In February 2013, GlaxoSmithKline (GSK) announced a commitment to further clinical transparency through the public disclosure of GSK Clinical Study Reports (CSRs) on the GSK Clinical Study Register.

The following guiding principles have been applied to the disclosure:

- *Information will be excluded in order to protect the privacy of patients and all named persons associated with the study*
- *Patient data listings will be completely removed* to protect patient privacy. Anonymized data from each patient may be made available subject to an approved research proposal. For further information please see the Patient Level Data section of the* ***GSK Clincal Study Register****.*
- *Aggregate data will be included; with any direct reference to individual patients excluded*

**Complete removal of patient data listings may mean that page numbers are no longer consecutively numbered*

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
Version identifier of the final study report	Study Report - Final
Date of last version of the final study report	17 March 2015
EU PAS Register Number	ENCEPP/SDPP/4584
Active substance	J07BM02-Papillomavirus (human types 16, 18)
Medicinal product	Cervarix®, Human Papillomavirus vaccine (Types 16, 18)
Product reference	EU/1/07/419
Procedure number	NA
Marketing Authorisation Holder(s)	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
Author	Coordinating author: • [redacted], Project Manager – Science Writing GSK contributors: • [redacted], Director, Head of Global Epidemiology • [redacted], Project Statistician • [redacted], Epidemiological Scientist/Statistician, [redacted], for GSK Biologicals

	• [redacted], Therapeutic Area Head (*Cervarix*), Safety Evaluation & Risk Management (SERM), VCSP • [redacted], Safety Physician, VCSP • [redacted], Global Regulatory Affairs, [redacted] for GSK Biologicals • [redacted], Global Regulatory Lead • [redacted], Study Delivery Lead (SDL) Non GSK contributing authors (external experts): • [redacted] (previously known as [redacted]), Senior Researcher, CPRD Research Team • [redacted], Director Epidemiology at • [redacted], • [redacted],

Marketing authorisation holder(s)

Marketing authorisation holder(s)	GlaxoSmithKline Biologicals Rue de l'Institut 89, 1330 Rixensart, Belgium
MAH contact person	[redacted], Director, Head of Global Epidemiology, GSK Biologicals

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ANNEXES

PAGE

1. 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

Keywords

Cervarix, cohort study, adolescent/young-adult women, autoimmune diseases

Rationale and background

Cervarix is a bivalent recombinant vaccine against human papillomavirus (HPV, types 16 and 18). UK *Cervarix* vaccination was started in September 2008 for 12-13 year old girls, with a catch-up programme for girls/women aged 14-18 years, with more than 6 million doses currently administered. GSK Biologicals had a post-marketing regulatory commitment with the US FDA to investigate the incidence of autoimmune diseases (AD) in subjects receiving *Cervarix*. An earlier US database study was started (EPI-HPV-015), but low recruitment meant that the study objectives could not be assessed in a reasonable timeframe. This UK study investigating incidence of AD using the CPRD GOLD database (Clinical Practice Research Datalink General Practitioner OnLine Database) was planned, and hence was proposed as an alternative study for addressing the US FDA commitment.

Research question and objectives

The primary objective was to assess whether *Cervarix* vaccination was associated with increased risk of 1) Neuroinflammatory/ophthalmic AD or 2) Other AD (co-primary objectives) within 12 months of first dose. The secondary objective was to investigate the incidence of pre-specified AD within 2, 6 or 12 months of first *Cervarix* dose.

Study design

Observational cohort study using CPRD GOLD, and a self-controlled case-series (SCCS) analysis for confirmed AD in the exposed female cohort.

Setting

CPRD GOLD containing coded records from UK general practices and its linked component of the Hospital Episode Statistics (HES).

Subjects and study size, including dropouts

The main study population included 259,876 subjects, of which 64,964 subjects were exposed to *Cervarix.*

Variables and data sources

The data sources were CPRD GOLD including "free-text" entries and HES. The endpoint for the primary objective was occurrence of one of the selected ADs within 12 months of the first *Cervarix* dose. Secondary endpoints were occurrence of selected ADs within 2, 6, or 12 months of first *Cervarix* dose. Algorithms were used to identify subjects with CPRD GOLD medcodes that corresponded to the selected ADs. After first patient profile review, individual data with doubt on the aetiology were reviewed by an expert.

Results

For the neuroinflammatory/ophthalmic diseases (confirmed cases), the adjusted Incidence Rate Ratio (IRR) for male cohorts was 0.95 [95%CI: 0.06-15.18] and it was not done for females due to only one case in the non-exposed cohort.

For the other autoimmune diseases (confirmed cases), the adjusted IRRs were 1.41 [95%CI: 0.86-2.31] for the exposed and non-exposed female cohorts and 1.77 [95%CI: 0.94-3.35] for the concurrent and historical male cohorts.

For autoimmune (AI) thyroiditis, the IRR was 3.75 [95%CI: 1.25-11.31] for confirmed cases in the two female cohorts meaning that the risk is higher in exposed women and the difference between exposed and non-exposed women was statistically significant. When considering confirmed and non-confirmed cases the IRR was 1.45 [95%CI: 0.79-2.64] and became non-significant.

Discussion

The present study did not show any evidence of increased risk for the two co-primary endpoints (neuroinflammatory/ophthalmic AD and other AD).

No cluster with regards to time to onset and *Cervarix* dosing was observed in the 49 cases of AI thyroiditis. The individual review of the cases concluded that it is not consistent with vaccination having caused the events, considering the known natural progression of AI thyroiditis, i.e. the cases observed could represent pre-existing conditions at the time of vaccination. Similar conclusions were published for the HPV4 vaccine.

Marketing Authorisation Holder(s)

GlaxoSmithKline Biologicals, Rue de l'Institut 89, 1330 Rixensart, Belgium

Names and affiliations of principal investigators

Not applicable

2. LIST OF ABBREVIATIONS

AD	Autoimmune disease(s)
ADEM	Acute Disseminated Encephalomyelitis
AI	Autoimmune
ANOVA	Analysis of Variance
CARS	Computer Aided Regulatory Submission
CI	Confidence interval
CPRD GOLD	Clinical Practice Research Datalink General Practitioner OnLine database
DH	Department of Health (UK)
EU	European Union
EXP	Exposed Cohort
FDA	Food and Drug Administration (US)
FU	Follow up
GBS	Guillain-Barré Syndrome
GP	General Practitioner
GPP	Good Pharmacoepidemiology Practices (Guidelines)
GSK	GlaxoSmithKline
HES	Hospital Episode Statistics
HIST	Non Exposed Historical Male Cohort
HPV	Human papillomavirus
ICD	International Classification of Diseases
IEC	Independent ethics committee
IRB	Institutional review board
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
LL	95% Lower exact confidence limit
MALE	Non Exposed Concurrent Male Cohort
MHRA	Medicines and Healthcare products Regulatory Agency
N	Number of subjects
NNEXP	Non Exposed Historical Female Cohort
NRES	National Research Ethics Service Committee
PASS	Post Authorization Safety Study
pIMD	potential immune-mediated disorders
QC	quality control
RA	Rheumatoid arthritis
RDE	Remote Data Entry
RR	Relative Risk
SCCS	Self-controlled case-series
SD	Standard Deviation
SDD	SAS Drug Development

SDL	Study Delivery Lead
SERM	Safety Evaluation and Risk Management
SLE	Systemic Lupus Erythematous
UK	United Kingdom
UL	95% Upper exact confidence limit
US	United States (of America)
VCSP	Vaccines Clinical Safety & Pharmacovigilance

3. ETHICS

3.1. Independent Ethics Committee (IEC) or Institutional Review Board (IRB)

The study protocol and other information that required pre-approval were reviewed and approved by the Independent Scientific Advisory Committee (ISAC) for Medicines and Healthcare products Regulatory Agency (MHRA) database research. ISAC protocol approval was given on 30 August 2012.

