical record as necessary to investigate suspected data collection errors.

As defined by the HIPAA, the data submitted from each center to the central study-wide database will constitute a limited data set, as the only PHI it will contain will be dates (date of birth and dates associated with clinical events). These dates are important to retain in the database for multiple reasons, the most important being the critical role that accurate postnatal age calculations play in the development of the predictive models. In addition, it will facilitate consideration of site-specific changes in neonatal care practices, ROP screening guidelines, or ROP treatment practices during the study period. Each center will sign a data use agreement with the Children’s Hospital of Philadelphia, as required by HIPAA.

Additional information on data management security is available in section 9.1.

7 STATISTICAL CONSIDERATIONS

7.1 Primary and Secondary Endpoints

The primary objective of this study is to develop a prognostic model using postnatal weight gain to identify premature infants who are likely to develop severe ROP. The primary endpoints will be (1) the sensitivity of the predictive model for predicting severe ROP and (2) the reduction in the number of infants requiring exams.
The secondary objective is to evaluate the relative cost-effectiveness of the prognostic model versus conventional ROP screening guidelines. The secondary endpoints will be incremental cost-effectiveness ratios comparing current screening guidelines with applications of the predictive model using different risk cut-point levels in (a) primary, short-run, analysis: cost per severe ROP case detected, and (b) Long-run analysis: cost per quality-adjusted life years (QALYs) inferred from visual acuity using previously published equations.

7.2 Clinical endpoint

The clinical outcome of severe ROP for a given subject is defined as the presence in either eye of (1) any stage 3 disease (2) any stage disease in zone 1, (3) plus disease, or (4) treated ROP. With regards to the data quality of the ROP data, it has been confirmed with each clinical center that all exams during the retrospective period were completed by fellowship-trained pediatric ophthalmologists or retinal specialists, with ROP expertise, and universally using ICROP terms.

Candidate predictors include demographic and clinical variables at birth, postnatal growth measurements, reported risk factors for ROP, factors affecting IGF1 levels and growth, and comorbidities of prematurity. Based upon our preliminary studies, we anticipate many predictors will fall out of the model if their influence on ROP is mediated via IGF-1. With regards to the data quality of the candidate predictors, such data are routinely and consistently charted across all centers and do not need to be inferred. In particular, the variables in the preliminary models (PINT ROP and CHOP ROP), BW, GA, and weight measurements, are consistently documented. Weights are measured at least every 2 days, often daily, with the standard equipment in use at each site, reflecting the common range of clinical practice. Based upon the initial studies, while daily weights improved performance slightly, even weekly weights can be used in the model.

Potential outliers are very important and will be checked with the original data source, but if not due to error, sensitivity analysis will be performed to assess the robustness of the prediction model with regard to outliers.

Association of each candidate predictor with severe ROP will be examined with univariate logistic regression. Data collected when or after severe ROP occurs will be excluded. Categorical variables may be collapsed into several levels, so the prediction model is stable. For continuous predictors, a fractional polynomial procedure will be used to detect and model non-linearity. Predictors associated with p<0.20 (to avoid missing important predictors) will be included in the multivariate regression. The model will be further reduced through backward selection, which is found to be preferable to forward selection. The final model will include predictors with p≤0.05 and the known predictors BW and GA. Other selection criteria (e.g., Akaike information criteria) and stepwise selection processes will be explored to assess the robustness of the model to different variable selection procedures.
Consideration will be given to multiple ways of representing weight gain as a predictor, such as absolute and percentage changes from birth or previous week (weight gain rate), accumulated departures from “normal” weight gain, and standardized (Z score) weights. The best performing method in the CHOP ROP study was to calculate the difference between the average of the prior week’s daily weights and the average of penultimate week’s daily weights. Since at least weekly weights or weekly averages of daily weights will be measured for each infant, repeated measure logistic regression will be used to model the association of weight gain with severe ROP. PMA when postnatal weight was measured will also be included in the prediction model, and the correlation from repeated measures of weight gain will be accounted for by using the robust sandwich estimate.

