NOTE TO THE EDITOR

THIS IS THE PROTOCOL OF EPI-HPV-040 VS UK (ETRACK: 116239)WHICH HAS BEEN PROVIDED BY AUTHORS AS A PART OF POSTING FULL PROTOCOL TO THE EMA PASS REGISTRY. SOME PARTS WITH PERSONALLY IDENTIFIABLE INFORMATION ARE RETRACTED FROM THE PROTOCOL AND THE PAGES WHERE RETRACTION HAS BEEN DONE IS MENTIONED IN PAGE 3

DATE: AUGUST 30, 2013 – HB

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

1. PASS INFORMATION

Title	An observational cohort study to assess the risk of autoimmune diseases in adolescent and young adult women aged 9 to 25 years exposed to Cervarix® in the United Kingdom
Protocol version identifier	116239 (EPI-HPV-040 VS UK)
Date of last version of the protocol:	FDA - EMA PASS Final Version 1: 09 July 2013
EU PAS Register No:	NA (Not applicable)
Active substance	Composition of active substance retracted to protect proprietary information
Medicinal product(s):	Cervarix®, Human Papillomavirus vaccine (Types 16, 18)
Product reference:	EU/1/07/419
Procedure number:	NA
Marketing Authorisation Holder	GlaxoSmithKline Biologicals Rue de l'Institut 89 1330 Rixensart, Belgium
Joint PASS	No
Research question and objectives	To assess the risk of neuroinflammatory/ophthalmic autoimmune diseases and other pre-specified autoimmune diseases within 12 months following the administration of the first dose of Cervarix
Country(-ies) of study	United Kingdom

Authors	Coordinating author: Coordinating authors' names retracted to protect Subject Privacy GSK contributors: Contributing authors' names retracted to protect Subject Privacy Other contributors: Contributing authors' names retracted to protect Subject Privacy

2. MARKETING AUTHORISATION HOLDER

Marketing authorisation holder(s)	GlaxoSmithKline Biologicals Rue de l'Institut 89, 1330 Rixensart, Belgium

NOTE TO THE EDITOR:

The following minor corrections (highlighted in the text) were made to the last version (30 August 2013) of this protocol: Composition of active substance retracted on page 2, co-ordinating authors' names are retracted on page 3, contributing authors' names retracted on page 3, name of the Epidemiologist retracted on page 10, sponsor signatory name retracted on page 62

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ANNEXES

PAGE

LIST OF ABBREVIATIONS

AI	Autoimmune
CBER	Centre for Biologics Evaluation and Research (US FDA)
CI	Confidence interval
CIN	Cervical Intraepithelial Neoplasia
CPRD GOLD	Clinical Practice Research Datalink General Practitioner OnLine database
CRO	Contract Research Organisation
DH	Department of Health (UK)
FDA	Food and Drug Administration (US)
GBS	Guillain-Barré Syndrome
GP	General Practitioner
GPP	Good Pharmacoepidemiology Practices (Guidelines)
GSK	GlaxoSmithKline
HES	Hospital Episode Statistics
HIRD	HealthCare Integrated Research Database
HPV	Human papillomavirus
ICD	International Classification of Diseases
IRR	Incidence rate ratio
ISAC	Independent Scientific Advisory Committee (for Medicines and Healthcare products Regulatory Agency database research)
ISPE	International Society for Pharmacoepidemiology
ITP	Idiopathic thrombocytopenic purpura
JRA	Juvenile rheumatoid arthritis
MHRA	Medicines and Healthcare products Regulatory Agency
MMR	Measles, mumps and rubella

NOAD	New onset of autoimmune disease(s)
PASS	Post Authorization Safety Study
PMS	Post-marketing surveillance
P-Y	Person-years
RA	Rheumatoid arthritis
RR	Relative Risk
SCCS	Self-control case-series
SERM	Safety Evaluation and Risk Management
SLE	Systemic Lupus Erythematous
TSS	Targeted safety study
UK	United Kingdom
US	United States (of America)
VCSP	Vaccines Clinical Safety & Pharmacovigilance

3. RESPONSIBLE PARTIES

GSK Biologicals has the overall responsibility for the conduct of the study.

Name retracted to protect Subject Privacy is the GSK Biologicals designated Head of Global Epidemiology and Lead Epidemiologist for this study.

4. ABSTRACT

Title	An observational cohort study to assess the risk of autoimmune diseases in adolescent and young adult women aged 9 to 25 years exposed to Cervarix® in the United Kingdom.
Rationale and background	Cervarix is GlaxoSmithKline (GSK) Biologicals' bivalent recombinant vaccine against human papillomavirus (HPV, types 16 and 18). To address a regulatory commitment made in 2009 to the US FDA, GSK initiated an observational cohort study in the USA (e-track: 113522, EPI-HPV-015) to assess the risk of new onset of autoimmune disease(s) (NOAD) within 12 months following the administration of at least one dose of Cervarix (exposed) versus a non-Cervarix vaccinated cohort (unexposed). Because of the current low Cervarix uptake in the USA which is anticipated to stay at a low level over the next few years, it will take significantly longer than the 3 years planned to complete accrual. The present protocol is submitted as an alternative epidemiological study using the Clinical Practice Research Datalink General Practitioner OnLine database (CPRD GOLD) in the UK to fulfil the post-marketing commitment. The UK has had sufficient Cervarix vaccination coverage to, in theory, enable data acquisition.
Research question and objectives	**Primary** • To assess the risk of neuroinflammatory/ophthalmic new onset of autoimmune disease(s) (NOAD) and other pre-specified NOAD within 12 months following the administration of the first dose of Cervarix **Secondary** • To describe individually the incidence of the pre-specified NOAD considering different time periods following the administration of the first dose of Cervarix: – Incidence of Guillain Barré syndrome (including Miller Fisher syndrome and other variants), and autoimmune haemolytic anaemia within two months following the administration of the first dose of Cervarix; – Incidence of idiopathic thrombocytopenic purpura (ITP) within six months following the administration of the first dose of Cervarix; – Incidence of multiple sclerosis, transverse myelitis, optic neuritis, other demyelinating diseases,

auto-immune uveitis, systemic lupus erythematous (SLE), rheumatoid arthritis (RA), juvenile rheumatoid arthritis (JRA), Still's disease, psoriatic arthritis, ankylosing spondylitis, type 1 diabetes mellitus, auto-immune thyroiditis (including Hashimoto's disease, Graves'/Basedows' disease), and inflammatory bowel / hepatic disease (Crohn's disease, ulcerative colitis and autoimmune hepatitis) within one year following the administration of the first dose of Cervarix.

Study design

- This is an observational cohort study using the CPRD GOLD data source in the UK.

  Four cohorts will be defined based on exposure to Cervarix and sex as recorded in the CPRD GOLD data source:

  – Cervarix vaccinated (exposed) female cohort

  – Unexposed historical female cohort

  – Unexposed concurrent male cohort

  – Unexposed historical male cohort

  – Study population:

    – Female population is composed of female subjects vaccinated with Cervarix between the ages of 9 to 25 years and unexposed female subjects of the same age, identified from historical data.

    – Male population is composed of 9- to 25-year-old male subjects not vaccinated with Cervarix.

  Female subjects included in the exposed cohort will have received at least one dose of GSK's vaccine Cervarix administered according to local practice.

  Female subjects in the unexposed historical cohort will be frequency matched for age and practice region identifier to the subjects included in the vaccinated (exposed) cohort.

  Comparison of the unexposed concurrent male cohort with the unexposed historical male cohort will be used as an internal control for changes over time in CPRD GOLD in reporting NOAD. The male subjects will be frequency matched for age and practice region identifier.

  A self-control case-series (SCCS) analysis for confirmed NOAD in the exposed female cohort will also be conducted, using a risk period of one year after the first Cervarix dose, a control period of one year and a

six-month buffer period between risk and control periods.

Data will be extracted from CPRD GOLD and will be validated using relevant free text for the date of onset and symptoms of the identified autoimmune diseases, and full hospital discharge statistics (HES, when available) abstracted by an independent entity (according to the CPRD GOLD process).

Population

Abstract Table 1 Study groups foreseen in the study (using CPRD GOLD, UK)

Study Groups	Number of subjects	Age*
Exposed cohort	65,000	9-25 years
Non-exposed cohorts:		
Historical female cohort	65,000	9-25 years
Concurrent male cohort	65,000	9-25 years
Historical male cohort	65,000	9-25 years

* In the European Union, Cervarix is indicated for use from the age of 9 years onwards whereas in the USA, Cervarix is indicated for use in females 9 through 25 years of age.

Variables

Primary Endpoint

- Occurrence of new onset of confirmed[1] autoimmune disease during the period of one year following administration of the first dose of Cervarix (risk period) among an exposed cohort and during an equivalent time period in the unexposed cohorts for the following two co-primary composite endpoints:

  [1] **Neuroinflammatory/ophthalmic autoimmune diseases:**

  - Multiple Sclerosis
  - Transverse myelitis
  - Optic neuritis
  - Guillain-Barré syndrome, including Miller Fisher syndrome and other variants
  - Other demyelinating diseases:
    - Acute disseminated encephalomyelitis, including site specific variants: e.g. non-infectious encephalitis, encephalomyelitis, myelitis, myeloradiculomyelitis
    - AI peripheral neuropathies and plexopathies (including chronic inflammatory demyelinating

[1] Autoimmune disease diagnosis ascertainment by an expert physician panel.

polyneuropathy, multifocal motor neuropathy and polyneuropathies associated with monoclonalgammopathy).

- Auto-immune uveitis

[2] **Other autoimmune diseases:**

- Systemic lupus erythematous
- Autoimmune (AI) disease with rheumatologic conditions:
  - Rheumatoid arthritis (RA)
  - Juvenile rheumatoid arthritis (JRA)
  - Still's disease
  - Psoriatic arthritis
  - Ankylosing Spondylitis
- AI haematological conditions:
  - Idiopathic thrombocytopenic purpura (ITP)
  - AI haemolytic anaemia
- AI endocrine conditions:
  - Type 1 diabetes mellitus
  - AI thyroiditis including Hashimoto's disease, Graves'/Basedows' disease
- Inflammatory bowel / hepatic diseases:
  - Crohn's diseases
  - Ulcerative colitis
  - Autoimmune hepatitis

Secondary Endpoint

Secondary endpoint is the occurrence of new onset of individual confirmed autoimmune disease during the following specific periods:

- Occurrence of Guillain Barré syndrome (including Miller Fisher syndrome and other variants), and autoimmune haemolytic anaemia within two months following the administration of the first dose of Cervarix;
- Occurrence of idiopathic thrombocytopenic purpura (ITP) within six months following the administration of the first dose of Cervarix;
- Occurrence of multiple sclerosis, transverse myelitis, optic neuritis, other demyelinating diseases, auto-immune uveitis, systemic lupus erythematous, rheumatoid arthritis (RA), juvenile rheumatoid arthritis (JRA), Still's disease, psoriatic arthritis, ankylosing spondylitis, type 1 diabetes mellitus, auto-immune thyroiditis (including Hashimoto's

	disease, Graves'/Basedows' disease), and inflammatory bowel / hepatic disease (Crohn's disease, ulcerative colitis and autoimmune hepatitis) within one year following the administration of the first dose of Cervarix.
Data sources	The CPRD GOLD is the world's largest computerised database of linked anonymised longitudinal medical records from primary care.
Study size	Please refer to Abstract Table 1 in the Population row above.
Data Analysis	Incidence rates for NOAD will be calculated as the number of cases divided by person-time. A Poisson regression model will estimate the exposed/unexposed risk ratio and its 95% confidence interval. The Poisson model will include the number of cases in each cohort as the dependent variable, the exposure status as a binary independent variable and the log-transformed total person-year as an offset. The same statistical model will be used to compare the two cohort. The cases of NOAD in exposed subjects will be analysed using Self-control case-series (SCCS) methods.
Milestones	Provisional milestones for the study which depend on timely approval of the study in by end of Q3 2013, are provided in the cover letter to the Regulatory Authorities.

