1-1-2018

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

Jeanne M. Pimenta
Real World Evidence (Epidemiology), GSK, Uxbridge, Middlesex, United Kingdom

Myla Ebeling
Medical University of South Carolina

Timothy H. Montague
Clinical Statistics, GSK, Philadelphia, Pennsylvania

Kathleen J. Beach
Maternal Neonatal Health Unit, GSK, Research Triangle Park, North Carolina

Jill Abell
Clinical Effectiveness and Safety Clinical Evaluation Science, GSK, Philadelphia, Pennsylvania

See next page for additional authors

Follow this and additional works at: https://researchrepository.wvu.edu/ctsi

Part of the Medicine and Health Sciences Commons

Authors

Jeanne M. Pimenta, Myla Ebeling, Timothy H. Montague, Kathleen J. Beach, Jill Abell, Michael T. O'Shea, Marcy Powell, and Thomas C. Hulsey

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

Jeanne M. Pimenta, PhD[1]   Myla Ebeling, BS[2]   Timothy H. Montague, PhD[3]   Kathleen J. Beach, MD[4]
Jill Abell, PhD[5,6]   Michael T. O'Shea, MD[7]   Marcy Powell, MD, FACOG[8]   Thomas C. Hulsey, ScD[2,9]

[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 $\geq$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

In 2010, 11% of live births worldwide were preterm (occurring before 37 weeks of gestation).[1] Variation in estimated preterm birth rates was seen across countries, from approximately 5% in Europe to 18% in Africa.[1] 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.[2] Preterm birth is the most common direct cause of neonatal death and indirectly affects infant mortality rates through associated morbidities.[3–5] Consequences for the neonate can persist into later life, with sequelae such as visual and hearing impairment, lung disease, cardiovascular disorders, and neurodevelopmental impairments.[3,6–9] 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).[7]

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.[10] 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,[11] but there is very limited evidence of a measurable benefit of tocolytics on neonatal outcomes.[7,12] 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.[13] 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.[13] 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.[14]

