Statistical Analysis Plan for Serum Assessment of Preterm Birth-- Outcomes Compared to Historical Controls: AVERT PRETERM TRIAL

ClinicalTrials.gov number: NCT03151330

Lead Study Investigator(s): Matthew Kendall Hoffman, MD, MPH-ChristianaCare

Funding: Sera Prognostics, Inc., ChristianaCare Health Services.

Title:	Serum Assessment of Preterm Birth: Outcomes Compared to Historical Controls: AVERT PRETERM TRIAL
Study Design:	Prospective cohort study of screened women compared to a historical control.
Study Objective:	To examine whether the women who are screened using the PreTRM® test around 20 weeks gestation will have a statistically significant reduction in adverse pregnancy outcomes.
Prepared by:	Zugui Zhang, PhD Matthew Kendall Hoffman, MD, MPH

TABLE OF CONTENTS

1 ACRONYMS and ABBREVIATIONS

ACOG	American College of Gynecologists
AE	Adverse Event
AFI	Amniotic fluid index
APS	Antiphospholipid syndrome
ASA	Acetylsalicylic acid
AUC	Area under the curve
BPD	Bronchopulmonary dysplasia
BPP	Biophysical profile
CC	Cubic Centimeter
CCHS	ChristianaCare Health System
CL	Cervical Length
CLASI	Community Legal Aid Society, Inc
CLASP	Collaborative Low-dose Aspirin Study in Pregnancy
CRF	Case report form
CRL	Crown rump length
DCC	Data Coordinating Center
dL	Deciliter
DMS	Data Management System
EAGER	Effects of Aspirin in Gestation and Reproduction Study
EDD	Estimated due date (or estimated date of delivery)
FDA	U.S. Food and Drug Administration
g	Gram
GA	Gestational age
GAB	Gestational age at birth
GC	Gonococcus
ICER	Incremental Cost Effectiveness Ration
IRB	Institutional Review Board
IBP4	Insulin-like growth factor binding protein 4
ITT	Intention to treat
IVH	Intraventricular hemorrhage
LBW	Low birth weight
LC-MS/MS	Liquid chromatography –tandem mass spectrometry
LDA	Low dose aspirin
LMP	Last menstrual period
mg	Milligram
miPTB	Medically-indicated preterm birth
mm	Millimeter
NEC	Necrotizing enterocolitis
NICHD	Eunice Kennedy Shriver National Institute of Child Health and Human Development
NICU	Neonatal Intensive Care Unit
NIH	National Institutes of Health
NMI	Neonatal morbidity and mortality

NNOLOS	Length of neonatal hospital stay (NNOLOS)
NICULOS	Length of neonatal NICU stay (NICULOS)
OHRP	Office of Human Research Protections
P	Probability Value
PECEP	Pesario Cervical para Evitar Prematuridad
PTL	Preterm Labor
PROM	Premature rupture of membranes
PPROM	Preterm premature rupture of membranes
PTB	Preterm birth
PVL	Periventricular leukomalacia
RCT	Randomized Clinical Trial
RDS	Respiratory distress syndrome
RR	Relative Risk
SAE	Serious adverse event
SAB	Spontaneous Abortion
SCF	Specimen Collection Form
SD	Standardized Deviation
SEAR	Statistical analysis plan for health economic analysis of AVERT PTB
SGA	Small for gestational age
SHBG	Sex hormone binding globulin
SOP	Standardized Operating Procedure
SPTB	Spontaneous preterm birth
TOPS	a randomized Trial of Pessary in Singletons
TVCL	Trans-vaginal cervical length
TVUS	Trans-vaginal ultrasound
UAE	Unexpected Adverse Event
UTI	Urinary Tract Infection
VLBW	Very low birth weight
WHO	World Health Organization
17-OHPC	17-hydroxy progesterone caproate

2. Abstract

Executive Summary:

AVERT PRETERM is a historically controlled trial to evaluate the safety and efficacy of a preterm birth (PTB) prevention strategy versus standard-of-care pregnancy management in reduction of the incidence of adverse pregnancy outcomes. Subjects in the prospective arm receive a commercially-available laboratory developed test, PreTRM® (Sera Prognostics, Inc). Those with higher-risk results are offered a multimodal intervention protocol; the remainder receive routine pregnancy management. Outcomes are compared between the prospective arm and historical controls.

The AVERT study addresses an enormous obstetrical problem having a large unmet need with the use of a novel risk assessment tool (PreTRM®) and established multimodal interventions administered within the single ChristianaCare hospital network. However, there are several potential limitations to the AVERT study. Comparisons to historical controls can be susceptible to selection bias and epoch effects. Subjects who don't accept interventions may add self-selection bias. As well, AVERT is subject to a research halt due to the spread of SARS-CoV-2, avoiding a pandemic epoch effect while restricting study size.

This statistical analysis plan addresses the study objective of evaluation of the PTB prevention strategy while overcoming potential study limitations. The original endpoints are replaced with more powerful endpoints aligned to the study size and faithful to the study's goal of reducing adverse outcomes. The primary analysis is in a modified ITT population enabling comparison of the fully consented PTB prevention strategy population to a population who received standard care. To adjust for any bias in consent to intervention, exploratory analyses compare subjects accepting and not accepting interventions, and address the efficacy of individual intervention components.

An estimated 1,453 prospective subjects reached a gestational age of $37^{0/7}$ weeks before local spread of SARS-CoV-2 in Delaware and the associated shut-down of non-COVID-19 research at ChristianaCare. Approximately 10,000 historical controls were selected

from a period of about 2 years immediately prior to study initiation. The study design shows at least 86% power for 5% alpha shared across two co-primary endpoints: neonatal morbidity and mortality index and neonatal length of total hospital stay.

The funder, Sera Prognostics, will remain blinded to all clinical data until a formal communication of topline results from this analysis is made.

Background: Preterm birth (PTB) remains the leading cause of neonatal mortality and long term disability throughout the world. Recently treatments early in pregnancy such as progesterone, cervical support and maternal support have been demonstrated to delay delivery amongst at risk women. Nonetheless, the majority of women who are at risk are not identified using current screening modalities.

Hypothesis: A cohort of pregnancies who are screened using the PreTRM® test around 20 weeks gestation in which a bundle of interventions is given for elevated PreTRM® risk will show either decreased neonatal morbidity/and mortality (measured as a composite score, "NMI"), or decreased length of neonatal stay in the hospital (NNOLOS). Secondarily, they will show an increase in gestational age at birth (GAB) and a reduction in length of neonatal NICU stay (NICULOS), compared to an unscreened historical control group.

Study Design Type: Prospective cohort study of screened pregnant women compared to a historical control.

Population: Prospective subjects are recruited from ChristianaCare's patient population as described in the study protocol. A comparable control population will be identified from ChristianaCare's database, as described in the study protocol. A total of approximately 10,000 women meeting the screening protocol will be identified with delivery prior to July 6th, 2018. Women will be sequentially identified retrospectively who qualify for the study based on applicable inclusion and exclusion criteria. As such cohort definition dates can only be estimated but are noted with a delivery of volume of approximately 6,000 deliveries a year, this would approximate a start date for the historical cohort of August 6th, 2016.

Intervention: Qualifying women will be screened using the PreTRM® test (Sera Prognostics, Inc.) at a large tertiary care center. Predicated upon the degree of risk based on a prespecified algorithm that includes height, weight, and measurement of two serum proteins (IBP4 and SHBG), women will be treated according to a pre-determined protocol. The outcomes of the screened cohort of women will be compared to those of the historical control group at the same tertiary care center.

Outcomes:

***Co-Primary outcomes*:** To determine whether a cohort of women who are screened with the PreTRM® test and then managed according to a prespecified protocol will have statistically significant reductions in either (a) composite neonatal morbidity and mortality (NMI score), or (b) length of neonatal hospital stay (NNOLOS), compared to a historical control group.

Secondary outcomes: To determine whether a cohort of women who are screened with the PreTRM® test and then managed according to a pre-specified treatment protocol will have statistically significant reduction in (a) length of NICU hospital stay for critical care of neonates from time of birth up to neonatal discharge (NICULOS), or (b) a statistically significant increase in duration of gestation compared to historical controls.

3 Statement of Problem and Study

3.1 Background and Rationale

Preterm delivery, defined as delivery prior to $37^{0/7}$ weeks gestation, remains the dominant contributor to neonatal morbidity and mortality throughout the world (March of Dimes Save the Children, WHO. 2012; Mathews, Menacker, and MacDorman 2004; Anderson and Smith 2003), including up to 50% of pediatric neurodevelopmental disorders (Goldenberg and Rouse 1998). Infants born prematurely are at increased risk for a variety of long term medical complications such as respiratory, gastrointestinal, cardiovascular, and metabolic disorders (McCormick & Richardson, 2002; Saigal & Doyle, 2008). Given the tremendous medical, financial and emotional burden of preterm birth, interventions that increase the GAB of premature infants, even marginally, would profoundly impact these children, their families and the health care system.

