STATISTICAL ANALYSIS PLAN

A PHASE 3B, MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND STUDY OF HYDROXYPROGESTERONE CAPROATE INJECTION, 250 MG/ML, VERSUS VEHICLE FOR THE PREVENTION OF PRETERM BIRTH IN WOMEN WITH A PREVIOUS SINGLETON SPONTANEOUS PRETERM DELIVERY

PROTOCOL NUMBER: 17P-ES-003

29 JANUARY 2019

NCT01004029

STATISTICAL ANALYSIS PLAN

Version 2.0

A PHASE 3B, MULTI-CENTER, RANDOMIZED, DOUBLE-BLIND STUDY OF HYDROXYPROGESTERONE CAPROATE INJECTION, 250 MG/ML, VERSUS VEHICLE FOR THE PREVENTION OF PRETERM BIRTH IN WOMEN WITH A PREVIOUS SINGLETON SPONTANEOUS PRETERM DELIVERY

Protocol: 17P-ES-003

Prepared for:
AMAG Pharma USA, Inc.
1100 Winter Street
Waltham, MA 02451

Prepared by:

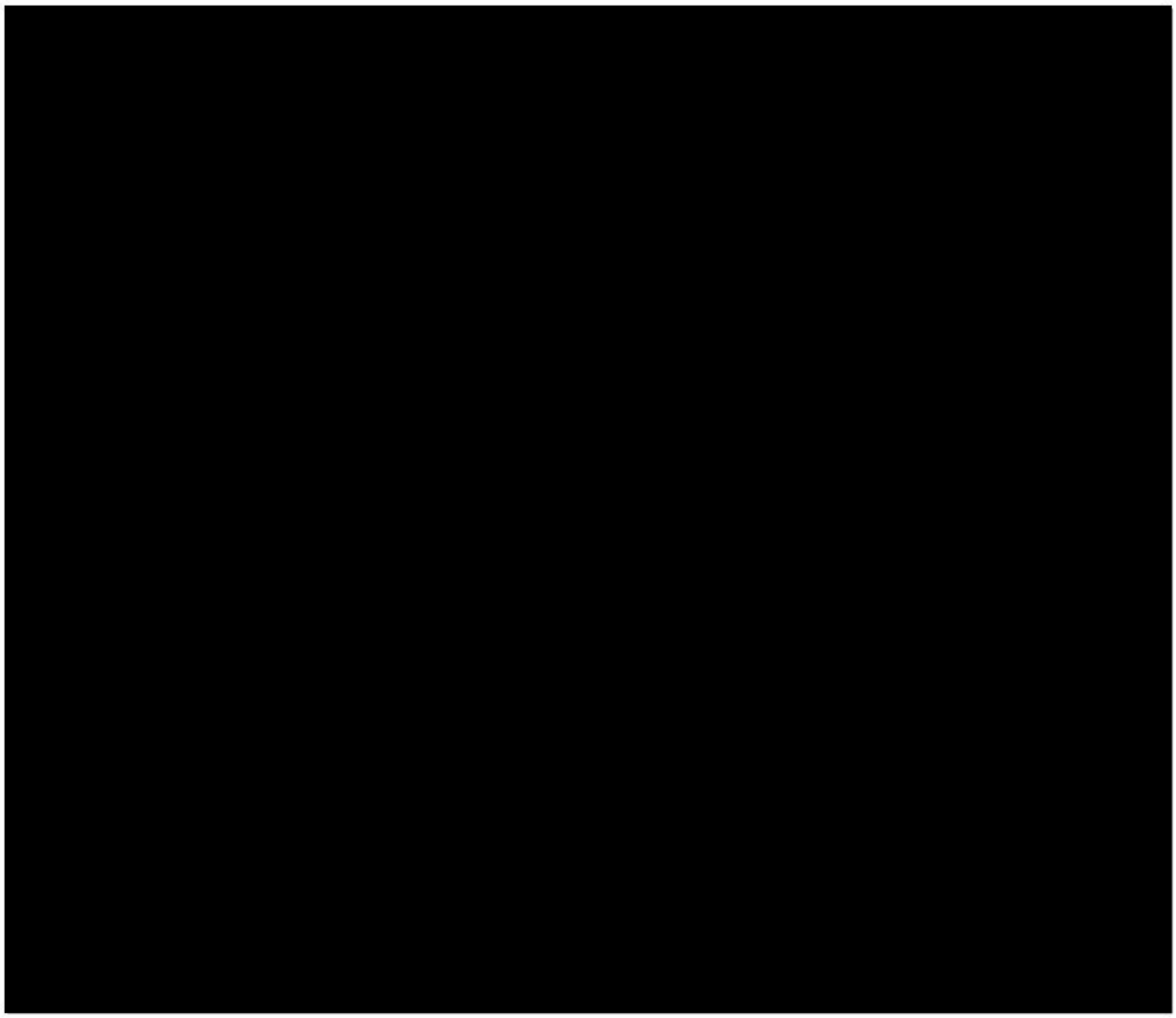
[redacted]

Revised by:

[redacted]

