A Retrospective Database Analysis of Neonatal Morbidities to Evaluate a Composite Endpoint for Use in Preterm Labor Clinical Trials

Jeanne M. Pimenta, PhD1 Myla Ebeling, BS2 Timothy H. Montague, PhD3 Kathleen J. Beach, MD4 Jill Abell, PhD5,6 Michael T. O’Shea, MD7 Marcy Powell, MD, FACOG8 Thomas C. Hulsey, ScD2,9

1 Real World Evidence (Epidemiology), GSK, Uxbridge, Middlesex, United Kingdom
2 Division of Neonatology, Department of Pediatrics, Medical University of South Carolina, Charleston, South Carolina
3 Clinical Statistics, GSK, Philadelphia, Pennsylvania
4 Maternal Neonatal Health Unit, GSK, Research Triangle Park, North Carolina
5 Clinical Effectiveness and Safety Clinical Evaluation Science, GSK, Philadelphia, Pennsylvania
6 Janssen Scientific Affairs, Real World Evidence, Philadelphia, Pennsylvania
7 Department of Pediatrics (Neonatology Section), University of North Carolina, Chapel Hill, North Carolina
8 Safety Evaluation and Risk Mitigation, GSK, Research Triangle Park, North Carolina
9 Department of Epidemiology, West Virginia University, Morgantown, West Virginia

Address for correspondence Jeanne M. Pimenta, PhD, Real World Evidence (Epidemiology), GSK, Stockley Park West, 1–3 Ironbridge Road, Uxbridge, Middlesex, UB11 1BT, United Kingdom (e-mail: jeanne.m.pimenta@gsk.com).


Abstract

Objective To propose and assess a composite endpoint (CE) of neonatal benefit based on neonatal mortality and morbidities by gestational age (GA) for use in preterm labor clinical trials.

Study Design A descriptive, retrospective analysis of the Medical University of South Carolina Perinatal Information System database was conducted. Neonatal morbidities were assessed for inclusion in the CE based on clinical significance/risk of childhood neurodevelopmental impairment, frequency, and association with GA in a mother–neonate linked cohort, comprising women with uncomplicated singleton pregnancies delivered at ≥24 weeks’ GA.

Results Among 17,912 mother–neonate pairs, neonates were at a risk of numerous severe but infrequent morbidities. Clinically important, predominantly rare events were combined into a CE comprising neonatal mortality and morbidities, which decreased in frequency with increasing GA. The highest CE frequency occurred at <31 weeks. High frequency of respiratory distress syndrome, bronchopulmonary dysplasia, and sepsis drove the CE. Median length of hospital stay was longer at all GAs in those with the CE compared with those without.

Conclusions Descriptive epidemiological assessment and clinical input were used to develop a CE to measure neonatal benefit, comprising clinically meaningful outcomes. These empirical data and CE allowed trials investigating tocolytics to be sized appropriately.

Keywords

► neonatal benefit
► composite endpoint
► preterm birth
► preterm labor
► gestational age
► tocolytics

received May 2, 2017
accepted after revision January 6, 2018
ISSN 2157-6998.

Copyright © 2018 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
Tel: +1(212) 584-4662.

License terms
In 2010, 11% of live births worldwide were preterm (occurring before 37 weeks of gestation). Variation in estimated preterm birth rates was seen across countries, from approximately 5% in Europe to 18% in Africa. While approximately one-third of preterm births result from medical intervention to resolve maternal or fetal medical indications, 40 to 45% result from spontaneous preterm labor, the underlying pathogenesis of which is not fully understood. Preterm birth is the most common direct cause of neonatal death and indirectly affects infant mortality rates through associated morbidities. Consequences for the neonate can persist into later life, with sequelae such as visual and hearing impairment, lung disease, cardiovascular disorders, and neurodevelopmental impairments. Preterm birth also imposes a substantial socioeconomic burden on families (cost of care, loss of wages, and emotional impact) and wider society (acute and chronic health-care costs).