5. AMENDMENTS AND UPDATES

None

6. MILESTONES

Provisional milestones for the study which depend on timely approval of the study by end of Q3 2013, are provided in the cover letter to the Regulatory Authorities.

Milestone	Planned date
Final protocol submitted to Regulatory Authorities	31 July 2013
Start of data collection	30 September 2013
End of data collection	31 May 2014
Planned analyses completed	30 June 2014
Projected study completion	30 September 2014
Final report of study results	31 March 2015

7. RATIONALE AND BACKGROUND

Cervarix is a GlaxoSmithKline (GSK) Biologicals' bivalent recombinant vaccine against human papillomavirus (HPV, types 16 and 18). It is currently licensed in more than 120 countries worldwide, including the European Union (EU) via the Centralised Procedure. Cervarix was granted approval by the European Medicines Agency (EMA) in September 2007 and the US Centre for Biologics Evaluation and Research (CBER) in October 2009. In the US, Cervarix is indicated for the prevention of cervical cancer, cervical intraepithelial neoplasia (CIN) grade 2 or worse and adenocarcinoma *in situ*, and cervical intraepithelial neoplasia (CIN) grade 1, caused by oncogenic human papillomavirus (HPV) types 16 and 18, in females 9 through 25 years of age. In the EU, Cervarix is indicated for use from the age of 9 years for the prevention of premalignant cervical lesions and cervical cancer causally related to certain oncogenic Human Papillomavirus (HPV) types.

Pre-licensure clinical studies provide key vaccine safety data, but their power to detect rare outcomes such as new onset of auto-immune disease(s) (NOAD) is limited due to their sample size, since incidence rates of different NOAD vary roughly from 1 to 20/100,000 per year [Cooper, 2003]. A pooled analysis of NOAD data from 68,000 subjects exposed to AS04-adjuvanted HPV-16/18, herpes simplex virus and hepatitis B vaccines in the GSK development programs did not suggest any excess risk associated with the AS04-adjuvanted vaccines compared to control vaccines [Verstraeten, 2008]. A pooled safety analysis of data from 57,580 adolescent and adult females aged 9 years and above, of whom 33,339 received at least one dose of HPV vaccine, showed the vaccine to be generally well tolerated in women of all ages [GSK confidential document. Prophylactic HPV-16/18 L1 VLP Vaccine Formulated with AS04. Investigator Brochure Edition 11 March 2012; Descamps, 2009].

Analysis of the end-of-study PATRICIA efficacy trial showed the vaccine to be generally well tolerated, which included the recorded incidence of NOAD in a broad range of women, including those of different nationalities and ethnicities [Lehtinen, 2012]. The percentage of subjects experiencing a NOAD as assessed by GSK or the investigators was low and comparable between the two groups (99 (1.1%) in the Cervarix group and 95 (1.0%) in the Hepatitis A (Havrix) group), and no imbalances between groups were observed for any event classified by the MedDRA Preferred Term.

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 [Medicines and Healthcare products Regulatory Agency (MHRA), 2012]. No new safety concerns were identified and the number and nature of ADR reports received was as expected after administration of at least 6 million doses of the vaccine in the UK.

Siegrist et al. [Siegrist, 2007] suggested the use of population-based data which allows identification of issues of potential concern, monitoring of the impact of large-scale interventions and rapid action if any vaccine safety issues occur, that could compromise vaccine programs. The Gardasil vaccine from Merck was approved by the US FDA in 2006 and a post-licensure commitment to the regulatory authorities was established to conduct a safety surveillance study to estimate NOAD. The post-licensure study has been published in 2011 showing the Gardasil vaccine was not associated with any "autoimmune safety signal" in a large Californian database study (Kaiser Permanente Southern California & Kaiser Permanente Northern California managed care organisation databases), and "no pre-specified autoimmune condition examined demonstrated any cluster of disease onset in relation to vaccination timing, dose sequence or age" [Chao, 2011]. In this study involving 189,629 women who received one of more doses of Gardasil between August 2006 and March 2008, the incidence of potential cases of autoimmune diseases after vaccination was investigated within 3 pre-specified categories of NOAD: rheumatologic/autoimmune disorders [2a], endocrine conditions [b], and neurological/ophthalmic conditions [c]. Overall, 1014 potential onset of new cases of the pre-specified autoimmune diseases were electronically identified: 719 were eligible for case review; 31-40% were confirmed as onset of new cases. No cluster of NOAD in relation to vaccination timing, dose sequence or subject age was found. No estimated incidence rate ratios (IRR) were elevated, except for Hashimoto's disease (IRR = 1.29,

---

[2] Autoimmune conditions of interest were pre-specified and composed of three groups:

a) Rheumatologic/autoimmune disorders: immunethrombocytopenia (ITP), autoimmune haemolytic anaemia, systemic lupus erythematous (SLE), rheumatoid arthritis (RA) and juvenile rheumatoid arthritis (JRA)

b) Autoimmune endocrine conditions: type 1 diabetes, Hashimoto's disease and Graves' disease/Basedows' disease

c) Autoimmune neurological/ophthalmic conditions: multiple sclerosis (MS), acute disseminated encephalomyelitis, other demyelinating diseases of the central nervous system, vaccine-associated demyelination, Guillain-Barre´ syndrome, neuromyelitis optica, optic neuritis and uveitis

95% confidence interval: 1.08-1.56) but the authors reported that "there was no consistent elevation in incidence for autoimmune thyroid conditions in the vaccinated cohort [IRR = 0.72 (0.50–1.01) for Graves' disease] and several confirmed new-onset autoimmune thyroid condition cases were likely pre-existing cases at the time of vaccination" [Chao, 2011].

GSK also committed to develop a post-licensure study of NOAD as outlined in the 2009 approval letter for Cervarix in the US:

"To conduct an observational study in a US managed care organization to evaluate the incidence of new onset autoimmune disease among at least 50,000 Cervarix recipients. The final protocol will be submitted by March 2010. Projected completion of patient accrual, subject to vaccine uptake, will be completed by March 2013. Projected study completion, subject to vaccine uptake, will be completed by September 2014. The final study report is projected to be submitted by March 2015 (6 months after study completion)".

In order to address this regulatory commitment, GSK initiated an observational cohort study to assess the risk of NOAD within 12 months following the administration of at least one dose of Cervarix (exposed) versus a non-Cervarix vaccinated cohort (unexposed). This study (e-track: 113522, EPI-HPV-015) planned to include 140,000 females, aged 9 to 25 years, enrolled in US health plans. Based on the low incidence of NOAD in this age-group, composite endpoints were defined and agreed with the FDA. Its primary objective is to evaluate whether there is an increased incidence of neuroinflammatory NOAD or other NOAD within 12 months following the administration of at least one dose of Cervarix. Data are retrieved from a large insurance administrative claims database (HealthCore Integrated Research Database (HIRD), HealthCore Inc.,WellPoint, New York, US) which includes data from 14 health plans distributed throughout the US, representing claims information from the largest commercially insured population in the US.

In the US, the commercial distribution of Cervarix began in November 2009. However, the uptake of Cervarix in the US is currently lower than initially expected (51,000 doses were distributed in 2009; 234,710 doses were distributed in 2010; 153,730 doses were distributed for 2011; 134,720 doses were distributed in 2012; and between 100,000 to 150,000 doses are projected for 2013). During the time period of 16 October 2009 to 30 April 2012, 851 females in the Cervarix® exposed and 851 females in the unexposed cohorts were accrued, who received a total of 1,516 cumulative doses of Cervarix®. This is 1.2% of the target number of 70,000 females in the Cervarix® exposed and unexposed cohorts, respectively, and 1.1% of the target of 135,000 cumulative Cervarix® doses. At this rate of accrual, it will take significantly longer than the 3 years planned to complete accrual in this study.

The present protocol is submitted as an alternative epidemiological study using the Clinical Practice Research Datalink General Practitioner OnLine database (CPRD GOLD) in the UK to fulfil the post-marketing commitment.

7.1. Autoimmune disease(s) incidence rates for UK and USA

The background incidence rates of auto-immune diseases in the UK and USA for male and female subjects were derived from published literature and have been tabulated by GSK Biologicals in Table 1, showing no difference in magnitude between the two countries for the age range from 9 to 25 years and for events for which data are available in both countries.

Table 1 Background incidence rates of NOAD in the UK and US*

Disease	Age (years)	Incidence rate UK (per 100,000/year)		Age (years)	Incidence rate US (per 100,000/year)		References
		Males	Females		Males	Females	
Diabetes mellitus type 1	0-14	15.4-26.8	15.3-25.9	0-14	14.1-19.1	15.1-16.4	UK: Diamond Project Group, 2006**;Harron, 2011; Imkampe, 2011 US: Diamond Project Group, 2006**
	10-19	35.0	26.0	10-17	6.7-33.1	6-28.2	UK: Gonzalez, 2009; US: Kostraba, 1992; Lipton, 1995; MacDonald, 1989; Allen, 1986; NCKP[$$]
	15-34	20.0	10.1	18-25	10.0-15.2	7.9-19.2	UK: Imkampe, 2011; US: Fishbein, 1982; Allen, 1986; NCKP[$$]
	NA	NA	NA	9-26	NA	18.0	US: Chao, 2011
Inflammatory bowel disease***	10-19	4.2-10.7	3.4-11.0	10-19	2.1-8.7	3.5-9.4	UK: Steed, 2010; Henderson, 2012 US: Herrinton, 2008; Abramson, 2010
	20-29	15.5	21.9	20-29	8.3-15.7	8.1-13.4	UK: Steed, 2010; Henderson, 2012; US: Herrinton, 2008
Multiple sclerosis	15-19	0	0.9	10-17	0	0-2.3	UK: Alonso, 2007[$]; US: Mayr, 2003; NCKP[$$]
	20-24	1.7	5.9	18-25	5.3	7.5-8.6	UK: Alonso, 2007[$]; US: Mayr, 2003; NCKP[$$]
	NA	NA	NA	9-26	NA	2.5	US: Chao, 2011
Immune or idiopathic thrombocytopenic purpura (ITP)	0-18	4.7	3.7	9-26	NA	5.9	UK: Yong, 2010; US: Chao, 2011
	6-17	2.1-2.6	2.7-3.4	10-17	NA	1.5-15	UK: Schoonen, 2009; US: Simpson, 1989; NCKP[$$]
	18-29	0.6-1.6	3.6-4.9	18-25	NA	3.3-15	UK: Abrahamson, 2009; US: Simpson, 1989, NCKP[$$]
Guillain-Barré syndrome	NA	NA	NA	10-17	NA	0.8-1.8	UK: Hughes, 2006 US: Shui, 2012; Beghi, 1996; Koobatian, 1991; Riggs, 1989; NCKP[$$]
	NA	NA	NA	5-17	1.1-1.9	0.8-1.2	
	15-24	0.6	1.1	18-25	1.4-2.2	0.4-2	
Systemic lupus erythematous	10-19	0.1	2.3	10-17	0-0.3	1.5-3.4	UK: Nightingale, 2006; US: Hochberg, 1985; McCarty, 1995; Naleway, 2005
	20-29	0.0	4.7	18-25	1.3-1.7	5.6-19.2	
	NA	NA	NA	9-26	NA	10.3	US: Chao, 2011
Rheumatoid arthritis	NA	NA	NA	9-34	3.6	7.0-13.8	US: Myasoedova, 2010; Chao, 2011
Autoimmune thyroiditis	NA [a]	NA	NA	10-17	NA	19.5	US: NCKP[$$]
	NA	NA	NA	18-25	NA	37.8	
Optic neuritis	NA	NA	NA	9-26	NA	3.9	US: Chao, 2011
Uveitis	NA	NA	NA	9-26	NA	11.9	US: Chao, 2011

