Statistical Analysis Plan

Study alias & e-track number(s): EPI-PERTUSSIS-037 VS BR (203153)

Detailed Title:	A post-marketing, observational, retrospective, cohort study to assess the safety of RefortrixTM (Tdap) when administered during pregnancy in a maternal immunization program in Brazil.
SAP version	Final
SAP date	01-APR-2017
Scope:	All data pertaining to the above study.
Co-ordinating author:	<PPD (Biostatistician)>
Other author(s):	
Adhoc reviewers:	<PPD (Safety physician)>
Approved by:	<PPD (Epidemiologist)>
	<PPD (Lead statistician)>

TABLE OF CONTENTS

PAGE

The complete statistical analysis plan and results presentation is divided into 2 parts: the first part detailing the analyses to be performed (known as SAP, current document) and a second part, annex (-es) (called TFL) describing the flow and format of tables, figures and listings to be annexed to the SR.

LIST OF ABBREVIATIONS

AE	Adverse event
ATP	According-To-Protocol
CI	Confidence Interval
CRF	Case Report Form
GSK	GlaxoSmithKline
LL	Lower Limit of the confidence interval
N.A.	Not Applicable
OR	Odds Ratio
PASS	Post-Authorisation Safety Study
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SR	Study Report
TC	Total Cohort
Tdap	Combined reduced-antigen-content diphtheria, tetanus and acellular pertussis (Tdap) vaccine
TSS	Targeted Safety Study
TFL	Tables Figures and Listing template annexed to SAP
UL	Upper Limit of the confidence interval

1. DOCUMENT HISTORY

Date	Description	Protocol Version
22-MAR-2017	Version 1	Final Version 1 11 September 2015

2. STUDY DESIGN

- Type of design: An observational, retrospective, cohort, single-centre study.
- This is a Targeted Safety Study (TSS) and a Post-Authorisation Safety Study (PASS).
- Study size: The planned sample size of this study is 2400 subjects. Using a two-sided (alpha = 0.01) test assuming the ratio of subjects in the Exposed cohort to the Unexposed cohort is 1:1 and assuming a background proportion of events in the Unexposed cohort to be 3%, a total of 2400 subjects [1200 subjects in each cohort], will be needed to have more than 80% power to detect a relative risk of 2 or higher.

The following group names will be used for the statistical analyses:

Group order in tables	Group label in tables	Group definition for footnote
1	Exposed Cohort	Pregnant women who had received Refortrix as part of the maternal immunization program in Brazil
2	Unexposed Cohort	Women pregnant before implementation of the maternal immunization program in Brazil who didn't receive Refortrix

The exposed cohort will be divided into co-vaccinated cohort and single-vaccinated cohort. The single-vaccinated cohort includes subjects who had received only Refortrix; and the co-vaccinated cohort includes subjects who had received Refortrix and other vaccination(s). This analysis will be performed depending on the number of subjects in each subgroup.

Group order in tables	Group label in tables	Group definition for footnote
1	Co-vaccinated exposed Cohort	Pregnant women who had received Refortrix and other vaccination(s) as part of the maternal immunization program in Brazil
2	Single-vaccinated exposed Cohort	Pregnant women who had received only Refortrix as part of the maternal immunization program in Brazil

3. OBJECTIVES

3.1. Co-primary objectives

- To compare the risk of gestational diabetes, pregnancy-related hypertension and pregnancy haemorrhage (ante-partum (after 24 weeks of gestation), intra-partum or post-partum) in a cohort of women following vaccination with *Refortrix* as part of the maternal immunization program in Brazil (Exposed cohort) with a historical cohort of unvaccinated pregnant women before the implementation of this immunization program (Unexposed cohort).
- To compare the risk of preterm birth and small for gestational age in neonates born to subjects in the Exposed cohort and to subjects in the Unexposed cohort.

3.2. Secondary objectives

- To describe the risk of pregnancy-related AEs/neonate-related events of interest (premature rupture of membranes, preterm premature rupture of membranes, premature uterine contraction, neonatal death, maternal death, still birth and neonatal hypoxic ischaemic encephalopathy) in the Exposed and Unexposed cohorts.
- To describe the risk of congenital anomalies in neonates in the Exposed and Unexposed cohorts.
- To describe the risk of pregnancy-related AEs and birth outcomes per calendar year in the Unexposed cohort.