7.3.2 Model performance assessment

Model performance will be assessed by evaluating calibration, discrimination, percentage reduction in babies requiring exams, and ROP alarm timing. Calibration will be evaluated using a calibration plot and the Hosmer-Lemeshow test. Discrimination will be assessed with sensitivity and specificity for detecting severe ROP using an alarm cut point value of predicted probability generated by the model. The cut point will be set to maximize sensitivity at the cost of lower specificity in order not to miss a case of ROP. However, multiple cut points will be assessed during the cost-effectiveness analysis (below). Using the model, a probability of severe ROP will be calculated for each weekly postnatal weight gain rate, resulting in multiple predicted probabilities per infant. If any predicted probability is equal to or greater than the cut point, an alarm indicating a need for clinical exams is sounded. If all predicted probabilities are less than the cut point value, the infant has no alarm. Sensitivity (Se) will be calculated as the proportion infants with alarms among infants with severe ROP. Specificity (Sp) will be calculated as the proportion of infants without alarm among infants without severe ROP. The 95% CI for Se and Sp will be calculated with traditional normal approximation methods or the Clopper-Pearson exact method if the value is close to 100%. Multiple positive (PPV) and negative (NPV) predictive values will be calculated based on the determined Se and Sp and a range of severe ROP prevalence rates observed at the clinical centers. Since the ultimate goal of the prediction model is to more accurately identify infants for eye exams, model performance will also be assessed by calculating the percentage reduction in the number of infants requiring eye exams for varying alarm levels. Performance will further be assessed using the time interval (in days) between severe ROP diagnosis and first model alarm. An effort will made to identify causes of non-physiologic weight gain that might exclude an infant from application of the model (such infants would cause a false negative signal but should receive exams). Finally, throughout the process, the parsimonious and simpler model will be preferred, as this will maximize the likelihood of physician acceptance and successful application to clinical practice.

7.3.3 Alternative strategies for model development

During model development, the cut-off alarm value, indicating a need for examinations can be lowered until all infants who develop severe ROP have alarms, but fewer infants may be spared exams. Alternative strategies include the following: (1) raise cut-off levels and assess the clinical and cost-effectiveness impacts of trade-offs between reduced exams and missed cases of severe ROP (for example, set the level to capture all lasers but not all cases of severe ROP); (2) identify infants with intermediate risk who can have a reduced frequency of exams; (3) identify and add to the model medical conditions that are causing false negative signals; (4) examine predictive values within certain subgroups of infants (e.g., define a cut-off level of BW or GA beneath which all infants receive exams, and limit the risk assessment to larger, lower-risk infants); and (5) develop a multi-tiered ROP
screening approach using additional modalities, such as a fundus imaging (e.g., predictive model determines a need for photos, and photos determine a need for exam by an ophthalmologist).

### 7.3.4 Cost-effectiveness analysis

Resource utilization and costs relevant to ROP exams will be applied in a stochastic decision tree simulation model in order to determine the economic implications of adopting the G-ROP prognostic models using various risk cut point levels.

Resource utilization will be measured the days prior to, of, and following eye exams and lasers using a checklist of neonatal interventions we previously validated in an economic study alongside a multi-center trial.67, 68 These data will be collected on a 500 event sub-sample, providing a reasonably precise estimate of the utilization proportion for each item. From our previous study101, we assume each item on the utilization checklist is used in 5 to 50% of infants, so a sample size of 500 provides a half-width of the binomial CI of 1.9% to 4.4%, respectively. Physician time may be inaccurately reflected in reimbursement and nursing time is difficult to separate from NICU per diem costs. Thus, clinician time will be measured in a time and motion study by trained bedside observers for a convenience sample of 50 events (exam or laser) at each of 3 centers, under separate IRB application.76

Unit costs for items of resource utilization will be obtained from previously collected patient-level hospital charges and converted to costs using Medicare cost to charge ratios, as previously reported by our group.77 For clinician time, an hourly wage rate derived from average national salary data will be applied. Longer-term resource utilization and costs for children with blindness will be obtained from published estimates.63, 64 To maintain a societal perspective, it is necessary to consider productivity losses and direct non-medical costs. In the short run, these are unlikely to differ between screening strategies. Long-term estimates for the cost of blindness will include costs such as time from work, out-of-pocket expenditures, and extra educational costs.63, 64 Total costs per patient will be the summed products of quantities of resources used multiplied by the unit prices for those resources. All costs will be discounted at the rate of 3% per annum and this value varied in sensitivity analyses. Costs will be in 2014 US dollars. Earlier costs will be converted using the medical care component of the consumer price index.78

Prevalence of RW-ROP and testing characteristics such as sensitivity and specificity will be determined using the retrospective study data. For long-term analysis, treatment efficacy will be based on ET-ROP results.79 Multiple risk cut points will be applied to the G-ROP predictive model to produce multiple ROP screening strategies with varying sensitivities and specificities for severe ROP. These strategies will be compared with each other and with current ROP guidelines. Presenting multiple strategies in this manner will enable readers to individually determine their acceptable trade-offs.

