Introduction to the revised application

This is the 2nd revision of 1 RO1 HD047234-01, Sepsis and Critical Illness in Babies ≥ 34 Weeks Gestation, reviewed on 2/12/04 and 2/17/05. The panel felt we had been responsive to its critique but had a few remaining concerns, to which we respond below. Changes in the proposal are in Times New Roman italic 11 font.

The approach is markedly complex and in places quite difficult to follow. Ultimately, the complexity stems from the retrospective nature of data collection. In the end, how will the multiple clinical prediction rules be brought together in a way that is accessible to the practicing clinician?

We recognize that our study is complex, but we do not agree that the complexity is driven by the retrospective nature of data collection. One major problem that has affected our study design is that strict definitions of bacterial infection or critical illness do not exist. This results in our having to employ different approaches to defining infection. It would, of course, be simpler to have separate study cohorts for each study aim, but, because of budgetary considerations, we have made every effort to have our cohorts overlap so as to decrease the amount of chart review while maximizing the use of electronic data collection methods. Last but not least, our analyses are complex because they employ sophisticated statistical techniques and take great care to quantify different types of uncertainty.

The fact that our analytic strategies are complex does not mean that our findings cannot be made accessible to practicing clinicians. All of the physicians in this project are experienced educators with strong track records in implementation. For example, in order to develop the Richardson score, Escobar and Newman conducted hundreds of bivariate and multivariate analyses but were able to condense these results into a carefully validated model that is simple enough for a nurse to assign in less than 3 minutes and that can fit into a 3” x 4” card (Kaiser Permanente nurses and physicians attach Richardson score cards to the same clip that has their ID badge; copies of these cards can be found in our Appendix). The Kaiser Permanente “rule out sepsis” guideline, also included in our Appendix, also provides another example of how these investigators can translate complex analytic models into accessible tools for practitioners. Drs. Escobar and Newman currently serve on a KPMCP hyperbilirubinemia guideline committee that is incorporating results of their studies on risk factors for severe hyperbilirubinemia1,2 into a standard neonatal jaundice data collection sheet. Finally, most relevant for this project, results of their work has convinced the KPMCP leadership of the need to enhance real-time reporting of bilirubin levels to include hour-specific percentiles and treatment recommendations. The major obstacle to this was getting information on the newborn’s date and time of birth into the laboratory computer. Once this has been accomplished, it will be easy to report age-specific percentiles for CBC parameters (total WBC, ANC, etc.) in the same way. Similar approaches can be used for other electronic medical record systems in the United States.

CBC results are not available on asymptomatic controls limiting the generalizability of any predictive CBC schema. The authors attempt to justify this, including the ethics of obtaining a CBC on healthy infants, but the heterogeneous basis for a clinician’s decision to perform a CBC mandates this.

We recognize the reviewer’s concerns. Obtaining CBC results at multiple specific chronological ages from a cohort of normal newborns ≥ 34 weeks would pose daunting logistic, ethical, and financial challenges. An alternative to this approach exists, however: use of simulation techniques, and we have added this to section D.7.b (Analytic strategy for Specific Aim 2). Employing data from the limited number of studies that have examined completely normal newborns’ ANCs (absolute neutrophil counts)3-13, Dr. Draper will create several simulated datasets using random effects modeling to accommodate between-study heterogeneity in the distribution of normal newborns’ ANC values. These simulated datasets, which will have different age-dependent ANC distributions, can permit us to test the stability of our estimated likelihood ratios given different assumptions as to what constitutes a “normal” ANC at a given chronological age. For example, as is described in Section B.4., it is well known that the neonatal ANC peaks between 6 and 12 hours of age, but less is known about its rate of fall between 12 and 60 hours of age. Simulated datasets can assume different rates of fall for the ANC after 12 hours, and it will also be possible for us to test the effects of increased or decreased variability in ANC distributions at all chronological ages. Using simulation techniques, we will also test the effects of case mix among the population of newborns considered to be “normal.” For example, maternal conditions such as pre-eclampsia may be associated with a decreased ANC in an otherwise healthy infant who
would not ordinarily have his or her CBC obtained. Use of simulation techniques will permit us to estimate the possible effects on our analyses of having different proportions of babies whose mothers had pre-eclampsia.

Use of simulation techniques is common in industry and is starting to occur in medicine, particularly with respect to the design of randomized trials, where the term “in silico” is starting to be employed \(^{14-18}\), and our use of simulated ANC datasets will serve as an additional form of sensitivity analysis. Dr. Draper is familiar with multiple types of simulation strategies, which he employs routinely in his work\(^{19,20}\).

**They plan to exclude babies with “congenital anomalies, chromosomal abnormalities, and inborn errors of metabolism;” their approach will be to use a panel of experts wherein judgment will be used to decide which anomalies might impact clinical decision making. This seems both arbitrary and somewhat circular and needs further delineation.**

We will exclude only congenital anomalies considered “major” by the Vermont Oxford Network (VON), a non-profit collaborative network that includes 485 neonatal intensive care units (including the 6 Kaiser Permanente units in this proposed study) that conducts randomized trials\(^{21}\) as well as retrospective health services research studies\(^{22}\). This defined set of anomalies, listed in Version 9.0 of the VON data collection protocol\(^{23}\), has recently been studied by two of us (Zupancic and Escobar) in a revalidation of Version II of the Score for Neonatal Acute Physiology involving 58 of the 485 VON units\(^{24}\). This defined set of anomalies has a quantified impact on mortality prediction in a cohort of 6,787 newborns (5,379 of whom had birth weight > 1500 grams), as is shown by the areas under the receiver-operator characteristic curves (SNAPPE-II = SNAP-II Perinatal Extension, which includes factors for birth weight, small for gestational age status, and Apgar scores):

<table>
<thead>
<tr>
<th>Area under ROC curve</th>
<th>Hosmer-Lemeshow goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNAP-II alone</td>
<td>0.86</td>
</tr>
<tr>
<td>SNAPPE-II</td>
<td>0.89</td>
</tr>
<tr>
<td>SNAPPE-II + defined major congenital anomaly</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Thus, a baby with transposition of the great vessels (VON code # 202) would be excluded, but a baby with an isolated ventricular septal defect would not. Since we will include babies with other anomalies, we will flag the presence of these anomalies, which will permit us to consider their effect in sensitivity analyses.

**The expansion of the outcome now includes blood gas determinants. At the same time, specific Aim 3 uses blood gas determinants as a major predictor. Again, this appears to be circular.**

The reviewers are correct. In our first submission, we excluded infants who became critically ill in the first 6 hours of age (only infants who became critically ill between 6 and 12 hours of age were eligible). When we expanded our definition in response to the reviewers' critique, we neglected to consider a small set of babies in whom our definition is, in fact, circular: babies with a rapidly progressing catastrophic illness who either die in the first 6 hours without the index test or whose first blood gas is such that they immediately meet our definition of critical illness (pH < 7.10 or PaO\(_2\)/FiO\(_2\) < 1.0). These babies need to be excluded from the study, and we anticipate that most of them will be < 6 hours old. Since these infants can be identified electronically, excluding them will not require additional chart review. Our revised definition (section D.1.b) now reads:

*To be included as a case of critical illness, a baby must meet the following criteria: 1) have a 5-minute Apgar score ≥6; 2) show signs of illness (generally respiratory distress) of sufficient magnitude to have at least one ABG measured in the first 6 hours after birth; 3) have objective physiologic evidence of critical illness ≥1 hour after birth, as manifested by: a) arterial pH < 7.10; b) PaO\(_2\)/FiO\(_2\) < 1.0; and/or c) death; and 4) not meet exclusion criteria. Exclusion criteria are: 1) presence of a major anomaly, as defined in the Vermont Oxford Network manual, Release 9.0 and/or 2) meeting the definition of critical illness at the first measurement (i.e., either dying in the first 6 hours with no ABG obtained or else having the first ABG with either an arterial pH < 7.10 or an arterial PaO\(_2\)/FiO\(_2\) < 1.0).*
One rationale (e.g. in Abstract) for the development of the model is to provide guidance for "infants with maternal risk factors for bacterial infection" however the nature of these maternal risk factors is not clarified. / The nature of the maternal risk factors needs more detailed explanation.

We are clarifying this in the proposal (Section D.1.e has been modified). Maternal risk factors that need to be considered by a clinician fall into 3 broad but somewhat overlapping categories: 1) those associated with immediate physiologic effects (e.g., placental abruption or overwhelming chorioamnionitis can lead to fetal death or a baby developing shock immediately after birth); 2) those associated with increased likelihood of developing infection even if an infant is initially asymptomatic (e.g., known carriage of group B streptococcus, time between rupture of membranes and delivery exceeding 24 hours); and 3) those that might alter the information value of predictors such as the absolute neutrophil count (e.g., pre-eclampsia). Our analytic strategy is explicit with respect to integrating this information into coherent models.

It is not clear how statistically the babies with asymptomatic bacteremia (positive culture without clinical symptoms of sepsis) will be stratified or excluded.

To address this issue, we have modified Sections D.1.a (Serious Bacterial Infection) and D.7.a (Analytic Strategy for Specific Aims 1 and 3). We will 1) start with a strict definition of “asymptomatic bacteremia”; 2) characterize infants who meet this definition with respect to maternal and neonatal characteristics (for example, we anticipate that many of these babies will be the offspring of mothers who received intrapartum antibiotics); 3) conduct our first set of analyses with these infants included; 4) conduct a second set of analyses with these infants excluded; and 5) compare the results of steps (3) and (4). Our definition of asymptomatic status will require that an infant be completely asymptomatic for 48 hours before and 48 hours after the positive culture.

(sic) In Aim 3, as noted by the authors, the inability to control for treatment effects or differences in testing (pg. 72).

We recognize that this is an important limitation of any study that employs retrospective data. Although we cannot eliminate this concern, we are going to address it using simulation techniques in the course of our dynamic Bayesian predictive modeling (for treatment effects) and by incorporating simulated ANC datasets (as noted above). From a mathematical modeling perspective, the major effect of treatments is that they alter the information content of predictors. For example, suppose one has two critically ill newborns, both of whom have an arterial pH of 7.23 at 14 hours of age, and one knows that newborn A received a rapid 20 mL/kg normal saline infusion 2 hours prior to the measurement, whereas newborn B did not. Using simulation techniques and input from content experts (in our case, our two neonatologists), it is possible to test various hypotheses as to which of these babies is sicker (one could argue, for example, that newborn A is sicker because, despite a fluid bolus, s/he is only able to achieve an arterial pH that the other baby could attain without such therapy) and thus alter the prior probabilities to be incorporated in subsequent iterations of a model.
A. SPECIFIC AIMS

Confirmed neonatal bacterial infections occur with a frequency of 1-5/1000 live births\textsuperscript{25-28}, but in the U.S. around 14% of newborns are evaluated for the presence of bacterial infection\textsuperscript{29} and between 4 and 10% (160,000 to 400,000 per year) are treated with systemic antibiotics\textsuperscript{29-32}. Textbooks provide recommendations for the initiation of antibiotic therapy in infants with definite signs of sepsis, shock, or respiratory failure. They do not provide evidence-based guidance for the evaluation and management of two large groups of term or near-term newborns. These groups are 1) infants with maternal risk factors for bacterial infection who are either asymptomatic or who have presentations that are considered equivocal\textsuperscript{31-35}, and 2) infants with varying degrees of respiratory distress, which occurs in approximately 2% of term and near-term infants\textsuperscript{36-41}. Clinicians evaluating term or near-term infants for the presence of infection must make 3 decisions. The first is whether to obtain any diagnostic tests, such as a complete blood count (CBC), blood culture, arterial blood gas (ABG), chest roentgenogram, and/or cerebrospinal fluid culture. The second is whether or not to initiate treatment with systemic antibiotics. Finally, since many infected newborns are very ill, clinicians must also make a decision regarding the appropriate level of care required. Some infants will need ambulance or helicopter transfer to a neonatal intensive care unit (NICU) that can provide prolonged assisted ventilation, inhaled nitric oxide (INO) therapy, and/or extracorporeal membrane oxygenation (ECMO).

Our long term goal is to support reaching five Healthy People 2010 objectives relating to decreased infant mortality and morbidity that can result from infectious diseases (14-16, 14-20, 14-21, 16-1, and 16-14; provided in our Appendix)\textsuperscript{42}. We propose to improve the evaluation and management of newborns ≥ 34 weeks gestation at risk for bacterial infection and/or critical illness by defining a comprehensive, evidence-based approach that integrates gestational age, baseline probability for infection, clinical signs, and chronological age-specific laboratory results. To achieve this goal, we have the following Specific Aims.

1. To develop a quantitative model to estimate the probability of early onset bacterial infection based on maternal risk factors and infants’ initial clinical examinations.

We will employ a nested case-control design employing two overlapping definitions of bacterial infection. Cases will be (definition 1) newborns ≥ 34 weeks gestation with culture-proven bacterial infection with onset prior to 72 hours of age; or (definition 2) babies who meet definition 1 or who meet a strict definition of critical illness with onset prior to 72 hours of age. Controls will be randomly selected newborns who are frequency-matched by hospital of birth and gestational age.

2. To estimate likelihood ratios for early onset bacterial infection (using definition 1 or definition 2) for components of the CBC, the most commonly employed diagnostic test in this setting.

We will employ a retrospective cross-sectional study. Subjects will be infants ≥ 34 weeks gestation who had a CBC and blood culture obtained together at < 72 hours of age. Predictor variables will be CBC results (e.g., ratio of immature neutrophils to total neutrophils) and covariables will include other predictors of sepsis or of CBC results (e.g., gestational age, maternal pre-eclampsia).

3. To develop a quantitative model to estimate the probability of newborns ≥ 34 weeks gestation developing a critical illness (defined by life-threatening arterial blood gas results) based on clinical findings and the results of previously obtained laboratory tests.

We will employ a nested case-control design. Eligible subjects will be newborns ≥ 34 weeks gestation who developed respiratory distress in the first 6 hours of age. Cases will be infants who died or met a strict definition of critical illness. Controls will be a random sample of eligible newborns who did not die or go on to develop a critical illness, frequency-matched to cases by gestational age and hospital of birth.