Clinical development of new tocolytics (such as retosiban currently in development by GlaxoSmithKline[15,16]) 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%.[17] 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 $24^{0/7}$ and $33^{6/7}$ 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 ►Fig. 1 for full list of maternal complications used as exclusion criteria). Neonates with prespecified severe or life-threatening congenital conditions were also excluded (►Fig. 1). 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
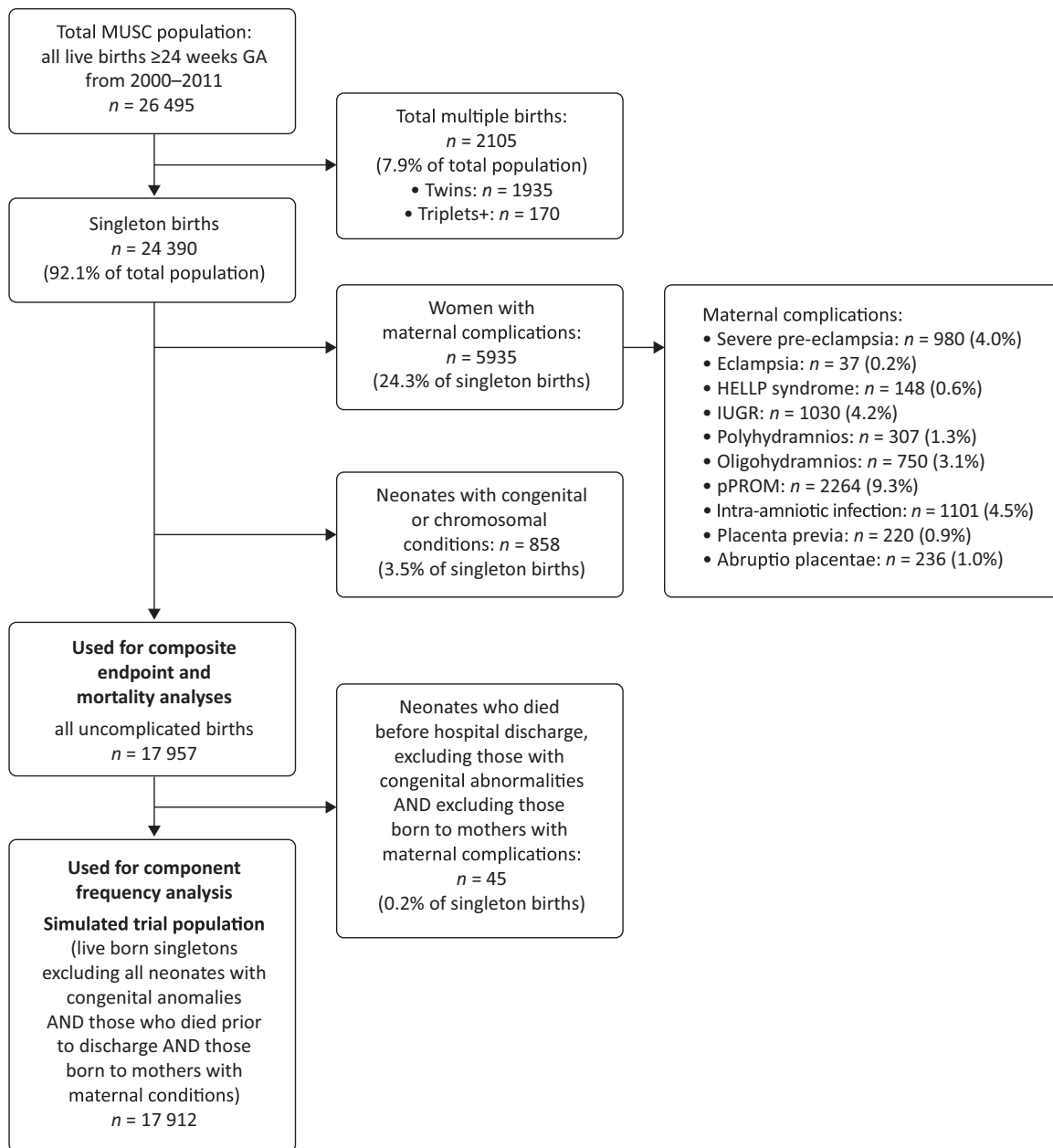
Fig. 1 Patient attrition from the complete MUSC PINS database population to the simulated trial population for analyses. Maternal complications (as listed in the correspondingly labeled box) were not mutually exclusive and therefore women could have more than one complication. Congenital abnormalities were achondrogenesis type II, anencephaly, arteriovenous malformation, autosomal recessive polycystic kidney disease, bilateral dysplastic kidneys, bilateral multicystic kidneys with renal failure, bilateral renal agenesis, congenital hydrocephaly, congenital renal abnormality/agenesis, diaphragmatic hernia, duodenal atresia, exencephaly, exstrophy of bladder, gastroschisis, genitourinary cystic/dysplastic obstruction, harlequin fetus, hydronephrosis, hydrops fetalis (immune and nonimmune), hypoplastic left heart, imperforate anus, inborn error of metabolism, iniencephaly, Meckel–Gruber's syndrome, mixed dysplastic/cystic/agenic disorders with kidney failure, myelomeningocele, esophageal atresia, omphalocele, osteogenesis imperfecta type II, Potter's syndrome, tetralogy of Fallot, total anomalous pulmonary venous return, tracheoesophageal fistula, transposition of the great arteries, trisomy 13, trisomy 18, and trisomy 21. GA, gestational age; HELLP syndrome, hemolysis, elevated liver enzymes, low platelet count; IUGR, intrauterine growth restriction (defined as fetal weight below the 10th percentile for GA); MUSC, Medical University of South Carolina; PINS, Perinatal Information System; pPROM, preterm prelabor rupture of membranes.