4. ENDPOINTS

4.1. Co-primary endpoints

- Occurrence of any of the following pregnancy-related AEs in Exposed and Unexposed subjects.
  - Gestational diabetes.
  - Pregnancy-related hypertension (including pre-eclampsia, eclampsia and HELLP syndrome).
  - Pregnancy haemorrhage (ante-partum (after 24 weeks of gestation), intra-partum or post-partum).
- Occurrence of any of the following outcomes in neonates from Exposed and Unexposed subjects.

- Preterm birth.
- Small for gestational age.

4.2. Secondary endpoints

- Occurrence of pregnancy-related AEs of interest/neonate-related events up to delivery in Exposed and Unexposed subjects.
  - Premature rupture of membranes.
  - Preterm premature rupture of membranes.
  - Premature uterine contraction.
  - Neonatal death.
  - Maternal death.
  - Still birth.
  - Neonatal hypoxic ischaemic encephalopathy.
- Occurrence of congenital anomalies in the neonates of Exposed and Unexposed subjects.
- Occurrence of pregnancy-related AEs and birth outcomes per calendar year in the Unexposed cohort.

5. STUDY POPULATION

5.1. Total Cohort (TC)

The TC will include all subjects enrolled in the study. All the information for these subjects will be entered in the eCRF.

5.2. According-To-Protocol (ATP) Cohort

The ATP cohort will include all the evaluable subjects (i.e. those meeting all eligibility criteria, complying with the procedures defined in the protocol).

The list of applicable elimination codes for each cohort can be found in the study specific form FORM-BIO-CLIN-9004-05 Criteria for eliminating subjects from the analyses.

Cohort	Elimination codes	Eli Type
ATP cohort for analysis	900, 2010	MA

6. STATISTICAL METHODS

All the statistical calculations will be done in SAS 9.2 or higher.

The primary analysis will be based on the Total cohort. Should the percentage of subjects excluded from the ATP cohort for analysis of safety be greater than 5%, a second analysis will be performed on the ATP cohort to complement the analysis of the Total cohort.

6.1. Analysis of demographics/baseline characteristics

The baseline and demographic characteristics of the Exposed and Unexposed cohorts will be tabulated in a summary of statistics.

- Frequency tables will be generated for categorical variables such as resident of the study area and previous pregnancies etc.

- Mean, median, standard deviation and range will be provided for continuous data such as maternal age and gestational age.

6.1.1. Demographic characteristics

The number of enrolled subjects as well as the number excluded from ATP analyses will be presented.

Demographic characteristics (age and resident of the study area) will be summarised using descriptive statistics (for all non-missing observations) overall and by group.

6.1.2. Clinical characteristics

The following tables will be generated for the following characteristics:

- General medical history

The percentage of subjects reporting any of the non-pregnancy related conditions, signs or symptoms or exanthematic diseases in the general medical history will be tabulated overall and by group. The diagnosis entry for before or during the pregnancy and the diagnosis entry details will be summarized overall and by group.

- Congenital anomalies

The percentage of subjects reporting any congenital anomalies (total, in the subjects, and/or for spouse, and/or for first degree relatives) will be described in a frequency table overall and by group.

- Obstetric history

The percentage of subjects who had previous pregnancies will be tabulated overall and by group. Among those who had previous pregnancies, the gravidity and parity number will be summarized and the percentage of subjects who had congenital anomalies, ectopic pregnancies, molar pregnancies, miscarriages will be tabulated.

The percentage of subjects who had events/complications (pre eclampsia, eclampsia, HELLP syndrome, infection, gestational diabetes, vaginal haemorrhage, premature rupture of membranes, preterm premature rupture of membranes, premature uterine contractions, pregnancy-related hypertension, neonatal death, neonatal hypoxic ischaemic encephalopathy) in previous pregnancies will also be tabulated overall and by group.

The percentage of subjects who had new born safety events (preterm babies, post term babies, new born with low birth weight, new born with birth weight > 4000 grams) will be summarized overall and by group in this section as well.