The global burden of preterm birth highlights the need for effective drugs to stop preterm labor, and, as such, tocolytic development is a discovery research priority of the World Health Organization's Every Newborn Action Plan to End Preventable Deaths. The goal of current tocolytics for the management of preterm labor is to prolong gestation, allowing time for the administration of antenatal corticosteroids or maternal transfer to another medical facility, but there is very limited evidence of a measurable benefit of tocolytics on neonatal outcomes. However, in considering atosiban for the treatment of preterm labor, the U.S. Food and Drug Administration stated that demonstrating prolongation of pregnancy was no longer sufficient for the approval of new tocolytics; randomized controlled trials (RCTs) of tocolytics should have a primary endpoint of neonatal benefit. Preterm birth study endpoints vary widely; a recent systematic review reported 72 different primary outcomes in 103 RCTs among diverse patient populations. The majority measured delay to delivery or perinatal mortality, and only five studies used a composite endpoint of neonatal morbidities and death, defined differently among each study. In addition, there are sparse empirical data available in the literature on the incidence of composite endpoints across the spectrum of gestational age (GA).

Small population sizes are common in preterm birth RCTs, and the infrequency of some severe morbidities underscores the need for a suitable composite endpoint to establish neonatal benefit. The European Network for Health Technology Assessment states that composite endpoints are appropriate in clinical trials in the absence of a single suitable primary endpoint. Moreover, there should be empirical and clinical evidence of the value of each component, and components should not be included in the composite if treatment is not expected to affect this variable.

Clinical development of new tocolytics (such as retosiban currently in development by GlaxoSmithKline) is required to demonstrate both prolongation of pregnancy and a benefit to the neonate. To inform the retosiban clinical trial program, the objective of this study was to use real-world empirical data to determine the frequencies of neonatal morbidities and mortality by GA in the patient population of interest and to propose a clinically meaningful composite endpoint to measure neonatal benefit in RCTs.

Material and Methods

Data Source and Patient Population

The study data source was the Medical University of South Carolina (MUSC) Perinatal Information System (PINS), a research-quality perinatal database for all women delivering at the MUSC, which is a regional tertiary referral hospital in the southeastern United States. An active surveillance system was used to populate the MUSC PINS, in which data were abstracted from clinical notes specifically for entry into the database. Data were entered by trained personnel, and logical error checks were performed routinely on all data; inter- and intrarater reliability both exceed 97%. The MUSC PINS database included detailed information on each mother’s medical history, linked to neonatal data (such as medical diagnoses, medications, surgical procedures, and laboratory tests) from delivery to hospital discharge; however, comprehensive antenatal care information from outside the hospital setting was not available. Ethical approval for this study was provided by the Institutional Review Board for Human Research at the MUSC.

A descriptive, retrospective cohort study of the MUSC PINS data was conducted from 2000 to 2011. The study cohort (termed simulated trial population) was created to match, as closely as possible within the constraints of the database, the population within the phase III retosiban clinical trials which was women with singleton uncomplicated pregnancies and intact membranes in preterm labor between 24 and 33 weeks of GA (Fig. 1). Women with multiple gestation pregnancies were excluded. Maternal pregnancy complications, as determined by trial exclusion criteria, were preterm prelabor rupture of membranes, hypertensive disorders of pregnancy (including severe preeclampsia and eclampsia-related conditions), placental conditions (such as placenta previa and placental abruption), and intra-amniotic infection or intrauterine growth restriction (see for full list of maternal complications used as exclusion criteria). Neonates with prespecified severe or life-threatening congenital conditions were also excluded. As antenatal care was not captured within the MUSC PINS, preterm labor diagnoses were not available and therefore did not form part of the inclusion criteria. All mother–neonate pairs who met the aforementioned inclusion and exclusion criteria and delivered at ≥24 weeks’ GA were included in the simulated trial population (Fig. 1).

Definition of Composite Endpoint

Clinical/Epidemiological Assessment of Data

The study investigated the frequency of approximately 100 neonatal outcomes (including neonatal medical diagnoses, surgical procedures, medication use, laboratory tests) by GA in the MUSC PINS database. The decision to consider further or to exclude each neonatal outcome in a potential composite
was made by the authors, additional members of the GlaxoSmithKline clinical team, and a consulting neonatologist (as detailed in the Acknowledgments section). They reviewed each of the outcomes and assessed its clinical significance, frequency of occurrence, association with GA, variability of diagnosis, and severity (particularly the associated risk of neurodevelopmental impairment), as well as published literature. Morbidities were considered for inclusion in the composite if they were of particular clinical relevance, such as predictors of long-term neurodevelopmental impairment; this was not dependent on their reported frequency.
Components and Calculation of the Composite Endpoint

The following morbidities were identified as the most appropriate components for a composite endpoint to investigate neonatal benefit of interventions to postpone preterm delivery based on their clinical significance and relationship with early GA: fetal/neonatal death, respiratory distress syndrome requiring continuous positive airway pressure or mechanical ventilation, bronchopulmonary dysplasia (BPD), severe intraventricular hemorrhage (grade 3 or 4), periventricular leukomalacia (with periventricular cyst or with porencephalic cyst or acquired hydrocephaly), retinopathy of prematurity (confirmed as ≥ stage 3 or any stage requiring laser treatment), confirmed sepsis, meningitis, and necrotizing enterocolitis (surgically treated).

