Study Report

P2-C3-004

DARWIN EU® - Effectiveness of Human Papillomavirus Vaccines (HPV) to prevent cervical cancer

05/11/2025

Version 8.0

CONTENTS

Study Title	DARWIN EU® - Effectiveness of Human Papillomavirus Vaccines (HPV) to prevent cervical cancer
Study Report Version	V8.0
Date	05/11/2025
EUPAS number	EUPAS1000000080
Active substance	Bivalent HPV vaccine (types 16, 18) Quadrivalent HPV vaccine (types 6, 11, 16, 18) 9-valent HPV vaccine (types 6, 11, 16, 18, 31, 33, 45, 52, 58)
Medicinal product	Cervarix Gardasil/Silgard Gardasil-9
Research question and objectives	What is the effectiveness of HPV vaccination in prevention of severe disease outcomes in women, including invasive cervical cancer and CIN2+ for the different licensed HPV vaccines in Europe. More specifically the study objectives are: Main objectives: 1. To assess the effectiveness of HPV vaccination in prevention of invasive cervical cancer, stratified by licenced vaccine brand. 2. To assess the effectiveness of HPV vaccination in prevention of CIN2+, stratified by licenced vaccine brand. 3. To assess the effectiveness of HPV vaccination in prevention of conisation, stratified by licenced vaccine brand. Secondary objectives: • To assess the effectiveness of HPV vaccination regardless of the brand or schedule for each of the three outcomes (i.e. invasive cervical cancer, CIN2+ and conisation) • To assess the effectiveness of HPV vaccination in prevention of invasive cervical cancer, CIN2+ and conisation separately in subgroups defined by number of doses, within each brand.
Country(-ies) of study	UK, Spain, Norway
Author	Daniel Prieto Alhambra, Albert Prats Uribe

1. DESCRIPTION OF STUDY TEAM

Study team role	Names	Organisation
Study Project Manager/Principal Investigator	Daniel Prieto Alhambra Albert Prats Uribe	University of Oxford
Data Scientist	Mike Du Marti Catala Sabate	
Epidemiologist	Daniel Prieto Alhambra	
Clinical Domain Expert	Albert Prats Uribe	
Data Partner*	**Names**	**Organisation**
SIDIAP	Talita Duarte Salles	IDIAP JGOL
SIDIAP	Anna Palomar	IDIAP JGOL
SIDIAP	Agustina Giuliodori Picco	IDIAP JGOL
CPRD GOLD	Antonella Delmestri	University of Oxford
NLHR	Hedvig Marie Egeland Nordeng	University of Oslo
NLHR	Nhung Trinh	University of Oslo

*Data partners' role is only to execute code at their data source, review and approve their results. They do not have an investigator role. Data analysts/programmers do not have an investigator role and thus declaration of interests (DOI) for them is not needed.

2. DATA SOURCES

Country	Name of database	Health Care setting	Type of data	Number of active subjects	Calendar period covered by each data source
United Kingdom	CPRD-GOLD	Primary care	EHR	17M	Sept 1987 – Dec 2023
Spain	SIDIAP	Primary care + linkage to hospital data	EHR	5.8M	Jan 2006 – June 2023
Norway	NLHR	Primary care + linkage to hospital data + vaccination registry	Linked Health Registry	5.7M	Jan 2018 – Dec 2023

3. ABSTRACT

Title

DARWIN EU® - Effectiveness of Human Papillomavirus Vaccines (HPV) to prevent cervical cancer

Rationale and Background

HPV vaccination programmes have been shown to reduce not only HPV infection but also the incidence of pre-cancerous lesions and cervical cancer. However, there is limited evidence on the real-world effectiveness of the different vaccine brands and dose schedules.

Research Questions and Objectives

To generate evidence from real-world data on the effectiveness of HPV vaccination in preventing severe disease outcomes in women, including invasive cervical cancer and CIN2+, for the different licensed HPV vaccines in Europe.

More specifically the study objectives are:

Main objectives:

1. To assess the effectiveness of HPV vaccination in prevention of **invasive cervical cancer,** stratified by licenced vaccine brand.
2. To assess the effectiveness of HPV vaccination in prevention of **CIN2+**, stratified by licenced vaccine brand.
3. To assess the effectiveness of HPV vaccination in prevention of **conisation**, stratified by licenced vaccine brand.

Secondary objectives:

- To assess the effectiveness of HPV vaccination regardless of brand for each of the outcomes separately (i.e. invasive cervical cancer, CIN2+ and conisation).
- To assess the effectiveness of HPV vaccination in prevention of invasive cervical cancer, CIN2+.

Study Design

New user matched cohort study. This study included data sources from UK (CPRD-GOLD), Spain (SIDIAP), and Norway (NLHR).

Population

We included all females born on or after 1993 (i.e. 15 years old or less in 2008, the year of the earliest launch of the vaccine in all countries included in the analysis). We then restricted to those in observation in the databases at least between 9 and 15 years old.

Further restrictions were made on a year-by-year basis. For each 1st of January, participants needed to: be in observation on that day, have at least 365 days of prior observation available, and be aged between 9 and 15 years old.

Setting

Data Sources:
Primary care records from the UK (Clinical Practice Research Datalink (CPRD-GOLD) and primary care records linked to hospital records from Catalonia, Spain (Information System for Research in Primary Care

(SIDIAP)), Population-based health registry data from Norway (NLHR) Norwegian Linked Health Registry data.

Study Period:

The study period began on the 1st of January 2008 as this date corresponds to the start date of the earliest roll-out of the HPV vaccination programme in these countries. For NLHR, the study period started in 2018 due to lack of prior data. For all databases, the end of the study period corresponded to the most recent data available, i.e., 2023.

Eligibility criteria

Females between 9 and 15 years old at any date after the launch of the vaccination programme in the corresponding country.

Follow-up

Follow up started at the administration of the first dose before the age of 15 years. For unvaccinated participants, the follow up started at the same date as their vaccinated matched counterpart. Follow-up extended until another vaccine dose or outcome event occurred, end of available follow-up, or death of any individual of the matched pair, whichever comes first.

Variables

Exposure

Assignment procedures: Vaccination status (brand and number of doses) was assigned as seen in the data at 15 years old. Unvaccinated status was assigned as not being vaccinated at 15 years old and censored if they get vaccinated later on.

Brand: For those vaccinated, brand was primarily assigned as brand of all the doses administered before 15 years old. Women with heterologous schedules (not the same brand for each dose) were excluded. If this information was not available, it was inferred, when possible, using each country's vaccination schedules.

Schedules: Unvaccinated, vaccinated with 1 dose, vaccinated with 2 doses, and vaccinated with 3 doses.

Outcome

The main outcome of interest was invasive cervical cancer. Two secondary outcomes were also considered: CIN2+ and Conisation.

Other variables

Year of birth, calendar year, age at vaccination, cytology results from smear test prior to the first dose of vaccine if available. For LASSO regression (propensity score estimation), all recorded features recorded in the database were included (i.e., socio-demographics, geographic location, healthcare resource use (measured as number of visits on the prior year), comorbidity, medicine/s use, previous smear testing, and number of previous vaccination/s).

Treatment of Intercurrent events:

For the unvaccinated, vaccination was dealt with a hypothetical strategy. To implement this, data from women in the unvaccinated group that received a vaccine after 15 years old was included in the analysis up to the time of vaccination. An additional sensitivity analysis not censoring women who at the time of receiving the vaccination were over 15 years old was performed.

For the vaccinated, treatment discontinuation (i.e. not receiving all scheduled doses) was dealt with a treatment policy strategy: All available data from these women were included in the analysis regardless of the number of doses received after 15 years old.

Data Analysis

We conducted a matched cohort design, where target and comparator cohort participants were matched up to 5:1. First, matching was done year by year based on year of birth and geographic region using nearest neighbour matching, with calliper width 0.2 standard deviations as is standard for propensity score matching. Large-scale PS was estimated for each person at the start of each year, using LASSO regression to estimate the probability of being in the target cohort (as specified below). After that, participants were also matched yearly on PS.

The following matched cohorts were compared:

Main comparisons:

Vaccinated vs unvaccinated per brand:

- Vaccinated with Gardasil/Silgard (target) (1 or more dose) vs unvaccinated (comparator)
- Vaccinated with Cervarix (target) (1 or more dose) vs unvaccinated (comparator)
- Vaccinated with Gardasil-9 (target) (1 or more dose) vs unvaccinated (comparator)

Secondary comparisons:

Vaccinated (target) (1 or more dose) (any brand) vs unvaccinated (comparator) overall.

Dose comparisons:

- Vaccinated with 2 or more doses (target) vs 1 dose (comparator) of the same brand.
- Vaccinated with 3 or more doses (target) vs 2 doses (comparator) of the same brand.

Vaccine effectiveness analyses

Incidence rates and incidence rate ratios (IRR) were calculated for the matched cohorts and outcomes at 5, 10, and 15 years of follow-up after index date (when available). Cox proportional hazard models were used to calculate hazard ratios (HR) for time-to-event analyses. Analyses were conducted separately for each database, and carried out in a federated manner, with effectiveness estimates meta-analysed and the I2 heterogeneity coefficient reported.

Subjects and study size

Restricting the population to those women in the study period and who were observed from 9 to 15 years old resulted in 191,376 vaccinated and 142,607 unvaccinated at 15 years old in CPRD-GOLD, in 262,364 vaccinated and 40,195 unvaccinated in SIDIAP, and 116,271 vaccinated and 176,133 unvaccinated in NLHR. After restricting to those for whom we were able to match 1:5 based on year of vaccination, year of birth, region or GP, and PS we achieved the “PS Matched” cohorts including 81,863 vaccinated and 46,357 unvaccinated in CPRD-GOLD; 148,214 vaccinated and 39,952 unvaccinated in SIDIAP; and 14,885 vaccinated and 4,073 unvaccinated in NLHR.

Results

In the CPRD-GOLD, vaccine coverage started very low for those born between 1993-1995 but rises to over 60% for those born after 1995, remaining stable between 60-70% coverage until it drops to 55.6% for those born in 2007-2008. In SIDIAP, coverage is consistently high, starting at 83.1% for the 1997 birth cohort and steadily increasing to 94.3% for those born in 2008. For NLHR, coverage data begins in 2005 at 80.1% and remains high, reaching 85.3% by 2005.

After matching, we arrived at a cohort of 81,863 vaccinated and 46,357 unvaccinated in CPRD-GOLD, 148,214 vaccinated and 39,952 unvaccinated in CPRD-GOLD, and 14,885 vaccinated and 4,073 unvaccinated in NLHR. The mean age at first vaccination was 13 years in CPRD-GOLD, 11 in SIDIAP, and 12 in NLHR. After vaccination, women were followed for an average of 7 years in CPRD-GOLD, 10 years in SIDIAP, and 5 years in NLHR, with maximum follow-up periods of 16, 15, and 6 years, respectively.

There were less than 5 invasive cancer cases in each of the databases, precluding the analysis of this outcome. The analysis focused on CIN2+ and conisation.

VE estimates regardless of brand against CIN2+ were of 41% in CPRD-GOLD and 42% in SIDIAP, with a metanalysis estimate of 42%. VE in NLHR could not be calculated in matched cohorts due to the low number of outcomes. Against conisation, VE in CPRD-GOLD was 41%, and in NLHR, with lower follow-up, the VE was of 60%. By brand, Cervarix had a metanalysis VE estimate of 38% against CIN2+ and 55% against conisation. Gardasil showed a metanalysis estimate against CIN2+ of 41% and 51% against conisation. Due to the small number of cases, it was not possible to conduct dose comparison analyses.

Discussion

We were unable to assess the causal effect of HPV vaccines against cervical cancer using a target trial emulation design due to the limited number of cases and limited available follow-up to account for the long cancer latency period post-vaccination. Vaccine effectiveness in preventing CIN2+ and conisation shows similar results in this European population to those seen in clinical trials (mostly conducted outside of Europe), although in the lower range. Our effectiveness data is potentially underestimated due to lower screening rates in the unvaccinated population.

