Statistical Analysis Plan

Study Title:	The safety of Boostrix following routine immunization of pregnant women
Protocol Number:	EPI-PERTUSSIS-047 VS US DB (207221)
Sponsor:	GlaxoSmithKline
Coordinating/Principal Investigator(s):	Hung Fu Tseng Department of Research and Evaluation Kaiser Permanente Southern California PPD USA
Epidemiologist:	PPD Epidemiology Lead DTP GSK Biologicals
Plan Prepared by:	PPD Research Scientist Biostatistician Kaiser Permanente Southern California

Approvals

Approvers, Organization & Role	Approval Status	Signature	Date
PPD KPSC Research Scientist Biostatistician	Approved	PPD PPD (Dec 2, 2019)	Dec 2, 2019
PPD GSK Biostatistics and Statistical Programming	Approved	PPD PPD (Dec 3, 2019)	Dec 3, 2019

Publication Record

Version	Date	Primary Author	Description
Draft 1.0	14 June 2019	PPD	1st Draft for GSK Review
Final 1.0	26 July 2019	PPD	Final version incorporating GSK comments/edits
Final 2.0	02 December 2019	PPD	Final version of the SAP amendment following the protocol amendment (signed 25NOV2019)

ABBREVIATIONS

ACIP: Advisory Committee on Immunization Practices

ACOG: American College of Obstetricians

ADaM: Analysis Data Model

AE: Adverse Event

ASD: Absolute Standardized Difference

BMI: Body mass index

CA: Congenital anomalies

CDISC: Clinical Data Interchange Standards Consortium

CI: Confidence interval

DMP: Data Management Plan

ED: Emergency Department

EDD: Estimated Date of Delivery

EHR: Electronic Health Record

GSK: GlaxoSmithKline

HIV: Human immunodeficiency virus

ICD: International Classification of Diseases

KPSC: Kaiser Permanente Southern California

MRN: Medical Record Number

PHI: Protected Health Information

PPROM: Preterm pre-labor rupture of membranes

PPV: Positive Predictive Value

R&E: Research and Evaluation

RR: Relative Risk

SAP: Statistical Analysis Plan

SAS: Statistical Analysis System

SDTM: Study Data Tabulation Module

SGA: Small for Gestational Age

SOP: Standard Operating Procedures

Tdap: Tetanus, diphtheria, acellular pertussis

TABLE OF CONTENTS

1. INTRODUCTION

Whooping cough, also referred to as pertussis, is a highly contagious respiratory disease primarily caused by the bacterium *Bordetella pertussis*. The disease is most severe in unvaccinated or incompletely vaccinated young infants. To increase the protection of newborns, in October 2012, the Advisory Committee on Immunization Practices (ACIP) recommended tetanus-diphtheria-acellular pertussis (Tdap) vaccine to be administered during pregnancy to all pregnant women, regardless of whether they have received Tdap previously. The optimal timing for Tdap administration is between 27 and 36 weeks of gestation.

Two Tdap vaccines (Adacel and Boostrix) were licensed for use in the United States in 2005. These vaccines were recommended for routine use in non-pregnant adolescents and adults when introduced. Both Adacel and Boostrix can be used interchangeably. There are limited post-marketing safety data on Boostrix during pregnancy. Through partnership between Kaiser Permanente Southern California (KPSC) and the sponsor, GlaxoSmithKline (GSK), we will conduct an observational, electronic health record (EHR) database study to assess the safety of Boostrix following routine immunization of pregnant women in a community setting.

This Statistical Analysis Plan (SAP) provides a detailed description of the data considerations, statistical methods, data analysis and results presentation as previously proposed in the protocol and Data Management Plan (DMP), where applicable. The SAP also describes quality assurance procedures and analytic data format standards.

2. STUDY OBJECTIVE

2.1. Primary objective

To rule out a two-fold increase (non-inferiority testing) in the incidence of each of the following maternal and infant adverse events (AEs):

- Preeclampsia and/or eclampsia
- Intra-uterine infections such as chorioamnionitis and endometritis
- Small for gestational age (SGA)
- Preterm delivery

among women who were vaccinated with Boostrix on or after the 1st day of the 27th week of gestation as compared to the incidence in a historical cohort of pregnant women (approximately January 1, 2012 to December 31, 2013) who were unvaccinated with any Tdap vaccine throughout their pregnancy.

The null hypothesis and alternative hypothesis for test of relative risk (RR) are:

$$\text{Ho: } RR \geq 2$$

$$\text{Ha: } RR < 2$$

2.2. Secondary objective

1. To explore differences (superiority testing) in the incidence rate of other maternal and infant AEs, namely:
   - Transfusion during delivery hospitalization
   - Poor fetal growth
   - Stillbirth/fetal death with or without congenital anomalies
   - Neonatal death (within 28 days of birth)
   - Maternal death (while pregnant or within 42 days of end of pregnancy)
   - Placental abruption
   - Preterm pre-labor rupture of membranes (PPROM)
   - Congenital anomalies (CA) at birth and through 6 months of age

   between women who were vaccinated with Boostrix on or after the 1st day of the 27th week of gestation and a historical cohort of pregnant women (approximately January 1, 2012 to December 31, 2013) who were unvaccinated with any Tdap vaccine throughout their pregnancy.

   The null hypothesis and alternative hypothesis for test of RR are:

   $$Ho: RR = 1$$

   $$Ha: RR \neq 1$$

2. To describe the incidence of maternal and infant AEs described above, as well as the incidence of spontaneous and therapeutic abortion with or without congenital anomalies, in pregnant women who were vaccinated with Boostrix before the 27th week of gestation.

3. STUDY DESIGN

3.1 Overview of study design

This is an observational retrospective matched cohort study. The study will be conducted among pregnant women who are members of KPSC.

The exposed cohort (from the 27th week of gestation) will consist of pregnant women who received the Tdap vaccine (Boostrix) on or after the 1st day of the 27th week of pregnancy during the vaccination period at ob-gyn clinics of seven selected medical centers. Approximately 15,000 women vaccinated with Boostrix will be included in this study over approximately a one-year period from January 2018 to January 2019.

The unexposed cohort will consist of women, pregnant sometime during the approximate estimated period from 1/1/2012 to 12/31/2013 who never received Tdap vaccine during pregnancy. The unexposed cohort of pregnant women will be matched 1:1 to the exposed cohort of pregnant women for comparisons of maternal and infant AEs between the two cohorts.

Subject-level data will be collected for pregnant women and their infants through the EHR. For each individual pregnant woman, analytical follow up will begin on the date they receive the Boostrix vaccine, or the index date for the unvaccinated cohort, and will end on the date of

disenrollment or the end of pregnancy, whichever came first. For the maternal death endpoint, pregnant women will be followed for 42 days after the end of pregnancy. Infants born in KPSC hospitals will be followed for 6 months.

In addition, women vaccinated with Boostrix before the 27th week of gestation, with membership at the date of vaccination, and who were not vaccinated with any other Tdap vaccine at any other time during the pregnancy will be included for a descriptive analysis (single arm).

3.2 Study population, exposures, outcomes and covariates

3.2.1 Study populations/subgroups

Study population for main analysis (primary objective and secondary objective #1)

- Pregnant women received prenatal care at KPSC
- Exposed cohort: vaccinated with Boostrix on the 1st day of the 27th week of pregnancy or later during the pregnancy and were not vaccinated with any other Tdap vaccine at any other time during the pregnancy or
- Unexposed cohort: pregnant at least one day during the approximate estimated period between 1/1/2012-12/31/2013 and did not receive any Tdap vaccine during the pregnancy
- Continuous membership (allowing up to a 31-day gap) between the 1st day of the 27th week of pregnancy and the index (vaccination) date

The unexposed cohort will be 1:1 matched to the exposed cohort on age at pregnancy start (± 1 year), race/ethnicity (White, Black, Hispanic, Asian, and Other), and multiple gestation (yes/no). Every unexposed woman will be assigned an index date that is determined by the number of days from pregnancy start to the Boostrix vaccination date of her matched exposed woman. The same number of days calculated in the exposed woman will be added to the pregnancy start of the unexposed woman to determine her index date.

Study population for descriptive analysis (secondary objective #2)

- Pregnant women received prenatal care at KPSC
- Vaccinated with Boostrix before the 27th week of gestation and were not vaccinated with any other Tdap vaccine at any other time during the pregnancy
- Membership at the index (vaccination) date

Subgroups for endpoint-specific analysis

For some maternal and infant AEs (endpoints), additional restriction criteria will be applied. We will exclude individuals who do not meet the restriction criteria, but not the matched pair. Analysis will be conducted on a subgroup of the population:

- Small for gestational age (SGA): Infants 1) live born, 2) singleton, and 3) born in KPSC hospitals. Rationale: The evaluation of SGA is limited to singleton births because the reference values for SGA definition are established based on singleton births. The birth weight of infants born outside KPSC hospitals is not captured in claims, and such infants are therefore excluded for this endpoint.

- Preterm delivery: Pregnant women 1) with live birth and 2) with index (vaccination) date prior to 37 weeks gestation. Rationale: Pregnant women with an index (vaccination) date on or after 37th week of gestation are not at risk of preterm delivery, and are therefore excluded for this endpoint.
- Preterm pre-labor rupture of membranes (PPROM): Pregnant women with index (vaccination) date prior to 37 weeks gestation. Rationale: Pregnant women with an index (vaccination) date on or after 37th week of gestation are not at risk of PPROM, and are therefore excluded for this endpoint.
- Transfusion during delivery hospitalization: Pregnant women with delivery hospitalization. Rationale: Transfusion can be captured by procedure codes during the delivery hospitalization.
- Stillbirth/fetal death: Pregnant women with pregnancy outcome. Rationale: Pregnant women with unknown pregnancy outcome are considered as missing data, and are therefore excluded for this endpoint.
- Neonatal death (within 28 days of birth) and congenital anomalies at birth and through 6 months of age: Infants 1) live born and 2) born in KPSC hospitals. Rationale: Infants born outside KPSC hospitals may not be captured and may not be able to be linked back to their pregnant mothers, and are therefore excluded for this endpoint.

Analysis for preeclampsia and/or eclampsia, intra-uterine infections, poor fetal growth, placental abruption, and maternal death will be conducted on all pregnant women eligible for the main analysis and descriptive analysis.

3.2.2 Exposures

Boostrix vaccination will be received as part of routine clinical care. Vaccine exposure will be identified retrospectively from the electronic health records, including vaccine, manufacturer, and lot number entered at the time of vaccination.

3.2.3 Primary and secondary endpoints

Primary endpoints (Table 1) were selected based on their medical relevance, also considering the background incidence or prevalence of these endpoints.

Secondary endpoints (Table 1) were differentiated from primary endpoints to allow evaluation of the safety of maternal vaccination with sufficient statistical power on the primary endpoints, and to allow for hypothesis generation/signal detection on secondary endpoints.

Table 1. List of primary and secondary endpoints

Primary endpoints	Preeclampsia and/or eclampsia
	Intra-uterine infections such as chorioamnionitis and endometritis
	Small for gestational age (SGA)
	Preterm delivery
Secondary endpoints	Transfusion during delivery hospitalization
	Poor fetal growth

	Stillbirth/fetal death with or without congenital anomalies
	Neonatal death (within 28 days of birth)
	Maternal death (while pregnant or within 42 days of end of pregnancy)
	Placental abruption
	Preterm pre-labor rupture of membranes (PPROM)
	Congenital anomalies at birth and through 6 months of age
	Congenital anomalies of nervous system
	Congenital anomalies of eye
	Congenital anomalies of ear, face, or neck
	Congenital anomalies of cardiovascular system
	Congenital anomalies of respiratory system
	Clefts
	Congenital anomalies of upper gastrointestinal system
	Congenital anomalies of lower gastrointestinal system
	Congenital anomalies of genital organs
	Congenital anomalies of renal system
	Congenital anomalies of musculoskeletal system
	Congenital anomalies of limb
	Congenital anomalies of integument
	Other and unspecified congenital anomalies
	Spontaneous abortion with or without congenital anomalies (only for women exposed to Boostrix before the 27th week of gestation)
	Therapeutic abortion with or without congenital anomalies (only for women exposed to Boostrix before the 27th week of gestation)

The algorithms for identification of primary and secondary endpoints are described in the DMP. Except SGA, preterm delivery, and transfusion during delivery hospitalization, AEs identified through the algorithms will be chart reviewed to confirm the diagnosis and the onset (when applicable). Analyses of endpoints will be based on confirmed cases. For AEs with a chart-reviewed onset date, the event start date will be the onset date; for AEs without a chart-reviewed onset date (unknown or inapplicable), the event start date will be the earliest confirmed diagnosis date. To be a confirmed case, the event start date needs to be after the index date. Some AEs might be diagnosed during the delivery hospitalization, and have an event start date after the delivery date. For pre-eclampsia and/or eclampsia, intra-uterine infections, and poor fetal growth, the event date needs to be on or before the delivery date. For placental abruption and PPROM, if the event start date is after the delivery date, the delivery date will be used as the event start date.