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

Characteristic	N	%[a]
Overall	N = 17,912	
Maternal age (y)[a]		
12–18	853	8.6
19–29	6,286	63.3
30–39	2,574	25.9
≥ 40	212	2.1
Unknown	7,987	
Maternal race		
White	4,997	27.9
Black	8,011	44.7
Hispanic	4,557	25.4
Other	347	1.9
Health insurance[a–c]		
Commercial/private	4,083	22.8
Self-pay	4,393	24.5
Medicaid/public/medically indigent	9,435	52.7
Education level[a]		
≤ 12 y	8,416	66.2
≥ 13 y	4,298	33.8
Unknown	5,198	
GA at birth (wk)		
24–27	136	0.8
28–33	445	2.5
34–36	888	5
≥ 37	16,443	91.8
Maternal medications[d]		
Tocolysis (included MgSO$_4$, terbutaline, indomethacin, nifedipine, ritodrine[e])	300	51.6
Antenatal steroids (included betamethasone or dexamethasone)	487	83.8

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.
[a]Percentage calculated only among those with known data. [b]One woman with unknown insurance status. [c]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). [d]Percentage calculated among the 581 women with neonates of <34 weeks' GA only (indicated population). [e]Ritodrine was discontinued in the United States in March 2010.

Table 2 Neonatal characteristics in the simulated trial population

	GA at birth (wk)[a]				
Neonatal variable	24–27 (n = 136)	28–33 (n = 445)	34–36 (n = 888)	≥37 (n = 16,443)	Total (N = 17,912)
Neonatal sex, n (%)					
Male	79 (58.1)	232 (52.1)	465 (52.4)	8,353 (50.8)	9,129 (51)
Female	57 (41.9)	213 (47.9)	423 (47.6)	8,090 (49.2)	8,783 (49)
Neonatal race, n (%)					
White	46 (33.8)	147 (33)	264 (29.7)	4,510 (27.4)	4,967 (27.7)
Black	82 (60.3)	268 (60.2)	470 (52.9)	7,196 (43.8)	8,016 (44.8)
Hispanic	8 (5.9)	27 (6.1)	146 (16.4)	4,385 (26.7)	4,566 (25.5)
Other	0	3 (0.7)	8 (0.9)	352 (2.1)	363 (2)
Highest level neonatal ward admission during hospital stay, n (%)					
Normal nursery (level I)	0	10 (2.2)	536 (60.4)	15,141 (92.1)	15,687 (87.6)
NICU level II[b]	0	149 (33.5)	246 (27.7)	938 (5.7)	1,333 (7.4)
NICU level III[c]	136 (100)	286 (64.3)	106 (11.9)	359 (2.2)	887 (5)
Length of hospital stay, n (%)					
< 3 d	0	8 (1.8)	443 (49.9)	12,899 (78.4)	13,350 (74.5)
3 d to 1 wk	1 (0.7)[d]	40 (9)	331 (37.3)	3,290 (20)	3,662 (20.4)
1 wk to 1 mo	0	245 (55.1)	106 (11.9)	212 (1.3)	563 (3.1)
1–2 mo	42 (30.9)	128 (28.8)	6 (0.7)	31 (0.2)	207 (1.2)
2–3 mo	58 (42.6)	20 (4.5)	0	7 (0)	85 (0.5)
≥ 3 mo	35 (25.7)	4 (0.9)	2 (0.2)	4 (0)	45 (0.3)

Abbreviations: GA, gestational age; NICU, neonatal intensive care unit.
[a]GA was analyzed by week but is summarized here by range for clarity. [b]NICU level II is a special care nursery providing care for neonates born at ≥32 weeks' GA and weighing ≥ 1,500 g who are physiologically immature or are moderately ill with problems that are expected to resolve rapidly without urgent subspecialty services. [c]NICU level III can provide sustained life support, provide comprehensive care for neonates born at <32 weeks' GA and weighing <1,500 g or neonates of any GA/birth weight with critical illness, give ready access to pediatric specialists including surgeons, can provide respiratory support including ventilation and inhaled nitric oxide, and perform emergency advanced imaging. [d]Neonate transferred to another hospital for continued care at day 7.