CONFIDENTIAL
January 29, 2019

ANALYSIS PLAN APPROVAL
Protocol Number: 17P-ES-003



ANALYSIS PLAN APPROVAL
Protocol Number: 17P-ES-003

ANALYSIS PLAN APPROVAL
Protocol Number: 17P-ES-003

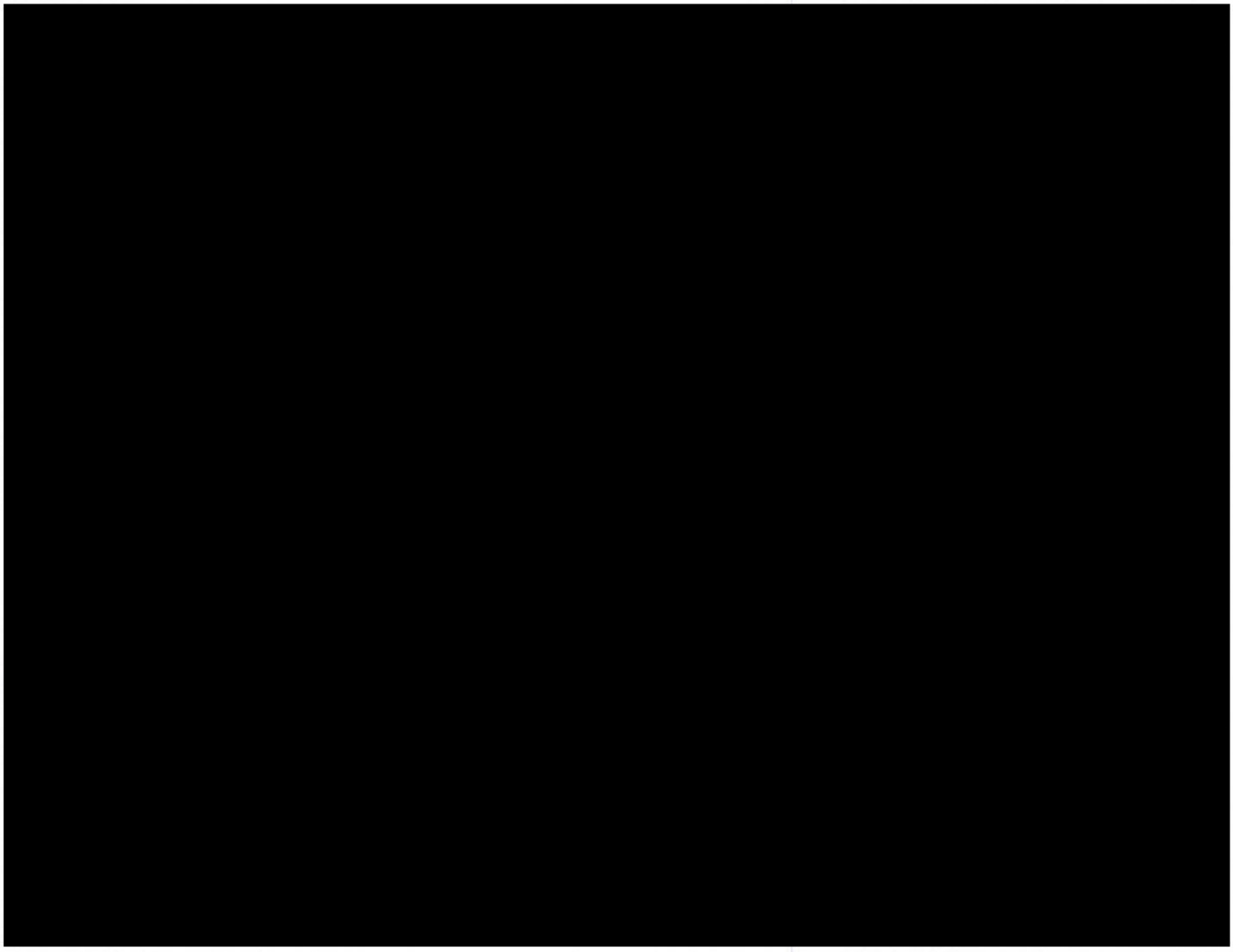


TABLE OF CONTENTS

1 Introduction

This Statistical Analysis Plan (SAP) provides the framework for the analysis and summarization of the clinical data from the study, "A Phase 3B, Multi-center, Randomized, Double-Blind Study of Hydroxyprogesterone Caproate Injection, 250 mg/ml, versus Vehicle for the Prevention of Preterm Birth in Women with a Previous Singleton Spontaneous Preterm Delivery" (protocol 17P-ES-003). Pharmacokinetic (PK) and pharmacodynamic (PD) analyses will not be presented in this SAP. Any changes made to the SAP after it has been signed-off but prior to database lock will be detailed in a Statement of Change document. Any changes made to the analyses after database lock will be described in the clinical study report (CSR).

2 Study Design

This study is a multi-center, randomized, double-blind, vehicle-controlled clinical trial in women with a singleton pregnancy, aged 18 years or older, with a history of a previous singleton spontaneous preterm delivery. A total of 1707 subjects will be randomized in a 2:1 ratio to receive either 17P or vehicle, respectively. Subjects who provide informed consent and are 15^0 to 20^3 weeks gestational age are eligible for screening. Potentially eligible subjects undergo the procedures as outlined in Appendix A at baseline (Visit 1) including receipt of a trial injection of vehicle. Approximately 7 days later and at 16^0 to 20^6 weeks gestational age, subjects who met all inclusion and have no exclusion criteria are randomized to study treatment. Subjects receive weekly injections of study drug from randomization (16^0 through 20^6 weeks of gestation) until 36^6 weeks of gestation or delivery, whichever occurs first. Randomized subjects will be followed for efficacy outcomes through the date of delivery and for adverse events up to the End of Treatment Period Visit defined as up to 35 ± 7 days after the last dose of study drug. If the End of Treatment Period Visit is prior to the date of delivery, maternal and fetal deaths will be reported until delivery. Neonates of randomized subjects will be followed until day 28 or the date of discharge from the NICU or equivalent, whichever occurs later. Discharge from the NICU is defined as discharge to any of the following: home, a non-medical facility, a chronic-care facility, or a step-down unit.

The schedule of assessments is provided in Appendix A.

3 Study Objectives

3.1 Primary Objective

The two co-primary objectives of this study are to:

- Determine if treatment with 17P reduces the rate of preterm birth < 35^0 weeks of gestation in women with a singleton pregnancy, aged 18 years or older, with a previous singleton spontaneous preterm delivery.

- Determine if 17P reduces the rate of neonatal mortality or morbidity. Neonatal mortality or morbidity is measured by a composite index comprised of:

    - Neonatal death.
    - Grade 3 or 4 intraventricular hemorrhage (IVH).
    - Respiratory distress syndrome (RDS).
    - Bronchopulmonary dysplasia (BPD).
    - Necrotizing enterocolitis (NEC).
    - Proven sepsis.

3.2 Secondary Objectives

The secondary objectives of this study are to:

- Exclude a doubling of the risk of fetal/early infant death, defined as spontaneous abortion/miscarriage (delivery from 16^0 through 19^6 weeks of gestation) or neonatal death (from minutes after birth until 28 days of life) occurring in liveborns born at less than 24 weeks of gestation or stillbirth (antepartum or intrapartum death from 20 weeks of gestation through term), in the 17P group compared to the vehicle group.

- Determine if 17P reduces the rate of preterm birth < 32^0 weeks of gestation.

- Determine if 17P reduces the rate of preterm birth < 37^0 weeks of gestation.

- Determine if 17P reduces the rate of stillbirth defined as all stillbirths/fetal deaths/in-utero fetal losses occurring from 20 weeks of gestation until term.

- Determine if 17P reduces the rate of neonatal death (from minutes after birth until 28 days of life) occurring in liveborns born at 24 weeks of gestation or greater.

- Evaluate the PK/PD of 17P in a subset of pregnant women (analysis not described in this SAP)

4 Gestational Age Determination

Gestational age is determined from the combination of a reliable menstrual history and measurements obtained at the subject's first ultrasound. It is determined at screening and is used to determine whether or not a subject is eligible for the study and as a randomization stratification factor using the categories 16^0 weeks – 17^6 weeks of gestation and 18^0 weeks – 20^6 weeks of gestation. For some subjects, after randomization, it was determined that gestational age was calculated incorrectly at screening, and the subject was assigned to the incorrect randomization stratum. Gestational age is then corrected; however, the subject's randomization stratum cannot be changed. To distinguish these two different gestational ages, the following terminology is used:

- Randomization gestational age stratum: refers to the gestational age stratum determined at screening and used for randomization. It had the categories 16^0 weeks

– 17^6 weeks of gestation and 18^0 weeks – 20^6 weeks of gestation. It does not vary over time and is not used for any purpose other than randomization.

- Project gestational age: refers to the correct gestational age calculated from the subject's menstrual history and measurements obtained at the subject's first ultrasound. In particular, project gestational age at the time of randomization will be used as an explanatory variable in statistical models (either as a continuous variable or as a categorical variable) or as a stratification variable in analyses (as a categorical variable, e.g., in Cochran-Mantel-Haenszel analyses), and project gestational age at the time of delivery will be used to define the primary and secondary efficacy outcomes (i.e., preterm birth prior to 35^0, 32^0, and 37^0 weeks of gestation). The terms "gestational age" and "weeks of gestation" will be used to refer to project gestational age.

5 Definitions of Outcome Measures

5.1 Efficacy Outcomes

5.1.1 Primary Efficacy Outcomes

There are two co-primary outcomes:

- Preterm birth prior to 35^0 weeks of gestation (as determined by project gestational age at time of delivery). All deliveries occurring from randomization up through 34^6 weeks of gestation, including miscarriages occurring from 16^0 through 19^6 weeks of gestation and elective abortions, will be included.

- Composite neonatal morbidity and mortality index, defined as a liveborn neonate with any of the following occurring at any time during the birth hospitalization up through discharge from the NICU (except as specified below):

    - Neonatal death occurring in a liveborn at any time through discharge from the NICU.
    - Grade 3 or 4 IVH: Bleeding from blood vessels in the periventricular matrix germinal matrix of the brain. Determined by cranial ultrasound, CT scan or MRI performed as part of routine clinical care. Grade III is defined as IVH with ventricular dilation and Grade 4 is defined as IVH with parenchymal extension.
    - RDS: A) PaO2 <50 mmHg in room air, central cyanosis in room air, a requirement for supplemental oxygen to maintain PaO2 >50 mmHg, or a requirement for supplemental oxygen to maintain a pulse oximeter saturation over 85% within the first 24 hours of life, AND B) A chest radiograph consistent with RDS within the first 24 hours of life.

- BPD: defined as having
    - Respiratory symptoms **and**
    - Chest radiograph abnormalities **and**
    - A requirement for supplemental oxygen in an infant born at <32 weeks who is now 36 weeks post menstrual age (PMA).
- NEC: defined by the presence of one or more of the following clinical signs: bilious gastric aspirate or emesis, abdominal distension, occult or gross blood in stool (no fissure) AND one or more of the following radiographic findings present: pneumatosis intestinalis (cystic or linear), hepato-biliary gas, pneumoperitoneum.
- Proven sepsis: If an infant has at least one of the following signs/symptoms or in the judgment of the investigator has other symptoms that are indicative of sepsis, the physician is to order appropriate cultures to document sepsis.