3.2. Ethical conduct of the study

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

3.3. Subject information and consent

No patient informed consent was needed, because the patient information in the Clinical Practice Research Datalink General Practice OnLine Database (CPRD GOLD) is fully coded and GlaxoSmithKline (GSK) Biologicals personnel were not able to make a link between the data and specific individuals.

The CPRD GOLD has obtained ethical approval from a National Research Ethics Service Committee (NRES) for purely observational research (i.e. studies that do not include patient involvement [CPRD GOLD, 2014a]).

4. INVESTIGATORS

Not applicable

5. OTHER RESPONSIBLE PARTIES

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

[redacted] was the GSK Biologicals designated Director and Head of Global Epidemiology.

The ISAC of MHRA reviewed the protocol and other information requiring pre-approval. The key roles of this committee are to provide expert advice on the medical aspects, statistical/epidemiological aspects and methodological aspects of studies involving CPRD GOLD [CPRD GOLD, 2014b].

The ascertainment of the etiologic diagnosis and date of disease onset for all identified autoimmune diseases was performed by Pallas, Health Research and Consultancy B.V to ensure the correct classification of each case (see Section 10.5.1). Pallas reviewed all subject data retrieved from CPRD GOLD including clinical diagnosis, vaccines history, laboratory testing, and drug prescription, the relevant "free text" and Hospital Episode Statistics (HES; including specific ICD-10 diagnostic codes), when available. Pallas assessed whether the aetiology of the autoimmune disease was confirmed or not and whether the date of disease diagnosis fell within the observation period of the study, which was one year after the reference date. The GSK safety physician and a Physician/Director of Epidemiology at RTI Health Solutions reviewed all suspected cases of autoimmune disease for which Pallas had some doubts.

In the event that the (date of) diagnosis, first symptom and/or first abnormal laboratory test could not be confirmed, a second review step was conducted with an expert physician to reach an agreement. A panel of 5 physicians from different medical specialty were constituted including experts in internal medicine (2), neurology (1), ophthalmology (1) and rheumatology (1). The experts were blinded with regards to HPV vaccine exposure. Furthermore, the experts reviewed a random sample of 10% of the cases per autoimmune disease.

6. MILESTONES

Milestone	Planned date	Actual date	Comments
ISAC protocol approval	Protocol submitted on 12 July 2012	Approval: 30 August 2012	None
Final protocol submitted to Regulatory Authorities	31 July 2013	31 July 2013	None
Registration in the EU PAS register	16 September 2013	01 October 2013	Registry number: ENCEPP/SDPP/4584
Start of data collection	30 September 2013	28 October 2013	
End of data collection	31 May 2014	12 August 2014	This corresponded to the date when the final case ascertainment was entered into RDE. The timelines were updated due to an unexpected workload during the case ascertainment process (additional review of patient profiles and requests for further expert consultation).
Projected study completion (Statistical analysis complete; SAC)	30 September 2014	16-October-2014	The date on which the statistical analysis for the database results was completed (SAC).
Final report of study results	31 March 2015	17 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 130 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 in females from 9 years of age onwards for the prevention of persistent infection, premalignant genital (cervical, vulvar and vaginal) lesions and cervical, vulvar and vaginal cancers (squamous-cell carcinoma and adenocarcinoma) caused by oncogenic Human Papillomaviruses (HPV).

Pre-licensure clinical studies provided key vaccine safety data, but their power to detect rare outcomes such as new onset of autoimmune disease(s) (AD) is limited due to their sample size, since incidence rates of different AD vary roughly from 1 to 20/100,000 per year [Cooper, 2003]. A pooled analysis of AD 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 almost 30,000 adolescent and adult females aged 9 years and above, of whom 16,142 received at least one dose of HPV vaccine, showed the vaccine to be generally well tolerated in women of all ages [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 AD in a broad range of women, including those of different nationalities and ethnicities [Lehtinen, 2012]. The percentage of subjects experiencing an AD 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. Another safety pooling was performed with a later data lock point of 30 April 2011. This last pooling included data from more than 42 completed or on-going controlled and uncontrolled studies conducted in 40 countries with more than 31,000 subjects in the HPV group and 24,000 subjects in Control groups (2,000 in the Coad group) [Angelo, 2014a]. The analysis of potential Immune Mediated Diseases (new approach used by the company which includes autoimmune and other diseases for which immune mediated pathophysiology is suspected) did not show statistical evidence of an increased risk of any immune-mediated disease, or of any specific syndrome after HPV-16/18 vaccination compared with other vaccines used as controls during a 1 year follow-up period and the entire study period [Angelo, 2014a].

Human papillomavirus universal immunisation using HPV-16/18 vaccine was initiated in the UK in September 2008 for 12–13 year old girls, with a catch-up programme that included girls/women 14–18 years of age, with more than 6 million doses administered. During the course of the four years that Cervarix was in use in the HPV immunisation programme, the UK Medicines and Healthcare products Regulatory Agency (MHRA) closely monitored safety. The ‘end of routine use’ review of safety data gathered to the end of July 2012 supports the conclusions that the balance of its benefits and risks remains clearly positive [MHRA, 2012a]. As part of the MHRA’s enhanced pharmacovigilance strategy for Cervarix, certain events were evaluated via statistical Observed/Expected analyses, including Guillain-Barré Syndrome (GBS), encephalitis, Bell’s palsy (VIIth nerve paralysis/facial palsy), complex regional pain syndrome and chronic fatigue syndrome / post viral fatigue syndrome. The available evidence suggests that the number of reports received by the MHRA of these events was no greater than expected and therefore consistent with chance, given the number of girls vaccinated and the natural incidence of these conditions in adolescent girls [MHRA, 2012a].

GSK was committed to develop a post-licensure study of AD 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 AD 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 AD in this age group, composite endpoints were defined and agreed with the FDA.

In the US, the commercial distribution of *Cervarix* began in November 2009. During the time period of 16 October 2009 to 28 February 2013, 1061 females in the *Cervarix* exposed and 1079 females in the unexposed cohorts were accrued [Holick, 2014]. All women in the *Cervarix* exposed cohort received the first dose of *Cervarix*, while 594 (56.0%) received the second dose and 301 (28.4%) received the third dose. This was 1.5% of the target number of 70,000 females in the *Cervarix* exposed and unexposed cohorts, respectively, which meant that it would have taken significantly longer than the planned 3 years to complete accrual [Holick, 2014].

Due to the low uptake of *Cervarix* in US, accrual of subjects exposed to *Cervarix* was limited in the EPI-HPV-015 study. The resulting sample size was not sufficient to draw conclusions. Therefore, GSK was already planning to conduct an alternative study in the UK, since it is the first country in Europe that implemented the HPV mass immunization program nationwide with over 4 million doses of *Cervarix* being administered since the start of the program in 2008 [EPI-HPV-015 VS US DB study report, 2014].

Post-licensure safety surveillance after more than 4 years of HPV-16/18 vaccine use confirms the acceptable benefit–risk of vaccination in adolescent girls and adult women. An analysis of potentially immune-mediated diseases after vaccination showed no patterns or trends for concern. The observed incidences of VIIth nerve (facial) palsy and Guillain-Barré syndrome were within the overall range of expected background incidence rates in the general population. GSK continues to closely monitor pIMDs and pregnancy outcomes, with no specific safety concern identified from more than 4 years of HPV-16/18 vaccine use in routine clinical practice [Angelo, 2014a].

8. RESEARCH QUESTION AND OBJECTIVES

8.1. Primary objective

- To assess the risk of neuroinflammatory/ophthalmic new onset of autoimmune disease(s) (AD) and other pre-specified AD 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
    - Autoimmune (AI) peripheral neuropathies and plexopathies (including chronic inflammatory demyelinating polyneuropathy, multifocal motor neuropathy and polyneuropathies associated with monoclonal gammopathy).
  - Autoimmune uveitis

  [2] **Other autoimmune diseases:**

  - Systemic lupus erythematous
  - 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 AD 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 [1], autoimmune uveitis, systemic lupus erythematous (SLE), rheumatoid arthritis (RA), juvenile rheumatoid arthritis (JRA), Still's disease, psoriatic arthritis, ankylosing spondylitis, type 1 diabetes mellitus, autoimmune 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 10.4 for the definition of the primary and secondary endpoints and the pre-specified list of AD.