3.2 Risk Identification

To date, our ability to identify women at risk for PTB have been limited to looking for three specific clinical circumstances in a pregnant woman: 1) Non-specific socio-economic factors (e.g. African-American race, lower socio-economic status, low BMI), 2) A history of a prior preterm birth, or 3) a short cervix on transvaginal sonography. Each of these characteristics have severe limitations as screening strategies and do not allow for the meaningful identification of the majority of at-risk pregnancies. Socioeconomic risk factors are not highly specific for preterm birth and have only demonstrated a sensitivity of 32% in a general population (Mercer, 1996). The strongest known predictor thus far for predicting preterm birth is having a prior preterm birth. As a sole risk factor in singleton pregnancy, its predictive value is limited by the fact that 40% of pregnancies occur in nulliparous women and only a minority of multiparous women will have this as a risk factor, resulting in a detection sensitivity of only 11% (Petrini, 2005). Universal cervical length screening has demonstrated to be a reasonably specific marker, though shortening of the cervix only occurs in approximately 1%-2% of the general pregnant population with at most a 6% sensitivity increase (Werner, Hamel,Orzechowski, Berghella, & Thung, 2015)(Orzechowski et al. 2014; Son et al. 2016) and has shown an AUC of only 0.61 when used in a nulliparous population at mid-trimester (Esplin et al. 2017). If meaningful change is to occur in reducing the effects of preterm birth, a more sensitive and specific assay needs to be developed and to be implemented at a point early enough in pregnancy that a rational treatment strategy can be invoked.

3.3 Study Design

We propose a prospective study where pregnant women are screened between $19^{1/7}$- $20^{6/7}$ wks gestation using the PreTRM® test (Sera Prognostics, Inc.). A predetermined risk of sPTB ≥ 16% is set as the lower limit of what constitutes elevated risk that would initiate the pre-set group of treatments as defined by a prespecified protocol. The outcomes of this cohort will be compared to a historical control at the same tertiary care center at a time in which PreTRM® testing was not available. The funder, Sera Prognostics, will

remain blinded to all clinical data until a formal communication of topline results from this analysis is made.

Additionally, a cumulative fiscal analysis of both cohorts will be covered by a separate SAP (SEAR).

The study design is illustrated in **Figure 1**.

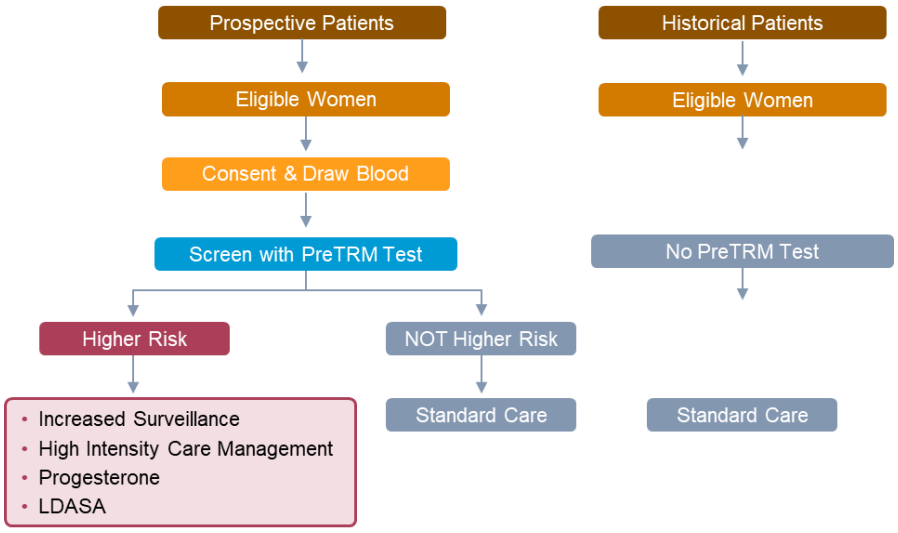

Figure 1. Design of the AVERT study

3.4 Study Population

We intend to study the effects of serum screening on women carrying a live singleton pregnancy between $19^{1/7}$ to $20^{6/7}$ weeks gestation. Inclusion criteria and exclusion criteria were shown in the study protocol.

Prospective Tested Group: as described in the prospective cohort study design. Briefly, prospective subjects will be enrolled during a routine prenatal visit during $19^{1/7}$-$20^{6/7}$. Prior to screening a potential subject, the Principal Investigator (PI) will have obtained written IRB approval of the informed consent form (ICF), and other related information. Eligibility criteria are given in the protocol.

Historical Control Group: The Historical Control Group of women was defined by inclusion and exclusion criteria similar to those which defined the prospective cohort. Women who were transferred from other facilities would be excluded due to lack of records. As an exception to exclusion criteria used in the prospective cohort, women who did not meet eligibility criteria related to assay requirements were not excluded from the historical control group as these criteria are only anticipated to affect the PreTRM®test.

3.5 Primary Hypothesis or Question

Co-primary hypotheses are that women who are screened with the PreTRM® test between $19^{1/7}$ weeks and $20^{6/7}$ weeks GA will have 1) a significant reduction in severe composite neonatal morbidity and mortality (NMI), or 2) a significant reduction in length of neonatal hospital stay (NNOLOS) in comparison to historical controls.

3.6 Secondary Hypothesis or Questions

Women who were screened with the PreTRM® test will have 1) a significant reduction in length of NICU hospital stay for neonates from time of birth up to neonatal discharge or 2) a significant increase in duration of gestation (GAB) in comparison to historical controls.

3.7 Definitions of Outcomes

Primary and secondary endpoints, including adverse events, will be collected, measured, calculated, and pre-defined according to the study design.

- Composite neonatal morbidity and mortality (NMI): neonatal morbidity and mortality (14). Neonates are assigned points according to the following criteria. 0 to 4 scale with NICU

  0 = no events,

  1 = one event for (RDS, BPD, grade III or IV IVH, PVL, proven sepsis, or NEC) or <5 days in the NICU, and no perinatal mortality;

  2 = two events or between 5 and 20 days in the NICU, and no perinatal mortality;

3 = three or more events or >20 days in the NICU, and no perinatal mortality;

4 = perinatal mortality

The component condition abbreviations are defined as follows:

RDS: respiratory distress syndrome

BPD: bronchopulmonary dysplasia

IVH: intraventricular hemorrhage

NEC: necrotizing enterocolitis

PVL: periventricular leukomalacia

Component conditions are further defined in Appendix 1.

- Length of neonatal hospital stay (NNOLOS): days from delivery to neonatal discharge home, or until death if occurring before discharge. This includes all levels of care before discharge home, including all transfers across levels of care. NNOLOS will be calculated in fractional days if times of birth and discharge are available and rounded to the nearest integer, or as the number of days of difference between dates if times are not available; times are not required. In cases of stillbirth or neonatal mortality prior to discharge home, NNOLOS is set to one day longer than the maximum length of stay recorded for any infant. NNOLOS will be truncated at $44^{0/7}$ weeks of gestational age as specified in the protocol.
- Length of neonatal NICU stay (NICULOS): days from admission to neonatal intensive care (NICU) to neonatal discharge home from the NICU or until death if occurring before discharge. This includes all levels of care designated by ChristianaCare as intensive care, including all transfers across levels of intensive care. Nursery and non-intensive care days are not included. This will be calculated in fractional days if times of admission and discharge are available and rounded to the nearest integer value of total days, or as the number of total days in intensive care if times are not available; times are not required. NICULOS is set to zero for babies never admitted to the NICU but admitted to the hospital. In cases of stillbirth or neonatal mortality in the NICU, NICULOS is set to one day longer than the maximum length of stay recorded for any infant. NICULOS will be truncated at $44^{0/7}$ weeks of gestational age as specified in the protocol.

- Duration of gestation: GAB is set to $40^{0/7}$ weeks for deliveries on the estimated due date (EDD) and adjusted by the difference in days between the date of delivery and the EDD for all other births. The EDD is established by ultrasound data or accurately known last menstrual period (LMP), according to ACOG guidelines, as described in the protocol.

Safety outcomes: monitoring of adverse events is described in the protocol.

4. Determination of Sample Size

This study has been terminated due to the advent of the SARS-CoV-2 pandemic in the ChristianaCare population. At study termination, 1,873 subjects were enrolled in the study. To avoid bias in comparison of pre-pandemic controls to prospective subjects who reached term during the pandemic period, the primary analysis of the study will be limited to an estimated 1,453 subjects who reached $37^{0/7}$ weeks gestation before local spread of SARS-CoV-2 in Delaware and the associated shut-down of non-COVID-19 research at ChristianaCare.. Approximately 10,000 historical controls were selected from a period of about 2 years immediately prior to study initiation.

From historical data on pregnancies at ChristianaCare Health System, we estimated a historical preterm birth rate of 9.1%. The sample size estimation is based on simulations of co-primary outcomes based on a simulated distribution of gestational age with a singleton preterm birth rate of 9.1%, and an effect of intervention based on literature data[26-28], with α-level spending of 0.05 shared between co-primaries using Holm's method.

Given shut-down of the study due to the SARS-CoV-2 pandemic, the sample size for the study is estimated to be approximately 1,453 subjects in the prospective arm, and approximately 10,000 in the historical control group. In the primary and alternative analyses, more than 10,000 subjects may meet the eligibility criteria for the historical control group and have the required data, and if so, they will be included in the study.

Power and sample size for this study are affected by factors including the proportion of prospective subjects receiving PreTRM® risk scores at or above the threshold of 16% risk (the study Screen Positive Rate (SPR)), the proportion of higher-risk subjects complying with interventions (compliance), and the rate at which subjects and their neonates remain in the study (retention). For the purpose of power and sample size estimation and based on study monitoring, we estimate that 34% of subjects in the pre-pandemic portion of the prospective arm had PreTRM® risk scores at or above the threshold of 16% risk (the SPR).