Sign/Symptom	**Blood Culture**	**Urine Culture**	**CSF Culture**
Fever (>38°C core)	X	X	X
Hypothermia (<37°C core)	X	X	X
Apnea	X	X	X
Bradycardia	X	X	X
Dysuria		X	
Lethargy		X	
Vomiting		X	
Stiff neck			X
Meningeal signs			X
Cranial nerve signs			X
Irritability			X

The infant must have a positive blood, urine or CSF culture to be recorded as proven sepsis. In a case of a positive blood culture, the infant must have either two cultures within a 3 calendar-day period showing the same organism or the investigator must document that in his/her judgment:

- Other causes of infection have been ruled out
- A single positive blood culture is not due to a skin contaminant

In the case of a positive urine culture, the culture must also show $\geq 10^5$ microorganisms per cc of urine and no more than two species of microorganisms.

If the infant does not have a positive blood, CSF or urine culture for the purposes of study documentation, a diagnosis of proven sepsis cannot be

made, even if the investigator makes a clinical diagnosis of sepsis and treats the subject with antibiotics.

5.1.2 Secondary Neonatal Efficacy Outcomes

Secondary neonatal outcomes that will be measured include:

- Neonatal death (from minutes after birth until 28 days of life) occurring in liveborns born at 24 weeks of gestation or greater.

5.1.3 Secondary Maternal Efficacy Outcomes

Secondary maternal efficacy outcomes that will be measured include:

- Preterm birth prior to 32^0 weeks of gestation (as determined by project gestational age at time of delivery). All deliveries occurring from randomization up through 31^6 weeks of gestation, including miscarriages occurring from 16^0 through 19^6 weeks of gestation and elective abortions, will be included.

- Preterm birth prior to 37^0 weeks of gestation (as determined by project gestational age at time of delivery). All deliveries occurring from randomization up through 36^6 weeks of gestation, including miscarriages occurring from 16^0 through 19^6 weeks of gestation and elective abortions, will be included.

5.1.4 Additional Neonatal Efficacy Outcomes

Additional neonatal efficacy outcomes that will be measured include:

- The following individual components of the neonatal morbidity and mortality index:
    - IVH
    - RDS
    - BPD
    - NEC
    - Proven sepsis
- Birth weight
- Seizures
- Retinopathy of prematurity (ROP)
- Patent ductus arteriosus (PDA)
- Infant hospital days, defined as the number of days from birth to hospital discharge.

- Number of days of neonatal respiratory therapy, defined as the number of days on ventilator support and/or oxygen therapy.

- Transient tachypnea

- Persistent pulmonary hypertension

5.1.5 Additional Maternal Efficacy Outcomes

Additional maternal efficacy outcomes that will be measured include:

- Spontaneous preterm birth prior to 37^0 weeks of gestation and prior to 35^0 weeks of gestation. Spontaneous delivery is defined as following preterm premature rupture of membranes (pPROM) or spontaneous labor from 20^0 weeks up through 36^6 weeks of gestation (for prior to 37^0 weeks) and 20^0 weeks up through 34^6 weeks of gestation (for prior to 35^0 weeks) or miscarriage from 16^0 weeks through 19^6 weeks of gestation.

- Indicated preterm birth prior to 37^0 weeks of gestation for either fetal or maternal indications. Elective abortions will also be defined as indicated preterm births.

- Gestational age at delivery. Gestational age at delivery is calculated from the Project Estimated Date of Confinement (Project EDC) captured on the CRF and the date of delivery as follows:

  280 – (Project EDC – Date of Delivery) and is presented in weeks, days

- Spontaneous abortion/miscarriage defined as delivery from 16^0 through 19^6 weeks of gestation.

5.2 Safety Outcomes

5.2.1 Primary Safety Outcome

The primary safety outcome is fetal/ early infant death.

Fetal/early infant death will be defined as:

- Spontaneous abortion/miscarriage (delivery from 16^0 up through 19^6 weeks of gestation) or;

- stillbirth (antepartum or intrapartum death) from 20 weeks of gestation through term, or;

- neonatal death (from minutes after birth until 28 days of life) occurring in liveborns born at less than 24 weeks of gestation.

5.2.2 Secondary Safety Outcomes

The secondary safety outcome is stillbirth defined as all stillbirths/fetal deaths/in-utero fetal losses occurring from 20 weeks of gestation until term.

Safety will also be assessed by analysis of the occurrence of treatment emergent adverse events (TEAEs) and maternal pregnancy complications. A TEAE is defined as an adverse event (AE) that starts or worsens at or during the time of the first blinded study medication injection (active drug or vehicle) through the End of Treatment Period Visit. Adverse events will also be provided for the time period starting from the trial injection and ending on the day of, but prior to, administration of the first study medication injection.

Maternal pregnancy complications include the following:

- Gestational diabetes: Any degree of glucose intolerance with onset or first recognition during pregnancy. A fasting plasma glucose level > 126 mg/dL (7.0 mmol/L) or random plasma glucose > 200 mg/dL (11.1 mmol/L) meets the threshold for the diagnosis of diabetes, if confirmed on a subsequent day, and precludes the need for any glucose challenge.[1]
- Oligohydramnios: Using ultrasound, amniotic fluid index (the sum of measurements of the deepest cord-free amniotic fluid pocket in each of the abdominal quadrants) less than 5 cm or if clinically diagnosed.
- Significant antepartum bleeding or hemorrhage: Bleeding after the subject was randomized, including placenta previa, abruption placenta, or threatened abortion. Spotting is not considered significant.
- Preeclampsia: Within a 24-hour period having blood pressure measurements ≥ 140/90 mmHg on at least 2 occasions (4 or more hours apart) and proteinuria of 0.3 g or greater in a 24-hour urine specimen or +1 or greater protein on urine dipstick.
- Eclampsia: Occurrence of generalized convulsion and/or coma in the setting of preeclampsia, with no other neurological condition.
- HELLP syndrome: Meets all three of the following conditions:

  Hemolysis: abnormal peripheral smear, increased bilirubin > 1.2 mg/dL

  Elevated liver enzymes: aspartate aminotransferase (AST) ≥72 IU/L, lactate dehydrogenase (LDH) > 600 IU/L; and

  Thrombocytopenia: platelet count < 100,000/mm^3.
- Gestational hypertension: Having 2 blood pressure measurements ≥ 140/90 mmHg on at least 2 occasions (4 or more hours apart) and a clinical diagnosis of gestational hypertension (without proteinuria or other signs of preeclampsia).
- Abruption: Clinical diagnosis of placental abruption (retroplacental hematoma). Does not include abruption diagnosed by pathologist's report.

- Chorioamnionitis: Clinical diagnosis of chorioamnionitis and body temperature of ≥ 100° F (37.8° C) and no other defined infection. Does not include chorioamnionitis diagnosed only by a placental pathology report.

6 Statistical Methods

6.1 Sample Size

In 3 studies of high-risk pregnant women, the rate of preterm birth < 35^0 weeks of gestation in women receiving vehicle ranged from 26.5% to 30%.[2,3,4] The NICHD study also found that 17.2% of live born infants in women receiving vehicle had the neonatal composite index. Using a 2:1 randomization of subjects to 17P and vehicle, respectively, a total of 1665 liveborn infants are required to detect a reduction of 35% in the rate of the composite index (from 17% to 11%) with a power of 90% (assuming a two-sided type I error of 5%). Assuming 2.5% of pregnancies will result in miscarriage or stillbirth, an additional 42 women need to be enrolled, for a total of 1707 women (1138 active and 569 placebo). A total sample size of 1707 subjects provides 98% power to detect a reduction of approximately 30% in the rate of preterm birth <35^0 weeks of gestation (from 30% to 21%) using a two-sided type I error of 5%. The effect size for the neonatal composite index as well as preterm birth <35^0 weeks gestation was chosen to represent a clinically significant reduction.

Since these outcome measures are co-primary outcomes, the power to detect statistically significant differences between the treatment groups for *both* outcome measures may be reduced. If the outcome measures are independent, the power is 88.2% and if the outcome measures are correlated as highly as possible, the power is 90%. Data from the NICHD Study indicate these outcome measures are highly correlated with 56% of liveborn infants of women who delivered <35^0 weeks of gestation with the neonatal composite index compared with 2% of liveborn infants of women who delivered ≥35^0 weeks of gestation. Thus, the power to detect differences between the treatment groups for *both* outcome measures is expected to be close to 90%.

There is also sufficient power to detect clinically significant reductions in the secondary outcomes of delivery <32^0 weeks of gestation and < 37^0 weeks of gestation as indicated in Table 1 Sample Size Calculation.