4. LIST OF ABBREVIATIONS

Abbreviation	Name
ASMD	Absolute Standard Mean Difference
CDM	Common Data Model
CIN	Cervical Intraepithelial Neoplasia
CIN2	Cervical intraepithelial Neoplasia grade 2
CIN3	Cervical intraepithelial Neoplasia grade 3
CPRD-GOLD	Clinical Practice Research Datalink
EHR	Electronic Health Record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
HPV	Human papillomavirus
IRR	Incidence rate ratio
NLHR	Norwegian Linked Health Registry
LSPS	Large-scale propensity scores
OMOP	Observational Medical Outcomes Partnership
SIDIAP	The Information System for Research in Primary Care
SNOMED	Systematized Nomenclature of Medicine
VE	Vaccine Effectiveness

5. AMENDMENTS AND UPDATES

The analyses described in the protocol have been revised to incorporate emigration data from the Norwegian Linked Health Registry (NLHR). This has resulted in updated study findings. Please see section **13. Deviations from original protocol** for further details.

6. MILESTONES

Study deliverable	Timeline (planned)	Timeline (actual)
Draft Study Protocol	October 2023	14th of January 2024
Final Study Protocol	January 2024	January 2024
Creation of Analytical code	January-July 2024	January-June 2024
Execution of Analytical Code on the data	February-July 2024	July 2024
Draft Study Report	31st July 2024	09th Aug 2024
Final Study Report	10th December 2024	10th December 2024
Amended Study Report with updated NLHR data	21st October 2025	21st October 2025

7. RATIONALE AND BACKGROUND

Cervical cancer ranks as the second most common cancer among women aged 15 to 44 years in the European Union (EU) and England (1, 2). Annually, there are approximately 33,000 patients diagnosed with cervical cancer in the EU, resulting in 15,000 fatalities (2). The primary cause of cervical cancer is persistent infection of the genital tract by specific strains of human papillomavirus (HPV). There are over 100 strains of HPV, 40 of which can infect the genital tract, and at least 14 of which are considered ‘high risk’ for cervical cancer. Around 70% of cases of cervical cancer are caused by HPV types 16 and 18 – the most common ‘high risk’ strains (2).

In 2018 the World Health Organisation (WHO) launched the ‘Cervical Cancer Elimination Initiative’ which has accelerated the implementation of HPV vaccination programmes (3). As a result, HPV vaccines are now licenced in more than 100 countries worldwide. There are currently three highly efficacious prophylactic vaccines that are approved for use in Europe and the UK: a bivalent (Cervarix), a quadrivalent (Gardasil/Silgard), and a 9-valent (Gardasil-9). Clinical trials have demonstrated each of these to provide protection against HPV-associated anogenital disease, including genital warts, intraepithelial neoplasia, and cervical cancer (4-6). Each of these protect against the most carcinogenic HPV strains, 16 and 18, and the quadrivalent and 9-valent vaccines provide additional protection against strains 6 and 11, which are typically responsible for non-cancerous genital warts, and the 9-valent against strains 31, 33, 45, 52 and 58, which have been associated with 20% of cervical cancers (7).

HPV vaccines provide greater advantages and enhanced protection when administered to preadolescent individuals. This is because the vaccines are more effective in individuals not previously exposed to the HPV types included in the vaccines, and research has shown that preadolescents tend to have a more robust immune response compared to adults (8).

Each of the 3 HPV vaccines are approved for use in females and males from the age of 9 years to protect against precancerous lesions (intraepithelial neoplasia) and cervical cancer (2). Males and females aged 9-13 years (Gardasil) or 9-14 years (Cervarix and Gardasil-9) are typically given two doses; and those aged 14+ (Gardasil) or 15+ years (Cervarix and Gardasil-9) a three-dose schedule (9-11).

Because HPV vaccines were approved for worldwide vaccination programmes starting in 2006, we may only now begin to see the long-term effect of the vaccination programmes on the incidence of cancer. Furthermore, given the known lag between HPV infection and the development of cervical lesions or cancer, longitudinal studies with long follow-up time are required to examine the impact of HPV vaccination on cancer risk.

Some observational studies have examined the impact of HPV vaccination programmes in Europe (12-15). One study in England examined the impact of the bivalent HPV vaccine in reducing incidence of HPV infection, showing substantial declines in HPV strains 16, 18, and cross-protection of strains 31, 33, and 45, 8 years following the start of the vaccination programme (14). One study in Scotland demonstrated an 89% reduction in prevalence of grade 3 cervical intraepithelial neoplasia (CIN3) or worse in girls vaccinated with the bivalent vaccine (Cervarix) compared to unvaccinated girls, and that the most protection was provided when girls were vaccinated at age 12-13 years compared to those aged 17 years (15). A meta-analysis conducted in 2021 compiling results from 65 articles across 14 countries, including both bivalent and quadrivalent vaccines (Gardasil/Silgard), demonstrated that between 5-8 years after the implementation of vaccination programmes the prevalence of HPV strains 16 and 18 were reduced by 83% in girls aged 13-19 years and by 66% in women aged 20-24 years. Between 5-9 years after vaccination, the occurrences of grade 2 cervical intraepithelial neoplasia (CIN2) or worse decreased by 51% in those screened at aged 15-19 years and by 31% in women screened at age 20-24 years (16). The first study to investigate the impact of the bivalent vaccine on incidence of cervical cancer and CIN3 used National Cancer Registry data in England and further investigated the impact of age at vaccination (12). Three cohorts of girls vaccinated with the bivalent vaccine in different calendar years were compared with unvaccinated cohorts from years prior to the vaccination programme roll-out. Girls vaccinated at age 12-13 years exhibited 87% reduction in cervical cancer rates; those vaccinated at age 14-16 years 62% reduction, and those vaccinated at 16-18 years 34% reduction (12) (note that age was classified by school year, with some overlapping ages). Rate reductions of CIN3 were even greater (97%, 75%, and 39% for those vaccinated at ages 12-13 years, 14-16 years, and 16-18 years, respectively).

In a Swedish cohort of adolescent girls, the incidence rate of cervical cancer in girls receiving at least one dose of quadrivalent vaccine was compared to unvaccinated girls. Vaccination was associated with a substantially reduced incidence of cervical cancer, particularly after adjusting for confounders including age at follow-up, calendar year, county of residence, parental education, household income, mother's country of birth, and maternal disease history (13). Similarly, the quadrivalent vaccine has been demonstrated to provide protection against the development of cancers of the anus (17); and a meta-analysis of both the bivalent and 9-valent HPV vaccines showed that vaccinated individuals were 80% less likely to develop HPV-16, which is a particular risk for oropharyngeal cancer (18).

HPV vaccination has been shown to be cost-effective globally (19), though there have been suggestions that one-dose may confer comparable protection to two- and three- dose schedules, which could make vaccination programmes more cost-effective both financially and logistically. There is evidence from observational, prospective cohort studies, and a few retrospective studies pointing to the effectiveness of a single HPV vaccine dose in providing strong protection against persistent HPV infections (20-23). For example, Sankaranarayanan and colleagues have illustrated that the immediate protection offered by one quadrivalent HPV vaccine dose is comparable to that of two or three doses (22). This level of protection is similar to what is achieved with a full three-dose regimen. Similar findings have been reported for the bivalent vaccine (21). Additionally, some studies have modelled the clinical and economic impact of one-dose vaccine schedules compared to no- or 2-dose schedules in reducing HPV infection and cervical cancer outcomes in numerous countries worldwide (24, 25). Yet, only a few observational studies have investigated the real-world impact of a single dose schedule on incidence of high-grade cervical lesions (CIN2, CIN3). A study of cancer registry and screening data in Australia has shown that one dose of the quadrivalent vaccine provides comparable effectiveness versus 2 or 3 doses in preventing CIN2 or CIN3 (26). A study in the US also demonstrated equivalent effectiveness of one, two, and three doses of the quadrivalent vaccine in reducing incidence of high-grade cervical lesions (27). However, there is a dearth of research investigating these trends in Europe and none, to our knowledge, which examine all vaccines approved in this region, underscoring the need for further investigation into the dosing schedule. Reducing the dosage can lead to cost savings, streamlined vaccine distribution, and enhanced vaccine accessibility, all while preserving effectiveness in preventing severe illness (25). Recently, the UK Joint Committee on Vaccination and Immunisation (JCVI) have recommended the use of one-dose vaccination nationally [link],(28) illustrating the relevance of and the need for research on this topic.

Based on all the above, the aim of the present study is to generate real world evidence on the long-term effectiveness of HPV vaccination to prevent cervical cancer, including the analysis of the different licensed HPV vaccines and observed dosing regimens in Europe, as feasible with the available data.

8. RESEARCH QUESTION AND OBJECTIVES

To generate evidence from real-world data on the effectiveness of HPV vaccination in preventing severe disease outcomes, i.e. invasive cervical cancer and CIN2+, for the different licensed HPV vaccines in Europe (UK, Spain, and Norway).

More specifically the main study objectives are:

1. To assess the effectiveness of HPV vaccination in prevention of **invasive cervical cancer,** stratified by licenced vaccine brand.
2. To assess the effectiveness of HPV vaccination in prevention of **CIN2+**, stratified by licenced vaccine brand.
3. To assess the effectiveness of HPV vaccination in prevention of **conisation**, stratified by licenced vaccine brand.

Secondary objectives:

- To assess the effectiveness of HPV vaccination for each outcome separately (invasive cervical cancer, CIN2+, and conisation) regardless of brand.
- To assess the effectiveness of HPV vaccination in prevention of invasive cervical cancer, CIN2+, and conisation in subgroups defined by number of doses, within each brand.

9. RESEARCH METHODS

9.1 Study type and study design

Table 9.1. Description of potential study types and related study designs.

Study type	Study design	Study classification
Comparative effectiveness study	New user cohorts	Complex

9.2 Study setting and data sources

This study was conducted using routinely collected data from 3 databases in 3 European countries. All databases had been previously mapped to the OMOP CDM: Clinical Practice Research Datalink (CPRD-GOLD), United Kingdom; Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) linked to hospital inpatient records (CMBD-AH for the acronym in Catalan language), Spain; and Norwegian Linked Health Registry data (NLHR), Norway. A priori, these 3 databases had a good capture of both the exposure and the outcome, as they are both linked to vaccination registries and to the data from screening services. However, limitations were found in CPRD-GOLD with the completeness of vaccination data, and the short follow up period on NLHR.

Clinical Practice Research Datalink GOLD [CPRD-GOLD], United Kingdom (University of Oxford)

The Clinical Practice Research Datalink (CPRD-GOLD) is a governmental, not-for-profit research service, jointly funded by the National Institute for Health and Care Research and the Medicines and Healthcare products Regulatory Agency, a part of the Department of Health, United Kingdom (UK) (https://cprd.com). CPRD-GOLD (29) comprises computerised records of all clinical and referral events in primary care in addition to comprehensive demographic information and medication prescription data in a sample of UK patients, with the most recent data being predominantly from Scotland (52% of practices) and Wales (28% of practices). The prescription records include information on the type of product, date of prescription, strength, dosage, quantity, and route of administration. Data from contributing practices are collected and processed into research databases. Quality checks on patient and practice level are applied during the initial processing. Data are available for 20M patients, including 3.2M currently registered patients. Approval for this study was granted via the Research Data Governance Process.

Information System for Research in Primary Care [SIDIAP], Spain (IDIAP Jordi Gol)

SIDIAP data is collected from EHR records of patients receiving primary care delivered through Primary Care Teams (PCT), consisting of GPs, nurses and non-clinical staff (30). The Catalan Health Institute manages 286 out of 370 such PCT with a coverage of 5.8M patients, out of 7.8M people in the Catalan population (74%). The database started to collect data in 2006. The mean follow-up is 10 years. The observation period for a patient can be the start of the database (2006), or when a person is assigned to a Catalan Health Institute primary care centre. Date of exit can be when a person is transferred out to a primary care centre that does not pertain to the Catalan Health Institute, or date of death, or date of end of follow-up in the database. Drug information is available from prescriptions and from dispensing records in pharmacies. Drugs not prescribed in the GP setting might be underreported; and disease diagnoses made at specialist care settings are not included. Approval for this study was granted by both SIDIAP's Scientific and Ethics Committee.