Power and sample size are also affected by the study population's baseline level of risk of adverse neonatal outcomes related to prematurity, the sensitivity of testing in identifying at-risk pregnancies, retention of subjects in the study and the efficacy of the interventions as implemented in the study. Literature[26-28] indicates that the effect of treatment is not constant across different values of GAB. Thus, there is no single value of the effect size in terms of proportion of subjects with NMI less than 3, or mean length of stay. Rather, the effect size is larger for subjects with earlier GAB, while the effect size is reduced for subjects with later GAB.

For these reasons, power and sample size were estimated based on simulations, using a range of possible values for the factors noted above. These simulations used the observed distribution of GAB from previous studies, adjusted to match the expected overall preterm birth rate of 9.1% in this study. Treatment effect was modeled as a shift in the GAB corresponding to the effect sizes reported in published studies of preterm birth prevention that were based on single treatment modalities, and with extrapolated treatment effects estimating additive effects of bundled interventions.

For the first co-primary endpoint, reduction of severe composite morbidity/mortality, a range of the possible expected effect size can be inferred from the simulations. In the simulations, the proportion of subjects with NMI ≥ 3 is expected to be between 0.023 and 0.02 in the prospective arm and near 0.036 in the control arm based on a previous clinical utility study (Sera Prognostics: data on file). Assuming these proportions, a one-

sided Fisher's Exact test, a sample size of approximately 1,453 subjects with outcomes in the prospective arm with 55% compliance among higher risk subjects and about 10,000 historical controls provides power of 0.7 – 0.9 (NCSS PASS 2020, module "Group-Sequential Tests for Two Proportions").

For the second co-primary test, length of neonatal hospital length of stay from time of birth up to neonatal discharge is assessed by Cox proportional hazards regression with stratification. The hazard ratio based on simulations is expected to be between 1.32 and 1.46 based on a previous clinical utility study (Sera Prognostics: data on file). Assuming these hazard ratios, a sample size of approximately 1,453 subjects with outcomes in the prospective arm with 55% compliance among higher risk subjects and about 10,000 historical controls provides power of 0.8 – 1.0 (NCSS PASS 2020, modules "Group-Sequential Tests for Two Hazard Rates" and "Tests for Two Survival Curves Using Cox's Proportional Hazards Model"). Stratification is indicated based on non-proportionality observed in previous studies.

The two co-primary endpoints provide independent measurements of improved neonatal outcomes. Due to their correlation, there may be modest power to be gained from having two co-primary endpoints. The estimated study power for the combination of the two co-primary endpoints is 0.8 – 1.0.

5 Statistical Analysis

Statistical tests will be performed at the 5% significance level unless otherwise specified. Such tests will be two-tailed unless otherwise specified as directional tests. Primary and secondary endpoints will be analyzed using one-tailed tests.

5.1 Analysis Populations

One control population will be considered in primary analyses: the historical control population. The historical control population includes all subjects who attended ChristianaCare during the eligibility period and who satisfy all eligibility criteria.

The historical control arm will be the control population for primary and alternative analysis of the primary, secondary and exploratory endpoints.

Seven subject populations will be considered in the analyses: an Intent-to-Treat (ITT) population, a modified ITT population, a completer(s) population, a compliant completers population, a non-intervention completers population and a safety population. The completers, compliant completers, non-intervention completers and modified ITT populations are subsets of the ITT population.

The ITT population includes all historical subjects selected and all prospective patients enrolled regardless of whether or not the co-primary outcomes were observed, or in the prospective arm, any PreTRM test result has been generated or intervention consent has been signed. Subjects with missing outcomes will be handled as described in the missing data sensitivity analyses.

The modified ITT (mITT) population includes all subjects for whom both co-primary outcomes are known; and who have either been selected for the historical control group, received a not-higher-risk PreTRM® test result, or consented to and initiated treatment before $24^{0/7}$ weeks gestation after receiving a higher risk PreTRM® test result.

The completers population includes subjects for whom both of the co-primary outcomes are known, and, in the prospective arm, for whom a PreTRM® test result was generated.

The compliant completers population includes ITT subjects who have missed no more than 20% of daily doses for both low-dose aspirin and vaginal progesterone, and participated in weekly care management calls with no more than 20% missed weeks. The non-intervention completers population includes all ITT subjects who received a higher-risk PreTRM® test result but neither consented to nor initiated treatment before $24^{0/7}$ weeks gestation.

The safety population includes all eligible subjects who gave consent to participate in the study and attempted a blood draw for the PreTRM® test, regardless of receipt of study interventions.

5.2 Descriptive statistics

Tables of demographic characteristics of participants in the PreTRM® and historical control groups will be created from the study data bases. Baseline subject characteristics include: demographic information, medical history, physical examination, and other laboratory tests. Comorbidities include risk factors, diabetes mellitus, or metabolic syndrome, among others. Tables of resource use of prospective subjects and historical control groups will be created as well. Variables will be summarized using means, medians, ranges, interquartile ranges and standard deviations for continuous data and counts and percentages for categorical data. In addition, we will also summarize the percentage of prospective subjects classified as higher and not-higher risk from the PreTRM® test results. Summaries will be created for both the mITT population and separately for the ITT population, and for the higher and not-higher risk groups in the mITT population.

5.3 Covariate Analyses

The prospective participants and historical controls in the study may differ in ways other than the interventions or exposures under investigation. Covariates will be used to adjust the regressions prespecified in the primary analysis for potential differences between the control and prospective arms, and in derivation of propensity scores used to adjust for biases between the two arms, either by weighting subjects in the alternative analysis or by matching subjects in exploratory analyses..

The following variables are candidates for inclusion as covariates in the primary analysis: regression with covariates. These baseline covariates are present before blood draw.

- Maternal age
- Categorical parity: multiparous vs. nulliparous
- Maternal substance use disorders assessed as Neonatal Opioid Withdrawal Syndrome

The following variables are candidates for inclusion as covariates in the alternative analysis using inverse probability score weighting, and in the exploratory analysis using propensity matching. These baseline covariates are present before blood draw.

- Maternal age
- Categorical parity (nulliparous vs. multiparous)
- Gravidity (integer)
- Number of term children
- Number of miscarriages
- Maternal race
- Pre-pregnancy continuous BMI
- Pre-pregnancy categorical low BMI<19
- Maternal height
- Maternal chronic hypertension
- Gestational diabetes
- Insurance status
- Smoking status
- Maternal substance use disorders assessed as Neonatal Opioid Withdrawal Syndrome
- Delivery outside of ChristianaCare

Inclusion of covariates in the covariate-adjusted primary, alternative and exploratory analyses will depend on the proportion of missing values for the covariate, as follows. For each candidate variable, if fewer than 1% of subjects have a missing value, that variable will be included as a covariate in the analysis. Sensitivity of results to missing data will be examined using missing data analysis methods.

The primary and alternative analyses will use cases with complete data for the included covariates, provided that no more than 5% of the total subjects are excluded from the analysis. If more than 5% of the total subjects would be excluded from the analysis,

candidate variables will be dropped from the analyses in order of the percent missing, until no more than 5% of the total subjects are excluded from the analysis.

Descriptive statistics on baseline covariates will be compared between the prospective arm and historical controls. This analysis includes covariates prespecified for both the primary and alternative analyses. If we can obtain information on some of the covariates that affect exposures and decisions to participate, we can assess our ability to reduce bias. Sociodemographic factors such as education and age are often considered in this role.

It is common that the basic characteristics may not be balanced between the prospective cohort and historical cohort groups. We will examine the potential sources of bias and address them with a classical and common technique—adjusting by covariates.

While an easy to implement strategy to control for selection bias in propensity scoring is to adjust for factors that can break the biasing paths linking the exposure and the outcome, regression with covariates benefits from adjustment for independent predictors of the outcome. We expect that adjustment for independently predictive covariates can result in decreased selection bias and increased precision.

The factors which impact the effect of intervention through selection and epoch biases will be assessed to demonstrate how well they address confounding in the analysis. Model selection techniques will be used to demonstrate an optimized model. Although a minimum set of covariates is prespecified for use in addressing confounding, there may be an advantage in exploratory analyses to using a staged approach in which groups of covariates are introduced sequentially leading to progressively greater adjustment. Using random forest and other machine learning methods, we will provide insight into which covariates have relatively greater influences on effect estimates, permitting comparison with known or expected associations or permitting the identification of possible mediator and/or intermediate variables.

In alternative and exploratory analyses, the propensity score model will model the propensity score: predicted probability of a participant who will be placed in the prospective group, via running logistic regression: the dependent variable,

y=1, if prospective group;

y=0, if historical group.

The propensity score model will be derived from logistic regression using the key clinical covariates available at the gestational age of PreTRM® blood draw in the historical cohort, as well basic demographic characteristics. The variables used to generate propensity scores are specified above.

The performance of the propensity score model will be evaluated by AUC/ROC c-index; potential interaction terms or high degree of terms may be included to improve the performance of the propensity model. We will obtain an estimation for the propensity score: predicted probability (p).

With propensity score methods, we will check whether covariates are balanced across the prospective and historical cohorts. We can examine the distribution of propensity score between groups. We will assess covariate balance between groups using p-value for comparison of two group, as well absolute standardized mean difference (ASMD), with an ASMD <0.1 indicating good balance.