To calculate the overall frequency of the composite endpoint, the data for each individual neonate were assessed. For any individual, if any of the morbidities were present, then the score was 1, otherwise the score was 0; the scores from all individuals were summed by week of GA to form a GA-specific count or numerator (by individual week). The denominator population was the simulated trial cohort (Fig. 1) but included neonates who died, as death was an outcome in the composite; denominators were also calculated by week of GA. Deaths were defined as all deaths before hospital discharge, including antepartum and intrapartum deaths.

Composite endpoint frequencies were also stratified by sex and race. This stratification allowed the endpoint to be evaluated against established neonatal morbidity associations.18,19

Statistical Software

All analyses were conducted in SAS version 9.3 (SAS Institute Inc., Cary, NC, United States).

Results

Study Population Characteristics

The overall MUSC population comprised 26,495 women; population attrition from the overall population extracted from the MUSC PINS database and reasons for patient exclusion (as well as details of complications and abnormalities resulting in exclusion from the cohort) are detailed in Fig. 1. From the overall population, 17,912 mother–neonate pairs were included in the simulated trial population (67.6%). Maternal demographics of the simulated trial population are shown in Table 1. Women in the study had a mean age of 26 years (± standard deviation: 6.1 years; range: 13–51 years), and almost half of the population were black (44.7%). Just over half were on government-funded social welfare for low-income individuals (Medicaid/public/medically indigent; 52.7%) and a third (33.8%) had ≥13 years education, equating to a university or other higher-level education. In the simulated trial population, 8.2% of the births were preterm in contrast to 28.5% of the overall population (data not shown), indicative of the high-risk referral status of the hospital.

Neonatal demographic characteristics and composite morbidities of the simulated trial population are shown in Table 2 and Fig. 2A, with rates of additional morbidities in Supplementary Table S1 (online only). As expected, a

Table 1 Characteristics of women in the simulated trial population in the MUSC PINS between 2000 and 2011

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>%a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>17,912</td>
<td></td>
</tr>
<tr>
<td>Maternal age (y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12–18</td>
<td>853</td>
<td>8.6</td>
</tr>
<tr>
<td>19–29</td>
<td>6,286</td>
<td>63.3</td>
</tr>
<tr>
<td>30–39</td>
<td>2,574</td>
<td>25.9</td>
</tr>
<tr>
<td>≥ 40</td>
<td>212</td>
<td>2.1</td>
</tr>
<tr>
<td>Unknown</td>
<td>7,987</td>
<td></td>
</tr>
<tr>
<td>Maternal race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>4,997</td>
<td>27.9</td>
</tr>
<tr>
<td>Black</td>
<td>8,011</td>
<td>44.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4,557</td>
<td>25.4</td>
</tr>
<tr>
<td>Other</td>
<td>347</td>
<td>1.9</td>
</tr>
<tr>
<td>Health insurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial/private</td>
<td>4,083</td>
<td>22.8</td>
</tr>
<tr>
<td>Self-pay</td>
<td>4,393</td>
<td>24.5</td>
</tr>
<tr>
<td>Medicaid/public/medically indigent</td>
<td>9,435</td>
<td>52.7</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 12 y</td>
<td>8,416</td>
<td>66.2</td>
</tr>
<tr>
<td>≥ 13 y</td>
<td>4,298</td>
<td>33.8</td>
</tr>
<tr>
<td>Unknown</td>
<td>5,198</td>
<td></td>
</tr>
<tr>
<td>GA at birth (wk)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24–27</td>
<td>136</td>
<td>0.8</td>
</tr>
<tr>
<td>28–33</td>
<td>445</td>
<td>2.5</td>
</tr>
<tr>
<td>34–36</td>
<td>888</td>
<td>5</td>
</tr>
<tr>
<td>≥ 37</td>
<td>16,443</td>
<td>91.8</td>
</tr>
<tr>
<td>Maternal medications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tocolysis (included MgSO4, terbutaline, indomethacin, nifedipine, ritodrine)</td>
<td>300</td>
<td>51.6</td>
</tr>
<tr>
<td>Antenatal steroids (included betamethasone or dexamethasone)</td>
<td>487</td>
<td>83.8</td>
</tr>
</tbody>
</table>

Abbreviations: GA, gestational age; MUSC, Medical University of South Carolina; PINS, The MUSC Perinatal Information System.