8.3. Exploratory objective

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

9. AMENDMENTS AND UPDATES

None.

[1] 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 monoclonal gammopathy).

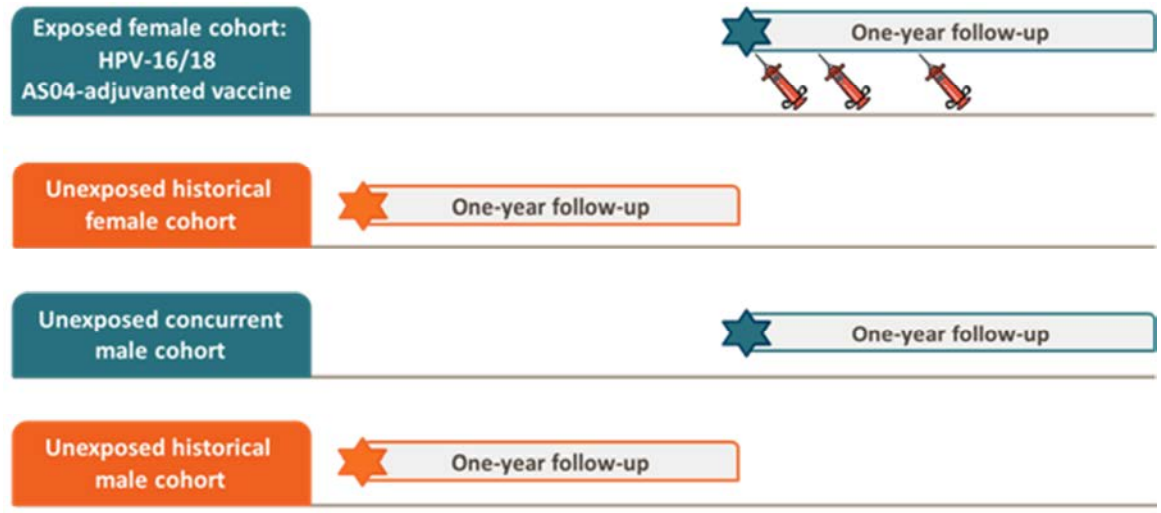
10. RESEARCH METHODS

10.1. Study design

10.1.1. Overview

- This was an observational cohort study using the CPRD GOLD data source in the UK.
- Four cohorts were defined based on exposure to *Cervarix* and sex as recorded in the CPRD GOLD data source (See Figure 1):
  1. *Cervarix* vaccinated (exposed) female cohort
  2. Unexposed historical female cohort
  3. Unexposed concurrent male cohort
  4. Unexposed historical male cohort

Figure 1 Cohort design



Reference date between 1 SEPTEMBER 2005 and 31 AUGUST 2007: female and male subjects with ≥1 general practitioner consultation.

Reference date between 1 SEPTEMBER 2008 and 31 AUGUST 2010: female subjects vaccinated with a first dose of vaccine and male subjects with ≥1 general practitioner consultation. Not all female subjects who received one *Cervarix* vaccine completed all three planned *Cervarix* vaccinations.

- Study population:
  - Female population was 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 was composed of 9- to 25-year-old male subjects not vaccinated with *Cervarix*.

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

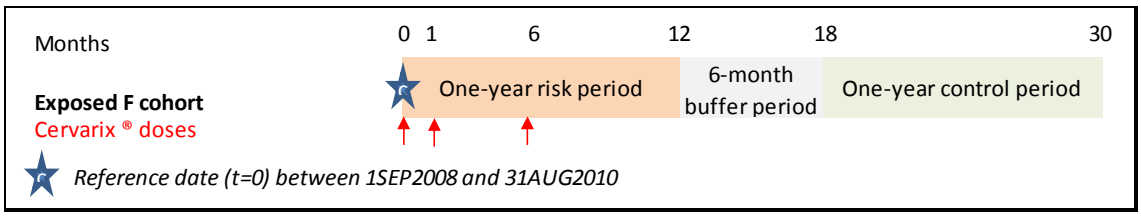
Female subjects in the unexposed historical cohort were frequency matched for age and practice region identifier to the subjects included in the vaccinated (exposed) cohort.

Study design:

Comparison of the unexposed concurrent male cohort with the unexposed historical male cohort was used as an internal control for changes over time in CPRD GOLD in reporting AD. The unexposed concurrent male cohort consisted 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 consisted of frequency age-matched and practice region-matched male subjects from the period before the introduction of *Cervarix*.

A self-controlled case-series (SCCS) analysis for confirmed AD in the exposed female cohort was also conducted, using a risk period of one year after the first *Cervarix* dose, a six-month buffer period after the risk period and a one-year control period after the buffer period (See Figure 2).

Figure 2 Self-control case-series analysis



Note: all three doses may not have been administered

10.2. Setting

10.2.1. The UK HPV National Immunization Programme and Cervarix coverage

The UK had sufficient *Cervarix* vaccination coverage to 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 matched the age range required by the FDA (9-25 years of age) for the post-licensure safety study. The bivalent vaccine was replaced in the programme by *Gardasil*® in September 2012.

10.2.2. The UK Clinical Practice Research Datalink General Practitioner OnLine database (CPRD GOLD)

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

The CPRD GOLD is operated on a non-profit basis by the UK Medicines and Healthcare products Regulatory Agency (MHRA), containing coded longitudinal medical records from general practices and more recently has been linked to data sources from hospital-based care (e.g., Hospital Episode Statistics, HES). Data quality is monitored continuously by the MHRA and data quality markers are provided at both patient and practice level. The linkage between CPRD GOLD primary care data and HES data was around 50% as of Q1 2013. The CPRD GOLD database was licensed in-house by GSK.

During the feasibility assessment the CPRD GOLD database (release 2012Q3) identified 123,085 female subjects vaccinated with *Cervarix* between 01-Sep-2008 and 31-Aug-2010. The number of vaccinated women in the relevant age-range, 9 to 25 years old, included in the CPRD GOLD was equal to 121,881. The number of HPV-vaccinated females in the CPRD GOLD appeared to be sufficient for studying potential AD adverse events.

One of the key criteria for the feasibility assessment was:

- To estimate the incidence of new cases of autoimmune disease (AD) in female and male subjects aged 9-25 years for the two study periods (01-SEP-2005 to 31-AUG-2007 and 01-SEP-2008 to 31-AUG-2010).

Table 1 illustrates the incidence rate range estimates in CPRD for the study co-primary endpoints, which confirmed that there were sufficient neuroinflammatory/ophthalmic autoimmune diseases and other autoimmune diseases recorded in CPRD GOLD to assess the causal association between *Cervarix* vaccination and these autoimmune adverse events.

Table 1 Incidence rates in CPRD GOLD for the study co-primary endpoints

Endpoints	Incidence Rate 01-Sep-2005 to 31-Aug-2007 /100,000 person-years	Incidence Rate 01-Sep-2008 to 31-Aug-2010 /100,000 person-years
[1] Neuroinflammatory/Ophthalmic autoimmune disease*	6.42 - 27.13	5.76 - 21.72
[2] Other autoimmune diseases	81.51	77.87

*6.42/100,000py and 5.76/100,000py are incidence rates including Multiple Sclerosis, Transverse Myelitis, Optic neuritis and Guillain Barre Syndrome. The incidence rates 27.13/100,000py and 21.72/100,000py are rates for Multiple Sclerosis, Transverse Myelitis, Optic neuritis, Guillain Barre Syndrome and 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 monoclonal gammopathy; and autoimmune uveitis). During the review and classification of the CPRD medical codes (medcodes), all the medcodes were classified as confirmed except for Other demyelinating diseases and uveitis where any medcodes were confirmed but classified as potential.

Table 2 illustrates the incidence rates for selected autoimmune diseases, which further confirms that CPRD GOLD is a suitable database to assess the association between certain autoimmune diseases and vaccination with *Cervarix*.