We propose Inverse Propensity score Weighting (IPW) method for the alternative analysis. IPW is a statistical technique for calculating statistics standardized to a pseudo-population different from that in which the data was collected. After obtaining propensity score, we define for the prospective arm:

$IPW = \frac{1}{PS}$; $IPW = \frac{1}{PS}$; for historical controls: $IPW = \frac{1}{1-PS}$.

$IPW = \frac{1}{1-PS}$. To avoid the IPW method's instability to extremes of propensity among participants with high propensity score weights with undue influence the results, we will use truncated scores and weights: 1) below the 1st or 5th percentile and above the

99th or 95th percentile; resulting in weights between reasonable intervals for all participants; 2) truncating propensity score <0.001 or >0.999.

5.4 Primary and Alternative Analyses

The primary and alternative analyses will be performed in the modified ITT (mITT) population on the primary, secondary and exploratory endpoints.
The primary analysis is regression with covariates as specified for each endpoint.

The alternative analysis is regression with covariates as specified for each primary, secondary and exploratory endpoint, with weighting by inverse propensity score to address potential selection bias.
Propensity-matched and unadjusted "raw" analyses are pre-specified as exploratory analyses. These analyses will be performed on primary and secondary endpoints.

In analysis combining primary, alternative and exploratory analyses, we will present the demographic and baseline characteristics differing between the prospective arm and historical controls, selected from among the covariates prespecified for the primary and alternative analyses. We will compare the primary analysis using prespecified covariates, the alternative analysis using IPW, and exploratory PSM un-adjusted (raw) data analyses (Table 2). The results from these four distinct methods are expected to show similarity in direction and potentially in magnitude of effect. If differences in direction are observed, further analysis such as double robust propensity score analysis will be conducted.

Variable	Covariate Regression			Inverse Probability Weighted			Matched data			Unadjusted "Raw" data		
	Prospec-tive (N=)	Historical (N=)	p-value	Prospec-tive (N=)	Historical (N=)	p-value	Prospec-tive	Historical	p-value			
Age												
Parity												

Table 2: Baseline Characteristics

For pre-specified analysis populations and for subgroups specified under Subgroup Analyses below, analyses will be conducted to examine the clinically important groups and unbalanced variables between the prospective and historical cohorts.

5.5 Primary Endpoints

Co-Primary Endpoint: neonatal morbidity and mortality (NMI)

Reduction in composite neonatal morbidity and mortality (NMI)(14; PREGNANT trial) in prospective subjects versus controls will be evaluated by test of proportions of NMI of ≥3 vs. other scores. The null hypothesis (H_{01}) is no difference between controls and the prospective arm. The alternative hypothesis is (H_{a1}) relative reduction in proportion of NMI of ≥3 in the prospective arm as compared to controls. Statistical testing for this endpoint will be performed using ordinal logistic regression model with covariates with weighting of NMI by NMI level value plus a constant such as 0.01.

The primary hypothesis is:

H_0: no difference between controls and the prospective arm

Odds of a higher vs. lower NMI in the prospective arm = Odds of a higher vs. lower NMI in historical controls

H_A: relative reduction in odds of a higher vs. lower NMI in the prospective arm as compared to historical controls.

Co-Primary Endpoint: length of neonatal hospital stay

Reduction in length of neonatal hospital stay for all admissions from time of birth up to neonatal discharge in prospective subjects versus controls will be evaluated using a time-to-event analysis in the top quantile of length of stay. To avoid significant contamination of short stays associated with term births, the strata will be adjusted by the preterm birth rate. If 1.2 times the preterm birth rate in the controls is less than 10%, the quantile will be set at 1.2 times the preterm birth rate. Otherwise the quantile will be set to 10%. The null hypothesis (H_{02}) is no difference between controls and the prospective arm. The alternative hypothesis (H_{a2}) is shorter length of stay in the prospective arm as compared

to controls. Statistical testing for this endpoint will be performed using a Cox proportional hazards regression model with covariates.

The primary hypothesis is:

H_0: no difference between controls and the prospective arm

NNOLOS in prospective arm =

NNOLOS in historical controls

H_A: shorter NNOLOS in the prospective arm

as compared to historical controls.

Specification of the co-primary endpoint of NNOLOS includes adjustment for non-proportionality. In case of significant non-proportionality remaining after the adjustment prespecified above, we will undertake additional exploratory analyses in the context of the Cox model. We will stratify any covariates with non-proportional effects and incorporate the covariates into the model as stratification factors rather than predictors; this includes discretizing quantitative variables. We will document any non-linear effects of continuous covariates associated with non-proportional effects. Further, we will partition the time axis to analyze if the proportional hazards assumption holds for specific time periods (say the first week vs. after the first) but not for the entire length of hospital stay.

Holm's method will be used to control for multiple hypothesis testing ($H_{01}\ vs\ H_{a1}$, $H_{02}\ vs\ H_{a2}$) of the co-primary endpoints in the study, controlling the family-wise error rate, to a significant α-level of < 0.05.

5.6 Secondary Endpoints

Reduction in length of NICU hospital stay (NICULOS) for all neonates from time of birth up to neonatal discharge in prospective subjects versus controls will be evaluated using a time-to-event analysis in the top quantile of length of stay. To avoid significant contamination of short stays associated with term births, the strata will be adjusted by the

preterm birth rate. If 1.2 times the preterm birth rate in the controls is less than 10%, the quantile will be set at 1.2 times the preterm birth rate. Otherwise the quantile will be set to 10%.

The secondary hypothesis is as follows:

H_0: no difference between controls and the prospective arm

NICULOS in PreTRM® test group =

NICULOS in historical group

H_A: shorter NICULOS in the PreTRM® prospective arm as compared to historical controls.

Statistical testing for this endpoint will be performed using a Cox proportional hazards regression model with covariates and stratification as specified above.

For increase in time of gestation in prospective subjects versus historical controls, we will evaluate it using a time-to-event analysis in the bottom quantile of gestational ages at birth. To avoid significant contamination of short stays associated with term births, the strata will be adjusted by the preterm birth rate. If 1.2 times the preterm birth rate in the controls is less than 10%, the quantile will be set at 1.2 times the preterm birth rate. Otherwise the quantile will be set to 10%.

The secondary hypothesis is as follows. :

H_0: no difference between controls and the prospective arm

time of gestation in PreTRM® test group =

time of gestation in historical group

H_A: longer time of gestation in the PreTRM® prospective arm as compared to historical controls.

Statistical testing for this endpoint will be performed using a Cox proportional hazards regression model with covariates and stratification as specified above.

A secondary endpoint will be met if the endpoint achieves significance with an alpha of 0.05. No adjustments will be made for the multiple hypothesis tests among the secondary endpoints, but conclusions will be interpreted with caution due to the multiple tests proposed.

Non-proportionality of secondary endpoints will be addressed using the pre-specified adjustments for non-proportionality, and the additional approaches provided above for the co-primary endpoint NNOLOS.

5.7 Exploratory Analyses

In AVERT, there is a challenge and opportunity in the moderate rate of acceptance of intervention among high-scoring women in the intervention arm. The exploratory analyses include options for addressing the non-accepting completers in the prospective arm, and comparing them to those in mITT and the compliant completers.

The exploratory outcomes will be summarized within the prospective arm and within historical controls using descriptive statistics and graphical displays where appropriate. Continuous outcomes will be summarized using sample size (n), means, medians, standard deviations, and ranges. Categorical outcomes will be summarized using frequency and percentages. Estimates will be presented with 95% confidence intervals. Exploratory summaries will be performed for the mITT population.
The following exploratory endpoints will be evaluated by a test of proportions in the prospective arm versus historical controls. The null hypothesis is no difference between historical controls and the prospective arm. The alternative hypothesis is relative reduction in proportion of occurrence of one or more conditions in the prospective arm as compared to historical controls. Statistical testing for these endpoints will be performed using a logistic regression model with covariates as specified above for the primary analysis, and with propensity scoring adjustment as specified for the alternative analysis.

- Reduction in occurrence of one or more major neonatal morbidities with high likelihood of major chronic illness (MNM): cystic periventricular leukomalacia,

grade 3 and 4 intraventricular hemorrhage, grade 3 or higher retinopathy of prematurity, bronchopulmonary dysplasia.

- Reduction in the rate of PTB <35 weeks, <32 weeks and <37 weeks of gestation. This analysis will be repeated for premature neonates delivered after spontaneous rupture of membranes or spontaneous onset of labor with gestational age at birth <32, <35 and <37 weeks vs. all other neonates.
- Reduction in NICU admission rates in the immediate neonatal period prior to initial discharge or neonatal death. This analysis will be repeated for all premature neonates, and for premature neonates delivered after spontaneous rupture of membranes or spontaneous onset of labor.

The following exploratory endpoints will be evaluated using a time-to-event analysis. The null hypothesis is no difference between historical controls and the prospective arm. The alternative hypothesis is shorter length of stay in the prospective arm as compared to historical controls. Statistical testing for this endpoint will be performed using a Cox proportional hazards regression model with covariates and stratification as specified above for the primary analysis, and with propensity scoring adjustment as specified for the alternative analysis. Length of hospital stay is adjusted to the maximum observed length of stay plus one for all instances of perinatal mortality. Length of NICU stay is adjusted to the maximum observed length of stay plus one for all instances of NICU mortality.