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**). While the
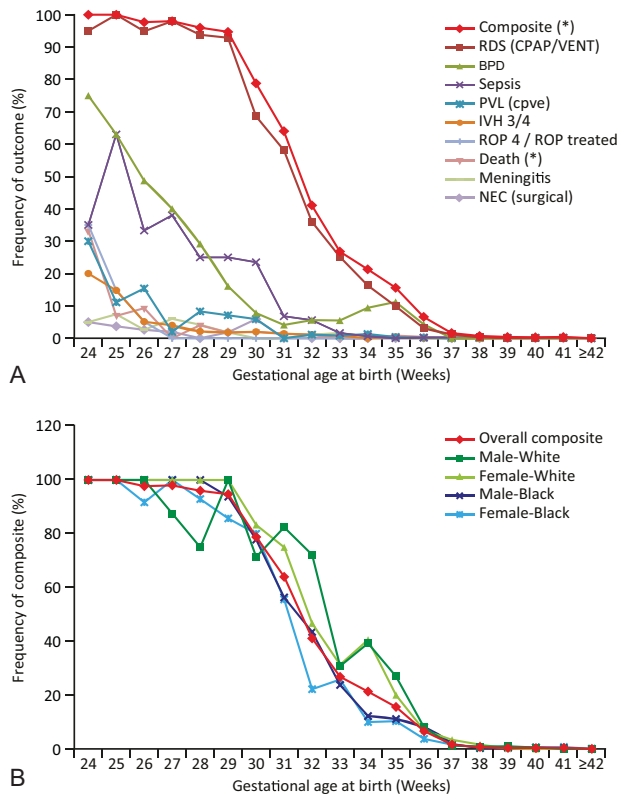
Fig. 3 Median length of stay in neonates with and without the composite. *Composite endpoint excludes death.

Fig. 2 (**A**) Frequency of individual neonatal morbidities and mortality in the composite. (**B**) Race/sex stratification of the composite in white and black neonates. BPD, bronchopulmonary dysplasia; CPAP, continuous positive airway pressure; CPVE, cysts or porencephaly and ventricular enlargement; IVH, intraventricular hemorrhage; NEC, necrotizing enterocolitis; PVL, periventricular leukomalacia; RDS, respiratory distress syndrome; ROP, retinopathy of prematurity; VENT, ventilation. *Frequency of composite endpoint and deaths calculated in 17,957 births following uncomplicated pregnancies; frequency of morbidities calculated in live uncomplicated births only (n = 17,912).

smaller sample sizes in the subgroups may partly explain the variability between GA groups, overall trends can be seen. The frequency of the composite in black neonates of either sex tended to be below the rate seen for the overall population, whereas white neonates had composite rates slightly higher than those in the overall population, though this was not statistically determined.

Median length of stay of the birth hospitalization was consistently longer at each week of GA among those with the composite endpoint compared with those without (►**Fig. 3**). It fell sharply from 102 days (interquartile range [IQR]: 88–116) at 24 weeks' GA to 6 to 8 days at 36 to 38 weeks' GA, after which a slight upturn was seen. The composite was not present in only seven neonates before 30 weeks' GA; median length of stay was 29 days (IQR: 21–34) at 30 weeks' GA, dropping to 2 days (IQR: 1–2) at 35 weeks and remaining at this level thereafter.