NA = not available; ** Study conducted in several parts of the UK and US. UK: the ranges represent data from Scotland, Leicestershire, Northern Ireland, Oxford, Plymouth and Yorkshire. US: The ranges represent data from Allegheny and Jefferson; *** Inflammatory bowel disease includes Crohn's disease, ulcerative colitis and non-Crohn's colitis (ulcerative colitis and unclassified inflammatory bowel disease combined); [$] Used data from the beginning of the GPRD, likely to be incomplete;
[$$] Northern California Kaiser Permanente. Only female data available from the previous GSK review. See EPI-HPV-015 (e-track: 113522)

[a] For all ages – Incidence in male = 22/100,000/year and in female = 99/100,000/year in Scotland, UK [Leese, 2008]

8. RESEARCH QUESTIONS AND OBJECTIVES

8.1. Primary objective

- To assess the risk of neuroinflammatory/ophthalmic new onset of autoimmune disease(s) (NOAD) and other pre-specified NOAD within 12 months following the administration of the first dose of Cervarix:

  [1] **Neuroinflammatory/ophthalmic autoimmune diseases:**

  - Multiple Sclerosis
  - Transverse myelitis
  - Optic neuritis
  - Guillain-Barré syndrome, including Miller Fisher syndrome and other variants
  - Other demyelinating diseases:
    - Acute disseminated encephalomyelitis, including site specific variants: e.g. non-infectious encephalitis, encephalomyelitis, myelitis, myeloradiculomyelitis
    - AI peripheral neuropathies and plexopathies (including chronic inflammatory demyelinating polyneuropathy, multifocal motor neuropathy and polyneuropathies associated with monoclonalgammopathy).
  - Auto-immune uveitis

  [2] **Other autoimmune diseases:**

  - Systemic lupus erythematous
  - Autoimmune (AI) disease with rheumatologic conditions:
    - Rheumatoid arthritis (RA)
    - Juvenile rheumatoid arthritis (JRA)
    - Still's disease
    - Psoriatic arthritis
    - Ankylosing Spondylitis
  - AI haematological conditions:
    - Idiopathic thrombocytopenic purpura (ITP)
    - AI haemolytic anaemia
  - AI endocrine conditions:
    - Type 1 diabetes mellitus
    - AI thyroiditis including Hashimoto's disease, Graves' /Basedows' disease
  - Inflammatory bowel / hepatic diseases:
    - Crohn's diseases
    - Ulcerative colitis
    - Autoimmune hepatitis

8.2. Secondary objectives

- To describe individually the incidence of the pre-specified NOAD considering different time periods following the administration of the first dose of Cervarix:
  - Incidence of Guillain Barré syndrome (including Miller Fisher syndrome and other variants), and autoimmune haemolytic anaemia within two months following the administration of the first dose of Cervarix;
  - Incidence of idiopathic thrombocytopenic purpura (ITP) within six months following the administration of the first dose of Cervarix;
  - Incidence of multiple sclerosis, transverse myelitis, optic neuritis, other demyelinating diseases [3], auto-immune uveitis, systemic lupus erythematous (SLE), rheumatoid arthritis (RA), juvenile rheumatoid arthritis (JRA), Still's disease, psoriatic arthritis, ankylosing spondylitis, type 1 diabetes mellitus, auto-immune thyroiditis (including Hashimoto's disease, Graves'/Basedows' disease), and inflammatory bowel / hepatic disease (Crohn's disease, ulcerative colitis and autoimmune hepatitis) within one year following the administration of the first dose of Cervarix.

Refer to Section 9.3 for the definition of the primary and secondary endpoints and the pre-specified list of NOAD.

8.3. Exploratory objective

- To evaluate if temporal clustering of the individual NOADs comprising the primary endpoint and the secondary endpoints (i.e., those NOADs on the pre-defined list) occurs following the administration of at least one dose of Cervarix® within the 12-month follow-up period.

9. RESEARCH METHODS

9.1. Study Design

9.1.1. Overview

- This is an observational cohort study using the CPRD GOLD data source in the UK.
- Four cohorts will be defined based on exposure to Cervarix and sex as recorded in the CPRD GOLD data source:

[3] Other demyelinating diseases:
- Acute disseminated encephalomyelitis, including site specific variants: e.g. non-infectious encephalitis, encephalomyelitis, myelitis, myeloradiculomyelitis
- AI peripheral neuropathies and plexopathies (including chronic inflammatory demyelinating polyneuropathy, multifocal motor neuropathy and polyneuropathies associated with monoclonalgammopathy).

1. Cervarix vaccinated (exposed) female cohort
2. Unexposed historical female cohort
3. Unexposed concurrent male cohort
4. Unexposed historical male cohort

- Study population:
  - Female population is composed of female subjects vaccinated with Cervarix between the ages of 9 to 25 years and unexposed female subjects identified from historical data.
  - Male population is composed of 9- to 25-year-old male subjects not vaccinated with Cervarix.

  Female subjects included in the exposed cohort will have received at least one dose of GSK's vaccine Cervarix administered according to local practice.

  Female subjects in the unexposed historical cohort will be frequency matched for age and practice region identifier to the subjects included in the vaccinated (exposed) cohort.

  Study design:

  Comparison of the unexposed concurrent male cohort with the unexposed historical male cohort will be used as an internal control for changes over time in CPRD GOLD in reporting NOAD. The male subjects will be frequency matched for age and practice region identifier as described in Section 9.2.

  A self-control case-series (SCCS) analysis for confirmed NOAD in the exposed female cohort will also be conducted, using a risk period of one year after the first Cervarix dose, a control period of one year and a six-month buffer period between risk and control periods.

9.1.2. Rationale for study design

NOAD represent a heterogeneous group of diseases with different clinical conditions and disease progression. Some NOAD present with a chronic disease pattern of relapse over time e.g. multiple sclerosis or systemic lupus erythematous, or an acute disease pattern (e.g. Guillain Barré Syndrome).

In addition to the comparison of the exposed vs. non-exposed cohort (cohort design), the confirmed NOAD in the exposed cohort will also be analysed using a self-control case-series (SCCS) analysis.

In the cohort design, specified NOAD will be collected over a period of one year following the administration of at least one dose of Cervarix in an exposed cohort and over a comparable period in the unexposed cohorts as shown in Figure 1. Four cohorts will be constituted as shown in Figure 1. The unexposed male cohorts will be enrolled in order to assess a possible change over time in the incidence rate of NOAD in CPRD GOLD independent of Cervarix introduction. The cohorts will be frequency

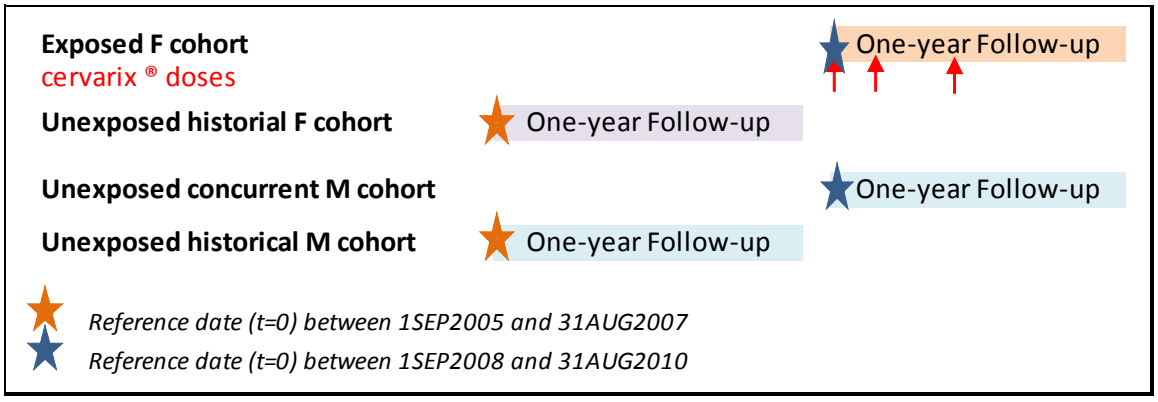
matched for the age (age class of one year) and practice region identifier at reference date (age at first dose of Cervarix).

An unexposed concurrent female cohort has been not chosen as the unvaccinated group for the following reasons:

- There is a under-reporting of HPV vaccination in CPRD GOLD as evidenced by the difference in vaccine coverage in CPRD GOLD population and the UK population (see Table 2), an exposed concurrent cohort would include both true and false unexposed subjects;
- Moreover, because of the large vaccine coverage in the UK, non-vaccinated women could have different health care behaviour compared to the vaccinated women, resulting in a difference in the probability to detect auto-immune disease.

Therefore, the reference date (time = 0) for the vaccinated (exposed) cohort will be the date of the first dose of Cervarix recorded in CPRD GOLD. The reference date for the unexposed (unvaccinated) cohorts will be a date randomly selected among the reference dates of the exposed subjects and minus 3 years for the historical cohorts.

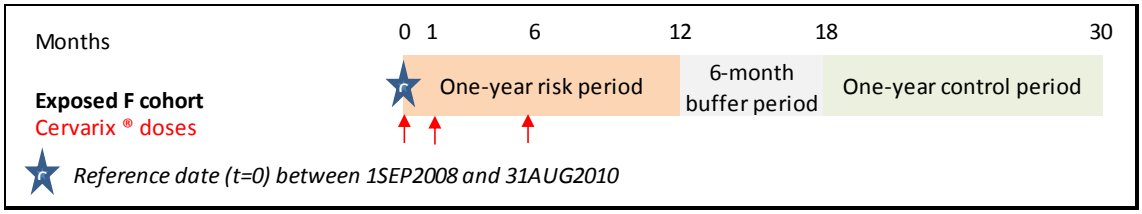
Figure 1 Cohort design



Note: all three doses may not have been administered

For the SCCS analysis, the exposed female cohort will be followed during a total period of 30 months from the first dose. The risk period will be defined as one year after the first dose, including the six months after the last dose when the full three-dose vaccination course is administered. A control period of the same duration will be defined, excluding the periods after subsequent doses. To control for a possible late effect of the vaccine, the control and the risks period will be separated by a six-month buffer period (Figure 2). For analysis of individual diseases, specific risk period could be defined.