SGA will be defined based on reference values derived by Talge et al. with a cut-off of <10th percentile (1). Talge at al used 2009-2010 US natality data and corrected for likely errors in gestational age dating to yield an up-to-date birth weight for gestational age reference. The

evaluation of SGA is limited to singleton births because the reference values for SGA definition are established based on singleton births.

Stillbirth/fetal death has a number of different and legally mandated definitions (2). This study will use ACOG's definition which defines stillbirth/fetal deaths at 20 completed weeks gestation or greater (if the gestational age is known). Maternal death will include all non-accidental deaths while pregnant or within 42 days of end of pregnancy. Neonatal death will include deaths within 28 days of birth among KPSC live born infants.

For a multiple gestation pregnancy, the pregnancy outcome will be recorded for each fetus, while the endpoint will be defined by the worst scenario. For examples, the woman will be considered to have a pre-term delivery, if any fetus was delivered pre-term after the index date. The woman will also be considered to have spontaneous abortion, therapeutic abortion, or stillbirth/fetal death if any fetus experienced these pregnancy outcomes after the index date.

For a multiple gestation pregnancy with multiple live births, only the infant(s) with infant AEs (neonatal death and congenital anomalies) will be chart reviewed. The endpoint will be defined at the infant level.

Additional birth outcomes will be captured among KPSC born infants for descriptive purposes, but will not be considered as AEs: birth length, head circumference, and Apgar scores.

3.2.4 Covariates

Covariates (Table 2) to be collected and explored include medical conditions indicative of a high-risk pregnancy and other relevant baseline characteristics. The algorithms for identification of high-risk medical conditions are described in the DMP.

Table 2. List of covariates (all covariates will be treated as categorical in the analyses)

High-risk medical condition
Thyroiditis (present/absent)
Diabetes (present/absent)
Other autoimmune diseases (present/absent)
Renal diseases(present/absent)
Cardiac diseases (present/absent)
Pre-pregnancy obesity (BMI>40: yes/no/missing)
Pre-existing hypertension (present/absent)
Rhesus sensitization (present/absent)
HIV infection (present/absent)
Syphilis infection (present/absent)
Cervical incompetence (present/absent)
Lifestyle factors
Smoking status from pregnancy start to index date (Yes/No)
Alcohol use from pregnancy start to index date (Yes/No)

Enrollment
Length of membership prior to 27 weeks gestation (but after pregnancy start) – having continuous membership from pregnancy start to index date (Yes/No)
Medicaid insurance (public health insurance: Indicator of low-income status) – any Medicaid enrollment from pregnancy start to index date (Yes/No)
Utilization
Health care utilization (receipt of medical care) from pregnancy start to index date Any hospitalization or ED visit (Yes/No) Number of outpatient visits (0-4/5-12/≥13)
Maternal Data
Number of previous pregnancies (0/1/≥2/Missing)
History of pregnancy loss (Yes/No)
Season of pregnancy start (September-March/April-August)
Vaccine
Receipt of other vaccines during pregnancy, e.g., influenza vaccine (Yes/No)
Receipt of vaccines containing diphtheria, tetanus toxoid or pertussis antigens within one year before pregnancy (Yes/No)
Matching variables
Race/Ethnicity (White/Black/Hispanic/Asian/Other)
Multiple gestation (Yes/No)
Age at pregnancy start (years)

Smoking status will be defined as current smoker or not. If patients reported they smoked any time from pregnancy start to index date, then they will be current smokers. Alcohol use will be defined in a similar fashion.

Length of membership prior to 27 weeks gestation (but after pregnancy start) will be defined as a dichotomous variable of having continuous membership (allowing up to a 31-day gap) from pregnancy start to index date.

Health care utilization will be defined as any hospitalization or emergency department (ED) visit, and number of outpatient visits (0-4, 5-12, ≥13) from pregnancy start to index date.

Receipt of other vaccines during pregnancy will be defined as any receipt of other vaccines from pregnancy start to index date.

The covariates included in adjusted analyses will be determined by scientific relevance, association with exposure and outcome, and data availability. Specifically, we will select covariates by following these steps:

1. The distribution of covariates will be reviewed. Those covariates in the same category with rare events will likely be combined as a single covariate. For example, we might consider grouping all chronic conditions into a single covariate.
2. The data availability will be checked. If a covariate contains a large proportion that is missing, and the proportion missing is differential by exposure, then the covariate may not be included.
3. Association of covariates with exposure will be assessed. We will use standardized difference to assess the balance of covariates between exposed and unexposed cohorts. Unlike p-values, for which magnitude is highly related to sample size, standardized difference is a unified approach to quantifying the magnitude of difference between groups regardless of sample size, where an absolute value less than 0.1 is considered a negligible difference (3). We will use a SAS macro %stddiff developed by PPD and PPD (4) to calculate the standardized differences for continuous and categorical variables. Potential confounders will be determined by absolute standardized difference (ASD) >0.1.
4. Association of potential confounders (identified from step 3) with each primary endpoint will be further assessed using a change-in-estimate approach (5). Diabetes, pre-existing hypertension (except for the analysis of the preeclampsia and/or eclampsia endpoint), Medicaid insurance, and health care utilization, which are considered as the most representative prognosis factors, will be kept in the adjusted model regardless of their association with exposure and primary endpoints. For a specific endpoint, the exposure coefficient estimate from a model adjusted for the 4 above covariates will be compared with the estimate from a model adjusted for one more covariate at a time. Covariates which result in a 10% change in the exposure coefficient estimate will be selected for final adjusted analysis.
5. All potential confounders (identified from step 3) will be included in the analyses of secondary endpoints. If the outcome is rare and the adjusted model cannot converge, problematic covariates will be removed from the model. If total counts from the two groups (i.e., sum of events in the exposed and unexposed cohorts) is less than 5, then no adjusted analysis will be performed.

3.3 Sample Size and Power Considerations

The power computation for the primary objective was performed using PASS non- inferiority of two independent proportions using Miettinen and Nurminen method. Although the power is not based on the Poisson regression model, it is expected to reflect the power on the planned analysis since inference on proportion and incidence rate provide similar results when incidence rates are low.

The following table (Table 3) shows the power to rule out a two-fold increase considering an assumed possible design bias leading to a 1.25 relative risk increase.

To be able to make independent conclusions for each of the four primary endpoints while controlling the type I error below 2.5%, a Bonferroni adjustment will be used, i.e., each primary endpoint will be evaluated using a one-sided 0.625% (2.5% type I error / 4 primary endpoints) nominal type I error. More specifically, the upper limit of the two-sided 98.75% confidence interval (CI) will be used to make conclusions.

Considering that the four primary endpoints are expected to have an incidence rate above 1%, the power to conclude that maternal immunization is non-inferior for any primary endpoint is at least 95% and the power to conclude that maternal immunization is non-inferior for all 4 primary endpoints is at least 80% (100% - 4*beta).

Table 3. Power to rule out a two-fold increase with 15,000 pregnant women in each cohort

Incidence rate in unvaccinated cohort	One-sided type I nominal error	True relative risk	Power	Beta
1%	0.625%	1.25	95.7%	4.3%

4. ANALYSES OF OBJECTIVE

All analyses will be performed by statisticians at KPSC using the SAS statistical software package (version 9.4 or later).

4.1 Description of study population

The size of the study population and subgroups for main analysis will be described, overall and by exposed and unexposed cohorts. See Appendix A, Table Shell 1 for details. The number and percentage of pregnant women with unknown pregnancy outcome, lost to follow up, and multiple gestation will also be reported.

Baseline characteristics will be described and compared between the exposed and unexposed cohorts. See Appendix A, Table Shell 2 for details. Bivariate analyses will be conducted for all covariates of interest. Continuous variables such as age in years will be summarized by mean, median, Q1, Q3, range, and standard deviation and compared using t-tests; categorical variables will be summarized by frequency and percentage and compared using chi-square tests or Fisher's exact test, as appropriate. Absolute standardized difference will be calculated to assess the balance of covariates. Potential confounders will be determined based on bivariate analyses and scientific relevance (see Section 3.2.4).

Birth outcomes captured among KPSC born infants (birth weight, length, head circumference, and Apgar scores) will be described for exposed and unexposed cohorts. See Appendix A, Table Shell 3 for details.

4.2 Analyses of primary endpoints

We will calculate crude incidence and 95% CI for each primary endpoint for the exposed and unexposed cohorts. Unadjusted and adjusted RR with two-sided 98.75% CI will be estimated comparing the exposed cohort to the unexposed cohort, accounting for multiple comparisons.

Analysis of primary endpoints will be used to evaluate the safety of Boostrix vaccination and concluded to be non-inferior if the upper limit of the 98.75% CI for the adjusted relative risk is below 2.

Analyses of preeclampsia and/or eclampsia and intra-uterine infection will be performed on all pregnant women eligible for main analysis. For events identified any time after the index date (e.g., preeclampsia and/or eclampsia and intra-uterine infection), the incidence will consist of the total number of women with the condition in the numerator and the total person time in the denominator. The person-year for a pregnant woman will be the time from the index date to the date of each adverse event, end of the pregnancy, or disenrollment, whichever comes first. The unadjusted and adjusted relative risk with 98.75% CI will be estimated by Poisson regression models accounting for the follow-up time without and with adjustment for potential confounders. If the outcome is common (i.e., incidence>2%), a Poisson regression model with a robust error variance will be used to estimate the relative risk (6). The robust Poisson regression will be performed using the SAS Proc GENMOD procedure with the REPEATED statement subject= Patient_ID.

Analysis of SGA will be performed on live singleton infants born in KPSC hospitals, while analysis of preterm delivery will be performed on pregnant women with live births, and with an index date prior to 37 weeks gestation. For events identified at delivery (e.g., SGA and preterm delivery), the incidence will consist of the total number of women or infants with the condition in the numerator and the number of women or infants for whom the condition can be assessed in the denominator. The unadjusted and adjusted relative risk with 98.75% CI will be estimated by Poisson regression models with robust error variances without and with adjustment for potential confounders. The planned methods of analysis by primary endpoint are outlined in Table 4. See Appendix A, Table Shell 4 for details.

Table 4. Analytic population and statistical methods by primary endpoint

Primary endpoints	Analytic population	Statistical Methods
Preeclampsia and/or eclampsia	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time, with robust variance estimate
Intra-uterine infections	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time, with robust variance estimate
SGA	KPSC live born infant, singleton only*	Adjusted Poisson regression model, with robust variance estimate
Preterm delivery	Pregnant women with live birth, with index date prior to 37 weeks gestation	Adjusted Poisson regression model, with robust variance estimate

*There are no well accepted SGA metrics for infants born to pregnant women with multiple gestation

Limiting analyses of endpoints to subgroups (e.g., KPSC live born infant) may cause matches to be broken, although we expect the proportions of non-live births and non-KPSC births to be small.

The matching variables will be included in the adjusted models of analyses of SGA and preterm delivery. Age at pregnancy start will be adjusted as a categorical variable (<24/25-34/≥35 years).

4.3 Analyses of secondary endpoints

The analysis of secondary endpoints will be used for hypothesis generation/signal detection. We will calculate crude incidence and 95% CI for each secondary endpoint for the exposed and unexposed cohorts. Unadjusted and adjusted RR with two-sided 95% CI will be estimated comparing the exposed cohort to the unexposed cohort. At least one case in each cohort is required to perform unadjusted and adjusted analyses. If total counts from the two groups is less than 5, then no adjusted analysis will be performed. An elevated risk will be detected if the lower limit of the 95% CI for the adjusted relative risk is above 1. The planned methods of analysis by secondary endpoint are outlined in Table 5. See Appendix A, Table Shell 4 for details.

Table 5. Analytic population and statistical methods by secondary endpoint

Secondary endpoints	Analytic population	Statistical Methods
Transfusion during delivery hospitalization	Pregnant women with delivery hospitalization	Adjusted Poisson regression model
Poor fetal growth	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time, with robust variance estimate
Stillbirth/fetal death*	Pregnant women with pregnancy outcome	Adjusted Poisson regression model
Neonatal death	KPSC live born infants	Adjusted Poisson regression model
Maternal death	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time
Placental abruption	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time
PPROM	All pregnant women, with index (vaccination) date prior to 37 weeks gestation	Adjusted Poisson regression model accounting for follow up time
Congenital anomalies (CA) by type	KPSC live born infants	Adjusted Poisson regression model, with robust variance estimate if outcome is common

*Stillbirth/fetal death will be described with or without congenital anomalies separately, but will be analyzed as one outcome.

