NOTE TO THE EDITOR

THIS IS THE PROTOCOL AMENDMENT OF EPI-HPV-018 VS UK DB (ETRACK: 114101)WHICH HAS BEEN PROVIDED BY AUTHORS AS A PART OF POSTING FULL PROTOCOL TO THE EUROPEAN NETWORK OF CENTRES FOR PHARMACOEPIDEMIOLOGY AND PHARMACOVIGILANCE WEBSITE. SOME PARTS WITH PERSONALLY IDENTIFIABLE INFORMATION ARE RETRACTED FROM THE PROTOCOL AND THE PAGES WHERE RETRACTION HAS BEEN DONE IS MENTIONED IN THE NEXT PAGE DATE: JUNE 14, 2013 – HB

Study Protocol
Sponsor:
GlaxoSmithKline Biologicals
Rue de l'Institut 89
1330 Rixensart, Belgium

eTrack study number and Abbreviated Title	114101 (EPI-HPV-018 VS UK DB)
Date of protocol	Final: 20 September 2012
Date of protocol amendment	Amendment 1 Final: 14 March 2013
Title	Post-marketing safety study to assess the risk of spontaneous abortions in women exposed to *Cervarix* in the United Kingdom
Detailed Title	An observational cohort study assessing the risk of spontaneous abortions during the first 23 weeks of gestation in women aged 15 to 25 years exposed to *Cervarix*, residing in the United Kingdom and reporting their last menstrual period between 30 days before and 45 days after any dose of *Cervarix*.
Co-ordinating authors	Co-ordinating authors' names retracted to protect Subject Privacy
Main contributing authors	Contributing authors' names retracted to protect Subject Privacy

eTrack study number and Abbreviated Title	114101 (EPI-HPV-018 VS UK DB)
Date of protocol	Final: 20 September 2012
Date of protocol amendment	Amendment 1 Final: 14 March 2013
Title	Post-marketing safety study to assess the risk of spontaneous abortions in women exposed to *Cervarix* in the United Kingdom
Detailed Title	An observational cohort study assessing the risk of spontaneous abortions during the first 23 weeks of gestation in women aged 15 to 25 years exposed to *Cervarix*, residing in the United Kingdom and reporting their last menstrual period between 30 days before and 45 days after any dose of *Cervarix*
Other contributing authors (GSK Biologicals) (Amended 14 March 2013)	Contributing authors' names retracted to protect Subject Privacy

eTrack study number and Abbreviated Title	114101 (EPI-HPV-018 VS UK DB)
Date of protocol	Final: 20 September 2012
Date of protocol amendment	Amendment 1 Final: 14 March 2013
Title	Post-marketing safety study to assess the risk of spontaneous abortions in women exposed to *Cervarix* in the United Kingdom
Detailed Title	An observational cohort study assessing the risk of spontaneous abortions during the first 23 weeks of gestation in women aged 15 to 25 years exposed to *Cervarix*, residing in the United Kingdom and reporting their last menstrual period between 30 days before and 45 days after any dose of *Cervarix*
Non GSK contributing authors (external experts)	Contributing authors' names retracted to protect Subject Privacy
CPRD *GOLD* contributing authors	Contributing authors' names retracted to protect Subject Privacy

(Amended 14 March 2013)

GSK Biologicals' protocol template for observational studies and interventional studies without administration of medicinal products as described in a research protocol based on the Protocol Document Standard version 13.2

NOTE TO THE EDITOR:

The following minor corrections (highlighted in the text) were made to the last version (14 March 2013) of this protocol amendment: Co-ordinating authors' names are retracted on pages 2 and 54. Contributing authors' names retracted on pages 2, 3, 4 and 55. Sponsor signatory name retracted on page 5.

Protocol Amendment 1 Sponsor Signatory Approval

eTrack study number and Abbreviated Title	114101 (EPI-HPV-018 VS UK DB)
Date of protocol amendment	Amendment 1 Final: 14 March 2013
Detailed Title	An observational cohort study assessing the risk of spontaneous abortions during the first 23 weeks of gestation in women aged 15 to 25 years exposed to *Cervarix*, residing in the United Kingdom and reporting their last menstrual period between 30 days before and 45 days after any dose of *Cervarix*.
Sponsor signatory	Sponsor signatory's name retracted to protect Subject Privacy
Signature	_____
Date	_____

Protocol Amendment 1 Rationale

Amendment number: Amendment 1

Rationale/background for changes:

This protocol has been amended for the following reasons:

- In section 5.2.4.2. Pregnancy outcomes and section 5.2.6. Final case ascertainment step, review of pregnancy outcomes by the GSK safety physician and by experts has been updated.
- In section 5.2.5.1. First day of LMP, the length of pregnancy has been modified to fit the United Kingdom (UK) definition (280 days) and not the French definition (283 days).
- Clinical Practice Research Datalink (CPRD) has been replaced by Clinical Practice Research Datalink General Practitioner OnLine Database (CPRD GOLD).
- Section 1.2. Rationale for the study has been updated with recent published data about *Cervarix* uptake in the United States (US) and about *Cervarix* safety in the UK.
- In section 3. Study design, the number of subjects in the exposed cohort in the epoch table has been corrected.
- Section 3.1. HPV vaccine coverage and pregnancy rate in the UK teenagers has been updated with recent published data about HPV vaccine coverage in the UK.
- In section 5.6.3.1. Cox proportional hazards model, enrolment from CPRD has been replaced by de-enrolment from CPRD GOLD.
- Contributing authors have been updated.

Other changes were made for simplification, clarification or consistency.

SYNOPSIS

(Amended 14 March 2013)

Detailed Title	An observational cohort study assessing the risk of spontaneous abortions during the first 23 weeks of gestation in women aged 15 to 25 years exposed to ***Cervarix***, residing in the United Kingdom and reporting their last menstrual period between 30 days before and 45 days after any dose of ***Cervarix***.
Rationale for the study	***Cervarix*** is not recommended for use during pregnancy due to insufficient pre-licensure data in pregnant women. Because ***Cervarix*** is indicated in women of child-bearing age, inadvertent exposure to ***Cervarix*** is possible during the earlier stages of pregnancy (i.e. first trimester). Based upon the results of a subgroup analysis of pre-licensure clinical trial data [GSK Biologicals Report 580299/008] suggesting a numerical imbalance (no statistically significant difference) in spontaneous abortions among ***Cervarix*** recipients whose gestations occurred around the time of vaccination (defined as the first day of Last Menstrual Period [LMP] occurring 30 days before until 45 days after vaccination), compared to control subjects (receiving a vaccine against hepatitis A), the US Food and Drug Administration (FDA) has determined that GlaxoSmithKline (GSK) Biologicals is required to conduct a post-marketing study pursuant to Section 505(o)(3)(B)(iii) of the federal Food, Drug, and Cosmetic Act (FDCA). The US FDA has also determined that an analysis of spontaneously reported adverse events, after product licensure, pursuant to subsection 505(k)(1) of the FDCA, will not be sufficient to identify an unexpected serious risk when available data indicate such potential [FDA, 2009]. In order to address this request, GSK has planned a prospective observational cohort study in women aged 15-25 years old, to assess the risk of spontaneous abortion during weeks 1-19 of gestation (≥ 20 weeks of gestation are defined as stillbirth in the US), and other pregnancy outcomes, in an exposed vaccinated cohort (women with first day of LMP between 30 days before and 90 days after any dose of ***Cervarix*** [risk period as agreed with the FDA]), when compared to the risk in a non-exposed vaccinated cohort (women with first day of LMP between 120 days and 18 months after the last ***Cervarix*** or *Gardasil* dose). The study has started recruiting subjects in the US in Q3 2011 (e-track: 114176, EPI-HPV-020), i.e. in a population comparable to the

target population of the ***Cervarix*** license application approved by the US FDA, using the Vaccines and Medications in Pregnancy Surveillance System (VAMPSS) through the Organization of Teratology Information Service (OTIS) network. The OTIS study's target sample size is 150 exposed and 300 non-exposed women within 2 years of recruitment.

In the US, the commercial distribution of ***Cervarix*** began in October 2009. However, the uptake of ***Cervarix*** in the US is lower than expected (234,710 doses were shipped in 2010; 153,730 ***doses*** in 2011***; 134,720 doses in 2012; and between 100,000 and 150,000 doses are projected for 2013).*** Hence, the EPI-HPV-020 study will be unable to reach the target sample size during the scheduled recruitment period (Projected completion of the study: Q2 2015). To address the regulatory requirement from the US FDA, GSK evaluated potential alternatives to the US study and identified a suitable alternative, namely the UK, a country with high vaccine coverage for ***Cervarix***, so that these data can be acquired, analyzed and reported to FDA in a timely matter. In the UK, ***Cervarix*** has been available since September 2008. It is recommended for women between 9 and 25 years of age [***Cervarix***, UK Package Insert, 2011a]. Because ***Cervarix*** was used exclusively in the public Human papillomavirus (HPV) immunization program in the UK from mid-2008 to 2011, this is currently the only country in Europe with sufficient vaccination coverage to undertake an alternative study to the US study.

In the UK, a public immunization programme targeting girls between 12-13 years of age including a catch-up programme for young women up to 18 years was undertaken during the academic year 2008/09. A phased catch-up programme for females born 1 September 1991 to 31 August 1995 during the 2008/09 academic year was completed by the end of the 2009/10 academic year. The programme was delivered largely through secondary schools [Crawford, 2009; Sheridan, 2009; Sheridan, 2010].

The present cohort study is an analysis of data collected in the Clinical Practice Research Datalink ***General Practitioner OnLine Database*** (CPRD ***GOLD***) in the UK. The CPRD ***GOLD*** is the world's largest computerized database of linked anonymised longitudinal medical records from primary care [Williams, 2012].

Objective(s)

Primary

- To assess the risk of spontaneous abortion during weeks 1 to 23 of gestation (UK definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 45 days after any dose of ***Cervarix***.

Secondary

- To assess the risk of spontaneous abortion during weeks 1 to 23 of gestation (UK definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 90 days after any dose of ***Cervarix***.
- To assess the risk of spontaneous abortion during weeks 1 to 19 of gestation (US definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 45 days after any dose of ***Cervarix***.
- To assess the risk of spontaneous abortion during weeks 1 to 19 of gestation (US definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 90 days after any dose of ***Cervarix***.
- To assess the risk of other pregnancy outcomes in women aged 15 to 25 years with the first day of LMP between 30 days before and 45 days after any dose of ***Cervarix***.
- To assess the risk of other pregnancy outcomes in women aged 15 to 25 years with the first day of LMP between 30 days before and 90 days after any dose of ***Cervarix.***

Study design

- This is an observational cohort study using the CPRD ***GOLD*** data source in the UK.
- Two cohorts will be defined based on the first day of gestation and exposure to ***Cervarix*** as recorded in the CPRD ***GOLD*** data source.
- This is a Targeted Safety Study (TSS) and a Post Authorisation Safety Study (PASS).
- Type of study: Self-contained.
- Epoch 001 (Synopsis Table 1): Data collection.