By taking a comprehensive approach to the evaluation and management of suspected bacterial infection, we are addressing an important problem that has been sidestepped by much of the “rule out sepsis” literature – what to do about babies evaluated for infection who have respiratory distress and a significant risk for clinical deterioration. Our proposed project builds on two population-based studies on the evaluation and treatment for neonatal infection, one of which (“Watchful Waiting” vs. “Antibiotics A.S.A.P.”, 6 MCJ-060803-01-1) was funded by the Federal Government’s Maternal and Child Health Bureau’s Research Program\textsuperscript{29, 43}, a population-based study on the epidemiology of neonatal assisted ventilation\textsuperscript{44}, studies on the effects of...
epidural anesthesia and maternal fever, evaluation of the triage process in neonatal intensive care, nested case-control studies, and work on designs and optimal analysis and reporting of studies of diagnostic tests. It continues a longstanding collaboration between the Kaiser Permanente Medical Care Program (KPMCP) Division of Research, the Department of Epidemiology and Biostatistics of the University of California, San Francisco, and Harvard University's Joint Program in Neonatology.

B. BACKGROUND AND SIGNIFICANCE

B.1. Epidemiology of Early Neonatal Bacterial Infection

The incidences of neonatal sepsis and meningitis – 1-5/1000 live births for sepsis and 0.4-0.5/1000 live births for meningitis – have remained constant over the past few decades. The most common causative organisms are group B streptococcus (Streptococcus agalactiae, GBS) and Eschericia coli. Largely as a result of advances in neonatal intensive care, mortality from sepsis and meningitis has fallen and is now in the 10 to 30% range. It has also been recognized that intrapartum antibiotic treatment can significantly decrease the prevalence of neonatal GBS infection. Finally, given the high prevalence (20-30%) of GBS carriage among pregnant women, emphasis is now given to strategies that optimize use of intrapartum antibiotics. Protocols based on universal antepartum screening for GBS are superior to those based on the presence of specific epidemiologic risk factors (e.g., prematurity) or clinical signs (e.g., maternal fever) so the Centers for Disease Control recently issued a new guideline recommending universal screening. Nonetheless, there is growing concern that the gains made against GBS could be offset by increases in infections due to other organisms, particularly E. coli.

“Rule out sepsis” remains the most common neonatal diagnosis and is the most frequent indication for admission to the neonatal intensive care unit (NICU) in the U.S., where approximately 10-12% of all newborns spend at least some time in an intensive care setting. Not surprisingly, widespread practice variation occurs. Richardson and Escobar documented a threefold variation in rates of admission of term infants to the NICU, much of it driven by variation in rates of sepsis evaluation, while Zupancic et al. found that sepsis evaluations were a major component of triage admissions to the NICU and that their frequency varied two-fold by hospital and even varied by the time of day.

Most of the epidemiology literature is based on culture-confirmed infections, but in recent years many investigators have become concerned that existing definitions of bacterial infections may be inadequate, particularly with respect to making immediate decisions regarding eligibility of infants for randomized controlled trials for novel therapies that have been tried in adults (e.g., activated protein C). An important factor has been that adult intensive care physicians began to re-define sepsis on the basis of the Systemic Inflammatory Response Syndrome (SIRS) and, more recently, on the basis of the PIRO model, which includes the SIRS concept. The PIRO concept was developed partly in response to the important critique that the SIRS criteria were too nonspecific to be helpful in either the clinical or research setting. The new adult definitions have been endorsed by multiple professional societies, such as the Society for Critical Care Medicine and the European Society of Intensive Care Medicine, but it is clear that these definitions remain working concepts only. In response to the changes in adult definitions of sepsis, pediatric researchers began to question existing definitions of serious neonatal bacterial infection with negative cultures, and the major recommendations of the September 2004 conference were that 1) more research needs to be conducted so that neonatal and pediatric definitions of bacterial infections with negative cultures are based on empirical evidence; 2) more empiric research must be conducted to define what constitutes neonatal SIRS; and 3) that provisional definitions for two categories of infection –“definite” and “probable”– should be employed. The “definite” category refers to culture-confirmed infections with definite clinical signs, while the “probable” category refers to culture-negative infections with definite clinical signs. The conference did not reach consensus on categorization of asymptomatic bacteremia in newborns (i.e., when the blood culture is positive but the baby is asymptomatic).
B.2. Epidemiology of Respiratory Distress and Critical Illness in Term and Near-term Infants

Respiratory distress, respiratory insufficiency, circulatory insufficiency, pulmonary hypertension, circulatory failure, and respiratory failure can precede or accompany sepsis and meningitis. However, the epidemiology of these conditions is not clearly understood. Some data are available on respiratory distress. These include: a) population-based studies conducted in Sweden in the 1980s and Italy in the 1990s, b) studies attempting to characterize the pathophysiology of “retained fetal lung fluid” and “transient tachypnea of the newborn” and c) some studies by the current investigators. In addition, an Australian team recently reported on the epidemiology of mechanical ventilation in term infants and a recent U.S. study reported on the outcomes of assisted ventilation in babies ≥ 34 weeks gestation. These studies concur in two respects. The first is on the overall frequency. The Swedish and Italian studies found that, among term and near-term infants, the frequency of some sort of respiratory distress is in the 2% range. This range is consistent with observations made by Escobar et al. (who found that 3% of live births in the KPMCP’s Northern California Region had a discharge diagnosis of “transient tachypnea of the newborn”) and the triage rate observed by Zupancic et al., who found that at least 4% of live births in two teaching hospitals had transient respiratory distress. The second is on outcome. All of these studies have found that, while mortality rates are low, diagnostic and resource burdens are high because many infants may have respiratory signs for periods of up to 6-12 hours and many infants do go on to develop more complicated problems such as pneumonia, pneumothorax, and pulmonary hypertension. Available studies, however, do not provide information on the progression of respiratory distress or on factors that could be used to estimate a given newborn’s risk for further deterioration. The two recent KPMCP population-based studies we conducted found that between 1 to 1.5% of term and near-term infants experience assisted ventilation, and that considerable overlap exists between “rule out sepsis” and respiratory difficulties. We have also found that the number of babies who experience assisted ventilation without a positive culture is around ten times as great as the number who have culture-confirmed sepsis or meningitis. The Australian study found a lower rate of assisted ventilation (0.25% of all live births) but this may be due to its excluding anomalies, babies < 37 weeks gestation, multiple births, and infants who were not intubated. Some data are available on variation in management and outcomes from pulmonary hypertension requiring extracorporeal membrane oxygenation, but there are no similar studies focusing on other aspects of neonatal critical illness in term or near-term infants. Little is known about what is associated with optimal outcomes from assisted ventilation in term and near-term infants. Phibbs et al. have shown that mortality among infants < 1500 grams birth weight, which is tightly linked to assisted ventilation, is less in NICUs with higher volumes but comparable data for heavier or more mature infants are unavailable. Not all birth hospitals have the capability to ventilate infants, let alone provide therapies such as INO or ECMO. Clearly, more information is needed on what predicts the need for intensive care for newborns with severe respiratory distress, respiratory failure, and other forms of critical illness.

B.3. Previous Studies of Maternal and Infant Risk Factors

The pathogenesis of neonatal bacterial infection, which primarily involves dissemination of infection acquired before birth, is well described in neonatology recent reviews. Unfortunately, the existing literature is less helpful with respect to the first decision clinicians must make – whether or not to evaluate an infant. Ideally, such a decision should be based on the prior probability of infection and on the presence of maternal risk factors and clinical signs. As noted above, in term and near-term infants, the prior probability of infection is around 2/1000 live births. However, few if any studies have integrated known risk factors (e.g., maternal fever, prolonged time of ruptured membranes) so that one could generate likelihood ratios (LRs) from a multivariate model. We have reviewed much of this literature and, when possible, estimated odds ratios (ORs) from the information provided by specific studies, some of which report ORs and some of which do not. There is evidence that for some predictors (e.g. prematurity and length of rupture of membranes) ORs for GBS may differ significantly from those for other (chiefly gram-negative) organisms, but in many cases only ORs for GBS are available. Having ORs for GBS only is of limited clinical utility because it is not possible to consider GBS in isolation – decisions such as whether or not to do additional tests or start antibiotics have to be made based upon the total risk of bacterial infection, not just GBS. Maternal chorioamnionitis is known to be a major risk factor, with an approximate OR of 2.4. Other known risk factors include the following: a) infant sex, approximate OR of 1.3 for male gender; b) gestational age, approximate OR (for GBS) of 6.0 for 34-36.9 weeks compared with ≥ 37 weeks; c) birth weight, approximate OR (for GBS) of 8.2 for 1500-2000 grams, and 4.0 for 2001-2500 grams, compared with > 2500 grams; d) highest antepartum maternal temperature,

PNIC 398/2590 (Rev. 05/01) Page 62 Continuation Format Page
approximate OR = 10 for ≥ 38°C\textsuperscript{101}; e) length of time maternal membranes have been ruptured, approximate OR for ≥ 18 hours, varies widely by study from approximately 1.5 for both GBS and non-GBS cases in one study\textsuperscript{28} to 26 for GBS cases only in another\textsuperscript{101}. Maternal treatment with intrapartum antibiotics is protective, reduces the risk of GBS infection by about 60-80\%\textsuperscript{60} and is also associated with increased numbers of infants being asymptomatic and fewer infants requiring assisted ventilation\textsuperscript{29}. While it is clear that the presence of any clinical signs is associated with infection\textsuperscript{29}, the existing literature does not permit calculation of LRs for specific signs (e.g., temperature instability, hypothermia, grunting, flaring, retractions, tachycardia, lethargy), so that these might be combined with each other and with maternal data to generate an estimated probability of infection. The same situation pertains to prediction of critical illness. Neonatology textbooks recommend various treatment modalities (e.g., placement of umbilical catheters or initiation of assisted ventilation) if clinicians suspect respiratory or circulatory failure\textsuperscript{103-109}. However, actual prognostic or outcomes data are not provided because they are unavailable. Consequently, the use of clinical signs in diagnosis, establishing prognosis, and management of infants remains very much an art, not a science.

**B.4. Use of the Complete Blood Count (CBC)**

The use of the CBC is universal in U.S. hospitals, and it is routinely recommended as a screening test for newborns suspected of having bacterial infection\textsuperscript{67}. While agreement exists that the test should be performed, it is unclear how one should interpret it, and some authors question how physicians actually use the results obtained\textsuperscript{10,111}. The major problem in the literature has been its focus on what constitutes the "right" cutoff to label a given CBC result as "normal" or "abnormal." Little thought has been given to the underlying assumptions involved in making such judgments or on the consequences of making such judgments. The test has not been employed in a sophisticated fashion that takes prior probability into account. Experimental animal studies as well as granulocyte transfusion studies involving human newborns, well summarized by Al-Mulla and Christensen\textsuperscript{112} clearly show that circulating neutrophils are the key cells of interest, since they are specifically equipped to combat bacterial infection. These studies also support the notion that lower absolute neutrophil counts (ANCs) are associated with bacterial infection, which includes both a predictive component (i.e., infection is more likely among babies with lower numbers of circulating neutrophils) and a prognostic component (i.e., persistence of neutropenia is strongly associated with the progress of proven infections)\textsuperscript{113}. These studies also support the notion that increased numbers of immature cells in the circulation signal an attempt by the body to combat infection. A small number of studies have performed serial circulating neutrophil count measurements on newborns\textsuperscript{3-13}, and it is possible to make certain inferences about neutrophil kinetics based on these studies. The 95% reference range for circulating neutrophils in normal term and near-term newborns probably ranges between 6,000 - 20,000/mm\textsuperscript{3} at birth, rising rapidly and steadily to 10,000 - 26,000/mm\textsuperscript{3} by 6-12 hours of age. After 12 hours of age, the ANC falls slowly, but fewer inferences can be made as to the rate of fall because of a paucity of published data. In the presence of bacterial infection (as well as other stressful maternal and neonatal conditions) the ANC falls, with sustained neutropenia observed in the presence of continuing or overwhelming infection. Elevated total white blood cell counts or neutrophil counts do not appear to have a strong association with the presence of early neonatal bacterial infection\textsuperscript{29,114}. It is not known whether variations in circulating neutrophil counts observed in children of different racial groups\textsuperscript{115} are evident in the neonatal period. The neutrophil count is lower in very premature infants, but it is not known whether significant differences exist between moderately premature (34-36 weeks gestation) infants, term infants, and healthy post-term infants.

Although many studies have been published advocating different cutoffs for different components of the CBC (ANC, ratio of immature to total neutrophils, platelet count, etc.), these studies – most of which have small sample sizes – suffer from significant methodological problems, which have been summarized by Fowlie et al.\textsuperscript{116} and Da Silva et al.\textsuperscript{117}. None of the existing studies have employed LRs (in fact, in many it is not even possible to calculate them) and very few have reported areas under the receiver operator characteristic curve, so it is difficult to combine their results with other studies in the literature. The reported differences in sensitivity and specificity could be due to chance, use of different cutoffs, differences in sample collection\textsuperscript{112,118} or biologic heterogeneity. True biologic heterogeneity could be due to different organisms causing sepsis (for example, in older infants LRs for the total white blood cell count are better for pneumococci than for other organisms causing bacteremia) or be due to the timing of when tests were performed\textsuperscript{119,120}. It could also be due to differences in how cohorts were assembled. Cohorts may vary with respect to the proportion of infants with meconium aspiration and/or asphyxia, which are associated with CBC abnormalities\textsuperscript{121}, or they may vary with respect to whether infants with apparent sepsis but with negative cultures were included or excluded.
Finally, cohorts may vary with respect to whether or not infants with early deteriorations were included. Such infants do not present a diagnostic dilemma, and if a given CBC parameter performs well in a study with many of these infants, the study may not address test characteristics in a less sick but more relevant population.