For poor fetal growth, placental abruption, PPROM, and maternal death, the incidence will consist of the total number of women with the condition in the numerator and the total person time in the denominator. For poor fetal growth, placental abruption, and PPROM, the person-time for a pregnant woman will be the time from the index date to the date of each adverse event, end of the pregnancy, or disenrollment, whichever comes first. For maternal death, the person-time for a pregnant woman will be the time from the index date to the date of death, 42 days after delivery, or disenrollment, whichever comes first. The unadjusted and adjusted relative risk with 95% CI will be estimated by Poisson regression models, or Poisson regression models with robust error variances if incidence>2%, accounting for the follow-up time without and with adjustment for potential confounders.

For stillbirth/fetal death and transfusion during delivery hospitalization, the incidence will consist of the total number of women with the condition in the numerator and the number of women for whom the condition can be assessed in the denominator. The unadjusted and adjusted relative risk with 95% CI will be estimated by Poisson regression models, without and with adjustment for potential confounders.

Neonatal death and congenital anomalies will be identified among infants born in KPSC hospitals. The incidence of neonatal death will be calculated as the number of neonatal deaths in the numerator and the total number of infants born in KPSC hospitals in the denominator. The prevalence of congenital anomalies (overall, by type) will be calculated as the number of infants with a congenital anomaly in the numerator and the total number of infants born in KPSC hospitals in the denominator. The proportion of live births without congenital anomalies will be calculated as 1 minus the proportion of live births with congenital anomalies among all KPSC live born infants. The unadjusted and adjusted relative risk with 95% CI will be estimated by Poisson regression models, or Poisson regression models with robust error variances if incidence>2%, without and with adjustment for potential confounders.

4.4 Analysis of pregnant women vaccinated before the 27th week of gestation

The analysis of pregnant women vaccinated with Boostrix before the 27th week of gestation will be descriptive. Due to the expected extremely small sample size, no formal hypothesis testing is planned.

The size of the study population and subgroups will be described. See Appendix A, Table Shell 6 for details. Baseline characteristics of pregnant women and their KPSC live born infants will be described. Continuous variables will be summarized by mean, median, Q1, Q3, range, and standard deviation; categorical variables will be summarized by frequency and percentage. See Appendix A, Table Shell 7 and Table Shell 8 for details.

The number of events for each primary and secondary endpoint in Table 1 among pregnant women vaccinated with Boostrix before the 27th week of gestation will be reported. The incidence will be calculated as the total number of women with the condition in the numerator and the number of women (or total person time) for whom the condition can be assessed in the denominator. Since we will likely observe no events or few events for most of the endpoints, the 95% Wilson score interval with continuity correction will be provided (7). See Appendix A, Table Shell 9 for details.

4.5 Signal investigation

If a primary endpoint analysis cannot rule out a 2-fold relative risk, or secondary endpoint analysis detects an elevated risk, further exploration of the data will be completed to investigate alternative potential sources for the Boostrix-outcome association.

Description of events

Events will be summarized by age at pregnancy start, gestational age at event date, receipt of other vaccines during pregnancy, and other risk factors associated with the signaled endpoint. Stratified incidence will be calculated as the number of events in a specific stratum in the numerator and the total number of persons or person-time of the stratum in the denominator for the exposed and unexposed cohorts. If the endpoint is chart reviewed, chart confirmation rates for the exposed and unexposed cohorts will be provided. See Appendix A, Table Shell 10 for details.

In addition, depending on the AE that signals, secular trends may be assessed. The KPSC team will work together with GSK team to conduct signal investigation, with consideration of timelines and data/resource availability.

5. STATISTICAL CONSIDERATIONS AND LIMITATIONS

5.1 Handling of missing data

Medical record review will be conducted on unknown pregnancy outcomes and KPSC live born singleton infants with unknown birthweight. We expect that there will be a very small amount of missing data after medical record review. Data with unknown pregnancy outcome will be excluded from the analysis of stillbirth; data with birthweight missing will be excluded from the analysis of SGA. The data can be kept for analyses of other endpoints if they have pregnancy end dates.

Medical record review will be performed on pregnant women who have continuous membership, but who are missing pregnancy end dates. If the pregnancy end date remains missing after medical record review, the last contact date with KPSC before the estimated date of delivery (EDD) will be used as the end date of follow up. The patient will be considered lost to follow up.

No imputation is planned for missing data in covariates. Unknown race/ethnicity is grouped into the "Other" category by convention, which can be matched between exposed and unexposed cohorts. Missing number of previous pregnancies and missing pre-pregnancy weight will be assessed. If they are included in the adjusted model, "Missing" will be a category for the covariates along with the other categories. Missing birth length, head circumference, and Apgar scores will be described as is.

5.2 Multiplicity considerations

Analyses of the primary endpoints will be conducted using a non-inferiority design with multiple comparison adjustment. Analyses of the secondary endpoints will be conducted using superiority testing without adjustment for multiple comparisons. The approach leads to a more conservative

conclusion regarding safety of vaccination, i.e., for a non-inferiority test, it is harder to reject Ho: RR≥2, and for a superiority test, it is easier to reject Ho: RR=1.

The analyses of secondary endpoints will be used for hypothesis generation/signal detection. No formal adjustment for multiple comparisons is common practice for this type of study. Since outcomes monitored in vaccine safety studies are usually serious, researchers do not want to miss the opportunity to detect potential safety findings.

5.3 Margin for non-inferiority testing

The non-inferiority design allows us to account for possible bias from an observational study design which may not be fully adjusted for in the analysis, and conclude that the risk is less than a threshold, say RR<2. The non-inferiority margin was determined based on prior experience from post-marketing vaccine safety studies, the magnitude of possible bias, and sample size considerations. While a non-inferiority margin as low as 1.8 or 1.5 might be preferable, it would require a much larger sample size.

5.4 Limitations

Secular confounding

Using a historical comparison group may introduce bias due to secular confounding. The events identified from a current exposed cohort and a historical unexposed cohort may differ in background rates due to care improvement, changes in diagnostic criteria, changes in health-seeking behavior, and changes in the coding system. This is an inherent limitation of using a historical comparison and needs to be acknowledged. The approach we propose, using a comprehensive list of codes to identify all potential events of interest, followed by a review of medical records to confirm the incident events, aims to minimize the bias caused by changing coding practices.

Confounding by indication

Compared to a concurrent unexposed cohort (assuming approximately 80% Tdap uptake among pregnant women), a historical unexposed cohort would be less selective and more representative of an unvaccinated cohort. However, confounding by indication is still possible. A number of covariates including high-risk medical conditions, health care utilization, pregnancy history, etc. will be collected and possibly adjusted to minimize bias.

6. CDISC

The analytic datasets will be formatted according to CDISC standards. The KPSC team will work with a CDISC standards consultant to develop Analysis Data Model (ADaM) specifications and create ADaM datasets in accordance with the Study Data Tabulation Module (SDTM) datasets, study report, and CDISC implementation guide; validate the ADaM datasets using an open source CDISC validation program (e.g., Pinnacle 21 or similar); and produce submission documentation including DEFINE.XML and Analysis Data Reviewer Guide (ADRG).

The development of ADaM specifications and ADaM datasets will parallel statistical analysis activities. Specifically, draft ADaM specifications and datasets will be developed according to the SAP before conducting analysis. The ADaM specifications and datasets will be modified if any analyses deviate from the pre-specified SAP which can cause a change in the analytic dataset. The ADaM specifications and datasets will be finalized once the results tables are finalized. Consistency of final ADaM datasets produced against the specifications will be evaluated, and compliance checks using a validation tool (e.g. Pinnacle 21 or similar) including validation of standard compliance, controlled terminology, and data quality will be conducted on the dataset to ensure compliance with CDISC standards for ADaM. Consistency of the results tables produced against the final ADaM datasets will be evaluated too. Additionally, KPSC will ensure and oversee that the CDISC consultancy has adequate procedures and controls in place to guarantee that the resulting CDISC package (i.e., datasets and supporting documentation) is of adequate quality.

7. QUALITY CONTROL

KPSC will follow the Department of Research and Evaluation's Analytical Research Project Standard Operating Procedures (SOP), including data extraction, data analysis, communication, and documentation. SAS programs used for the analysis will be provided to GSK for review. Statistical analysis (results tables) will be discussed in the multisite meetings with GSK prior to study report finalization.

Quality control steps

During the analysis, when analytic data are created and analyzed, the following quality control steps will be performed by the study statisticians as appropriate:

1. Read study protocol, DMP, and the statistical analysis plan. Communicate with the principal investigator, senior statistician, and study team with any questions regarding study design and analysis plan. Document any decisions related to statistical analysis.

2. Check data quality (sample size, duplicates, range of values, number of missings, proper data type and format, inconsistency, etc.). Any aberrant data that have been missed by data cleaning and discovered by the statistician will be described in the results table.

3. Participate in 2 days of CDISC training on SDTM, ADaM, and Define.xml. Participate in 1-day CDISC study-specific workshop. Become familiar with SDTM and ADaM data standards. Work with CDISC standards consultant to develop ADaM specifications and create ADaM datasets according to SAP.

4. Develop SAS programs for analysis. Programs will include a program header and section headers. The program header contains information on the name, location, and purpose of the program, a list of prerequisite programs if applicable, a list of input datasets with full path, a list of output datasets, and a history log. Revision of the program will be noted in the history log to capture the change made, the change date, identification of the individual making the change, and type of change (e.g. modification, addition, removal). Each program might have several sections

executing various analytic steps. Programs will be annotated with comments that describe the intent or purpose.

5. Perform statistical analysis. Use appropriate procedures and tests. SAS logs and outputs will be reviewed for any warnings, error messages, and model convergence issues. Organize the output into results tables. When possible, automate the generation of results tables to minimize the need to copy/paste. Proofread the results tables to prevent any copy/paste and editing errors. SAS logs and outputs will be saved.

6. Use established SAS macro. Use established SAS macros for analyses, including but not limited to the macro for creating baseline covariate tables and the macro for calculating standardized difference.

Program review

All SAS analysis programs, logs and outputs will be reviewed by a second statistician. The second statistician will review the program logic against the analysis requirements. Comments from the second statistician will be documented and followed up until resolved.

Quality control findings

A record of quality problems and resolutions will be kept in the form of an Excel spreadsheet or other method appropriate to the circumstances (e.g. commented code, detailed correspondence).

Analysis deviation

Any analyses differing from the protocol, DMP and pre-specified analyses in this SAP will be documented. The rationale for the change in approach will be provided.

8. STUDY REPORTS

The final report will include all data generated from this study as specified in this SAP.

9. LIST OF REPORT TABLES

	Description	Population
Table 1	Study population and subgroups for main analysis	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 2	Baseline characteristics of Boostrix exposed and unexposed pregnant women	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 3	Characteristics of KPSC live born infants with Boostrix exposed and unexposed mothers	KPSC live born infants with exposed and unexposed mother

		with index (vaccination) date after the 27th week of gestation
Table 4*	Incidence and relative risk of primary endpoints comparing Boostrix exposed and unexposed cohorts	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 5*	Incidence and relative risk of secondary endpoints comparing Boostrix exposed and unexposed cohorts	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 6	Study population and subgroups for descriptive analysis	Pregnant women vaccinated with Boostrix before the 27th week of gestation
Table 7	Baseline characteristics of pregnant women vaccinated with Boostrix before the 27th week of gestation	Pregnant women vaccinated with Boostrix before the 27th week of gestation
Table 8	Characteristics of KPSC live born infants with mothers vaccinated with Boostrix before the 27th week of gestation	KPSC live born infants with mothers vaccinated with Boostrix before the 27th week of gestation
Table 9	Incidence of primary and secondary endpoints among pregnant women vaccinated with Boostrix before the 27th week of gestation	Pregnant women vaccinated with Boostrix before the 27th week of gestation
Table 10	Summary of signaled events	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation

* The details of the model for each endpoint will be shared with GSK, but will not be included in the report tables.

10. REFERENCES

1. Talge NM, Mudd LM, Sikorskii A, Basso O. United States birth weight reference corrected for implausible gestational age estimates. Pediatrics. 2014;133(5):844-853.

2. Tavares Da Silva F, Gonik B, McMillan M, et al. Stillbirth: Case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine. 2016;34(49):6057–6068.

3. Austin PC**.** An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.

4. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum, Statistics and Data Analysis. 2012;Paper 335-2012.

5. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340-349.

6. Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702-6.

7. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine. 1998 17 (8): 857–872.