Table 2 Incidence rates in CPRD GOLD for selected autoimmune disease by sex and age classes

Autoimmune diseases	Ages Classes (years)	Incidence rate 01-Sep-2005 to 31-Aug-2007 (per 100,000 person-years)		Incidence rate 01-Sep-2008 to 31-Aug-2010 (per 100,000 person-years)	
		Males	Females	Males	Females
Type 1 Diabetes Mellitus	[9-18[	36.42	30.82	31.11	25.28
	[18-25]	26.05	16.46	15.71	13.37
Autoimmune Thyroiditis	[9-18[	1.79	5.53	1.22	5.52
	[18-25]	1.88	12.78	0.90	8.30
Crohn's disease	[9-18[	15.52	12.54	21.76	14.02
	[18-25]	26.52	31.21	27.16	29.05
Ulcerative colitis	[9-18[	7.56	4.68	9.76	6.59
	[18-25]	19.95	18.43	21.55	20.98

After the feasibility assessment, this study protocol to be conducted using CPRD GOLD was submitted to the US FDA as an alternative epidemiological study to the originally planned US database study to fulfil the post-marketing commitment.

10.2.3. Autoimmune diseases incidence rates for UK, USA, and recorded in the CPRD GOLD

10.2.3.1. Background Tables

The background incidence rates of autoimmune diseases in the UK and USA for male and female subjects were derived from published literature and have been tabulated by Pallas in Table 3 and Table 4, 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.

In the feasibility assessment for the study, the estimates for individual autoimmune disease incidence rates were close to those published in the literature, which means the algorithms for autoimmune disease diagnosis in CPRD GOLD were correctly defined for the studied age groups (see feasibility report assessment: Report modular appendices).

Table 3 Background incidence rates and prevalence rates of neuroinflammatory/ophthalmic AD in the UK and US

Disease	Age (years)	Incidence rate UK (per 100,000/year) Or Prevalence rate UK (bold italic text; per 100,000 population)			Age (years)	Incidence rate US (per 100,000/year) Or Prevalence rate US (bold italic text; per 100,000 population)			References
		Overall	Males	Females		Overall	Males	Females	
Multiple sclerosis: Incidence	10-19		0-0.4	0.9-1.4	10-17		0	0-2.3	UK: Alonso, 2007[$]; Mackenzie, 2014; US: Mayr, 2003; Deussing, 2012; NCKP[$$]
	20-29		1.7-3.9	5.9-11.6	18-25		5.3-6.7	7.5-8.6	UK: Alonso, 2007[$]; Mackenzie, 2014; US: Mayr, 2003; Deussing, 2012; NCKP[$$]
	0-29	9.64			0-29	12.9	9.6	2.5-32.0	US: Chao, 2011
Multiple sclerosis: Prevalence	***0-14***		***0.0***	***0.0-2.6***					UK: Hirst, 2009; Gray, 2008
	10-19		***2.4***	***3.0***					UK: Mackenzie, 2014
	15-24		***0.0-9.4***	***30.3-48.8***					UK: Hirst, 2009; Gray, 2008
	20-29		***19.6***	***58.4***					UK: Mackenzie, 2014
Optic neuritis					9-26			3.9	US: Chao, 2011
Guillain-Barré syndrome					10-17			0.8-1.8	UK: Hughes, 2006
					5-17		1.1-1.9	0.8-1.2	US: Shui, 2012; Beghi, 1996; Koobatian, 1991; Riggs, 1989; NCKP[$$]
	15-24		0.6	1.1	18-25		1.4-2.2	0.4-2	
AI Uveitis					9-26			11.9	US: Chao, 2011

[$] 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)

Table 4 Background incidence rates and prevalence rates of Other AD in the UK and US

Disease	Age (years)	Incidence rate UK (per 100,000/year) Or Prevalence rate UK (bold italic text; per 100,000 population)			Age (years)	Incidence rate US (per 100,000/year) Or Prevalence rate US (bold italic text; per 100,000 population)			References
		Overall	Males	Females		Overall	Males	Females	
Systemic lupus erythematous: Incidence	10-19		0.1	2.3	9-17	1.9-6.4	0-1.8	1.5-11.0	UK: Nightingale, 2006; US: Hochberg, 1985; McCarty, 1995; Naleway, 2005; Hiraki, 2012
	20-29		0.0	4.7	18-25		1.3-1.7	5.6-19.2	
					9-26			10.3	US: Chao, 2011
Systemic lupus erythematous: Prevalence					***3-18***	***9.7***	***3.1***	***16.4***	UK: Nightingale, 2007 US: Hiraki, 2012
	0-9		***0.9-1.4***	***0.0-0.5***					
					9-11	***4.9***	***1.7***	***8.2***	
					12-14	***11.2***	***3.7***	***19.0***	
					15-17	***18.9***	***5.7***	***31.0***	
	10-19		***0.0-1.8***	***5.4-10.6***					
	20-29		***2.5-4.3***	***26.1-39.3***					
Rheumatoid arthritis	15-24		0	18.6					UK: Humphreys, 2013
					9-34		3.6	7.0-13.8	US: Myasoedova, 2010; Chao, 2011
Juvenile idiopathic arthritis: Incidence					0-15	11.9	7.7	16.4	US: Harrold, 2013
					11-15	15.4	9.2	21.7	
Juvenile idiopathic arthritis: Prevalence					***0-15***	***44.7***	***28.6***	***61.6***	US: Harrold, 2013
					11-15	***67.3***	***45.7***	***89.7***	
Immune or idiopathic thrombocytopenic purpura (ITP)	0-18		4.7	3.7	9-26			5.9	UK: Yong, 2010; US: Chao, 2011
	6-17		2.1-2.6	2.7-3.4	10-17			1.5-15	UK: Schoonen, 2009; US: Simpson, 1989; NCKP[$$]
	18-29		0.6-1.6	3.6-4.9	18-25			3.3-15	UK: Abrahamson, 2009; US: Simpson, 1989, NCKP[$$]

Disease	Age (years)	Incidence rate UK (per 100,000/year) Or Prevalence rate UK (bold italic text; per 100,000 population)			Age (years)	Incidence rate US (per 100,000/year) Or Prevalence rate US (bold italic text; per 100,000 population)			References
		Overall	Males	Females		Overall	Males	Females	
Diabetes mellitus type 1: Incidence	0-14	25.0	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; Hodgson, 2012 US: Diamond Project Group, 2006**
	10-19	35.7	35.0	26.0	10-17		6.7-33.1	6-28.2	UK: Gonzalez, 2009; Hodgson, 2012; US: Kostraba, 1992; Lipton, 1995; MacDonald, 1989; Allen, 1986; NCKP[$$]
	15-34		20.0	10.1	18-25		10.0-15.2	7.9-19.2	UK: Imkampe, 2011; US: Fishbein, 1982; Allen, 1986; NCKP[$$]
					9-26			18.0	US: Chao, 2011
Diabetes mellitus type 1: Prevalence					***0-20***		***193**#***	***193**#***	US: Dabelea, 2014
					10-14	***269**#***			
					15-19	***322**#***			
Idiopathic thrombocytopenic purpura: Prevalence					***6-11***	***7.3***			US: Segal, 2006
					11-14	***4.1***			
					15-18	***5.6***			
					19-24	***4.1***			
Combined Hashimoto's and Graves' disease: Incidence	10-14		1.1-22.0	3.4-99.0[€]	9-26			107.1	UK: Williamson, 2010; Leese, 2008 US: Chao, 2011
Inflammatory bowel disease: Incidence ***	10-19		4.2-16.4	3.4-11.8					UK: Steed, 2010; Henderson, 2012
	20-29		15.5	21.9					UK: Steed, 2010; Henderson, 2012;
	0-18		2.2-8.8	1.9-6.8	0-18	5.7			UK: Henderson, 2012; US: Deneau, 2013
Inflammatory bowel disease: Prevalence					0-18	22.3*			US: Deneau, 2013;