Model structure and analysis: An incremental cost-effectiveness (CE) analysis will be undertaken, in which the difference in costs between alternative screening strategies is divided by the difference in effects. Incremental CE ratios for each strategy relative to the next most desirable option will be estimated. Using a decision-tree structure, a simulation of a one-year US birth cohort will be performed to determine the policy implications of alternative screening strategies at a national level. Test characteristics and patient-level resource utilization inputs will be derived from the retrospective G-ROP dataset and supplemented with secondary data for the long-run analysis, as described above. Uncertainty in non-distributional inputs such as choice of price weights, background prevalence of ROP and discount rate will be assessed through sensitivity analysis, by varying the factors through their plausible ranges and reporting changes in the CE.
estimates. Uncertainty in stochastic variables such as costs will be assessed through second-order simulation, in which the input distributions will be sampled repeatedly, each time followed by a cohort simulation, to generate a series of 1000 CE ratios that collectively reflect the uncertainty in the inputs. The frequency with which these ratios overlap with CE thresholds of $0 to $200,000 per case of RW-ROP detected or QALY attained will be reported using cost-effectiveness acceptability curves. 80

Analytic horizon and study perspective: For optimal use of data with the fewest assumptions, the primary model will determine medical cost implications through hospital discharge. Comparison to other health-related and non-health related social programs will be facilitated by performing a secondary analysis over a lifetime time horizon, from a comprehensive societal viewpoint, in which all relevant medical and non-medical costs will be included, regardless of the parties to whom they accrue. Relevance to secondary audiences will be enhanced by presenting results in a disaggregated cost-consequence format, allowing those with different perspectives to synthesize the information according to their individual needs.

7.4 Sample Size and Power

(1) Assumptions: Sample size is driven by the precision of the point estimate of sensitivity for predicting severe ROP, as represented by the half-width of the 95% CI around that point estimate. This half-width CI, which determines the lower boundary of the CI, must be very small, i.e., less than 1%, as current screening guidelines will not be changed without assurance that the model continues to accurately identify infants who may require treatment to prevent blindness (see section 1.3 above). Further, the CI must remain very small across the range of sensitivities that may be observed in the study. With regards to this range, as we develop the model, its alarm threshold will be set low enough to attain very high sensitivity for severe ROP, preferably 100%, but we must allow for slightly lower discrimination (98% or 99%) when calculating sample size. Based upon our preliminary studies, we assume that 85.5% of enrolled infants will be evaluable (90% will have a known ROP outcome, of which 95% will have sufficient growth and medical data) and a severe ROP rate of 9.5%.

(2) Calculation: As demonstrated in the table below, a sample size of 9,850 enrolled infants will provide a half width of the 95% CI of less than 1% for sensitivities ranging from 100% down to 98%.

<table>
<thead>
<tr>
<th>Sensitivity of model</th>
<th>Sample size: ENROLLED</th>
<th>EVALUABLE</th>
<th># of cases severe ROP</th>
<th>Lower bound of sensitivity for 95% CI**</th>
<th>Upper bound of 95% CI</th>
<th>Half width of 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>9,850</td>
<td>8,422</td>
<td>800</td>
<td>99.5%</td>
<td>100%</td>
<td>0.24%</td>
</tr>
<tr>
<td>99%</td>
<td>9,850</td>
<td>8,422</td>
<td>800</td>
<td>98.0%</td>
<td>99.5%</td>
<td>0.73%</td>
</tr>
<tr>
<td>98%</td>
<td>9,850</td>
<td>8,422</td>
<td>800</td>
<td>96.8%</td>
<td>98.8%</td>
<td>0.99%</td>
</tr>
</tbody>
</table>

Definitions for enrolled and evaluable infants appear in section 3.3.3 above.

8 SAFETY MANAGEMENT

This study is a retrospective, non-interventional chart review and presents no greater than minimal risk to the subjects.