11. APPENDIX A: TABLE SHELLS

Appendix A: Table Shells

Table 1: Study population and subgroups for main analysis

	Exposed	Unexposed	Total
Pregnant women eligible per inclusion/exclusion criteria, n			
With pregnancy outcome, n (%)			
With unknown pregnancy outcome, n (%)			
Lost to follow up, n (%)			
Pregnant women with multiple gestation, n (%)			
Pregnant women with delivery hospitalization, n (%)			
Pregnant women with index date prior to 37 weeks gestation, n (%)			
With live birth, n (%)			
KPSC live born infants, n			
Singleton only, n (%)			

Table 2. Baseline characteristics[a] of Boostrix exposed and unexposed pregnant women

		Exposed N =	Unexposed N =	Total N =	p value
	Maternal age at pregnancy st				N/A[d]
mean (sd)					
median					
Q1, Q3					
range					
	Race/ethnicity, n (%)				N/A
White					
Black					
Hispanic					
Asian					
Other					
	Gestational age in weeks at i				N/A
27-30					
31-36					
≥37					
Current smoker, n (%)					
Yes					
No					
Alcohol use, n (%)					
Yes					
No					
Continuous membership, n (%)					
Yes					
No					
Medicaid insurance, n (%)					
Yes					
No					
Any hospitalization or ED visit, n (%)					
Yes					
No					
Number of outpatient visits, n (%)					
0-4					
5-12					
≥13					
Number of previous pregnancies, n (%)					
0					
1					
≥2					
Missing					
History of pregnancy loss, n (%)					
Yes					
No					
Season of pregnancy start, n (%)					
September-March					
April-August					
Received other vaccines[b], n (%)					
Yes					
No					
Received vaccines containing diphtheria, tetanus toxoid, or pertussis antigens within one year before pregnancy, n (%)					
Yes					
No					
High-risk medical condition, n (%)					
	Multiple gestation				N/A
Thyroiditis					
Diabetes					
Other autoimmune diseases					
Renal diseases					
Cardiac diseases					
Pre-pregnancy obesity[c]					
Missing					
Pre-existing hypertension					
Rhesus sensitization					
HIV infection					
Syphilis infection					
Cervical incompetence					

[a] Baseline characteristics were defined from pregnancy start to index date, unless otherwise specified

[b] Other vaccines included: inactivated influenza vaccine (xx%), hepatitis A/hepatitis B vaccine (xx%), human papillomavirus vaccine (xx%), and other (xx%).

[c] BMI>40 within 6 months prior to pregnancy start

[d] N/A= not applicable, for matching variable

Appendix A: Table Shells

Table 3. Characteristics of KPSC live born infants with Boostrix exposed and unexposed mothers

	Exposed N =	Unexpose N =
Sex, n (%)		
F		
M		
Gestational age in weeks at birth, n (%)		
27-33		
34-36		
37-39		
≥40		
Birth weight, g		
mean (sd)		
median		
Q1, Q3		
range		
Missing		
Birth length, cm		
mean (sd)		
median		
Q1, Q3		
range		
Missing		
Head circumference, cm		
mean (sd)		
median		
Q1, Q3		
range		
Missing		
Apgar 1 score, n (%)		
≤3		
4-6		
≥7		
Missing		
Apgar 5 score, n (%)		
≤3		
4-6		
≥7		
Missing		

Table 4. Incidence and relative risk of primary endpoints comparing Boostrix exposed and unexposed cohorts

	Exposed				Unexposed				Relative Risk[e] (Wald 98.75% CI)	
	Number of Persons	Number of Person years	Number of adverse events	Incidence per 1000 persons or person-years (95% CI)	Number of Persons	Number of Person years	Number of adverse events	Incidence per 1000 persons or person-years (95% CI)	Unadjusted	Adjusted
Pre-eclampsia and eclampsia[a]										
Intra-uterine infections[a]										
Small for gestational age [b]		N/A*				N/A				
Preterm delivery [c]		N/A				N/A				

[a] Identified among all pregnant women eligible for main analysis
[b] Identified among KPSC live born infants, singleton only
[c] Identified among pregnant women with live birth, with index (vaccination) date prior to 37 weeks gestation
[d] Adjusted for covariates XXX (place holder, covariates may differ for each endpoint)
[e] Poisson regression model with robust variance estimate
* N/A=Not Applicable

Table 5. Incidence and relative risk of secondary endpoints comparing Boostrix exposed and unexposed cohorts

	Exposed				Unexposed				Relative Risk (Wald 95% CI)	
	Number of Persons	Number of Person years	Number of events	Incidence per 1000 persons or person-years (95% CI)	Number of Persons	Number of Person years	Number of events	Incidence per 1000 persons or person-years (95% CI)	Unadjusted	Adjusted
Poor fetal growth, IUGR[a]										
Placental abruption[a]										
Preterm pre-labor rupture of membranes (PPROM)[b]										
Maternal death [a]										
Stillbirth/fetal death[c]		N/A*				N/A				
With congenital anomalies		N/A				N/A			N/A	N/A
Transfusion during delivery hospitalization[d]		N/A				N/A				
Neonatal death [e]		N/A				N/A				
Congenital anomalies at birth and through 6 months of age[d]		N/A				N/A			N/A	N/A
Congenital anomalies of nervous system		N/A				N/A				
Congenital anomalies of eye		N/A				N/A				
Congenital anomalies of ear, face, or neck		N/A				N/A				
Congenital anomalies of cardiovascular system		N/A				N/A				
Congenital anomalies of respiratory system		N/A				N/A				
Clefts		N/A				N/A				
Congenital anomalies of upper gastrointestinal system		N/A				N/A				
Congenital anomalies of lower gastrointestinal system		N/A				N/A				
Congenital anomalies of genital organs		N/A				N/A				
Congenital anomalies of renal system		N/A				N/A				
Congenital anomalies of musculoskeletal system		N/A				N/A				
Congenital anomalies of limb		N/A				N/A				
Congenital anomalies of integument		N/A				N/A				
Other and unspecified congenital anomalies		N/A				N/A				
Number of congenital anomalies (based on body system) per infant		N/A		N/A		N/A		N/A	N/A	N/A
0		N/A		N/A		N/A		N/A	N/A	N/A
1		N/A		N/A		N/A		N/A	N/A	N/A
2		N/A		N/A		N/A		N/A	N/A	N/A
3		N/A		N/A		N/A		N/A	N/A	N/A
≥4		N/A		N/A		N/A		N/A	N/A	N/A

[a] Identified among all pregnant women eligible for main analysis
[b] Identified among pregnant women with index (vaccination) date prior to 37 weeks gestation
[c] Identified among pregnant women with pregnancy outcome
[d] Identified among pregnant women with delivery hospitalization
[e] Identified among KPSC live born infants
[f] Adjusted for covariates XXX (place holder, covariates may differ for each endpoint)
[g] Poisson regression model with robust variance estimate (place holder)
* N/A=Not Applicable

Appendix A: Table Shells

Table 6: Study population and subgroups for descriptive analysis

	Exposed
Pregnant women eligible per inclusion/exclusion criteria, n	
With pregnancy outcome, n (%)	
With unknown pregnancy outcome, n (%)	
Lost to follow up, n (%)	
Pregnant women with multiple gestation, n (%)	
Pregnant women with delivery hospitalization, n (%)	
Pregnant women with index date prior to 37 weeks gestation, n (%)	
With live birth, n (%)	
KPSC live born infants, n	
Singleton only, n (%)	

Table 7. Baseline characteristics[a] of pregnant women vaccinated with Boostrix before the 27th week of gestation

	Exposed N =
Maternal age at pregnancy start, years	
mean (sd)	
median	
Q1, Q3	
range	
Race/ethnicity, n (%)	
White	
Black	
Hispanic	
Asian	
Other	
Gestational age in weeks at index date, n (%)	
27-30	
31-36	
≥37	
Current smoker, n (%)	
Yes	
No	
Alcohol use, n (%)	
Yes	
No	
Continuous membership, n (%)	
Yes	
No	
Medicaid insurance, n (%)	
Yes	
No	
Any hospitalization or ED visit, n (%)	
Yes	
No	
Number of outpatient visits, n (%)	
0-4	
5-12	
≥13	
Number of previous pregnancies, n (%)	
0	
1	
≥2	
Missing	
History of pregnancy loss, n (%)	
Yes	
No	
Season of pregnancy start, n (%)	
September-March	
April-August	
Received other vaccines[b], n (%)	
Yes	
No	
Received vaccines containing diphtheria, tetanus toxoid, or pertussis antigens within one year before pregnancy, n (%)	
Yes	
No	
High-risk medical condition, n (%)	
Multiple gestation	
Thyroiditis	
Diabetes	
Other autoimmune diseases	
Renal diseases	
Cardiac diseases	
Pre-pregnancy obesity[c]	
Missing	
Pre-existing hypertension	
Rhesus sensitization	
HIV infection	
Syphilis infection	
Cervical incompetence	

[a] Baseline characteristics were defined from pregnancy start to index (vaccination) date, unless otherwise specified

[b] Other vaccines included: inactivated influenza vaccine (xx%), hepatitis A/hepatitis B vaccine (xx%), human papillomavirus vaccine (xx%), and other (xx%).

[c] BMI>40 within 6 months prior to pregnancy start

Table 8. Characteristics of KPSC live born infants with mothers vaccinated with Boostrix before the 27th week of gestation

	Exposed N =
Sex, n (%)	
F	
M	
Gestational age in weeks at birth, n (%)	
27-33	
34-36	
37-39	
≥40	
Birth weight, g	
mean (sd)	
median	
Q1, Q3	
range	
Missing	
Birth length, cm	
mean (sd)	
median	
Q1, Q3	
range	
Missing	
Head circumference, cm	
mean (sd)	
median	
Q1, Q3	
range	
Missing	
Apgar 1 score, n (%)	
≤3	
4-6	
≥7	
Missing	
Apgar 5 score, n (%)	
≤3	
4-6	
≥7	
Missing	

Table 9. Incidence of primary and secondary endpoints among pregnant women vaccinated with Boostrix before the 27th week of gestation

	Exposed			
	Number of Persons	**Number of Person years**	**Number of events**	**Incidence per 1000 persons or person-years (95% CI)**
Primary endpoints				
Pre-eclampsia and eclampsia[a]				
Intra-uterine infections[a]				
Small for gestational age [b]		N/A*		
Preterm delivery [c]		N/A		
Secondary endpoints				
Poor fetal growth, IUGR[a]				
Placental abruption[a]				
Preterm pre-labor rupture of membranes (PPROM)[d]				
Maternal death [a]				
Stillbirth/fetal death[e]		N/A		
With congenital anomalies		N/A		
Spontaneous abortion [e]		N/A		
With congenital anomalies		N/A		
Therapeutic abortion[e]		N/A		
With congenital anomalies		N/A		
Transfusion during delivery hospitalization[f]		N/A		
Neonatal death [g]		N/A		
Congenital anomalies at birth and through 6 months of age[f]		N/A		
Congenital anomalies of nervous system		N/A		
Congenital anomalies of eye		N/A		
Congenital anomalies of ear, face, or neck		N/A		
Congenital anomalies of cardiovascular system		N/A		
Congenital anomalies of respiratory system		N/A		
Clefts		N/A		
Congenital anomalies of upper gastrointestinal system		N/A		
Congenital anomalies of lower gastrointestinal system		N/A		
Congenital anomalies of genital organs		N/A		
Congenital anomalies of renal system		N/A		
Congenital anomalies of musculoskeletal system		N/A		
Congenital anomalies of limb		N/A		
Congenital anomalies of integument		N/A		
Other and unspecified congenital anomalies		N/A		
Number of congenital anomalies (based on body system) per infant		N/A		N/A
0		N/A		N/A
1		N/A		N/A
2		N/A		N/A
3		N/A		N/A
≥4		N/A		N/A

[a] Identified among all pregnant women eligible for descriptive analysis

[b] Identified among KPSC live born infants, singleton only

[c] Identified among pregnant women with live birth, with index (vaccination) date prior to 37 weeks gestation

[d] Identified among pregnant women with index (vaccination) date prior to 37 weeks gestation

[e] Identified among pregnant women with pregnancy outcome

[f] Identified among pregnant women with delivery hospitalization

[g] Identified among KPSC live born infants

[h] Wilson score interval with continuity correction

* N/A=Not Applicable

Table 10. Summary of signaled events

	Exposed			Unexposed		
	Number of events	Number of Persons (or person years)	Incidence per 1000 persons or person-years (95% CI)	Number of events	Number of Persons (or person years)	Incidence per 1000 persons or person-years (95% CI)
Number of confirmed events		N/A	N/A		N/A	N/A
Number of potential cases identified by the algorithm		N/A	N/A		N/A	N/A
Confirmation rate		N/A	N/A		N/A	N/A
Age at pregnancy start, years						
<18						
18-24						
25-34						
≥35						
Received other vaccines during pregnancy, n (%)						
Yes						
No						
High-risk medical condition, n (%)						
Thyroiditis						
Yes						
No						
Diabetes						
Yes						
No						
…						
Gestational age at event date						
27th week		N/A	N/A		N/A	N/A
28th week		N/A	N/A		N/A	N/A
29th week		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
ICD diagnosis code and description						
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A