Figure 2 Self-control case-series analysis



Note: all three doses may not have been administered

The SCCS uses each subject as its own control, preventing virtually all potential confounding factors which do not vary with time (e.g. socio-economic status, gender, location) [Whitaker, 2006]. Additionally, fewer cases are usually required, as compared to a case-control design. The control period is chosen subsequent to the at-risk period. Age bias is not anticipated because the follow-up period is short compared to the age effect on incidence of autoimmune diseases (see Table 1).

9.1.3. HPV vaccine coverage in UK and in CPRD GOLD

The submitted study is in the UK, which is a country with a high Cervarix vaccine coverage. The recommended schedule in the UK for HPV vaccination of 12-13 year old girls involves three doses of Cervarix given over at least a 6-month period [DH, 2008; HPA, 2010].

In the UK public HPV immunization program (12-13 year olds), 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].

Vaccination records in the CPRD GOLD cover only part of the Cervarix-vaccinated population in the UK. For instance, 70.7% of the 01-SEP-1996 – 31-AUG-1997 birth cohort were reported as HPV vaccinated in CPRD GOLD versus 85.9% reported by the UK Health Protection Agency. Corresponding numbers for the 01-SEP-1994 – 31-AUG-1995 birth cohort are 70.6% versus. 81.9% (Table 2).

Table 2 HPV vaccination coverage in CPRD GOLD versus HPA-DH UK data

	CPRD†			UK‡
Birth cohort	N vaccinated	N total	%	%
01 SEP 1996 and 31 AUG 1997	16028	22685	70.7%	85.9%
01 SEP 1994 and 31 AUG 1995	16343	23143	70.6%	81.9%

† Data extracted from CPRD GOLD version ffgprd_smart_2012Q3
‡ From Health Protection Agency Department of Health [Health Protection Agency, 2012]

A first exploration of the CPRD GOLD database identified 148,731 subjects vaccinated with Cervarix between 2007 and 2010. The number of vaccinated women in the relevant age-range included in the CPRD GOLD is provided in Table 3. The number of HPV-vaccinated females in the CPRD GOLD appears to be sufficient for studying adverse events which have a low incidence after vaccination with Cervarix.

Table 3 Number of HPV-vaccinated females in the relevant age range included in the CPRD GOLD

Age range [years]	Number of subjects
[9-10]	16
[11-15]	82326
[16-20]	65080
[21-25]	866

Created by GSK Biologicals', 2012
Date extracted from CPRD GOLD version ffgprd_smart_2012Q3

9.2. Setting

Setting and study population

9.2.1. The UK HPV National Immunization Programme

The submitted study is in the UK, which is a country with a high Cervarix vaccine coverage. The recommended schedule in the UK for routine HPV vaccination of all girls at 12 to 13 years of age is [DH, 2008; HPA, 2010]:

- Vaccination at this age starts with a first dose of 0.5ml of Cervarix HPV vaccine
- The second dose of 0.5ml follows one to two months after the first dose
- A third dose of 0.5ml follows at least six months after the first dose

The UK has had sufficient Cervarix vaccination coverage to, in theory, enable data acquisition. A public immunisation 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]. In the UK public HPV immunization program (12-13 year olds), 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 recommended age range for the UK programme matches the age range required by the FDA (9-25 years of age) for the post-licensure safety study.

The study population will be composed of female and male subjects, 9 to 25 years of age, registered in the CPRD GOLD.

The exposed cohort will be composed of female subjects vaccinated with at least one dose of Cervarix, with or without other recommended vaccines.

The unexposed historical female cohort will consist of frequency age-matched and practice region-matched female subjects from the period before the introduction of Cervarix. The unexposed concurrent male cohort will consist of frequency age-matched and practice region-matched by one-year classes (15, 16, 17, etc.) male subjects from the

period after the introduction of Cervarix. The unexposed historical male cohort will consist of frequency age-matched and practice region-matched male subjects from the period before the introduction of Cervarix.

9.2.2. Cohort identification and creation

The exposed eligible cohort have been identified based on the stepwise approach defined in Annex 5. Among eligible exposed subjects, 65,000 subjects will be randomly selected using the RANUNI function of SAS. The RANUNI function returns a number that is generated from the uniform distribution on the interval (0, 1). The corresponding subject number will be computed as *random_subj_number=ranuni(seed)*. The subjects will be ordered according the *random_subj_number* and the first 65,000 subjects will be included in the exposed cohort.

The unexposed eligible cohorts have been identified based on the stepwise approach define in Annex 5.

All the unexposed subjects who matched exposed subject for age (birth cohort) and region (frequency matching) will be identified. This represents 234 combinations of birth cohort-region (18 birth cohorts and 13 regions in CPRD GOLD). In each combination, the subjects will be randomly selected based on the distribution in the exposed cohort.

A reference date is randomly attributed to all potential unexposed subjects. For the concurrent male cohort, the reference date is a random date between 01-SEP-2008 and 31-AUG-2010. For the historical cohorts, the reference date is a random date between 01-SEP-2005 and 31-AUG-2007 (reference dates for exposed cohort – 3 years).

The random reference dates in each ‘birth cohort-region” combination will be attributed randomly using the RANUNI function of SAS.

The age at reference date will be calculated for the unexposed cohort. After applying the exclusion criteria, the eligible subjects (65,000 subjects in each cohort) will be randomly selected in each of the combinations ‘birth cohort-region’ taking into account the distribution in the exposed cohort. The random selection will also use the RANUNI function of SAS.

9.2.3. Number of subjects

For the cohort design, the target sample size is 65,000 subjects for each cohort. Refer to Section 9.5 for a detailed description of the estimation of the sample size.

9.2.4. Inclusion criteria

Note: Other vaccines are allowed in this study regardless of the time of administration and the time interval between subsequent doses.

9.2.4.1. Inclusion criteria for the exposed female cohort

Exposed females must satisfy ALL the following criteria at study entry:

- Female aged from 9 to 25 years at the reference date (01 September 2008 through 31 August 2010)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- The first dose of Cervarix received between 01 September 2008 through 31 August 2010, Full date (day/month/year) of Cervarix vaccination(s) available
- Subject defined as acceptable in CPRD GOLD

9.2.4.2. Inclusion criteria for the unexposed historical female cohort

Unexposed females must satisfy ALL the following criteria at study entry:

- Female aged 9 to 25 years at the reference date (01 September 2005 through 31 August 2007)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- Subject defined as acceptable in CPRD GOLD

9.2.4.3. Inclusion criteria for the unexposed concurrent male cohort

Unexposed concurrent males must satisfy ALL the following criteria at study entry:

- Male aged 9 to 25 years at the reference date (01 September 2008 through 31 August 2010)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- Subject defined as acceptable in CPRD GOLD

9.2.4.4. Inclusion criteria for the unexposed historical male cohort

Unexposed historical males must satisfy ALL the following criteria at study entry:

- Male aged 9 to 25 years at the reference date (01 September 2005 through 31 August 2007)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- Subject defined as acceptable in CPRD GOLD

9.2.5. Exclusion criteria

9.2.5.1. Exclusion criteria for all cohorts

- Subjects with a diagnostic code of any auto-immune disease during the year prior to the reference date.
- Subjects who received at least one dose of unspecified HPV vaccine or Gardasil at any time before the reference date.
- Subjects who have been included in the other cohort.

9.2.5.2. Exclusion criteria for the non-exposed cohorts

- Subjects who received any dose of Cervarix at any time before the reference date.

9.3. Variables

9.3.1. Primary endpoint

- Occurrence of new onset of confirmed[4] autoimmune disease during the period of one year following administration of the first dose of Cervarix (risk period) among an exposed cohort and during an equivalent time period in the unexposed cohorts for the following two co-primary composite endpoints:

  [1] **Neuroinflammatory/ophthalmic autoimmune diseases:**

  - Multiple Sclerosis
  - Transverse myelitis
  - Optic neuritis
  - Guillain-Barré syndrome, including Miller Fisher syndrome and other variants
  - Other demyelinating diseases:
    - Acute disseminated encephalomyelitis, including site specific variants: e.g. non-infectious encephalitis, encephalomyelitis, myelitis, myeloradiculomyelitis
    - AI peripheral neuropathies and plexopathies (including chronic inflammatory demyelinating polyneuropathy, multifocal motor neuropathy and polyneuropathies associated with monoclonalgammopathy).
  - Auto-immune uveitis

  [2] **Other autoimmune diseases:**

  - Systemic lupus erythematous
  - Autoimmune (AI) disease with rheumatologic conditions:
    - Rheumatoid arthritis (RA)

[4] Auto-immune disease diagnosis ascertainment by an expert physician panel (Section 9.4.2).

  - Juvenile rheumatoid arthritis (JRA)
  - Still's disease
  - Psoriatic arthritis
  - Ankylosing Spondylitis
- AI haematological conditions:
  - Idiopathic thrombocytopenic purpura (ITP)
  - AI haemolytic anaemia
- AI endocrine conditions:
  - Type 1 diabetes mellitus
  - AI thyroiditis including Hashimoto's disease, Graves' /Basedows' disease
- Inflammatory bowel / hepatic diseases:
  - Crohn's diseases
  - Ulcerative colitis
  - Autoimmune hepatitis

9.3.2. Secondary endpoints

Secondary endpoint is the occurrence of new onset of individual confirmed autoimmune disease during the following specific periods:

- Occurrence of Guillain Barré syndrome (including Miller Fisher syndrome and other variants), and autoimmune haemolytic anaemia within two months following the administration of the first dose of Cervarix;
- Occurrence of idiopathic thrombocytopenic purpura (ITP) within six months following the administration of the first dose of Cervarix;
- Occurrence of multiple sclerosis, transverse myelitis, optic neuritis, other demyelinating diseases (see the two sub-bullets for these diseases in Section 9.3.1), auto-immune uveitis, systemic lupus erythematous (SLE), rheumatoid arthritis (RA), juvenile rheumatoid arthritis (JRA), Still's disease, psoriatic arthritis, ankylosing spondylitis, type 1 diabetes mellitus, auto-immune thyroiditis (including Hashimoto's disease, Graves'/Basedows' disease), and inflammatory bowel / hepatic disease (Crohn's disease, ulcerative colitis and autoimmune hepatitis) within one year following the administration of the first dose of Cervarix.