- Reduction in length of NICU hospital stay of all NICU admissions for premature neonates. For clarity, this analysis includes only PTBs with NICU admission. This analysis will be repeated for all premature neonates delivered after spontaneous rupture of membranes or spontaneous onset of labor.
- Reduction in length of NICU hospital stay (including zero-length stays for those not admitted) for all premature neonates from time of birth up to neonatal discharge. For clarity, this analysis includes PTBs independent of NICU admission, with stays for the non-admitted set to zero days. This analysis will be repeated for all premature neonates delivered after spontaneous rupture of membranes or spontaneous onset of labor.

- Reduction in length of neonatal hospital stay for all premature neonates from time of birth up to neonatal discharge. This analysis will be repeated for all premature neonates delivered after spontaneous rupture of membranes or spontaneous onset of labor.

As exploratory analyses, the primary and alternative analyses will be replicated, with the modification that the exploratory analyses will use multiple imputation and will include the candidate variables for which more than 5% of subjects have a missing value.

As exploratory analyses, the primary analysis will be replicated as a "raw" unadjusted analysis, with the modification that no covariates will be included in regressions.

As an exploratory analysis, a test of the proportional odds assumption for NMI data will be conducted. If the proportional odds assumption does not hold, we will examine the data using a set of separate logistic regressions or partial proportional odds models, to explicitly see how the odds ratios for the explanatory variables vary at the different thresholds. Alternatively, we will also apply the generalized logit model, which treats the response as nominal (unordered) rather than ordinal and has a full set of parameters for each generalized logit. We will then compare the results from proportional, fully nonproportional, and partial proportional odds models fitted to data.

As exploratory analyses, the primary analysis will be replicated, with the modification that the analysis will be restricted to prospective subjects with at least one propensity-matched control, and historical controls will be restricted to those propensity-matched to prospective subjects. At most 4 historical controls will be matched to each prospective subject. Matching algorithms will be chosen from among nearest neighbor matching, caliper matching with comparison units within a certain width of the propensity score of the treated units get matched, where the width is generally a fraction of the standard deviation of the propensity score; Mahalanobis metric matching in conjunction with propensity score matching (PSM), stratification matching, difference-in-differences matching (kernel and local linear weights), or exact matching. The matching method will

be chosen by the distribution of the propensity scores and the balance status of baseline characteristics.

As exploratory analyses, the primary and alternative analyses will be replicated using the ITT population. Subjects with missing data for required variables will be excluded from the analyses. To test if covariates are known to be misclassified under some approaches, an "intention to treat" analysis will be conducted that assumes that each participant continues to be exposed once they have received an initial treatment. Originally used in the analysis of randomized trials, this approach has been used in this hybrid study. It can be worthwhile to do a sensitivity analysis on studies that use an "intention to treat" approach to see how different an "as treated" analysis would be even if intention to treat is the main estimate of interest.

As exploratory analyses, the primary and alternative analyses will be replicated using the completers population with treatment covariates. Covariates for treatment modalities inclusive of use of aspirin or progesterone, will be included in the analyses if showing record of use between LMP and delivery in both the prospective arm and historical controls. Subjects with missing data for required variables will be excluded from the analyses. Subjects in the higher-risk group accepting interventions will be compared to non-intervention completers and to historical controls using descriptive statistics as described above. If use of any treatment is significantly associated with variables specified for the alternative analysis, this analysis will be repeated with inclusion of such variables as covariates in the analysis. Subjects with missing data for required variables will be excluded from this analysis.

As an exploratory analysis, the primary and alternative analyses will be replicated using the compliant completers population plus the not-higher-risk group as the preterm prevention arm for comparison to historical controls. Additional subgroup analysis of the compliant completers population will apply the primary, secondary and exploratory endpoints to subjects compliant with each component of the intervention versus all other compliant completers, and versus non-intervention completers. If compliance with any

treatment is significantly associated with variables specified for the alternative analysis, this analysis will be repeated with inclusion of such variables as covariates in the analysis. Subjects with missing data for required variables will be excluded from this analysis.

The use of antenatal corticosteroids will be assessed in relation to primary, secondary and exploratory endpoints, and will be compared between the preterm prevention arm and historical controls, and between higher risk and not higher risk subjects in the preterm prevention arm. Optimal timing of dosing of antenatal corticosteroids is herein defined as 2 to 7 days (inclusive) prior to delivery at less than 37 weeks of gestation. Use of antenatal corticosteroids will be expressed as a binary variable: were steroids given in the optimal interval between dose and delivery? Significance of interaction between optimally timed use of antenatal corticosteroids and the specified treatment groups will be assessed in regression as an interaction term: (binary variable) * group. This analysis will be performed without adjustment, adjusted for covariates as specified for the primary analysis, and adjusted via propensity scoring as specified for the alternative analysis. Missing values will be addressed as previously specified for exploratory analyses: by omission of subjects with missing values and by replicating the analysis with multiple imputation of missing values.

Primary, secondary and exploratory endpoints will be assessed in the prospective portion of the mITT population as dependent responses to nonrandom quantitative assignment of treatment[42]. The independent quantitative assignment variable will be the PreTRM® risk score. The functional relationship between endpoints and the risk score will be estimated for the prospective arm. A discontinuity at the intervention threshold in the relationship between outcome and risk score observed in the prospective arm will demonstrate intervention effect, measured as the estimated magnitude of the discontinuity and re-assessed as the estimated risk-score dependent difference between compliant completers and non-intervention completers. As this analysis involves only prospective subjects, it is independent of any selection biases or epoch effects present in comparisons of the prospective arm relative to historical controls. To adjust for any biases present within the

prospective arm, the analysis will be performed without adjustment, adjusted for covariates as specified for the primary analysis, and adjusted via propensity scoring as specified for the alternative analysis.

The primary, secondary and exploratory endpoints will be applied to comparisons within the mITT population.

a) The higher-risk group will be compared to the not-higher-risk group within the prospective arm, for tests of equivalence. This comparison is independent of any selection biases or epoch effects present in comparisons of the prospective arm relative to historical controls.
b) The not-higher-risk group of the prospective arm will be compared to historical controls to assess test sensitivity. It is expected that there will be significant differences in baseline characteristics between the two groups. The analysis will be performed without adjustment, adjusted for covariates specified for the primary analysis, and adjusted via propensity scoring as specified for the alternative analysis.
c) The higher-risk group in the prospective arm will be compared to a control comparator group estimated as the difference between the control group and the not-higher-risk group, to assess efficacy of the intervention strategy. For count endpoints, direct subtraction of counts (adjusted for arm size) will be used to estimate the comparator group. For non-count endpoints, a number of most similar outcomes proportional to arm size will be removed from the control comparator group for each subject in the not-higher-risk group. It is expected that there will be significant differences in baseline characteristics between the two groups. The analysis will be performed without adjustment, adjusted for covariates specified for the primary analysis, and adjusted via propensity scoring as specified for the alternative analysis.

The following exploratory outcomes will be examined using only the descriptive statistics described above.

- Birthweight
- Number of days of mechanical ventilation

- Amount of surfactant administered
- Occurrence of birthweight <1500g and <2500g
- Individual components of composite morbidity/mortality index
- Whether or not a neonate received surfactant
- Occurrence of pneumonia
- Occurrence of 5-minute Apgar<7
- Occurrence of asphyxia
- Occurrence of preeclampsia
- Occurrence of gestational diabetes mellitus
- Rate of PPROM
- Rate of preterm delivery not associated with PPROM
- Baseline and on-treatment progesterone levels determined by LC-MS

Sensitivity analyses will be performed to check uncertainties in our analysis, as well as address basic assumption validation and the effect of unmeasured/unbalanced confounding factors on primary outcomes. The analyses will compare inclusion and omission of covariates in the analyses of primary, secondary and exploratory endpoints in the mITT population. As PTB is an outcome whose consequences are dominated by the lower tail of the gestational age distribution, sensitivity analyses will also examine the effect on effect size estimates of omitting outliers from the analyses.

A sensitivity analysis to control for reduction in subject number in the prospective arm of the mITT population will be performed. The primary and alternative analyses will be replicated as sensitivity analysis with bootstrap sampling of historical controls to the same number of subjects as the prospective arm. If exclusion of any completers from the mITT population is significantly associated with variables specified for the alternative analysis, this analysis will be repeated with such variables included as covariates. Subjects with missing data for required variables will be excluded from the covariate analysis.

A Bayesian Meta-Analysis will be performed on the data from this study and relevant previous studies. This will be described in a separate document.

Health Economic Outcomes:

Total cost of hospital care for each of the mother and neonate beginning at initiation of care through primary delivery and a gestational age of $40^{0/7}$ weeks in prospective vs. historical subjects. Analysis of these outcomes is covered in a separate SAP entitled SEAR.

5.8 Safety Analyses

Drawing of blood and the interventions planned for women diagnosed as higher risk have a known and established safety record. However, we will collect and analyze safety information for the mother. The definitions for AEs, UAEs, and SAEs are included in the protocol. These adverse events will be summarized and descriptive statistics provided. All adverse events (AEs) will be coded using the Medical Dictionary for Regulatory Activities(MedDRA). Frequencies of adverse events will be presented by system organ class, preferred term, treatment group, and severity level. Safety analyses will be performed in the safety population.

5.9 Subgroup analyses

Subgroup analyses may be performed if the sample size in this study can provide sufficient power.