Disease	Age (years)	Incidence rate UK (per 100,000/year) Or Prevalence rate UK (bold italic text; per 100,000 population)			Age (years)	Incidence rate US (per 100,000/year) Or Prevalence rate US (bold italic text; per 100,000 population)			References
		Overall	Males	Females		Overall	Males	Females	
Crohn's disease: Incidence					5-9		1.1-1.5	1.0-1.4	US: Abramson, 2010; Herrinton, 2008
					10-14		4.5-5.7	3.5-5.2	
					15-19		3.9-6.8	6.1-7.4	
Crohn's disease: Prevalence					***5-9***		***2.0-7.3***	***5.6-6.2***	US: Abramson, 2010; Herrinton, 2008
					10-14		***21.1-24.7***	***14.8-19.7***	
					15-19		***34.0-45.2***	***28.5-79.2***	
Ulcerative colitis : Incidence					5-9		1.2-1.9	1.2-1.4	US: Abramson, 2010; Herrinton, 2008
					10-14		2.1-4.2	4.0-5.0	
					15-19		8.2-8.7	7.7-9.4	
Ulcerative colitis : Prevalence					***5-9***		***7.0-8.7***	***4.6-8.3***	US: Abramson, 2010; Herrinton, 2008
					10-14		***11.3-27.3***	***29.0-30.4***	
					15-19		***49.3-50.2***	***32.4-52.7***	
Autoimmune hepatitis: Incidence					0-18	0.4			US: Deneau, 2013
Autoimmune hepatitis: Prevalence					***0-18***	***3.0****			US: Deneau, 2013

** Study conducted in several parts of the UK and US. UK: the ranges represent data from Scotland (upper end), Leicestershire (lower end), Northern Ireland, Oxford, Plymouth and Yorkshire. US: The ranges represent data from Allegheny (upper end) and Jefferson (lower end);

*** 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)

*# Crude prevalence – Prevalence rates were converted to per 100,000 population by Pallas.

€ For the Leese, 2008 study, the incidence rate of 99.0 per 100,000 per year was a point estimate for hyperthyroidism from a Scottish population including adults and children. For the Williamson, 2010 study, the incidence rate of 3.4 per 100,000 per year was for the UK and Ireland population and age range of 10-14 year old girls.

10.2.4. Study Period

The cohort design for the study illustrating the different periods of subject selection is illustrated in Figure 1. Subjects receiving *Cervarix* vaccinations were followed for a period of one year with the reference date for Cervarix vaccination being a date between 01-SEP-2008 and 31-AUG-2010. For subjects in the three unexposed cohorts, either unexposed historical female, unexposed concurrent male or unexposed historical male, the reference date was chosen as explained in 10.3.4. These subjects were also followed for one year, either over the same timeframe of 01-SEP-2008 and 31-AUG-2010, or 3 years earlier for the two historical cohorts: 01-SEP-2005 to 31-AUG-2007.

10.3. Subjects

10.3.1. Number of subjects

For the cohort design, the target sample size was 65,000 subjects for each cohort. Refer to Section 10.7 for a detailed description of the estimation of the sample size.

10.3.2. Inclusion criteria

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

10.3.2.1. Inclusion criteria for the exposed female cohort

Exposed females must have satisfied ALL the following criteria at study entry:

- Female aged from 9 to 25 years at the reference date (*Cervarix* vaccination date between 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 (reference date)
- Subject defined as acceptable in CPRD GOLD

10.3.2.2. Inclusion criteria for the unexposed historical female cohort

Unexposed females must have satisfied ALL the following criteria at study entry:

- Female aged 9 to 25 years at the reference date (GP consultation between 01 September 2005 through 31 August 2007; refer to Section 10.3.4 for selection procedure)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- Subject defined as acceptable in CPRD GOLD

10.3.2.3. Inclusion criteria for the unexposed concurrent male cohort

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

- Male aged 9 to 25 years at the reference date (GP consultation between 01 September 2008 through 31 August 2010; refer to Section 10.3.4 for selection procedure)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- Subject defined as acceptable in CPRD GOLD

10.3.2.4. Inclusion criteria for the unexposed historical male cohort

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

- Male aged 9 to 25 years at the reference date (GP consultation between 01 September 2005 through 31 August 2007; refer to Section 10.3.4 for selection procedure)
- Recorded in the CPRD GOLD for at least 12 months before the reference date
- Subject defined as acceptable in CPRD GOLD

10.3.3. Exclusion criteria

10.3.3.1. Exclusion criteria for all cohorts

- Subjects with a diagnostic code of any autoimmune 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 any other cohort

10.3.3.2. Exclusion criteria for the non-exposed cohorts

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

10.3.4. Generation of the four cohorts

The exposed eligible cohort was identified based on the stepwise approach defined in Annex 5 of the Protocol. Amongst the eligible exposed subjects, 65,000 subjects were randomly selected using the RANUNI function of SAS. The RANUNI function generated streams of random numbers from an initial starting point (called *seed*) and returned numbers that were generated from the uniform distribution on the interval (0, 1). A random subject number was computed for each subject as

random_subj_number=ranuni(seed). The subjects were ordered according the *random_subj_number* and the first 65,000 subjects were included in the exposed cohort.

The unexposed eligible cohorts were identified based on the stepwise approach defined in the Protocol (Annex 5 of the Protocol).

All the unexposed subjects who matched to exposed subjects for age (birth cohort) and region (frequency matching) were identified. This represented 72 combinations of birth cohort-region (18 birth cohorts and 13 regions categorised in 4 groups in CPRD GOLD). In each combination, the subjects were randomly selected based on the distribution in the exposed cohort.

The reference date was the first *Cervarix* vaccination date for subjects in the exposed cohort between 01-SEP-2008 and 31-AUG-2010. For the three unexposed cohorts, one reference date was selected per subject. The reference date for the subjects in the three unexposed cohorts was a random date selected among the reference dates of the matched exposed subjects, but the random selection was conducted so that there was an even distribution of reference dates across the time period within each of the four cohorts. For the concurrent male cohort, the reference date was between 01-SEP-2008 and 31-AUG-2010. For the historical non-exposed female and male cohorts, the reference date was between 01-SEP-2005 and 31-AUG-2007 (reference dates for exposed cohort minus 3 years).

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

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

10.4. Variables

10.4.1. Primary endpoint

- Occurrence of new onset of confirmed[2] 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
    - Autoimmune (AI) peripheral neuropathies and plexopathies (including chronic inflammatory demyelinating polyneuropathy, multifocal motor neuropathy and polyneuropathies associated with monoclonalgammopathy).
  - AI uveitis

  [2] **Other autoimmune diseases:**

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

[2] Autoimmune disease diagnosis ascertainment by an expert physician panel (Section 10.5.1).

– AI hepatitis

10.4.2. Secondary endpoints

Secondary endpoint was 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 10.4.1), autoimmune uveitis, systemic lupus erythematous (SLE), rheumatoid arthritis (RA), juvenile rheumatoid arthritis (JRA), Still's disease, psoriatic arthritis, ankylosing spondylitis, type 1 diabetes mellitus, autoimmune 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*.

10.5. Data sources and measurement

10.5.1. Data source for case ascertainment

CPRD GOLD is based on data from GPs, while most autoimmune diseases are probably diagnosed in specialist settings. Consequently, the number of autoimmune diseases, the quality of the information, and the diagnostic certainty might be limited. In particular the specific information related to the onset of clinical symptoms, and radiological and biological data associated with the etiologic diagnosis of autoimmune diseases may not all be available in the CPRD GOLD database and associated resources. Besides, not all general practices participating to CPRD GOLD consented to the linkage between CPRD GOLD primary care data and HES data . Specific algorithms for each outcome of interest were developed (See the Protocol: Annex 5), individual subject profiles were produced, and available "free text" that related to autoimmune disease diagnoses were requested from CPRD GOLD when needed.