Table 1: Sample Size Calculation

Secondary Outcome	Outcome Rate in Vehicle Group	Percentage Reduction	Power
Delivery < 32^0 weeks of gestation	20%	33%	92%
Delivery < 37^0 weeks of gestation	40%	33%	>99%

Assuming a 4% fetal/early infant death rate in both treatment arms with a two-sided alpha of 5%, a sample size of 1707 subjects provides 82.8% power to rule-out a doubling in the risk of fetal/early infant death (i.e., the upper bound of the confidence interval for the relative risk of 17P compared to vehicle will be <2.0).

A fetal/early infant death rate of 4.0% is based on the results of Study 17P-CT-002 (the NICHD 17P trial).

6.2 Randomization and Blinding

Subjects will be randomly assigned in a 2:1 ratio to 17P or vehicle treatment (two 17P subjects to each vehicle subject) using a blocked randomization sequence stratified by study site and randomization gestational age stratum (16^0 weeks - 17^6 weeks gestation and 18^0 weeks – 20^6 weeks gestation). Randomization is stratified by study site to ensure balance between the 2 treatment groups with respect to anticipated differences in the clinic population and possible site-to-site differences in subject management. The rationale for 2:1 randomization is that because the study involves weekly injections, subjects may be more willing to participate if they know that they have a 2 to 1 chance of receiving active medication. Randomization will be performed through the use of an interactive voice response system (IVRS).

This will be a double-blind study. The subjects, clinical site staff, and sponsor will not be aware of the treatment assignment. A Data and Safety Monitoring Board (DSMB) will review unblinded summary safety data.

6.3 Interim Analysis

During the trial, an external and independent DSMB will meet periodically to review safety data. Because the DSMB will not be reviewing efficacy data to determine whether the study should be stopped for efficacy, no adjustment to the alpha level is required. The timing of the DSMB reviews and the scope of the safety review are detailed in the DSMB Charter. Unblinded data are reviewed by the DSMB.

6.4 Analysis Populations

Five analysis populations are defined:

All Enrolled Population: The All Enrolled Population will consist of all subjects who receive a trial injection.

Intent-to-Treat (ITT) Population: The ITT Population will consist of all randomized subjects. All subjects will be analyzed in the treatment group to which they were randomized regardless of whether the subject received the correct blinded study medication (active or vehicle).

Per-Protocol (PP) Population: The PP Population will consist of all subjects who are compliant with the study protocol. Each subject will be classified as compliant or not

compliant with the protocol based on the following compliance criteria: subject did not have a major protocol deviation potentially affecting efficacy or the evaluation of efficacy as determined by the Sponsor in a blinded review, received the correct blinded study medication for the majority of the duration of study drug receipt, at least 90% compliant with study medication (based on receipt of study medication through 36^{6} weeks of gestation or delivery, whichever occurs first), and had outcome data available.

Safety Population: The Safety Population will consist of all subjects who received at least one dose of blinded study medication. Subjects who received the wrong blinded study medication will be analyzed in the treatment group based on the medication received.

Liveborn Neonatal Population: The Liveborn Neonatal Population will consist of all babies of randomized women (ITT Population) who were liveborn and have morbidity data available.

6.5 Comments on Statistical Analysis

- Inferential statistical analyses as specified will be conducted and all comparisons will be between the 17P and vehicle treatments.

- An alpha level of 0.05 will be used for the primary and secondary efficacy analyses. Additional efficacy analyses are provided as supportive analyses to the primary and secondary efficacy analyses and no conclusion of statistical significance will be made. Thus, no adjustment to the alpha level is required for the additional efficacy analyses. Unless specified otherwise, all p-values will be presented to three significant digits.

- Unless specified otherwise, data will be pooled across study sites for all statistical analyses.

- Descriptive statistics, including the numbers and percentages for categorical variables, and the number of subjects, mean, standard deviation, median, minimum, and maximum for continuous variables will be provided.

- Listings of individual subject's clinical data will be provided.

- Study day will be calculated as the date of the AE/medication/assessment minus the date of the first dose of blinded study medication (active or vehicle), plus one.

- Version 9.4 (or higher) of SAS statistical software package will be used to provide all summaries, listings, and statistical analyses.

6.6 Handling of Missing Data

- All missing and partial dates for events occurring after randomization or for medications received after randomization will be queried for a value. If no value can be obtained, substitutions will be made as follows:

- For start dates, missing months and days will be defined by "01", as long as this occurs on or after the first dose of blinded study medication. If the algorithm produces a date prior to the first dose of blinded study medication, the date of the first dose of blinded study medication will be used for the partial date. For stop dates, missing months will be defined as "12" and days will be defined by the last day of the respective month. If the date is associated with a maternal event and the algorithm produces a date after delivery, the date of delivery will be used for the partial date. If the date is associated with a neonatal event and the algorithm produces a date after 28 days of life or discharge from the NICU, whichever is later, the date associated with 28 days of life or discharge from the NICU will be used for the partial date. These substitutions will be used in calculations; however, the actual value recorded on the CRF will be used for all listings.

- Missing data for the severity and causality assessment for adverse events and maternal complications will be queried for a value. If severity remains missing, the adverse event/maternal complication will be considered severe. If causality remains missing, the adverse event/maternal complication will be considered definitely related to study drug.

- The primary efficacy analysis of delivery $<35^0$ weeks gestation will utilize the ITT Population. Missing primary outcome data will be imputed with multiple imputation methods as follows:

  If the data initially have a monotone missing data pattern, the missing data will be imputed using SAS® PROC MI with the MONOTONE REG option for continuous variables and the MONOTONE LOGISTIC option for categorical variables. One hundred imputed datasets will be created.

  If the data do not initially have a monotone missing data pattern, a two-step imputation procedure will be performed. First, enough missing data will be imputed for the independent variables using the MCMC option of SAS® PROC MI to produce a monotone missing data pattern. One hundred imputed datasets will be created in this first step. Next, the missing data in the 100 resulting datasets will be imputed using SAS® PROC MI with the MONOTONE REG option for continuous variables and the MONOTONE LOGISTIC option for categorical variables. A single imputation for each of the 100 datasets from the first step will be done in the second step resulting in a total of 100 datasets with complete data.

  The model for the MONOTONE REG/LOGISTIC option(s) will include terms for treatment, project gestational age at randomization, the interaction of treatment and project gestational age at randomization, the number of previous preterm deliveries, and gestational age at qualifying delivery included as predictive variables. Project gestational age at randomization will be treated as a categorical variable with categories of week, i.e., 16^0 - 16^6, 17^0 - 17^6, 18^0 - 18^6, 19^0 - 19^6, and 20^0 - 20^6. Subjects with a project gestational age at randomization $< 16^0$ and subjects with a project gestational age at randomization $> 20^6$ will be included in the 16^0 - 16^6 and 20^0 - 20^6 categories, respectively. If a subject is known to still be pregnant on or after 35^0 weeks gestation, she will be considered not to

have delivered prior to 35 weeks even if her delivery date is unknown. The logistic regression analysis used to assess whether there is an interaction between treatment and project gestational age at the time of randomization and the logistic regression analysis including prognostic factors for which there is an imbalance at baseline between the treatment groups will be based on these same imputed datasets.

- The primary efficacy analysis of neonatal composite index will utilize the Liveborn Neonatal Population. By definition, neonates in this population must have morbidity data available. If any component of the neonatal composite index is missing and the non-missing component data do not definitively indicate the presence or absence of the index, a multiple imputation analysis will be used to impute the missing composite index value as follows:

  If the data initially have a monotone missing data pattern, the missing data will be imputed using SAS® PROC MI with the MONOTONE REG option for continuous variables and the MONOTONE LOGISTIC option for categorical variables. One hundred imputed datasets will be created.

  If the data do not initially have a monotone missing data pattern, a two-step imputation procedure will be performed. First, enough missing data will be imputed for the independent variables using the MCMC option of SAS® PROC MI to produce a monotone missing data pattern. One hundred imputed datasets will be created in this first step. Next, the missing data in the 100 resulting datasets will be imputed using SAS® PROC MI with the MONOTONE REG option for continuous variables and the MONOTONE LOGISTIC option for categorical variables. A single imputation for each of the 100 datasets from the first step will be done in the second step resulting in a total of 100 datasets with complete data.