The use of regression methods simplifies analysis of subgroups, or the impact of other factors such as age, or other comorbidities on the effectiveness of the PreTRM® strategy. These can easily be included as covariates in a regression model, including treatment (prospective arm or historical control) by covariate interactions. All covariates included in the alternative analysis in the Covariate Analyses section will be included in subgroup analyses.

In prespecified subgroup analyses, interaction terms will include age, chronic hypertension and gestational diabetes. Covariates categorical BMI, categorical parity and

insurance status will be used to defined subgroups. The remaining covariates defined for the alternative analysis will be used as covariates in these analyses.

5.10 Missing data

We need to evaluate the quality control procedures in historical control dataset, with respect to missing data and outlying values. We expect that most variables that will enter the analyses will have few missing values. In cases for which more than one percent of the observations have missing values, we have used various approaches, all depending on the "missing at random" (MAR) assumption. If missingness can be considered to be random after conditioning on observed data, then the mechanism is said to be missing at random.[42-44] Both mixed-effects models and multiple imputation are appropriate under these circumstances. When the missingness depends on the values of the missing data, however, the missing value mechanism is said to be non-ignorable, and standard statistical procedures are no longer valid. As discussed by Verbeke and Molenbergh, virtually all strategies for dealing with this type of data are based on a number of "untestable and unverifiable" assumptions, which limit their utility.[45] Clearly, the best defense is to simply avoid missing data as far as humanly possible.

Missing data patterns of both prospective subjects and historical controls will be assessed by age, and other appropriate characteristics.[46] Missing data patterns of the strategies will be compared to determine similarity. Analyses will be made comparing number of missing values of the prospective and historical control dataset. In this case, the outcome is dichotomous, missing or not missing, and the event rates of the two groups will be compared by Poisson regression, or negative binomial regression if the Poisson model is over-dispersed. This analysis will include covariates specified for the primary and alternative analyses. Similar to logistic regression, the exponentiated regression coefficients of the model are incident rate ratios, comparing event rates of the intervention vs. historical cohort groups.

6 Bias in the Hybrid Study

6.1 Potential bias in the hybrid study

In this study, it is crucial to address selection bias, epoch bias, bias from self-selection of compliant completers in the higher-risk group, as well bias due to COVID-19 / SARS-CoV-2 exposure in the last ~400 subjects. Selection bias is the bias introduced by the selection of individuals, groups or data for analysis in such a way that proper randomization is not achieved, thereby ensuring that the sample obtained is not representative of the population intended to be analyzed. In this study, prospective intervention group and historical cohort study group are not allocated through randomization, therefore selection bias may be one of the important issues to be addressed. As the prospective arm and historical controls received care in different periods, epoch effects from trends in patient care and in public health may also prove an important source of bias. Finally, incomplete acceptance of interventions by some higher risk subjects in the prospective arm provides a source of bias that cannot be directly adjusted, as no higher risk group defined by PreTRM® risk score is present amongst the historical controls. Bias due to comparison of historical controls from a pre-pandemic period to prospective subjects reaching term during the SARS-CoV-2 pandemic has been ameliorated by restriction of the study to subjects reaching term before detection of the first SARS-CoV-2 positive case in Delaware.

Bias from any source will cause the distortion of a statistical analysis. Bias can have varying effects, and the magnitude of its impact and the direction of the effect are often hard to determine. If bias is not taken into account, the results and conclusions of the study may be false or misleading. Therefore, it is necessary to consider the sources of bias in the outcomes analysis.

6.2 Potential differences between historical and prospective cohorts

The assessment of selection bias involves assumptions regarding inclusion or participation by potential subjects, and results can be highly sensitive to assumptions. Even with external validation data, which may work for unmeasured confounders, it is difficult to account for more than a trivial amount of selection bias. Self-selection bias in consenting of higher-risk subjects to interventions further depends on PreTRM® score, a

factor measured only in the prospective arm. Assessment of epoch bias adds assumptions regarding the composition of standard care and incidence of comorbid conditions due to public health trends. Since the study population is a hybrid of prospective cohort and historical cohort, instead of randomized clinical trial, bias is the key issue in the analysis of outcomes. Using appropriate methods to address bias is critical in the study. Different approaches have been applied, including multivariate analysis with covariate adjustment in the studies in which selection bias needs to be handled, and more recently, propensity scoring. Propensity score analysis, includes subgroups defined by propensity score, propensity score used as a covariate in multivariate analysis, matched patients in two groups by propensity score, and inverse probability weighting. Methods based on propensity scores have become important in addressing selection bias in recent years. Other methods are instrumental variable and regression discontinuity design (RDD).

Comparison of diverse adjustment methods as specified in the primary, secondary and exploratory analyses will enable assessment of the degree of selection and epoch bias in the study, and the quality of prespecified and exploratory adjustments for bias.

As well, exploratory analyses within the prospective arm will provide additional estimates of intervention efficacy without reference to historical controls.

Statistical Analysis Plan for AVERT Economic Study (SEAR, Separate Document)

REFERENCES

1 Anderson, Robert N, and Betty L Smith. 2003. “Deaths: Leading Causes for 2001.” *National Vital Statistics Reports : From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* 52 (9): 1–85. http://www.ncbi.nlm.nih.gov/pubmed/14626726.

2 Berghella, Vincenzo, Timothy J Rafael, Jeff M Szychowski, Orion A Rust, and John Owen. 2011. “Cerclage for Short Cervix on Ultrasonography in Women with Singleton Gestations and Previous Preterm Birth: A Meta-Analysis.” *Obstetrics and Gynecology* 117 (3): 663–71. http://www.ncbi.nlm.nih.gov/pubmed/21446209.

3 Boelig, Rupsa C, Kelly M Orzechowski, and Vincenzo Berghella. 2016. “Cervical Length, Risk Factors, and Delivery Outcomes among Women with Spontaneous Preterm Birth.” *The Journal of Maternal-Fetal & Neonatal Medicine : The Official*

Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians 29 (17): 2840–44. doi:10.3109/14767058.2015.1105957.

4 Christian, M S, R L Brent, and P Calda. 2007. "Embryo-Fetal Toxicity Signals for 17alpha-Hydroxyprogesterone Caproate in High-Risk Pregnancies: A Review of the Non-Clinical Literature for Embryo-Fetal Toxicity with Progestins." *The Journal of Maternal-Fetal & Neonatal Medicine : The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians* 20 (2). Argus International, Horsham, PA 19044, USA. mschristian@comcast.net: 89–112. doi:10.1080/14767050601178758.

5 "CLASP: A Randomised Trial of Low-Dose Aspirin for the Prevention and Treatment of Pre-Eclampsia among 9364 Pregnant Women. CLASP (Collaborative Low-Dose Aspirin Study in Pregnancy) Collaborative Group." 1994. *Lancet* 343 (8898): 619–29. http://www.ncbi.nlm.nih.gov/pubmed/7906809.

6 "Committee Opinion No. 561: Nonmedically Indicated Early-Term Deliveries." 2013. *Obstetrics and Gynecology* 121 (4). United States: 911–15. doi:10.1097/01.AOG.0000428649.57622.a7; 10.1097/01.AOG.0000428649.57622.a7.

7 Corporation, Ther Rx. n.d. "Makena Prescribing Information."

8 Duley, L, D J Henderson-Smart, M Knight, and J F King. 2004. "Antiplatelet Agents for Preventing and Treating Pre-Eclampsia." *Cochrane Database of Systematic Reviews (Online)*, no. 4: CD000492. doi:10.1002/14651858.CD000492.

9 Esplin, M. Sean, Michal A. Elovitz, Jay D. Iams, Corette B. Parker, Ronald J. Wapner, William A. Grobman, Hyagriv N. Simhan, et al. 2017. "Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women." *JAMA* 317 (10). American Medical Association: 1047. doi:10.1001/jama.2017.1373.

10 Fonseca, E B, E Celik, M Parra, M Singh, K H Nicolaides, and Fetal Medicine Foundation Second Trimester Screening Group. 2007. "Progesterone and the Risk of Preterm Birth among Women with a Short Cervix." *The New England Journal of Medicine* 357 (5). Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, London, United Kingdom.: Massachusetts Medical Society: 462–69. doi:10.1056/NEJMoa067815.

11 Goldenberg, Robert, and Dwight Rouse. 1998. "Prevention of Premature Birth." *New England Journal of Medicine* 339 (5): 313–20.

12 Goya, M, L Pratcorona, C Merced, C Rodo, L Valle, A Romero, M Juan, et al. 2012. "Cervical Pessary in Pregnant Women with a Short Cervix (PECEP): An Open-Label Randomised Controlled Trial." *Lancet*, April. Maternal Fetal Medicine Unit, Department of Obstetrics, Hospital Universitari Vall d'Hebron, Universitat Autonoma de Barcelona, Barcelona, Spain; Maternal and Child Health and Development Network, Instituto Salud Carlos III, Madrid, Spain.: Elsevier Ltd. doi:10.1016/S0140-6736(12)60030-0.

13 Hamilton, Brady E, Joyce A Martin, Michelle J K Osterman, and Sally C Curtain. 2015. “Births: Preliminary Data for 2014.” *National Vital Statistics Reports : From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* 64 (6): 1–19. http://www.ncbi.nlm.nih.gov/pubmed/26114874.