Expert case review of medical records has been proposed for autoimmune safety studies, and case identification could be expanded by use of laboratory test results and other relevant measures in addition to specific ICD-10 diagnosis codes [Chao, 2012]. In the current study, free text in CPRD GOLD was searched to gather information on clinical symptoms, medical diagnosis, suitable laboratory test results, other diagnostic test results and data on possible first symptoms to ascertain if the specific ICD-10 codes were correct. The case ascertainment process which is explained below was applied consistently to enable the maximum number of cases to be confirmed, although only a proportion of identified cases were eventually confirmed.

The ascertainment of the etiologic diagnosis and date of disease onset for all identified autoimmune diseases was performed by Pallas, a GSK identified reviewer (Pallas, health research and consultancy) to ensure the correct classification of each case. Pallas was asked to review all subject profiles and to collect relevant information from these profiles to import into the Remote Data Entry (RDE) computer application. The subject profile consisted of data retrieved from CPRD GOLD including clinical diagnosis, laboratory testing, drug prescription, and HES data (including specific ICD-10 diagnostic codes). The relevant free-text was identified by Pallas and requested when necessary for disease confirmation. The free-text request was sent by GSK to CPRD GOLD and extracted free-text was de-identified by the CPRD GOLD research group before sending to GSK. The free-text was reviewed by Pallas. Pallas assessed whether the aetiology of the autoimmune disease was confirmed or not and whether the date of disease diagnosis fell within the observation period of the study, which was one year [3] after the reference date. The GSK safety physician and a medical Epidemiologist at RTI Health Solutions reviewed all subjects for which Pallas had some doubts.

In the event that the (date of) diagnosis, first symptom and/or first abnormal laboratory test could not be confirmed, a second review step was to be conducted with an expert physician panel as described in Section 10.5.3.

A report was prepared by Pallas providing the decisions taken after reviewing the subject profiles, review of the requested free-text, describing how the expert review was conducted and how relevant data was imported in RDE – see Pallas Methodology Report (Report Annex 5).

The final study database consisted of data extracted from CPRD GOLD, HES and additional data from free-text review. The final study database was frozen on 26th August 2014 and stored by GSK Biologicals' data management.

10.5.2. Collected data

10.5.2.1. Subjects characteristics

The following data were extracted for the analysis population (study database population):

- Demographic characteristics: birth month and birth year, sex, region, practice region identifier, date of death (if applicable) and CPRD patient acceptability flag. Patients are labelled 'acceptable' if their data did not contain obvious errors such as first registration date being empty, year of birth labelled as missing, current registration date in CPRD GOLD being before first registration date, gender being listed as neither female, male or indeterminate etc. For a full explanation of this acceptability

[3] Note: for the purpose of the Self-Control Case-Series analysis the observation period was extended to 30 months after reference date for the cases exposed to Cervarix vaccination.

flag please refer to the Data Quality Section (pages 3-4) of the publication by [Williams, 2012].

- 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 until the end of follow-up was collected: date of vaccination, medcodes and immunization type were extracted from the immunisation file.
- Health care resource utilization: number of times a primary care resource was utilized during the year before the reference date (it refers to all available information for a patient in the database including GP visits, prescriptions, laboratory tests, etc.).

10.5.2.2. Clinical outcomes

Occurrence of autoimmune diseases defined as study endpoints were identified in CPRD GOLD by using defined algorithms (see the protocol: Annex 5).

For each case, the following data were extracted - from the reference date up to end of the follow-up period (12 months for non-exposed cohorts and 30 months for exposed female cohort)-:

- Medical code(s)
- Date of event

The algorithms were able to identify and extract free-text identifiers. A preliminary review step of the subject profiles (all records from one year before reference date to the end of the follow-up period) was performed by Pallas, which was the GSK-identified reviewer, in order to identify all relevant free-texts, which were additional to those pre-identified by the algorithm. For all the identified cases of autoimmune diseases, the subject profile and associated free-text were reviewed by GSK-identified reviewers for confirmation and determination of the aetiology and the date of first symptoms of AD.

10.5.2.3. Date of first symptom

The first clinical sign or symptom from the patient profiles or associated free text was considered as the onset of the disease.

The following rules were applied for the determination of the symptom date:

1. If there was no free text belonging to the event considered to be the first symptom, the event date of that first symptom was used;
2. If some free text was available with a time indication, the date of first symptom was derived according specific rules - see Pallas Methodology Report (Report Annex 5).

10.5.2.4. Other derived variables

The following variables were derived from the CPRD GOLD data:

- If the birthdate day was missing, the Subject's date of birth was defined as the 15th of the birth month and birth year. If the birth day and month were missing, the birth date was defined as the 30th of June of the birth year.
- Incomplete dates (except for vaccination date which was the reference date) were substituted as follows for calculation of age and/or time to event; if the day was missing the date was defined as the 15th of the month, if both the day and the month were missing, the date was defined as 30th of June of the year.
- Age at a specific event was computed as the difference between the date of the event and the date of birth.

10.5.3. Final case ascertainment step

After the first patient profile review which was performed by Pallas, the GSK safety physician and a medical Epidemiologist from RTI, individual data with a doubt on the aetiology were reviewed by an expert. The panel of experts consisted of five physicians in the fields of rheumatology, neurology, ophthalmology and internal medicine. Each physician reviewed the case profiles according to their specialty, which included data of first symptom, date of diagnosis, date of first abnormal laboratory test and final classification of AD. Additionally among the autoimmune disease cases with no doubt on the aetiology, a random 10% sample from each of the categories of autoimmune disease were reviewed by the experts as a quality check. During this quality check, agreement on the date of first symptom, date of diagnosis and classification of AD existed for all rheumatology and neurology cases and for most of the ophthalmology and internal medicine cases. For autoimmune uveitis, however, the expert decided to include an additional symptom (i.e. conjunctivitis/episcleritis) that was not used by Pallas, GSK and RTI until then. For IBD, CD and UC, the expert suggested other criteria to determine the date of diagnosis and confirmation of the diagnosis. All uveitis, IBD, CD and UC cases were therefore reviewed again by Pallas, applying the new rules accordingly. Furthermore, after review by the expert of the SLE subjects, the expert proposed other criteria to determine the diagnosis and its confirmation. The expert reviewed all remaining cases and applied these criteria.

For further details of the expert review of cases, please refer to the Pallas Methodology Report – Section 5.1 (p19-21) which is located in Report Annex 5. The experts were blinded with regards to HPV vaccine exposure, and their opinions were captured into RDE.

10.6. Bias

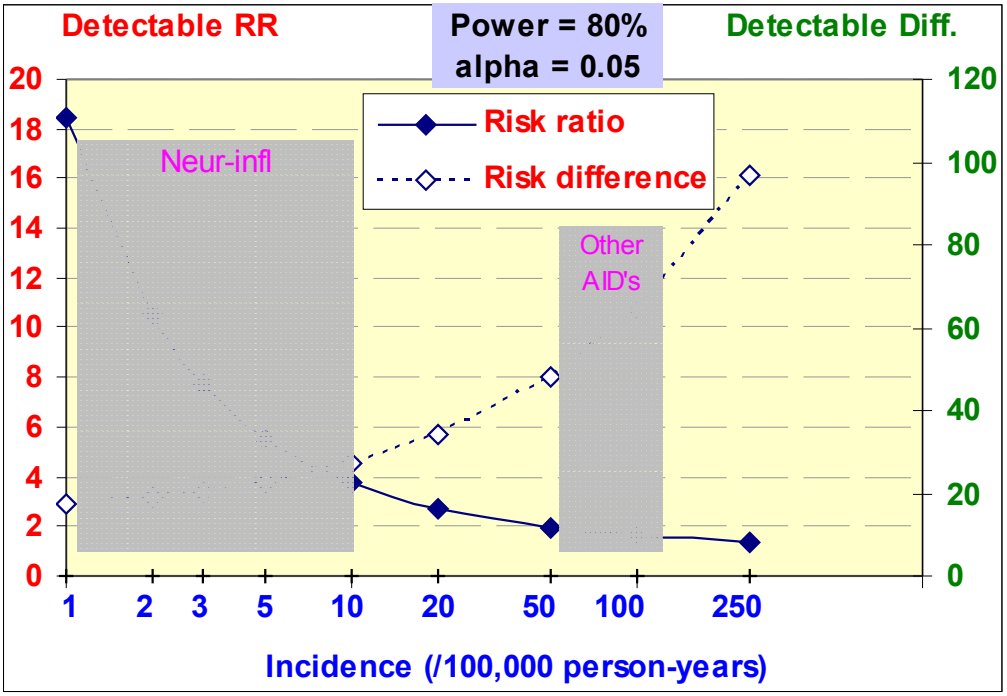
Please refer to Section 12.2 Study limitations.