Norwegian Linked Health Registry data [NLHR], Norway (University of Oslo)

Norway has a universal public health care system consisting of primary and specialist health care services covering a population of approximately 5.4M inhabitants. Many population-based health registries were established in the 1960s with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness. We harmonised data from the following registries: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR). Linkage between the registries was facilitated using project-specific person ID generated from unique personal identification assigned at birth or immigration for all legal residents in Norway. In brief: MBRN stores information about the pregnancy, the mother, father, and child. NPR records diagnosis in secondary care (e.g., hospital). KUHR contains information about diagnosis and contact in primary care (e.g., GPs and outpatient specialists). NorPD recorded all medications dispensed outside of hospitals. MSIS collects test results of communicable diseases (e.g., Sars-Cov-2) and SYSVAK recorded vaccinations. The current data cut only has data from patients that were present in the database from 2015 to 2018.

9.3 Study period

Study period started on 1st of January 2008 as this date corresponds to the start date of the earliest roll-out of the HPV vaccination programme in UK, Spain, and Norway. NLHR only had data from 2018, so the start of the study period was 1st of January 2018. The end of the study period was the last available date of data collection for each contributing dataset: 15th Dec 2023 for CPRD-GOLD, 30th June 2023 for SIDIAP, and 31st Dec 2023 for NLHR.

9.4 Follow-up

For all analyses, follow-up time started from the index date. For vaccinated participants, the index date was defined as the date they received the first HPV vaccine dose before the age of 15. For unvaccinated, the index date was imputed as the midpoint in their matched vaccinated counterparts (that is, the mean date their matched vaccinated pairs have received their first vaccine dose).

End of follow-up was the end of a person’s observation time (i.e. date of data extraction, death), or the date of outcome event, whichever comes first. Additionally, if any individual in the unvaccinated cohort received their first dose after the age of 15, the entire matched group was censored. A sensitivity analysis was conducted with neither censoring the unvaccinated group upon receiving their first dose nor the matched vaccinated group if their corresponding unvaccinated match was censored.

For the secondary analyses (comparison between the number of doses), the matched groups were censored if any of the participants had received a further dose after the age of 15, and an additional analysis without this censoring was also performed.

9.5 Study population with inclusion and exclusion criteria

The study population comprised females born in or after 1993 (aged 15 years old or less in 2008, which is the earliest launch of HPV vaccination in all countries). This population was further restricted to those in observation in the database when they turned 15, and in observation in the database when they turned 9.

Further restrictions were applied on a year-by-year basis for the entire study period, applied on the 1st of January of each year. For each year, participants needed to be in observation on the 1st of January of that year, have 365 days of prior observation available, and be aged between 9 and 15 years old.

Target (vaccinated for the main analysis and a specific dosing regimen for the dose analyses) and comparator (unvaccinated for the main analysis and the rest of the dosing regimens for the dose analyses) cohort participants were matched up to 5:1 based on exact matching by year of birth and geographic region or GP practice (when available); and on PS (nearest neighbour within a 0.2 calliper width).

9.6 Variables

Concept lists used for the identification of exposures and outcomes are included as Supplementary Documents in **Appendix I**. These were produced following the DARWIN EU® Phenotyping standard processes, which involve the review of code lists by clinical experts and the review of phenotype diagnostics after their execution in the participating databases, to ensure completeness and quality of the definitions in all the participant data partners.

9.6.1 Exposure /s

HPV Vaccination

HPV vaccination status was defined by the number of doses received (0, 1, 2, or 3 or more) before 15 years old and by vaccine brand (Bivalent: Cervarix; Quadrivalent: Gardasil/Silgard; and 9-valent: Gardasil-9). For vaccine brand, we followed different strategies to identify it, as recording varied by database. In SIDIAP, recording of vaccine composition was complete so we used it (bi, quadri, 9-valent). In CPRD-GOLD, the brand or valency of the vaccine was not specified, so we used the date of administration as a proxy for vaccine brand, as only one vaccine brand was given at a certain point in time, as per Public Health England (PHE) recommendations: Cervarix from 2008 to 2011, Gardasil from 2012 to 2021, and Gardasil-9 from 2022 to 2023. In Norway, we identified the brand by valency, and for those participants with missing information, we used Cervarix, as it is the brand offered in the Norwegian public health system throughout the time period. Codes used to identify the vaccines in the databases are shown in **Appendix I**.

9.6.2 Outcome/s

The primary outcomes of interest were: (1) Invasive Cervical Cancer, defined as any occurrence of a clinical diagnosis code of invasive cervical cancer; (2) CIN2+, defined as any occurrence of one of the codes diagnosing CIN2+; (3) and conisation, defined as having a procedure coded as conisation of the cervix or cold knife cone (CKC) or loop diathermy. Codes used to identify the outcomes are specified in **Appendix I**. Different variants of phenotype (a broad one, more sensitive, and a narrow, more specific) were used for each outcome, with only using the broad variant for conisation, and the narrow variant for invasive cancer.

Additional outcome variables were identified to investigate intermediary procedures that may increase the likelihood of occurrence of the outcomes (i.e., whether a patient underwent a cervical cancer screening).

9.6.3 Other covariates, including confounders, effect modifiers, and other variables

These variables were used for the characterisation of study cohorts, matching (e.g. geography), stratification (e.g. by age), and to minimise confounding through their inclusion as potential covariates in large-scale propensity scores.

Demographics

We calculated the age at the index date.

GP practice / geographic region

For matching, exact matching was done by GP practice when this variable was available (CPRD-GOLD) and by geographic region when not (SIDIAP, NLHR).

Cytology results

Cytology results indicating HPV status from Pap smear test prior to index date was accounted for as potential covariate in the large-scale propensity score.

Healthcare resource use

Prior number of visits to GPs or any other specialists as recorded in the year before the matching year was used as a proxy for healthcare resource use. This resource was accounted for as a potential covariate in the large-scale propensity score.

Health conditions pre-index date

Individuals´ history of comorbidities was identified over three time periods prior to the start of the matching year, and was used for summary characterisation and calculation of large-scale propensity scores:

1) 30 days prior to one day prior index date,
2) 365 days prior to one day prior index date,
3) all available days observed up to one day prior to index date.

A range of health conditions were assessed using the time windows above, as depicted in **Figure A**.

Medications pre-index date

Pre-existing medication use was identified using 2-time windows defined as 365 days to one day prior to the start of the matching year, and 30 days to 1 day prior to the start of the matching year, and they were used to provide summary characterisation for patients and calculation of large-scale propensity scores.

HIV status pre-index date

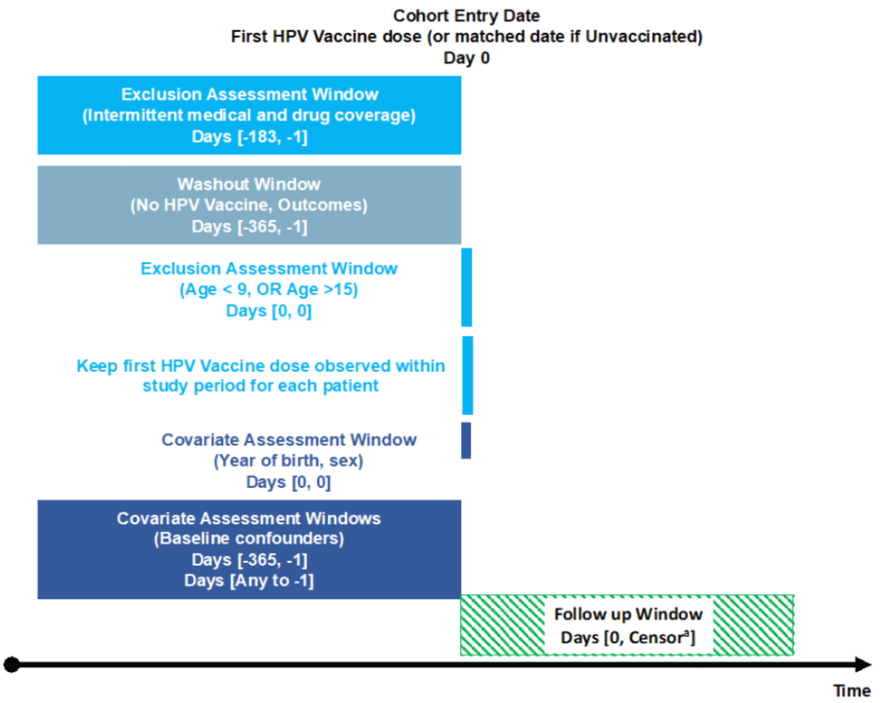
Presence/absence of HIV/AIDS any time in history prior to the start of the matching year was used as a potential covariate in the large-scale propensity score.

Previous Papanicolaou smear Testing

Number of Papanicolaou smear tests (cytology tests) prior to the start of the matching year was used as a potential covariate in the large-scale propensity score.

Previous Vaccinations

Number of vaccinations (any vaccine) prior to the start of the matching year was used as potential covariate in the large-scale propensity score.



a. Censoring at date when vaccine exposure status changes, from unvaccinated to vaccinated

HPV = Human Papillomavirus Vaccine

Figure A. HARPER diagram of study design and covariate assessment.

9.7 Study size

For each database, all individuals that satisfy the eligibility criteria for a study cohort were included. Assuming a vaccine effectiveness against cervical cancer of 88%, with 60% vaccination coverage (a mean ratio unvaccinated to vaccinated of 0.67), a 10-year cumulative incidence of 94/100,000 based on a previous study (13), and for a 95% CI we calculated sample size needed for different precision values:

IRR	Lower limit of 95%CI	Upper limit of 95%CI	Relative precision (%)	Sample size total
0.12	0.11	0.13	9	884,672
0.12	0.10	0.14	20	201,492
0.12	0.09	0.16	33	80,930
0.12	0.08	0.18	50	40,740
0.12	0.07	0.21	71	23,055
0.12	0.06	0.24	100	13,940
0.12	0.05	0.29	140	8,738
0.12	0.04	0.36	200	5,550

Contributing data sources range from 20,000 to 80,000 people vaccinated against HPV, so we expected a relative precision of 33-71.

9.8 Data transformation

Analyses were conducted separately for each database. Before study initiation, test runs of the analytics were performed on a subset of the data sources or on a simulated set of patients and quality control checks were performed. After all the tests were passed, the final package was released in the version-controlled Study Repository for execution against all the participating data sources. The data partners locally executed the analytics against the OMOP CDM in R Studio and reviewed and approved the results. The study results of all data sources were checked after they were made available to the DARWIN EU® Coordination Centre. All results were locked and timestamped for reproducibility and transparency

9.9 Statistical methods

9.9.1 Exposure and outcome diagnostics

We ran diagnostics in all the involved data partners for both the exposure and outcome cohorts to validate the definitions. These included occurrence of code counts with correspondence to source codes, cohort counts, overlap between cohorts, age and sex distribution, distribution measures of time in the database before and after the index date. It also included prevalence and incidence of the cohorts in a sample of the database, and a large-scale characterisation in a sample of the patients and a matched sample of same age and sex patients. For the exposure, we also measured the coverage of the HPV vaccine, the number of doses, and the timing between doses for each of the databases.

9.9.2 Main statistical methods

The analyses in this study are shown in **Table 9.2**.

Table 9.2. Description of study types and types of analysis.

Study type	Study classification	Type of analysis
Comparative effectiveness study	Complex	New user cohort design: Large-scale characterisation of participants in the target and comparator cohorts. Large-scale propensity scores (LSPS) were estimated. Incidence rate/s of each of the outcomes of interest in the target and comparator cohorts. Diagnostic/s: Covariate balance, Equipoise, Power, residual confounding/systematic error (optional). Rate Ratios or Hazard Ratio/s and 95% confidence intervals using Poisson or Cox models, respectively.

We used a PS-matched cohort design. The matching process was conducted on a year-by-year basis, starting by matching individuals vaccinated in 2008 (the beginning of the study period), followed by those vaccinated in 2009, and continued until the end of the observation period in the databases. The process for matching within each year was as follows:

Step 1: All subjects in the vaccinated (target) cohort were exact-matched by year of birth and geographic region (or GP identifier in CPRD-GOLD) to all potential unvaccinated (comparator) matches not belonging in the target cohort. Both the target and comparator cohort needed to meet specific criteria to be included: they had to be in observation, have at least 365 days of prior history, and be aged between 9 to 15 years at the start of the matching year.