Statistical Analysis Plan

Study Title:	The safety of Boostrix following routine immunization of pregnant women
Protocol Number:	EPI-PERTUSSIS-047 VS US DB (207221)
Sponsor:	GlaxoSmithKline
Coordinating/Principal Investigator(s):	Hung Fu Tseng Department of Research and Evaluation Kaiser Permanente Southern California [redacted] PPD [redacted] [redacted] USA
Epidemiologist:	PPD [redacted] Epidemiology Lead DTP GSK Biologicals
Plan Prepared by:	PPD [redacted] Research Scientist Biostatistician Kaiser Permanente Southern California

Approvals

Approvers, Organization & Role	Approval Status	Signature	Date
PPD KPSC Research Scientist Biostatistician	Approved	PPD PPD (Jul30, 2019)	Jul 30, 2019
PPD GSK Biostatistics and Statistical Programming	Approved	PPD PPD (Jul 31, 2019)	Jul 31, 2019

Publication Record

Version	Date	Primary Author	Description
Draft 1.0	14 June 2019	PPD	1st Draft for GSK Review
Final 1.0	26 July 2019	PPD	Final version incorporating GSK comments/edits

ABBREVIATIONS

ACIP: Advisory Committee on Immunization Practices

ACOG: American College of Obstetricians

ADaM: Analysis Data Model

AE: Adverse Event

ASD: Absolute Standardized Difference

BMI: Body mass index

CA: Congenital anomalies

CDISC: Clinical Data Interchange Standards Consortium

CI: Confidence interval

DMP: Data Management Plan

ED: Emergency Department

EDD: Estimated Date of Delivery

EHR: Electronic Health Record

GSK: GlaxoSmithKline

HIV: Human immunodeficiency virus

ICD: International Classification of Diseases

KPSC: Kaiser Permanente Southern California

MRN: Medical Record Number

PHI: Protected Health Information

PPROM: Preterm pre-labor rupture of membranes

PPV: Positive Predictive Value

R&E: Research and Evaluation

RR: Relative Risk

SAP: Statistical Analysis Plan

SAS: Statistical Analysis System

SDTM: Study Data Tabulation Module

SGA: Small for Gestational Age

SOP: Standard Operating Procedures

Tdap: Tetanus, diphtheria, acellular pertussis

TABLE OF CONTENTS

1. INTRODUCTION

Whooping cough, also referred to as pertussis, is a highly contagious respiratory disease primarily caused by the bacterium *Bordetella pertussis*. The disease is most severe in unvaccinated or incompletely vaccinated young infants. To increase the protection of newborns, in October 2012, the Advisory Committee on Immunization Practices (ACIP) recommended tetanus-diphtheria-acellular pertussis (Tdap) vaccine to be administered during pregnancy to all pregnant women, regardless of whether they have received Tdap previously. The optimal timing for Tdap administration is between 27 and 36 weeks of gestation.

Two Tdap vaccines (Adacel and Boostrix) were licensed for use in the United States in 2005. These vaccines were recommended for routine use in non-pregnant adolescents and adults when introduced. Both Adacel and Boostrix can be used interchangeably. There are limited post-marketing safety data on Boostrix during pregnancy. Through partnership between Kaiser Permanente Southern California (KPSC) and the sponsor, GlaxoSmithKline (GSK), we will conduct an observational, electronic health record (EHR) database study to assess the safety of Boostrix following routine immunization of pregnant women in a community setting.

This Statistical Analysis Plan (SAP) provides a detailed description of the data considerations, statistical methods, data analysis and results presentation as previously proposed in the protocol and Data Management Plan (DMP), where applicable. The SAP also describes quality assurance procedures and analytic data format standards.

2. STUDY OBJECTIVE

2.1. Primary objective

To rule out a two-fold increase (non-inferiority testing) in the incidence of each of the following maternal and infant adverse events (AEs):

- Preeclampsia and/or eclampsia
- Intra-uterine infections such as chorioamnionitis and endometritis
- Small for gestational age (SGA)
- Preterm delivery

among women who were vaccinated with Boostrix on or after the 1st day of the 27th week of gestation as compared to the incidence in a historical cohort of pregnant women (approximately January 1, 2012 to December 31, 2013) who were unvaccinated with any Tdap vaccine throughout their pregnancy.

The null hypothesis and alternative hypothesis for test of relative risk (RR) are:

$$\text{Ho: } RR \geq 2$$

$$\text{Ha: } RR < 2$$

2.2. Secondary objective

1. To explore differences (superiority testing) in the incidence rate of other maternal and infant AEs, namely:
   - Post-partum hemorrhage
   - Poor fetal growth
   - Stillbirth/fetal death with or without congenital anomalies
   - Neonatal death (within 28 days of birth)
   - Maternal death (while pregnant or within 42 days of end of pregnancy)
   - Placental abruption
   - Preterm pre-labor rupture of membranes (PPROM)
   - Congenital anomalies (CA) at birth and through 6 months of age

   between women who were vaccinated with Boostrix on or after the 1st day of the 27th week of gestation and a historical cohort of pregnant women (approximately January 1, 2012 to December 31, 2013) who were unvaccinated with any Tdap vaccine throughout their pregnancy.

   The null hypothesis and alternative hypothesis for test of RR are:

   $$Ho: RR = 1$$

   $$Ha: RR \neq 1$$

2. To describe the incidence of maternal and infant AEs described above, as well as the incidence of spontaneous and therapeutic abortion with or without congenital anomalies, in pregnant women who were vaccinated with Boostrix before the 27th week of gestation.

3. STUDY DESIGN

3.1 Overview of study design

This is an observational retrospective matched cohort study. The study will be conducted among pregnant women who are members of KPSC.

The exposed cohort (from the 27th week of gestation) will consist of pregnant women who received the Tdap vaccine (Boostrix) on or after the 1st day of the 27th week of pregnancy during the vaccination period at ob-gyn clinics of seven selected medical centers. Approximately 15,000 women vaccinated with Boostrix will be included in this study over approximately a one-year period from January 2018 to January 2019.

The unexposed cohort will consist of women, pregnant sometime during the approximate estimated period from 1/1/2012 to 12/31/2013 who never received Tdap vaccine during pregnancy. The unexposed cohort of pregnant women will be matched 1:1 to the exposed cohort of pregnant women for comparisons of maternal and infant AEs between the two cohorts.

Subject-level data will be collected for pregnant women and their infants through the EHR. For each individual pregnant woman, analytical follow up will begin on the date they receive the Boostrix vaccine, or the index date for the unvaccinated cohort, and will end on the date of

disenrollment or the end of pregnancy, whichever came first. For the maternal death endpoint, pregnant women will be followed for 42 days after the end of pregnancy. Infants born in KPSC hospitals will be followed for 6 months.

In addition, women vaccinated with Boostrix before the 27th week of gestation, with membership at the date of vaccination, and who were not vaccinated with any other Tdap vaccine at any other time during the pregnancy will be included for a descriptive analysis (single arm).

3.2 Study population, exposures, outcomes and covariates

3.2.1 Study populations/subgroups

Study population for main analysis (primary objective and secondary objective #1)

- Pregnant women received prenatal care at KPSC
- Exposed cohort: vaccinated with Boostrix on the 1st day of the 27th week of pregnancy or later during the pregnancy and were not vaccinated with any other Tdap vaccine at any other time during the pregnancy or
- Unexposed cohort: pregnant at least one day during the approximate estimated period between 1/1/2012-12/31/2013 and did not receive any Tdap vaccine during the pregnancy
- Continuous membership (allowing up to a 31-day gap) between the 1st day of the 27th week of pregnancy and the index (vaccination) date

The unexposed cohort will be 1:1 matched to the exposed cohort on age at pregnancy start (± 1 year), race/ethnicity (White, Black, Hispanic, Asian, and Other), and multiple gestation (yes/no). Every unexposed woman will be assigned an index date that is determined by the number of days from pregnancy start to the Boostrix vaccination date of her matched exposed woman. The same number of days calculated in the exposed woman will be added to the pregnancy start of the unexposed woman to determine her index date.

Study population for descriptive analysis (secondary objective #2)

- Pregnant women received prenatal care at KPSC
- Vaccinated with Boostrix before the 27th week of gestation and were not vaccinated with any other Tdap vaccine at any other time during the pregnancy
- Membership at the index (vaccination) date

Subgroups for endpoint-specific analysis

For some maternal and infant AEs (endpoints), additional restriction criteria will be applied. We will exclude individuals who do not meet the restriction criteria, but not the matched pair. Analysis will be conducted on a subgroup of the population:

- Small for gestational age (SGA): Infants 1) live born, 2) singleton, and 3) born in KPSC hospitals. Rationale: The evaluation of SGA is limited to singleton births because the reference values for SGA definition are established based on singleton births. The birth weight of infants born outside KPSC hospitals is not captured in claims, and such infants are therefore excluded for this endpoint.

- Preterm delivery: Pregnant women 1) with live birth and 2) with index (vaccination) date prior to 37 weeks gestation. Rationale: Pregnant women with an index (vaccination) date on or after 37th week of gestation are not at risk of preterm delivery, and are therefore excluded for this endpoint.
- Preterm pre-labor rupture of membranes (PPROM): Pregnant women with index (vaccination) date prior to 37 weeks gestation. Rationale: Pregnant women with an index (vaccination) date on or after 37th week of gestation are not at risk of PPROM, and are therefore excluded for this endpoint.
- Post-partum hemorrhage: Pregnant women with KPSC delivery. Rationale: Delivery information such as blood loss is not available for women who deliver outside KPSC, and such women are therefore excluded for this endpoint.
- Stillbirth/fetal death: Pregnant women with pregnancy outcome. Rationale: Pregnant women with unknown pregnancy outcome are considered as missing data, and are therefore excluded for this endpoint.
- Neonatal death (within 28 days of birth) and congenital anomalies at birth and through 6 months of age: Infants 1) live born and 2) born in KPSC hospitals. Rationale: Infants born outside KPSC hospitals may not be captured and may not be able to be linked back to their pregnant mothers, and are therefore excluded for this endpoint.

Analysis for preeclampsia and/or eclampsia, intra-uterine infections, poor fetal growth, placental abruption, and maternal death will be conducted on all pregnant women eligible for the main analysis and descriptive analysis.

3.2.2 Exposures

Boostrix vaccination will be received as part of routine clinical care. Vaccine exposure will be identified retrospectively from the electronic health records, including vaccine, manufacturer, and lot number entered at the time of vaccination.

3.2.3 Primary and secondary endpoints

Primary endpoints (Table 1) were selected based on their medical relevance, also considering the background incidence or prevalence of these endpoints.

Secondary endpoints (Table 1) were differentiated from primary endpoints to allow evaluation of the safety of maternal vaccination with sufficient statistical power on the primary endpoints, and to allow for hypothesis generation/signal detection on secondary endpoints.

Table 1. List of primary and secondary endpoints

Primary endpoints	Preeclampsia and/or eclampsia
	Intra-uterine infections such as chorioamnionitis and endometritis
	Small for gestational age (SGA)
	Preterm delivery
Secondary endpoints	Post-partum hemorrhage
	Poor fetal growth

	Stillbirth/fetal death with or without congenital anomalies
	Neonatal death (within 28 days of birth)
	Maternal death (while pregnant or within 42 days of end of pregnancy)
	Placental abruption
	Preterm pre-labor rupture of membranes (PPROM)
	Congenital anomalies at birth and through 6 months of age
	Congenital anomalies of nervous system
	Congenital anomalies of eye
	Congenital anomalies of ear, face, or neck
	Congenital anomalies of cardiovascular system
	Congenital anomalies of respiratory system
	Clefts
	Congenital anomalies of upper gastrointestinal system
	Congenital anomalies of lower gastrointestinal system
	Congenital anomalies of genital organs
	Congenital anomalies of renal system
	Congenital anomalies of musculoskeletal system
	Congenital anomalies of limb
	Congenital anomalies of integument
	Other and unspecified congenital anomalies
	Spontaneous abortion with or without congenital anomalies (only for women exposed to Boostrix before the 27th week of gestation)
	Therapeutic abortion with or without congenital anomalies (only for women exposed to Boostrix before the 27th week of gestation)

The algorithms for identification of primary and secondary endpoints are described in the DMP. Except SGA and preterm delivery, AEs identified through the algorithms will be chart reviewed to confirm the diagnosis and the onset (when applicable). Analyses of endpoints will be based on confirmed cases. For AEs with a chart-reviewed onset date, the event start date will be the onset date; for AEs without a chart-reviewed onset date (unknown or inapplicable), the event start date will be the earliest confirmed diagnosis date. To be a confirmed case, the event start date needs to be after the index date. Some AEs might be diagnosed during the delivery hospitalization, and have an event start date after the delivery date. For pre-eclampsia and/or eclampsia, intra-uterine infections, and poor fetal growth, the event date needs to be on or before the delivery date. For placental abruption and PPROM, if the event start date is after the delivery date, the delivery date will be used as the event start date.