Synopsis Table 1 Study groups and epochs foreseen in the study

Study Groups	Number of subjects	Age (Min/Max)	Epochs
			Epoch 001
exposed cohort	Approximately 240 (-30 to +45 day window) and ***380*** (-30 to +90 day window)	15-25 years	x
non-exposed cohort	Approximately 660	15-25 years	x

Discussion of study design	The study will be conducted in the UK because of the high HPV vaccine coverage as a national immunization program and catch-up campaign have started in 2008. A feasibility analysis of HPV vaccination records and pregnancy outcomes data in CPRD ***GOLD*** database was performed by GSK and it was concluded that the proposed study could be performed using CPRD ***GOLD*** data source. To overcome the limitation related to incomplete vaccination records, the non-exposed cohort will include ***Cervarix*** vaccinated women who have had the first day of gestation between 120 days and approximately 18 months after the last ***Cervarix*** dose. This is also the design agreed with the FDA for the study in the US (EPI-HPV-020). Contrary to the lack of sensitivity, a lack of specificity in the vaccine records is very unlikely, because vaccination is an important event. Subjects in the non-exposed cohort are believed to be comparable to those in the exposed cohort as both have received ***Cervarix*** and a possible risk linked to vaccination could be reasonably excluded because of the delay between vaccination and first day of gestation. Hence, the non-exposed cohort may be slightly older than the exposed cohort. Although it is expected that the difference will be approximately one year, the analysis of primary endpoint will be adjusted for the age at first day of LMP.
Number of subjects	It is planned to include all eligible pregnant women in each of the two cohorts between September 1st 2008 and June 30th 2011.
Endpoints	**Primary** • Occurrence of spontaneous abortion during weeks 1-23 of gestation **Secondary** • Occurrence of spontaneous abortion during weeks 1-19 of gestation • Occurrence of other pregnancy outcomes : – Induced/therapeutic and other abortions – Stillbirth

  - Birth defects identified among all pregnancies with known outcome classified as live births, stillbirths and abortions. For live births, birth defects identified within the first 12 weeks of life will be included
  - Small/Large for gestational age at birth
  - Pre-term and post-term delivery
- Baby's death in the first 12 weeks of life.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF APPENDICES

PAGE

LIST OF ABBREVIATIONS

(Amended 14 March 2013)

ACIP	Advisory Committee on Immunization Practices
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CPRD *GOLD*	Clinical Practice Research Datalink ***General Practitioner OnLine Database***
CSP	Cervical Screening Program
EDD	Estimated date of delivery
EU	European Union
FDA	Food and Drug Administration
FDCA	Food, Drug, and Cosmetic Act
GP	***General Practitioner***
GPP	Good Pharmacoepidemiology practice
GSK	GlaxoSmithKline
HES	Hospital Episode Statistics
HPA	Health Protection Agency
HPV	Human Papillomavirus
HR	Hazard Ratio
ISAC	Independent Scientific Advisory Committee ***(for Medicines and Healthcare products Regulatory Agency database research)***
LMP	Last Menstrual Period
MACDP	Metropolitan Atlanta Congenital Defects Program
MBL	Mother-Baby link
MHRA	Medicines and Healthcare products Regulatory Agency (UK)
MREC	Multi-centre Research Ethics Committee

NHS	National Health Service (UK)
ONS	Office for National Statistics (England& Wales)
OR	Odds Ratio
OTIS	Organization of Teratology Information Service
PASS	Post Authorisation Safety Study
RDE	Remote Data Entry
SD	Standard Deviation
TSS	Targeted Safety Study
UK	United Kingdom
US	United States
VAMPSS	Vaccines and Medications in Pregnancy Surveillance System

GLOSSARY OF TERMS

(Amended 14 March 2013)

Birth defect:	A congenital anomaly defined as any morphological, functional and/or biochemical developmental disturbance in the embryo or foetus whether detected at birth or not.
Coded:	Information is associated with a subject number i.e. a code number. Coded information can only be linked back to the individual via a key code i.e. a listing of the research participants and their code. Within the pharmaceutical industry coding data is the usual mechanism used for protecting an individual's research data. The key code is kept secure, usually by the investigator, and GSK researchers cannot identify the research individual other than in exceptional and controlled circumstances.
Cohort study:	A form of epidemiology study where subjects in a study population are classified according to their exposure status/disease and followed over time (prospective / retrospective) to ascertain the outcome(s).
Database study:	A study using a pre-existing internal or external 'database' as data for the trial (includes meta-analysis).
Eligible:	Qualified for enrolment into the study based upon strict adherence to inclusion/exclusion criteria.
Epoch:	An epoch is a well-defined part of a protocol that covers a set of consecutive time points. Generally, an epoch is self-contained and allows to perform a data analysis to address some of the trial objectives (e.g. primary, booster, yearly follow-ups, retrospective data collection, prospective data collection).
eTrack:	GSK's tracking tool for clinical/epidemiology trials.
First trimester of pregnancy:	Time period extending from the first day of the last menstrual period (LMP) through 13 completed weeks of gestation.
Full-term delivery:	Delivery occurring ≥ 37 and ≤ 42 gestational weeks.

Gestational age:	Foetal age calculated from the first day of the LMP. It could be adjusted based on ultrasound at the first trimester of pregnancy. This involves the assumption that conception (fertilization) occurred 14 days later.
Large for gestational age:	Large newborn size defined as more than, or equal to, the 90^{th} percentile for sex and age on birth weight or length.
HPV vaccination catch-up programme:	Vaccination of a cohort of young women in addition to those targeted in the universal mass vaccination programme recommended by national Governments, as appropriate.
Independent Scientific Advisory Committee (ISAC):	Independent Scientific Advisory Committee ***for Medicines and Healthcare products Regulatory Agency database research.*** All studies that will be released into the public domain require ISAC approval. Every member is appointed by NHS.
Last menstrual period (LMP):	Considered as the first day of the LMP before conception (fertilization) onset. The first day of LMP is equal to first day of gestation. The estimated day of conception (fertilization) is calculated as the first day of LMP plus 14 days.
Live birth:	The foetus is born alive.
Non-interventional (observational) Human Subject Research:	Studies where medicinal products, should they be administered, are prescribed in normal (routine) medical practice. No medical care or medical/scientific procedures as required in a research protocol are administered to participants except as part of routine medical care.
Post Authorisation Safety Study:	A pharmacoepidemiological study or a clinical trial carried out in accordance with the terms of the marketing authorisation, conducted with the aim of identifying or quantifying a safety hazard relating to an authorised medicinal product. This includes all GSK sponsored non-interventional studies and clinical trials conducted anywhere in the world that are in accordance with the terms of the European marketing authorisation and where the investigation of safety is the specific stated objective. Note: The phrase, 'In accordance with the terms of the European marketing authorisation' means that the product is used according to the European label (e.g. within the recommended dose range, the approved formulation, indication etc.).

Post-term delivery:	Delivery after 42 completed gestational weeks (≥ 43 gestational weeks).
Pregnancy onset:	Pregnancy onset is defined as the first day of LMP plus 14 days (two weeks).
Pre-term delivery:	Delivery at less than 37 completed gestational weeks.
Small for gestational age:	Small newborn size defined as less than, or equal to, the 10^{th} percentile for sex and age on birth weight or length.
Spontaneous abortion:	Spontaneous loss of pregnancy during the first 23 weeks of gestation (UK definition). Spontaneous loss of pregnancy during the first 19 weeks of gestation (US definition).
Stillbirth:	Intra-uterine death of foetus in pregnancies 24 weeks or later of gestation (UK definition) Intra-uterine death of foetus in pregnancies 20 weeks or later of gestation (US definition)
Subject:	Term used throughout the protocol to denote an individual who has been contacted in order to participate or participates in the clinical/ epidemiology study, or a person about whom some medical information has been recorded in a database.
Targeted Safety Study:	Studies specifically planned or conducted to examine an actual or hypothetical safety concern in a product marketed anywhere in the world. This includes any GSK sponsored pharmaco-epidemiology study or clinical trial conduced anywhere in the world with the aim of identifying or quantifying a safety hazard. Although all clinical trials collect safety information as a matter of routine, only those initiated to examine a specific safety concern are considered a targeted safety study.

TRADEMARKS

The following trademarks are used in the present protocol.

Note: In the body of the protocol (including the synopsis), the names of the vaccines will be written without the superscript symbol ™ or ® and in italics.

Trademarks of the GlaxoSmithKline group of companies	Generic description
Cervarix®	Bivalent human papillomavirus (types 16, 18) recombinant vaccine

Trademarks not owned by the GlaxoSmithKline group of companies	Generic description
Gardasil® (Merck & CO., Inc.)	Quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine

1. INTRODUCTION

1.1. Background

1.1.1. Human papillomavirus and cancer

Genital human papillomavirus (HPV) infections are considered the most common sexually transmitted infections worldwide [Burd, 2003; Bosch, 2002]. High-risk HPV types are the main cause of cervical cancer and pre-cancerous lesions [Munoz, 2004; Walboomers, 1999; Stanley, 2010]. Of the approximately 40 HPV types that infect the anogenital region, 13 are considered high-risk (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 66) [Cogliano, 2005]. Among these high-risk types, HPV-16 and -18 alone cause globally about 70% of cervical cancer cases [Smith, 2007; de Sanjose, 2010].

It is estimated that in the United States (US) about 80% of women will have an HPV infection by the age of 50 years [CDC, 2009a]. HPV infection occurs rapidly after the onset of sexual debut; approximately 50% of HPV negative adolescents acquired HPV within 3-4 years of onset of sexual activity [Moscicki, 2001; Winer, 2003].

In the United Kingdom (UK), the National Health Service (NHS) has implemented a highly efficient cervical screening program (CSP). In 2007, 2,828 new cases of cervical cancer were diagnosed in the UK, making it the eleventh most common cancer in women and accounting for around 2% of all female cancers. The lifetime risk of a female developing cervical cancer was estimated at 1 in 136, and the age-standardised annual incidence rate of cervical cancer was 8.4 per 100,000 females. In 2008, there were 957 deaths from cervical cancer in the UK giving an age-standardized death rate of 2.4 per 100,000 females [Cancer Research UK , 2008].

1.1.2. Spontaneous abortion

A spontaneous abortion, sometimes called miscarriage or pregnancy loss, is a spontaneous loss of pregnancy during the first weeks of gestation. The first day of last menstrual period (LMP) is equal to first day of gestation. The cut-off between spontaneous abortion and stillbirth varies by country. The definition of spontaneous abortion includes spontaneous loss of pregnancy during the first 23 weeks of gestation in the UK (definition by NHS, UK), and during the first 19 weeks in the US (definition by CDC, US).

1.2. Rationale for the study

The Advisory Committee on Immunization Practices (ACIP) of the US Centers for Disease Control and Prevention (CDC) recommends vaccination with either the bivalent HPV vaccine (*Cervarix*) or the quadrivalent HPV vaccine (*Gardasil*) in females 9 through 26 years of age for prevention of HPV 16 or 18 related cervical cancers, precancers and dysplastic lesions [CDC, 2009b]. ***The*** quadrivalent HPV vaccine has also been recommended for vaccination of males [CDC, 2011]. *Cervarix* is approved for use in females 9 through 25 years of age in the US [*Cervarix*, US Package Insert, 2011b].