Discussion

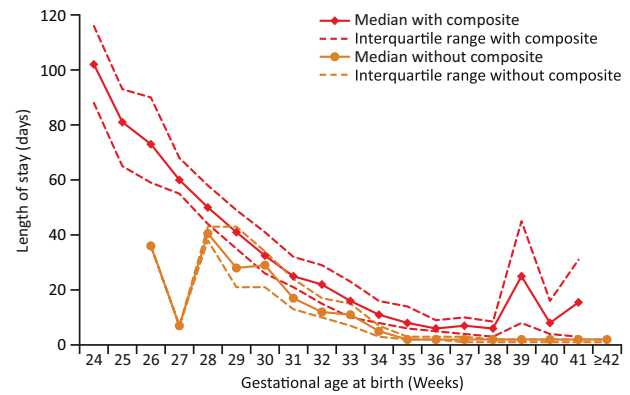Current tocolytic agents do not necessarily demonstrate neonatal benefit, despite often successfully prolonging preg-

nancy.[11] As a result, it is now recommended that preterm labor RCTs should have an additional primary endpoint of neonatal benefit.[20] Despite this, there are no standard endpoints in clinical trials involving women who might deliver preterm neonates.[13,17,21] 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.[22] 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 retosiban 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.[23] 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 neurodevelopment impairment such as cerebral palsy, developmental delay, and visual/hearing impairments. For example, BPD[24] and necrotizing enterocolitis[25] 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 retosiban 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 tocolytics required to demonstrate neonatal benefit, and the empirical data generated were used to directly inform sample size calculations for such trials.

References
1 Blencowe H, Cousens S, Oestergaard MZ, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 2012;379(9832):2162–2172
2 Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008;371(9606):75–84
3 Blencowe H, Cousens S, Chou D, et al; Born Too Soon Preterm Birth Action Group. Born too soon: the global epidemiology of 15 million preterm births. Reprod Health 2013;10(Suppl 1):S2
4 Lawn JE, Cousens S, Zupan J; Lancet Neonatal Survival Steering Team. 4 million neonatal deaths: When? Where? Why?. Lancet 2005;365(9462):891–900
5 Liu L, Johnson HL, Cousens S, et al; Child Health Epidemiology Reference Group of WHO and UNICEF. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 2012;379 (9832):2151–2161
6 Greenough A. Long-term respiratory consequences of premature birth at less than 32 weeks of gestation. Early Hum Dev 2013;89 (Suppl 2):S25–S27

7   Institute of Medicine Committee on Understanding Premature Birth. Assuring Healthy Outcomes. The National Academies Collection: Reports funded by National Institutes of Health. In: Behrman RE, Butler AS, eds. Preterm Birth: Causes, Consequences, and Prevention. Washington, DC: National Academies Press/National Academy of Sciences; 2007

8   Marlow N, Hennessy EM, Bracewell MA, Wolke D; EPICure Study Group. Motor and executive function at 6 years of age after extremely preterm birth. Pediatrics 2007;120(04):793–804

9   Mwaniki MK, Atieno M, Lawn JE, Newton CR. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. Lancet 2012;379(9814):445–452

10 World Health Organization. Every Newborn: an action plan to end preventable deaths. 2014. https://www.everynewborn.org/Documents/Full-action-plan-EN.pdf. Accessed October 19, 2016

11 National Institute for Health and Care Excellence (NICE). Preterm labour and birth. NICE guideline [NG25]. Published 20 November 2015 www.nice.org.uk/guidance/ng25. Accessed February 16, 2018

12 Smith GN, Walker MC, Ohlsson A, O'Brien K, Windrim R; Canadian Preterm Labour Nitroglycerin Trial Group. Randomized double-blind placebo-controlled trial of transdermal nitroglycerin for preterm labor. Am J Obstet Gynecol 2007;196(01):37.e1–37.e8

13 Meher S, Alfirevic Z. Choice of primary outcomes in randomised trials and systematic reviews evaluating interventions for preterm birth prevention: a systematic review. BJOG 2014;121(10): 1188–1194, discussion 1195–1196

14 European Network for Health Technology Assessment. Endpoints used for relative effectiveness assessment: composite endpoints 2015. http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/Endpoints%20used%20for%20Relative%20Effectiveness%20Assessment%20Composite%20endpoints_Amended%20JA1%20Guideline_Final%20Nov%202015_0.pdf. Accessed October 18, 2016