SGA will be defined based on reference values derived by Talge et al. with a cut-off of <10th percentile (1). Talge at al used 2009-2010 US natality data and corrected for likely errors in gestational age dating to yield an up-to-date birth weight for gestational age reference. The

evaluation of SGA is limited to singleton births because the reference values for SGA definition are established based on singleton births.

Stillbirth/fetal death has a number of different and legally mandated definitions (2). This study will use ACOG's definition which defines stillbirth/fetal deaths at 20 completed weeks gestation or greater (if the gestational age is known). Maternal death will include all non-accidental deaths while pregnant or within 42 days of end of pregnancy. Neonatal death will include deaths within 28 days of birth among KPSC live born infants.

For a multiple gestation pregnancy, the pregnancy outcome will be recorded for each fetus, while the endpoint will be defined by the worst scenario. For examples, the woman will be considered to have a pre-term delivery, if any fetus was delivered pre-term after the index date. The woman will also be considered to have spontaneous abortion, therapeutic abortion, or stillbirth/fetal death if any fetus experienced these pregnancy outcomes after the index date.

For a multiple gestation pregnancy with multiple live births, only the infant(s) with infant AEs (neonatal death and congenital anomalies) will be chart reviewed. The endpoint will be defined at the infant level.

Additional birth outcomes will be captured among KPSC born infants for descriptive purposes, but will not be considered as AEs: birth length, head circumference, and Apgar scores.

3.2.4 Covariates

Covariates (Table 2) to be collected and explored include medical conditions indicative of a high-risk pregnancy and other relevant baseline characteristics. The algorithms for identification of high-risk medical conditions are described in the DMP.

Table 2. List of covariates (all covariates will be treated as categorical in the analyses)

High-risk medical condition
Thyroiditis (present/absent)
Diabetes (present/absent)
Other autoimmune diseases (present/absent)
Renal diseases(present/absent)
Cardiac diseases (present/absent)
Pre-pregnancy obesity (BMI>40: yes/no/missing)
Pre-existing hypertension (present/absent)
Rhesus sensitization (present/absent)
HIV infection (present/absent)
Syphilis infection (present/absent)
Cervical incompetence (present/absent)
Lifestyle factors
Smoking status from pregnancy start to index date (Yes/No)
Alcohol use from pregnancy start to index date (Yes/No)

Enrollment
Length of membership prior to 27 weeks gestation (but after pregnancy start) – having continuous membership from pregnancy start to index date (Yes/No)
Medicaid insurance (public health insurance: Indicator of low-income status) – any Medicaid enrollment from pregnancy start to index date (Yes/No)
Utilization
Health care utilization (receipt of medical care) from pregnancy start to index date Any hospitalization or ED visit (Yes/No) Number of outpatient visits (0-4/5-12/≥13)
Maternal Data
Number of previous pregnancies (0/1/≥2/Missing)
History of pregnancy loss (Yes/No)
Season of pregnancy start (September-March/April-August)
Vaccine
Receipt of other vaccines during pregnancy, e.g., influenza vaccine (Yes/No)
Receipt of vaccines containing diphtheria, tetanus toxoid or pertussis antigens within one year before pregnancy (Yes/No)
Matching variables
Race/Ethnicity (White/Black/Hispanic/Asian/Other)
Multiple gestation (Yes/No)
Age at pregnancy start (years)

Smoking status will be defined as current smoker or not. If patients reported they smoked any time from pregnancy start to index date, then they will be current smokers. Alcohol use will be defined in a similar fashion.

Length of membership prior to 27 weeks gestation (but after pregnancy start) will be defined as a dichotomous variable of having continuous membership (allowing up to a 31-day gap) from pregnancy start to index date.

Health care utilization will be defined as any hospitalization or emergency department (ED) visit, and number of outpatient visits (0-4, 5-12, ≥13) from pregnancy start to index date.

Receipt of other vaccines during pregnancy will be defined as any receipt of other vaccines from pregnancy start to index date.

The covariates included in adjusted analyses will be determined by scientific relevance, association with exposure and outcome, and data availability. Specifically, we will select covariates by following these steps:

1. The distribution of covariates will be reviewed. Those covariates in the same category with rare events will likely be combined as a single covariate. For example, we might consider grouping all chronic conditions into a single covariate.
2. The data availability will be checked. If a covariate contains a large proportion that is missing, and the proportion missing is differential by exposure, then the covariate may not be included.
3. Association of covariates with exposure will be assessed. We will use standardized difference to assess the balance of covariates between exposed and unexposed cohorts. Unlike p-values, for which magnitude is highly related to sample size, standardized difference is a unified approach to quantifying the magnitude of difference between groups regardless of sample size, where an absolute value less than 0.1 is considered a negligible difference (3). We will use a SAS macro %stddiff developed by PPD and PPD (4) to calculate the standardized differences for continuous and categorical variables. Potential confounders will be determined by absolute standardized difference (ASD) >0.1.
4. Association of potential confounders (identified from step 3) with each primary endpoint will be further assessed using a change-in-estimate approach (5). Diabetes, pre-existing hypertension (except for the analysis of the preeclampsia and/or eclampsia endpoint), Medicaid insurance, and health care utilization, which are considered as the most representative prognosis factors, will be kept in the adjusted model regardless of their association with exposure and primary endpoints. For a specific endpoint, the exposure coefficient estimate from a model adjusted for the 4 above covariates will be compared with the estimate from a model adjusted for one more covariate at a time. Covariates which result in a 10% change in the exposure coefficient estimate will be selected for final adjusted analysis.
5. All potential confounders (identified from step 3) will be included in the analyses of secondary endpoints. If the outcome is rare and the adjusted model cannot converge, problematic covariates will be removed from the model. If total counts from the two groups (i.e., sum of events in the exposed and unexposed cohorts) is less than 5, then no adjusted analysis will be performed.

3.3 Sample Size and Power Considerations

The power computation for the primary objective was performed using PASS non- inferiority of two independent proportions using Miettinen and Nurminen method. Although the power is not based on the Poisson regression model, it is expected to reflect the power on the planned analysis since inference on proportion and incidence rate provide similar results when incidence rates are low.

The following table (Table 3) shows the power to rule out a two-fold increase considering an assumed possible design bias leading to a 1.25 relative risk increase.

To be able to make independent conclusions for each of the four primary endpoints while controlling the type I error below 2.5%, a Bonferroni adjustment will be used, i.e., each primary endpoint will be evaluated using a one-sided 0.625% (2.5% type I error / 4 primary endpoints) nominal type I error. More specifically, the upper limit of the two-sided 98.75% confidence interval (CI) will be used to make conclusions.

Considering that the four primary endpoints are expected to have an incidence rate above 1%, the power to conclude that maternal immunization is non-inferior for any primary endpoint is at least 95% and the power to conclude that maternal immunization is non-inferior for all 4 primary endpoints is at least 80% (100% - 4*beta).

Table 3. Power to rule out a two-fold increase with 15,000 pregnant women in each cohort

Incidence rate in unvaccinated cohort	One-sided type I nominal error	True relative risk	Power	Beta
1%	0.625%	1.25	95.7%	4.3%

4. ANALYSES OF OBJECTIVE

All analyses will be performed by statisticians at KPSC using the SAS statistical software package (version 9.4 or later).

4.1 Description of study population

The size of the study population and subgroups for main analysis will be described, overall and by exposed and unexposed cohorts. See Appendix A, Table Shell 1 for details. The number and percentage of pregnant women with unknown pregnancy outcome, lost to follow up, and multiple gestation will also be reported.

Baseline characteristics will be described and compared between the exposed and unexposed cohorts. See Appendix A, Table Shell 2 for details. Bivariate analyses will be conducted for all covariates of interest. Continuous variables such as age in years will be summarized by mean, median, Q1, Q3, range, and standard deviation and compared using t-tests; categorical variables will be summarized by frequency and percentage and compared using chi-square tests or Fisher's exact test, as appropriate. Absolute standardized difference will be calculated to assess the balance of covariates. Potential confounders will be determined based on bivariate analyses and scientific relevance (see Section 3.2.4).

Birth outcomes captured among KPSC born infants (birth weight, length, head circumference, and Apgar scores) will be described for exposed and unexposed cohorts. See Appendix A, Table Shell 3 for details.

4.2 Analyses of primary endpoints

We will calculate crude incidence and 95% CI for each primary endpoint for the exposed and unexposed cohorts. Unadjusted and adjusted RR with two-sided 98.75% CI will be estimated comparing the exposed cohort to the unexposed cohort, accounting for multiple comparisons.

Analysis of primary endpoints will be used to evaluate the safety of Boostrix vaccination and concluded to be non-inferior if the upper limit of the 98.75% CI for the adjusted relative risk is below 2.

Analyses of preeclampsia and/or eclampsia and intra-uterine infection will be performed on all pregnant women eligible for main analysis. For events identified any time after the index date (e.g., preeclampsia and/or eclampsia and intra-uterine infection), the incidence will consist of the total number of women with the condition in the numerator and the total person time in the denominator. The person-year for a pregnant woman will be the time from the index date to the date of each adverse event, end of the pregnancy, or disenrollment, whichever comes first. The unadjusted and adjusted relative risk with 98.75% CI will be estimated by Poisson regression models accounting for the follow-up time without and with adjustment for potential confounders. If the outcome is common (i.e., incidence>2%), a Poisson regression model with a robust error variance will be used to estimate the relative risk (6). The robust Poisson regression will be performed using the SAS Proc GENMOD procedure with the REPEATED statement subject= Patient_ID.

Analysis of SGA will be performed on live singleton infants born in KPSC hospitals, while analysis of preterm delivery will be performed on pregnant women with live births, and with an index date prior to 37 weeks gestation. For events identified at delivery (e.g., SGA and preterm delivery), the incidence will consist of the total number of women or infants with the condition in the numerator and the number of women or infants for whom the condition can be assessed in the denominator. The unadjusted and adjusted relative risk with 98.75% CI will be estimated by Poisson regression models with robust error variances without and with adjustment for potential confounders. The planned methods of analysis by primary endpoint are outlined in Table 4. See Appendix A, Table Shell 4 for details.

Table 4. Analytic population and statistical methods by primary endpoint

Primary endpoints	Analytic population	Statistical Methods
Preeclampsia and/or eclampsia	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time, with robust variance estimate
Intra-uterine infections	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time, with robust variance estimate
SGA	KPSC live born infant, singleton only*	Adjusted Poisson regression model, with robust variance estimate
Preterm delivery	Pregnant women with live birth, with index date prior to 37 weeks gestation	Adjusted Poisson regression model, with robust variance estimate

*There are no well accepted SGA metrics for infants born to pregnant women with multiple gestation

Limiting analyses of endpoints to subgroups (e.g., KPSC live born infant) may cause matches to be broken, although we expect the proportions of non-live births and non-KPSC births to be small.

The matching variables will be included in the adjusted models of analyses of SGA and preterm delivery. Age at pregnancy start will be adjusted as a categorical variable (<24/25-34/≥35 years).

4.3 Analyses of secondary endpoints

The analysis of secondary endpoints will be used for hypothesis generation/signal detection. We will calculate crude incidence and 95% CI for each secondary endpoint for the exposed and unexposed cohorts. Unadjusted and adjusted RR with two-sided 95% CI will be estimated comparing the exposed cohort to the unexposed cohort. At least one case in each cohort is required to perform unadjusted and adjusted analyses. If total counts from the two groups is less than 5, then no adjusted analysis will be performed. An elevated risk will be detected if the lower limit of the 95% CI for the adjusted relative risk is above 1. The planned methods of analysis by secondary endpoint are outlined in Table 5. See Appendix A, Table Shell 4 for details.

Table 5. Analytic population and statistical methods by secondary endpoint

Secondary endpoints	Analytic population	Statistical Methods
Post-partum hemorrhage	Pregnant women with KPSC delivery	Adjusted Poisson regression model, with robust variance estimate
Poor fetal growth	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time, with robust variance estimate
Stillbirth/fetal death*	Pregnant women with pregnancy outcome	Adjusted Poisson regression model
Neonatal death	KPSC live born infants	Adjusted Poisson regression model
Maternal death	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time
Placental abruption	All pregnant women eligible for main analysis	Adjusted Poisson regression model accounting for follow up time
PPROM	All pregnant women, with index (vaccination) date prior to 37 weeks gestation	Adjusted Poisson regression model accounting for follow up time
Congenital anomalies (CA) by type	KPSC live born infants	Adjusted Poisson regression model, with robust variance estimate if outcome is common

*Stillbirth/fetal death will be described with or without congenital anomalies separately, but will be analyzed as one outcome.