- Current pregnancy

The characteristics for current pregnancy will be described overall and by group. The characteristic include gestational age in weeks at the time of the present delivery, assisted fertilization, multiparity ultrasound, hospitalization, type of delivery, risk factors, habits, the use of illicit drugs, the use of chronic medications, etc.

- DTAP vaccination history and concomitant vaccination history

The percentage of subjects who had DTAP vaccination history, including the number of doses and gestational age at the time of vaccination will be summarized. Similar analysis will be performed for concomitant vaccination history. The percentage for co-vaccinated exposed cohort and single exposed cohort will also be tabulated.

6.2. Analysis of co-primary endpoints

The main analysis for co-primary objectives will contain only the subjects with vaccination date in the Exposed cohort and subjects from the Unexposed cohort. If more than 10% of subjects have missing vaccination date in the Exposed cohort, a sensitivity analysis will be performed using the imputed vaccination date to 27 completed weeks of gestational age to evaluate if this has any potential impact on the results.

The risk for each primary endpoint (gestational diabetes, pregnancy-related hypertension, pregnancy haemorrhage, preterm birth and small for gestational age) will be calculated for both exposed and unexposed cohorts. Similar analyses will be performed for co-vaccinated exposed cohort, single-vaccinated exposed cohort and unexposed cohort.

For each specific endpoint, the number of subjects where the event occurred [between the index date (refer to Annex 2 in protocol for definition of index date) and the date of the delivery] will be divided by the total number of subjects at risk for both the Exposed and Unexposed cohorts respectively, together with its exact 99% confidence interval (CI).

The co-primary endpoints of pregnancy (gestational diabetes, pregnancy-related hypertension and pregnancy haemorrhage) will be pooled together and of birth outcome (preterm birth and small for gestational age) will be pooled together in addition. The analysis of the risk for the pooled endpoints together with its exact 95% CI will be performed using the same method as for the separate co-primary endpoints.

The comparison of the risk with its two sided 99% CI of each primary endpoint between the Exposed cohort and the Unexposed cohort will be obtained by means of logistic regression model, using the exposure status as a binary independent variable in the model.

Univariate analyses will describe the association between the each primary endpoint and all potential risk factors such as: Tdap vaccination status, maternal age at the start of the pregnancy, parity and gestational age, congenital anomalies (in parents or first degree relatives), maternal comorbidities and complications during previous pregnancies, etc. In addition, potential interaction terms will be added if deemed necessary after exploring the data.

Afterwards, a multiple logistic regression model will be fitted with a backward selection to identify the possible confounding factors which were detected from the univariate analysis using an alpha level of 0.1. Adjusted odds ratio (OR) and its 95% CI will be derived from the final model.

6.3. Analysis of secondary endpoints

The risk for each secondary endpoint (pregnancy-related AEs and birth outcomes) will be calculated by the number of subjects with at least one of the events occurring between the index date (refer to Annex 2 in protocol for definition of index date) and the date of the delivery, corresponding to that endpoint divided by the total number of subjects at risk for both the Exposed and Unexposed cohort, together with its exact 95% CI. If more than 10% of subjects have missing vaccination date in the exposed cohort, a sensitivity analysis will be performed as for the co-primary endpoints.

In addition, the risk of all the co-primary and secondary endpoints will be calculated by calendar year as well to evaluate the stability among the Exposed (total, co-vaccinated and single-vaccinated) and Unexposed cohorts.

6.4. Sensitivity analysis

A sensitivity analysis will also be performed if more than 10% of the measurements are missing for each endpoint. In the first instance, missing outcomes will be imputed with a value of '0'. In the second instance, all missing outcomes will be imputed with a value of '1'. The risk for each endpoint will then be analysed using similar methods as mentioned in sections 6.2 and 6.3 for all the co-primary and secondary endpoints.

6.5. Additional analysis not in the protocol

- The index date was defined as 27 weeks completed (28 weeks) gestational age in the unexposed cohort. However, in the exposed cohort, some of the subjects had the vaccination dates at 27 gestational weeks. Therefore, a sensitivity analysis for the primary analysis will be performed by excluding these subjects who had vaccination date at 27 gestational weeks.