8.1 Unanticipated Problems
Unanticipated problems involving risks to subjects and others will be monitored throughout the study by the PI, DCC, and G-ROP Steering committee.

8.2 Adverse Event Reporting

Since the study procedures are not greater than minimal risk, serious adverse events (SAEs) are not expected. If any unanticipated problems related to the research involving risks to subjects or others happen during the course of this study (including SAEs) these will be reported to the IRB in accordance with CHOP IRB SOP 408: Unanticipated Problems Involving Risks to Subjects. AEs that are not serious but that are notable and could involve risks to subjects will be summarized in narrative or other format and submitted to the IRB at the time of continuing review.

9 STUDY ADMINISTRATION

9.1 Data Collection and Management

Overview

Data abstraction will occur locally at each study center. Study data will be entered onto written case report forms (CRF) by a data abstractor, with the option for non-PHI data and dates to be entered directly into a computer database, as described below. These primary study documents will remain at the study centers, kept in locked file cabinets in the offices of the local investigator(s). CRFs containing ROP and medical data will be coded by subject ID number, study center number, and an “alpha code,” a 4 digit alphanumeric identifier. The combination of these codes will be unique to each subject across all study centers. All PHI except for dates (i.e., subject name, MRN, and mother’s name) will be kept on a separate participant ID log, which will map this information to subject ID number, study center number, and alpha code. This PHI will remain at the study center (see PHI, 6.1.4). The participant ID log will include all enrolled patients.

Data from the CRF’s will be entered via secure web-based interface into a computer database housed at the study Data Coordinating Center (DCC) in Philadelphia. Non-PHI data and dates may also be entered directly into the computer database by local centers if desired. The data entered into the database will constitute a Limited Data Set by HIPAA terminology, as the only included PHI will be dates. The DCC will be based at the University of Pennsylvania Center for Preventive Ophthalmology and Biostatistics (CPOB) and Clinical Research Computing Unit (CRCU), which is coordinating data collection and management, and the Children's Hospital of Philadelphia, which is overseeing the distribution of study materials to and study activities of each clinical site contributing data to the study. Abstractors will have web-based training and certification. Manual quality checks will monitor for incomplete or inconsistent data, queries will be sent to clinical centers for review and correction, and many checks will be built into the web entry process as well. Data audits will occur on a small random sample subjects, for which redacted medical records will be compared to the database (see Data Quality Assurance and Monitoring below for details). All data will be housed on secure back-upped servers, as detailed below. Privacy will be protected by the use of coded data, the keys for which will remain at the Study Centers, as described above.

A data management system (DMS) will be developed and overseen by the CRCU, which will promote data security and integrity. The system will be developed using Oracle
Corporation’s suite of pharmaceutical applications, including Oracle Clinical and Oracle Clinical Remote Data Capture. Electronic audit trails of changes to database contents are incorporated into the design and will capture and record those changes automatically. In addition to the study database where actual results will be maintained, a development database will be created. The development database is a working environment that facilitates the development, testing, troubleshooting, enhancement, and training for the DMS without adversely affecting the integrity of the collected project data. Prior to deployment and use by the research staff, the database and DMS will be subjected to extensive functional testing.

Data Security

Paper CRF’s will be kept in locked file cabinets in the offices of the local investigators. Through the web-based interface, all research data for the study will be stored in an electronic Oracle database that is managed by a certified Oracle Database Administrator (DBA). The database will be hosted on secure CRCU computing servers and will be restricted to only those individuals who are authorized to work on the study. Individual user accounts with passwords will be used to restrict access to the database. Specific privilege assignments within the database will also be employed to limit the types of functions that authorized users can perform to those functions that are appropriate for their role in the study. Additional measures to prevent unauthorized external access to the database environment will be employed using network firewall technologies.

The DBA will maintain the database in an appropriate manner for the retention period required by regulation. Database administration includes user account maintenance, database security, performance monitoring, and database change management. CRCU staff will also perform operational tasks as necessary. Such tasks may include electronic data loading, database software management, and process documentation.

Data Processing

The CRCU will configure a module to allow remote data entry from the clinical research sites, using Oracle Clinical Remote Data Capture (RDC), resulting in the immediate storage of the data in the central database. The RDC module will be available to any computer with a persistent internet connection and will be run using standard web browser software. The data entry screens will look like the data collection forms as closely as possible to allow visual referencing during data entry. Data entry checks will be included in the entry screen designs where appropriate to limit the opportunity for erroneous entries due to mistyping. Such data entry checks would include value range comparisons, valid data type checks, required value checks, and skip pattern enforcement. They will aid identification of missing, illogical, or inconsistent values at the time of data entry.