Cervarix is not recommended for use during pregnancy due to insufficient pre-licensure data in pregnant women. Because *Cervarix* is indicated in women of child-bearing age, inadvertent exposure to *Cervarix* is possible during the earlier stages of pregnancy (i.e. first trimester). Based upon the results of a subgroup analysis of pre-licensure clinical trial data [(GSK Biologicals) Clinical Report 580299/008] suggesting a numerical imbalance (no statistically significant difference) in spontaneous abortions among *Cervarix* recipients whose gestations occurred around the time of vaccination (defined as the first day of LMP occurring 30 days before until 45 days after vaccination), compared to control subjects (receiving a vaccine against hepatitis A), the US Food and Drug Administration (FDA) has determined that GlaxoSmithKline (GSK) Biologicals is required to conduct a post-marketing study pursuant to Section 505(o)(3)(B)(iii) of the federal Food, Drug, and Cosmetic Act (FDCA).

The US FDA has also determined that an analysis of spontaneously reported adverse events, after product licensure, pursuant to subsection 505(k)(1) of the FDCA, will not be sufficient to identify an unexpected serious risk when available data indicate such potential [FDA, 2009].

In order to address this request, GSK has planned a prospective observational cohort study in women aged 15-25 years old, to assess the risk of spontaneous abortion during weeks 1-19 of gestation (≥ 20 weeks of gestation are defined as stillbirth in the US), and other pregnancy outcomes, in an exposed vaccinated cohort (women with first day of LMP between 30 days before and 90 days after any dose of *Cervarix* [risk period as agreed with the FDA]), when compared to the risk in a non-exposed vaccinated cohort (women with first day of LMP between 120 days and 18 months after the last *Cervarix* or *Gardasil* dose). The study has started recruiting subjects in the US in Q3 2011 (e-track: 114176, EPI-HPV-020), i.e. in a population comparable to the target population of the *Cervarix* license application approved by the US FDA, using the Vaccines and Medications in Pregnancy Surveillance System (VAMPSS) through the Organization of Teratology Information Service (OTIS) network. The OTIS study's target sample size is 150 exposed and 300 non-exposed women within 2 years of recruitment.

In the preparation period of the EPI-HPV-020 study, the independent combined analysis on pregnancies from two multicentre, phase III randomised controlled trials (HPV-008 and HPV-009) performed by the National Cancer Institute was published [Wacholder, 2010]. The estimated rate of spontaneous abortion, defined as spontaneous loss of pregnancy within 20 weeks after LMP, was 11.5% in pregnancies in women in the *Cervarix* arm and 10.2% in the control arm. The authors concluded that there is no evidence overall for an association between HPV vaccination and risk of spontaneous abortion. Nonetheless, the authors stated that they "could not completely rule out the possibility of an increased risk among pregnancies conceived within 3 months of vaccination".

Moreover, a recent publication by the UK Medicines and Healthcare products Regulatory Agency (MHRA) reviewed the safety profile of Cervarix use in the UK from September 2008 to July 2012. No new safety concerns were identified and the number and nature of adverse drug reaction reports received was as expected after administration of at least 6 million doses of the vaccine in the UK. There was no

indication that being vaccinated with Cervarix during pregnancy results in any congenital abnormality or is associated with a risk of spontaneous abortion (or miscarriage) or other adverse pregnancy outcomes [MHRA, 2012a; MHRA, 2012b].

In the US, the commercial distribution of *Cervarix* began in October 2009. However, the uptake of *Cervarix* in the US is lower than expected (234,710 doses were shipped in 2010; 153,730 ***doses*** in 2011***; 134,720 doses in 2012; and between 100,000 and 150,000 doses are projected for 2013***). Hence, the EPI-HPV-020 study will be unable to reach the target sample size during the scheduled recruitment period (Projected completion of the study: Q2 2015). To address the regulatory requirement from the US FDA, GSK evaluated potential alternatives to the US study and identified a suitable alternative, namely the UK, a country with high vaccine coverage for *Cervarix*, so that these data can be acquired, analyzed and reported to FDA in a timely matter. In the UK, *Cervarix* has been available since September 2008. It is recommended for women between 9 and 25 years of age [*Cervarix*, UK Package Insert, 2011a]. Because *Cervarix* was used exclusively in the public HPV immunization program in the UK from mid-2008 to 2011, this is currently the only country in Europe with sufficient vaccination coverage to undertake an alternative study to the US study.

In the UK, a public immunization programme targeting girls between 12-13 years of age including a catch-up programme for young women up to 18 years was undertaken during the school year 2008/09. A phased catch-up programme for females born 1 September 1991 to 31 August 1995 during the 2008/09 academic year was completed by the end of the 2009/10 academic year. The programme was delivered largely through secondary schools [Crawford, 2009; Sheridan, 2009; Sheridan, 2010].

The present cohort study is an analysis of data collected in the Clinical Practice Research Datalink ***General Practitioner OnLine Database*** (CPRD ***GOLD***) in the UK. The CPRD ***GOLD*** is the world's largest computerised database of linked anonymised longitudinal medical records from primary care [Williams, 2012].

(Amended 14 March 2013)

2. OBJECTIVES

2.1. Primary objective

- To assess the risk of spontaneous abortion during weeks 1 to 23 of gestation (UK definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 45 days after any dose of *Cervarix*.

Refer to 5.1.1 for the definition of the primary endpoint.

2.2. Secondary objective

- To assess the risk of spontaneous abortion during weeks 1 to 23 of gestation (UK definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 90 days after any dose of *Cervarix*.

- To assess the risk of spontaneous abortion during weeks 1 to 19 of gestation (US definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 45 days after any dose of *Cervarix*.
- To assess the risk of spontaneous abortion during weeks 1 to 19 of gestation (US definition) in women aged 15 to 25 years with the first day of LMP between 30 days before and 90 days after any dose of *Cervarix*.
- To assess the risk of other pregnancy outcomes in women aged 15 to 25 years with the first day of LMP between 30 days before and 45 days after any dose of *Cervarix*.
- To assess the risk of other pregnancy outcomes in women aged 15 to 25 years with the first day of LMP between 30 days before and 90 days after any dose of *Cervarix*.

Refer to 5.1.2 for the definition of the secondary endpoints.

3. STUDY DESIGN

- This is an observational cohort study using the CPRD ***GOLD*** data source in the UK.
- Two cohorts will be defined based on the first day of gestation and exposure to *Cervarix* as recorded in the CPRD ***GOLD*** data source.
- This is a Targeted Safety Study (TSS) and a Post Authorisation Safety Study (PASS).
- Type of study: Self-contained.
- Epoch 001 (Table 1): Data collection.

Table 1 Study groups and epochs foreseen in the study

Study Groups	Number of subjects	Age (Min/Max)	Epochs
			Epoch 001
exposed cohort	Approximately 240 (-30 to +45 day window) and ***380*** (-30 to +90 day window)	15-25 years	x
non-exposed cohort	Approximately 660	15-25 years	x

(Amended 14 March 2013)

3.1. HPV vaccine coverage and pregnancy rate in the UK teenagers

In the UK public HPV immunization program (12-13 year olds), HPV vaccination coverage in the UK for 2009 was 88.1%, 86.0% and 80.1% for the first, second and third dose respectively. The coverage in the catch-up program for women (15-18 years old) was 62.2%, 54.2% and 31.8% for the first, second and third dose, respectively [Sheridan, 2009]. ***HPV vaccination coverage in the UK for 2010/11 was 89.0%, 87.6% and 83.8% for the first, second and third dose respectively [Health Protection Agency, 2012].***

The teenage pregnancy rate in the UK is reported to be around 40,000 teenage pregnancies each year [ONS, 2009; ONS, 2010]. For girls aged 15-17 years, the

provisional number of conceptions in England was 35,966 in 2009, reflecting a conception rate of 38.2 per 1000 person years. The pregnancy rate in females 15-17 years in the US is estimated at 42 per 1000 in 2004 [Ventura, 2008].

(Amended 14 March 2013)

3.2. Feasibility analysis and limitations of the study design

A feasibility analysis of HPV vaccination records and pregnancy outcomes data in CPRD ***GOLD*** database was performed by GSK**.**

Immunization with HPV vaccine is coded in the CPRD ***GOLD*** with the following codes: 93489, 93621, and 95554 for the 1st, 2nd and 3rd dose, respectively.

Exploration of the CPRD ***GOLD*** database identified 148,731 vaccinated subjects (Table 2) and 83,686 aged between 15 and 25 years (Table 3).

Table 2 Counts of HPV vaccination per year in the CPRD *GOLD*

Year of vaccination	N
2007	247
2008	29,301
2009	87,233
2010	29,364
2011	2,586

Created by GSK Biologicals, 2012
Data extracted from CPRD ***GOLD*** version ffcprd_smart_2012Q3

Table 3 Number of HPV vaccinations per age range in the CPRD *GOLD*, 2007-2011

Age Range	N subjects HPV
9-14	64,667
15-19	82,382
20-25	1,304

Created by GSK Biologicals, 2012
Data extracted from CPRD ***GOLD*** version ffcprd_smart_2012Q3

However, data on HPV vaccination status is not captured for all vaccinees in CPRD ***GOLD*** (lack of sensitivity) and absence of a vaccination record does not imply absence of vaccination. For instance, 70.7% of the 01SEP1996 – 31AUG1997 birth cohort was reported as HPV vaccinated in CPRD ***GOLD*** vs. 85.9% reported by the UK Health Protection Agency (HPA). Corresponding numbers for the 01SEP1994 – 31AUG1995 birth cohort are 70.6% vs. 81.9% (Table 4).

Table 4 HPV vaccination coverage in CPRD *GOLD* and HPA data

	CPRD *GOLD* **			UK*
Birth Cohort	N vaccinated	N total	%	%
01SEP1996 and 31AUG1997	16,028	22,685	70.7%	85.9%
01SEP1994 and 31AUG1995	16,343	23,143	70.6%	81.9%

Created by GSK Biologicals, 2012
*From Health Protection Agency Department of Health [Health Protection Agency, 2012]
** Data extracted from CPRD ***GOLD*** version ffcprd_smart_2012Q3

To overcome this limitation related to incomplete vaccination records, the non-exposed cohort will include *Cervarix* vaccinated women who have had the first day of gestation between 120 days and approximately 18 months after the last *Cervarix* dose. That would increase the specificity in the control group since it would be very unlikely for women to be vaccinated again after completing the three HPV doses. This is also the design agreed with the FDA for the study in the US (EPI-HPV-020).

Contrary to the lack of sensitivity in the overall population, a lack of specificity in the vaccine records is very unlikely, because vaccination is an important event.

Subjects in the non-exposed cohort are believed to be comparable to those in the exposed cohort as both have received *Cervarix* and a possible risk linked to vaccination could be reasonably excluded because of the delay between vaccination and first day of gestation. Hence, the non-exposed cohort may be slightly older than the exposed cohort. Although it is expected that the difference will be approximately one year, the analysis of primary endpoint will be adjusted for the age at first day of LMP.

The number of subjects in CPRD ***GOLD*** with at least one pregnancy in the study period from January 1st 2007 to June 30th 2011 by age range, among HPV vaccinated and non-vaccinated women, are provided in Table 5.