Step 2: Large-scale propensity scores were estimated using LASSO regression to estimate the probability of being in the target cohorts at the beginning of the matching year. The resulting equations were manually inspected by two clinical epidemiologists to identify any strong instrumental variables. Up to 5 matches were found in the target cohort for each participant in the comparator cohort using PS matching with nearest neighbour matching with a calliper width of 0.2. Matches were sampled from the pool of target cohort participants identified as potential matches in the first step. Then, the index date of the target cohort participant (or the average time point, if more than one) was applied to all the identified comparator cohort matches. Participants from the comparator cohort that had been matched were removed from the pool of future potential matches.

For the secondary objectives involving dose schedules, matching was also done following a yearly basis. That means we first matched individuals vaccinated at the start of the study period and then continued with the following years. As both the comparator and target cohorts had already an index date assigned (date of first dose vaccine), the specific criteria from step 1, and propensity scores in step 2 were applied and calculated at the index date for target and comparator cohorts.

The following matched cohorts were compared:

Main comparisons (Primary objectives):

Vaccinated vs unvaccinated per brand:

- Vaccinated with Gardasil/Silgard (target) (1 or more dose) vs unvaccinated (comparator)
- Vaccinated with Cervarix (target) (1 or more dose) vs unvaccinated (comparator)
- Vaccinated with Gardasil-9 (target) (1 or more dose) vs unvaccinated (comparator)

Secondary comparisons (Secondary objectives):

Vaccinated (target) (1 or more dose) (any brand) vs unvaccinated (comparator) overall.

Dose comparisons:

- Vaccinated with 2 or more doses (target) vs 1 dose (comparator) of the same brand.
- Vaccinated with 3 or more doses (target) vs 2 doses (comparator) of the same brand.

In all the matched cohorts, people were followed up from their index date until the earliest of end of their observation (i.e. date of data extraction, death) or the date of outcome event, whichever comes first. We also censored unvaccinated individuals that received a vaccine dose after the age of 15. For secondary analyses regarding dose number comparisons, we censored all participants if they received a further dose of the vaccine after the age of 15. In both analyses, if any individual in the comparator cohort received an extra dose after the age of 15, the entire matched group was censored.

We reported summary descriptive analyses including age, sex, and key variables for matching and conditions and medication pre-index date for characterisation.

We calculated Incidence rates and incidence rate ratios (IRR), for the unmatched and matched cohorts, of outcomes at 5, 10, and 15 years after vaccination using Poisson regression. We used Cox proportional hazard models to calculate hazard ratios (HR) for the outcomes in both unmatched and matched cohorts.

Two study diagnostics were used to minimise the risk of reporting biased results. First, any analyses with evidence of residual observed confounding after matching, as defined by Absolute Standard Mean Difference (ASMD) >0.1 for any covariate, was inspected manually by two clinical epidemiologists (31). If any of these variables were deemed as a confounder, all subsequent analyses stopped. Additionally, negative control outcomes were be used to assess residual unobserved confounding. A previously validated list of negative control outcomes was utilised and refined to identify potential outcomes not associated with outcome risk but sharing similar confounders as the association between HPV vaccination and outcomes. A list of the outcomes used as negative controls can be found in **Appendix I Table 3**.

Kaplan-Meier plots were used to illustrate time-to-event analyses. Log-log plots were visually inspected to identify scenarios with a violation of the proportional hazards assumption. If these plots showed evidence of violation, we didn't report the results from the Cox regression and only reported incidence rates and incidence rate ratios.

Cell counts <5 were suppressed to comply with the database's privacy protection regulations.

Table 9.3. Primary, secondary, and subgroup analysis specification.

A. Primary analysis

Hypothesis:	HPV vaccine decreases the risk of CIN2+, conisation, and invasive cervical cancer
Exposure contrast:	HPV Vaccine (each brand) vs unvaccinated
Outcome:	CIN2+, conisation, and invasive cervical cancer
Analytic software:	R
Model(s):	Incidence rates, incidence rate ratios, Cox proportional Hazards models, Kaplan-Meier Time-to-event.
Confounding adjustment method:	Among those participants in the target and comparator cohorts who met the inclusion criteria, target participants were matched 5:1 to a comparator participant, based on year of birth, calendar year of vaccination, geographic region, and large-scale propensity scores using the nearest neighbour matching, with calliper width 0.2 standard deviations. Large-scale propensity scores were estimated using LASSO regression to estimate the probability of being in the target cohorts. Covariates were included as all recorded features in the database, including socio-demographics, geographic location, healthcare resource use, comorbidity, medicine/s use, previous Papanicolaou testing, and previous vaccination/s. Among those, covariates with a prevalence below 0.5% in the study population were omitted. Logistic regression with LASSO regularisation was then be used for variable selection. The list of selected covariates was manually screened by 2 epidemiologists/clinical domain experts to exclude potential instrumental variables.
Missing data methods:	None

A. Secondary analysis 1

Hypothesis:	HPV vaccine decreases the risk of CIN2+, conisation, and invasive cervical cancer
Exposure contrast:	HPV Vaccine(overall) vs unvaccinated
Outcome:	Same as Primary Analysis
Analytic software:	Same as Primary Analysis
Model(s):	Same as Primary Analysis
Confounding adjustment method:	Same as Primary Analysis
Missing data methods:	Same as Primary Analysis

A. Secondary analysis 2

Hypothesis:	Higher number of doses decreases more the risk of CIN2+, conisation, and invasive cervical cancer
Exposure contrast:	HPV Vaccine with 1 vs 2 or more doses (secondary objective 1); HPV Vaccine with 2 vs 3 or more doses (secondary objective 2)
Outcome:	Same as Primary Analysis
Analytic software:	Same as Primary Analysis
Model(s):	Same as Primary Analysis
Confounding adjustment method:	Same as Primary Analysis
Missing data methods:	Same as Primary Analysis

9.9.3 Missing values

Missingness in exposure was evaluated by comparing coverage of HPV vaccination as observed in the data with the coverage reported by national public health agencies. We included a sensitivity analysis to deal with situations where we expected missingness in the exposure. For the outcomes, follow up was censored at the moment they stopped being observed in the database, thus reducing the possibility of missed outcomes.

9.9.4 Sensitivity analysis

Sensitivity analyses are summarised in **Table 9.4.**

Table 9.4. Sensitivity analyses – rationale, strengths, and limitations.

What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary analysis	Limitations of the sensitivity analysis compared to the primary analysis
Study population and follow-up, not censoring unvaccinated subjects who are vaccinated after the age of 15	To assess the potential impact of selection bias related to the censoring of subjects vaccinated after age 15	Does not exclude potentially higher risk subjects vaccinated in later life	Misclassification of exposure (vaccination status)
In databases with incomplete vaccination data, we restricted the population to women in practices/region and birth cohort with more than 60% coverage.	Restricting the analyses where we believe vaccination data is properly registered, we reduced the risk of exposure misclassification	Minimises the number of subjects without information on vaccination.	Overestimation of exposure prevalence, selection bias.

We performed two sensitivity analyses:

In the first sensitivity analysis, we did not censor individuals from the comparator group (either for the main and secondary analysis) if they have had a dose of the vaccine after 15 years old, to mimic an intention to treat analysis.

A second sensitivity analysis was performed in CPRD-GOLD, where we had incomplete information on vaccination for some GP practices for some birth cohorts. For this analysis, we restricted the analysis to women vaccinated in GP practices that have more than a 60% coverage of HPV vaccination for their year of birth cohort. We decided to use this threshold by establishing a minimal coverage reported by the local public health agency (Public Health England/UKHSA/OHID) on their Fingertips database [link].

This number was obtained by getting the minimum coverage achieved by area, using the smallest area data available (Upper tier local authorities), before the COVID pandemic. This coverage was for Kensington and Chelsea in 2014-15, of 67.6% 95%CI (63.6% - 71.3%). We then decided to truncate the figure to 60%, to account for the possible variability introduced by us having smaller areas.

10. DATA MANAGEMENT

All databases have previously mapped their data to the OMOP common data model. This enables the use of standardised analytics and using DARWIN EU tools across the network since the structure of the data and the terminology system is harmonised. The OMOP CDM was developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org.

The analytic code for this study was written in R, and we used standardised analytics wherever possible. Each data partner executed the study code against their database containing patient-level data and then returned the results (csv files) which only contained aggregated data. The results from each of the contributing data sites was then be combined in tables and figures for the study report.

Packages used for this study included:

In addition to all the packages developed as part of the DARWIN EU project, this study used the MatchIt package for matching, the EpiR package for calculating incidence rates, the survival package for Kaplan-Meier analyses, log-log plots, and calculations of hazard ratios, and the glmnet package for estimating incidence rate ratios for the main analysis. All the dependencies and the versions of the packages used can be found in the lock file from the GitHub repository of this study [link].

10.1 Data storage and protection

For this study, participants from various EU member states processed personal data from individuals which was collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses were run, which generate non-identifiable aggregate summary results.

The output files are stored in the DARWIN Remote Research Environment. These output files do not contain any data that allow identification of subjects included in the study. The RRE implements further security measures in order to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

11. QUALITY CONTROL

General database quality control

A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, data partners ran the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness, and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

Vaccine exposure status, clinical diagnoses, and conisation procedures were identified from the data using code-lists reviewed by clinicians. When defining conditions for outcomes of interest, i.e. CIN2+, cervical cancer, a systematic search of possible codes for inclusion was conducted using CodelistGenerator R package (32). Clinicians reviewed the resulting code lists to exclude irrelevant codes, such as for persisting disease or complications. In addition, vaccine coverage and cohort diagnostics were run to assess the use of different codes across the databases contributing to the study and identify any codes potentially omitted in error.

12. RESULTS

12.1 Participants

Figure 12.1 describes the number of study subjects who entered the study and those who were excluded. The number of females born on or after 1993 were 2,013,936 in CPRD-GOLD, 1,070,348 in SIDIAP, and 1,028,584 in NLHR. After limiting to those in observation at least from 9 years old to 15 years old the population size was restricted to 333,983 in CPRD-GOLD, 302,559 in SIDIAP, and 292,404 in NLHR. This resulted in 191,376 vaccinated and 142,607 unvaccinated at 15 years old in CPRD-GOLD, 262,364 vaccinated and 40,195 unvaccinated in SIDIAP, and 116,271 vaccinated and 176,133 unvaccinated in NLHR.