14 Hassan, S S, R Romero, D Vidyadhari, S Fusey, J K Baxter, M Khandelwal, J Vijayaraghavan, et al. 2011. “Vaginal Progesterone Reduces the Rate of Preterm Birth in Women with a Sonographic Short Cervix: A Multicenter, Randomized, Double-Blind, Placebo-Controlled Trial.” *Ultrasound in Obstetrics & Gynecology : The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 38 (1): 18–31. doi:10.1002/uog.9017.

15 Hui, Shuk Yi Annie, Chung Ming Chor, Tze Kin Lau, Terence T Lao, and Tak Yeung Leung. 2013. “Cerclage Pessary for Preventing Preterm Birth in Women with a Singleton Pregnancy and a Short Cervix at 20 to 24 Weeks: A Randomized Controlled Trial.” *American Journal of Perinatology* 30 (4): 283–88. doi:10.1055/s-0032-1322550.

16 Keirse, M J. 1990. “Progestogen Administration in Pregnancy May Prevent Preterm Delivery.” *British Journal of Obstetrics and Gynaecology* 97 (2): 149–54. http://www.ncbi.nlm.nih.gov/pubmed/2138496.

17 Knight, M, L Duley, D J Henderson-Smart, and J F King. 2000. “Antiplatelet Agents for Preventing and Treating Pre-Eclampsia.” *Cochrane Database of Systematic Reviews (Online)*, CD000492. doi:10.1002/14651858.CD000492.

18 Kozer, Eran, Shekoufeh Nikfar, Adriana Costei, Rada Boskovic, Irena Nulman, and Gideon Koren. 2002. “Aspirin Consumption during the First Trimester of Pregnancy and Congenital Anomalies: A Meta-Analysis.” *American Journal of Obstetrics and Gynecology* 187 (6): 1623–30. http://www.ncbi.nlm.nih.gov/pubmed/12501074.

19 Maher, Mohammad Ahmed, Ahmed Abdelaziz, Mohamed Ellaithy, and Mohamed Fayez Bazeed. 2013. “Prevention of Preterm Birth: A Randomized Trial of Vaginal Compared with Intramuscular Progesterone.” *Acta Obstetricia et Gynecologica Scandinavica* 92 (2): 215–22. doi:10.1111/aogs.12017.

20 Mahoney, E M; Chu H. 2003. *Cardiovascular Health Care Economics. Cost-Effectiveness Analysis alongside Clinical Trials: Statistical and Methodologic Issues*. Edited by Weintraub WS. TOTOWA: Humana Press.

21 March of Dimes Save the Children, WHO., PMNCH. 2012. *Born Too Soon: The Global Action Report on Preterm Birth.* . Edited by M V Kinney CP Howson JE Lawn. Geneva: World Health Organization.

22 Marret, Stéphane, Laetitia Marchand, Monique Kaminski, Béatrice Larroque, Catherine Arnaud, Patrick Truffert, Gérard Thirez, Jeanne Fresson, Jean-Christophe Rozé, and Pierre-Yves Ancel. 2010. “Prenatal Low-Dose Aspirin and Neurobehavioral Outcomes of Children Born Very Preterm.” *Pediatrics* 125 (1): e29-34. doi:10.1542/peds.2009-0994.

23 Mathews, T J, Fay Menacker, and Marian F MacDorman. 2004. “Infant Mortality Statistics from the 2002 Period: Linked Birth/infant Death Data Set.” *National Vital*

Statistics Reports : From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System 53 (10): 1–29. http://www.ncbi.nlm.nih.gov/pubmed/15622996.

24 McCormick, Marie C, and Douglas K Richardson. 2002. “Premature Infants Grow Up.” *The New England Journal of Medicine* 346 (3): 197–98. doi:10.1056/NEJM200201173460310.

25 McNamara, Helen Christine, Rachael Wood, James Chalmers, Neil Marlow, John Norrie, Graeme MacLennan, Gladys McPherson, Charles Boachie, and Jane Elizabeth Norman. 2015. “STOPPIT Baby Follow-up Study: The Effect of Prophylactic Progesterone in Twin Pregnancy on Childhood Outcome.” *PloS One* 10 (4): e0122341. doi:10.1371/journal.pone.0122341.

26 Hillemeier MM, Domino ME, Wells R, et al. Effects of maternity care coordination on pregnancy outcomes: propensity-weighted analyses. Matern Child Health J. 2015;19(1):121-127.

27 Manuck TA, Henry E, Gibson J, et al. Pregnancy outcomes in a recurrent preterm birth prevention clinic. Am J Obstet Gynecol. 2011;204(4):320 e321-326.

28 Romero R, Conde-Agudelo A, Da Fonseca E, O'Brien JM, Cetingoz E, Creasy GW, Hassan SS, Nicolaides KH. Vaginal progesterone for preventing preterm birth and adverse perinatal outcomes in singleton gestations with a short cervix: a meta-analysis of individual patient data. Am J Obstet Gynecol. 2018 Feb;218(2):161-180.

29 Meis, P J, M Klebanoff, E Thom, M P Dombrowski, B Sibai, A H Moawad, C Y Spong, et al. 2003. “Prevention of Recurrent Preterm Delivery by 17 Alpha-Hydroxyprogesterone Caproate.” *The New England Journal of Medicine* 348 (24). Department of Obstetrics and Gynecology, Wake Forest University, Winston-Salem, NC 27157, USA. pmeis@wfubmc.edu: Massachusetts Medical Society: 2379–85. doi:10.1056/NEJMoa035140.

30 Newman, Roger B, Scott A Sullivan, M Kathryn Menard, Charles S Rittenberg, Amelia K Rowland, Jeffrey E Korte, and Heather Kirby. 2008. “South Carolina Partners for Preterm Birth Prevention: A Regional Perinatal Initiative for the Reduction of Premature Birth in a Medicaid Population.” *American Journal of Obstetrics and Gynecology* 199 (4): 393.e1-8. doi:10.1016/j.ajog.2008.07.047.

31 Nicolaides, Kypros H, Argyro Syngelaki, Liona C Poon, Gemma Picciarelli, Natasa Tul, Aikaterini Zamprakou, Evdoxia Skyfta, Mauro Parra-Cordero, Ricardo Palma-Dias, and Jesus Rodriguez Calvo. 2016. “A Randomized Trial of a Cervical Pessary to Prevent Preterm Singleton Birth.” *The New England Journal of Medicine* 374 (11): 1044–52. doi:10.1056/NEJMoa1511014.

32 Nold, Christopher, Monique Maubert, Lauren Anton, Steven Yellon, and Michal A Elovitz. 2013. “Prevention of Preterm Birth by Progestational Agents: What Are the Molecular Mechanisms?” *American Journal of Obstetrics and Gynecology* 208 (3): 223.e1-7. doi:10.1016/j.ajog.2013.01.020.

33 Norgard, B, E Puho, A E Czeizel, M V Skriver, and H T Sorensen. 2005. “Aspirin Use during Early Pregnancy and the Risk of Congenital Abnormalities: A Population-Based Case-Control Study.” *Am J Obstet Gynecol* 192: 922–23.

34 Norman, Jane Elizabeth, Neil Marlow, Claudia-Martina Messow, Andrew Shennan, Phillip R Bennett, Steven Thornton, Stephen C Robson, et al. 2016. “Vaginal Progesterone Prophylaxis for Preterm Birth (the OPPTIMUM Study): A Multicentre, Randomised, Double-Blind Trial.” *The Lancet* 387 (10033): 2106–16. doi:10.1016/S0140-6736(16)00350-0.

35 Northen, A T, G S Norman, K Anderson, L Moseley, M Divito, M Cotroneo, M Swain, et al. 2007. “Follow-up of Children Exposed in Utero to 17 Alpha-Hydroxyprogesterone Caproate Compared with Placebo.” *Obstetrics and Gynecology* 110 (4). Department of Obstetrics and Gynecology, Center for Women’s Reproductive Health, University of Alabama at Birmingham, Birmingham, Alabama 35294-0024, USA. anorthen@uab.edu: 865–72. doi:10.1097/01.AOG.0000281348.51499.bc.

36 Orzechowski, Kelly M, Rupsa C Boelig, Jason K Baxter, and Vincenzo Berghella. 2014. “A Universal Transvaginal Cervical Length Screening Program for Preterm Birth Prevention.” *Obstetrics and Gynecology* 124 (3): 520–25. doi:10.1097/AOG.0000000000000428.

37 Owen, J, G Hankins, J D Iams, V Berghella, J S Sheffield, A Perez-Delboy, R S Egerman, et al. 2009. “Multicenter Randomized Trial of Cerclage for Preterm Birth Prevention in High-Risk Women with Shortened Midtrimester Cervical Length.” *American Journal of Obstetrics and Gynecology* 201 (4). Department of Obstetrics and Gynecology, University of Alabama at Birmingham, Birmingham, AL 35249-7333, USA.: 375.e1-375.e8. doi:10.1016/j.ajog.2009.08.015.

38 Reddy, Uma M, Alfred Z Abuhamad, Deborah Levine, and George R Saade. 2014. “Fetal Imaging: Executive Summary of a Joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-Fetal Medicine, American Institute of Ultrasound in Medicine, American College of Obstetricians and Gynecolog.” *Obstetrics and Gynecology* 123 (5): 1070–82. doi:10.1097/AOG.0000000000000245.

39 Rittenberg, Charles, Roger B Newman, Niki B Istwan, Debbie J Rhea, and Gary J Stanziano. 2009. “Preterm Birth Prevention by 17 Alpha-Hydroxyprogesterone Caproate vs. Daily Nursing Surveillance.” *The Journal of Reproductive Medicine* 54 (2): 47–52. http://www.ncbi.nlm.nih.gov/pubmed/19301566.