Table 5 Pregnancies in the CPRD *GOLD* within the study period (≥ Jan 1st, 2007)

Age Range	N pregnant and HPV vaccinated women	N total pregnant women
[11 - 15]	2,155	5,066
[16 - 20]	8,232	60,438
[21 - 25]	181	84,932
[26 - 30]	36	103,803
[31 - 35]	14	87,153
[36 - 40]	6	46,224
[41 - 45]	5	13,861
[46 - 50]	6	4,174

Based on CPRD ***GOLD*** version ffCPRD_smart_2012Q3.
Procedure: This analysis is a feasibility using a preliminary list of pregnancy identifiers, no use of free text. We selected the subject identifiers and the pregnancy identifier event date from the clinical data in CPRD ***GOLD*** and matched on immunization codes from HPV (93489, 93621, 95554), restricting the event date from January 1st 2007 to June 30th 2011. For determination of the total pregnancy number, the same procedure was used, except for the matching on

immunization codes. These data are then matched to each female subject's year of birth to select the age at onset of the pregnancy identifiers. The number of subjects vaccinated and the total number of pregnant subjects are tabulated by 5 year age-classes from 11 to 50. If there are multiple pregnancy indicators per subject then the one with the earliest date is used.

Based on the recent CPRD ***GOLD*** release (2012Q3), 2440 subjects out of 78,111 subjects (3.1%) 15-25 years of age and exposed to *Cervarix* were identified with a simultaneous *Cervarix* vaccination and a LMP date between September 1st 2008 and June 30th 2011. Among these potential eligible subjects (not taking into account all inclusion/inclusion criteria as defined below in this protocol), 243 were identified as being exposed to *Cervarix* vaccine (first day of LMP between 30 days before and 45 days after any dose), 379 (first day of LMP between 30 days before and 90 days after any dose) and 667 as non-exposed (first day of LMP between 120 days and 18 months after the last dose of *Cervarix* vaccine).

The estimate for spontaneous abortions in the CPRD ***GOLD*** of 11.6% is in line with earlier literature reports in US and UK [e.g. 13.8%, Saraiya, 1999; 7.0%, Seamark, 2001; 14.0%, Ventura, 2008; 11.5% and 10.2%, Wacholder, 2010; 14.9%, Devine, 2010].

Based on the numbers obtained above, the proposed study could be performed using CPRD ***GOLD*** data source. The risk of spontaneous abortions in the exposed cohort will be compared to that in the non-exposed cohort.

(Amended 14 March 2013)

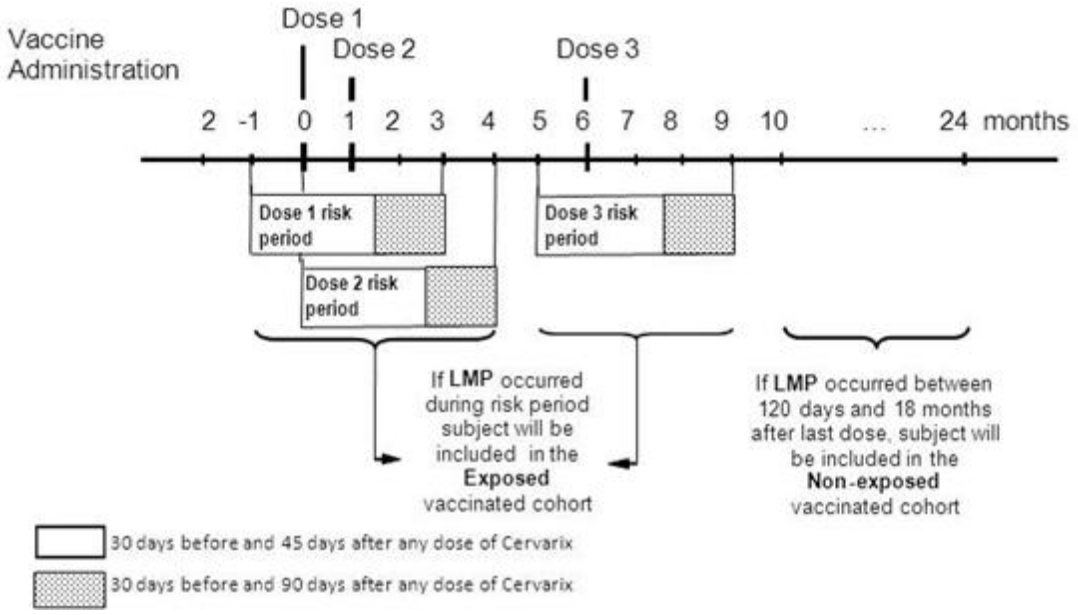
4. STUDY POPULATION

4.1. Study cohorts

Two cohorts of pregnant women will be defined based on exposure to *Cervarix* as recorded in the CPRD ***GOLD*** (Figure 1) **(Amended 14 March 2013)**:

- The **exposed cohort** will include women with the first day of LMP: between 30 days before and 45 days after any *Cervarix* dose, as requested by the FDA, and between 30 days before and 90 days after any *Cervarix* dose, as defined in the study EPI-HPV-020 and as used for the pooled analysis by Wacholder [Wacholder, 2010].
- The **non-exposed cohort** will include women with the first day of LMP between 120 days and 18 months after their last *Cervarix* dose (and no further *Cervarix* dose before the outcome).

Figure 1 Definition of exposure and non-exposure



Note: Last dose can be first or second dose if no subsequent dose was administered.

4.2. Cohorts identification and follow-up

Cohort identification will be done in a stepwise approach using defined algorithms (see Appendix A).

4.2.1. Screened population

The screened population will be comprised of all women that have been vaccinated with HPV vaccine. Medical codes for HPV vaccination are: 93489 (first dose), 93621 (second dose), 95554 (third dose).

4.2.2. Eligible population

The eligible population will be comprised of all subjects included in CPRD ***GOLD*** who fulfil to the inclusion/exclusion criteria.

(Amended 14 March 2013)

4.2.2.1. Inclusion criteria for exposed cohort

All subjects must satisfy all of the following criteria at entry into the exposed cohort:

- Full date (day/month/year) of first day of LMP available in the database or calculated from Estimated date of delivery (EDD) (day/month/year)
- Female, aged between, and including, 15 and 25 years of age at LMP
- Recorded in the CPRD ***GOLD*** for at least 12 months at LMP

- First day of LMP available in the database or calculated from EDD (day/month/year) between September 1st 2008 and June 30th 2011
- At least one dose of *Cervarix* received between September 1st, 2008 and June 30th, 2011
- Full date (day/month/year) of *Cervarix* vaccination(s) available
- First day of LMP available in the database or calculated from EDD between 30 days before and 90 days after any dose of *Cervarix*
- Subject defined as acceptable in CPRD ***GOLD*** [1]

(Amended 14 March 2013)

4.2.2.2. Inclusion criteria for non-exposed cohort

All subjects must satisfy all of the following criteria at entry into the non-exposed cohort:

- Full date (day/month/year) of first day of LMP available in the database or calculated from EDD (day/month/year)
- Female, aged between, and including, 15 and 25 years of age at LMP
- Recorded in the CPRD ***GOLD*** since at least 12 months at LMP
- First day of LMP available in the database or calculated from EDD (day/month/year) between September 1st 2008 and June 30th 2011
- At least one dose of *Cervarix* received between September 1st 2008 and June 30th 2011
- Full date (day/month/year) of *Cervarix* vaccination(s) available
- First day of LMP available in the database or calculated from EDD between 120 days and 18 months after the last dose of *Cervarix*
- Subject defined as acceptable in CPRD ***GOLD***

(Amended 14 March 2013)

4.2.2.3. Exclusion criterion for exposed cohort

The following criterion should be checked at the time of entry into the exposed cohort. If this exclusion criterion applies, the subject must not be included in this cohort:

- Female with the first day of LMP between 30 days before and 90 days after at least one dose of unspecified HPV vaccine and/or *Gardasil*.

[1] Defined in Appendix A. CPRD ***GOLD*** indicates whether the patient has met certain quality standards (based on an 'internal CPRD ***GOLD*** check') **(Amended 14 March 2013)**

4.2.2.4. Exclusion criteria for non-exposed cohort

The following criteria should be checked at the time of entry into the non-exposed cohort. If those exclusion criteria apply, the subject must not be included in this cohort:

- Female with the first day of LMP between 30 days before and 90 days after at least one dose of unspecified HPV vaccine and/or *Gardasil*.
- Female included for a previous pregnancy in the exposed cohort.

4.3. Included population

It is planned to include all eligible pregnant women in each of the two cohorts.

All included subjects will be followed until: end of pregnancy, subject de-enrolment from CPRD ***GOLD*** or study termination.

(Amended 14 March 2013)

5. DATA EVALUATION: CRITERIA FOR EVALUATION OF OBJECTIVES

5.1. Endpoints

5.1.1. Primary endpoint

- Occurrence of spontaneous abortion[2] during weeks 1-23 of gestation

5.1.2. Secondary endpoints

- Occurrence of spontaneous abortion during weeks 1-19 of gestation
- Occurrence of other pregnancy outcomes:
  - Induced/therapeutic and other abortions (according to UK and US definitions of spontaneous abortion)
  - Stillbirth (according to UK and US definitions of stillbirth)
  - Birth defects identified among all pregnancies with known outcome classified as live births, stillbirths and abortions. For live births, birth defects identified within the first 12 weeks of life will be included
  - Small/Large for gestational age at birth
  - Pre-term and post-term delivery
- Baby's death in the first 12 weeks of life.

[2] Refer to GLOSSARY OF TERMS for definition of spontaneous abortion and of the other pregnancy outcomes

5.2. Data collection

5.2.1. The Clinical Practice Research Datalink *General Practitioner OnLine Database* (CPRD *GOLD*)

The CPRD ***GOLD*** is the world's largest computerised database of linked anonymised longitudinal medical records from primary care. The data are drawn from the computer systems used by general practitioners to maintain the clinical records within their practices. As of March 2011, CPRD ***GOLD*** contained records from over 12 million patients contributing 64 million person-years of prospectively recorded high-quality primary healthcare data [Williams, 2012].

The CPRD ***GOLD*** is operated on a non-profit basis by the MHRA, containing coded longitudinal medical records from general practices and more recently from hospital-based care (e.g., Hospital Episode Statistics, HES). The current linkage between CPRD ***GOLD*** primary care data and HES data is around 40% as of Q1 2012. The CPRD ***GOLD*** database is licensed in-house by GSK. Data quality is monitored continuously by the MHRA and practices that fail to maintain the required standards are removed from the database.

The latest update provided by the CPRD ***GOLD*** team in Q3 of 2012 (second release of 2012) contains data for 10,547,532 research standard patients, drawn from 644 practices throughout the UK. A total of 4,621,799 patients from ***546*** practices are currently active in the database. The CPRD ***GOLD*** population closely matches the age and gender distribution of the UK population as a whole. Mean follow-up is 6.8 years (median 5.0 years). Recorded data include demographic information, prescription details, clinical events, preventive care provided, specialist referrals, hospital admissions and their major outcomes. Data are retrieved by means of the READ classification system; READ codes are a coded thesaurus of clinical terms, which are the basic means by which clinicians record patient findings and procedures in health and social care IT systems across primary and secondary care (e.g. ***general practitioner [***GP***]*** surgeries and pathology reporting of results). The Medcodes are the abbreviated terms which mean CPRD ***GOLD*** medical codes. Medcodes consisting of READ codes are used to enter medical diagnosis in the CPRD ***GOLD*** database.