For poor fetal growth, placental abruption, PPROM, and maternal death, the incidence will consist of the total number of women with the condition in the numerator and the total person time in the denominator. For poor fetal growth, placental abruption, and PPROM, the person-time for a pregnant woman will be the time from the index date to the date of each adverse event, end of the pregnancy, or disenrollment, whichever comes first. For maternal death, the person-time for a pregnant woman will be the time from the index date to the date of death, 42 days after delivery, or disenrollment, whichever comes first. The unadjusted and adjusted relative risk with 95% CI will be estimated by Poisson regression models, or Poisson regression models with robust error variances if incidence>2%, accounting for the follow-up time without and with adjustment for potential confounders.

For stillbirth/fetal death and post-partum hemorrhage, the incidence will consist of the total number of women with the condition in the numerator and the number of women for whom the condition can be assessed in the denominator. The unadjusted and adjusted relative risk with 95% CI will be estimated by Poisson regression models, or Poisson regression models with robust error variances if incidence>2%, without and with adjustment for potential confounders.

Neonatal death and congenital anomalies will be identified among infants born in KPSC hospitals. The incidence of neonatal death will be calculated as the number of neonatal deaths in the numerator and the total number of infants born in KPSC hospitals in the denominator. The prevalence of congenital anomalies (overall, by type) will be calculated as the number of infants with a congenital anomaly in the numerator and the total number of infants born in KPSC hospitals in the denominator. The proportion of live births without congenital anomalies will be calculated as 1 minus the proportion of live births with congenital anomalies among all KPSC live born infants. The unadjusted and adjusted relative risk with 95% CI will be estimated by Poisson regression models, or Poisson regression models with robust error variances if incidence>2%, without and with adjustment for potential confounders.

4.4 Analysis of pregnant women vaccinated before the 27th week of gestation

The analysis of pregnant women vaccinated with Boostrix before the 27th week of gestation will be descriptive. Due to the expected extremely small sample size, no formal hypothesis testing is planned.

The size of the study population and subgroups will be described. See Appendix A, Table Shell 6 for details. Baseline characteristics of pregnant women and their KPSC live born infants will be described. Continuous variables will be summarized by mean, median, Q1, Q3, range, and standard deviation; categorical variables will be summarized by frequency and percentage. See Appendix A, Table Shell 7 and Table Shell 8 for details.

The number of events for each primary and secondary endpoint in Table 1 among pregnant women vaccinated with Boostrix before the 27th week of gestation will be reported. The incidence will be calculated as the total number of women with the condition in the numerator and the number of women (or total person time) for whom the condition can be assessed in the denominator. Since we will likely observe no events or few events for most of the endpoints, the 95% Wilson score interval with continuity correction will be provided (7). See Appendix A, Table Shell 9 for details.

4.5 Signal investigation

If a primary endpoint analysis cannot rule out a 2-fold relative risk, or secondary endpoint analysis detects an elevated risk, further exploration of the data will be completed to investigate alternative potential sources for the Boostrix-outcome association.

Description of events

Events will be summarized by age at pregnancy start, gestational age at event date, receipt of other vaccines during pregnancy, and other risk factors associated with the signaled endpoint. Stratified incidence will be calculated as the number of events in a specific stratum in the numerator and the total number of persons or person-time of the stratum in the denominator for the exposed and unexposed cohorts. If the endpoint is chart reviewed, chart confirmation rates for the exposed and unexposed cohorts will be provided. See Appendix A, Table Shell 10 for details.

In addition, depending on the AE that signals, secular trends may be assessed. The KPSC team will work together with GSK team to conduct signal investigation, with consideration of timelines and data/resource availability.

5. STATISTICAL CONSIDERATIONS AND LIMITATIONS

5.1 Handling of missing data

Medical record review will be conducted on unknown pregnancy outcomes and KPSC live born singleton infants with unknown birthweight. We expect that there will be a very small amount of missing data after medical record review. Data with unknown pregnancy outcome will be excluded from the analysis of stillbirth; data with birthweight missing will be excluded from the analysis of SGA. The data can be kept for analyses of other endpoints if they have pregnancy end dates.

Medical record review will be performed on pregnant women who have continuous membership, but who are missing pregnancy end dates. If the pregnancy end date remains missing after medical record review, the last contact date with KPSC before the estimated date of delivery (EDD) will be used as the end date of follow up. The patient will be considered lost to follow up.

No imputation is planned for missing data in covariates. Unknown race/ethnicity is grouped into the "Other" category by convention, which can be matched between exposed and unexposed cohorts. Missing number of previous pregnancies and missing pre-pregnancy weight will be assessed. If they are included in the adjusted model, "Missing" will be a category for the covariates along with the other categories. Missing birth length, head circumference, and Apgar scores will be described as is.

5.2 Multiplicity considerations

Analyses of the primary endpoints will be conducted using a non-inferiority design with multiple comparison adjustment. Analyses of the secondary endpoints will be conducted using superiority testing without adjustment for multiple comparisons. The approach leads to a more conservative

conclusion regarding safety of vaccination, i.e., for a non-inferiority test, it is harder to reject Ho: RR≥2, and for a superiority test, it is easier to reject Ho: RR=1.

The analyses of secondary endpoints will be used for hypothesis generation/signal detection. No formal adjustment for multiple comparisons is common practice for this type of study. Since outcomes monitored in vaccine safety studies are usually serious, researchers do not want to miss the opportunity to detect potential safety findings.

5.3 Margin for non-inferiority testing

The non-inferiority design allows us to account for possible bias from an observational study design which may not be fully adjusted for in the analysis, and conclude that the risk is less than a threshold, say RR<2. The non-inferiority margin was determined based on prior experience from post-marketing vaccine safety studies, the magnitude of possible bias, and sample size considerations. While a non-inferiority margin as low as 1.8 or 1.5 might be preferable, it would require a much larger sample size.

5.4 Limitations

Secular confounding

Using a historical comparison group may introduce bias due to secular confounding. The events identified from a current exposed cohort and a historical unexposed cohort may differ in background rates due to care improvement, changes in diagnostic criteria, changes in health-seeking behavior, and changes in the coding system. This is an inherent limitation of using a historical comparison and needs to be acknowledged. The approach we propose, using a comprehensive list of codes to identify all potential events of interest, followed by a review of medical records to confirm the incident events, aims to minimize the bias caused by changing coding practices.

Confounding by indication

Compared to a concurrent unexposed cohort (assuming approximately 80% Tdap uptake among pregnant women), a historical unexposed cohort would be less selective and more representative of an unvaccinated cohort. However, confounding by indication is still possible. A number of covariates including high-risk medical conditions, health care utilization, pregnancy history, etc. will be collected and possibly adjusted to minimize bias.

6. CDISC

The analytic datasets will be formatted according to CDISC standards. The KPSC team will work with a CDISC standards consultant to develop Analysis Data Model (ADaM) specifications and create ADaM datasets in accordance with the Study Data Tabulation Module (SDTM) datasets, study report, and CDISC implementation guide; validate the ADaM datasets using an open source CDISC validation program (e.g., Pinnacle 21 or similar); and produce submission documentation including DEFINE.XML and Analysis Data Reviewer Guide (ADRG).

The development of ADaM specifications and ADaM datasets will parallel statistical analysis activities. Specifically, draft ADaM specifications and datasets will be developed according to the SAP before conducting analysis. The ADaM specifications and datasets will be modified if any analyses deviate from the pre-specified SAP which can cause a change in the analytic dataset. The ADaM specifications and datasets will be finalized once the results tables are finalized. Consistency of final ADaM datasets produced against the specifications will be evaluated, and compliance checks using a validation tool (e.g. Pinnacle 21 or similar) including validation of standard compliance, controlled terminology, and data quality will be conducted on the dataset to ensure compliance with CDISC standards for ADaM. Consistency of the results tables produced against the final ADaM datasets will be evaluated too. Additionally, KPSC will ensure and oversee that the CDISC consultancy has adequate procedures and controls in place to guarantee that the resulting CDISC package (i.e., datasets and supporting documentation) is of adequate quality.

7. QUALITY CONTROL

KPSC will follow the Department of Research and Evaluation's Analytical Research Project Standard Operating Procedures (SOP), including data extraction, data analysis, communication, and documentation. SAS programs used for the analysis will be provided to GSK for review. Statistical analysis (results tables) will be discussed in the multisite meetings with GSK prior to study report finalization.

Quality control steps

During the analysis, when analytic data are created and analyzed, the following quality control steps will be performed by the study statisticians as appropriate:

1. Read study protocol, DMP, and the statistical analysis plan. Communicate with the principal investigator, senior statistician, and study team with any questions regarding study design and analysis plan. Document any decisions related to statistical analysis.

2. Check data quality (sample size, duplicates, range of values, number of missings, proper data type and format, inconsistency, etc.). Any aberrant data that have been missed by data cleaning and discovered by the statistician will be described in the results table.

3. Participate in 2 days of CDISC training on SDTM, ADaM, and Define.xml. Participate in 1-day CDISC study-specific workshop. Become familiar with SDTM and ADaM data standards. Work with CDISC standards consultant to develop ADaM specifications and create ADaM datasets according to SAP.

4. Develop SAS programs for analysis. Programs will include a program header and section headers. The program header contains information on the name, location, and purpose of the program, a list of prerequisite programs if applicable, a list of input datasets with full path, a list of output datasets, and a history log. Revision of the program will be noted in the history log to capture the change made, the change date, identification of the individual making the change, and type of change (e.g. modification, addition, removal). Each program might have several sections

executing various analytic steps. Programs will be annotated with comments that describe the intent or purpose.

5. Perform statistical analysis. Use appropriate procedures and tests. SAS logs and outputs will be reviewed for any warnings, error messages, and model convergence issues. Organize the output into results tables. When possible, automate the generation of results tables to minimize the need to copy/paste. Proofread the results tables to prevent any copy/paste and editing errors. SAS logs and outputs will be saved.

6. Use established SAS macro. Use established SAS macros for analyses, including but not limited to the macro for creating baseline covariate tables and the macro for calculating standardized difference.

Program review

All SAS analysis programs, logs and outputs will be reviewed by a second statistician. The second statistician will review the program logic against the analysis requirements. Comments from the second statistician will be documented and followed up until resolved.

Quality control findings

A record of quality problems and resolutions will be kept in the form of an Excel spreadsheet or other method appropriate to the circumstances (e.g. commented code, detailed correspondence).

Analysis deviation

Any analyses differing from the protocol, DMP and pre-specified analyses in this SAP will be documented. The rationale for the change in approach will be provided.

8. STUDY REPORTS

The final report will include all data generated from this study as specified in this SAP.

9. LIST OF REPORT TABLES

	Description	Population
Table 1	Study population and subgroups for main analysis	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 2	Baseline characteristics of Boostrix exposed and unexposed pregnant women	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 3	Characteristics of KPSC live born infants with Boostrix exposed and unexposed mothers	KPSC live born infants with exposed and unexposed mother

		with index (vaccination) date after the 27th week of gestation
Table 4*	Incidence and relative risk of primary endpoints comparing Boostrix exposed and unexposed cohorts	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 5*	Incidence and relative risk of secondary endpoints comparing Boostrix exposed and unexposed cohorts	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation
Table 6	Study population and subgroups for descriptive analysis	Pregnant women vaccinated with Boostrix before the 27th week of gestation
Table 7	Baseline characteristics of pregnant women vaccinated with Boostrix before the 27th week of gestation	Pregnant women vaccinated with Boostrix before the 27th week of gestation
Table 8	Characteristics of KPSC live born infants with mothers vaccinated with Boostrix before the 27th week of gestation	KPSC live born infants with mothers vaccinated with Boostrix before the 27th week of gestation
Table 9	Incidence of primary and secondary endpoints among pregnant women vaccinated with Boostrix before the 27th week of gestation	Pregnant women vaccinated with Boostrix before the 27th week of gestation
Table 10	Summary of signaled events	Exposed and unexposed pregnant women with index (vaccination) date after the 27th week of gestation

* The details of the model for each endpoint will be shared with GSK, but will not be included in the report tables.

10. REFERENCES

1. Talge NM, Mudd LM, Sikorskii A, Basso O. United States birth weight reference corrected for implausible gestational age estimates. Pediatrics. 2014;133(5):844-853.

2. Tavares Da Silva F, Gonik B, McMillan M, et al. Stillbirth: Case definition and guidelines for data collection, analysis, and presentation of maternal immunization safety data. Vaccine. 2016;34(49):6057–6068.

3. Austin PC**.** An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.

4. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum, Statistics and Data Analysis. 2012;Paper 335-2012.

5. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340-349.

6. Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004;159(7):702-6.

7. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine. 1998 17 (8): 857–872.