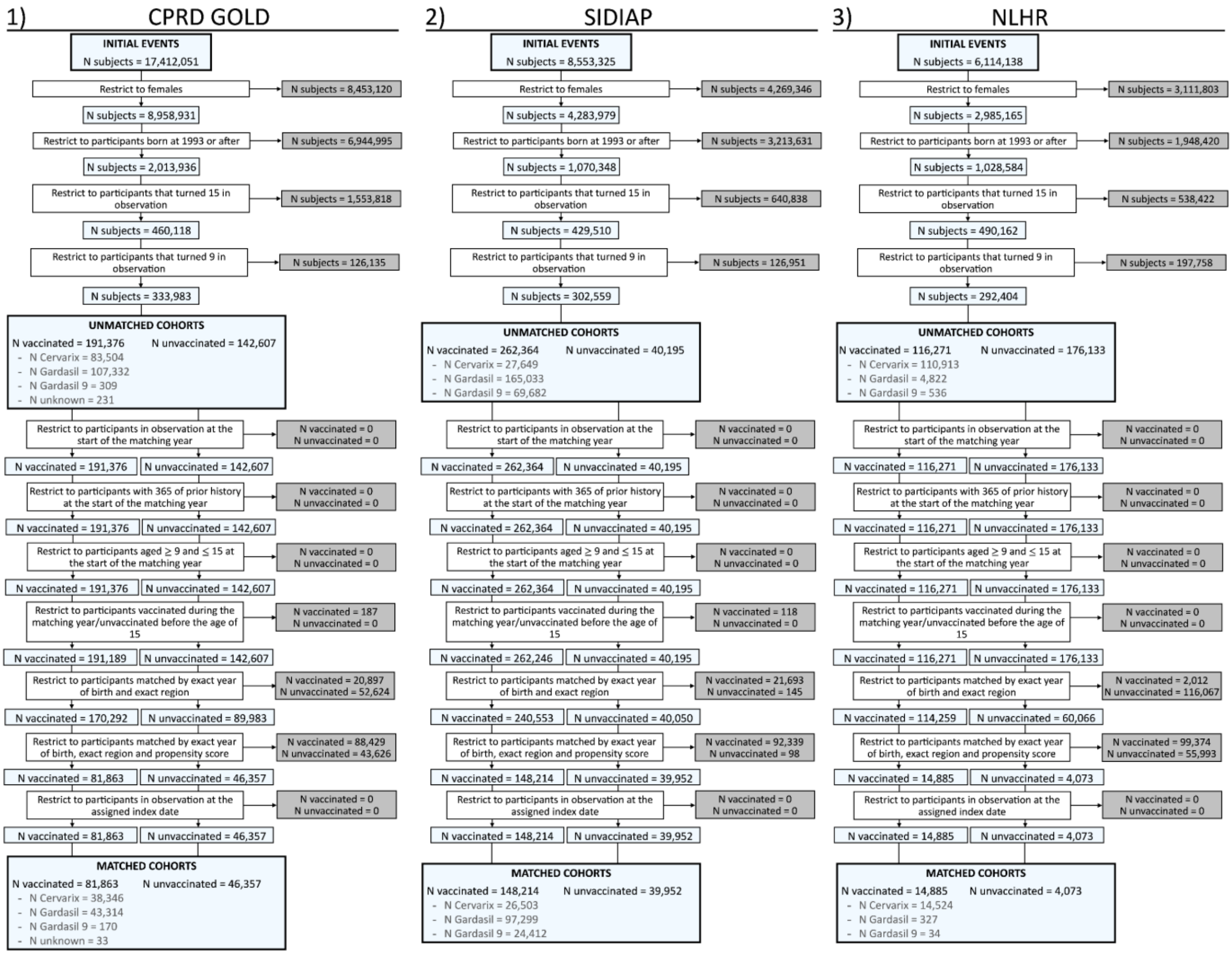


Figure 12.1. Flowchart of inclusion criteria for matched and unmatched cohorts, by data partner.

12.1.1 Exact matching and PS matching

Figure 12.1 also describes the attrition due to the matching process. Further restrictions were done for the matching on a year-by-year basis, applied on the 1st of January of each year. For each year, participants needed to be in observation on the 1st of January of that year, to have 365 days of prior observation available, and to be aged between 9 and 15 years old to be considered eligible for inclusion on a certain year. These restrictions amounted to excluding 187 women in CPRD-GOLD, 118 in SIDIAP, and none in NLHR. The remaining population was further matched (vaccinated to unvaccinated) using exact matching by calendar year (on the 1st of January of each year) and geographic region (SIDIAP, NLHR) or GP practice (CPRD-GOLD).This resulted in a further exclusion of 73,521 participants in CPRD-GOLD, 21,838 in SIDIAP, and 118,079 in NLHR, for whom we were unable to find a match.

In this age-region-year exact-matched cohort, a PS of being vaccinated on a certain calendar year was calculated. The shiny app (https://data-dev.darwin-eu.org/P2-C3-004_Update/) shows the coefficients for this model, for each vaccination year and data partner. After review, we did not identify any suspected instrumental variables, and the top contributing covariates were mostly confounders (e.g. exposure to other vaccines) or risk factors related to healthcare use (e.g. prescription of medicines like ibuprofen or coding of acute non-serious infections).

We matched for the resulting large-scale propensity score 1 unvaccinated to up to 5 vaccinated participants every year, to the nearest neighbour within a 0.2 calliper width. We were not able to match and therefore excluded an additional 132,055 people from CPRD-GOLD (88,432 vaccinated and 43,626 unvaccinated), 92,437 from SIDIAP (92,339 vaccinated and 58 unvaccinated), and 155,367 from NLHR (99,374 vaccinated and 55,993 unvaccinated). The final “Matched” cohorts included 81,863 vaccinated and 46,357 unvaccinated in CPRD-GOLD; 148,214 vaccinated and 39,952 unvaccinated in SIDIAP; and 14,885 vaccinated and 4,073 unvaccinated in NLHR.

In **Figure 12.2** we show the ASMD for previous condition occurrences and drug prescriptions in the month and year before the vaccination date. This compares the balance between vaccinated and unvaccinated cohorts for all these covariates before (X axis) and after matching (Y axis). **Table 12.1** shows the top 10 imbalanced variables for each data partner before matching. These relate to acute respiratory tract conditions, like tonsilitis or common cold, or related treatments, like ibuprofen, acetaminophen, or amoxicillin. In CPRD-GOLD we also observed imbalances on the previous uptake of the influenza vaccine. Overall, the matching achieves its goal and improves the balance, with the only exposure failing to achieve the goal balance of <0.1 ASMD being the prescription of oral solution of ibuprofen in SIDIAP. We did not consider this to be a relevant confounder, as there is no known effect of ibuprofen on the occurrence of the study outcomes.

The balance on covariate occurrence or prescription not limited to the previous year (any time before the index date) before and after matching is shown in the shiny app (https://data-dev.darwin-eu.org/P2-C3-004_Update/) and shows a few additional imbalanced drugs and conditions, none of them deemed to be substantial confounders.

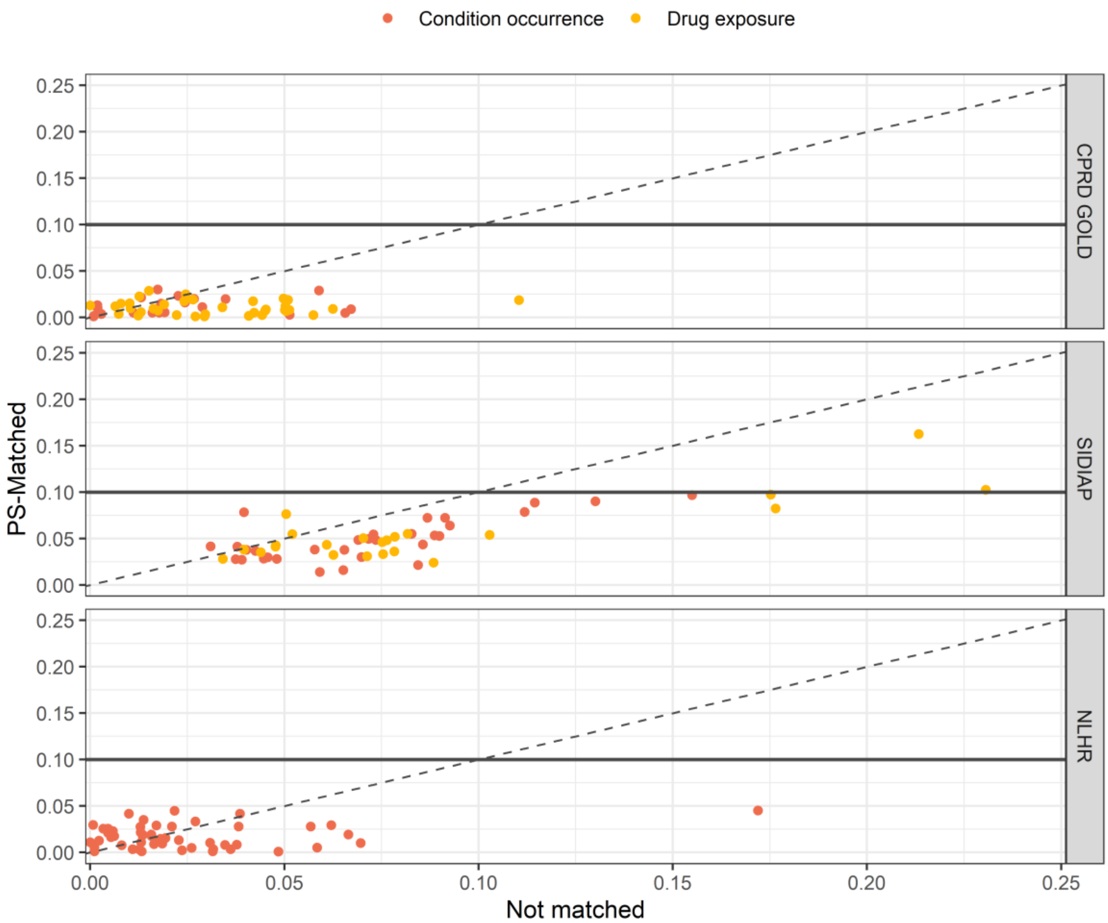


Figure 12.2. Absolute standardised mean difference (ASMD) before and after ps-matching. Dots in red indicate conditions/diagnoses, whereas yellow dots indicate drugs and vaccines, all recorded in the year before index.

Table 12.1. Table showing the top 10 imbalanced variables before matching for the previous year time window and their ASMD before and after matching.

Variable	ASMD	
	Not matched	PS-Matched
CPRD GOLD		
influenza nasal, unspecified formulation	**0.11**	0.01
Asthma not disturbing sleep	0.07	0.01
Asthma not limiting activities	0.07	0.01
influenza virus vaccine, unspecified formulation	0.07	0.01
Sore throat symptom	0.06	0.03
Beclomethasone 0.05 MG/ACTUAT Inhalant Powder [Clenil Modulite] by Chiesi	0.06	0.00
Asthma daytime symptoms	0.06	0.01
penicillin V potassium 50 MG/ML Oral Solution	0.05	0.03
1000 ML bicarbonate ion 0.017 MMOL/ML / POLYETHYLENE GLYCOL 3350 105 MG/ML / Potassium 0.0054 MMOL/ML / Sodium 0.065 MMOL/ML Oral Powder [Movicol] by Norgine	0.05	0.03
cetirizine hydrochloride 1 MG/ML Oral Solution	0.05	0.01
SIDIAP		
ibuprofen 20 MG/ML Oral Suspension	**0.23**	**0.10**
ibuprofen 40 MG/ML Oral Suspension	**0.21**	**0.16**
acetaminophen 100 MG/ML Oral Solution	**0.18**	0.08
amoxicillin 50 MG/ML Oral Suspension	**0.18**	**0.10**
Common cold	**0.15**	**0.10**
Acute tonsillitis	**0.13**	0.09
Acute pharyngitis	**0.11**	0.09
Traumatic or non-traumatic injury	**0.11**	0.08
albuterol	**0.10**	0.05
Gastrointestinal infection	0.09	0.06
NLHR		
Illness	**0.17**	0.05
Upper respiratory tract infection caused by Influenza virus	0.07	0.01
Abdominal pain	0.07	0.02
Joint pain	0.06	0.03
Pain in limb	0.06	0.01
Constipation	0.06	0.03
Hypermetropia	0.05	0.00
Acute atopic conjunctivitis	0.04	0.04
Allergic rhinitis	0.04	0.03
Chronic nasopharyngitis	0.04	0.01

12.2 Descriptive data

12.2.1 Vaccine exposure

Figure 12.3 shows coverage as percentage of women with at least one dose HPV vaccination at 15 years old, by birth cohort and data partner, for those women observed in the database at least since age 9. For CPRD-GOLD, vaccine coverage starts with some very low coverage for those born between 1993-1995 and rises to more than 60% for those born after 1995, the first birth cohort that becomes eligible for systematic vaccination in schools. Coverage in CPRD-GOLD remains stable between 60 and 70% for all birth cohorts until those born in 2007 and 2008, when coverage decreased to 55.6%. In SIDIAP there is a high coverage throughout, the first birth cohort year with a high coverage being 1997, with 83.1% uptake, increasing steadily until those born in 2008, with a 94.3% coverage. As for NLHR, first birth year with coverage data was 2005 as the data made available only started in 2018, where it was 80.1%, and remained high until 2008, with 85.3%.