9.3.3. Data to be collected

9.3.3.1. Subjects characteristics

The following data will be extracted for the analysis population:

- Demographic characteristics: birth month and birth year, sex, region, practice region identifier, date of death (if applicable) and acceptable patient flag

- CPRD GOLD information: CPRD GOLD start date, first registration date, current registration date, registration gaps, registration status, transfer-out date, transfer-out reason
- HES information: Linkage to HES data
- Vaccines:

  Administration of any other vaccine from one year before the reference date and until the end of follow-up will be collected: date of vaccination, medcodes and immunization type will be extracted from the immunisation file. Cross-tabulation of medcodes and vaccine names/class is detailed in Annex 5.
- Health care resource utilization: number of primary care resource utilization during the year before the reference date.

9.3.3.2. Clinical outcomes

Occurrence of auto-immune diseases defined as study endpoints will be identified using defined algorithms (see Annex 5).

For each case, the following data will be extracted:

- Medcode(s)
- Date of event

The associated "free text" (event text ID) if any will be identified. For all the identified cases of auto-immune diseases, the associated free text will be reviewed by a GSK-identified reviewer for confirmation and determination of the date of first symptoms of NOAD.

9.3.3.3. Other derived variables

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

- Subject'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 (except for vaccination date which is the reference date) 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 a specific event will be computed as the difference between the date of the event and the date of birth

9.4. Data Sources

9.4.1. The UK 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 (GPs) to maintain the clinical records within their practices. As of March 2011, CPRD GOLD contains 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 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). The current linkage between CPRD GOLD primary care data and HES data is around 50% as of Q1 2013. 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 Q1 of 2013 (first release of 2013) contains data for 10,960,947 research standard patients, drawn from 660 practices throughout the UK. A total of 4,727,669 patients from 548 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 6.9 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. 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.

9.4.2. Data source for case ascertainment

First, the CPRD GOLD is based on data from GPs, while most auto-immune diseases would probably be diagnosed in specialist settings. Consequently, the number of auto-immune diseases, the quality of the information, and the diagnostic certainty might be lower compared to other databases that include hospital data only. In particular the specific information related to the onset of clinical symptoms, and radiological and biological data associated with the etiologic diagnosis of auto-immune diseases may not all be available in the CPRD GOLD database and associated resources. Besides, not all general practices participating to CPRD GOLD have consented to the linkage between CPRD GOLD primary care data and HES data (current linkage around 50% as of Q3 2012). Specific algorithms for each outcome of interest have been developed (Annex 5)

and available "free text" that is related to auto-immune disease diagnosis will be requested from CPRD GOLD, when needed.

Moreover, in a recent study by [Chao & Jacobsen, 2012], the authors recommended that expert case review of medical records is used in autoimmune safety studies, and case identification can be expanded by use of laboratory test results and other relevant measures in addition to specific ICD-10 diagnosis codes.

The ascertainment of the etiologic diagnosis and date of disease onset for all identified auto-immune diseases will be performed by a contract research organisation (CRO) to ensure the correct classification of each case. The CRO will review all subject data retrieved from CPRD GOLD including medcodes (including clinical, laboratory, and treatment files), the relevant "free text" and HES (including specific ICD-10 diagnostic codes), when available. They will assess whether the aetiology of the auto-immune disease is confirmed or not and whether the date of disease onset falls within the observation period of the study, which is one year after the reference date.

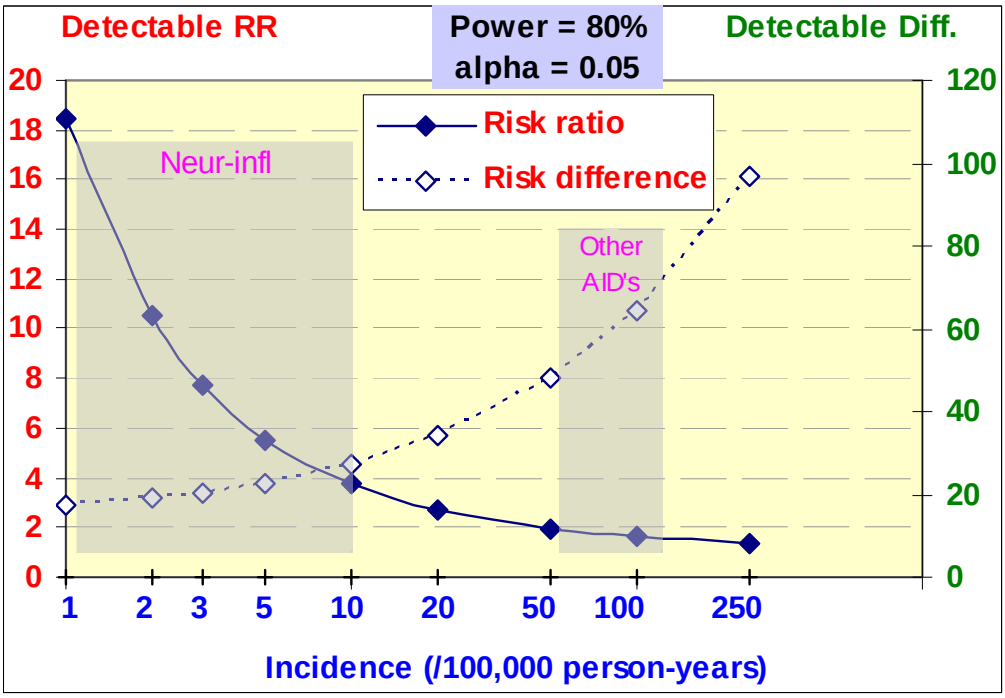
In the event that the aetiology or the date of onset could not be confirmed, a second review step will be conducted with an expert physician panel to reach an agreement about the case ascertainment, and, if they cannot reach an agreement, then the different clinical opinion of the experts will be listed. The experts will be blinded with regards to HPV vaccine exposure.

9.5. Study size

9.5.1. Sample size for cohort design

The target sample size is 50,000 subjects in each cohort. The relative risk (RR) that would be detected with 80% power and alpha = 0.05 is given in Figure 3 versus the incidence rate in the (Cervarix) unexposed cohort. The detectable difference in incidence rate (= additional cases per 100,000 person-years) is also depicted.

Figure 3 **Detectable relative risk and difference versus the incidence rate in the (Cervarix) unexposed cohort**



(Method: Comparison of two independent proportions using a likelihood ratio test, PASS 2005)

Cohorts of 50,000 subjects each should allow detection, with 80% power, of a RR between 3.7 and 18.7 for the neuro-inflammatory NOAD (incidence rate between 10 and 1/100,000 person-years) and between 1.6 and 2.0 for other NOAD (incidence rate between 100 and 50/100,000 person-years).

Because of risk of lost to follow-up, the sample size is increased by approximately 30% in each cohort to approx. 65,000 subjects.

9.5.2. Sample size for self-control case-series

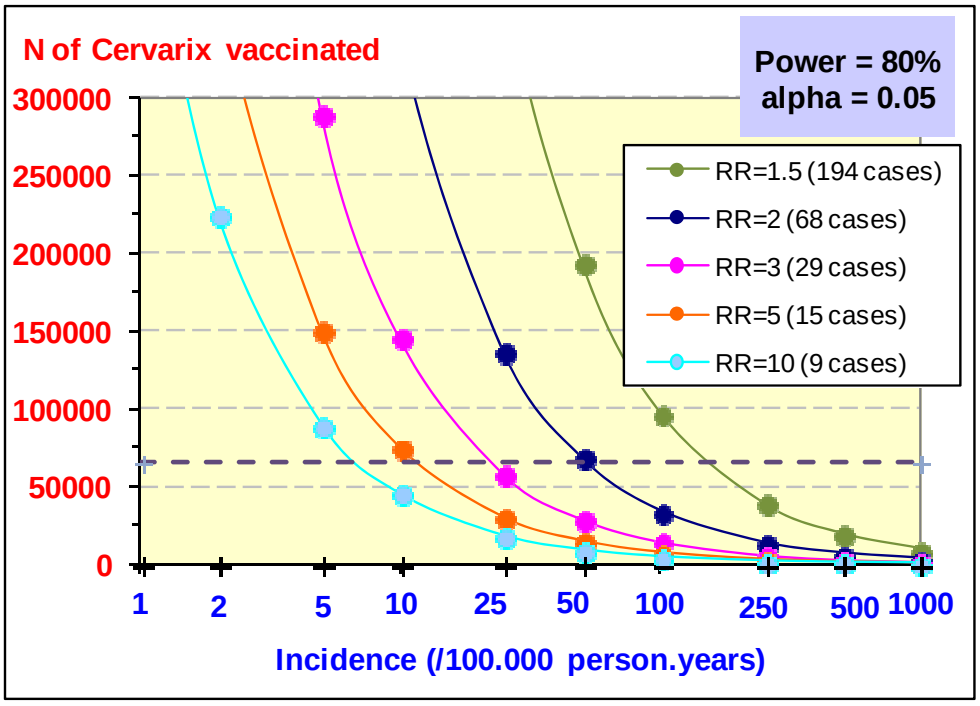
The power of the SCCS analysis depends on the number of cases and the ratio between the duration of the risk and the control periods. With a risk and a control period of 12 months each, the number of cases needed versus the detectable risk ratio (incidence rate ratio) is summarized in Table 4. Figure 4 shows the total number of vaccinated subjects needed to follow-up for 30 months after vaccination versus the number of cases and the background incidence.

Table 4 Sample size for a SCCS analysis - Number of cases in vaccinated subjects versus the incidence rate ratio [a]

Incidence rate ratio	Total number of cases
1.5	194
2	68
3	29
5	15

Method: sample for case-series analysis based on the signed root likelihood ratio [Musonda, 2006]
[a] 80% power using a two-sided test and alpha = 0.05

Figure 4 Population size for a SCCS analysis versus the incidence rate ratio and the background incidence in the general population



Dotted line: target sample size of the exposed cohort

9.6. Data Management

9.6.1. Remote Data Entry instructions

Remote Data Entry (RDE), using a validated computer application 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.

9.6.2. Final study database

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 according to GSK Biologicals Standard Procedures.

9.7. Data Analysis

Examples of statistical tables and figure templates are given in Annex 6.

9.7.1. Hypotheses

9.7.1.1. Hypotheses for the cohort analysis

Null hypothesis (H0): the incidence of neuroinflammatory/ophthalmic autoimmune diseases (other autoimmune diseases) in the exposed female cohort is equal to the incidence in the historical non-exposed female cohort (historical cohort).

Alternative hypothesis (H1): the incidence of neuroinflammatory/ophthalmic autoimmune diseases (other autoimmune diseases) in the exposed female cohort is different from the incidence in the historical unexposed female cohort (historical cohort).

These hypotheses will be tested separately for each of the two co-primary endpoints. No alpha adjustment will be done.

The same hypotheses will be tested between the two male cohorts.

9.7.1.2. Hypotheses for the self-control case-series analysis

Null hypothesis (H0): the incidence rate of neuroinflammatory/ophthalmic autoimmune diseases (other autoimmune diseases) in the exposed female cohort is the same during the risk period and the control period.

Alternative hypothesis (H1): the incidence rate of neuroinflammatory/ophthalmic autoimmune diseases (other autoimmune diseases) in the exposed female cohort is different during the risk period and during the control period.