11. APPENDIX A: TABLE SHELLS

Appendix A: Table shells

Table 1: Study population and subgroups for main analysis

	Exposed	Unexposed	Total
Pregnant women eligible per inclusion/exclusion criteria, n			
With pregnancy outcome, n (%)			
With unknown pregnancy outcome, n (%)			
Lost to follow up, n (%)			
Pregnant women with multiple gestation, n (%)			
Pregnant women with KPSC delivery, n (%)			
Pregnant women with index date prior to 37 weeks gestation, n (%)			
With live birth, n (%)			
KPSC live born infants, n			
Singleton only, n (%)			

Table 2. Baseline characteristics[a] of Boostrix exposed and unexposed pregnant women

		Exposed N =	Unexposed N =	Total N =	p value
	Maternal age at pregnancy st				N/A[d]
mean (sd)					
median					
Q1, Q3					
range					
	Race/ethnicity, n (%)				N/A
White					
Black					
Hispanic					
Asian					
Other					
	Gestational age in weeks at i				N/A
27-30					
31-36					
≥37					
Current smoker, n (%)					
Yes					
No					
Alcohol use, n (%)					
Yes					
No					
Continuous membership, n (%)					
Yes					
No					
Medicaid insurance, n (%)					
Yes					
No					
Any hospitalization or ED visit, n (%)					
Yes					
No					
Number of outpatient visits, n (%)					
0-4					
5-12					
≥13					
Number of previous pregnancies, n (%)					
0					
1					
≥2					
Missing					
History of pregnancy loss, n (%)					
Yes					
No					
Season of pregnancy start, n (%)					
September-March					
April-August					
Received other vaccines[b], n (%)					
Yes					
No					
Received vaccines containing diphtheria, tetanus toxoid, or pertussis antigens within one year before pregnancy, n (%)					
Yes					
No					
High-risk medical condition, n (%)					
	Multiple gestation				N/A
Thyroiditis					
Diabetes					
Other autoimmune diseases					
Renal diseases					
Cardiac diseases					
Pre-pregnancy obesity[c]					
Missing					
Pre-existing hypertension					
Rhesus sensitization					
HIV infection					
Syphilis infection					
Cervical incompetence					

[a] Baseline characteristics were defined from pregnancy start to index date, unless otherwise specified

[b] Other vaccines included: inactivated influenza vaccine (xx%), hepatitis A/hepatitis B vaccine (xx%), human papillomavirus vaccine (xx%), and other (xx%).

[c] BMI>40 within 6 months prior to pregnancy start

[d] N/A= not applicable, for matching variable

Table 3. Characteristics of KPSC live born infants with Boostrix exposed and unexposed mothers

	Exposed N =	Unexpose N =
Sex, n (%)		
F		
M		
Gestational age in weeks at birth, n (%)		
27-33		
34-36		
37-39		
≥40		
Birth weight, g		
mean (sd)		
median		
Q1, Q3		
range		
Missing		
Birth length, cm		
mean (sd)		
median		
Q1, Q3		
range		
Missing		
Head circumference, cm		
mean (sd)		
median		
Q1, Q3		
range		
Missing		
Apgar 1 score, n (%)		
≤3		
4-6		
≥7		
Missing		
Apgar 5 score, n (%)		
≤3		
4-6		
≥7		
Missing		

Appendix A: Table shells

Table 4. Incidence and relative risk of primary endpoints comparing Boostrix exposed and unexposed cohorts

	Exposed				Unexposed				Relative Risk[e] (Wald 98.75% CI)	
	Number of Persons	Number of Person years	Number of adverse events	Incidence per 1000 persons or person-years (95% CI)	Number of Persons	Number of Person years	Number of adverse events	Incidence per 1000 persons or person-years (95% CI)	Unadjusted	Adjusted
Pre-eclampsia and eclampsia[a]										
Intra-uterine infections[a]										
Small for gestational age [b]		N/A*				N/A				
Preterm delivery [c]		N/A				N/A				

[a] Identified among all pregnant women eligible for main analysis
[b] Identified among KPSC live born infants, singleton only
[c] Identified among pregnant women with live birth, with index (vaccination) date prior to 37 weeks gestation
[d] Adjusted for covariates XXX (place holder, covariates may differ for each endpoint)
[e] Poisson regression model with robust variance estimate
* N/A=Not Applicable

Table 5. Incidence and relative risk of secondary endpoints comparing Boostrix exposed and unexposed cohorts

	Exposed				Unexposed				Relative Risk (Wald 95% CI)	
	Number of Persons	Number of Person years	Number of events	Incidence per 1000 persons or person-years (95% CI)	Number of Persons	Number of Person years	Number of events	Incidence per 1000 persons or person-years (95% CI)	Unadjusted	Adjusted
Poor fetal growth, IUGR[a]										
Placental abruption[a]										
Preterm pre-labor rupture of membranes (PPROM)[b]										
Maternal death [a]										
Stillbirth/fetal death[c]		N/A*				N/A				
With congenital anomalies		N/A				N/A			N/A	N/A
Post-partum hemorrhage[d]		N/A				N/A				
Neonatal death [e]		N/A				N/A				
Congenital anomalies at birth and through 6 months of age[d]		N/A				N/A			N/A	N/A
Congenital anomalies of nervous system		N/A				N/A				
Congenital anomalies of eye		N/A				N/A				
Congenital anomalies of ear, face, or neck		N/A				N/A				
Congenital anomalies of cardiovascular system		N/A				N/A				
Congenital anomalies of respiratory system		N/A				N/A				
Clefts		N/A				N/A				
Congenital anomalies of upper gastrointestinal system		N/A				N/A				
Congenital anomalies of lower gastrointestinal system		N/A				N/A				
Congenital anomalies of genital organs		N/A				N/A				
Congenital anomalies of renal system		N/A				N/A				
Congenital anomalies of musculoskeletal system		N/A				N/A				
Congenital anomalies of limb		N/A				N/A				
Congenital anomalies of integument		N/A				N/A				
Other and unspecified congenital anomalies		N/A				N/A				
Number of congenital anomalies (based on body system) per infant		N/A		N/A		N/A		N/A	N/A	N/A
0		N/A		N/A		N/A		N/A	N/A	N/A
1		N/A		N/A		N/A		N/A	N/A	N/A
2		N/A		N/A		N/A		N/A	N/A	N/A
3		N/A		N/A		N/A		N/A	N/A	N/A
≥4		N/A		N/A		N/A		N/A	N/A	N/A

[a] Identified among all pregnant women eligible for main analysis
[b] Identified among pregnant women with index (vaccination) date prior to 37 weeks gestation
[c] Identified among pregnant women with pregnancy outcome
[d] Identified among pregnant women with KPSC delivery
[e] Identified among KPSC live born infants
[f] Adjusted for covariates XXX (place holder, covariates may differ for each endpoint)
[g] Poisson regression model with robust variance estimate (place holder)
* N/A=Not Applicable

Table 6: Study population and subgroups for descriptive analysis

	Exposed
Pregnant women eligible per inclusion/exclusion criteria, n	
With pregnancy outcome, n (%)	
With unknown pregnancy outcome, n (%)	
Lost to follow up, n (%)	
Pregnant women with multiple gestation, n (%)	
Pregnant women with KPSC delivery, n (%)	
Pregnant women with index date prior to 37 weeks gestation, n (%)	
With live birth, n (%)	
KPSC live born infants, n	
Singleton only, n (%)	

Table 7. Baseline characteristics[a] of pregnant women vaccinated with Boostrix before the 27th week of gestation

	Exposed N =
Maternal age at pregnancy start, years	
mean (sd)	
median	
Q1, Q3	
range	
Race/ethnicity, n (%)	
White	
Black	
Hispanic	
Asian	
Other	
Gestational age in weeks at index date, n (%)	
27-30	
31-36	
≥37	
Current smoker, n (%)	
Yes	
No	
Alcohol use, n (%)	
Yes	
No	
Continuous membership, n (%)	
Yes	
No	
Medicaid insurance, n (%)	
Yes	
No	
Any hospitalization or ED visit, n (%)	
Yes	
No	
Number of outpatient visits, n (%)	
0-4	
5-12	
≥13	
Number of previous pregnancies, n (%)	
0	
1	
≥2	
Missing	
History of pregnancy loss, n (%)	
Yes	
No	
Season of pregnancy start, n (%)	
September-March	
April-August	
Received other vaccines[b], n (%)	
Yes	
No	
Received vaccines containing diphtheria, tetanus toxoid, or pertussis antigens within one year before pregnancy, n (%)	
Yes	
No	
High-risk medical condition, n (%)	
Multiple gestation	
Thyroiditis	
Diabetes	
Other autoimmune diseases	
Renal diseases	
Cardiac diseases	
Pre-pregnancy obesity[c]	
Missing	
Pre-existing hypertension	
Rhesus sensitization	
HIV infection	
Syphilis infection	
Cervical incompetence	

[a] Baseline characteristics were defined from pregnancy start to index (vaccination) date, unless otherwise specified

[b] Other vaccines included: inactivated influenza vaccine (xx%), hepatitis A/hepatitis B vaccine (xx%), human papillomavirus vaccine (xx%), and other (xx%).

[c] BMI>40 within 6 months prior to pregnancy start

Table 8. Characteristics of KPSC live born infants with mothers vaccinated with Boostrix before the 27th week of gestation

	Exposed N =
Sex, n (%)	
F	
M	
Gestational age in weeks at birth, n (%)	
27-33	
34-36	
37-39	
≥40	
Birth weight, g	
mean (sd)	
median	
Q1, Q3	
range	
Missing	
Birth length, cm	
mean (sd)	
median	
Q1, Q3	
range	
Missing	
Head circumference, cm	
mean (sd)	
median	
Q1, Q3	
range	
Missing	
Apgar 1 score, n (%)	
≤3	
4-6	
≥7	
Missing	
Apgar 5 score, n (%)	
≤3	
4-6	
≥7	
Missing	

Table 9. Incidence of primary and secondary endpoints among pregnant women vaccinated with Boostrix before the 27th week of gestation

	Exposed			
	Number of Persons	Number of Person years	Number of events	Incidence per 1000 persons or person-years (95% CI)
Primary endpoints				
Pre-eclampsia and eclampsia[a]				
Intra-uterine infections[a]				
Small for gestational age [b]		N/A*		
Preterm delivery [c]		N/A		
Secondary endpoints				
Poor fetal growth, IUGR[a]				
Placental abruption[a]				
Preterm pre-labor rupture of membranes (PPROM)[d]				
Maternal death [a]				
Stillbirth/fetal death[e]		N/A		
With congenital anomalies		N/A		
Spontaneous abortion [e]		N/A		
With congenital anomalies		N/A		
Therapeutic abortion[e]		N/A		
With congenital anomalies		N/A		
Post-partum hemorrhage[f]		N/A		
Neonatal death [g]		N/A		
Congenital anomalies at birth and through 6 months of age[f]		N/A		
Congenital anomalies of nervous system		N/A		
Congenital anomalies of eye		N/A		
Congenital anomalies of ear, face, or neck		N/A		
Congenital anomalies of cardiovascular system		N/A		
Congenital anomalies of respiratory system		N/A		
Clefts		N/A		
Congenital anomalies of upper gastrointestinal system		N/A		
Congenital anomalies of lower gastrointestinal system		N/A		
Congenital anomalies of genital organs		N/A		
Congenital anomalies of renal system		N/A		
Congenital anomalies of musculoskeletal system		N/A		
Congenital anomalies of limb		N/A		
Congenital anomalies of integument		N/A		
Other and unspecified congenital anomalies		N/A		
Number of congenital anomalies (based on body system) per infant		N/A		N/A
0		N/A		N/A
1		N/A		N/A
2		N/A		N/A
3		N/A		N/A
≥4		N/A		N/A

[a] Identified among all pregnant women eligible for descriptive analysis

[b] Identified among KPSC live born infants, singleton only

[c] Identified among pregnant women with live birth, with index (vaccination) date prior to 37 weeks gestation

[d] Identified among pregnant women with index (vaccination) date prior to 37 weeks gestation

[e] Identified among pregnant women with pregnancy outcome

[f] Identified among pregnant women with KPSC delivery

[g] Identified among KPSC live born infants

[h] Wilson score interval with continuity correction

* N/A=Not Applicable

Table 10. Summary of signaled events

	Exposed			Unexposed		
	Number of events	Number of Persons (or person years)	Incidence per 1000 persons or person-years (95% CI)	Number of events	Number of Persons (or person years)	Incidence per 1000 persons or person-years (95% CI)
Number of confirmed events		N/A	N/A		N/A	N/A
Number of potential cases identified by the algorithm		N/A	N/A		N/A	N/A
Confirmation rate		N/A	N/A		N/A	N/A
Age at pregnancy start, years						
<18						
18-24						
25-34						
≥35						
Received other vaccines during pregnancy, n (%)						
Yes						
No						
High-risk medical condition, n (%)						
Thyroiditis						
Yes						
No						
Diabetes						
Yes						
No						
…						
Gestational age at event date						
27th week		N/A	N/A		N/A	N/A
28th week		N/A	N/A		N/A	N/A
29th week		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
ICD diagnosis code and description						
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A
…		N/A	N/A		N/A	N/A