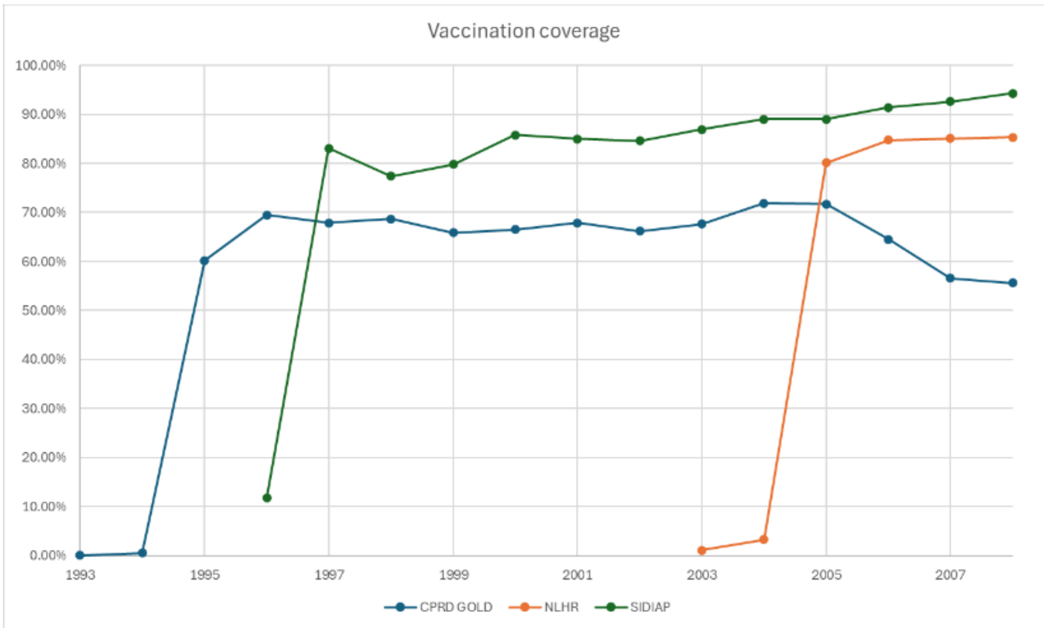


Figure 12.3. Coverage by birth date and data partner at age 15, for women in observation in the data since 9 years old.

The number of women included in the different cohorts: vaccinated, unvaccinated, by brand, and by number of doses is shown in **Table 12.2**. Of those women vaccinated, 44% had Cervarix, 56% Gardasil, and <1% Gardasil-9 in CPRD-GOLD. In SIDIAP, 11% received Cervarix, 63% Gardasil, and 27% Gardasil-9. In NLHR, 95% of women had Cervarix, 4% had Gardasil, and <1% had Gardasil-9.

As for number of doses, most women vaccinated with Cervarix received 3 doses in CPRD-GOLD (74%) and SIDIAP (96%), and 2 doses in NLHR (74%). For Gardasil, the most common schedule was two doses in CPRD-GOLD (46%), three in SIDIAP (54%), and one in NLHR (54%). As for Gardasil 9, one dose was mostly administered in CPRD-GOLD (86%), two in SIDIAP (94%), and NLHR (64%).

Table 12.2. Number of women unvaccinated and vaccinated at age 15 included in the matched and unmatched cohorts, by number of doses and brand.

		Before Matching			After Matching 5:1		
Brand	Number of doses	CPRD-GOLD	SIDIAP	NLHR	CPRD-GOLD	SIDIAP	NLHR
All	Zero	142,607	40,195	176,133	46,357	39,952	4,073
	Any	191,376	262,364	116,271	81,863	148,214	14,885
	One	42,615	7,258	30,390	9,453	5,658	2,014
	Two	63,598	138,665	84,946	8,238	6,122	917
	Three or more	85,163	116,441	935	24,154	20,568	925
Cervarix	Zero	142,607	40,195	176,133	21,578	7,095	3,716
	Any	83,504	27,649	110,913	38,346	26,503	14,524
	One	6,946	244	27,676	3,471	681	1,984
	Two	14,459	905	82,432	5,733	1,377	908
	Three or more	62,099	26,500	805	16,990	6,317	916
Gardasil	Zero	142,607	40,195	176,133	24,607	27,588	323
	Any	107,332	165,033	4,822	43,314	97,299	327
	One	35,354	3,809	2,585	5,979	3,369	27
	Two	49,044	72,273	2,171	2,505	4,032	8
	Three or more	22,934	88,951	66	7,164	13,537	8
Gardasil 9	Zero	142,607	40,195	176,133	140	5,269	34
	Any	308	69,682	536	170	24,412	34
	One	264	3,205	129	3	1,608	3
	Two	44	65,487	343		713	1
	Three or more		990	64		714	1

12.2.2 Description of the participants

Table 12.3 shows the characteristics of the participants included in the main analyses by vaccination status at age 15. The start date corresponds to the vaccination date for those vaccinated and the vaccination date of the matched pair for those unvaccinated. Both matched cohorts start in 2008 for CPRD-GOLD and SIDIAP, and in 2018 for NLHR. Mean age at the time of first vaccination was 13 years old in CPRD-GOLD, 11 in SIDIAP and 12 in NLHR. After the vaccination date, women were followed for a mean of 7 years and a maximum of 16 years in CPRD-GOLD, a mean of 10 years and a maximum of 15 in SIDIAP, and a shorter follow-up mean of 5 years and a maximum of 6 in NLHR.

Table 12.3. Baseline characteristics of the resulting year, age, region, and propensity score matched cohort. Index date in the matched cohort refers to the first vaccination (or assigned first vaccination for the unvaccinated) date. Age, prior observation, and future observation are reported in years.

Variable		Database name: CPRD GOLD – Unvax	CPRD GOLD – Vax	SIDIAP – Unvax	SIDIAP – Vax	NLHR – Unvax	NLHR – Vax
Number records	**N**	46,357	81,863	39,952	148,214	4,073	14,885
Number subjects	**N**	46,357	81,863	39,952	148,214	4,073	14,885
Cohort start date	**Min, max**	2008-01-17 - 2023-08-04	2008-01-01 - 2023-08-04	2008-01-01 - 2019-12-09	2008-01-01 - 2019-12-31	2018-01-02 - 2023-11-20	2018-01-02 - 2023-11-20
Age	**Mean (SD)**	12.68 (0.72)	12.69 (0.68)	10.90 (0.33)	10.90 (0.31)	12.58 (1.03)	12.24 (0.67)
Prior observation	**Mean (SD)**	9.83 (2.77)	9.97 (2.73)	6.85 (2.81)	7.39 (2.82)	10.31 (1.49)	10.35 (1.30)
	Min, max	3 - 15	2 - 15	1 - 13	2 - 14	2 - 16	2 - 15
Future observation	**Mean (SD)**	6.67 (3.44)	6.95 (3.50)	9.49 (3.28)	9.55 (3.15)	5.13 (0.90)	5.25 (0.64)
	Min, max	0 - 16	0 - 16	2 - 15	3 - 15	0 - 6	0 - 6

SD = Standard deviation; Vax = Vaccinated; Unvax = Unvaccinated.

12.2 Outcome data

Table 12.4 shows the incidence rates of the main outcomes in the PS-matched vaccinated and unvaccinated participants for each vaccine brand in the analysis population.

Invasive cervical cancer

There were less than 5 cases of invasive cancer per database after a total follow up of ~700,000 person-years in CPRD, 1.5M in SIDIAP, and 90,000 in NLHR.

CIN2+

In CPRD-GOLD, there were 14 cases of CIN2+ in vaccinated and 14 in unvaccinated participants, amounting to an incidence rate of 3.26 (1.78 to 5.47) per 100,000PYs in vaccinated and 5.57 (3.04 to 9.34) per 100,000PYs in unvaccinated. In SIDIAP, there were 15 cases in unvaccinated and 31 in vaccinated, with an incidence of 2.47 (1.68 to 3.51) per 100,000PYs in vaccinated and 4.24 (2.37 to 6.99) per 100,000PYs in unvaccinated. As for NLHR, there were no cases recorded.

Conisation

Conisation data was only available in CPRD-GOLD and NLHR.

In CPRD-GOLD, we found less than 5 women with conisation in each vaccination group, with incidences of 0.23 (0.01 to 1.3) in vaccinated and 0.40 (0.01 to 2.22) in unvaccinated. In NLHR, 13 vaccinated women and 9 unvaccinated had conisation with incidences of 18.72 (9.97 to 32.01) and 46.87 (21.43 to 88.97) respectively.

Table 12.4. 5, 10, 15y incidence rate of study outcomes in the PS-matched cohorts according to vaccination status and stratified by database.

Year	Variable	CPRD GOLD		SIDIAP		NLHR	
		Unvax	Vax	Unvax	Vax	Unvax	Vax
Cervical cancer							
5 years	**Number events (N)**	0	0	0	0	0	0
	Person Year	186,937	327,343	194,566	719,174	17,949	65,063
	IR (95% CI)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
10 years	**Number events (N)**	0	0	0	0	0	0
	Person Year	241,679	414,560	315,759	1,137,151	19,229	69,504
	IR (95% CI)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
15 years	**Number events (N)**	0	0	0	<5	0	0
	Person Year	251,490	429,213	354,094	-	19,229	69,504
	IR (95% CI)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.08 (0, 0.24)	0 (0, 0)	0 (0, 0)
All	**Number events (N)**	0	0	0	<5	0	0
	Person Year	251,539	429,280	354,098	-	19,229	69,504
	IR (95% CI)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.08 (0, 0.24)	0 (0, 0)	0 (0, 0)
CIN2+							
5 years	**Number events (N)**	<5	0	<5	<5	0	0
	Person Year	-	327,342	-	-	17,949	65,063
	IR (95% CI)	0.53 (0, 1.6)	0 (0, 0)	0.51 (0, 1.54)	0.14 (0, 0.42)	0 (0, 0)	0 (0, 0)
10 years	**Number events (N)**	<5	5	5	12	0	0
	Person Year	-	414,540	315,719	1,136,991	19,229	69,504
	IR (95% CI)	1.66 (0.41, 3.31)	1.21 (0.24, 2.41)	1.58 (0.32, 3.17)	1.06 (0.53, 1.67)	0 (0, 0)	0 (0, 0)
15 years	**Number events (N)**	14	14	15	31	0	0
	Person Year	251,429	429,116	353,990	1,254,182	19,229	69,504
	IR (95% CI)	5.57 (2.78, 8.75)	3.26 (1.63, 5.13)	4.24 (2.26, 6.5)	2.47 (1.67, 3.35)	0 (0, 0)	0 (0, 0)
All	**Number events (N)**	14	14	15	31	0	0
	Person Year	251,478	429,182	353,993	1,254,192	19,229	69,504

Year	Variable	CPRD GOLD		SIDIAP		NLHR	
		Unvax	Vax	Unvax	Vax	Unvax	Vax
	IR (95% CI)	5.57 (2.78, 8.75)	3.26 (1.63, 5.13)	4.24 (2.26, 6.5)	2.47 (1.67, 3.35)	0 (0, 0)	0 (0, 0)
Conisation							
5 years	Number events (N)	0	0	0	0	7	11
	Person Year	186,937	327,343	194,566	719,174	17,932	65,027
	IR (95% CI)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	39.04 (15.7, 80.43)	16.92 (8.44, 30.27)
10 years	Number events (N)	0	<5	0	0	9	13
	Person Year	241,678	-	315,759	1,137,151	19,203	69,444
	IR (95% CI)	0 (0, 0)	0.24 (0, 0.72)	0 (0, 0)	0 (0, 0)	46.87 (21.43, 88.97)	18.72 (9.97, 32.01)
15 years	Number events (N)	<5	<5	0	0	9	13
	Person Year	-	-	354,095	1,254,561	19,203	69,444
	IR (95% CI)	0.4 (0, 1.19)	0.23 (0, 0.7)	0 (0, 0)	0 (0, 0)	46.87 (21.43, 88.97)	18.72 (9.97, 32.01)
All	Number events (N)	<5	<5	0	0	9	13
	Person Year	-	-	354,099	1,254,572	19,203	69,444
	IR (95% CI)	0.4 (0, 1.19)	0.23 (0, 0.7)	0 (0, 0)	0 (0, 0)	46.87 (21.43, 88.97)	18.72 (9.97, 32.01)

12.3 Main results

12.3.1 Vaccinated vs unvaccinated

Figure 12.4 shows the risk ratio of the outcomes in vaccinated vs unvaccinated PS-matched participants at 5, 10, and 15 years after vaccination date and Hazard Ratios for the entire period.

Invasive cervical cancer

Because of the low number of outcomes, invasive cervical cancer (primary outcome) could not be modelled, so we only show results for CIN2+ and conisation.