  The model for the MONOTONE REG/LOGISTIC option(s) will include terms for treatment, project gestational age at randomization, the interaction of treatment and project gestational age at randomization, the number of previous preterm deliveries, and gestational age at qualifying delivery included as predictive variables. Project gestational age at randomization will be treated as a categorical variable with categories of week, i.e., 16^0 - 16^6, 17^0 - 17^6, 18^0 - 18^6, 19^0 - 19^6, and 20^0 - 20^6. Subjects with a project gestational age at randomization < 16^0 and subjects with a project gestational age at randomization > 20^6 will be included in the 16^0 - 16^6 and 20^0 - 20^6 categories, respectively. The logistic regression analysis used to assess whether there is an interaction between treatment and project gestational age at the time of randomization and the logistic regression analysis including prognostic factors for which there is an imbalance at baseline between the treatment groups will be based on these same imputed datasets. Sensitivity analyses will be conducted whereby neonates with missing data are assumed not to have the neonatal composite index and with these neonates excluded from the analysis.

- For the secondary efficacy analyses of delivery prior to 37^0 weeks of gestation and delivery prior to 32^0 weeks of gestation, and additional analyses of spontaneous preterm birth prior to 37^0 weeks of gestation, spontaneous preterm birth prior to 35^0 weeks of

gestation, and indicated preterm birth prior to 37^0 weeks of gestation, missing data will be imputed using the same methodology as for the primary efficacy outcome analysis.

- The safety analysis of fetal/early infant death will be conducted in the ITT Population. Multiple imputation methods assuming a monotone missing data pattern will be used to impute the missing data for fetal/early infant death. One hundred imputed datasets will be created using a logistic regression model with treatment, project gestational age at randomization, the interaction of treatment and project gestational age at randomization, and the number of previous preterm deliveries included as predictive variables. Project gestational age at randomization will be treated as a categorical variable with categories of week, i.e., 16^0 - 16^6, 17^0 - 17^6, 18^0 - 18^6, 19^0 - 19^6, and 20^0 - 20^6. Subjects with a project gestational age at randomization < 16^0 and subjects with a project gestational age at randomization > 20^6 will be included in the 16^0 - 16^6 and 20^0 - 20^6 categories, respectively. Sensitivity analyses will be conducted whereby subjects with missing data are assumed not to have a fetal/early infant death and with these subjects deleted from the analysis. The analyses will also be conducted in the Safety Population.

- Missing values for other individual data points will remain as missing. For these data points, missing values will not be imputed and only observed values will be used in data analyses and presentations.

- Where individual data points are missing, categorical data will be summarized based on reduced denominators (i.e., only subjects with available data will be included in the denominators).

7 Statistical Analyses

7.1 Subject Population and Baseline Characteristics

The total number of subjects who received the trial injection (enrolled subjects) and who were randomized will be summarized by geographic region, country, and site, by treatment group and overall. A summary of the reasons for not randomizing subjects who received the trial injection will also be provided for all enrolled subjects. The demographic characteristics (age, ethnicity, and race) of all enrolled subjects will be presented separately for subjects who were randomized and subjects who were not randomized. The number of randomized subjects included in each of the analysis populations (All Enrolled, ITT, PP, Safety, and Liveborn Neonatal), and the primary reason for exclusion from the PP population will be summarized by treatment group.

The total number of subjects in the ITT Population with a protocol deviation will be summarized by geographic region, site, and treatment group. A distribution of the type of protocol deviation (e.g., inclusion criteria, exclusion criteria, study drug, etc.) will be summarized by treatment group. Protocol deviations will also be provided in a listing. The number and percentage of randomized subjects who withdrew from the study and the primary reason for withdrawal will be presented by treatment group. A subject is considered withdrawn from the study if the subject delivery data are not obtained. The number and

percentage of randomized subjects who discontinued study medication prior to 37^0 weeks of gestation or delivery (whichever came first) and the primary reason for discontinuation will be presented by treatment group. Statistically significant differences between treatment groups in the percentages of subjects who withdrew from the study and who discontinued study medication prior to 37^0 weeks of gestation or delivery will be determined using Fisher's exact test.

Demographics (age, ethnicity, race, pre-pregnancy weight, and pre-pregnancy BMI) will be summarized by treatment group for the ITT Population, as will the subject's social history (marital status, years of education, and use of cigarettes, alcohol and street drugs during pregnancy). Subject's obstetrical characteristics, project gestational age at randomization (as a continuous variable and also categorized as $<16^0$ weeks, $16^0 – 17^6$ weeks, $18^0 – 20^6$ weeks and $>20^6$ weeks), randomization gestational age stratum (categorized as $16^0 – 17^6$ weeks and $18^0 – 20^6$ weeks), and cervical length by ultrasound (for those patients with data available) and history of douching will be summarized by treatment group. Project gestational age at randomization is calculated from the Project Estimated Date of Confinement (Project EDC) captured on the CRF and the date of randomization as follows:

(280 – (Project EDC – Date of Randomization))/7.

A table will provide data on previous pregnancy history including descriptive statistics of the number of previous preterm deliveries, number of previous spontaneous preterm deliveries, miscarriages, and stillbirths, as well as the percentage of patients with 1 and >1 previous preterm deliveries, none, 1 and >1 miscarriages, and none, 1 and >1 stillbirths. The percentage of subjects with a prior spontaneous preterm delivery (following PTL or pPROM) and a prior indicated (for maternal or fetal indications) preterm delivery will be provided. The gestational age of the qualifying delivery and the percentage of subjects with a qualifying delivery at <37 weeks, <35 weeks, <32 weeks, and <28 weeks will be provided. Descriptive statistics of the gestational age at the earliest (i.e., earliest gestational age) prior spontaneous preterm delivery will also be provided.

The chi-square test or Fisher's exact test will be used for dichotomous variables and the Wilcoxon Rank Sum test for ordinal and continuous variables to test for differences between treatment groups.

Obstetrical complications occurring prior to randomization and during the current pregnancy, relevant and significant obstetrical or gynecological surgical procedures occurring prior to randomization, and other relevant and significant medical history will be summarized in the ITT Population by MedDRA system organ class and higher level or preferred term, as appropriate. A summary of abnormal findings from the physical examination at Visit 2 will be provided by body site and treatment group.

7.2 Efficacy Analyses

7.2.1 Primary Efficacy Analyses

The primary hypothesis for efficacy compares the percentage of subjects with a preterm delivery < 35^0 weeks of gestation between the 17P and vehicle treatment groups in the ITT Population and the percentage of neonates with the neonatal composite index between the 17P and vehicle treatment groups in the Liveborn Neonatal Population. The null hypothesis (H_O) is that 1) there is no difference between the 17P ($\prod_{17P}$) and vehicle ($\prod_V$) treatments in the percentage of subjects with a preterm delivery < 35^0 weeks of gestation in the ITT population; and 2) there is no difference between the 17P (P_{17P}) and vehicle (P_V) treatments in the percentage of neonates with the neonatal composite index in the Liveborn Neonatal Population, while the alternative hypothesis (H_A) is that there is a difference between the treatments for at least one of the endpoints. The null and alternative hypotheses are as follows:

H_{O1}: $\prod_{17P} = \prod_V$

$H_{A1:}$ $\prod_{17P} \neq \prod_V$, and

H_{O2}: $P_{17P} = P_V$

$H_{A2:}$ $P_{17P} \neq P_V$

An alpha level of 0.05 will be used for the primary analysis of both primary outcomes, as an adjustment for multiple comparisons is not required for testing the null hypotheses when stated as above.

Statistically significant differences between the 17P and vehicle treatments in the percentage of subjects who deliver prior to 35^0 weeks gestation will be determined using a Cochran-Mantel-Haenszel (CMH) test stratified by project gestational age at randomization (16^0 weeks - 17^6 weeks gestation and 18^0 weeks – 20^6 weeks gestation). Subjects with missing delivery data will have delivery <35^0 weeks and ≥ 35^0 weeks of gestation imputed with multiple imputation methods as described in Section 6.6. The combined CMH statistic from the imputed datasets will be determined using the Wilson-Hilferty transformation.

For each treatment group, the number and percentage of subjects with a preterm delivery at < 35^0 weeks of gestation will be presented for each project gestational age at randomization stratum and overall. Kaplan-Meier curves of the time from randomization to delivery will also be provided for each project gestational age at randomization stratum and overall for each treatment group.