These hypotheses will be tested separately for each of the two co-primary endpoints. No alpha adjustment will be done. These hypotheses will be tested for individual diseases using the specific risk periods provided that at least 10 cases be recorded (with 10 cases for SCCS analysis, we will have 80% power to detect incidence rate of 6 to 8 depending on the risk period versus control period ratio).

9.7.2. Analysis Population

9.7.2.1. Population for the cohort design

The study population for the cohort design will comprise all exposed and unexposed subjects that satisfy the inclusion criteria.

9.7.2.2. Population for the SCCS analyses

Only the cases of NOAD recorded in the exposed cohort during either the risk or the control periods will be included in the SCCS analysis.

9.7.3. Subject disposition

Subject disposition will be summarized by cohort and overall by computing:

- Number of screened subjects.
- Number (%) of non-eligible subjects for each of the following reasons of non-eligibility:
  - Diagnostic code of NOAD during the year prior to the reference date (for all cohorts);
  - Subject not actively registered with the practice during the study period (for all cohorts);
  - Subject not flagged as acceptable in CPRD GOLD (for all cohorts);
  - Subject not recorded for at least 12 months within CPRD GOLD at reference date;
  - At least one dose of unspecified HPV vaccine or Gardasil at any time before the reference date(for all cohorts);
  - At least one dose of Cervarix vaccine at any time before the reference date (for unexposed cohorts);
  - The first dose of Cervarix received before 01 September 2008 or after 31 August 2010 (for the exposed female cohort).
- Number of eligible subjects in each cohort.
- After frequency matching for age and for practice-region, number of included subjects in each cohort.

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

9.7.4. Demographic and baseline characteristics

Demographic and baseline characteristics of all included subjects (age at reference date, region (GP practice), other vaccination during the previous year) 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.

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

9.7.5. Analysis of co-primary endpoints

9.7.5.1. Cohort analysis

The primary analysis will compare the incidence rates of the primary outcomes of interest between the Cervarix exposed female cohort and the historical unexposed female cohort. Results will be presented as the incidence rate ratio and the incidence difference. Exposed person-time will be defined as the period between the reference date and the earliest of the following events:

- End of study period (defined as 12 months after the reference date);
- Date of de-enrolment from CPRD GOLD;
- Date of unspecified HPV vaccine or Gardasil or Cervarix for unexposed cohort;
- Date of first diagnosis of the outcome of interest.

Incidence rates for auto-immune diseases will be calculated by dividing the number of cases by person-time. A Poisson regression model will estimate the risk ratio and its 95% confidence interval. The Poisson model will include the number of cases in each cohort as the dependent variable, the exposure status as a binary independent variable, the age-group ([9-18],[18-25]) as a covariate, and the log-transformed total person-year as an offset.

Sensitivity analyses

Sensitivity analyses will be performed:

- Incidence rates for NOAD in the exposed cohort will be calculated after each dose by dividing the number of cases by the total person-time. Exposed person-time will be defined as the period between the date of the dose administration and the earliest of the following events:
  - End of risk period (defined as 6 months after each dose);
  - Date of the next Cervarix dose;
  - Date of de-enrolment from CPRD GOLD;

  - Date of unspecified HPV vaccine or Gardasil;
  - Date of first diagnosis of the outcome of interest.

- Separate analysis will be performed for subjects younger/older than 18 years.
- Analysis including possible confounding factors (if available): other vaccination, age, region, healthcare resource utilization, categorization will be determined based on the available data.
- In case of more than 5% of NOAD with unknown/incomplete date, a sensitivity analysis will be done including these events as occurring during the risk period. This imputation will be done for the four cohorts.
- The same comparison will be done between the two male cohorts. In case of significant difference between these two male cohorts, the primary analysis will be adjusted for time effect other than Cervarix (see Section 9.7.5.1 for the detailed statistical models).

For the SCCS analysis, the incidence rates in the exposed cohort during the risk period will be compared with the incidence rates in the exposed cohort during the control period using a conditional Poisson regression model [Whitaker, 2006].

Analysis of individual disease could also be performed depending on the number of cases (at least 10 cases in both exposed and non-exposed cohorts for each defined risk period).

9.7.6. Secondary endpoints

The new cases of individual autoimmune disease during the specific period (see Section 9.3.2) will be analysed by descriptive statistics per cohort. Incidence rate during the specific period will be computed per cohort for each individual disease as the total number of new cases divided by the total person-year as for the primary endpoint.

In case of more than 10 cases in the exposed female and unexposed female cohorts, a Poisson regression model will estimate the risk ratio and its 95% confidence interval. The Poisson model will include the number of cases in each cohort as the dependent variable, the exposure status as a binary independent variable and the log-transformed total person-year as an offset. Same analysis will be performed for the two male cohorts.

A SCCS analysis of individual disease will be carried out if there are at least 10 cases of disease in the exposed female cohort and during the total of the risk and the control period. The risk period of each disease is the disease specific period : 2 months, 6 months, or 1 year depending on the disease (see Section 9.3.2), the control period is, 22 months, 18 months, or 12 months, respectively. The risk and the control periods are separated by a 6-month buffer period.

These analyses will not be considered as confirmatory analysis.

9.7.7. Exploratory analysis

We will employ the SaTScan software package to investigate temporal clustering of adverse events. Using this package, we will analyze data to test whether the events are randomly temporal distributed. This analysis will be done for each of the four cohorts.

9.7.8. 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.

9.7.8.1. Handling of missing data

Missing data will not be substituted.

9.7.8.2. Descriptive statistics

Age at reference date will be summarized by descriptive statistics per cohort and overall: n of subjects, mean, SD, median, minimum and maximum and compared among the four cohorts using a one-way ANOVA. In case of an overall significant difference, pair-wise comparison using t-test will be carried out.

Exposure to other vaccines (Yes/No) during the year prior to the reference date and during the one year after will be summarized in frequency tables (n, %) per cohort and overall. The four cohorts will be compared using Chi-square test. In case of an overall difference, pair-wise comparison will be carried out using Chi-square test.

All auto-immune diseases 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 subjects

All the cases of auto-immune disease will be listed with data about exposure status and demographic characteristics.

9.7.9. Statistical models

9.7.9.1. Poisson regression

Poisson regression will be computed using the SAS GENMOD procedure. The dependent variable is the number of events (Y). The main model (Model 1) will include the exposure status (exposed ($X=1$) vs. non-exposed ($X=0$)) as a binary independent variable, the age-group ([9-18],[18-25]) as a covariate, and the log-transformed total person-time (PY) of each exposed and unexposed cohort as an offset.

Female cohort model: $$\ln(Y) = \beta_0 + \beta_1 X + \beta_2 Z + \ln(PY)°$$

The coefficients β_1 and β_2 are the coefficients associated to exposure effect and age-group, respectively. The risk ratio (exposed/unexposed) will be derived as the exponential of the coefficient associated with the exposure status and its 95% Wald confidence interval.

The SAS code is:

```
PROC  GENMOD data=<filename> ;
  MODEL Y= X  Z / offset=Ln_PY dist=poisson link=log ;
RUN;
```

The same model will be run for comparing the two male cohorts (concurrent vs. historical)

Male cohort model: $$\ln(Y) = \beta_{0'} + \beta_{1'} X + \beta_{2'} Z + \ln(PY)°$$

The coefficient $\beta_{1'}$ is the time effect (concurrent vs. historical) in males.

The exposure effect in females adjusted for temporal effect in males model will be:

$$\ln(Y) = \beta_0 + \beta_{11} X_1 + \beta_{12} X_2 + \beta_{13} X_3 + \beta_2 Z + \ln(PY)°$$

Where:

X_1 is '1' for the exposed female cohort and '0' for the other cohorts;
X_2 is '1' for the concurrent male cohort and '0' for other cohorts;
X_3 is '1' for the historical male cohort and '0' for the other cohort.

The Cervarix effect in females adjusted for the temporal effect in males will be computed as :

$$\beta_{1*} = \beta_{11} - (\beta_{12} - \beta_{13})$$

The SAS code is:

```
PROC  GENMOD data=<filename> ;
  Class X (ref=' 1');
  MODEL Y= X  Z / offset=Ln_PY dist=poisson link=log ;
  Contrast "vaccine adjusted effect in Females" X 1 -1 +1/ / estimate=exp;;
RUN;
```

Note: X=1 for the non-exposed female cohort, 2 for the exposed female cohort, 3 for concurrent male cohort, and 4 for the historical male cohort.

Sensitivity analysis: A Poisson regression model including, in addition to the exposure and the age-group, other covariates

- Region (class variable: 13 regions defined in CPRD)
- Vaccination during the year prior to reference data (2 classes: yes, no)

- Use of healthcare resources during the previous year (categories will be quartiles computed from all cohorts)

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. If the number of subjects or the number of events is too low in some categories (for example regions), categories could be grouped.

9.7.9.2. Self-control case-series

Background

The self-control case-series method (SCCS) was developed to investigate associations between vaccination and acute potential adverse events [Farrington, 1996]. The SCCS is based only on cases, and provides consistent estimates of the relative incidence. It controls implicitly for all fixed confounders, that is to say, confounders that do not vary with time over the observation period, e.g. genetics, location, socio-economic status, gender, individual frailty, severity of underlying disease.

The effect estimate is calculated as the ratio of the rate (or hazard) of events in a given post-exposure period (risk period), to the rate of events in the absence of the exposure (control period).

Risk and control periods

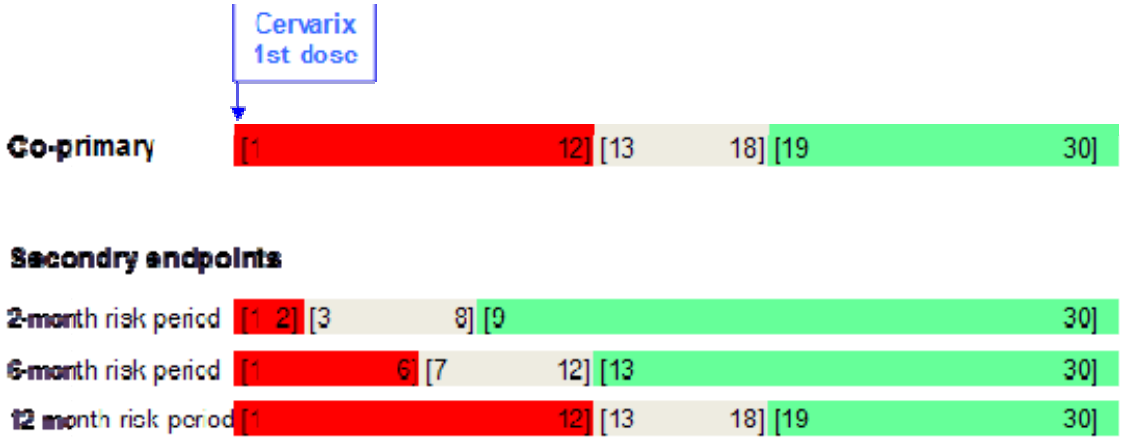
For the two co-primary endpoint analysis, the period at risk will be 12 months after the first dose of Cervarix. The control period will also be a period of 12 months after the risk period but separated by a buffer period of 6 months. The new events occurring during the buffer period will not be included in the model (Figure 5). A buffer period is defined to control the risk of late effect of vaccine (risk period for many auto-immune diseases is not clearly defined). We have judged reasonably that the vaccine effect if any will be absent after 18 months.