CIN2+

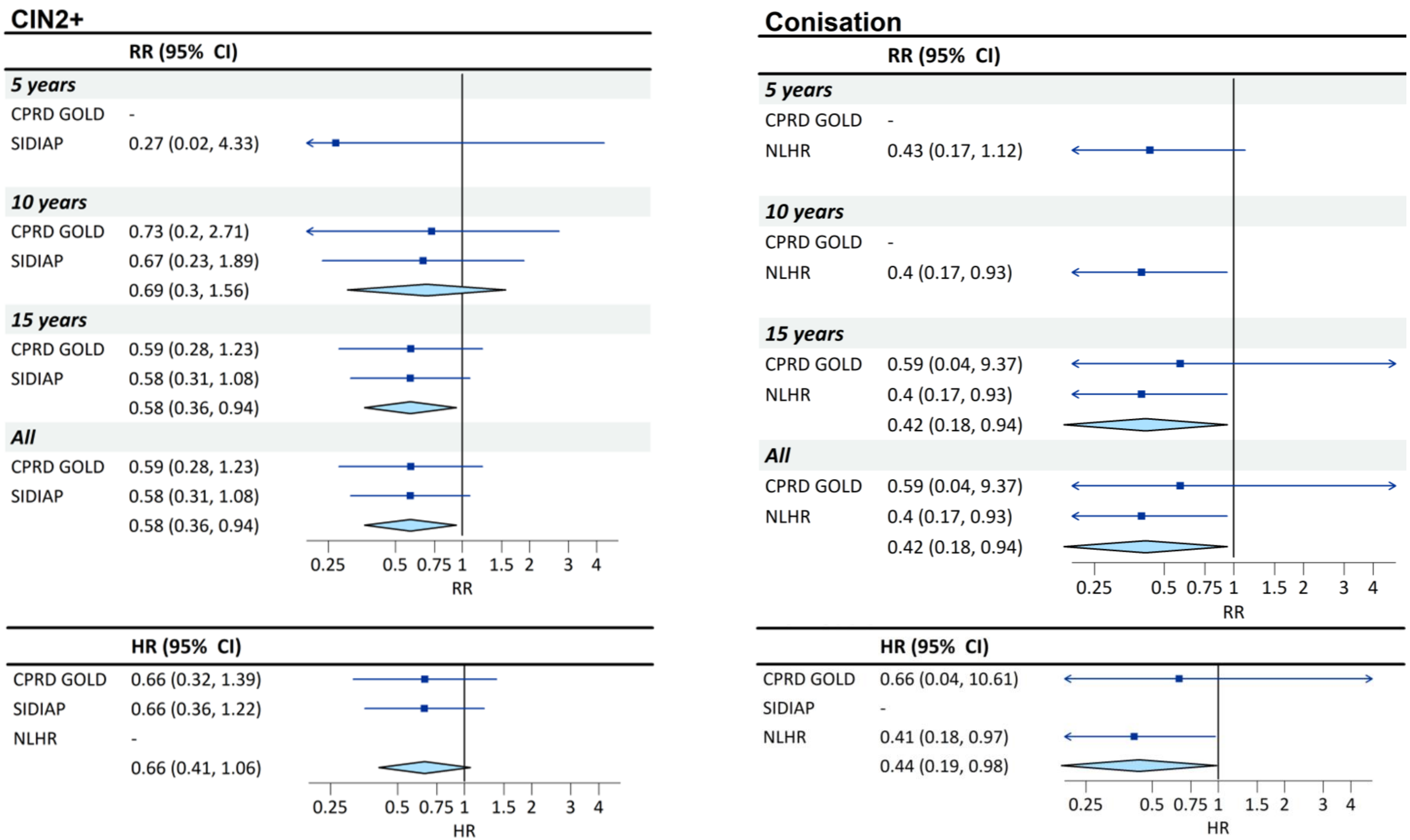
Overall vaccine effectiveness (VE) against CIN2+ in 15 years calculated as 1-RR for all vaccine brands combined was of 41% CI95% (-23% to 72%) in CPRD-GOLD and 42% CI95% (-8% to 69%). We didn't observe enough events to calculate VE in NLHR. The meta-analytic estimate of VE was of 42% CI95% (6% to 64%). The VE estimate using time-to-event analyses (1-HR) was similar, with a pooled meta-analytic estimate of 34% CI95% (-6 to 59%), again based on CPRD-GOLD and SIDIAP (excluding NLHR).

When stratifying by brand, the pooled VE against CIN2+ (calculated as 1-RR in 15 years) for Cervarix was of 38% (-26% to 97%) and of 41% (-11% to 69%) for Gardasil. VE calculated as 1-HR was similar.

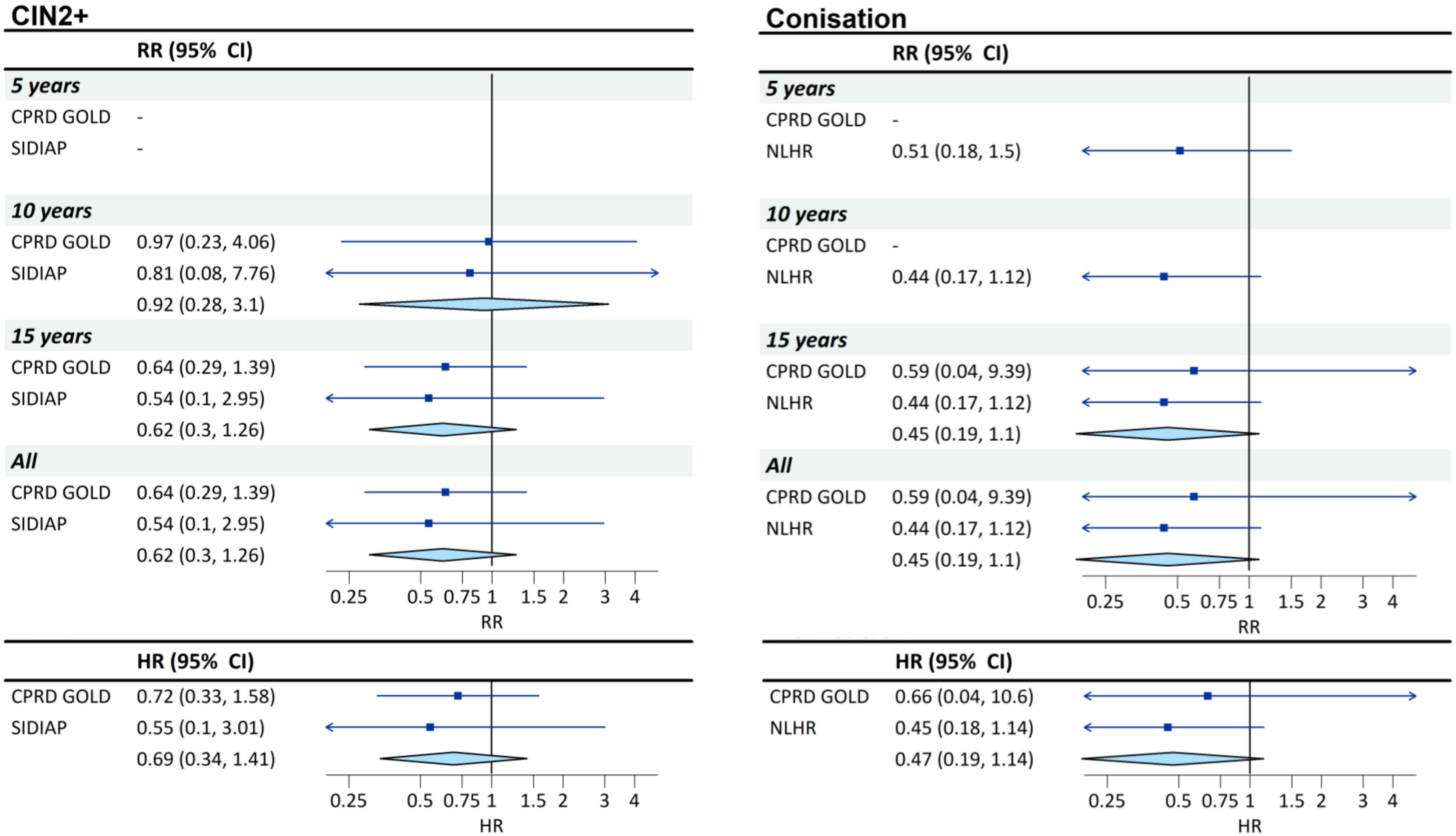
Conisation

As for conisation, in CPRD-GOLD we found a VE of 41% with large confidence intervals (-837% to 96%) and in NLHR the 15-year VE was 60% (7% to 83%). SIDIAP did not contribute data on conisation and the pooled meta-analytic estimate of VE against conisation based on CPRD-GOLD and NLHR was of 58% (6% to 82%). The VE estimate using Cox regression (1-HR) was similar, with a pooled estimate of 56% (2% to 81%).

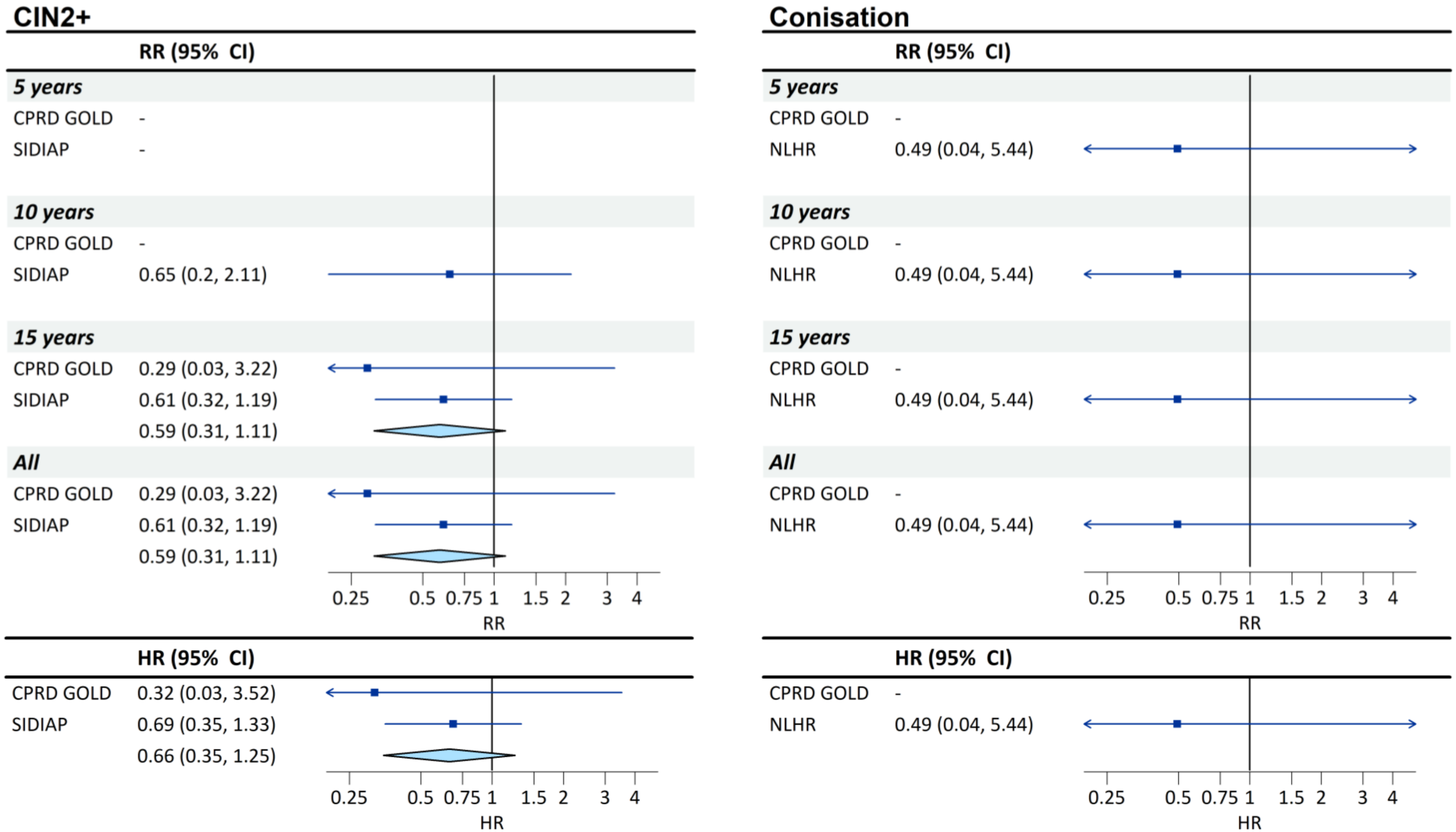
When stratifying by brand, the pooled VE against Conisation (calculated as 1-RR in 15 years) for Cervarix was 55% (-10% to 81%). VE calculated as 1-HR was similar. For Gardasil, VE calculated as 1-RR in 15 years was 51% (-444% to 96%) and VE calculated as 1-HR was similar.