The use of the CPRD ***GOLD*** database for research on pregnancies has been reported in a few publications. Devine *et al* have developed an algorithm for identification of pregnancies [Devine, 2010]. Hardy *et al* have used CPRD ***GOLD*** records, for linking mother and baby automated records and to describe the frequency, type and pregnancy risk level of medications prescribed prior to and during early pregnancy [Hardy, 2006]. The CPRD ***GOLD*** database has been also assessed as an alternative for pregnancy registries to study drug exposure [Charlton, 2008], major congenital malformations [Charlton, 2010], and identification of potential teratogens [Charlton, 2011].

Data will be automatically extracted from the second release of 2012 (2012Q3) available in GSK, using the algorithms defined in the Appendix A. In addition, specific free text associated with endpoints (e.g. pregnancy outcomes), will be requested where necessary for endpoint confirmation. The free text request will be sent by GSK to CPRD ***GOLD***.

Extracted free text will be de-identified before sending to GSK. Review of free text will be done by a GSK identified reviewer (Pallas, Health research and consultancy B.V) for completing extracted data; a final case ascertainment step will be performed as described in section ***5.2.4.2,*** 5.2.4.3 and 5.2.6.

The final study database will consist of data extracted from CPRD ***GOLD*** and additional data from free text review. The study database will be locked and stored by GSK Biologicals' data management.

(Amended 14 March 2013)

5.2.2. Validation of the CPRD *GOLD* database

The CPRD ***GOLD*** database is licensed in-house by GSK. GSK has not audited CPRD ***GOLD***, but CPRD ***GOLD*** has internal procedures to ensure the quality of data. The use of the CPRD ***GOLD*** database is well documented and it has been used by both MHRA and FDA.

(Amended 14 March 2013)

5.2.3. Study period

Study period will include women being vaccinated with *Cervarix* and with the first day of LMP between September 1st 2008 and June 30th 2011 and followed for pregnancy outcomes until June 25th 2012 (considering 9 months of pregnancy and 12 weeks after delivery). The 2012Q3 database is up to June 25th 2012.

5.2.4. Collected data

Defined algorithms are detailed in Appendix A.

5.2.4.1. Women characteristics

The following data will be extracted for the analysis population:

- Demographic characteristics: birth month and birth year, sex, region, practice identifier, family number, marital status, date of death (if applicable) and acceptable patient flag
- CPRD ***GOLD*** information: CPRD ***GOLD*** start date, current registration date, registration gaps, registration status
- History of previous pregnancies: parity and number of previous pregnancies
- Lifestyle during pregnancy: smoking, alcohol consumption
- Medical conditions during pregnancy: high blood pressure, diabetes
- Vaccines:

  Administration of any other vaccine from 90 days before to 90 days after first day of gestation will be collected: date of vaccination, medcode and immunization type will

be extracted from the immunization file. Cross-tabulation of medcode and vaccine names/class is detailed in Appendix A.

- Vaccination with another vaccine within 90 days before first day of gestation or within the 90 days after first day of gestation (binary variable).
- Contra-indicated drugs during gestation:

  The drugs used during the first trimester of gestation will be classified as: non contra-indicated drug, manufacturer advice "to use with caution" drugs, manufacturer mentioned advice "to avoid" drugs (this classification will be done respectively on the list of all drugs used during gestation reported in CPRD ***GOLD*** for each subject).

(Amended 14 March 2013)

5.2.4.2. Pregnancy outcomes

Pregnancy outcomes will be identified using defined algorithms (see Appendix A).

Definition of pregnancy outcomes (according to the UK or US definition):

1. Abortion defined as a loss of pregnancy with the first 23 weeks (≤ 161 days) or within the first 19 weeks (≤ 133 days) of gestation, with the following sub-categories:
   1. Spontaneous abortion
   2. Induced / therapeutic abortion
   3. Other abortion
2. Stillbirth defined as intra-uterine death of foetus after 23 weeks (>161 days) or after 19 weeks (> 133 days) of gestation
3. Live births, with the following sub-categories:
   1. Small for gestational age at birth
   2. Large for gestational age at birth
   3. Pre-term delivery (< 37 weeks of gestation)
   4. Full term delivery (between 37 weeks and 42 completed weeks of gestation)
   5. Post-term delivery (> 42 ***completed*** weeks of gestation)

(Amended 14 March 2013)

Confirmation of pregnancy outcomes (according to the UK or US definition):

For each pregnancy, the gestational age at outcome will be computed in days as:

- Gestational_age (days) = (date of outcome) – (date of first day of gestation) +1

All abortions and stillbirths will be confirmed using this gestational age:

- if gestational_ age ≤ 161 days or ≤ 133 days, → abortion is confirmed

- if gestational_age > 161 days or > 133 days, → stillbirth is confirmed

For all the identified cases of spontaneous abortion, other abortion, and stillbirth, the associated free text will be reviewed by a GSK-identified reviewer for sub-category classification (spontaneous, induced/therapeutic, and others). As far as possible, other abortion will be re-classified as spontaneous or induced/therapeutic abortion.

Cases of live birth will be classified as pre-term, full term and post-term delivery based on the gestational age at delivery:

- if gestational_ age < 259 days then pre-term delivery is confirmed
- if gestational_age ≥ 259 and ≤ 294 days then full term delivery is confirmed
- if gestational_age > 294 days then post-term delivery is confirmed

Cases of live birth for which confirmation of pre-term, full term or post-term delivery is not possible because the gestational age is missing will be classified as non-confirmed pre-term, full term or post-term delivery, respectively.

Confirmation of cases of "small for gestational age at birth" and of "large for gestational age at birth" will be done as follows by comparing the weight and/or length at birth with the age and sex specific normal weight or length distributions. If the measured weight or length at birth is ≤ 10^{th} percentile of normal weight or length, the "small for gestational age at birth" will be confirmed. If the measured weight or length at birth is ≥ 90^{th} percentile of normal weight or length, the "large for gestational age at birth" will be confirmed.

Cases of "small for gestational age at birth" and of "large for gestational age at birth" identified in the codes but for which confirmation is not possible because both weight and length at birth are missing will be classified as non-confirmed "small for gestational age at birth" or non-confirmed "large for gestational age at birth".

The GSK Safety physician will also review the following pregnancy outcomes: abortions (spontaneous abortions, induced/therapeutic abortions, other abortions), stillbirths, neonatal death, for life births: pre-term and post-term delivery, small for gestational age at birth or large for gestational age at birth.

(Amended 14 March 2013)

5.2.4.3. Baby (live birth) characteristics and confirmation of birth defects

The baby will be identified using the Mother-Baby link (MBL) available in CPRD ***GOLD***. In case of multiple pregnancy, all babies will be followed.

For the live births, the following baby characteristics at birth will be extracted:

- Gender
- Weight
- Length

- Apgar score (at 1 minute and 5 minutes of life)

Codes related to baby's death within 12 weeks of life will be extracted with the associated free text.

Birth defects reported in the mother's file and/or in the baby's files within 12 weeks after delivery will also be extracted with the associated free text.

All cases with an identified birth defect will be reviewed by a GSK Safety physician using the birth defects classification by the CDC Metropolitan Atlanta Congenital Defect Program or MACDP [Centers for Disease Control and Prevention, 2004 and Centers for Disease Control and Prevention, 2007]. Cases with apparent birth defects will be described in detail. Link between medical codes in CPRD ***GOLD*** and MACDP terminology is detailed in Appendix A.

(Amended 14 March 2013)

5.2.5. Derived variables

5.2.5.1. First day of LMP

Definition of first day of LMP will be as follows:

1. The date of first day of LMP as reported in the database
2. If date of first day of LMP is missing or LMP date is equal to system date (± 2 days) but the EDD is completed in CPRD ***GOLD***, the first day of LMP will be calculated as follows: EDD-***280*** days ***(280 days equal to the median of gestational age: 40 weeks)***.

(Amended 14 March 2013)

5.2.5.2. First day of gestation

Definition of first day of gestation will be as follows:

1. the day reported as the first day of her LMP,
2. the day adjusted based on available ultrasound scan test data performed at week 12 (± 2 weeks) after first day of LMP. The first day of gestation will be adjusted only if there is a discrepancy between the first day of LMP and the ultrasound-based first day of gestation of seven days or more. Associated free text will be reviewed for the ultrasound scan test performed to adjust for the first day of gestation when needed,
3. if ultrasound scan test data is not available at week 12 (± 2 weeks) after first day of LMP, other ultrasound scan tests will be used to adjust for the first day of gestation knowing that accuracy of gestational age estimation via ultrasound decreases over time as shown in Table 6.

Table 6 Adjustment of the first day of gestation using ultrasound scan test

Ultra-sound completed (Wk)	Ultrasound accuracy (days)	Subject-reported first day of gestation will be adjusted if discrepancy is greater than:
]0-14]	5 - 7	7 days compared to ultrasound-based LMP date
]14-26]	10 - 14	14 days compared to ultrasound-based LMP date
≥ 27	21	21 days compared to ultrasound-based LMP date

All studied pregnancies will be confirmed by the occurrence of at least one other pregnancy identifier between the first day of gestation and the date of end of pregnancy. List of codes (read codes) of pregnancy identifiers are detailed in Appendix A.

Based on the defined first day of gestation, the exposed status (exposed vs. non-exposed) could be adjusted and the subjects will be reassigned to the correct cohort.

5.2.5.3. Time from first day of gestation to spontaneous abortion (UK and US definition)

The primary and first secondary endpoints are the occurrence of spontaneous abortion during the first 23 weeks of gestation (UK definition) and during the first 19 weeks of gestation (US definition). The time (day) from first day of gestation to spontaneous abortion will be computed as the difference between the date of spontaneous abortion and the date of first day of gestation + 1 day.

Time from first day of gestation to spontaneous abortion will be censored at the first of the following events:

- At the end of week 23 (or week 19) for women still pregnant at the beginning of week 24 (or week 20)
- At the time of induced or therapeutic abortion
- At the time of de-enrolment from CPRD ***GOLD***, if during the first 23 weeks (or 19 weeks) of gestation
- At the time of death if during the first 23 weeks (or 19 weeks) of pregnancy.

(Amended 14 March 2013)

5.2.5.4. Other derived variables

The following variables will be derived from the CPRD ***GOLD*** data:

- Woman's date of birth will be defined as the 15th of the birth month and birth year. If the birth month is missing, the birth date will be defined as the 30th June of the birth year
- Incomplete dates will be substituted as follows for calculation of age and/or time to event; if the day is missing the date will be defined as the 15th of the month, if both the day and the month are missing, the date will be defined as 30th June of the year
- Age at first day of gestation will be computed as the difference between the first day of gestation and the date of birth and will be presented in completed year

- Age at a specific event will be computed similarly
- The time from first day of gestation to a specific event (e.g. spontaneous abortion) will be computed as the difference between the date of event (Date1) and the date of first day of gestation (Date0) + 1 and it will be represented in days.

(Amended 14 March 2013)

5.2.6. Final case ascertainment step

After the first review by the GSK Safety physician (see section 5.2.4.2), individual data related to pregnancy outcomes defined as ***spontaneous and other*** abortions, stillbirths, ***and*** baby's death, will be reviewed by two pharmaco-epidemiologists specialised in teratology (see contributors) for final case ascertainment. The experts will be blinded with regards to HPV vaccine exposure. In the event that the experts disagree, a second review step will be conducted and, if they cannot reach an agreement, then the different clinical opinion of the experts will be listed.