Note: Rounding of percentages of patients may result in characteristic groups summing to >100.

*Percentage calculated only among those with known data. bOne woman with unknown insurance status. In the United States, health insurance is required to pay for medical expenses (commercial/private insurance is employer-funded, self-pay is payment at the point of care, Medicaid/public/medically indigent insurance is government-funded, social welfare for low-income individuals). cPercentage calculated among the 581 women with neonates of <34 weeks GA only (indicated population). dRitodrine was discontinued in the United States in March 2010.
clear relationship was seen with GA for both neonatal ward use and length of hospital stay (birth hospitalization). All extremely preterm babies (24–27 weeks’ GA) required level III neonatal intensive care unit (NICU) treatment, and 25.7% were hospitalized for ≥3 months. Similarly, almost 98% of moderately preterm babies (28–33 weeks’ GA) required level II/III NICU care, and >30% remained in the hospital for ≥1 month.

**Neonatal Morbidities and Composite Endpoint Development**

In total, 3.4% of the population were positive to the composite endpoint. Of those with the composite endpoint, 84.9% were born preterm compared with 5.6% without the composite. The frequency of the composite endpoint and the individual frequencies of neonatal mortality and morbidities by individual week of GA at delivery are shown in – Fig. 2A. The frequency of the composite and all component morbidities decreased steeply with increasing GA, with the highest frequency of morbidity seen before 31 weeks, dropping to very few events in the full-term population (≥37 weeks). For the majority of other key neonatal morbidities, a similar inverse relationship was seen with increasing GA, with most events among the extremely and moderately preterm infants. The exception to this was hypoglycemia, which occurred in approximately 15 to 22% in each of the preterm groups (– Supplementary Table S1, online only).

The frequency of the composite was driven by the frequency of respiratory distress syndrome, BPD, and sepsis. Respiratory distress syndrome was the most frequent morbidity at every GA week, apart from 35 and 36 weeks, where BPD was more common. Other components were rarer, occurring between 3 and 12% of the extremely premature (24–27 weeks) group, and 1 and 3% of the moderately preterm group (28–33 weeks’ GA). When race (black, white) and sex stratification by GA was performed, there was considerable variability in the composite endpoint frequency, particularly between 27 and 36 weeks’ GA in each subgroup; after 37 weeks, the composite frequency was similar between these subgroups (– Fig. 2B).
Composite Endpoint of Neonatal Benefit

Pimenta et al.

Discussion

Current tocolytic agents do not necessarily demonstrate neonatal benefit, despite often successfully prolonging pregnancy. As a result, it is now recommended that preterm labor RCTs should have an additional primary endpoint of neonatal benefit. Despite this, there are no standard endpoints in clinical trials involving women who might deliver preterm neonates. The Core Outcomes for Prevention of Preterm birth project led by the Core Outcomes in Women’s Health initiative aims to obtain expert consensus about the most relevant and clinically meaningful endpoints in this field to facilitate effective synthesis of evidence in RCTs and help develop robust recommendations for clinical practice. These initiatives will be valuable in providing expert clinical opinion regarding the core outcome measures of neonatal benefit in preterm birth; however, it is also important to provide empirical evidence of the frequency of the proposed endpoints in similar populations to that of an RCT, hence the use of real-world data in this study.

Data from this study were used to directly inform sample size planning for the ongoing rotopsiban clinical trials, as the database hospital site closely resembles the study centers for the RCT program. We took into consideration that, because of the stipulated population characteristics, the MUSC rates are likely to represent the upper limit of plausibility of those seen in RCTs. However, this similarity between the populations should maximize the likelihood of appropriate study power to determine neonatal benefit; indeed, an earlier preterm birth intervention study noted that the use of historical data from dissimilar study centers might have been at least partly responsible for negative results. Replication of the frequency of this composite endpoint and individual neonatal morbidities is ongoing using other U.S. and European databases.