Careful review of the physiologic and pathophysiologic literature on circulating neutrophils strongly supports the notion that one should employ more sophisticated approaches to the use of the CBC. The most commonly cited source for “normal ranges” of circulating neutrophils, the study by Manroe et al.\textsuperscript{114}, was based on only 108 newborns, many of them ill. In 1994, Schelonka et al. showed that if one applied Manroe’s criteria for “abnormal” total white blood cell count to the 193 entirely normal term infants they studied, 43\% would have been classified as being at risk for sepsis; applying the Manroe neutrophil criteria to these normal infants would have classified 63\% of them as being at risk for sepsis\textsuperscript{13}. Another important concern is that using the ratio of immature to total neutrophils is problematic because of problems with inter-rater agreement\textsuperscript{122} and because the number of cells examined to establish it is only 100, giving very broad confidence intervals\textsuperscript{123}. Not surprisingly, one finds studies reporting fatal cases of sepsis with “normal” CBC results\textsuperscript{124}. Also, it is clear from the literature that certain factors unrelated to infection (e.g., maternal hypertension, asphyxia, respiratory distress, mode of delivery) may influence the number of circulating neutrophils\textsuperscript{116}.

**B.5. Use of the Arterial Blood Gas (ABG)**

The use of ABGs for the diagnosis of respiratory failure or critical illness is common in medicine, but quantitative use of such data in neonatology is limited. Some ABG-specific recommendations for the management of very premature infants exist, as evidenced by the use of the arterial-alveolar oxygen gradient in some randomized trials of surfactant administration in respiratory distress syndrome\textsuperscript{125-129}. In more mature infants, the oxygenation index, calculated by multiplying the mean airway pressure times the fraction of inspired oxygen (FiO\textsubscript{2}) and dividing by the arterial oxygen concentration (PaO\textsubscript{2}) has been used as an entry criterion for treatment with extracorporeal membrane oxygenation and/or inhaled nitric oxide\textsuperscript{130-134}. However, there are no published prognostic data for term or near-term infants prior to receipt of assisted ventilation, nor are there any studies that provide prognostic information in combination with maternal risk factors (e.g., chorioamnionitis), newborn clinical findings (e.g., respiratory rate), or other laboratory tests (e.g., having an ANC that is less than the 10\textsuperscript{th} percentile for chronological age). A second use of the ABG, which we are incorporating in this project, was pioneered by the late Dr. Douglas K. Richardson, with whom several of these investigators have collaborated. Richardson et al. showed that the ABG components with a strong relationship to mortality are the arterial pH and the ratio of oxygen in arterial blood (PaO\textsubscript{2}) to that being provided to a patient (FiO\textsubscript{2}). An arterial pH < 7.20 in the first 12-24 hours of age is associated with at least a 7\% mortality risk in newborns admitted to the NICU, while a PaO\textsubscript{2} to FiO\textsubscript{2} ratio of < 2.5 is associated with at least a 5\% mortality risk\textsuperscript{135, 136}. The PaO\textsubscript{2}:FiO\textsubscript{2} ratio is one of the components of the SNAP\textsuperscript{136} and is calculated by dividing the arterial oxygen concentration in mm Hg by the fraction of inspired oxygen and dividing by 100. A healthy term newborn with a PaO\textsubscript{2} of 100 breathing 21\% oxygen would have a PaO\textsubscript{2}:FiO\textsubscript{2} ratio of 4.76; a very ill newborn with a PaO\textsubscript{2} of 80 despite breathing 100\% oxygen would have a PaO\textsubscript{2}:FiO\textsubscript{2} ratio of 0.8.

**B.6. Other Diagnostic Tests**

We will not study tests such as C-reactive protein, erythrocyte sedimentation rate, and interleukins. These tests have enthusiastic proponents\textsuperscript{137, 138} but are not uniformly used in the evaluation of term and near-term infants suspected of bacterial infection. One reason for this is that these tests are not always available (none of the 14 hospitals in this study consider them to be part of the standard of care), and a second is that they may also take longer to perform and interpret (for example, optimal use of the C-reactive protein requires serial measurements). This does not mean that we discount these tests. It is clear, for example, that use of C-reactive protein for the diagnosis of nosocomial neonatal infections is well supported by the literature\textsuperscript{139, 140}. We also are not studying two other commonly employed diagnostic tests: chest films and lumbar punctures. Chest films are commonly obtained to diagnose pneumonia, respiratory distress syndrome, and pneumothorax. However, it is difficult to distinguish between GBS pneumonia and other forms of respiratory distress in term and near-term infants\textsuperscript{141-144}; inter-rater agreement is poor\textsuperscript{145} and lack of availability of computerized, coded results makes use of x-rays for this study impractical. The routine performance of lumbar puncture as part of a “sepsis work-up” is controversial\textsuperscript{146, 147} and is not the standard of care at the 14 hospitals included in this project. In these hospitals, lumbar punctures are done primarily in infants in whom the decision to initiate
antibiotics has already been made; examination of cerebrospinal fluid thus primarily affects decisions on the duration of antibiotic treatment, which are not the topic of this proposal.

B.7. Strategies for Predicting Neonatal Risk for Adverse Outcomes

Although some studies propose specific strategies to decide which newborn should be treated with systemic antibiotics, few studies have attempted to frame the issue within a robust methodological and phenomenological context. With respect to methodological aspects, the majority of studies have only employed bivariate comparisons. From a phenomenological perspective, attempts to rationalize the “sepsis work up” have generally ignored four aspects of clinicians’ experience: 1) the fact that decisions about treatment with antibiotics are intertwined with the management of other problems, of which respiratory distress is the most important; 2) the dimension of time, which usually means making decisions within a 1-2 hour period in the first 6-12 hours of age; 3) variations in risk due to variations in local epidemiology; and 4) the fact that most human beings—even those who are statistically sophisticated, a minority of practicing clinicians—cannot juggle multifactorial predictive models in their heads, especially at night and under pressure. Textbooks provide lists of risk factors for and clinical signs associated with sepsis/meningitis. Multiple studies describe the epidemiology of neonatal bacterial infections. However, with the exception of the study conducted by the principal investigator, studies have not employed multivariate techniques that permit one to assess the independent contributions of various predictors. The only other study we identified that employed multivariate techniques for prediction of sepsis (as opposed to estimating the impact of maternal intrapartum prophylaxis) was restricted to GBS. A large literature exists that describes tests and combinations of tests to predict for bacterial infection, but the methodological quality of most studies has been poor, and none of them address the issue of respiratory distress and critical illness. Fowlie and Schmidt recently reviewed 670 articles published between 1966 and 1995 reporting the assessment of a diagnostic “test” predicting the presence or absence of bacterial infection in infants up to 90 days of age and found that, of the 194 studies that met their inclusion criteria, less than one third presented data in such a way that LRs could be calculated. Only one of the neonatal studies we found published after this review actually reported LRs.

In summary, many key risk factors for bacterial infection (e.g., history of GBS carriage, maternal fever, maternal chorioamnionitis, low birth weight, prematurity) are known and have been well-described in the literature. Although not as well quantified as those for bacterial infection, risk factors for respiratory difficulties (e.g., placental abruption, umbilical cord prolapse, prematurity, pre-eclampsia) are also known. It is also known that certain factors, such as whether or not a mother experienced epidural anesthesia and developed fever as a result, can confuse the issue, while maternal treatment significantly decreases the risk of both sepsis/meningitis and is associated with a decreased risk for a critical illness. It is also known that a normal physical examination is strongly associated with a lower risk of subsequently having a positive culture. What are needed are studies that a) pay careful attention to sampling, b) carefully describe their populations and risk factors, c) employ multivariate techniques, and d) explicitly provide LRs so that outcome probabilities can be calculated in different situations. Such studies are beginning to appear in the literature on evaluating older febrile infants but are still needed for newborns.

C. PRELIMINARY STUDIES

C.1. The KPMCP Division of Research Perinatal Research Unit (DOR PRU)

Prior to 1991, perinatal research in the KPMCP was very limited in scope. In 1991, Dr. Escobar was appointed to the Division of Research (DOR) with the express assignment of developing an integrated perinatal research program, the DOR PRU. The DOR PRU’s activities have been endorsed by all relevant governing and advisory bodies in the KPMCP, including the Chiefs of Obstetrics, Chiefs of Pediatrics, Nursery Directors, the Perinatologists’ Peer Group, the Perinatal Council, and the Perinatal Substance Abuse Task Force. The DOR PRU has access to all KPMCP perinatal information sources, including privileged information sources. It is the principal entity conducting perinatal outcomes research in the KPMCP’s Northern California Region and tracks the program’s 33,000 yearly deliveries at 12 birth facilities. The DOR PRU maintains a research database, the Kaiser Permanente Neonatal Minimum Data Set, which tracks all NICU admissions and has been used in multiple studies, which are described below. In addition to projects involving the NICU, the DOR PRU has conducted randomized studies involving discharge protocols for healthy term infants.
C.2. Research on Early Onset Neonatal Bacterial Infection


This is a population-based study based on data from 2 months of deliveries in the KPMCP (5709 live births). It defined a decision rule permitting clinicians to decide which infants evaluated for sepsis could be eligible for a shortened course of systemic antibiotics.


This Federally funded project is one of the largest population-based studies ever conducted on “rule out sepsis.” It involved 18,299 newborns ≥ 2000 grams birth weight, of whom 2785 were evaluated for sepsis during their birth hospitalization; of these latter infants, all but 10 were followed to one week after discharge. Significantly, this study documented that a) both over- and under-treatment with systemic antibiotics of mothers and newborns occurs, b) that the number of newborns with critical illness but with negative cultures is one order of magnitude greater than the number of newborns with culture-proven bacterial infection, c) that > 90% of infants who became critically ill did so within 12 hours of age, d) that initial asymptomatic status is an extremely strong predictor for ultimately having a negative blood culture, and e) that the use of the CBC with dichotomous cutoffs has limited value.


In this paper, Dr. Escobar described a) the rationale for the two previously described population-based studies on “rule out sepsis,” b) the need to consider clinicians’ experience when defining evaluation and management strategies, c) how data from published studies and clinical decision rules are integrated into clinical guidelines and recommendations in a large integrated health care system, and d) desirable methodologic characteristics of future studies.

C.3. Published Research on Early Neonatal Critical Illness


This study validated the SNAP-I outside its development setting and showed the feasibility of employing severity adjustment outside academic medical centers.


In addition to describing the rationale for development of neonatal severity of illness scores, this paper describes how these scores were used by Escobar’s team to address practice variation in the KPMCP.


This population-based study describes a) the basic epidemiology of assisted ventilation among newborns of all gestations and b) the degree to which the SNAP-I and other variables can predict length of assisted ventilation among newborns of different gestational ages.

This study describes the development and validation of SNAP-II and SNAPPE-II on a very large (25,429 NICU admissions) cohort from Canada, New England, and the KPMCP.

C.4. Pilot Study on Critical Illness Conducted by Drs. Escobar, Newman, and Zupancic


In this study, whose results are now being used in KPMCP nurseries to assist in triage decisions, we developed a new physiologic score. This score, which we have named in honor of the late Dr. Richardson, employs data collected to assign the SNAP-II in the first 12 hours of age to predict adverse respiratory outcomes in term and near-term infants. In the KPMCP, physicians and nurses have been provided with small laminated cards showing how to calculate the score (see Appendix). The study employed data from 2276 infants ≥ 34 weeks gestation who had respiratory distress and who received supplemental oxygen for ≥ 1 hour. Of these infants, 203 (9.3%) experienced the study adverse outcome (length of assisted ventilation > 3 days, use of INO or ECMO, and/or death).

Figure 1 shows how we employed recursive partitioning to define cutoffs for subsequent regression models. It shows results of a model containing a) gestational age, b) whether or not a baby had an Apgar score of < 7 at 5 minutes, and c) 5 physiologic predictors from the SNAP-II. Worst \( \frac{\text{PaO}_2}{\text{FiO}_2} \) yields 2 enriched nodes at different cutoffs (22.8% at ≤ 2.4, 33.1% at < 1.6). Combination of worst \( \frac{\text{PaO}_2}{\text{FiO}_2} \) and gestational age ≥ 37 weeks yields a low risk node where only 0.4% of the infants had the outcome of interest. In the figure, we have removed the “No” branch at a \( \frac{\text{PaO}_2}{\text{FiO}_2} \) of ≤ 1.6 for clarity. Our final regression model had a c statistic of 0.85 and a Hosmer-Lemeshow p value of 0.19. We then employed the regression (beta) coefficients from this model to define a very simple integer score with a maximum value of 8. In employing the beta coefficients for the score, which is intended for actual clinical use, we emphasized simplicity (as structured, the score can be assigned in less than 5 minutes on the basis of one arterial blood gas and one blood pressure measurement). The score performs well, with an area under the receiver operator characteristic (ROC) curve of 0.84 in the derivation dataset, 0.83 in the validation dataset, and 0.84 in the entire dataset. Zupancic then conducted a separate validation of this score using a dataset that was not explicitly designed for this study (data from 1427 Harvard inborn infants ≥ 34 weeks gestation who were part of the SNAP-II development cohort). Remarkably, the area under the ROC curve in this second validation dataset was 0.80.

While we are very pleased with the performance of this score, it has a number of deficiencies that led us to propose a slightly different design in this proposal. The first is the time frame, which is based on 12 hours (because those were the only data available to us). Such a time frame is probably appropriate when the goal is risk adjustment and the outcome of interest is death. A different time frame (6 hours) is more appropriate when the goal is to predict deterioration. Moreover, in this pilot study we were restricted to very few physiologic variables and could not employ other predictors of great interest, such as pulse oximetry or CBC results. We also could not test the value of changes in parameters over time.
C.5. Work conducted by Dr. Thomas B. Newman


Dr. Newman has been interested in studies of diagnostic tests, particularly those utilizing computer databases for 15 years. The first two studies listed above are examples of nested case-control studies similar to the what we propose here, from which he created (described in the first paper above) and then validated (in the second paper above) a risk index for total serum bilirubin levels of ≥ 25 mg/dL. This is part of a large body of research on neonatal jaundice, which has included a re-analysis of the Collaborative Perinatal Project. He has also studied other laboratory tests using large electronic database, including immunoglobulin, the reticulocyte count, and the erythrocyte sedimentation rate, and has described how studies on diagnostic tests should be performed and interpreted.