For birth defects (see section 5.2.4.3), when the GSK Safety physician is not able to classify a case, the two experts will be asked to review it and to reach an agreement.

(Amended 14 March 2013)

5.3. Hypotheses

Applicable, null hypothesis: the risk of spontaneous abortion in exposed women is equal to the risk of spontaneous abortion in women from the non-exposed cohort.

Alternative hypothesis: the risk of spontaneous abortion in exposed women is higher than the risk of spontaneous abortion in women from the non-exposed cohort.

5.4. Sample size calculation

Using a two-sided log-rank test with type I error rate of 5% (Software: PASS 2005) and results from the feasibility analysis which has estimated the potential eligible subjects (see section 3.2), the study would have the following power to detect a relative risk (Hazard ratio, HR) of 2.0 of spontaneous abortion between the exposed and non-exposed subjects: 98% power if the first day of LMP is between 30 days before and 45 days after and 99% power if the first day of LMP is between 30 days before and 90 days after any dose of *Cervarix*. The HR detectable with 80% power would be 1.69 and 1.60, respectively.

Those calculations are based on the following assumptions:

- Rate of spontaneous abortion: approximately 11.5% of pregnancies (see section 3.2, feasibility analysis).
- Proportion of dropped out subjects before the first 23 weeks of gestation: 20%.

5.5. Statistical methods

Examples of statistical tables and figure templates are given in Appendix B.

5.5.1. Subject disposition

Subject disposition will be summarized by computing:

- Number of screened subjects.
- Number (%) of non-eligible subjects for each of the following reasons of non-eligibility:
  - No certainty of *Cervarix* vaccination
  - All *Cervarix* doses outside September 1th 2008 and June 30th 2011
  - No pregnancy identifiers
  - First day of LMP outside September 1th 2008 and June 30th 2011
  - Age at first day of LMP outside 15 – 25 years
  - Start in CPRD ***GOLD*** less than one year before first day of LMP
  - Subject not flagged as acceptable in CPRD ***GOLD***
  - Neither in the exposed nor non-exposed cohort.
- Number of eligible subjects in each cohort.
- Number of included subjects in each cohort.

A detailed, comprehensive list of reasons for elimination from exposed and non-exposed cohort analyses will be established at the time of data cleaning.

(Amended 14 March 2013)

5.5.2. Demographic and baseline characteristics

Demographic and baseline characteristics of all included subjects (age at first day of gestation, age at vaccination, marital status, smoking, alcohol consumption, region (GP practice)) will be summarized per cohort and overall, using descriptive statistics.

Frequency tables will be generated for categorical variables.

Mean, standard error, median and range, will be provided for continuous variables.

Medical condition during pregnancy (diabetes, high blood pressure) will be summarized in a frequency table.

Availability of data in CPRD ***GOLD*** will be summarized by descriptive statistics: proportion of Baby and Mother links.

Proportion of subjects with first day of LMP before, during and after H1N1 pandemic in the UK (June 1st 2009 to February 28th 2010 according to Pebody [Pebody, 2010]).

The two cohorts will be compared for their demographic and baseline characteristics using Fisher's exact test or Student t-test.

(Amended 14 March 2013)

5.5.3. Pregnancy outcomes

5.5.3.1. Endpoints related to spontaneous abortion

Proportion of spontaneous abortion will be calculated for the exposed and the non-exposed cohorts as the number of pregnant women with spontaneous abortion divided by the total number of included pregnant women.

The risk of abortion according to gestational time will be depicted using Kaplan-Meier curves.

The primary analysis will be the comparison of the risk of spontaneous abortion during the first 23 weeks of gestation and during the first 19 weeks of gestation in the exposed cohort and the non-exposed cohort using a Cox regression model with spontaneous abortion as the dependent variable, a dichotomous exposure as the independent variable and with age at first day of gestation as a covariate. The HR and its 95% confidence interval (CI) will be derived. The analysis will be done for two risk periods for the exposed subjects: the 30 days before to 45 days after any dose of *Cervarix* period and the 30 days before to 90 days after any dose of *Cervarix* period. The main analysis (primary objective) is the -30 to +45 day risk period.

For subjects with no spontaneous abortion, time-to-event will be censored at: the end of week 23 (or week 19), date of induced or therapeutic abortion, date of death, date of last available pregnancy data whichever occurs first.

The following sensitivity analyses will be performed:

- Cox regression including in addition to the age at the first day of gestation other covariates (region, smoking, alcohol consumption, vaccination, diabetes, high blood pressure, number of previous pregnancies, other vaccination, use of contra-indicated drugs, exposure to H1N1 pandemic)
- Analysis of all abortions (defined as induced, therapeutic and other abortion) using the same model as for the primary analysis

5.5.3.2. Other endpoints

All other endpoints are binary variables: stillbirths, live births, full term delivery, pre-term and post-term delivery, small for gestational age, large for gestational age, birth defects and baby's death.

The proportion of each outcome will be computed for the exposed and non-exposed cohorts as the number of cases divided by the total number of subjects with known pregnancy outcome. An exact 95% CI will be computed.

The exposed and non-exposed cohorts will be compared using logistic regression with the pregnancy outcome as the dependent variable, a dichotomous exposure as the independent variable and the age at first day of gestation as a covariate. The odds ratio (OR) and its 95% CI will be derived.

Baby characteristics will be summarized per cohort using descriptive statistics: n of non-missing observations, mean, standard deviation, minimum and maximum for continuous variables (birth weight and length) and frequency table (n, %) for categorical variables (sex, Apgar score at 1 and 5 min).

5.6. Statistical calculations

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

All the statistical tests will be two-sided at alpha level of 0.05.

5.6.1. Handling of missing data

Missing data will not be substituted.

5.6.2. Descriptive statistics

Age at first day of gestation will be summarized by descriptive statistics per cohort (exposed and non-exposed) and overall: n of subjects, mean, standard deviation (SD), median, minimum and maximum and compared between the two cohorts using a student t-test.

Smoking status (smoker vs. non-smoker), alcohol consumption (Yes/No) will be summarized in frequency tables (n, %) per cohort and overall. The two cohorts will be compared using a Fisher's exact test.

Diabetes and high blood pressure will be summarized in frequency tables (n, %) per cohort and overall. The two cohorts will be compared using a Fisher's exact test.

First day of LMP occurring during the H1N1 pandemic (01JUN2009 to 28FEB2010) will be summarized in frequency tables (n, %) per cohort and overall. The two cohorts will be compared using a Fisher's exact test.

Exposure to other vaccines will also be summarized in frequency tables per cohort and period (within 3 months before first day of gestation and during the 1st trimester of gestation).

Number of subjects exposed to contra-indicated drugs during the 1st trimester of gestation will be tabulated per cohort and overall.

All the pregnancy outcomes (see section 5.2.4.2) and birth defects (see section 5.2.4.3) will be summarized by descriptive statistics per cohort. The statistics will be computed:

- Number of cases
- Proportion computed as the number of cases divided by the total number of pregnancies

All the cases of pregnancy outcome including outcomes with apparent birth defects and baby's deaths will be listed with data about exposure status, age at first day of gestation, gestational age at the event date, smoking status, alcohol consumption, diabetes, high blood pressure, exposure to other vaccines, to contra-indicated drugs and H1N1 pandemic.

5.6.3. Statistical models

5.6.3.1. Cox proportional hazards model

Cox model will be computed using the SAS PHREG procedure. The dependent variable is the time between first day of gestation and the event or at censoring.

For the spontaneous abortion analysis, time to event for subject with no spontaneous abortion will be censored at the earliest of the following time-points:

- 161 days (end of week 23) or 133 days (end of week 19)
- Lost-to-follow up (***de-***enrolment from CPRD ***GOLD***) **(Amended 14 March 2013)**
- Mother's death
- Pregnancy termination

The main model (Model 1) will include the exposure status (exposed vs. non-exposed) as a binary independent variable and the age at first day of gestation as a continuous covariate. The adjusted HR will be derived as the exponential of the coefficient associated with the exposure status and its 95% Wald CI.

Survival curve will be depicted for the exposed and non-exposed cohort (using the Option Plots(overlay)=survival in the SAS PHREG procedure).

The following models will also be computed as sensitivity analyses:

Model 2: A Cox model including, in addition to the age at first day of gestation, other covariates

- Smoking during pregnancy (3 classes: yes, no, unknown)
- Exposure to H1N1 pandemic (yes, no)
- Region (class variable)
- Alcohol consumption during pregnancy (3 classes: yes, no, unknown)
- Diabetes during pregnancy (3 classes: yes, no, unknown)

- High blood pressure during pregnancy (3 classes: yes, no, unknown)
- Number of previous pregnancies (sum of the number of previous births and the number of previous abortions)
- Vaccination with another vaccine within 90 days before first day of gestation or within the 90 days after first day of gestation (binary variable)
- Use of contra-indicated drugs during the first trimester of pregnancy (three classes: non contraindicated drug, manufacturer advice"to-use with caution" drugs, manufacturer mentioned advice "to-avoid" drugs)

Covariates occurring in less than 5% of the subjects (percentage will be computed over both exposed and non-exposed cohorts) will not be included in the model.

Model 3: A Cox model will compare the exposed cohort according to the number of *Cervarix* doses (3 classes: 1 dose, 2 doses, 3 doses) vs. the non-exposed cohort, using the same model as for the primary analysis (model 1). In this model, the exposed women with more than 3 doses of *Cervarix* will not be included.

Model 4: A Cox model identical to Model 1 will be performed for any abortion occurring before week 24 of gestation. Any abortion will include spontaneous abortion, therapeutic/induced abortion and other abortion.

5.6.3.2. Logistic model

The following pregnancy outcomes will be analysed using logistic regression model:

- Stillbirth
- Live birth
- Full term delivery
- Pre-term delivery
- Post-term delivery
- Small for gestational age at birth
- Large for gestational age at birth
- Baby's death during the first 12 weeks of life
- Birth defect

Logistic regression models will be computed using the SAS LOGISTIC procedure. The models will included the pregnancy outcome as a binary dependent variable, the exposure status as a binary independent variable and the age at first day of gestation as a continuous covariate. The OR and its 95% Wald CI will be derived. Other covariates will not be included in the model.

In case of less than 10 cases of a specific outcome, the logistic regression analysis will not be carried out on this outcome.

5.7. Conduct of analyses

Any deviation(s) or change(s) from the original statistical plan outlined in this protocol will be described and justified in the final study report.

5.7.1. Sequence of analyses

All the analyses will be done on the final database.

5.7.2. Statistical considerations for interim analyses

There is no interim analysis.

5.8. Changes from planned analyses

Not applicable.

6. SERIOUS ADVERSE EVENTS

This study intends to collect data only on pregnancies and their outcomes. Where required, the results of this study will be communicated to regulators when the final study report becomes available.

7. CONDUCT OF THE STUDY

7.1. Regulatory and ethical considerations, including the informed consent process

The study will be conducted in accordance with all applicable regulatory requirements.

The study will be conducted in accordance with the Guidelines for Good Pharmacoepidemiology Practices (GPP) [ISPE, 2007] all applicable subject privacy requirements and the guiding principles of the Declaration of Helsinki.