Data Quality Assurance and Monitoring

A data quality module will be developed to assess data entered into the database in relation to a set of rules that describe expectations for those data items. This set of data validation rules will be defined by CRCU clinical data management personnel, working closely with study investigators, to identify data items that may have been collected incorrectly or entered into the database inaccurately. The module will run automatically to inspect all newly entered or modified data. The CRCU research staff will review the results of the data validation and take any required corrective action for invalid data. Queries will be recorded and tracked in the data quality module. Corrections identified for individual data items will be managed by the research staff. All changes made will be recorded in an electronic audit
trail and documented using change control procedures. Resolution of queries will be required before data reaches the actual validated database.

Monitoring of study progress will be accomplished, in part, through the use of standard reports. The CRCU will program a set of standard enrollment, tracking, and quality review reports. Additional reports may be developed where regular feedback is desirable.

Data audits will occur after the first few participants are enrolled and periodically during the study to detect errors in data entry. Each clinical center will submit to the study headquarters or DCC redacted medical records for a small random sample of 5-10 subjects from that center. These records will be coded by subject ID, and all PHI will be obscured except for dates of events. Only a small portion of the medical record will be submitted, not the entire medical record. This portion will include the admission note containing birth history, all eye exam notes, a nursing flow sheet containing multiple days’ growth measurements, and one week of progress notes through which medical events, such as sepsis, etc., could be identified. The redacted records will be compared to the data entered in the study database by the study headquarters research staff. Any errors will be investigated, resolved, and a plan will be implemented to prevent further errors. All redacted records will be destroyed at the earliest possible time after the errors are resolved. If the error rate is high (e.g., >5% of data points) for a specific clinical center, additional records may be reviewed after the initial errors are addressed.

**Training and certification procedures**

The CRCU will conduct comprehensive training sessions to instruct the research staff in all aspects of data collection procedures and use of the electronic web-based data management system. Training sessions consisting of initial training, certification, and retraining will be conducted over the course of sessions via webinar before the start of the study and as needed during the study.

**De-identification**

The coding enrollment lists at each study center will be maintained for at least 5 years after publication of the results of the primary and secondary study endpoints. Maintenance of the coding lists is necessary should data checks be necessary (for example, from queries arising from study publications or presentations) or if the primary study analysis leads to secondary studies requiring additional data collection under separate IRB submission. After 5 years, the coding enrollment lists will either be destroyed or additional IRB application will be made to extend the period they are kept. Their destruction will de-identify the dataset.

**9.2 Confidentiality**

All data and records generated during this study will be kept confidential in accordance with Institutional policies and HIPAA on subject privacy. The investigators and other site personnel at each study center will not use such data and records for any purpose other than conducting the study.

Safeguards are described under Data Collection and Management.
9.3 Regulatory and Ethical Considerations

9.3.1 Data and Safety Monitoring Plan

See "Data Quality Assurance and Monitoring" and "Data Security" under section 9.1 above for details. The study-wide PI and G-ROP Steering Committee will monitor data collection and analysis with the aid of reports from the Data Coordinating Center in order to identify and address unanticipated problems.

9.3.2 Risk Assessment

Risks are not greater than minimal. The risks to the subjects and families are limited to breach of privacy and confidentiality, and these risks will be minimized with the safeguards described above under Data Collection and Management.

9.3.3 Potential Benefits of Study Participation

This retrospective study provides no direct benefit to the study subjects. However, improved ROP risk assessment may have broad-reaching benefits in neonatology, ophthalmology, and public health, and affect care for infants both at low and high risk for blindness from ROP. With further validation of the predictive model(s), these benefits include the development of new ROP screening guidelines to reduce the number of children requiring exams, the frequency of exams for lower-risk infants, and the total number of exams; better allocation of professional and infrastructure resources; improved cost effectiveness of ROP screening; spurring and supporting of other ROP research by identifying infants who may benefit from preventive measures, such as exogenous IGF1 supplementation and nutritional management; and global applications, such as using the same modeling techniques to develop country- and region-specific growth-based ROP screening algorithms to better target limited resources to high-risk infants and reduce blindness from ROP.