The number and percentage of neonates in the Liveborn Neonatal Population with the neonatal composite index will be presented by project gestational age at randomization stratum and overall for each treatment group. Statistically significant differences between the 17P and vehicle treatments will be determined using the Cochran-Mantel-Haenszel test stratified by project gestational age at randomization. Subjects with missing neonatal

composite index data will have their data imputed with multiple imputation methods as described in Section 6.6 and the combined CMH statistic from the imputed datasets will be determined using the Wilson-Hilferty transformation.

7.2.1.1 Additional Analyses of Primary Efficacy Outcome Measures

An analysis will be conducted in the ITT population where the neonatal composite index is defined with all deaths including miscarriages, abortions and stillbirths included as death in the composite index. The number and percentage of neonates with the neonatal composite index will be presented by project gestational age at randomization stratum and overall for each treatment group. Additional sensitivity analyses will be conducted whereby neonates with missing data are assumed not to have an event in the neonatal composite index and neonates with missing data are excluded from the analysis. For each sensitivity analysis, statistically significant differences between the 17P and vehicle treatments will be determined using the Cochran-Mantel-Haenszel test stratified by project gestational age at randomization.

A logistic regression model of preterm delivery < 35^0 weeks of gestation with terms for treatment, project gestational age at randomization stratum, and treatment-by-project gestational age at randomization stratum interaction will be used to assess whether there is an interaction between treatment and project gestational age at the time of randomization. A similar analysis will be performed for the neonatal composite index. For these analyses, subjects with missing delivery data or neonatal composite index data will have their data imputed with multiple imputation methods as described in Section 6.6.

For each treatment group, the number and percentage of subjects with a delivery at <35^0 weeks by gestational age at the qualifying delivery (20^0 –<28 weeks, 28^0 – <32 weeks, 32^0 – <35 weeks, and 35^0 – <37 weeks), by gestational age of the earliest prior preterm birth (20^0 – <28 weeks, 28^0 – <32 weeks, 32^0 – <35 weeks, and 35^0 – <37 weeks), by number of previous preterm deliveries (1, 2 and ≥3 prior preterm deliveries), by country, and by region (US and non-US) will be presented. For each treatment group, the number and percentage of neonates with the composite index by gestational age at the qualifying delivery, by gestational age at the earliest prior preterm birth, by the number of previous preterm deliveries, by country, and by geographic region will also be presented. The qualifying delivery is the previous singleton spontaneous preterm delivery that meets inclusion criterion number 4. If more than one previous preterm delivery meets inclusion criterion number 4, the qualifying delivery is the most recent preterm delivery.

If there are baseline imbalances between the treatment groups with respect to the prognostic factors (gestational age at qualifying delivery and number of previous preterm deliveries), an adjusted analysis of preterm delivery <35^0 weeks of gestation will be conducted using a logistic regression model. The logistic regression model will include terms for project gestational age at randomization, the prognostic factor(s), and treatment group. The p-value for treatment from the logistic regression model will be provided. Similarly, for the neonatal composite index, a logistic regression analysis will be conducted including terms for project gestational age at randomization, the prognostic factor(s), and treatment group. The p-value

for treatment from the logistic regression model will be provided. For these analyses, subjects with missing delivery data or neonatal composite index data will have their data imputed with multiple imputation methods as described in Section 6.6.

The percentage of subjects with a preterm birth $<35^0$ weeks of gestation will also be analyzed for the PP Population using the same analytic method as described for the primary efficacy analysis in the ITT Population. The percentage of neonates with the neonatal composite index will also be analyzed for the PP Population using the same analytic method as described for the primary efficacy analysis in the Liveborn Neonatal Population.

Exploratory subgroup analyses may also be conducted. These analyses will be performed for black and nonblack subjects and also for subjects with a cervical length by ultrasound <25 mm at baseline and those for whom the cervical length is ≥25 mm. The same analytic method as described for the primary efficacy analysis will be utilized.

7.2.2 Secondary Neonatal Efficacy Analyses

The number and percentage of neonatal deaths (from minutes after birth until 28 days of life) for neonates in the Liveborn Neonatal Population born at 24 weeks gestational age or greater will be presented by project gestational age at randomization stratum (16^0 weeks - 17^6 weeks gestation and 18^0 weeks – 20^6 weeks gestation) and overall, for each treatment group. Statistically significant differences between the 17P and vehicle treatments will be determined using the Cochran-Mantel-Haenszel test stratified by project gestational age at randomization.

7.2.3 Secondary Maternal Efficacy Analyses

Analyses of the secondary maternal outcomes (delivery prior to 37^0 weeks and prior to 32^0 weeks of gestation) will be conducted using the ITT and PP Populations. The proportions of subjects with a preterm birth $<37^0$ weeks and $<32^0$ weeks of gestation will be presented for each project gestational age at randomization stratum (16^0 weeks - 17^6 weeks gestation and 18^0 weeks – 20^6 weeks gestation)and overall, by treatment group. Statistically significant differences between treatments will be determined using the Cochran-Mantel-Haenszel test stratified by project gestational age at randomization. Subjects with missing delivery data will have preterm birth $<37^0$ weeks and $<32^0$ weeks of gestation imputed with multiple imputation.

The number and percentage of deliveries $<37^0$ weeks of gestation and $<32^0$ weeks of gestation will be presented by country.

7.2.4 Additional Neonatal Efficacy Analyses

A summary of neonatal morbidity data will be provided for the Liveborn Neonatal Population. The number and percentage of neonates with IVH (any IVH and grade III/IV IVH), RDS, BPD, NEC, proven sepsis, confirmed pneumonia, seizures, ROP (any ROP and grade 3/4/5 ROP), PDA, transient tachypnea, cystic periventricular leukomalacia, other

intercranial hemorrhage, persistent pulmonary hypertension and neonatal hypoglycemia will be provided by treatment group. Liveborn neonates not admitted to the NICU will be defined as not having any of the above listed morbidities. Statistically significant differences between the 17P and vehicle treatments will be determined using the Cochran-Mantel-Haenszel test stratified by project gestational age at randomization stratum. A distribution of the final status of the neonate (i.e., NICU outcome) will also be presented. The number and percentage of neonates on ventilator support or receiving supplemental oxygen will be provided for each treatment group, as will descriptive statistics of the number of days of respiratory therapy. Continuous variables will be tested using the Van Elteren test stratified by project gestational age at randomization stratum.

Descriptive statistics (mean, standard deviation, median, minimum and maximum) will be calculated by treatment group for birth weight, length, head circumference, APGAR scores (1, 5 and 10 minutes), the number of total infant hospital days (only for infants who did not die) and the number of NICU days. The number of total infant hospital days is defined as the time (in days) from birth to discharge to home, a chronic care facility, or a non-medical facility. For those infants who die, the number of NICU days will be set to the longest number of days. Since the analysis is nonparametric, this codes the deaths as the worst possible outcome. Continuous variables (other than the number of NICU days) will be tested using the Van Elteren test stratified by project gestational age at randomization stratum. The number of NICU days is defined as the time in days from admission to the NICU to discharge. If a neonate is admitted more than once to the NICU, the number of NICU days for each NICU stay will be calculated separately; the neonate's number of NICU days will then be calculated as the sum of NICU days from all NICU stays. Neonates not admitted to the NICU will be assigned a value of 0 days, and neonates who died will be censored at the date of death. A Kaplan-Meier analysis will be conducted, and statistically significant differences between treatments will be determined using a stratified (for project gestational age at randomization stratum) log-rank test. The numbers and percentages of neonates admitted to the NICU and with a major congenital malformation will be provided.

7.2.5 Additional Maternal Efficacy Analyses

Analyses of the secondary maternal outcomes of spontaneous preterm birth prior to 37^0 weeks of gestation, spontaneous preterm birth prior to 35^0 weeks of gestation, and indicated preterm birth prior to 37^0 weeks of gestation will be conducted in the ITT and PP Populations. The proportion of subjects within each preterm birth category will be presented for each project gestational age at randomization stratum and overall. Missing data will be imputed as detailed for the primary efficacy outcome. Statistically significant differences between treatments will be determined using the Cochran-Mantel-Haenszel test stratified by project gestational age at randomization stratum.