10.7. Study size

10.7.1. Sample size for cohort design

The target sample size was 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 AD (incidence rate between 10 and 1/100,000 person-years) and between 1.6 and 2.0 for other AD (incidence rate between 100 and 50/100,000 person-years).

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

10.7.2. Sample size for self-controlled case-series

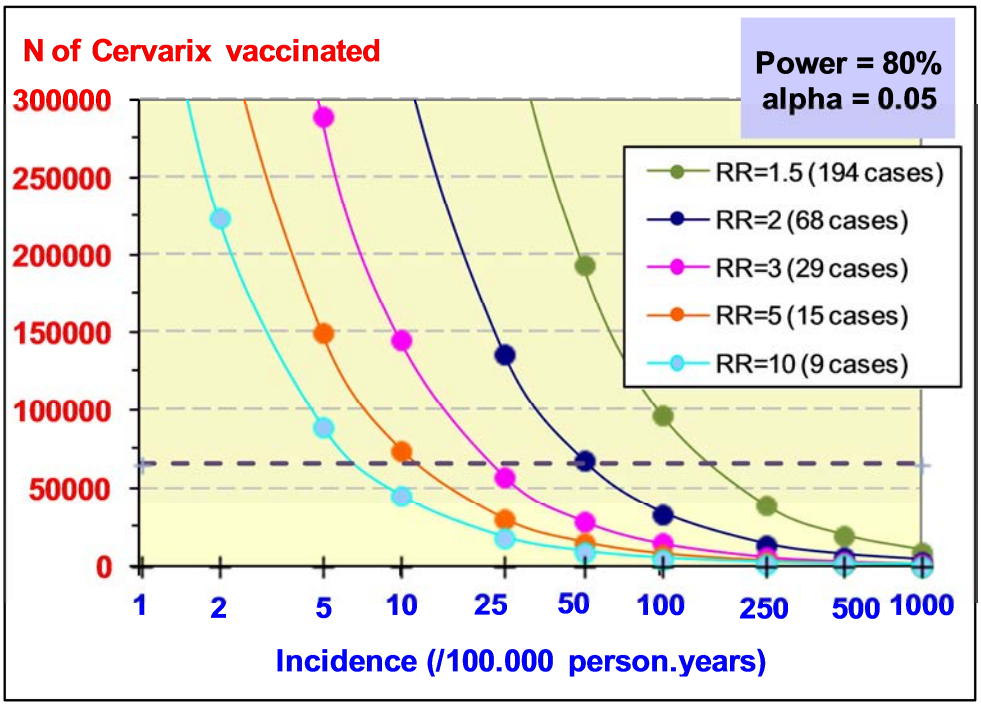
The power of the SCCS analysis depended on the number of cases and the ratio between the duration of the risk and the control periods. With a risk and a control period of 12 months each, the number of cases needed versus the detectable risk ratio (incidence rate ratio) is summarized in Table 5. Figure 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 5 Sample size for a SCCS analysis - Number of cases in vaccinated subjects versus the incidence rate ratio that could be detected [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

10.8. Data transformation

The CPRD GOLD data source was used in order to create the final study database. Specific study variables (please refer to the study Case Report Form) were defined based on the original database using algorithms and manual data review.

10.9. Statistical methods

10.9.1. Main summary measures

10.9.1.1. Subject disposition

Subject disposition was summarized for each cohort by computing:

- Number of screened subjects.
- Number (%) of non-eligible subjects for each of the following reasons of non-eligibility:
  - Diagnostic code of AD 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);
  - Subjects non-included after frequency matching for birth cohort and practice-region.
- Number of included subjects in each cohort.

A detailed, comprehensive list of reasons for elimination for the analyses was established at the time of data cleaning.

10.9.1.2. Case definitions

Three case definitions were defined leading to three analyses:

1. Main analysis included all the autoimmune cases with a known date of first symptom;
2. Imputed dates analysis included all the autoimmune cases with a known date of symptom plus all the autoimmune cases with an unknown date of symptom. In case of missing/unknown date of onset (date of first symptom), a date was imputed. The imputed date was computed as the date of diagnosis minus the median number of days between the date of diagnosis and the known date of

disease onset. The imputation was performed specifically for each disease and used all the known symptom dates (confirmed and non-confirmed cases) for computation of the median number of days;

3. Analysis based on known date of diagnosis.

10.9.1.3. Demographic and baseline characteristics

Demographic and baseline characteristics of the overall study population and autoimmune disease cases (age at reference date, region [GP practice], availability of data in CPRD GOLD [follow-up time within CPRD GOLD at reference date, proportion of HES linkage] and healthcare resources utilization the year prior to the reference date) were summarized per cohort and overall, using descriptive statistics: n of subjects, mean, standard deviation (SD), median, minimum and maximum; or (n, %).

Exposure to other vaccines was summarised in frequency tables (n, %) per cohort and period (the year prior the reference date and during the follow-up period).

Number of cases with an autoimmune disease drug prescription was tabulated per cohort and overall (n, %).

The two female cohorts and the two male cohorts were compared for their demographic and baseline characteristics using a Chi-Square test or Student t-test.

The four cohorts were compared for the age at reference date using a one-way Analysis of Variance (ANOVA).

10.9.1.4. Analysis of autoimmune diseases - co-primary endpoints and individual diseases

The proportion of autoimmune diseases (co-primary endpoints and individual diseases) was calculated for the exposed female cohorts and the non-exposed cohorts, as the number of autoimmune cases divided by the total number of included subjects in each cohort.

The incidence rates of autoimmune diseases (co-primary endpoints and individual diseases) during the 1-year follow-up period were calculated by dividing the number of cases by the total person-time. The comparison of the incidence rates of autoimmune diseases was done using a Poisson regression model (see Section 10.9.2.3. The incidence difference was calculated as incidence of exposed cohort minus incidence of non-exposed cohort. These analyses were also performed for the two male cohorts (concurrent and historical cohorts).

The incidence of autoimmune diseases (co-primary endpoints and individual diseases) was calculated after each *Cervarix* dose for the exposed female cohort by dividing the number of cases by the total person-time. The person-time was defined as the period between the *Cervarix* dose and the end of risk period (6 months after each dose), date of the next *Cervarix* dose, subject's date of death, date of unspecified HPV vaccine or *Gardasil*, or date of autoimmune diseases onset whichever occurred first.

A Self-Controlled Case-Series (SCCS) analysis for both co-primary endpoints and individual diseases was performed for the exposed female cohort. The SCCS is only based on cases, and provides consistent estimates of the relative incidence. This method controls implicitly for potential confounders which do not vary with time. The relative incidence was calculated as the ratio of the incidence rate during the risk period and the incidence rate during the control period.

Sensitivity analyses were performed as described in Section 10.9.2.3.

An exploratory analysis was performed assessing the presence of a temporal clustering of the individual AD or the two co-primary endpoints during the follow-up period for each cohort (30 months for the exposed female cohort and 12 months for non-exposed cohorts). A variable time window was used with respectively, a maximal scan window of 60 days (2 months) and 120 days (4 months), and a time aggregation of 1 day was applied. Poisson-based likelihood was used to compare the expected number of cases and actual number of cases within the window. Significance was evaluated with 9999 replications. The analysis was performed by using the SatScan tool (SaTScan version 9.3.1, 2014).