A sensitivity analysis of the co-primary endpoints will be carried out taking into account individual risk and control periods. For each subject the risk period is the period from the first dose of Cervarix until 6 months after the last dose, and the control period is the period from 6 months after the end of the risk period until month 30. This sensitivity analysis is an SCCS adjusted for the individual variability in the vaccination schedule.

For individual diseases (secondary endpoint), specific risk periods will be defined (Figure 5). SCCS analysis of individual diseases will be carried out only if there are at least 10 cases in the risk and the control periods.

Censoring will not be applied for this analysis, since date of death from the population register will not be available at the time.

Figure 5 Risk and control periods for the various endpoints



Statistical Calculations

The statistical calculation will be done using the specific SAS macro developed by Whitaker et al. [Whitaker, 2006] and available online from http://statistics.open.ac.uk/sccs

Mathematical model

Because of relatively short duration of the total period of observation (30 months) compared to a possible age effect, no age effect will not be included in the model.

Each individual ***i*** is observed during a time **[a_i, b_i].** This interval is the observation period for the individual ***i.*** The observation period for individual ***i*** is then partitioned into k=0,1 periods. Risk period, ***k = 1,*** correspond to an increased risk relative to control period which is coded ***k = 0.***

Conditioning on the exposure history over the entire observation period, we assume that events of interest for individual ***i*** arises as a non-homogeneous Poisson process with rate λ_{ik}. If n_{ik} is the number of events arising for individual ***i*** and risk period ***k***, then

$$n_{ik} \approx \text{Poisson}\left(\lambda_{ik} e_{ik}\right)$$

where e_{ik} is the time spent by subject ***i*** in period ***k***.

Conditioning on the total number of events $n_i = \varphi_i + \beta_k$ arising in **[a_i, b_i],** the log-likelihood contribution of individual ***i*** is

$$l_i = \sum_k n_k \log\left(\frac{\lambda_{ik} e_{ik}}{\sum_s \lambda_{is} e_{is}}\right)$$

With a log-linear model for the Poisson rate of the form

$$\log(\lambda_{ik}) = \varphi_i + \beta_k$$

Where φ_i is an individual effect, and β_k is the exposure effect associated with risk period. The parameter β_k is the log relative incidence.

The log-likelihood estimate of β_k is

$$l(\beta) = \sum_i n_{ik} \log\left(\frac{\exp(\beta_k) e_{ik}}{\sum_r \exp(\beta s) e_{ir}}\right)$$

9.7.9.3. Scan Statistics

Scan statistics will be used to detect temporal clusters of cases. This is done by gradually scanning a window across time noting the number of observed and expected observations inside the window. The scanning window will be an interval of time. Two time window sizes will be used: 2 months and 4 months. The window with the maximum likelihood is the most likely cluster, that is, the cluster least likely to be due to chance. A p-value is assigned to this cluster. Since two time windows will be used, p-values will be compared to a Bonferroni adjusted alpha ($\alpha/2 = 0.025$). Scan statistics use a different probability model depending on the nature of the data. Under the null hypothesis, the observed events occur randomly following a uniform distribution according to a discrete Poisson model during the total observation period.

This scan statistics analysis will be done in each of the four cohorts to detect possible clustering of the two co-primary composite endpoints and of the individual diseases.

9.7.10. Conduct of analyses

9.7.10.1. Sequence of analyses

The study feasibility assessment is intended to confirm that the data are of sufficient quality to confirm the diagnoses, and that the target sample sizes can be reached. A brief report of the study feasibility assessment will tabulate the numbers of confirmed diagnoses and the number of vaccinees in the periods considered for the exposed and historical cohort.

The final analyses will be performed according to a two-step schedule:

1. Between-cohort analysis will be done when the primary and secondary endpoints occurring during the one-year follow-up period will be available for all subjects and the corresponding database will be frozen.
2. The SCCS analysis will be carried out when the primary and secondary endpoints occurring during the 30-month follow-up period will be available for all exposed female subjects and the corresponding database will be frozen.

9.7.10.2. Statistical considerations for interim analyses

There is no interim analysis.

9.7.10.3. Changes from planned analyses

Not applicable.

9.8. Quality control

Validation of clinical outcomes is described in Section 9.3.3 and Section 9.4.2.

Clinical data management will be performed in accordance with applicable GSK standards and data cleaning procedures.

The final study dataset will be archived and stored on a secured, access limited, computer platform SAS Drug Development (SDD) according to GSK Biological Standard Procedures. Specific statistical programs will be written in SAS 9.2 (or higher) and validated according to the GSK standard procedures. The validation of the quality control (QC) of the statistical analysis will be documented. All statistical programs, output files and QC documentation will be saved as read-only files on SDD.

The final study protocol and possible amendments, the final statistical report and the QC document, and the final study report(s) will be archived on a Document management system based on the Documentum platform: Computer Aided Regulatory Submission (CARS).

9.9. Limitations of the research methods

Limitations and recommendations for the research methods

As mentioned previously, Cervarix vaccination administration was done in the UK in the schools and administration of Cervarix is not reported for all subjects in CPRD. The vaccination record in the CPRD GOLD may not be complete; if for a certain subject no vaccination code for Cervarix is registered, vaccination status is uncertain. Due to the potential for under-reporting of HPV vaccination, an unexposed historical cohort (before introduction of Cervarix in the UK) was chosen instead of an unexposed concurrent cohort.

Although, diagnosis and/or coding of auto-immune diseases could change over time. Moreover, a link with HES was recently implemented in CPRD GOLD since 2009. The HES data coverage period is actually longer than 2009: currently April 1997 to March 2012. As such, historical controls will have the potential to have linked data in the same way as the more recent exposed group. In order to assess the possible impact of changes in auto-immune diseases diagnosis/coding in CPRD GOLD and HES over time, two male cohorts will also be enrolled: a cohort concurrent to the exposed female cohort and a historical cohort selected at the same time period as the unexposed historical female cohort. However, it is recognized that incidence of auto-immune diseases might vary according to sex as well as according to the age group. These two cohorts will be useful to study potential change over time in CPRD GOLD due to recent access to HES data that might increase the number of auto-immune diseases coded in the database.

A multitude of medcodes are used by the GP to enter diagnosis of autoimmune disease in CPRD GOLD. There is no ICD-10 code mapping. As far as possible , an exhaustive list of medcodes has been defined for each disease and specific algorithms have been developed. In addition, associated “free text” will be reviewed. However, risk of false positive cases (lack of specificity) could not be totally excluded. Lack of specificity can bias the risk ratio estimate to the null hypothesis. There can also be a risk of false negative (lack of sensitivity) however for a rare event, a lack of sensitivity does not bias the risk ratio estimation. Determination of the time to onset of first symptoms is also a limitation. Time between diagnosis and first symptoms is largely variable and depends on the disease and this is why not only medcodes will be used, but also HES and the free text when available.

The definition of the risk period for autoimmune disease is challenging. Too short a risk period would underestimate the actual risk whereas too long a risk period would dilute the actual risk. A period of one year after the first dose of Cervarix will be used for the co-primary composite endpoints. Analysis of the incidence rate of individual diseases will be used for a specific-disease risk period.

These limitations were addressed during a feasibility assessment: with the planned number of subjects for each cohort (65,000), the expected number of autoimmune disease cases to be reviewed should be around 280 cases, which is achievable with the developed methodology for case reviews.

10. PROTECTION OF HUMAN SUBJECTS

10.1. 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 internal Independent Scientific Advisory Committee (ISAC) favourable approval of study protocol and any subsequent amendments. This approval has been obtained on August 30, 2012 (ISAC reference 12_086R, see Annex 4).

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 [Clinical Practice Research Datalink (CPRD GOLD) Website, 2012]).

10.2. 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 not available online and never shared with external parties. When GSK requests "free text" to CPRD GOLD, CPRD GOLD has internal processes to secure the maintenance of confidentiality concerning subject identifiers. Identifiers will never be transferred to GSK.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS / ADVERSE REACTIONS

This study intends to collect data only on auto-immune diseases recorded in the CPRD GOLD. Where required, the results of this study will be communicated to regulators when the final study report becomes available.

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

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

12.1. 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.

12.2. Ownership and publication

12.2.1. 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.

12.2.2. 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, an external expert from the London School of Hygiene and Tropical Medicine (LSHTM), 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).

13. REFERENCES

Abrahamson PE, Hall SA, Feudjo-Tepie M, Mitrani-Gold FS, Logie J. The incidence of idiopathic thrombocytopenic purpura among adults: a population-based study and literature review. *Eur J Haematol* 2009;83(2):83-89.

Abramson O, Durant M, Mow W, Finley A, Kodali P, Wong A, et al. Incidence, prevalence, and time trends of pediatric inflammatory bowel disease in Northern California, 1996 to 2006. *J Pediatr* 2010;157(2):233-239.e1.

Allen C, Palta M, D'Alessio DJ. Incidence and differences in urban-rural seasonal variation of type 1 (insulin-dependent) diabetes in Wisconsin. *Diabetologia* 1986;29(9):629-633.

Alonso A, Jick SS, Olek MJ, Hernan MA. Incidence of multiple sclerosis in the United Kingdom : findings from a population-based cohort. *J Neurol* 2007;254(12):1736-41.

Beghi E, Bogliun G. The Guillain-Barre syndrome (GBS). Implementation of a register of the disease on a nationwide basis. Italian GBS Study Group. *Ital J Neurol Sci* 1996;17(5):355-61.

Burnham JM, Shults J, Weinstein R, Lewis JD, Leonard MB. Childhood onset arthritis is associated with an increased risk of fracture: a population based study using the General Practice Database. *Ann Rheum Dis.* 2006;65(8):1074-1079.

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

Chao C, Klein NP, Velicer CM, Sy LS, Slezak JM, Takhar H, et al. Surveillance of autoimmune conditions following routine use of quadrivalent human papillomavirus vaccine. *Journal of Internal Medicine.* 2011;271(2):193-203.

Chao C & Jacobsen S. Evaluation of autoimmune safety signal in observational vaccine safety studies. *Human Vaccines & Immunotherapeutics* 2012;8(9):1302-1304.
URL:
http://www.landesbioscience.com/journals/vaccines/2012HV0049.pdf?nocache=2093950154

Clinical Practice Research Datalink (CPRD GOLD) Website. ISAC (Independent Scientific Advisory Committee) – Submission of CPRD GOLD Protocols to ISAC. http://www.cprd.com/ISAC. Accessed: 09 May 2012.

Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. *Autoimmun Rev.* 2003;2:119-125.

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

Department of Health (DH). *Immunisation against infectious disease.* The Green Book, DH publication; 2008 Gateway reference: 7523. The 'Green Book' chapter on Human Papillomavirus (HPV); dh_087787.

Descamps D, Hardt K, Spiessens B, Izurieta P, Verstraeten T, Breuer T, Dubin G. Safety of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine for cervical cancer prevention: A pooled analysis of 11 clinical trials. *Hum Vaccine.* 2009;5:332-340.