The third study listed above is of particular interest for this application. This tutorial demonstrates both graphically and mathematically how the value of diagnostic tests is enhanced when results are categorized and likelihood ratios for each category are used, rather than simply dichotomizing results and labeling them "normal" or "abnormal." The enhanced value derives from the possibility that extreme test results may be associated with likelihood ratios much smaller or larger than those obtained when test results are dichotomized. This widens the range of prior probabilities for which the test result can alter management, and hence provide value. The example used for this tutorial is the White Blood Cell count in febrile infants. These babies are somewhat older than those that will be studied for this application, but the principles are exactly applicable to newborns.

C.6. Work Conducted by Drs. Ellice Lieberman and Eric Eichenwald


This study reviewed all confirmed cases of early onset GBS disease at the Brigham and Women’s Hospital from 1/97 through 5/03. Puopolo and Eichenwald found that, during this time period, 14 of the 18 cases who were ≥ 34 weeks gestation had negative maternal GBS cultures, and that 13 of these infants showed clinical signs within 24 hours of age. The clinical characteristics of these infants were similar to those of infants with culture-confirmed sepsis described in the studies of Escobar et al. and Bromberger et al. – this supports the notion that GBS screening is associated with decreases in prevalence but not necessarily with decreases in illness severity or changes in the time of clinical presentation.


Dr. Lieberman has had a long-standing interest in sepsis evaluation among term infants. She was the first to report the large increase in maternal fever that accompanies epidural use (15% epidural, 1% no epidural) and to determine that 90% of intrapartum fever during term labor was attributable to epidural use. In that same study, she reported a 4-fold increase in neonatal sepsis evaluation and antibiotic use among infants whose mothers had received epidural for pain relief during labor. While much of this increase in evaluation was due to maternal fever, a higher rate of sepsis evaluation was also found among the infants of afebrile women. The higher rate in afebrile women occurred because criteria such as long rupture of membranes also contributed to the decision about which babies to evaluate for sepsis. Dr. Lieberman’s laboratory has also performed work examining the role of intrapartum maternal fever in identifying infants with early onset neonatal sepsis, a critical question in infants who are asymptomatic at birth. The study, which was conducted among all term infants with sepsis at Brigham and Women’s Hospital over a seven-year period, indicated that ignoring maternal intrapartum fever could result in failure to identify 36% of asymptomatic infants with sepsis. In addition, Dr. Lieberman’s group has also evaluated changes in the rate of early onset neonatal sepsis due to implementation of GBS screening and intrapartum treatment. Her team has found a significant decrease in the incidence of sepsis due to GBS with no increase in the incidence of sepsis due to non-GBS or ampicillin-resistant organisms. Dr. Eichenwald, a neonatologist with a strong interest in neonatal sepsis, its evaluation, and its relationship to the development of critical illness, has collaborated with Dr. Lieberman in several studies.

C.7. Work conducted by Dr. John Zupancic


Dr. Zupancic is a decision analyst whose research expertise focuses on computer simulation approaches to neonatal intensive care, and on resource utilization. He has had an interest in the low acuity, high volume population of infants and their impact on neonatal resource use. Along with Dr. Douglas Richardson, he published the first systematic description of the short-term evaluation and management of infants after delivery, known as neonatal triage. In a sample of 2486 infants admitted to the NICU for less than 24 hours, 34% underwent evaluation for sepsis risk and 23% had transient respiratory distress. Severity of illness was minimal, with 70% having SNAP scores of 0, indicating no derangement. Despite their minimal illness acuity, these infants accounted for 9.5% of total NICU aggregate costs. Dr. Zupancic has extended the analysis of high-volume, low-acuity care in decision analytic studies of discharge criteria for infants with resolving respiratory immaturity. Dr. Zupancic has also developed expertise in the methodology of measurement of neonatal illness severity. He was a co-developer of the Transport Risk Index of Physiologic Stability (TRIPS), a tool to measure physiologic instability over a shorter period of time than is required by SNAP, in order to facilitate risk adjustment prior to transport. More recently, he has revalidated the SNAP-II using a different population.
D. RESEARCH DESIGN AND METHODS

D.1. DEFINITIONS

D.1.a. Serious Bacterial Infection (SBI)

In this proposal, we are employing two definitions of Serious Bacterial Infection (SBI), and, when the term “SBI” is employed, it should be assumed that both definitions are being considered (e.g., for analyses, we will define likelihood ratios using both definitions and see whether significant differences exist). SBI definition 1 refers to confirmed bacterial infection with onset prior to 72 hours of age. Our definition includes a) sepsis or bacteremia with potentially pathogenic organisms; b) meningitis with a positive cerebrospinal fluid culture which was not felt to be a contaminant by treating physicians; c) any bacterial infection confirmed by positive culture from normally sterile sites (other than urine), e.g., positive pleural fluid culture for group B streptococcus in a baby suspected of having pneumonia, or an autopsy culture. We are not including newborns with bacteriuria because urine cultures are seldom indicated in sepsis evaluations in infants < 72 hours old\(^\text{167, 168}\) and because of difficulties interpreting urine cultures depending on collection methods. Our definition will not be based on CBC results or any variables being studied as predictors (e.g., chorioamnionitis). Babies meeting our alternative definition for SBI, which corresponds to the “probable” category defined in recent consensus conferences, will be those who 1) meet definition 1, above, or 2) meet the following definition of being critically ill in the first 72 hours of age: a) arterial pH < 7.10; b) \(\text{PaO}_2/\text{FiO}_2\) < 1.0; and/or c) death.

We will define asymptomatic bacteremia as follows: 1) having a positive blood culture with a pathogenic organism, 2) having a normal physical examination for 48 hours before and 48 hours after the positive blood culture. A normal physical examination is defined as: 1) absence of signs of respiratory distress (grunting, flaring, retracting, pallor, cyanosis, oxygen saturation < 95% on room air); 2) absence of apnea, seizure activity, increased tone, or decreased tone; and 3) having vital signs within the normal range (temperature 98.0 – 99.0 °F, heart rate 80 – 159 beats per minute, and respirations 40 – 70 breaths per minute). This group cannot be analyzed separately for studies of prediction from physical findings because by definition physical findings will be absent in this group.

Our study will not include bacterial infections acquired or with an onset after 72 hours of age, or after a newborn was sent home from the birth hospitalization if the baby was discharged prior to 72 hours of age. As was done by Schuchat et al.\(^\text{28}\) we will include bacteremias due to coagulase-negative staphylococci only if the same organism was isolated from two or more cultures of normally sterile sites. We will consider an infection to have had an onset prior to 72 hours if the baby had clinical signs (e.g., respiratory distress) that led clinicians to obtain a blood culture prior to 72 hours of age or – if the baby was asymptomatic – if a baby had a blood culture obtained prior to 72 hours of age because of maternal risk factors (e.g., chorioamnionitis). We will not include nosocomial infections (e.g., \textit{Staphylococcus epidermidis} central line infection).

D.1.b. Critical Illness

Many infants who are extremely ill have negative culture results. Conversely, many infants with positive blood cultures are asymptomatic, raising the concern that their positive culture was either a contaminant or the result of transient bacteremia. Confronted with a term infant with respiratory distress, some physicians might label such an infant as “critically ill” and go on to intubate and initiate mechanical ventilation immediately, while other physicians may elect to provide oxygen by hood or using nasal continuous positive airway pressure and obtain arterial blood gas measurements prior to initiating more intensive care. Some term and near-term babies who are intubated often improve dramatically after brief periods of assisted ventilation (for example, over 80% of KPMCP ventilated term infants require < 24 hours of assisted ventilation).

Consequently, we are not defining critical illness in terms of whether or not a specific procedure – e.g., initiation of assisted ventilation or treatment with vasopressors – occurred. This is because many maneuvers that might be initiated by one clinician could be considered discretionary by another. In addition, if our outcome variable were based on a clinician decision, the study would be subject to bias, because clinicians making the decisions might base them partly on the variables we wish to study as predictors. We are basing our definition of a critical illness based on experience with objective severity of illness scoring using the Score for Neonatal Acute Physiology, version II (SNAP-II)\(^\text{136}\) We are also explicitly excluding infants who become critically ill in the first hour of life because they do not present a diagnostic dilemma.
To be included as a case of critical illness, a baby must meet the following criteria: 1) have a 5-minute Apgar score ≥ 6; 2) show signs of illness (generally respiratory distress) of sufficient magnitude to have at least one ABG measured in the first 6 hours after birth; 3) have objective physiologic evidence of critical illness ≥ 1 hour after birth, as manifested by: a) arterial pH < 7.10; b) PaO$_2$/FiO$_2$ < 1.0; and/or c) death; and 4) not meet exclusion criteria. Exclusion criteria are: 1) presence of a major anomaly, as defined in the Vermont Oxford Network manual, Release 9.0$^{23}$ and/or 2) meeting the definition of critical illness at the first measurement (i.e., either dying in the first 6 hours with no ABG obtained or else having the first ABG with either an arterial pH < 7.10 or an arterial PaO$_2$/FiO$_2$ < 1.0).

Based on work with the SNAP-II, we know that an arterial pH of < 7.10 or a PaO$_2$/FiO$_2$ < 1.0 in the first 12 hours are each associated with an approximate mortality risk of 16%. Based on our clinical experience and multiple internal data reviews, we also know that these ABG results are excellent proxies for shock, respiratory failure, severe pulmonary hypertension, and organ failure. We also know from the results of Escobar's population-based study that most babies who experience respiratory distress do so early: in that study, 21% of newborns evaluated for sepsis required supplemental oxygen therapy in the first 6 hours of age and only 3% had such therapy initiated after 6 hours of age. Finally, our definition is designed so that objective electronic techniques (as opposed to chart-review) can be employed for initial patient identification.

### D.1.c. Appropriate Level of Care

Not all birth hospitals can care for a very ill infant. A predictive model that quantifies the risk for critical illness gives clinicians very useful guidance with respect to whether a baby should remain at a hospital with minimal (ability to give intravenous fluids, antibiotics, and supplemental oxygen), moderate (ability to sustain a baby on assisted ventilation for a few hours until arrival of a transport team), tertiary (ability to maintain assisted ventilation and use of vasopressors for long periods), or quaternary (ability to provide inhaled nitric oxide, extracorporeal membrane oxygenation, and/or other forms of surgery) capability.

### D.1.d. Prior and Posterior Probabilities and Likelihood Ratios

Information from history and physical examination findings and laboratory tests results changes probabilities of disease from what they were before the test results were known (prior probabilities) to values that reflect new information (posterior probabilities). The degree to which various findings and test results alter prior probabilities can be quantified with the likelihood ratio (LR), which is equal to the probability of the finding in people with the disease divided by the probability of the finding in people without the disease. The prior and posterior probabilities and LRs are related by the following equation, where the symbol | means “given”:

$$\text{prior odds} \times \text{LR} = \text{posterior odds}$$

where $P =$ Probability, odds = $P/(1-P)$, and $\text{LR} = P(\text{Result} \mid \text{disease})/P(\text{Result} \mid \text{no disease})$.

Our project is explicit about two characteristics of clinical and diagnostic information: a) the value of information varies with time and is highly contingent on the clinical situation, and b) many diagnostic tests lose information when dichotomized. For Specific Aim 1, we will start with the prior probability of developing SBI based on an infant’s gestational age. We will then determine how maternal predictors and the newborn’s clinical examination affect the probability of having SBI (i.e., we will determine LRs). For Specific Aim 2, we will start with the prior probability of SBI given maternal and neonatal risk factors and then determine LRs for CBC results. Finally, for Specific Aim 3, we will start with the prior probability of developing a critical illness given the presence of respiratory distress and then determine LRs for laboratory (ABG and CBC) results as well as components of a newborn’s clinical examination (e.g., blood pressure, pulse oximetry).

### D.1.e. Other Predictors

Demographic predictors are data elements that could be obtained immediately prior to a mother’s admission (e.g., maternal age, race, parity, insurance coverage, and so forth). Clinical predictors are data elements that could be obtained by physical examination or by in-hospital observation by clinicians. They include maternal predictors (e.g., highest temperature in the 12 hours prior to delivery) and neonatal predictors (e.g., respiratory rate; whether or not a baby had emesis). Maternal risk factors that need to be considered by a clinician, and which we will capture by chart review, fall into 3 broad but somewhat overlapping categories: 1) those associated with immediate physiologic effects (e.g., placental abruption or overwhelming chorioamnionitis can lead to...
fetal death or a baby developing shock immediately after birth); 2) those associated with increased likelihood of
developing infection even if an infant is initially asymptomatic (e.g., time between rupture of membranes and delivery
exceeding 24 hours, known carriage of group B streptococcus); and 3) those that might alter the information value of
predictors such as the absolute neutrophil count (e.g., pre-eclampsia). One must be judicious in selecting clinical
predictors. For example, for Specific Aim 1, we will not employ blood pressure a clinical predictor because it is
not routinely measured among healthy infants (controls). In contrast, for Specific Aim 3, blood pressure is
included because the study cohort only includes babies who reached a threshold level of physiologic
derangement, and it is standard clinical practice to measure blood pressure in such infants. We will emphasize
the use of objective predictors (e.g., using mean arterial blood pressure and pulse oximetry results rather than
a narrative description that a baby was “poorly perfused”). We will employ results of two tests as laboratory
predictors: the CBC and the ABG. Each of these tests provides several results and, in the case of arterial
blood gases, some test results are affected by treatment within very short time frames (e.g., oxygenation as
measured by the PaO$_2$ is highly dependent on how much oxygen is provided to the infant, the FiO$_2$). With both
of these tests, it is important to determine the predictive value of specific ratios (e.g., the ratio of immature to
total neutrophils, the PaO$_2$/FiO$_2$). For reasons noted in Section B.6, above, we have made an explicit decision
not to include other laboratory tests (e.g., C-reactive protein, procalcitonin).