Conduct of the study includes, but is not limited to, the following: CPRD ***GOLD***'s Independent Scientific Advisory Committee (ISAC) review and favourable opinion/approval of study protocol and any subsequent amendments. Of note, ISAC has approved the study design on 18 January, 2012 (see Appendix C).

No patient informed consent will be obtained. The patient information in the CPRD ***GOLD*** database is fully coded and GSK Biologicals personnel will not be able to make a link between the data and specific individuals.

The CPRD ***GOLD*** has an ethical approval from a Multi-centre Research Ethics Committee (MREC) for purely observational research (i.e. studies that do not include patient involvement, [CPRD ***GOLD***, 2012]).

(Amended 14 March 2013)

7.2. Data privacy

The CPRD ***GOLD*** database is a fully coded, MHRA-approved database with an international reputation in the field of drug safety signal evaluation [Williams, 2012].

GSK has a licence to use this database from CPRD ***GOLD***, in order to perform analyses. GSK has access to an online extract from CPRD ***GOLD*** which is continuously updated. Data will be not identifiable by GSK as the key-codes are maintained by CPRD ***GOLD*** and are not available online and are never shared with external parties. When GSK requests medical chart review to CPRD ***GOLD***, CPRD ***GOLD*** has internal processes to secure the maintenance of confidentiality concerning subject identifiers. Identifiers will never be transferred to GSK.

(Amended 14 March 2013)

7.3. Biological Sample handling and analysis

No samples will be collected in this study.

7.4. Remote Data Entry instructions

Remote Data Entry (RDE), using a validated computer application developed by GSK (En@ble), will be used by the GSK identified reviewer to enter the information obtained from the free text review and final case ascertainment classification.

In all cases, subject initials will not be collected nor transmitted to GSK. Subject data necessary for analysis and reporting will be entered/transmitted into a validated database or data system. Clinical data management will be performed in accordance with applicable GSK standards and data cleaning procedures.

No monitoring will be done. The GSK identified reviewer remains accountable for the data entry.

8. ADMINISTRATIVE MATTERS

To comply with GPP or other applicable guidelines administrative obligations relating to data collection, archiving data, audits, confidentiality and publications must be fulfilled.

8.1. Audits

To ensure compliance with GPP or other applicable guidelines and all applicable regulatory requirements, GSK may conduct a quality assurance audit. Regulatory agencies may also conduct a regulatory inspection of this study. Such audits/inspections can occur at any time during or after completion of the study. If an audit or inspection occurs, the database owner agrees to allow the auditor/inspector direct access to all relevant documents and to allocate his/her time and the time of his/her staff to the auditor/inspector to discuss findings and any relevant issues.

8.2. Posting of information on public registers

Study information from this protocol will be posted on public registers (e.g. GSK Clinical Study Register, clinicaltrials.gov) before the start of analysis as applicable.

8.3. Ownership and publication

8.3.1. Ownership

The source data are the property of the UK Secretary of State. GSK has received the authorisation to use this data for study purposes. All information provided by GSK and data generated as a result of the analysis are property of GSK.

8.3.2. Posting to the clinical trials registers and publication

The results summary will be posted to the GSK Clinical Study Register and other public registers as applicable, in accordance with regulatory and policy mandated timelines. In addition, a manuscript will be submitted to a peer reviewed journal for publication within the policy defined timelines. The manuscript will be co-authored by the CPRD ***GOLD*** Research Group and coordinated by GSK. When manuscript publication in a peer-reviewed journal is not feasible, further study information will be posted to the GSK Clinical Study Register (e.g. write-up).

(Amended 14 March 2013)

9. REFERENCES

Bosch FX and Munoz N. The viral etiology of cervical cancer. *Virus Res*. 2002; 89: 183-190.

Burd EM. Human papillomavirus and cervical cancer. *Clin Microbiol Rev*. 2003; 16: 1-17.

Cancer Research UK - Cervical cancer statistics UK. ***2008.*** http://info.cancerresearchuk.org/cancerstats/types/cervix/

Centers for Disease Control and Prevention (CDC). Metropolitan Atlanta Congenital Defects Program – 2004 Annual report. Available on: http://www.cdc.gov/ncbddd/birthdefects/MACDP.html

Centers for Disease Control and Prevention (CDC). Metropolitan Atlanta Congenital Defects Program. 40th Anniversary Edition Surveillance Report. Reporting Birth Defects Surveillance Data 1968-2003 (2007). Available on: http://www.cdc.gov/ncbddd/birthdefects/MACDP.html

Centers for Disease Control and Prevention (CDC). Epidemiology and Prevention of Vaccine-Preventable Diseases. Atkinson W, Wolfe S, Hamborsky J, McIntyre L, eds. 11th ed. Washington DC: Public Health Foundation, 2009a.

Centers for Disease Control and Prevention (CDC). Advisory Committee on Immunization Practices (ACIP). http://www.cdc.gov/vaccines/recs/provisional/downloads/hpv-vac-dec2009-508.pdf. Accessed: 01 December 2009b.

Centers for Disease Control and Prevention (CDC). Advisory Committee on Immunization Practices (ACIP). Recommendations on the Use of Quadrivalent Human Papillomavirus Vaccine in Males, 2011 http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6050a3.htm. Accessed: February 2012

Cervarix (Human Papillomavirus vaccine [Types 16, 18] (Recombinant, adjuvanted, adsorbed)) EU Product Information. February, 2011(a).

Cervarix (Human Papillomavirus Bivalent (Types 16 and 18) Vaccine, Recombinant) US Product Information. July, 2011(b).

Charlton R, Cunnington M, de Vries C, Weil J. Data Resources for Investigating Drug Exposure during Pregnancy and Associated Outcomes: The General Practice Research Database (GPRD) as an Alternative to Pregnancy Registries. *Drug Safety* 2008; 31: 39-51.

Charlton R, Weil J, Cunnington M, de Vries C. Identifying Major Congenital Malformations in the UK General Practice Research Database (GPRD): A Study Reporting on the Sensitivity and Added Value of Photocopied Medical Records and Free Text in the GPRD. *Drug Saf* 2010; 33: 741-750.

Charlton R,Weil J, Cunnington M, Ray S, de Vries C. Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as Tools for Postmarketing Teratogen Surveillance: Anticonvulsants and the Risk of Major Congenital Malformations *Drug Saf* 2011; 34: 157-171.

Clinical Practice Research Datalink ***General Practitioner OnLine Database*** (CPRD ***GOLD***) Website. ISAC (Independent Scientific Advisory Committee) – Submission of CPRD ***GOLD*** Protocols to ISAC. http://www.cprd.com/ISAC/otherinfo.asp. Accessed: 14 August 2012.

Cogliano V, Baan R, Straif K *et al*. Carcinogenicity of human papillomaviruses. *Lancet Oncol*. 2005; 6: 204.

Crawford C, White J. Health Protection Agency survey of Primary Care Trust teenage vaccination programmes: December 2009. Immunization, Hepatitis and Blood Safety Department HPA Centre for Infections

de Sanjose S, Quint WG, Alemany L *et al* and HPV Time Trends Study Group. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol.* 2010; 11: 1048-56.

Devine S, West S, Andrews E, Tennis P, Hammad TA, Eaton S, Thorp J, Olshan A. The identification of pregnancies within the general practice research database. *Pharmacoepidemiology and drug safety* 2010; 19: 45-50.

Food and Drug Administration (FDA). Vaccines, Blood & Biologics. Vaccines. Approved Products. CERVARIX Approval Letter. http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm186959.htm. Accessed: 16 October 2009.

GlaxoSmithKline Biologicals (GSK Biologicals) Clinical Report 580299/008. A phase III, double-blind, randomized, controlled, multi-center study to evaluate the efficacy of GlaxoSmithKline Biologicals' HPV-16/18 VLP/AS04 vaccine compared to hepatitis A vaccine as control in prevention of persistent HPV-16 or HPV-18 cervical infection and cervical neoplasia, administered intramuscularly according to a 0, 1, 6 month schedule in healthy females 15-25 years of age. Data on file

Hardy JR, Leaderer BP, Holford TR, Hall GC, Bracken MB. Safety of medications prescribed before and during early pregnancy in a cohort of 81,975 mothers from the UK General Practice Research Database. *Pharmacoepidemiology and drug safety* 2006; 15: 555-564.

Health Protection Agency (HPA; Immunization section). Annual HPV vaccine coverage in England in 2010/11. 22 March 2012. ***Available at: http://immunisation.dh.gov.uk/annual-hpv-vaccine-coverage-in-england-in-201011-report/***

International Society for Pharmacoepidemiology (ISPE). Guidelines for Good Pharmacoepidemiology Practices. GPP. ***Revision 2, 2007***. http://www.pharmacoepi.org/resources/guidelines_08027.cfm. Accessed: 13 September 2012.

Medicines and Healthcare products Regulatory Agency (MHRA). Drug safety update 2012(a); 6(4):H2. http://www.mhra.gov.uk/home/groups/dsu/documents/publication/con207196.pdf. Accessed: 11 March 2013.

Medicines and Healthcare products Regulatory Agency (MHRA). MHRA Public Assessment Report: Cervarix HPV vaccine: update on UK safety at end of 4 years use in the HPV routine immunisation programme. December 2012(b). http://www.mhra.gov.uk/home/groups/plp/documents/websiteresources/con213228.pdf. Accessed: 11 March 2013.

Moscicki AB, Hills N, Shiboski S *et al*. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. *JAMA*. 2001; 285: 2995-3002.

Munoz N, Bosch FX, Castellsague X *et al*. Against which human papillomavirus types shall we vaccinate and screen? The international perspective. *Int J Cancer*. 2004; 111: 278-285.

Office for National Statistics (ONS). Births in England and Wales, Characteristics of birth tables 2009. http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14408 Accessed: 10 August 2011.

Office for National Statistics (ONS). Births in England and Wales, Birth summary tables 2010. http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14408 Accessed: 10 August 2011.

Pebody RG, McLean E, Zhao H *et al*. Pandemic Influenza A (H1N1) 2009 and mortality in the United Kingdom: risk factors for death, April 2009 to March 2010. *Euro Surveill*. 2010 May 20;15(20). pii: 19571.

Saraiya M, Berg CJ, Shulman H *et al*. Estimates of the annual number of clinically recognized pregnancies in the United States, 1981–1991. *Am J Epidemiol* 1999; 149: 1025–9.

Seamark C. Design or accident? The natural history of teenage pregnancy. *J R Soc Med*. 2001; 94: 282-285.

Sheridan A and White J. Annual HPV vaccine uptake in England: 2008/09: Routine programme for year 8 girls (12-13 years old) and catch-up campaign for year 13 girls (17-18 years old). Health Protection Agency Department of Health. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_111675

Sheridan A and White J. Annual HPV vaccine coverage in England in 2009/2010. Health Protection Agency Department of Health. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_***123795***

Smith JS, Lindsay L, Hoots B *et al*. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer*. 2007; 121: 621-632.

Stanley M. Pathology and epidemiology of HPV infection in females. *Gynecol Oncol.* 2010; 117(2 Suppl): S5-10.

Ventura SJ, Abma JC, Mosher WD *et al*. Estimated pregnancy rates by outcome for the United States, 1990–2004. *National vital statistics reports* 2008; 56: 1-28.