9.3.4 Risk-Benefit Assessment

The risks are outweighed by the benefits, as the risks to the subject are no greater than minimal, and the benefits of improved ROP risk stratification to future premature infants and society are potentially great.

9.4 Recruitment Strategy (or Case Ascertainment)

See section 3.3.3 above.

9.5 Informed Consent/Assent

9.5.1 Waiver of Consent

The investigators request a waiver of consent for this chart review study. The study qualifies for a waiver based upon the following characteristics:

(1) The research involves no more than minimal risk to the subjects, as detailed in section 9.3.2.

(2) The waiver will not adversely affect the rights and welfare of the subjects. The study involves only retrospective data collection, which will be kept and analyzed in a coded fashion. Identifying information will be kept separately.
(3) The research could not practicably be carried out without the waiver. Data pertains only to past clinical encounters at the study centers during a period extending back six years. It would not be possible to request individual consent from each subject or family, as many of the subjects do not continue to receive care at the study center institutions.

(4) No additional pertinent information is expected to arise from this retrospective data collection. All of these children have been fully evaluated and treated for ROP as indicated

9.5.2 Waiver of HIPAA authorization

The study investigators request a waiver of HIPAA authorization. The study satisfies the following criteria:

(A) The use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals, based upon the following:

(1) an adequate plan exists to protect the identifiers from improper use and disclosure (see section 9.1 above).

(2) an adequate plan exists to destroy the identifiers at the earliest opportunity consistent with conduct of the research (See section 9.1 above)

(3) the PHI will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research project, or for other research for which the use or disclosure of PHI would be permitted by this subpart

(B) The research could not practicably be carried out without the waiver or alteration. Data pertains only to past clinical encounters at the clinical centers. It would be impractical, given both the large sample size and that many subjects may not continue to receive care at the clinical centers to request individual consent from each subject or family.

(C) The research could not be practicably be conducted without access to and use of the PHI. Name and MRN are required to permit data collection across multiple sources, facilitate data quality checks, and allow return to the medical record as necessary to investigate suspected data collection errors. Dates are important for multiple reasons, including recognizing changes in ROP screening guidelines and treatment practices during the study period, as well as the critical role that accurate postnatal age calculations play in the development of the predictive models.

9.6 Payment to Subjects/Families

N/A

9.7 G-ROP Study Group

9.7.1 Steering Committee

The Steering Committee will direct the study, set policy decisions, develop and modify protocols, initiate primary manuscripts, and prioritize the use of study data. The Steering
committee includes representation from the disciplines of ophthalmology, neonatology, and biostatistics. The committee is chaired by the Study PI.

9.7.2 Study Headquarters

The Children's Hospital of Philadelphia, Philadelphia PA (Gil Binenbaum MD MSCE, PI)

9.7.3 Clinical Centers

The study group currently (10/30/2012) includes 19 clinical centers, 18 in the United States and 1 in Canada. CHOP/HUP is included as a single clinical center in this total. Some academic centers cover more than one hospital NICU. An estimated 34 NICU's will therefore be represented in the dataset.

9.7.4 Data Coordinating Center

The Data Coordinating Center (DCC) is at both the Children's Hospital of Philadelphia, which will provide study activity oversight, and at the University of Pennsylvania Center for Preventive Ophthalmology and Biostatistics (CPOB) and University of Pennsylvania Center for Clinical Epidemiology and Biostatistics' Clinical Research Computing Unit (CRCU), which will actively receive, manage, and store study data. See section 9.1, Data Collection and Management, for additional details.

10 PUBLICATION

Publications and presentations will summarize the primary and secondary findings. Target journals and meetings include those in the fields of pediatrics, ophthalmology, clinical epidemiology and biostatistics, and health economics. No PHI will be published.

The preparation and submission of publications will be overseen by the study principal investigator and G-ROP Steering Committee. Manuscripts will be written by writing committees established on an ad hoc basis. Authorship will list writing committee members and specify “on behalf of the G-ROP Study Group.” Publication of results of ancillary studies using de-identified G-ROP data will be pursued based upon proposals submitted to and approved by the G-ROP Steering Committee. Such analyses will be completed by the G-ROP Data Coordinating Center. No identifiable or coded data will be released to investigators outside the G-ROP study group.

11 REFERENCES