D.2. STUDY HYPOTHESIS

Since our goal is to develop predictive models rather than to examine the presence or absence of
associations, we will not perform traditional statistical hypothesis testing with a goal of accepting or rejecting
hypotheses. Instead, our goal is to create tools that best utilize clinical information to predict SBI or critical
illness. An underlying assumption is that better estimates of these probabilities will improve outcomes of
newborns. We believe this is the case because the benefits of testing, treating, and transferring infants will
depend directly on the risk of SBI or critical illness. Thus, the ability to quantify these risks can help ensure that
those receiving the interventions are those in whom the benefits are most likely to exceed the risks and costs.

D.3. PREDICTOR-OUTCOME RELATIONSHIPS

Our goal is to develop models that will permit clinicians to employ demographic, clinical, and laboratory
predictors to estimate probabilities for the occurrence of SBI and critical illness.

Global relationships between predictors and outcomes are provided in Figure 2, below, and summarized in Table 3. In Figure
2, thin black arrows show how a baby’s underlying biological risk progresses over
time to one of several outcomes. Thin arrows
pointing to text in italics show where we will
perform measurements that can be used to
establish the probability of the outcomes. These measurements involve either paper chart abstraction (e.g., for maternal clinical
history items or neonatal vital signs) or
electronic capture of laboratory test results.
Possible outcomes are: no SBI, no critical
illness, SBI with critical illness, critical illness
without SBI, and critical illness with SBI.
Figure 2 also shows that babies with a critical
illness can die, recover without sequelae, or recover with sequelae (e.g., neurological damage due to hypoxia
and/or shock). Thick black arrows show that the baby’s underlying biological risk, which is primarily determined
by gestational age, is strongly affected by maternal infection and/or other complications, while clinical
progression is strongly affected by treatment.
D.4. DATA COLLECTION (see Appendix for examples of investigators’ instruments)

All team members have considerable experience in designing perinatal chart review instruments that capture time-specific data points, and we will base our instruments on those employed in previous published studies. Our Appendix shows two sample data collection instruments, one of which was employed in Dr. Escobar’s Federally funded population-based study\(^{29}\), and one which has been used at multiple locations for assigning the SNAP-II\(^{136}\). Both of these instruments have been validated by these investigators and have detailed coding protocols for use by professional medical records abstractors. Data abstraction will emphasize capturing hour-specific data so that models we develop are only based on information available at a given point in time. Based on our studies, we also know that these data can be found in the participating institutions’ charts. With respect to electronically captured data, all of the participating sites store laboratory, microbiology, demographic, and hospitalization data in clinical data repositories. The three senior investigators – Escobar, Newman, and Lieberman – have considerable experience in using electronic databases for clinical research.

The two principal study outcomes – SBI and critical illness – are based on chart-based confirmation following identification of patients through electronic scanning of clinical and laboratory databases. The same will be the case with respect to a number of maternal and neonatal predictors, such as maternal age, maternal race, newborn birth weight, and newborn gestational age. We will obtain maternal and neonatal diagnoses (ICD codes) through electronic scanning, but the occurrence and timing of key procedures (e.g., cesarean section, neonatal assisted ventilation) will be confirmed by chart review. Specific physiologic measurements (e.g., highest maternal temperature in the 12 hours prior to delivery, lowest mean arterial blood pressure in a given time interval) will be obtained by chart review, while laboratory tests (e.g., CBC’s, ABG’s) will be obtained electronically. Some laboratory data will be supplemented by manual chart review. Examples of this include ABG data (where it is essential to know whether certain therapies, such as assisted ventilation or fluid boluses, had occurred before or were occurring at the time of treatment) as well as information regarding GBS carriage, where additional information might have been obtained at time of delivery (e.g., history of GBS carriage in a previous pregnancy). Many variables will need to be derived using algorithms to combine variables obtained from chart review or electronic scanning. Examples of these include small for gestational age status and PaO\(_2\):FiO\(_2\).

D.5. STUDY DESIGN

D.5.a. Study Design for Specific Aim 1

For Specific Aim 1, To develop a quantitative model to estimate the probability of early onset bacterial infection based on maternal risk factors and infants’ initial clinical examinations, we will employ a nested case-control study design. For this study, babies will be eligible as Cases if they meet the following criteria: 1) gestational age ≥ 34 weeks; and 2) meet one of our two SBI definitions (Section D.1.a); and 3) no chromosomal abnormality, congenital anomaly, or inborn error of metabolism (as defined by the Vermont Oxford Neonatal Network) was present. Babies will be eligible as Controls if they meet criteria 1 and 3.

For SBI definition 1 (confirmed bacterial infection) we will employ the 1998-2005 birth cohorts from 1) the KPMCP’s Northern California Region’s 12 delivery centers (approximately 230,000 newborns ≥ 34 weeks gestation); 2) the Brigham and Women’s Hospital in Boston (approximately 76,000 eligible newborns); and 3) the Beth Israel Deaconess Hospital in Boston (approximately 34,000 eligible newborns). Given a birth cohort of 340,000, we anticipate having approximately 350 culture- or autopsy-confirmed cases of sepsis. Our SBI estimate of 1/1000 live births is based on the study of Escobar et al., which found a rate of 1.2/1000 live births ≥ 2000 grams in the KPMCP in 1995-96\(^{29}\) the study of Chen et al., which found a rate of 1.6/1000 among term deliveries at the BWH between 1990 and 1996\(^{169}\) and a more recent audit by Escobar, which found that the rate among all live births in the KPMCP in 2000 was 0.9/1000 live births. We will randomly select 1050 controls, who will be frequency-matched with respect to gestational age and hospital of birth. The specific approach to be employed for identification of cases and controls is described in Section D.6, Population Description and Sampling Plan. For SBI definition 2 (“probable” bacterial infection), we will employ the 2005 birth cohorts from the participating institutions, with an additional 3 randomly selected controls per case for each culture-negative critically ill baby.
We will collect the following maternal predictor variables: maternal age, gravidity, parity, race, highest antepartum temperature during the 12 hours prior to delivery, length of time membranes were ruptured, whether or not premature rupture of membranes occurred, history of urinary tract infection in the 3 months prior to delivery, clinical diagnosis of chorioamnionitis (i.e., is there a physician note to the effect that chorioamnionitis was present), objective findings of chorioamnionitis (e.g., documentation of intrauterine tenderness or foul smelling amniotic fluid), treatment with systemic intrapartum antibiotics (including number of doses and timing of treatment), use and timing of epidural anesthesia, mode of delivery, whether or not major obstetric complications (e.g., placental abruption) were present, and GBS carriage. We also will collect the following neonatal predictor variables as isolated data elements: gender, birth weight, gestational age, 1- and 5-minute Apgar scores, and presence of meconium staining. We will collect the following neonatal predictor variables for each of four 6 hour intervals in the first 24 hours of age: highest and lowest respiratory rate; highest and lowest heart rate; highest and lowest temperature; whether or not grunting, flaring, or retractions were present, and whether observers rated these as severe; whether or not apnea occurred, and what interventions were required to treat it; and whether or not vomiting or seizures occurred.

D.5.b. Strengths and Weaknesses of Study Design Selected for Specific Aim 1

With the exception of culturing and treating GBS-positive mothers to reduce the risk of infection in their newborns, which already has been studied in randomized trials most of the variables that predict the risk of sepsis and critical illness in newborns are not subject to manipulation by investigators. Randomized clinical trials are not feasible, so an observational design is required. The principal strength of a nested case-control design, compared with a cohort study, is its greater efficiency. Given the rarity of SBI, the time and expense that would be required for a cohort study are prohibitive, since it would require collecting data on about 100 (if SBI definition 2 were employed) to 1000 (if SBI definition 1 were employed) newborns not destined to suffer from SBI for each one that did.

Compared with other case-control designs, the strength of nested case control studies is their reduced susceptibility to sampling bias. Because the population from which the cases arose has already been enumerated, controls can be selected at random, thus ensuring that controls represent the population that gave rise to the cases. As is true with any retrospective study, the main weakness of the nested case-control design is that some variables we wish to study may not have been optimally measured. For example, we will record whether the obstetrician indicated a diagnosis of “chorioamnionitis” in the mother’s record, but we are not able to specify criteria for that notation as we could if we were doing the study prospectively. The main effect of this lack of control over some of the variables in the study is likely to be a slight diminution of the magnitude of observed associations between predictors and case status. On the other hand, if measurements for this study were much more precise than commonly occur in practice, the magnitude of observed associations would be greater than in the populations to which we would wish to generalize.

D.5.c. Study Design for Specific Aim 2

For Specific Aim 2, To estimate likelihood ratios for early onset bacterial infection for components of the complete blood count (CBC), we will employ a retrospective cross-sectional study design. We refer to the design as “retrospective” because the measurements of variables have already been made and “cross-sectional” because measurements of predictor (CBC) and outcome (blood culture) were made at the same point in time and sampling for the study is not related to the results of either one. Subjects will be infants ≥ 34 weeks gestation who had a CBC and blood culture obtained prior to 72 hours of age. Predictor variables will be CBC results (e.g., ANCs, ratio of immature neutrophils to total neutrophils). Covariables (possible confounders and effect modifiers) will include other clinical predictors of SBI (e.g., gestational age). They will also include other factors that can affect CBC results (e.g., maternal pre-eclampsia). The outcome of interest will be SBI, as defined in Section D.1.a. For this study, babies will be eligible if they meet the following criteria: 1) gestational age ≥ 34 weeks; 2) they had a first CBC and blood culture obtained < 1 hour apart prior to 72 hours of age. In children with multiple CBC/blood culture pairs only the first pair will be considered.

We expect that the infants with SBI (according to definition 1 above) will be a large subset of the infants enrolled in the nested case-control study described in Section D.5.a, but a few infants eligible for (definition 1) case status for Specific Aim #1 will not be eligible for this part of the study because their bacterial infection will
have been diagnosed from a sample obtained ≥ 1 hour following the first CBC or because they have negative blood cultures but have other positive cultures that cannot be linked to their initial CBC (e.g., an infant with SBI with a negative blood culture at 3 hours of age who went on to have a positive pleural fluid culture for GBS at 28 hours of age). These infants need to be excluded to avoid biases arising from clinicians using the CBC result to decide which other cultures to obtain (e.g. whether or not to do a lumbar puncture).

The primary predictors will be individual CBC components as well as combinations, z-scores, and ratios of these components. These will include: total white blood cell count; ANC; absolute band (immature neutrophil) count; ratio of immature to total neutrophils (i:T ratio); and platelet count. We will consider CBC components that are manually estimated (e.g., the i:T ratio) as well as those that are calculated electronically. Covariables of interest are neonatal or maternal conditions known or suspected to be associated with alterations in leukocyte kinetics (e.g., neonatal age, race, birth asphyxia, gestational age, pre-eclampsia). The specific approach to be employed for identification of study subjects and their laboratory test results is described in Section D.6, Population Description and Sampling Plan. We have conducted a number of studies at the participating institutions29, 43, 48, 135, 136 and they concur with respect to the fact that approximately 14% of newborns born at ≥ 34 weeks will have the test combination of interest (CBC + blood culture drawn less than 1 hour apart) during the neonatal period. Given a birth cohort of 340,000, we would thus expect to have approximately 47,600 test results. We also know from Escobar et al.’s population study that 77% of these tests are likely to be obtained before 12 hours of age, 12% between 12-23.9 hours of age, and 11% after 24 hours.

D.5.d. Strengths and Weaknesses of Study Design Selected for Specific Aim 2

An important feature of this study is that CBC results will be available only on infants for whom a CBC was ordered, rather than on the entire population of infants. We believe that if this has any effect, it is to strengthen the results. One possibility is that, once infection is taken into account, factors related to selection for a CBC are unrelated to the results of the CBC. If this were the case, CBC results available for the study would approximate those that would have been obtained on the entire population, and there would be no problem. It is more likely that some factors associated with having a CBC obtained (e.g., maternal fever) are also associated with CBC results, independent of the increased risk of sepsis. If this is the case, it is in fact desirable to estimate the LRs for different CBC results using a “non-diseased” group composed of infants in whom sepsis was considered, rather than the general population of non-diseased infants.

This principle – that discrimination of diagnostic tests is best evaluated on a spectrum of patients similar to those in whom the tests will be used – is well documented in the literature. Our previous studies of reticulocyte counts and red cell morphology164 and direct bilirubin measurements in jaundiced infants170 used only infants in whom the tests were done to define clinically relevant reference ranges. One of the key questions to ask when critically appraising studies of diagnostic tests is whether the test was evaluated on a spectrum of patients similar to those in whom it would be used in practice171, 172, i.e., in those in whom clinicians faced diagnostic uncertainty173. Studies of diagnostic tests in which the “nondiseased” group is excessively healthy (e.g., asymptomatic newborns with no risk factors) will tend to be biased towards suggesting diagnostic tests are more helpful than they will be in clinical practice.

Our analyses can be further refined by considering that infants evaluated for SBI represent a heterogenous group, with various combinations of risk factors for SBI that might differentially affect the CBC. For example, the moderately premature infants in our study (those who are 34-36 weeks gestation) are at higher risk of SBI25-28 and are therefore more likely to have a CBC obtained. But even in infants who are not septic, lower gestational age is associated with lower mean neutrophil counts – for example, we re-analyzed data from the population-based study by Escobar et al. and found that the mean first ANC was 12,200 in infants ≥ 37 weeks gestation, compared with 8000 in infants < 37 weeks (P < 0.0001). By using gestational age specific CBC z-scores, we can factor out the effect of gestational age on CBC results.