Diamond Project Group. Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999. *Diabet Med* 2006;23(8):857-66.

Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol.* 1996;143:1165-1173.

Fishbein HA, LaPorte RE, Orchard TJ, Drash AL, Kuller LH, Wagener DK. The Pittsburgh insulin-dependent diabetes mellitus registry: seasonal incidence. *Diabetologia* 1982;23(2):83-85.

Gonzalez EL, Johansson S, Wallander MA, Rodriguez LA. Trends in the prevalence and incidence of diabetes in the UK: 1996-2005. *J Epidemiol Community Health* 2009;63(4):332-336.

GSK confidential document. Prophylactic HPV-16/18 L1 VLP Vaccine Formulated with AS04. Investigator Brochure Edition 11. March 2012.

Harron KL, McKinney PA, Feltbower RG, Bodansky HJ, Norman PD, Campbell FM, et al. Incidence rate trends in childhood type 1 diabetes in Yorkshire, UK 1978-2007: effects of deprivation and age at diagnosis in the South Asian and non-South Asian populations. *Diabet Med* 2011;28(12):1508-13.

Health Protection Agency Department of Health (HPA). *Routine programme for year 8 girls (12-13 years old) and catch-up campaign for year 13 girls (17-18 years old).* Sheridan A and White J, London, UK; 2010. Annual HPV vaccine uptake in England: 2008/09.

Health Protection Agency (HPA; Immunisation section). Annual HPV vaccine coverage in England in 2010/11. 22 March 2012.

Henderson P, Hansen R, Cameron FL, Gerasimidis K, Rogers P, Bisset WM, et al. Rising incidence of pediatric inflammatory bowel disease in Scotland. *Inflamm Bowel Dis* 2012; 18:999-1005.

Herrinton LJ, Liu L, Lewis JD, Griffin PM, Allison J. Incidence and prevalence of inflammatory bowel disease in a Northern California managed care organization, 1996-2002. *Am J Gastroenterol* 2008;103(8):1998-2006.

Hochberg MC. The incidence of systemic lupus erythematosus in Baltimore, Maryland, 1970-1977. *Arthritis Rheum* 1985;28(1):80-86.

Hughes RA, Charlton J, Latinovic R, Gulliford MC. No association between immunization and Guillain-Barré syndrom in the United Kingdom, 1992 to 2000. *Arch Intern Med.* 2006;166(12):1301-1304.

Imkampe A-K & Gulliford MC. Trends in Type 1 diabetes incidence in the UK in 0- to 14-year-olds and in 15- to 34-year-olds, 1991-2008. *Diabetic Medicine* 2011;28(7):811-4

International Society for Pharmacoepidemiology (ISPE). *Guidelines for Good Pharmacoepidemiology Practices.* GPP, http://www.pharmacoepi.org/resources/guidelines_08027.cfm; 2007. Accessed 23 October 2011.

Koobatian TJ, Birkhead GS, Schramm MM, Vogt RL. The use of hospital discharge data for public health surveillance of Guillain-Barre syndrome. *Ann Neurol* 1991;30(4):618-21.

Kostraba JN, Gay EC, Cai Y, Cruickshanks KJ, Rewers MJ, Klingensmith GJ, et al. Incidence of insulin-dependent diabetes mellitus in Colorado. *Epidemiology* 1992;3(3):232-238.

Leese GP, Flynn RV, Jung RT, MacDonald TM, Murphy MJ, & Morris AD. Increasing prevalence and incidence of thyroid disease in Tayside, Scotland: The Thyroid Epidemiology, Audit and Research Study (TEARS). *Clinical Endocrinology* 2008;68:311-316.

Lehtinen M, Paavonen J, Wheeler CM, Jaisamrarn U, Garland SM, Castellsague X et al. Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *The Lancet.* 2012;13:89-99.

Lipton RB, Fivecoate JA. High risk of IDDM in African-American and Hispanic children in Chicago, 1985-1990. *Diabetes Care* 1995;18(4):476-482.

MacDonald MJ, Johnson SD, Gottschall JL, Hunter JB, Winter KL, Grieshop RJ, et al. Epidemiology of insulin dependent diabetes before age 20 in Wisconsin, with particular reference to seasonality. *Diabetes Res* 1989;12(4):151-60.

Mayr WT, Pittock SJ, McClelland RL, Jorgensen NW, Noseworthy JH, Rodriguez M. Incidence and prevalence of multiple sclerosis in Olmsted County, Minnesota, 1985-2000. *Neurology* 2003;61(10):1373-1377.

McCarty DJ, Manzi S, Medsger TA, Jr., Ramsey-Goldman R, LaPorte RE, Kwoh CK. Incidence of systemic lupus erythematosus. Race and gender differences. *Arthritis Rheum* 1995;38(9):1260-70.

Medicines and Healthcare products Regulatory Agency (MHRA). Drug Safety Update 2012;6(4):H2. URL: http://www.mhra.gov.uk/Safetyinformation/Safetywarningsalertsandrecalls/UKsafetyPublicAssessmentReports/CON221607 (click on .pdf document at bottom of page; last accessed on 06-FEB-2013).

Musonda P, Farrington CP, Whitaker HJ. Sample sizes for self-controlled case series studies. *Stat Med.* 2006;25:2618-2631.

Myasoedova E, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. *Arthritis Rheum* 2010;62(6):1576-1582.

Naleway AL, Davis ME, Greenlee RT, Wilson DA, McCarty DJ. Epidemiology of systemic lupus erythematosus in rural Wisconsin. *Lupus* 2005;14(10):862-866.

Nightingale AL, Farmer RD, de Vries CS. Incidence of clinically diagnosed systemic lupus erythematosus 1992-1998 using the UK General Practice Research Database. Pharmacoepidemiol Drug Saf 2006;15(9):656-61.

Riggs JE, Gutmann L, Whited JD. Guillain-Barre syndrome. Another immune-mediated disease with a predilection for young women? *West Virginia Med J* 1989;85(9):382-383.

Schoonen WM, Kucera G, Coalson J, Li L, Rutstein M, Mowat F, Fryzek J, Kaye JA. Epidemiology of immune thrombocytopenic purpura in the General Practice Research Database. *Br J Haematol.* 2009;145:235-244.

Sheridan A and White J. Annual HHPV 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, 2009.

Sheridan A and White J. Annual HPV vaccine coverage in England in 2009/2010; Health Protection Agency Department of Health, 2010.

Shui IM, Rett MD, Weintraub E, Marcy M, Amato AA, Sheikh SI et al. Guillain-Barré Syndrome incidence in a large United States cohort (2000-2009). *Neuroepidemiology* 2012;39:109-115.

Siegrist CA, Lewis EM, Eskola J, Evans SJ, Black SB. Human papilloma virus immunization in adolescents and young adults: a cohort study to illustrate what events might be mistaken for adverse reactions. *Pediatr Infect Dis J.* 2007;26:979-984.

Simpson KN, Coughlin CM, Eron J, Bussel JB. Idiopathic thrombocytopenia purpura: Treatment patterns and an analysis of cost associated with immunoglobulin and anti-D therapy. *Seminars in Hematology* 1989;35:58-64.

Steed H, Walsh S, Reynolds N. Crohn's disease incidence in NHS Tayside. *Scott Med J* 2010;55(3):22-25.

Verstraeten T, Descamps D, David MP, Zahaf T, Hardt K, Izurieta P, Dubin G, Breuer T. Analysis of adverse events of potential autoimmune aetiology in a large integrated safety database of AS04 adjuvanted vaccines. *Vaccine.* 2008;26:6630-6638.

Whitaker HJ, Farrington CP, Spiessens B and Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med.* 2006;25:1768-1797.

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.

Yong M, Schoonen WM, Li L, Kanas G, Coalson J, Mowat F, et al. Epidemiology of paediatric immune thrombocytopenia in the General Practice Research Database. *Br J Haematol* 2010;149(6):855-864.

Annex 1 List of stand-alone documents

No.	Document Reference No	Date	Title
1	116239 (EPI-HPV-040 VS UK)	09-JUL-2013	List of stand-alone documents
2	116239 (EPI-HPV-040 VS UK)	11-FEB-2013	ENCePP Checklist for study
3	116239 (EPI-HPV-040 VS UK)	11-FEB-2013	Glossary of terms
4	116239 (EPI-HPV-040 VS UK)	11-FEB-2013	ISAC evaluation of protocols for research involving CPRD GOLD
5	116239 (EPI-HPV-040 VS UK)	11-FEB-2013	Algorithms
6	116239 (EPI-HPV-040 VS UK)	11-FEB-2013	Example of table and figure templates
7	116239 (EPI-HPV-040 VS UK)	11-FEB-2013	Trademarks
8	116239 (EPI-HPV-040 VS UK)	09-JUL-2013	Protocol sponsor signatory approval

Annex 2 ENCePP Checklist for study protocols

Annex 3 GLOSSARY OF TERMS

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 participant 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 and followed over time (prospective / retrospective) to ascertain the outcome(s) (disease).
Eligible:	Qualified for enrolment into the study based upon strict adherence to inclusion/exclusion criteria.
eTrack:	GSK's tracking tool for clinical/epidemiological trials.
Medcode	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.
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-Authorization Safety Study (PASS)	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.

Self-control case-series (SCCS):	Statistical method for assessing the association between a transient exposure and an adverse event. The method was developed to study adverse reactions to vaccines. The method uses only cases; no controls are required as the cases act as their own controls. Each case's given observation time is divided into control and risk periods. Risk periods are defined during or after the exposure. The method estimates a relative incidence rate, that is, the incidence in risk periods relative to the incidence in control periods. Time-varying confounding factors such as age can be allowed for by dividing up the observation period further into age categories. An advantage of the method is that confounding factors that do not vary with time, such as genetics, location, socio-economic status are controlled for implicitly.
Study population:	Sample of population of interest.
Subject:	Term used throughout the protocol to denote an individual who has been contacted in order to participate or participates in the clinical/epidemiological study, or a person about whom some medical information have been recorded in a database.
Subject number:	A unique number identifying a subject, assigned to each subject consenting to participate in the study.
Targeted Safety Study (TSS)	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 conducted 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.

Annex 4 ISAC Evaluation Of Protocols For Research Involving CPRD GOLD

Annex 5 Algorithms

Annex 6 Example of table and figure templates

Annex 7 TRADEMARKS

The following trademarks are used in the present study outline. Note: In the remainder of the document, the names of the vaccines will be written without the superscript symbol ™ or ®.

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.)	Recombinant human papillomavirus quadrivalent (Types 6, 11, 16 and 18) vaccine

Annex 8 Protocol Sponsor Signatory Approval

Protocol Sponsor Signatory Approval

eTrack study number and Abbreviated Title	116239 (EPI-HPV-040 VS UK)
Date of protocol	FDA – EMA PASS Final Version 1: 09 July 2013
Title	An observational cohort study to assess the risk of autoimmune diseases in adolescent and young adult women aged 9 to 25 years exposed to Cervarix® in the United Kingdom
Sponsor signatory	Sponsor signatory names retracted to protect Subject Privacy
Signature	_____
Date	_____