In this study, data were descriptively analyzed from a cohort of almost 18,000 mother–neonate pairs identified from the MUSC PINS database, and components of the proposed composite were selected following examination of approximately 100 neonatal morbidities. Neonates were at a risk of a range of severe but predominantly infrequent morbidities, which suggested that it would be appropriate to combine clinically meaningful rare events into a single composite measure, particularly as morbidities rarely occurred singly but were commonly clustered.
The composite represents a set of important comorbidities inversely associated with GA at birth and resulting in a longer length of hospital stay. These morbidities are directly related to recognized health burdens. Most of the included morbidities substantially increase the risk of later neurodevelopmental impairment such as cerebral palsy, developmental delay, and visual/hearing impairments. For example, BPD24 and necrotizing enterocolitis25 are direct risk factors for neurodevelopment impairment, and intraventricular hemorrhage, through its strong association with white matter damage, indirectly increases the risk of neurodevelopment impairment.26

A strong inverse relationship of the composite with increasing GA was seen, and this declining incidence of morbidities is in agreement with previous studies in wider populations.27,28 The largest variability in sex and race, also observed between weeks 30 and 35, is consistent with previous studies.18,19 It is well established that male neonates have a higher mortality rate compared with female neonates,19 which was demonstrated in this study. Moreover, black preterm neonates (>28 weeks’ GA) have the lowest mortality rate compared with white or Hispanic neonates,18 which was also reflected within this composite analysis.

This study has several limitations. The MUSC database draws from a single center based in the southeastern United States, with a predominantly black and lower socioeconomic status population. Marked variation in perinatal and neonatal outcomes have been observed across different regions and countries;29 therefore, the study population used in this analysis may not be broadly generalizable to that of a global RCT. However, the MUSC site does resemble the proposed study centers in the retosibran RCT program (i.e., hospitals with high-level NICU capabilities).15,16 The study period spanned 11 years, over which time many components of neonatal care have improved, which may lead to reduced event frequency in more recently collected data. A recent U.S. study indicated that mortality and rates of several of the included morbidities (sepsis, intraventricular hemorrhage, periventricular leukomalacia, and retinopathy of prematurity) have decreased slightly between 1993 and 2012, though the rates of BPD did not.30 Further limitations of the study include the risk for bias during chart abstraction from medical records into the database; this was minimized by using data abstractors with training in medical terminology and implementation of strict guidelines on priority of data abstraction when multiple sources were available. In addition, data validity was further assessed by having a 10% random sample of charts abstracted by a different neonatologist who was not working in the hospital at the time of original data collection. The study population included all women with singleton, uncomplicated pregnancies but not those with a diagnosis of preterm labor (as would be included in the trial) as this was not captured within the database. Finally, within the clinical trial setting, outcomes will be defined using more specific clinical criteria than were able to be captured within a database, which may influence the frequency of the morbidities (Supplementary Table S2 [online only] for trial versus database definitions).

Neonates in this descriptive analysis of a real-world single-center database cohort were at a risk of several clinically meaningful but predominantly infrequent morbidities. While maintaining clinical relevance, we combined eight of these morbidities with mortality into a composite endpoint. Use of a composite endpoint may improve the statistical efficiency of RCTs for tocotropics required to demonstrate neonatal benefit, and the empirical data generated were used to directly inform sample size calculations for such trials.

Conflicts of Interest
At the time this work was performed, J.M.P., T.H.M., K.J.B., M.P., and J.A. were full-time employees of GSK R&D, and all owned stocks and shares in the company at the time of manuscript submission. J.A. has since left the company. M.E. is an employee of the MUSC and was contracted and paid by GSK to conduct the study but received no compensation for her role as an author. T.C.H. was an employee of the MUSC at the time of the study and was contracted by GSK to conduct the study, for which the MUSC received payment, but T.C.H. and the MUSC received no compensation for his role as an author. T.C. H. has since left the MUSC. T.M.O. received payment for consultancy from GSK.

Funding
This study was funded by GSK R&D (study number WEUSRTP4174/EPI40607).

Acknowledgments
The authors would like to thank their consulting neonatologist Ricki Goldstein, MD, Duke University, for her clinical input; the wider GSK clinical team, including Jerry Snidow, PharmD, Feng Liu, PhD, Yusuke Komatsu, MD, PhD, and Libby Black, PharmD, for interpretation and review. Editorial support during the development of this manuscript was provided by Gillian Groeger, PhD, from Fishawack Indicia, United Kingdom, and funded by GSK.

References
Composite Endpoint of Neonatal Benefit

Pimenta et al.