Studies in which large numbers of normal newborns have CBCs obtained purely for research purposes are rare and pose daunting ethical problems. The best (and largest) study, that of Schelonka et al., only had 193 infants, all of whom had a CBC measured at exactly 4 hours of age; most of the other studies, including that of Monroe et al., employed samples of convenience. Since our goal is not to establish what constitute normal CBC results (our goal is to help clinicians sort out what to do a posteriori) it is reasonable to base our LRs on results obtained from CBCs obtained for the same purpose.
D.5.e. Study Design for Specific Aim 3

For Specific Aim 3, To develop a quantitative model to estimate the probability of newborns developing a critical illness (defined by life-threatening arterial blood gas results) based on clinical findings and the results of laboratory tests, we will employ a nested case-control study design. Eligible infants will be infants ≥ 34 weeks gestation who 1) had a 5 minute Apgar score ≥ 6; 2) had at least 1 ABG obtained in the first 6 hours of age; 3) did not have a congenital anomaly, chromosomal anomaly, or inborn error of metabolism (as defined by the Vermont Oxford Neonatal Network); and 4) did not meet our definition of a critical illness in the first 1 hour of age. Cases will be randomly selected from among eligible infants meeting our definition of a critical illness (Section D.1.b) after 1 hour of age. Controls will be randomly selected from among eligible infants that did not go on to develop a critical illness as defined in Section D.1.b. As noted earlier, newborns who became critically ill prior to 1 hour of age will be excluded. Identification of cases and controls is described in Section D.6.

Figure 3: Population Context for Specific Aim 3

Figure 3 places cases and controls in a population context. It is based on our previous studies, which have included a review of all deaths in the KPMCP174 as well as on the pilot study described in Section C.4, which employed data from the 1997-2001 birth cohort at 6 KPMCP facilities. This pilot study found that, among 88,951 babies ≥ 34 weeks, 2,276 (2.6%) required supplemental oxygen for at least 1 hour. The figure shows that, in this gestational age group, overall mortality is low (about 1.6/1000 live births) and that most deaths occur among babies who deteriorate very early (either immediately after birth or within 6 hours after birth).

For Specific Aim 3, we will collect the same maternal and neonatal predictor variables for Specific Aims 1 and 2. With respect to the CBC, we will collect the same predictor variables and covariates described above. We will also collect information on delivery room events (e.g., did an infant receive bag and mask ventilation in the delivery room) as well as additional physiologic and treatment data. Additional physiologic data we will collect include blood pressure measurements, response of blood pressure and heart rate to fluid bolus and pressor administration, results of non-invasive measurement of oxygenation status, effects of treatment on acid-base status (e.g., response of pH and base deficit to the administration of sodium bicarbonate), nature of respiratory support, and neurological status (e.g., seizures, whether or not these responded to pharmacologic treatment). Since we wish to predict outcome after 1 hour of age based on information available in the first 6 hours of age, we will capture data using 1 hour time frames for data collection.

D.5.f. Strengths and Weaknesses of Study Design Selected for Specific Aim 3

It would be highly desirable to have randomized trial data permitting one to make rational decisions with respect to which babies ≥ 34 weeks with respiratory distress should be intubated, receive pressors, and so forth. However, the existing literature base does not yet permit one to conduct such studies ethically. Evidence-based guidelines or prognostic models that would permit a physician to quantify the risk of deterioration in a term or near-term infant with early respiratory distress (and then decide whether or not to implement specific therapies) do not exist. Consequently, we selected the nested case control design primarily for its efficiency and lack of susceptibility to sampling bias. Again, the main limitation is in availability of precise measurements of relevant predictor variables. For example, the FiO₂ and mean airway pressure may not have been optimally recorded at the exact time of the arterial blood gas measurement. Although critical illness as defined here is more common than culture-proven SBI, it is still too uncommon to be studied efficiently with a prospective design. Figure 3 also shows that, although the number of deaths in our cohort appears relatively low when compared against the group of deaths occurring very early, it is actually an important group of preventable
deaths, because many of the other deaths (those that occur with low Apgar scores or among infants who become critically ill in the first hour of age) either occur despite transfer to a tertiary center or are due to anomalies, over which clinicians have little control. We also recognize that we will not always be able to control for treatment effects or for differences in testing (i.e., for the same degree of respiratory distress, one clinician may order ABGs every hour X 3 whereas another may only order a single test, or no test).

It is worth pointing out that in our experience the more commonly taken approach to the problem of predicting critical illness and the need for transport is anecdotal: a newborn develops critical illness and dies, not having been transported to a tertiary care center in a timely manner. At that point clinicians, employing 20/20 hindsight, try to identify factors that should have been apparent to treating clinicians before the infant’s deterioration. Involved clinicians are then alert to this particular symptom or constellation of symptoms for some time, until gradually the memory of the event fades or is superseded by another “bad baby” with a different constellation of predictors. We believe our systematic approach, which emphasizes careful statistical analysis using a large dataset assembled over multiple years and institutions, will arrive at a more valid and stable risk prediction method. Moreover, the results of our study could be employed to design future randomized trials.

D.6. POPULATION DESCRIPTION AND SAMPLING PLAN

D.6.a. Study Settings (See also Evaluation and Treatment Protocols in the Appendix)

The KPMCP's Northern California Region is a group-model managed care organization. Under a mutual exclusivity contract, approximately 4,000 physicians of The Permanente Medical Group, Inc., provide care for 3,200,000 Kaiser Foundation Health Plan, Inc. members at facilities owned by Kaiser Foundation Hospitals, Inc. The KPMCP provides most perinatal care required by its members. Providers include perinatologists, neonatologists, neurosurgeons, pediatric surgeons, and geneticists. Approximately 32,000 births/year take place in the 12 health plan birth facilities included in this study; these facilities have delivery volumes ranging between 1,200 to 4,000 births per year. Of the 12 facilities included in this study, 6 (the Alta Bates, Hayward, Sacramento, San Francisco, Santa Clara, and Walnut Creek facilities) have level III NICUs staffed by board certified neonatologists and 6 (the Fresno, Redwood City, Santa Rosa, Santa Teresa, South Sacramento, and Vallejo) are level II units that transport ill newborns to the 6 level III NICUs. The Beth Israel Deaconess Medical Center (BIDMC) is a large tertiary birthing center, with approximately 5,000 births per year, a level III NICU with 35 licensed beds, and 24 hour in-hospital coverage by neonatologists and nurse practitioners. The Brigham and Women's Hospital (BWH) is one of the largest tertiary birthing hospitals in the country, with approximately 10,000 deliveries annually. Their 40-bed NICU is staffed 24 hours per day by pediatric residents, neonatology fellows and an attending neonatologist.

All four participating institutions and all investigators have impressive track records in conducting perinatal research. However, significant differences exist between the study sites. Both BWH and BIDMC are part of the Harvard Newborn Medicine Program and have employed a standardized sepsis work-up protocol for over 10 years (see Appendix for evaluation and treatment protocols employed by the study sites). In contrast, no standardized protocol was in place at the KPMCP sites until 1999, when a formal guideline (also in the Appendix) was developed based on the Northern California KPMCP study of Escobar et al. and the Southern California KPMCP study by Bromberger et al. While the KPMCP serves an insured population, both the BWH and BIDMC sites have approximately 18% of their deliveries from indigent patients. There are also significant differences in practice with respect to GBS screening. The KPMCP, which adopted a risk-based strategy in 1994, screened only 10-15% of women for GBS carriage prior to 2003, when the CDC recommendations for universal screening were adopted. In contrast, in recent years the BWH and BIDMC sites have attempted to implement protocols to improve screening efficiency of pregnant women so that transition to the new CDC guidelines did not represent as much of a change as in the KPMCP.

D.6.b. Description of the Populations of the 3 Study Sites

Table 1, below, provides population characteristics for the study sites. We do not have detailed socioeconomic status data for our populations, but we can make some general inferences. The KPMCP’s overall population distribution is racially similar to that of the general Northern California population but has fewer patients in the upper and lower ends of the income distribution. Two recent randomized controlled
trials on early newborn discharge conducted by the KPMCP team have found that the average education level in the program is somewhat higher than the general Northern California population. In contrast, the BWH and BIDMC have a significant proportion of indigent (internal reviews show that 17% of patients fall into this category) patients as well as approximately 1% uninsured patients.

TABLE 1: POPULATION CHARACTERISTICS

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Sites</th>
<th>BWH</th>
<th>BIDMC</th>
<th>KP (all hospitals combined)</th>
<th>KP (range across 12 hospitals)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N, births in 2000</td>
<td>9721</td>
<td>5,016</td>
<td></td>
<td>30,747</td>
<td>1,339 - 3,667</td>
</tr>
<tr>
<td>N, 34-36 wks</td>
<td>725</td>
<td>273</td>
<td></td>
<td>1,833</td>
<td>48 - 266</td>
</tr>
<tr>
<td>N, 37+ wks</td>
<td>8551</td>
<td>4,681</td>
<td></td>
<td>28,914</td>
<td>1,229 - 3,328</td>
</tr>
<tr>
<td>% White</td>
<td>64%</td>
<td>62%</td>
<td></td>
<td>41.8%</td>
<td>14.8% - 66.5%</td>
</tr>
<tr>
<td>% African American</td>
<td>13%</td>
<td>11%</td>
<td></td>
<td>8.1%</td>
<td>1.2% - 24.0%</td>
</tr>
<tr>
<td>% Asian</td>
<td>5%</td>
<td>10%</td>
<td></td>
<td>17.7%</td>
<td>5.3% - 31.6%</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>11%</td>
<td>5%</td>
<td></td>
<td>26.2%</td>
<td>15.8% - 43.1%</td>
</tr>
<tr>
<td>% other races</td>
<td>7%</td>
<td>12%</td>
<td></td>
<td>6.2%</td>
<td>1.3% - 18.2%</td>
</tr>
</tbody>
</table>

D.6.c. Identification and Verification of Sepsis and Control or Non-Infected Status

Dr. Escobar and Dr. Lieberman have concluded that it is not possible to verify the presence of SBI using only electronic information sources. There are two reasons for this. First, there are many false positive blood or cerebrospinal fluid culture results (in Escobar et al’s cohort study, the ratio of true positive to false positive cultures was approximately 1 to 1). Second, it is difficult to distinguish between true and false positive results based on data that are available electronically. This is because multiple, non-standardized, text strings are employed (e.g., no growth, no growth so far, no growth at 24/72/96 hours, S. epi, S. epidermidis, etc.). Consequently, in order to identify newborns with SBI, we will need to screen newborn records meeting the following criteria: 1) positive blood or cerebrospinal fluid culture results, and/or 2) presence of International Classification of Diseases codes indicating infection in a sterile site (e.g., codes 038.0 and 511.1, which would indicate group B streptococcal infection that was confirmed by pleural culture without a positive blood or cerebrospinal fluid culture) AND either prolonged length of stay (> 72 hours) or death. Newborns not meeting criteria for SBI (including newborns whose records were screened found not to have SBI) will be eligible to be controls (Specific Aim 1) or, if they had a CBC, will be classified as non-infected for Specific Aim 2.

D.6.d. Capture of CBC data for Specific Aim 2

At all 3 study sites, hospitalization databases will be employed to identify eligible infants based on gestational age criteria (≥ 34 weeks) as well as the absence of ICD codes for hematologic disorders or major anomalies as defined by the Vermont Oxford Network. We will then link these records to laboratory databases to obtain CBC results for those infants who had a CBC and a blood culture obtained within 1 hour of each other during the study time frame.

D.6.e. Sample size and power for Specific Aims 1 and 2

For the calculations below, we use the more restrictive definition of sepsis (definition 1). Sample size and power will be greater when probable sepsis is included (definition 2). We estimate that we will have about 350 cases of confirmed sepsis available for study. Because some risk factors we plan to study will be quite uncommon in the controls (e.g. significant tachypnea or high maternal fevers) we plan on including 3 controls per case for specific aim 1. Given this number of cases and controls, at alpha = 0.05, two-sided and beta = 0.2, the remaining determinant of the minimum detectable odds ratio (MDOR) is the prevalence of the risk factor (or risk indicator) in the control group. We present just a few examples of MDORs in Table 2, in order to show the range of MDORs over a wide range of risk factor prevalences (from about 51% to 1%). Over this range, the MDORs range from 1.4 to 3.5. Table 2 is based on 2/3 of the sample (233 cases, 699 controls).
TABLE 2: MINIMUM DETECTABLE ODDS RATIOS (MDORs) FOR SPECIFIC AIM 1

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Prevalence in controls</th>
<th>MDOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex</td>
<td>0.51</td>
<td>1.44</td>
</tr>
<tr>
<td>Gestational age &lt; 38 wk</td>
<td>0.19</td>
<td>1.54</td>
</tr>
<tr>
<td>Maximum RR ≥ 60</td>
<td>0.10</td>
<td>1.72</td>
</tr>
<tr>
<td>Maternal T ≥ 100° F</td>
<td>0.07</td>
<td>1.85</td>
</tr>
<tr>
<td>Maternal T ≥ 101° F</td>
<td>0.02</td>
<td>2.63</td>
</tr>
<tr>
<td>Maternal T ≥ 102° F</td>
<td>0.01</td>
<td>3.50</td>
</tr>
</tbody>
</table>

TABLE 3: PREDICTED LRs AND THEIR CONFIDENCE INTERVALS FOR SPECIFIC AIMS 1 AND 2

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Prevalence in controls</th>
<th>Expected prevalence in cases</th>
<th>Predicted LR</th>
<th>95% confidence interval for predicted LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Maternal Temp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;100.0°F</td>
<td>0.930</td>
<td>0.500</td>
<td>0.54</td>
<td>0.45 - 0.65</td>
</tr>
<tr>
<td>100.0-100.9°F</td>
<td>0.048</td>
<td>0.210</td>
<td>4.38</td>
<td>2.44 - 7.84</td>
</tr>
<tr>
<td>≥101.0°F</td>
<td>0.022</td>
<td>0.290</td>
<td>13.18</td>
<td>6.21 - 27.99</td>
</tr>
<tr>
<td>Maximum RR 0-12 hr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>0.640</td>
<td>0.200</td>
<td>0.31</td>
<td>0.22 - 0.45</td>
</tr>
<tr>
<td>50-59</td>
<td>0.260</td>
<td>0.130</td>
<td>0.50</td>
<td>0.30 - 0.83</td>
</tr>
<tr>
<td>60-69</td>
<td>0.075</td>
<td>0.250</td>
<td>3.33</td>
<td>2.06 - 5.40</td>
</tr>
<tr>
<td>≥70</td>
<td>0.025</td>
<td>0.420</td>
<td>16.80</td>
<td>8.45 - 33.40</td>
</tr>
<tr>
<td>First ANC* (per mm$^3$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 2000</td>
<td>0.014</td>
<td>0.140</td>
<td>10.00</td>
<td>6.24 - 16.03</td>
</tr>
<tr>
<td>2000-3999</td>
<td>0.050</td>
<td>0.140</td>
<td>2.80</td>
<td>1.77 - 4.42</td>
</tr>
<tr>
<td>4000-5999</td>
<td>0.100</td>
<td>0.140</td>
<td>1.40</td>
<td>0.89 - 2.21</td>
</tr>
<tr>
<td>6000-7999</td>
<td>0.140</td>
<td>0.190</td>
<td>1.36</td>
<td>0.93 - 1.98</td>
</tr>
<tr>
<td>8000-19999</td>
<td>0.623</td>
<td>0.360</td>
<td>0.58</td>
<td>0.45 - 0.74</td>
</tr>
<tr>
<td>≥ 20000</td>
<td>0.073</td>
<td>0.030</td>
<td>0.41</td>
<td>0.15 - 1.16</td>
</tr>
</tbody>
</table>

*ANC = absolute neutrophil count. ANC results are presented in absolute terms, as these will be more familiar to reviewers. As discussed in Section D.5.c, they will be transformed to z-scores. The LRs and their 95% CI will then depend on the width of z-score categories, which may be similar to what is shown here. The number of “controls” available for analyses (after splitting the sample) is more than 20,000, so precision is completely limited by the number of cases, set at 117 for the validation data set (33% of 350).