Descriptive statistics of gestational age at delivery will be provided by treatment group in the ITT and PP Populations, and the Van Elteren test stratified by project gestational age at randomization stratum will be used to test for statistically significant differences between treatments.

7.3 Safety Analyses

7.3.1 Primary Safety Analysis

Analysis of the safety outcome of fetal/early infant death will be conducted in the ITT Population. For each project gestational age at randomization stratum and overall, the percentage of subjects with a fetal/early infant death will be provided. The relative risk of fetal/early infant death for the 17P treatment relative to the vehicle treatment will be determined using the Cochran-Mantel-Haenszel procedure stratified by project gestational age at randomization stratum. A two-sided 95% confidence interval (CI) for the relative risk will be constructed using the Cochran-Mantel-Haenszel method adjusted for project gestational age at randomization stratum. If the upper bound of the 95% CI is less than 2.0, a doubling in the risk of fetal/early infant death can be ruled out.

7.3.1.1 Additional Analyses of Primary Safety Outcome Measure

A logistic regression model including terms for treatment, project gestational age at randomization stratum, and treatment-by-project gestational age at randomization stratum interaction will be used to assess whether there is an interaction between treatment and project gestational age at randomization.

The association of project gestational age at randomization as a continuous variable and fetal/early infant death will also be explored in a logistic regression model with terms for treatment, project gestational age at randomization (continuous), and treatment-by-project gestational age at randomization (continuous) interaction. In addition, the effect of project gestational age at randomization as a continuous variable on fetal/early infant death will be modeled as an additive rather than multiplicative effect by fitting a generalized linear model with terms for treatment, project gestational age at randomization as a continuous variable, and the interaction of treatment and project gestational age at randomization as a continuous variable.

Two sensitivity analyses will be completed. In the first sensitivity analysis, subjects with missing data will be assumed not to have a fetal/early infant death. The second sensitivity analysis will utilize only subjects with available data, i.e., subjects with missing data will not be considered. For both analyses, the percentage of subjects with a fetal/early infant death will be provided for each project gestational age at randomization stratum and overall, by treatment group. The relative risk of fetal/early infant death for the 17P treatment relative to the vehicle treatment will be determined using the Cochran-Mantel-Haenszel procedure stratified by project gestational age at randomization.

The analysis of fetal/early infant death as described above for the primary safety analysis will also be conducted in the Safety Population.

7.3.2 Secondary Safety Analyses

Secondary safety analyses will be conducted in the Safety Population unless otherwise indicated. Subjects who received a complete course of the wrong study medication will be analyzed in the treatment group based on the medication received.

7.3.2.1 Miscarriages and Stillbirths

The numbers and percentages of subjects with a stillbirth and with a miscarriage will be presented by treatment group and by project gestational age at randomization. The Cochran-Mantel-Haenszel test stratified by project gestational age at randomization stratum will be used to test for statistically significant differences between treatments.

7.3.2.2 Adverse Events and Maternal Complications

An AE is defined as an untoward medical event not present prior to exposure to study medication or any event already present that worsens in either intensity or frequency following exposure to study medication from trial injection at Baseline through the End of Treatment Period Visit, regardless of its causal relationship to study treatment. A treatment-emergent AE (TEAE) is defined as any event not present prior to exposure to study medication or any event already present that worsens in either intensity or frequency following exposure to study medication through the End of Treatment Period Visit.

Adverse events will be coded using Version 21.1 of MedDRA. A summary table by treatment group will provide the number of AEs and the number and percentage of subjects with an AE from the trial injection to the date of randomization for the All Enrolled Population. The table will also provide the number of TEAEs and the number and percentage of subjects with a TEAE from randomization through the follow-up visit, the number of maternal complications and the number and percentage of subjects with a maternal complication, the number and percentage of subjects with a TEAE/maternal complication that led to study drug discontinuation, and the number and percentage of subjects with a serious TEAE/maternal complication for the Safety Population.

The number and percentage of subjects in each treatment group in the Safety Population reporting a TEAE from the date of randomization (inclusive) through the End of Treatment Period Visit will be tabulated by MedDRA system organ class and preferred term; by geographic region (US and non-US), system organ class and preferred term; by country, system organ class and preferred term; by system organ class, preferred term, and severity (mild, moderate, and severe); and by system organ class, preferred term, and relationship (unrelated or related) to study medication. Unrelated TEAEs are defined as definitely not or probably not related to study medication. Related TEAEs are defined as definitely, probably, or possibly related to study medication. For all analyses of AEs, if the same AE (based on preferred term) is reported more than once for the same subject, the AE will be counted only once for that preferred term and at the highest severity and strongest relationship to study drug. All TEAEs leading to discontinuation of study medication will be summarized by system organ class and preferred term. All serious TEAEs will be summarized by system

organ class and preferred term. Treatment emergent injection site reactions will also be presented by MedDRA higher level and preferred term.

Maternal pregnancy complications will be coded using Version 14.1 of MedDRA. The number and percentage of subjects in the Safety Population in each treatment group reporting a pregnancy complication will be tabulated by MedDRA system organ class and preferred term; by geographic region (US and non-US), system organ class and preferred term; by country, system organ class and preferred term; by system organ class, preferred term, and severity (mild, moderate, and severe); and by system organ class, preferred term, and relationship (unrelated or related) to study medication. Unrelated pregnancy complications are defined as definitely not or probably not related to study medication. Related pregnancy complications are defined as definitely, probably, or possibly related to study medication. All maternal pregnancy complications leading to discontinuation of study medication will be summarized by system organ class and preferred term. All serious maternal complications will be summarized by system organ class and preferred term.

TEAEs and maternal pregnancy complications will be combined and the tables noted above for TEAEs will be provided. In addition, a table will provide all TEAEs and maternal complications, summarized by preferred term, occurring in at least 2% of subjects (based on preferred term) in either treatment group.

A listing of all TEAEs/maternal complications leading to discontinuation of study medication will be provided and will include subject ID, treatment group, date and Study Day of onset for the TEAE/maternal complication, system organ class, preferred term, verbatim term, severity, relationship to study medication, action taken, and seriousness. A listing of all serious TEAEs/maternal complications will also be provided and will include subject ID, treatment group, date and Study Day of onset of the serious TEAE, system organ class, preferred term, verbatim term, severity, relationship to study medication, and action taken.

7.4 Other Analyses

7.4.1 Study Drug Administration

Dosing information will be summarized by the number of injections received and compliance with the expected dosing regimen. Compliance is defined as the number of injections received divided by the number of expected injections multiplied by 100. Injections are expected to occur at least every 7 (-2/+3) days from randomization until delivery or through 36^6 weeks of gestation, whichever occurs first. Full compliance is defined as at least 90% compliance. An additional definition of compliance will be defined as the number of injections received divided by the number of expected injections based on a 7 day injection schedule. The percentages of subjects with <80%, 80%-120%, and >120% compliance will be presented. Statistically significant differences between treatments in the number of injections and in compliance will be determined using the Wilcoxon Rank Sum test and in the percentage of subjects who are fully compliant using the chi-square test.

7.4.2 Prior and Concomitant Medications

The World Health Organization (WHO) drug dictionary (Version March 2009) will be used to classify all medications. Prior and concomitant medications, other than tocolytics, will be summarized by ATC Class 2 and preferred term, separately by treatment group for the ITT population. Subjects receiving the same medication more than once will be counted only once for a particular medication class and medication. A prior medication is defined as any medication taken during the current pregnancy (i.e., from 40 weeks before the project EDC up to the date of randomization). A concomitant medication is defined as any medication, excluding tocolytic medications, taken on or after the date of randomization through the End of Treatment Period Visit. Concomitant tocolytic medications will be summarized by ATC Class 2 and preferred term, separately by treatment group for the ITT population. Subjects receiving the same medication more than once will be counted only once for a particular medication class and medication.

7.4.3 Post-randomization Cervical Length by Gestational Age

Descriptive statistics of post-randomization cervical length (mm) will be provided by gestational age at the time of the cervical ultrasound ($\leq 20^6$ weeks, 21^0 - 27^6 weeks, and $\geq 28^0$ weeks) and treatment group for all subjects in the ITT Population. Cervical length will also be defined categorically (<25 mm vs. ≥25 mm), and the number and percentage of subjects in each category will be presented. Only cervical length measurements obtained from ultrasounds after the date of randomization will be included in these analyses. The Wilcoxon Rank Sum test will be used to test for statistically significant differences between the 17P and vehicle treatments for continuous variables, and the chi-square test or Fisher's exact test will be used for dichotomous variables.