All the analyses were performed for the two co-primary endpoints and for the individual disease if more than 10 cases were observed in both female cohorts.

The individual disease analyses (main and sensitivity) were performed according to a time schedule that was disease-specific: 2 months for GBS and Haemolytic anaemia, 6 months for ITP and 12 months for all the other autoimmune diseases.

10.9.2. Main statistical methods

10.9.2.1. Hypotheses

10.9.2.1.1. 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.

Alternative hypothesis (H1): the incidence of neuroinflammatory/ophthalmic autoimmune diseases (other autoimmune diseases) in the exposed female cohort is not equal to the incidence in the historical non-exposed female cohort.

10.9.2.1.2. Hypotheses for the self-controlled case-series analysis

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

These hypotheses were tested separately for each of the two co-primary endpoints. These hypotheses were also tested for individual diseases if more than 10 cases were recorded in both exposed and non-exposed cohorts.

10.9.2.2. Statistical calculations

All the statistical calculations were done in SAS 9.2.

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

10.9.2.3. Statistical models

Poisson regression

The dependent variable was the number of events. The log transformed person-time was included in the model as an offset. The end of the follow-up time was defined as the end of study period (12 months after reference date), date of death of the subjects, date of unspecified HPV vaccine or *Gardasil* or *Cervarix* vaccine (for non-exposed cohort) or the date of disease onset whichever occurred first.

The main model (Model 1) included the exposure status (exposed vs. non-exposed) as a binary independent variable and the age in groups ([9-17],[18-25]) as a covariate.

The incidence risk ratio (exposed/non-exposed) was derived as the exponential of the coefficient associated with the exposure status and its 95% Wald confidence interval (CI).

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

The following models were computed as sensitivity analyses:

Model 2: A Poisson regression model including, in addition to the exposure status and age-group, other covariates:

- Region (class variable);
- Vaccination during the year prior to reference date (binary variable);
- Use of healthcare resources during the year prior to reference date (4 classes according to quartile).

Model 3: A Poisson regression model adjusted for time effect was performed where a significant difference in incidence rates was observed between the two male cohorts. This model is identical to Model 1 with the inclusion of a contrast for the temporal effect.

Model 4: A Poisson regression model stratified by age group ([9-17] and [18-25]) was performed. The model included the event as dependent variable and the exposure status as binary independent variable plus log transformed person-time as an offset.

Self-controlled case-series

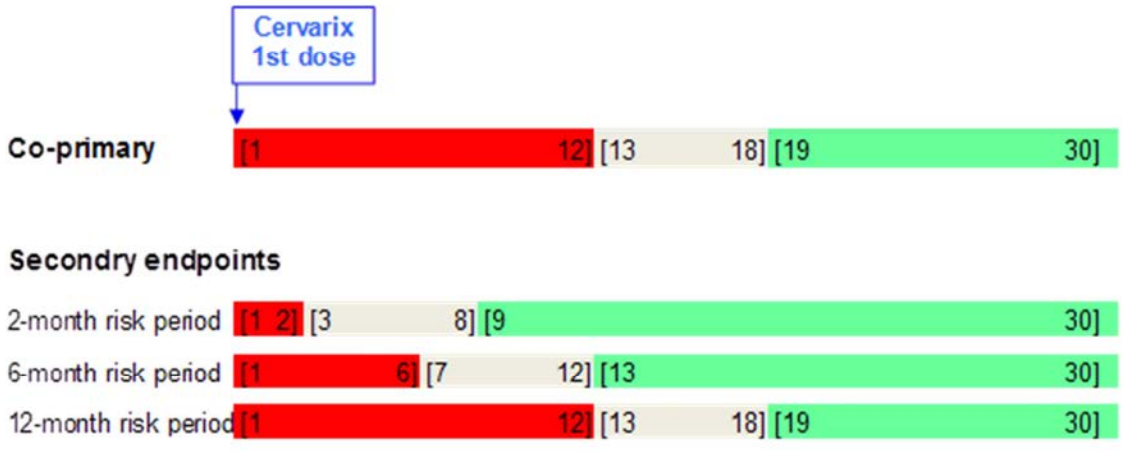
The statistical calculation was done by using the specific SAS macro developed by Whitaker et al. [Whitaker, 2006] and available online from http://statistics.open.ac.uk/sccs.

For the main SCCS the risk period was defined as (reference date +364 days), the buffer period was defined as (end of the risk period + 180 days) and the control period was defined as (end of the buffer period + 365 days). The relative risk was calculated for the co-primary endpoints and for the diseases with more than 10 cases in both risk and control periods as the ratio of the incidence in the risk period versus incidence in the control period.

For the sensitivity SCCS the risk period was defined as a period from the first dose of Cervarix until 6 months after the last dose. A buffer period covered a period of 6 months after the end of risk period, the control period starting after the buffer period until 30 months after the first dose of Cervarix (910 days). The relative risk was calculated between risk and control period.

An exclusion criterion was applied on the cases, if the disease onset was within the defined risk period but the date of diagnosis was after the risk period then the case was excluded from the SCCS analysis. A similar rule was applied for cases occurring in the control period. The reason of that rule was to avoid a bias in the number of cases occurring in the risk period. Indeed the exposed cohort was followed up to 30 months in the CPRD database, with this FU period we could expected to have more cases with a first symptom date in the risk period than in the control period.

Figure 5 Risk and control periods for the various endpoints



10.9.3. Missing values

Missing data was not substituted, except for the unknown date of first symptom where an imputation method was applied (see Section 10.9.1.2).

10.9.4. Amendments to the statistical analysis plan

- The formula for the incidence rate calculation adjusted for temporal effect was corrected as follows:

  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 +1/ estimate=exp;
  RUN;

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

- Categorization of variables:
  - Three regional categories were created for inferential statistics:

  1. North England (North East - North West - Yorkshire & The Humber);
  2. South England (East of England - South West - Central England - London - South Coast) and;
  3. Midlands (East Midlands - West Midlands) & Northern Ireland – Scotland – Wales.

Further to presentation of the results of the planned analyses at the GSK Safety Review Committee, a few additional analyses (post-hoc) were requested:

- Descriptive analysis of time-to-onset of all the confirmed cases and especially the autoimmune diseases cases;
- Descriptive analysis of the geographical distribution of the cases;
- An additional patient profile review was conducted for the confirmed autoimmune thyroiditis cases and included in the main analysis to clarify the diseases aetiology (hypo-or hyperthyroiditis). The incidence risk ratio for hypo-/hyperthyroiditis was calculated. The analysis based on the cases with a known date of onset was repeated for the cases of hypothyroiditis;
- The incidence rate ratios excluding the region Northern Ireland – Scotland - Wales were calculated;
- The analysis based on the cases with a known date of onset was repeated on the two co-primary endpoints and the three selected individual diseases after exclusion of subjects from Northern Ireland, Scotland and Wales.

10.10. Quality control

Validation of clinical outcomes is described in Section 10.5.2 and Section 10.5.1.

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

The final study dataset was archived and stored on a secured, access limited, computer platform SAS Drug Development (SDD) according to GSK Biologicals Standard Procedures. Specific statistical programs were to 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 was to be documented. All statistical programs, output files and QC documentation were to 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) were to be archived on a Document management system based on the Documentum platform: Computer Aided Regulatory Submission (CARS).

10.10.1. Remote Data Entry instructions

Remote Data Entry (RDE) using a validated computer application was used by the GSK identified reviewer to enter the information obtained from different data sources, and conduct the final case ascertainment classification and validation by the experts (when necessary).

Subject data necessary for analysis and reporting were entered/transmitted into a validated database or data system. Clinical data management was performed in accordance with applicable GSK standards and data cleaning procedures.

No monitoring was done. The GSK identified reviewer remained accountable for the data entry.

10.10.2. Final study database

The final study database consisted of data extracted from CPRD GOLD, HES and additional data from free-text review. The study database was locked and stored by GSK Biologicals' data management according to GSK Biologicals Standard Procedures.