The MDORs in the third column of Table 2 represent a traditional approach to power calculations for a case-control study with fixed sample size and dichotomous predictor variable, involving hypothesis testing and the specification of alpha and beta. Table 3, based on 1/3 of the sample, provides a more sophisticated approach to sample size. Table 3, which is especially relevant for Specific Aim 2, acknowledges that in the case of diagnostic tests like the CBC, where the goal is not identification of causal relationships but enhancing prediction of disease, statistical significance is necessary but not sufficient for clinical usefulness. The goal of such studies is estimation, rather than hypothesis testing. The main parameters we wish to estimate in the current study are LRs. Furthermore, we have emphasized the loss of information that results when predictor variables are dichotomized. Therefore, another way of approaching the sample size and power discussion is to show the projected width of the confidence intervals for the LRs we plan to estimate treating predictor variables as categorical rather than dichotomous variables. Examples of these are shown in Table 3. Again, the confidence intervals associated with a wide range of risk factor prevalences and projected LRs are presented. The estimates of prevalence of the various maternal temperatures in the controls for Tables 2 and 3 come from a study that abstracted data from 400 randomly selected KPMCP term deliveries, while the estimates for respiratory rates are from a previous nested-case control study, for which controls were randomly selected from the KPMCP birth cohort. The estimated prevalences of clinical risk factors among the cases and prevalences of the different ANC values in both cases and controls are from our previous study of infants receiving sepsis work-ups. Our sample size is adequate to allow upper and lower confidence intervals for the likelihood ratios of all but the most uncommon values of predictors to be less than a factor of 2 from one another with the conservative approach of using only the validation dataset to estimate confidence intervals. Given the low prior probability of SBI, this should provide adequate precision of likelihood ratio estimation.
D.6.f. Identification and Verification of Significant Respiratory Distress and Critical Illness Status

We will employ the KPMCP, BWH, and BIDMC laboratory databases to identify infants who meet initial entry criteria (having had at least one ABG at < 6 hours of age). These babies’ records will then be linked to hospitalization databases and NICU databases, which will permit removal of babies who meet the exclusion criteria listed in section D.1.b. After removing records of babies meeting exclusion criteria, we will identify all infants meeting our definition of a critical illness at all study sites during the 2002-2005 calendar years. We will then employ 2/3 of the cohort for a derivation dataset and 1/3 as a validation dataset. Thus, the total sample size for Specific Aim 3 will be 267 cases and 533 controls for the derivation dataset and 133 cases and 267 controls for the validation dataset.

D.6.g. Sample size justification for Specific Aim 3

There are no extant population based data on what predictors available in the first 6 hours of age can predict development of a neonatal critical illness (whether defined by ABG values or by any other proxy). Consequently, our decision to employ a sample size of 400 cases and 800 controls for our derivation dataset is somewhat arbitrary and is partly based on budgetary considerations.

### TABLE 4: LIKELIHOOD RATIOS (LRs) AND THEIR CONFIDENCE INTERVALS FOR SPECIFIC AIM 3

<table>
<thead>
<tr>
<th>Exposure (Measured at Age&lt; 6 hours)</th>
<th>Prevalence in controls</th>
<th>Expected prevalence in cases</th>
<th>Predicted LR</th>
<th>95% confidence interval for predicted LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PaO2:FiO2 ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;2.5</td>
<td>0.64</td>
<td>0.25</td>
<td>0.39</td>
<td>0.34 – 0.49</td>
</tr>
<tr>
<td>1.5 – 2.49</td>
<td>0.20</td>
<td>0.10</td>
<td>0.50</td>
<td>0.34 - 0.74</td>
</tr>
<tr>
<td>1.0-1.49</td>
<td>0.16</td>
<td>0.65</td>
<td>4.06</td>
<td>3.28 – 5.03</td>
</tr>
<tr>
<td>Lowest mean arterial blood pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;40mmHg</td>
<td>0.45</td>
<td>0.35</td>
<td>0.78</td>
<td>0.64 – 0.94</td>
</tr>
<tr>
<td>30 – 40 mmHg</td>
<td>0.45</td>
<td>0.45</td>
<td>1.00</td>
<td>0.85 – 1.18</td>
</tr>
<tr>
<td>&lt;30mmHg</td>
<td>0.10</td>
<td>0.20</td>
<td>2.00</td>
<td>1.41 – 2.84</td>
</tr>
<tr>
<td>Lowest pH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥7.30</td>
<td>0.70</td>
<td>0.40</td>
<td>0.57</td>
<td>0.49 – 0.67</td>
</tr>
<tr>
<td>7.25 – 7.299</td>
<td>0.15</td>
<td>0.20</td>
<td>1.33</td>
<td>0.97 – 1.82</td>
</tr>
<tr>
<td>7.10 – 7.249</td>
<td>0.15</td>
<td>0.40</td>
<td>2.67</td>
<td>2.08 – 3.42</td>
</tr>
</tbody>
</table>

However, based on our experience with the SNAP-II and the pilot study described in Section C.4, we have estimated the prevalence of certain ABG related predictors among infants with respiratory distress < 6 hours old who do and who do not go on to develop a critical illness at ≥ 1 hour. These frequencies have been used for the estimates shown in Table 4, above, which are based on 267 cases and 533 controls. At this point, we cannot estimate the expected frequencies of other predictors (e.g., pulse oximetry < 90%, or ANC < 10%ile). However, the above frequencies provide useful ranges that permit estimation of the precision of our LR estimates, given our available sample size.

D.6.h. Summary of Patient Samples

This section summarizes our sampling and shows where Specific Aims 1 and 3 have some overlap in cases and controls. Specific Aim 2 is not discussed here because it employs the same cases as Specific Aim 1. Our final chart numbers are as follows (for the sake of simplicity, only neonatal charts are listed here; the number of maternal charts will be slightly different due to multiple births and multiple pregnancies). The “Analyses” column refers to the number of charts used for analytic purposes, while the “Unique” column refers to the number of charts actually pulled.

<table>
<thead>
<tr>
<th>Analyses</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific Aim 1, definition 1</td>
<td>1400</td>
</tr>
<tr>
<td>Specific Aim 1, definition 2</td>
<td>1180</td>
</tr>
<tr>
<td>Specific Aim 3</td>
<td>1200</td>
</tr>
</tbody>
</table>
From the population study conducted by Escobar et al\textsuperscript{29}, we anticipate that, among the 45 babies with positive cultures born in a typical year of our study, 10 will be critically ill, 15 will be symptomatic but will not have a critical illness, and 20 will be asymptomatic. From our other studies, which are summarized in Figure 3, we anticipate that, over an 8 year period, we will have 2100 babies with a critical illness (approximately 260 per year). Of these 2100 babies, we expect that 950 will have had a deterioration within the first hour of age. These infants are eligible for Specific Aim 1 (definition 2) but not for Specific Aim 3. Another group of 600 babies, those that deteriorate between 1.0 and 5.9 hours of age, who were previously ineligible for Specific Aim 3, are now eligible for that aim. Lastly, there is the group of babies who deteriorate at ≥ 6 hours, who are eligible for all specific aims.

(i) Specific Aim 1, definition 1: only culture(+) babies

350 cases
1050 randomly selected controls

(ii) Specific Aim 1: definition 2: culture(+) babies OR critically ill babies

Since we are employing a one year cohort (the 2005 calendar year), we would have 260 critically ill babies with 780 randomly selected controls and 45 babies with positive cultures with 135 randomly selected controls. However, 10 of the critically ill babies would also have had a positive culture (and would have had 30 controls reviewed as well), so the net new number of charts is 250 cases and 750 controls. For analytic purposes, the actual number of cases will be 260 + 45 – 10, or 295 and the number of controls will be 780 + 135 – 30, or 885. This gives a total of 1180.

(iii) Specific Aim 3: critically ill babies

If we take approximately 4 years’ of babies (the 2002-2005 year cohort) who became critically ill after 1 hour, we would have 400 cases and 800 controls. However, some of the cases are also being reviewed for Specific Aim 1:

40 culture(+) babies who also became critically ill, previously reviewed for Specific Aim 1, definition 1
120 babies who were critically ill after 1 hour of age, previously reviewed for Specific Aim 1, definition 2

Thus, for Specific Aim 3, the number of unique charts being reviewed is 1200 minus 160, or 1040 charts.

D.7. PLAN FOR DATA ANALYSIS

D.7.a. Analytic Strategy for Specific Aims 1 and 3

The basic tasks and hence the analytic approaches for the two case-control studies are similar. For both studies, we will have a large number of candidate predictor variables, many of them ordinal, categorical, or continuous, and a dichotomous outcome variable. We will begin with exploratory bivariate analyses of predictor variables that have been identified in previous studies, categorizing them as has been done previously. We will then use two approaches to selecting variables and how they will be coded for a final model: stepwise logistic regression and recursive partitioning (also known as classification and regression trees or CART analysis)\textsuperscript{158-161}. Logistic regression has the advantage that the resulting model is simpler and more easily summarized – we will estimate odds ratios for predictor variables, from which we can obtain LRs as discussed below. For variables for which we expect a dose-response relationship (e.g., maternal temperature) we will investigate whether the logistic model fits by comparing odds ratios per unit change in predictor variables (e.g., the odds ratio per degree for maternal temperature) with those obtained by creating indicator variables for several levels of the risk factor. If the logistic model fits, the odds ratio for a maternal temperature of 101 to 101.9 °F should be about the square of the odds ratio for 100-100.9 °F. If the logistic model fits these relationships poorly, we will dichotomize continuous predictor variables or replace them with indicator variables for different categories. We will assess the discrimination of various competing logistic models using the c-statistic (equivalent to the area under the receiver operator characteristic curve), and assess the goodness of fit of overall logistic models using the Hosmer-Lemeshow method.

The other approach to identifying predictors and quantifying prediction is recursive partitioning.

Recursive partitioning has the advantage that no assumptions are made about how predictor variables are
related to outcome, and non-monotonic relationships between predictors and disease risk (e.g. with gestational age, pH, or ANC) can be identified. We plan a variety of exploratory analyses using recursive partitioning, setting the ratio of costs of false negatives to false positives between 1:10 and 1:200. The end result of both logistic analyses and recursive partitioning will be a mechanism that will allow us to use values of clinical variables available at the time of first concern about SBI to estimate a particular infant’s probability of SBI. This probability can be used to help clinicians decide whether additional tests (i.e., CBC and blood culture) or monitoring (e.g., more frequent vital signs) are needed. If a CBC is done, this probability estimate can then be used as the prior probability in calculations that will use LRs for different results on the CBC to refine the estimate of the infant’s probability of SBI, to a posterior probability (after the CBC result), which can be used to assist with decisions about whether to initiate antibiotic treatment.

The logistic model will provide multivariate odds ratios. Because SBI and critical illness are rare, these will closely approximate risk ratios, which are much more readily interpretable. However, in clinical situations when the prior probability is known and the goal is to estimate posterior probability, even risk ratios are difficult to employ. In this situation, what would be most useful are LRs. The ease of use of LRs is enhanced in this case because the rarity of the disease allows clinicians to skip the conversion to odds and just multiply prior probabilities by LRs. The conversion between risk ratios and LRs is straightforward algebraically given the prevalence of each level of the risk factor and the population incidence of the disease, but we have not found any articles describing this process, nor is it a currently available feature in the STATA statistical package. If after additional searching and consultation with colleagues we do not find a reference in which this conversion is presented in a way accessible to clinicians, one task of this proposal (to be headed by Dr. Newman) will be to produce such a paper. In addition, we will create utilities within the STATA statistical package and (at least for dichotomous predictors) for hand-held personal digital assistants to facilitate this conversion. The formula for converting risk ratios to likelihood ratios, derived by Drs. Hudes and Newman for this application, is included in the Appendix.

Section D.7.a (addendum): Dynamic Bayesian Predictive Modeling

The first approach to developing a predictive model for critical illness is one with which we are quite familiar, as is shown by our recent work with the Richardson score and our past work with the SNAP. This approach, which we describe as “static,” because of the way it handles missing information and changing information, has become traditional in severity of illness scoring. In the traditional approach, two important assumptions are made in the course of modeling. The first is that 1) when no data are available in the relevant time frame (e.g., the first 24 hours in the adult intensive care unit for the APACHE-III score, or the first 12 hours in the NICU for the SNAP-II or Richardson score), the assumption is made that the value for the missing data element is “normal;” and 2) when multiple values for a data element are available in the time frame, the “worst” (furthest from clinical normality) value is selected. Our second approach will be to treat the multiple readings for each newborn as a time series and use a method called dynamic linear modeling, which is a generalization of the multi-process Kalman filter, a technique that has been used in fields such as engineering for many decades to perform adaptive prediction of phenomena which change over time. To see why adaptive forecasting may be advantageous in this situation, consider a hypothetical 12 hour severity of illness score for newborns ≥ 34 weeks gestation that only consisted of two predictors (arterial pH and mean arterial blood pressure, or MAP) in a universe of only two newborns with respiratory distress.