15 GSK. NCT02292771: a randomized study comparing the efficacy and safety of retosiban versus atosiban for women in spontaneous preterm labour. 2015. https://www.clinicaltrials.gov/ct2/results?term=NCT02292771. Accessed October 19, 2016

16 GSK. NCT02377466: a phase III efficacy and safety study of intravenous retosiban versus placebo for women in spontaneous preterm labor. 2015. https://www.clinicaltrials.gov/ct2/results?term=NCT02377466. Accessed October 19, 2016

17 Mercer BM, Rabello YA, Thurnau GR, et al; NICHD-MFMU Network. The NICHD-MFMU antibiotic treatment of preterm PROM study: impact of initial amniotic fluid volume on pregnancy outcome. Am J Obstet Gynecol 2006;194(02):438–445

18 Alexander GR, Kogan M, Bader D, Carlo W, Allen M, Mor J. US birth weight/gestational age-specific neonatal mortality: 1995-1997 rates for whites, Hispanics, and blacks. Pediatrics 2003;111(01): e61–e66

19 Matthews TJ, MacDorman MF, Thoma ME. Infant mortality statistics from the 2013 period linked birth/infant death data set. Natl Vital Stat Rep 2015;64(09):1–30

20 US Food and Drug Administration. Advisory Committee for Reproductive Health Drugs. NDA 22–139 Progesterone gel (8%). 2012. http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/reproductivehealthdrugsadvisorycommittee/ucm287987.pdf. Accessed October 19, 2016

21 Zephyrin LC, Hong KN, Wapner RJ, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal–Fetal Medicine Units (MFMU) Network. Gestational age-specific risks vs benefits of multicourse antenatal corticosteroids for preterm labor. Am J Obstet Gynecol 2013;209(04):330.e1–330.e7

22 van 't Hooft J, Duffy JM, Daly M, et al; Global Obstetrics Network (GONet). A core outcome set for evaluation of interventions to prevent preterm birth. Obstet Gynecol 2016;127 (01):49–58

23 Combs CA, Garite T, Maurel K, Das A, Porto M; Obstetrix Collaborative Research Network. 17-hydroxyprogesterone caproate for twin pregnancy: a double-blind, randomized clinical trial. Am J Obstet Gynecol 2011;204(03):221.e1–221.e8

24 Short EJ, Kirchner HL, Asaad GR, et al. Developmental sequelae in preterm infants having a diagnosis of bronchopulmonary dysplasia: analysis using a severity-based classification system. Arch Pediatr Adolesc Med 2007;161(11):1082–1087

25 Hintz SR, Kendrick DE, Stoll BJ, et al; NICHD Neonatal Research Network. Neurodevelopmental and growth outcomes of extremely low birth weight infants after necrotizing enterocolitis. Pediatrics 2005;115(03):696–703

26 O'Shea TM, Allred EN, Kuban KC, et al; ELGAN Study Investigators. Intraventricular hemorrhage and developmental outcomes at 24 months of age in extremely preterm infants. J Child Neurol 2012;27(01):22–29

27 Rijken M, Stoelhorst GM, Martens SE, et al. Mortality and neurologic, mental, and psychomotor development at 2 years in infants born less than 27 weeks' gestation: the Leiden follow-up project on prematurity. Pediatrics 2003;112(02):351–358

28 Shapiro-Mendoza CK, Tomashek KM, Kotelchuck M, et al. Effect of late-preterm birth and maternal medical conditions on newborn morbidity risk. Pediatrics 2008;121(02):e223–e232

29 Shah PS, Lui K, Sjörs G, et al; International Network for Evaluating Outcomes (iNeo) of Neonates. Neonatal outcomes of very low birth weight and very preterm neonates: an international comparison. J Pediatr 2016;177:144–152.e6

30 Stoll BJ, Hansen NI, Bell EF, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993–2012. JAMA 2015; 314(10):1039–1051