- The index date for exposed cohort will be the real vaccination date which is around 27-36 gestation weeks; however, the index date for unexposed cohort will be imputed to 27 completed gestation weeks; it means that the unexposed cohort will have longer mean follow up period .Therefore, there is a risk of higher number of observed events in the unexposed vs. the exposed cohorts which could underestimate the risk ratio estimate. Therefore, a sensitivity analysis will be added as to summarize the total Follow-up period between two cohorts and to calculate the total IR = total number of events/ total person-years.

- Summary table for the cross-tabulation of each endpoint (primary and secondary) between previous pregnancies and current pregnancy will be presented.

- Incidence tables will be presented for pregnancy-related adverse events and birth outcomes with/without concomitant vaccination overall and separately, diphtheria-tetanus vaccination, hepatitis B vaccination in the exposed cohort.

7. STATISTICAL CALCULATIONS

7.1. Methodology for computing CI

All CI will be two sided 99% CI for co-primary objectives analysis and 95% CI for the other analyses.

7.1.1. Confidence interval for risk

The analysis of the risk for each specific endpoint will be performed using exact 99% confidence interval (CI).

If *n* is the number of subjects where the event occurred among the total number of subjects *N*, the risk can be estimated by *n/N*. Its exact *(1-α)%* confidence interval is obtained from:

*CINV(α /2,2*n)/2/N* as the lower boundary

and

CINV((1-α /2),2(n+1))/2/N* as the upper boundary.

where *CINV(probability, degrees of freedom)* returns the inverse of the chi-squared probability distribution and α is the type I error rate.

The comparison of the risk with its two sided 99% CI of each specific endpoint between the Exposed cohort and the Unexposed cohort will be obtained by means of logistic regression model, using the exposure status as a binary independent variable in the model.

Logit P=a+bx, *x* is exposure status

Where regression coefficient b is a log OR (it means that *exp(b)* is OR), 99%CI of OR will be exp(b ± 2.58*se(b)).

7.2. Number of decimals

The following decimal description will be used for the analyses.

Parameters	Number of decimal digits
% of count, including LL & UL of CI	2
p-value	3
Minimum, maximum, range	Number of decimals in the raw data
Mean, median	Number of decimals in the raw data +1
SD	Number of decimals in the raw data +2

LL = Lower Limit UL = Upper Limit CI = Confidence Interval
SD = Standard deviation

7.3. Handling of missing data

Missing or non-evaluable primary and secondary outcome measurements will not be replaced. Therefore, the main analysis will exclude subjects with missing or non-evaluable data.

For subjects in the Exposed cohort whose date of the vaccination is not available, it will be imputed to the 27th completed gestational week as the recommended start time for the vaccination for the sensitivity analysis.

7.4. Derived and transformed data

- Index date: For the Exposed cohort, the index date will be the date of Refortrix vaccination given as part of the maternal immunization program in Brazil. For the Unexposed cohort, the gestational age of 27 completed weeks will be considered as the index date.

- For Primary pregnancy-related AEs endpoints: Gestational diabetes, Pregnancy-related hypertension (including pre-eclampsia, eclampsia and HELLP syndrome), and Pregnancy haemorrhage (ante-partum (after 24 weeks of gestation), intra-partum or post-partum), only the events after the index date will be selected for the analysis for this co-primary objective.

- Subjects had at least one pre-eclampsia and/or eclampsia and/or HELLP syndrome will be counted once for the analysis for pregnancy-related hypertension endpoint. The worst event among pre-eclampsia, eclampsia and HELLP will be counted for summary of the event. For example, if a subject had one episode of pre-eclampsia and one episode of eclampsia, this subject will be summarized to have the episode of eclampsia as this is the worst event during the study period.

8. CONDUCT OF ANALYSES

8.1. Sequence of analyses

The statistical analyses will be performed when all data are available. All analyses will be performed on final and clean data.

Description	Analysis ID	Disclosure Purpose	Reference for TFL
Final Analysis	E1_01	Statistical analysis report	All tables from TFL dated xxxxxx

8.2. Statistical considerations for interim analyses

No interim analyses are planned for this study.