<table>
<thead>
<tr>
<th>Age (hours)</th>
<th>Baby 1</th>
<th>Baby 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pH</td>
<td>MAP (mm Hg)</td>
</tr>
<tr>
<td>0:55</td>
<td>7.41</td>
<td>45</td>
</tr>
<tr>
<td>1:10</td>
<td>--</td>
<td>43</td>
</tr>
<tr>
<td>3:45</td>
<td>7.29</td>
<td>40</td>
</tr>
<tr>
<td>4:10</td>
<td>--</td>
<td>38</td>
</tr>
<tr>
<td>6:15</td>
<td>7.28</td>
<td>39</td>
</tr>
<tr>
<td>9:20</td>
<td>--</td>
<td>30</td>
</tr>
<tr>
<td>11:30</td>
<td>7.20</td>
<td>24</td>
</tr>
<tr>
<td>11:45</td>
<td>Chest compressions required</td>
<td>--</td>
</tr>
</tbody>
</table>

Using the traditional, static approach to severity of illness scoring, both of these babies would be assigned the same score based on a lowest pH of 7.20 and a lowest MAP of 24 mm Hg. However, if one takes the
D.7.b. Addendum: Analytic Strategy for Babies with Asymptomatic Bacteremia

For Study Aims 1 and 2, we will first conduct our analyses using all infants (i.e., including babies who meet our strict definition of asymptomatic bacteremia). We will then characterize the infants with asymptomatic bacteremia with respect to key predictors (e.g., did their mothers receive intrapartum antibiotics?) and descriptors (e.g., do these infants tend to be of greater birth weight?). We will then repeat our analyses to determine the effect of removing these infants.

D.7.b. Analytic strategy for Specific Aim 2

Our initial step in analyzing CBC results will be to convert results for each relevant component of the CBC (total white blood cell count, ANC, etc.) to an hour-specific z-score using a moving average technique that employs a 3-hour window. Thus, to establish z-scores for age 1 hour, we will employ all CBCs obtained up to 2.5 hours; for ranges for age 2 hours, we will employ all CBCs obtained from 0.5 to 3.5 hours; and so forth. We are choosing to use z-scores rather than percentiles, which would be more familiar to clinicians, because their distribution will be much closer to Gaussian and because they will be more stable at the extremes of the distribution where, in spite of the large sample size, we will have relatively few subjects. Since we wish to conduct numerous exploratory analyses, we will split the sample, using 2/3 of the subjects as a derivation dataset and reserving 1/3 for validation. We will use ROC curves to 1) examine the relationship between SBI and both the raw CBC components and their z-scores; 2) determine which CBC components best predict SBI; and 3) see whether the z-score transformation improves discrimination. In so doing we will be mindful of the possibility of ROC curves whose slopes do not increase monotonically. If this is observed, it indicates that the area under the ROC curve is not a good measure of the discrimination of the test because the distribution of the predictor among those with sepsis may be bimodal (e.g., if newborns with SBI have total white blood cell counts that are both higher and lower than control infants). In that case, the different components of the CBC

D.7.b. Addendum: Analytic Strategy for Babies with Asymptomatic Bacteremia

For Study Aims 1 and 2, we will first conduct our analyses using all infants (i.e., including babies who meet our strict definition of asymptomatic bacteremia). We will then characterize the infants with asymptomatic bacteremia with respect to key predictors (e.g., did their mothers receive intrapartum antibiotics?) and descriptors (e.g., do these infants tend to be of greater birth weight?). We will then repeat our analyses to determine the effect of removing these infants.

D.7.b. Analytic strategy for Specific Aim 2

Our initial step in analyzing CBC results will be to convert results for each relevant component of the CBC (total white blood cell count, ANC, etc.) to an hour-specific z-score using a moving average technique that employs a 3-hour window. Thus, to establish z-scores for age 1 hour, we will employ all CBCs obtained up to 2.5 hours; for ranges for age 2 hours, we will employ all CBCs obtained from 0.5 to 3.5 hours; and so forth. We are choosing to use z-scores rather than percentiles, which would be more familiar to clinicians, because their distribution will be much closer to Gaussian and because they will be more stable at the extremes of the distribution where, in spite of the large sample size, we will have relatively few subjects. Since we wish to conduct numerous exploratory analyses, we will split the sample, using 2/3 of the subjects as a derivation dataset and reserving 1/3 for validation. We will use ROC curves to 1) examine the relationship between SBI and both the raw CBC components and their z-scores; 2) determine which CBC components best predict SBI; and 3) see whether the z-score transformation improves discrimination. In so doing we will be mindful of the possibility of ROC curves whose slopes do not increase monotonically. If this is observed, it indicates that the area under the ROC curve is not a good measure of the discrimination of the test because the distribution of the predictor among those with sepsis may be bimodal (e.g., if newborns with SBI have total white blood cell counts that are both higher and lower than control infants). In that case, the different components of the CBC
will be assessed by replacing results or z-scores with up to 7 indicator variables, entering these into a logistic model, and comparing the discrimination of different models using the area under the ROC curve for the logistic model.

Using the above method of quantifying discrimination, we will then test various possible refinements of the process, including consideration of gestational age, race, and other factors when computing ANC z-scores. We will also experiment with indices created from more than one component of the CBC. For example, since both a low ANC and a high immature:total (I:T) neutrophil ratio predict SBI, one might expect that combining the two – e.g., by dividing the I:T ratio by the ANC – would improve discrimination. Finally, we will explore possible interactions. For example, we will investigate whether the total white blood cell count performs better with some organisms than others. We will also look for differences between newborns with SBI and those with bacteremia only, since it is possible that newborns with SBI are more likely to have abnormal neutrophil kinetics than newborns with bacteremia only. Inspection of the shape of the ROC curves will also help us choose categories for the raw ANC and ANC z-scores, the ideal cut-points for categorizing values being those at which the slope of the ROC curve changes. As the penultimate step, after we have determined the best way to combine or transform the CBC results, we will compute LRs for each category using the portion of the data reserved for validation. Since it is not possible for us to employ ANC datasets obtained from healthy newborns, the final step in our analyses will be to test the stability of our LR estimates using simulated ANC datasets. These simulated datasets will be based on distributions from the small number of studies that have studied ANC kinetics among completely healthy newborns.

D.7.c. Overall Strengths of this Project

The greatest strength of our study is that it combines a comprehensive response to a clinical need with rigorous methodology. Existing recommendations are not tailored to clinical realities – textbooks are too vague, while other recommendations are either too restrictive (e.g., those confined to the management of GBS only) or too broad (e.g., those that lump data from very premature infants with data from term babies). In contrast, our target population is one commonly encountered by general pediatricians throughout the developed world. Armed with our data, clinicians could make more rational decisions about which infants to observe, which to treat, and which to refer. Clearer criteria also can permit auditing compliance with evidence based guidelines, something which is currently impossible. Knowing which babies are at extremely high risk of “crashing” is also helpful in terms of staffing and planning for inter-hospital transport. Since some babies still experience delays in diagnosis and receipt of antibiotic treatment, clearer criteria would also decrease adverse outcomes due to delayed diagnosis. Further, using our findings, it would be possible to develop handheld computer-based applications that could be used in real time, where all a clinician would need to do is enter a few data points – gestational age, chronological age, CBC results, ABG results – to obtain probabilities for key outcomes and recommendations for testing, treatment, or referral. Our study also addresses an important problem that has been sidestepped by much of the “rule out sepsis” literature – what to do about babies with negative culture results but with respiratory distress and risk for clinical deterioration. It also draws from a broad population that is diverse from both a racial as well as a socioeconomic perspective. Finally, it is large enough to permit stratified analyses for different organisms, not just GBS.

D.8. TIMELINE FOR PROJECT COMPLETION

<table>
<thead>
<tr>
<th>Month</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design and Preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRB Approvals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form design and pilot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification of cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical record abstraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data entry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linkage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Cleaning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manuscript preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
E. HUMAN SUBJECTS

E.1. Institutional Review Board Approval

This project has been approved by 4 Institutional Review Boards (IRBs) for the protection of human subjects (at BIDMC, BWH, UCSF and at the KPMCP, which has one IRB with jurisdiction over all 12 KPMCP hospitals participating in this study). Copies of these approvals will be provided to the NICHD upon request. We are also submitting this proposal to the Institutional Review Board of the University of California, Santa Cruz.

E.2. Protection of Human Subjects

E.2.a. Risks to the Subjects

As described in sections A-D, above, this study involves paper and electronic data from 1) women who delivered a liveborn infant ≥ 34 weeks gestation at one of the 14 study hospitals and 2) infants born to these mothers. Table 4 and Section D.6.b summarize the characteristics of our study population. The only risk to participants is loss of privacy. As described in Section D, above, it is not feasible to conduct this research using other methods.

E.2.b. Adequacy of Protection Against Risks

During the data collection phase, we will employ standard operating procedures, which are already in place at our participating institutions, to protect patient confidentiality. Only the study staff will have access to data collected as part of the study. All study staff have signed, or, if not yet hired, will sign, a confidentiality oath. All key personnel have successfully completed and new study staff will complete IRB-approved Protection of Human Research Subjects Training. Certificates of completion will be submitted upon request. No identifying information will be used in any report or publication that is produced as part of this study. All data collected as part of this study will be maintained in locked file cabinets or encrypted computer files with limited password access.

E.2.c. Recruitment and Informed Consent

This will be a records-only study. We do not need patient identifiers to conduct our analyses. Therefore, after data cleaning, we will strip our datasets of all patient identifiers and replace them with study ID numbers, retaining only those identifiers permitted for a Limited Data Set for Research under the Privacy Rule of the Health Insurance Portability and Accessibility Act. Given these precautions, the risk to study subjects is extremely low and our 4 IRBs have exempted us from obtaining consent from study subjects.

E.2.d. Potential Benefits of the Proposed Research to the Subjects and Others

Most mothers and infants whose records are involved in this study are unlikely to reap significant benefits from this research. A small proportion could benefit from improved management of future offspring. In contrast, the benefits to society could be substantial, as this study’s findings could lead to significant improvement in the outcomes of newborns ≥ 34 weeks gestation.

E.2.e. Importance of the Knowledge to be Gained

This study’s findings could lead to significant improvements in the health of newborns. Given the extremely low risk to participants, we feel that the risk benefit ratio is extremely low.

Inclusion of women
E.2.f. Inclusion of Women, Children, and Minorities

This study specifically targets women and newborns as its study population. As shown in Section D.6.b., our study population is racially and ethnically diverse.

Targeted/Planned Enrollment Table

This report format should NOT be used for data collection from study participants.

**Study Title:** Sepsis and Critical Illness in Babies ≥ 34 Weeks Gestation  
**Total Planned Enrollment:** 47,600

<table>
<thead>
<tr>
<th>TARGETED/PLANNED ENROLLMENT: Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ethnic Category</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
</tr>
<tr>
<td><strong>Ethnic Category: Total of All Subjects</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Racial Categories</th>
<th>Females</th>
<th>Males</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian/Alaska Native</td>
<td>430</td>
<td>430</td>
<td>860</td>
</tr>
<tr>
<td>Asian</td>
<td>3,950</td>
<td>3,950</td>
<td>7,180</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>1,900</td>
<td>1,900</td>
<td>3,800</td>
</tr>
<tr>
<td>Black or African American</td>
<td>2,980</td>
<td>2,980</td>
<td>5,960</td>
</tr>
<tr>
<td>White</td>
<td>14,900</td>
<td>14,900</td>
<td>29,800</td>
</tr>
<tr>
<td><strong>Racial Categories: Total of All Subjects</strong></td>
<td>23,800</td>
<td>23,800</td>
<td>42,600</td>
</tr>
</tbody>
</table>

* The "Ethnic Category: Total of All Subjects" must be equal to the "Racial Categories: Total of All Subjects."

F. VERTEBRATE ANIMALS

Not applicable.
G. LITERATURE CITED


H. CONSORTIUM/CONTRACTUAL ARRANGEMENTS

See included materials for the following Consortia:

Beth Israel Deaconess Medical Center, Boston Massachusetts
Brigham & Women’s Hospital, Boston, Massachusetts
University of California, San Francisco, California
University of California, Santa Cruz, California

We are not planning to employ any consultants in this project.

I. RESOURCE SHARING

1) Data Sharing Plan
As required by the NIH (http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html) we will prepare datasets that are compiled as part of this proposed study in accordance with PHS policy. If requested by the NIH, the prepared dataset will be sent to the appropriate data depository. Any dataset that will be shared will not contain identifiers. Kaiser’s Institutional Review Board approval may be required for requests.

2) Sharing Model Organisms:

N/A

J. LETTERS OF SUPPORT

Please see following pages 96 to 101.

9. APPENDIX [5 sets included]

APPENDIX 1: Healthy People 2010 objectives 14-16, 20, 21 and 16-1,14

APPENDIX 2: Educational tools based on new Richardson score (score developed in the pilot study described in C.4.)


APPENDIX 4: Kaiser Permanente Clinical Practice Guideline on Prevention and Screening of Neonatal Sepsis

APPENDIX 5: Harvard Joint Program in Neonatology Guidelines for “well appearing” newborns at risk for sepsis

APPENDIX 6:

Samples of a) recent neonatal and maternal charts from the study sites, highlighting data routinely collected for newborns with respiratory distress and/or suspected sepsis, and b) data collection forms for the “Unstudied Infants: Low Risk Babies in a High Risk Place” project, which was completed at the study sites on August 1, 2003, and which demonstrates data availability.

APPENDIX 7: General formula for conversion from relative risk (RR) to likelihood ratio (LR)