40 Rittenberg, Charles, Scott Sullivan, Niki Istwan, Debbie Rhea, Gary Stanziano, and Roger Newman. 2008. “Women Receiving 17-Alpha-Hydroxyprogesterone Caproate Hospitalized for Preterm Labor at Less than 34 Weeks Benefit from Daily Perinatal Nursing Surveillance.” *American Journal of Obstetrics and Gynecology* 199 (4): 389.e1-4. doi:10.1016/j.ajog.2008.06.070.

41 Roberge, S, K H Nicolaides, S Demers, P Villa, E Bujold, Roberge S., Nicolaides K.H., Demers S., Villa P., and Bujold E. 2013. “Prevention of Perinatal Death and Adverse Perinatal Outcome Using Low-Dose Aspirin: A Meta-Analysis.” *Ultrasound in Obstetrics & Gynecology : The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 41 (January): 491–99. doi:10.1002/uog.12421.

42 Finkelstein MO, Levin B and Robbins H. 1996. Clinical and prophylactic trials with assured new treatment for those at greater risk: I. A design proposal. Am J Public Health 85:691-695.

43 Saade, George R, Kim A Boggess, Scott A Sullivan, Glenn R Markenson, Jay D Iams, Dean V Coonrod, Leonardo M Pereira, et al. 2016. "Development and Validation of a Spontaneous Preterm Delivery Predictor in Asymptomatic Women." *American Journal of Obstetrics and Gynecology* 214 (5): 633.e1-633.e24. doi:10.1016/j.ajog.2016.02.001.

44 Saigal, Saroj, and Lex W Doyle. 2008. "An Overview of Mortality and Sequelae of Preterm Birth from Infancy to Adulthood." *Lancet* 371 (9608): 261–69. doi:10.1016/S0140-6736(08)60136-1.

45 Schisterman, Enrique F., Robert M. Silver, Laurie L. Lesher, David Faraggi, Jean Wactawski-Wende, Janet M. Townsend, Anne M. Lynch, Neil J. Perkins, Sunni L. Mumford, and Noya Galai. 2014. "Preconception Low-Dose Aspirin and Pregnancy Outcomes: Results from the EAGeR Randomised Trial." *The Lancet* 384: 29–36. doi:10.1016/S0140-6736(14)60157-4.

46 Seegmiller, R E, G W Nelson, and C K Johnson. 1983. "Evaluation of the Teratogenic Potential of Delalutin (17 Alpha-Hydroxyprogesterone Caproate) in Mice." *Teratology* 28 (2). UNITED STATES: 201–8. doi:10.1002/tera.1420280208.

47 Silver, Robert M, Katherine Ahrens, Luchin F Wong, Neil J Perkins, Noya Galai, Laurie L Lesher, David Faraggi, et al. 2015. "Low-Dose Aspirin and Preterm Birth: A Randomized Controlled Trial." *Obstetrics and Gynecology* 125 (4): 876–84. doi:10.1097/AOG.0000000000000736.

48 Son, Moeun, William A Grobman, Nina K Ayala, and Emily S Miller. 2016. "A Universal Mid-Trimester Transvaginal Cervical Length Screening Program and Its Associated Reduced Preterm Birth Rate." *American Journal of Obstetrics and Gynecology* 214 (3): 365.e1-5. doi:10.1016/j.ajog.2015.12.020.

49 Stinnett, A A, and J Mullahy. 1998. "Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis." *Medical Decision Making : An International Journal of the Society for Medical Decision Making* 18 (2 Suppl). ICOM Health Economics, Johnson & Johnson, Raritan, New Jersey, USA.: S68-80.

50 Voellmin, Annette, Sonja Entringer, Nora Moog, Pathik D Wadhwa, and Claudia Buss. 2013. "Maternal Positive Affect over the Course of Pregnancy Is Associated with the Length of Gestation and Reduced Risk of Preterm Delivery." *Journal of Psychosomatic Research* 75 (4): 336–40. doi:10.1016/j.jpsychores.2013.06.031.

51 Wadhwa, P D, T J Garite, M Porto, L Glynn, A Chicz-DeMet, C Dunkel-Schetter, and C A Sandman. 2004. "Placental Corticotropin-Releasing Hormone (CRH), Spontaneous Preterm Birth, and Fetal Growth Restriction: A Prospective Investigation." *American Journal of Obstetrics and Gynecology* 191 (4). Department of Psychiatry and Human Behavior, University of California-Irvine, Irvine, CA 92697-4260, USA. pwadhwa@uci.edu: 1063–69. doi:10.1016/j.ajog.2004.06.070.

52 Werler, Martha M, Jane E Sheehan, and Allen A Mitchell. 2002. "Maternal Medication Use and Risks of Gastroschisis and Small Intestinal Atresia." *American Journal of Epidemiology* 155 (1): 26–31. http://www.ncbi.nlm.nih.gov/pubmed/11772781.

53 Werner, Erika F, Maureen S Hamel, Kelly Orzechowski, Vincenzo Berghella, and Stephen F Thung. 2015. "Cost-Effectiveness of Transvaginal Ultrasound Cervical Length Screening in Singletons without a Prior Preterm Birth: An Update." American Journal of Obstetrics and Gynecology, June. doi:10.1016/j.ajog.2015.06.020.

54 Gold MR, Siegel JE, Russel LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

55 Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Am J Public Health. 1987;77:1417-26.52. Caro J, Ishak KJ, Migliaccio-Walle K. Estimating survival for cost-effectiveness analysis: a case study in atherothrombosis. Value in Health. 2004;7:627-635.

56 Little RJA, Rubin D.B. *Statistical Analysis with Missing Data*, New York: John Wiley & Sons, Inc. 1987

57 Rubin DB. *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc. 1987.

58 Schafer JL. *Analysis of Incomplete MultivariateData*, New York: Chapman and Hall. 1997.

59 Verbeke G, Melenbergh G. *Linear Mixed Models for Longitudinal Data*. Springer, 2009.

60 Van Buuren SV, Boshuizen HC, Knook DL. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis," *Statistics in Medicine*, 1999;18:681–694.

APPENDIX I:

DEFINITIONS OF NEONATAL MORBIDITY/MORTALITY

Neonatal Death and Stillbirth: Intrauterine fetal demise at a viable gestational age or neonatal death within 28 days of delivery

Intraventricular Hemorrhage (IVH): determined by cranial ultrasound or computed tomography

Grade I subependymal hemorrhage
Grade II intraventricular hemorrhage, uncomplicated
Grade III intraventricular hemorrhage with ventricular dilatation
Grade IV intraventricular hemorrhage with ventricular dilatation and parenchymal extension

Periventricular Leukomalacia (PVL): all PVL is determined by cranial ultrasound

- Any PVL
- Cystic PVL

Necrotizing Enterocolitis (NEC)

Stage I: Other – Suspect
•Treatment was observation
Stage II: Clinical – Definite
•Treatment was medical
Stage III: Surgical – Advanced
•Treatment was surgical

Respiratory Distress Syndrome (RDS): requires both diagnosis and oxygen therapy

- For the purpose of the study diagnosis must include:
  - oxygen therapy (FiO2 ≥ 0.40) until infant death or ≥ 24 hours **and**
  - a clinical diagnosis of RDS

Retinopathy of Prematurity (ROP)

- Stage I Ophthalmoscopic demarcation line of normal and abnormal vessels
- Stage II Intra-retinal ridge (ridge that rises from the retina as a result of the growth of the abnormal vessels)
- Stage III Ridge with extraretinal fibrovascular proliferation (the ridge grows from the spread of the abnormal vessels and extends into the vitreous)
- Stage III+ Stage III and "plus disease" meaning that the blood vessels of the retina have become enlarged and twisted, indicating a worsening of the disease.
- Stage IV Partially detached retina.
- Stage V Complete retinal detachment.

Bronchopulmonary Dysplasia (BPD)

- Treatment with > 21% oxygen for at least 28 days, **or**
- Oxygen dependence after 36 weeks post-conceptional age

Sepsis: Must include

- Blood culture proven sepsis, **and**
- a Clinically ill infant with infection defined as:
    - Bacterial sepsis of the newborn
    - Streptococcal sepsis
    - Severe sepsis

Neonatal Seizure

- Any incident(s) documented as evidence of seizure/epileptic activity by the neonatal staff,

DEFINITIONS OF COMPOSITE PERINATAL MORTALITY/NEONATAL MORBIDITY OUTCOME SCORES:

0 to 4 scale with NICU* (NMI): This score is defined as the following:
0 = no events,
1 = one event for (RDS, BPD, grade III or IV IVH, any PVL, proven sepsis, or NEC) or <1-4 days in the NICU, and no perinatal mortality;
2 = two events or between 5 and 20 days in the NICU, and no perinatal mortality;
3 = three or more events or >20 days in the NICU, and no perinatal mortality;
4 = perinatal mortality.

Morbidities with high likelihood of major chronic illness (MNM): This score is defined as the following:
0 = no events,
1 = one or more events for (cystic PVL, BPD, grade III or IV IVH, stage III or higher ROP, NEC requiring surgical treatment)

*Adapted from Hassan SS, et al. Ultrasound Obstet Gynecol 2011; 38:18-31
Supplementary Information