Wacholder S, Chen BE, Wilcox A, Macones G, Gonzalez P, Befano B, Hildesheim A, Rodríguez AC, Solomon D, Herrero R, Schiffman M; CVT group. Risk of miscarriage with bivalent vaccine against human papillomavirus (HPV) types 16 and 18: pooled analysis of two randomised controlled trials. *BMJ.* 2010; 340: c712.

Walboomers JM, Jacobs MV, Manos MM *et al*. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*. 1999; 189: 12-19.

Williams T, van Staa TP, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Therapeutic Advances in Drug Safety* 2012; 3(2): 89-99.

Winer RL, Lee SK, Hughes JP *et al*. Genital human papillomavirus infection: incidence and risk factors in a cohort of female university students. *Am J Epidemiol*. 2003; 157: 218-226.

(Amended 14 March 2013)

10. APPENDICES

Appendix A Algorithms and medical codes for data extraction

Appendix B Templates of tables and figures

Appendix C Feedback from ISAC

Appendix D Amendments and administrative changes to the protocol

GlaxoSmithKline Biologicals Clinical Research & Development **Protocol Amendment 1**	
eTrack study number and Abbreviated Title	114101 (EPI-HPV-018 VS UK DB)
Amendment number:	Amendment 1
Amendment date:	14 March 2013
Co-ordinating author:	Co-ordinating authors' names retracted to protect Subject Privacy

Rationale/background for changes:
This protocol has been amended for the following reasons:

- In section 5.2.4.2. Pregnancy outcomes and section 5.2.6. Final case ascertainment step, review of pregnancy outcomes by the GSK safety physician and by experts has been updated.
- In section 5.2.5.1. First day of LMP, the length of pregnancy has been modified to fit the United Kingdom (UK) definition (280 days) and not the French definition (283 days).
- Clinical Practice Research Datalink (CPRD) has been replaced by Clinical Practice Research Datalink General Practitioner OnLine Database (CPRD GOLD).
- Section 1.2. Rationale for the study has been updated with recent published data about *Cervarix* uptake in the United States (US) and about *Cervarix* safety in the UK.
- In section 3. Study design, the number of subjects in the exposed cohort in the epoch table has been corrected.
- Section 3.1. HPV vaccine coverage and pregnancy rate in the UK teenagers has been updated with recent published data about HPV vaccine coverage in the UK.
- In section 5.6.3.1. Cox proportional hazards model, enrolment from CPRD has been replaced by de-enrolment from CPRD GOLD.
- Contributing authors have been updated.

Other changes were made for simplification, clarification or consistency.

Amended text has been included in *bold italics* in the following sections:

Cervarix has been written in italics throughout the synopsis.

Clinical Practice Research Datalink has been replaced by Clinical Practice Research Datalink General Practitioner OnLine Database throughout the document.

CPRD has been replaced by CPRD GOLD throughout the document, including Appendices A and B.

- **Other contributing authors (GSK Biologicals)**

Contributing authors' names retracted to protect Subject Privacy

- **List of abbreviations**

GP: General Practitioner

ISAC: Independent Scientific Advisory Committee ~~from CPRD~~**(*for Medicines and Healthcare products Regulatory Agency database research*)**

- **Glossary of terms**

Independent Scientific Advisory Committee (ISAC): ~~CPRD's~~ Independent Scientific Advisory Committee ***for Medicines and Healthcare products Regulatory Agency database research***.

- **Section 1.2. Rationale for the study**

~~Recently the~~***The*** quadrivalent HPV vaccine has also been recommended for vaccination of males [CDC, 2011].

Moreover, a recent publication by the UK Medicines and Healthcare products Regulatory Agency (MHRA) reviewed the safety profile of Cervarix use in the UK from September 2008 to July 2012. No new safety concerns were identified and the number and nature of adverse drug reaction reports received was as expected after administration of at least 6 million doses of the vaccine in the UK. There was no indication that being vaccinated with Cervarix during pregnancy results in any congenital abnormality or is associated with a risk of spontaneous abortion (or miscarriage) or other adverse pregnancy outcomes [MHRA, 2012a; MHRA, 2012b].

- **Section 1.2. Rationale for the study and synopsis (Rationale for the study)**

In the US, the commercial distribution of *Cervarix* began in October 2009. However, the uptake of *Cervarix* in the US is lower than expected (234,710 doses were shipped in 2010; 153,730 ***doses*** in 2011; ***134,720 doses in 2012; and between 100,000 and 150,000 doses are projected for 2013***). Hence, ~~there is a high risk that~~ the EPI-HPV-020 study will be unable to reach the target sample size during the scheduled recruitment period (Projected completion of the study: Q2 2015).

- **Section 3. Study design and synopsis (Study design)**

Table 1 Study groups and epochs foreseen in the study

Study Groups	Number of subjects	Age (Min/Max)	Epochs
			Epoch 001
exposed cohort	Approximately 240 (-30 to +45 day window) and ~~280~~***380*** (-30 to +90 day window)	15-25 years	x
non-exposed cohort	Approximately 660	15-25 years	x

- **Section 3.1. HPV vaccine coverage and pregnancy rate in the UK teenagers**

In the UK public HPV immunization program (12-13 year olds), HPV vaccination coverage in the UK for 2009 was 88.1%, 86.0% and 80.1% for the first, second and third dose respectively. The coverage in the catch-up program for women (15-18 years old) was 62.2%, 54.2% and 31.8% for the first, second and third dose, respectively [Sheridan, 2009]. ***HPV vaccination coverage in the UK for 2010/11 was 89.0%, 87.6% and 83.8% for the first, second and third dose respectively [Health Protection Agency, 2012].***

- **Section 3.2. Feasibility analysis and limitations of the study design and synopsis (Discussion of study design)**

This is also the design agreed with the FDA for the ~~ongoing~~ study in the US (EPI-HPV-020).

- **Section 5.2.1. The Clinical Pratcice Research Datalink *General Practitioner OnLine Database* (CPRD *GOLD*)**

The CPRD ***GOLD*** is operated on a non-profit basis by the ~~UK Medicines and Healthcare products Regulatory Agency (~~MHRA~~)~~, containing coded longitudinal medical records from general practices and more recently from hospital-based care (e.g., Hospital Episode Statistics, HES).

A total of 4,621,799 patients from ~~646~~ ***546*** practices are currently active in the database.

Data are retrieved by means of the READ classification system; READ codes are a coded thesaurus of clinical terms, which are the basic means by which clinicians record patient findings and procedures in health and social care IT systems across primary and secondary care (e.g. ***general practitioner [***GP***]*** surgeries and pathology reporting of results).

Review of free text will be done by a GSK identified reviewer (Pallas, Health research and consultancy B.V) for completing extracted data; a final case ascertainment step will be performed as described in section ***5.2.4.2,*** 5.2.4.3 and 5.2.6.

- **Section 5.2.4.2. Pregnancy outcomes**

Post-term delivery (~~≥~~> 42 ***completed*** weeks of gestation)

The GSK Safety physician will also review the following pregnancy outcomes: abortions (spontaneous abortions, induced/therapeutic abortions, other abortions), stillbirths, neonatal death, for life births: pre-term and post-term delivery, small for gestational age at birth or large for gestational age at birth.

- **Section 5.2.4.3. Baby (live birth) characteristics and confirmation of birth defects**

All cases with an identified birth defect will be reviewed by a ~~blinded~~ GSK Safety physician using the birth defects classification by the CDC Metropolitan Atlanta Congenital Defect Program or MACDP [Centers for Disease Control and Prevention, 2004 and Centers for Disease Control and Prevention, 2007].

- **Section 5.2.5.1. First day of LMP**

If date of first day of LMP is missing or LMP date is equal to system date (± 2 days) but the EDD is completed in CPRD ***GOLD***, the first day of LMP will be calculated as follows: EDD-~~283~~***280*** days ~~(283 days equal to the median of gestational age: 40 weeks + 3 days)~~ ***(280 days equal to the median of gestational age: 40 weeks)***.

- **Section 5.2.6. Final case ascertainment step**

After the first review by the GSK Safety physician (see section 5.2.4.2), ~~I~~***i***ndividual data related to pregnancy outcomes defined as ***spontaneous and other*** abortions, stillbirths, ***and*** baby's death, ~~pre-term and post-term delivery, small for gestational age at birth or large for gestational age at birth,~~ will be reviewed by two pharmaco-epidemiologists specialised in teratology (see contributors) for final case ascertainment.

- **Section 5.6.3.1. Cox proportional hazards model**

Lost-to-follow up (***de-***enrolment from CPRD ***GOLD***).

- **Section 9. References**

Cancer Research UK - Cervical cancer statistics UK. ***2008.*** http://info.cancerresearchuk.org/cancerstats/types/cervix/

Health Protection Agency (HPA; Immunization section). Annual HPV vaccine coverage in England in 2010/11. 22 March 2012. ***Available at: http://immunisation.dh.gov.uk/annual-hpv-vaccine-coverage-in-england-in-201011-report/***

International Society for Pharmacoepidemiology (ISPE). Guidelines for Good Pharmacoepidemiology Practices. GPP. ***Revision 2, 2007***. http://www.pharmacoepi.org/resources/guidelines_08027.cfm. Accessed: 13 September 2012.

Medicines and Healthcare products Regulatory Agency (MHRA). Drug safety update 2012(a); 6(4):H2. http://www.mhra.gov.uk/home/groups/dsu/documents/publication/con207196.pdf. Accessed: 11 March 2013

Medicines and Healthcare products Regulatory Agency (MHRA). MHRA Public Assessment Report: Cervarix HPV vaccine: update on UK safety at end of 4 years use in the HPV routine immunisation programme. December 2012(b). http://www.mhra.gov.uk/home/groups/plp/documents/websiteresources/con213228.pdf. Accessed: 11 March 2013

Sheridan A and White J. Annual HPV vaccine coverage in England in 2009/2010. Health Protection Agency Department of Health. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_***123795*** ~~118052~~

- **Appendix A, Section 2, Variables derived from CPRD *GOLD***

32. First day of LMP:

First day of LMP=EDD-283*****

****283 was replaced by 280 during the review of the pregnant women profiles by Pallas to fit the UK definition of the length of pregnancy (40 weeks).***

In Clinical or Referral file where medcode equals to 8879 (=Estimated date of delivery) and First day of LMP=EDD-283*****

****283 was replaced by 280 during the review of the pregnant women profiles by Pallas to fit the UK definition of the length of pregnancy (40 weeks)***

33. First day of gestation:

Day of Gestation= EDD-283***** OR First Day of Gestation calculated from data_field_2/3 (=estimated size in weeks + Unit). The EDD data field will be also retrieved. ****283 was replaced by 280 during the review of the pregnant women profiles by Pallas to fit the UK definition of the length of pregnancy (40 weeks)***

34. Date of delivery: If no information available, the delivery date ~~form~~***from*** CPRD ***GOLD*** MBL will be used or in *Patient file* mob/yob under baby identifier OR Entity type =115

44. Date of HPV Cervarix vaccination: If at least one dose of Cervarix is between 01Sep2008 and ~~31DEC2010~~ ***30JUN2011***, the subject will be included in the eligible population.

- **Appendix B, Table 16 Cox Proportion Hazard analysis of spontaneous abortions: Sensitivity analysis (adjusted for other covariates)**

Other sensitivity analysis ~~with~~ ***will*** be reported in the same table layout