7.4.4 Gestational Diabetes

The number and percentage of subjects in the ITT Population who were diagnosed with gestational diabetes will be summarized by treatment group, along with the primary basis of the diagnosis. The number and percentage of subjects in the ITT Population who were screened, but not diagnosed, with gestational diabetes will also be summarized by treatment group. The chi-square test or Fisher's exact test will be used to determine statistically significant differences between the treatments with respect to the proportion of subjects diagnosed with gestational diabetes and the proportion of subjects screened, but not diagnosed, with gestational diabetes.

7.4.5 Maternal Delivery Data

Maternal delivery data will be summarized by treatment group. This includes maternal length of stay in the hospital, the percentage of subjects with at least one episode of preterm labor, type of labor (spontaneous, induced, spontaneous-augmented, no labor), indication for induced or augmented labor, delivery route (normal vaginal, assisted vaginal, cesarean section), and indication for cesarean section.

7.4.6 Gestational Age at Delivery and Neonatal Outcome

A logistic regression model of the neonatal composite index with covariate terms for treatment and gestational age at delivery as a continuous variable will be conducted. The odds ratio and 95% CI for the odds ratio for each covariate will be presented.

8 Changes from Analyses as Described in the Protocol

Six analysis populations are defined in the protocol. The Modified ITT population is not used for analyses and PK analyses are not described in this SAP. Thus, both the Modified ITT and PK populations are not discussed in this SAP.

Statistically significant differences between treatments for preterm delivery at $<35^0$ weeks, $<32^0$ weeks, and $<37^0$ weeks will be determined using a Cochran-Mantel-Haenszel test stratified by project gestational age at randomization (16^0 weeks - 17^6 weeks gestation and 18^0 weeks – 20^6 weeks gestation). Multiple imputation will be used to address missing data. The staggered entry Kaplan-Meier method will not be used because of issues with this approach. One issue is that subjects do not all have the same survival curve, as is assumed by this approach. Subjects entering the study at 16 weeks and those entering at 17^6 weeks, for example, do not have the same survival curve, because the former are treated from 16 to 18 weeks and the latter are not.

The protocol states that the percentages of subjects who deliver prior to 35^0 weeks gestation (primary outcome) and who deliver prior to 32^0 and prior to 37^0 weeks gestation (secondary outcomes) will be analyzed for the Safety Population, but these analyses will not be done.

The protocol states that, if there are baseline imbalances between the treatment groups with respect to prognostic factors such as the number of previous preterm deliveries, an adjusted analysis of the primary outcome measures will be conducted using the Cochran-Mantel-Haenszel procedure (for the neonatal composite index) and/or a Cox regression model (for preterm delivery $<35^0$ weeks gestation). Instead, logistic regression analyses will be performed for both endpoints. The protocol also states that an additional analysis of the primary efficacy outcomes will be performed to determine if there is a treatment-by-site interaction, but this will not be done.

9 Appendix A – Schedule of Events

Table 1 Schedule of Events

Procedures	Baseline[a]	Treatment Period[b]	Delivery and Hospitalization	Neonate Follow-Up[c]	End of Treatment Period Visit[d]
	Visit 1	Visits 2 to 36^6 Weeks of Gestation or Delivery			
Informed consent[e]	X				
Medical records release[f]	X				
Medical/obstetrical history	X				
Demographic information/social history	X				
Ultrasound (14^0 through 20^3)	X[g]				
Document previous preterm delivery	X				
Brief physical examination[h]	X[h]	X[h]			
Height	X				
Weight	X	X			
Prior medications[i]	X	X			
Concomitant medications[j]		X	X		X
Determine project gestational age and estimated date of confinement	X				
Schedule randomization visit	X				
Trial injection	X				
Randomization[k]		X			
Collect blood sample for pharmacokinetic analysis		X[l]			
Study drug administration		X[m]			
Record adverse events (AEs)[d]	X[n]	X	X		X
Record presence or absence of POME symptoms	X	X			
Record pregnancy complications		X	X		
Record additional risk factors of miscarriage	X		X		
Maternal delivery information			X		
Neonatal information[c]			X	X	

[a] Visit will occur wherever possible within 7 days before randomization.

[b] Subject will report to the clinical site weekly for study drug administration until 36^6 weeks of gestation or delivery, whichever occurs first.

[c] Neonates born to randomized subjects will be followed through 28 days of life or discharge from the NICU whichever is later. Therefore, the status of all neonates (alive or dead), regardless of when they are delivered and discharged from the hospital will be obtained at least 28 days after delivery (this contact may occur at later than 28 days after delivery as long as the status of the infant on Day 28 after delivery is determined). If the neonate has been discharged from the birth hospitalization, the subject may be contacted by telephone to obtain the neonate's status. If the infant is still in the NICU or equivalent 28 days after delivery, then their status (alive or dead) will be determined upon discharge from the NICU or equivalent. This status may be obtained after the infant is discharged as long as the status of the infant on the day of discharge is determined

[d] All randomized subjects, regardless of when they deliver, should be contacted for an End of Treatment Period Visit to obtain AE information including medications to treat AE(s). The contact can be either in person or by telephone and should occur 35 ± 7 days after the last dose of study drug.

[e] To be completed before performing any baseline procedures. Informed consent may be obtained at a gestational age of 10 weeks or greater to facilitate obtaining records from the qualifying delivery. At centers participating in the 17P-FU-004 study, every effort will be made to obtain informed consent for the 17P-FU-004 study from all randomized subjects while they are pregnant. If it is not possible to obtain consent during the pregnancy, consent may be obtained up to the point that their child reaches one year of age.

[f] Must be signed by subject in order to obtain medical records of previous deliveries.

[g] If a 14^0 to 20^3 weeks of gestation ultrasound to rule out fetal anomalies has not been performed as part of standard prenatal care; one must be performed prior to randomization.

[h] Conduct a brief physical examination including a brief head-to-toe visual inspection at either Baseline (Visit 1) or Visit 2 only.

[i] Prior medications include all medications taken since the start of this pregnancy, defined as 40 weeks before the project EDC and before the study drug is randomly assigned. Prior medications collection ends at Visit 2.

[j] Concomitant medications must be recorded in the case report form through the End of Treatment Period Visit.

[k] Between 16^0 and 20^6 weeks of gestation.

[l] Three blood samples will be drawn from the PK population at the following times: (1) Before study drug dosing at either Visit 6 or 7 (i.e., dose 5 or 6). (2) Before dosing at either Visit 8 or 9 (i.e., Dose 7 or 8). (3) At a separate, non-dosing visit 1 to 6 days after Visit 9, 10, or 11 (i.e., 1 to 6 days after Doses 8, 9, or 10). Subjects will be stratified 2:1 (17P: vehicle) such that an even distribution of samples will be drawn on day 1, day 2, day 3, day 4 or day 5/6 post-dose.

[m] Subjects will receive one injection for each week they are pregnant from randomization through 36^6 weeks gestation or delivery, whichever occurs first. As much as possible injections should be given on the same day each week. However, if a schedule change is required, in general injections should be at least 5 days apart and no more than 10 days apart.

[n] AEs are recorded from administration of the trial injection through the End of Treatment Period Visit including medications to treat the AE. Preterm birth is an anticipated outcome and is not considered an AE. Maternal or fetal deaths will be recorded through delivery.

10 References

[1] American Diabetes Association. Gestational diabetes mellitus. *Diabetes Care.* 2004;27(1): S88-90.

[2] Gonzalea-Quintero VH. 17 Alpha-hydroxyprogesterone Caproate Perinatal Mortality and Pregnancy Outcomes. *Obstet Gynecol.* 2007;109 (4) Supplement: 1S-2S.

[3] O'Brien JM, Adair CD, Lewis DF, et al. Progesterone vaginal gel for the reduction of recurrent preterm birth: primary results form a randomized, double-blind, placebo-controlled trial. *Ultrasound Obstet Gynecol*. 2007;30:687-696.

[4] Meis PJ, Klebanoff M, Thom E, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *N Engl J Med*. 2003;348(24):2379-2385.

[5] Ratitich B, Lipkovich I, O'Kelly M. Combining analysis results from multiply imputed categorical data. *Proceedings PharmaSUG 2019*, Paper SP03.