Any brand, vaccinated vs unvaccinated



Cervarix, vaccinated vs unvaccinated



Gardasil, vaccinated vs unvaccinated

Figure 12.4. 5, 10, 15y RR, and Total RR & HR of vaccinated vs unvaccinated after PS-matching for each brand separately, all vaccine brands together and for CIN2+ and conisation outcomes.

12.3.2 Dose comparisons

Table 12.5 shows the risk ratio of the outcome in those having 2 or more doses vs those having 1 dose, and of 3 or more doses vs those having 2 at 5, 10, and 15 years after vaccination date and Hazard Ratios for the entire period. Because of the low number of cases, most analyses couldn't be performed and those that could be modelled had very wide confidence intervals. We were also not able to pool analyses due to heterogeneity.

Relative VE (rVE) of 2 doses vs 1 dose against CIN2+ in 15y was of 76% with large confidence intervals (-103% to 94%) in SIDIAP, and of 3 doses vs 1 dose was of 88% (-16% to 99%) in CPRD-GOLD. Conisation could not be assessed because there were not enough number of events.

Table 12.5. 5, 10, 15y RR, and Total RR & HR for different dose schedules for each brand separately, all vaccine brands together and for CIN2+ and conisation outcomes.

All brands

Database name	Outcome	Matched				
		5y RR (95% CI)	10y RR (95% CI)	15y RR (95% CI)	All RR (95% CI)	All HR (95% CI)
Two or more vs one						
1) CPRD-GOLD	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
2) SIDIAP	CIN 2+	-	-	0.34 (0.06, 2.03)	0.34 (0.06, 2.03)	0.4 (0.07, 2.39)
	Conisation					
3) NLHR	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
Three or more vs two						
1) CPRD-GOLD	CIN 2+	-	-	0.12 (0.01, 1.16)	0.12 (0.01, 1.16)	0.16 (0.02, 1.54)
	Conisation	-	-	-	-	-
2) SIDIAP	CIN 2+	-	-	2.14 (0.27, 17.14)	2.15 (0.27, 17.14)	1.7 (0.21, 13.61)
	Conisation	-	-	-	-	-
3) NLHR	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-

Cervarix

Database name	Outcome	Matched				
		5y RR (95% CI)	10y RR (95% CI)	15y RR (95% CI)	All RR (95% CI)	All HR (95% CI)
Two or more vs one						
1) CPRD-GOLD	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
2) SIDIAP	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
3) NLHR	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
Three or more vs two						
1) CPRD-GOLD	CIN 2+	-	-	0.12 (0.01, 1.15)	0.12 (0.01, 1.15)	0.16 (0.02, 1.53)
	Conisation	-	-	-	-	-
2) SIDIAP	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
3) NLHR	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-

Gardasil

Database name	Outcome	Matched				
		5y IRR (95% CI)	10y IRR (95% CI)	15y IRR (95% CI)	All IRR (95% CI)	All HR (95% CI)
Two or more vs one						
1) CPRD-GOLD	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
2) SIDIAP	CIN 2+	0.96 (0, Inf)	0.97 (0, Inf)	0.34 (0.06, 2.02)	0.34 (0.06, 2.02)	0.4 (0.07, 2.41)
	Conisation	-	-	-	-	-
3) NLHR	CIN 2+	-	-	-	-	-
	Conisation	-	-	-	-	-
Three or more vs two						
1) CPRD-GOLD	CIN 2+	-	-	-	-	-
	Conisation (broad)	-	-	-	-	-
2) SIDIAP	CIN 2+	-	-	1.38 (0.16, 11.78)	1.38 (0.16, 11.78)	1.08 (0.13, 9.32)
	Conisation	-	-	-	-	-
3) NLHR	CIN 2+	-	-	-	-	-
	Conisation (broad)	-	-	-	-	-

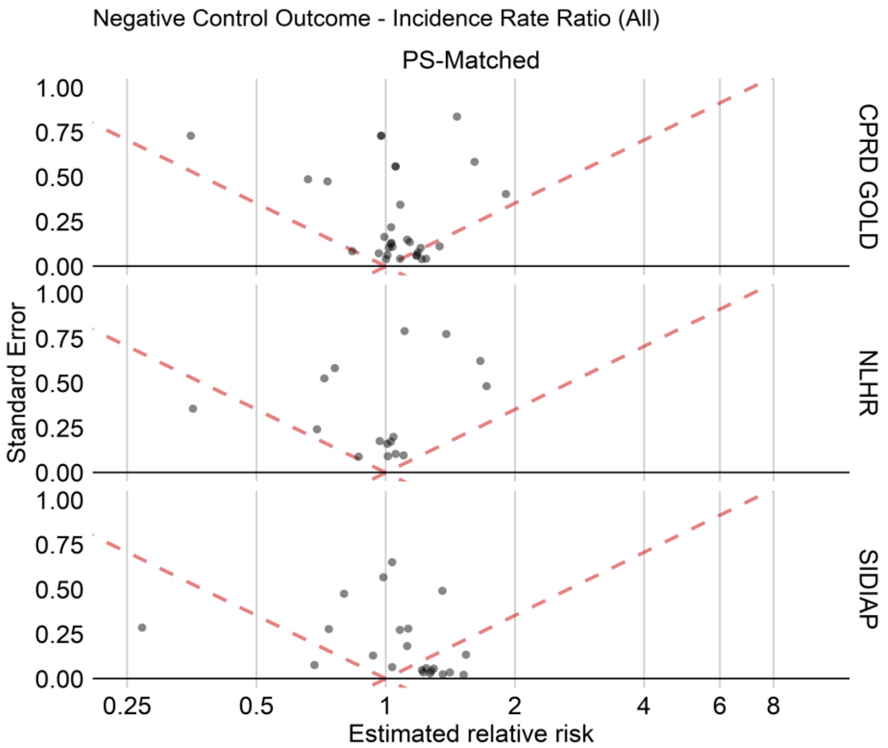
12.4 Additional outcomes

12.4.1 Negative control outcomes

We repeated the matched analyses with 38 different outcomes that are a priori unrelated to HPV, or HPV vaccination, to act as negative controls. **Figure 12.5** shows the results of these analyses.

In CPRD-GOLD, RR analyses show some potential residual confounding, with 82.5% of estimates including null and a tendency towards overestimating RR. Similar results are shown in the estimation of HR. In SIDIAP, the results suggest more residual confounding, with 70% of estimates including null and a tendency towards overestimating RR. Similar results are shown in the estimation of HR. In NLHR, both in HR and RR, estimates are centred around the null, suggesting little systematic bias.

Matched Incidence Ratios



Matched Hazard Ratios

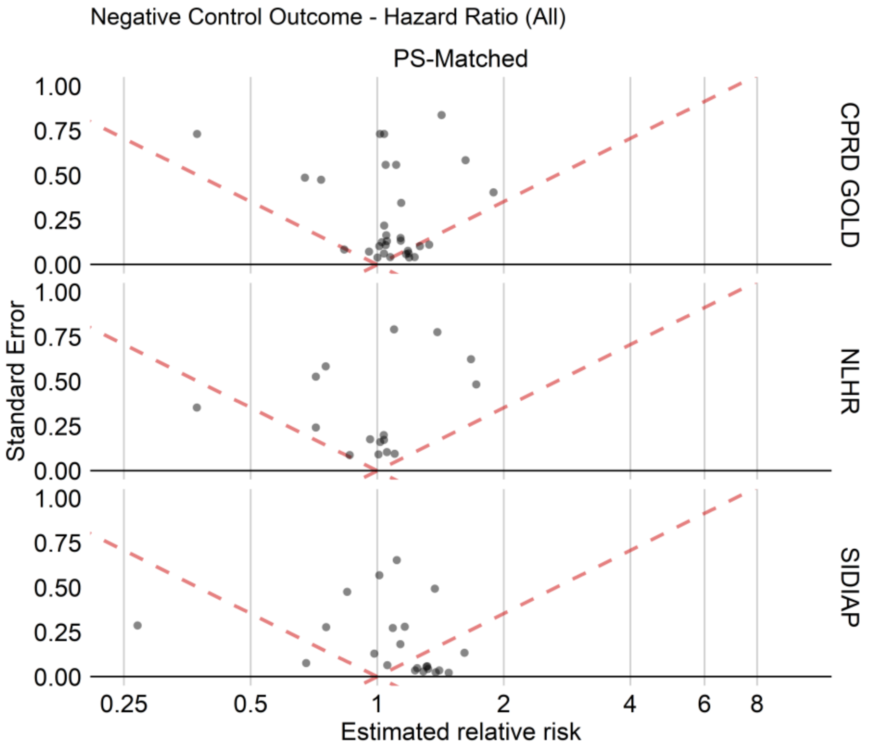


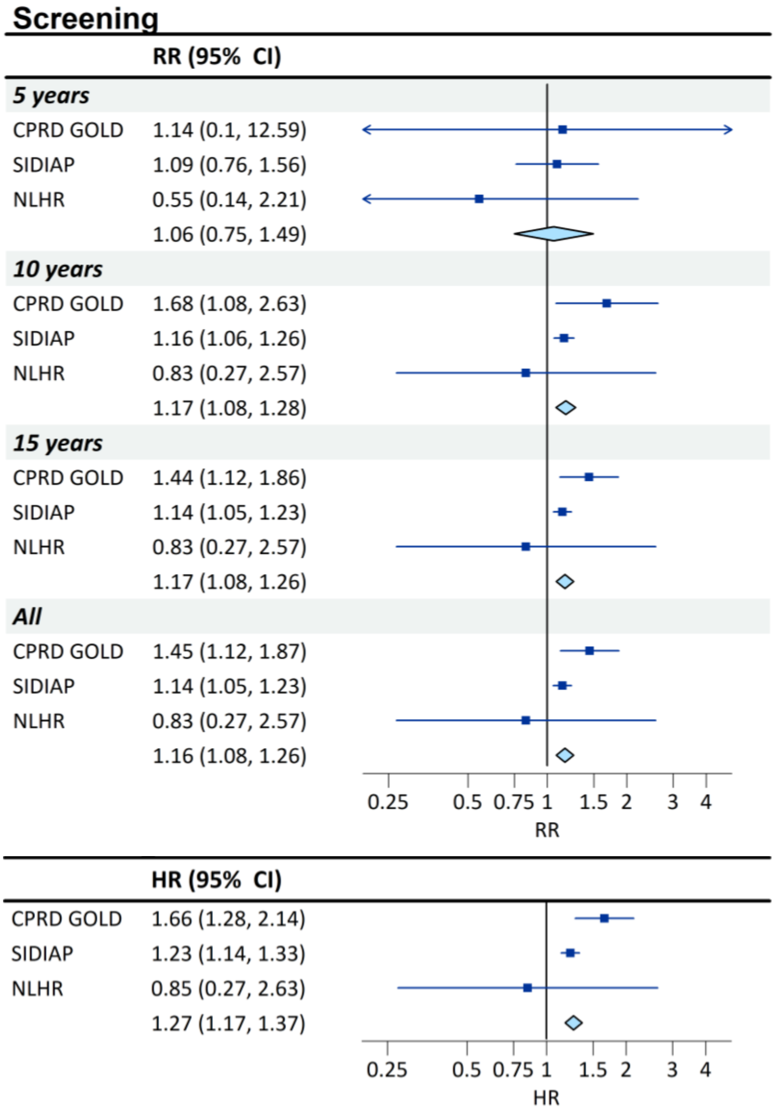
Figure 12.5. Systematic error plot of NCOs.

12.4.2 Screening

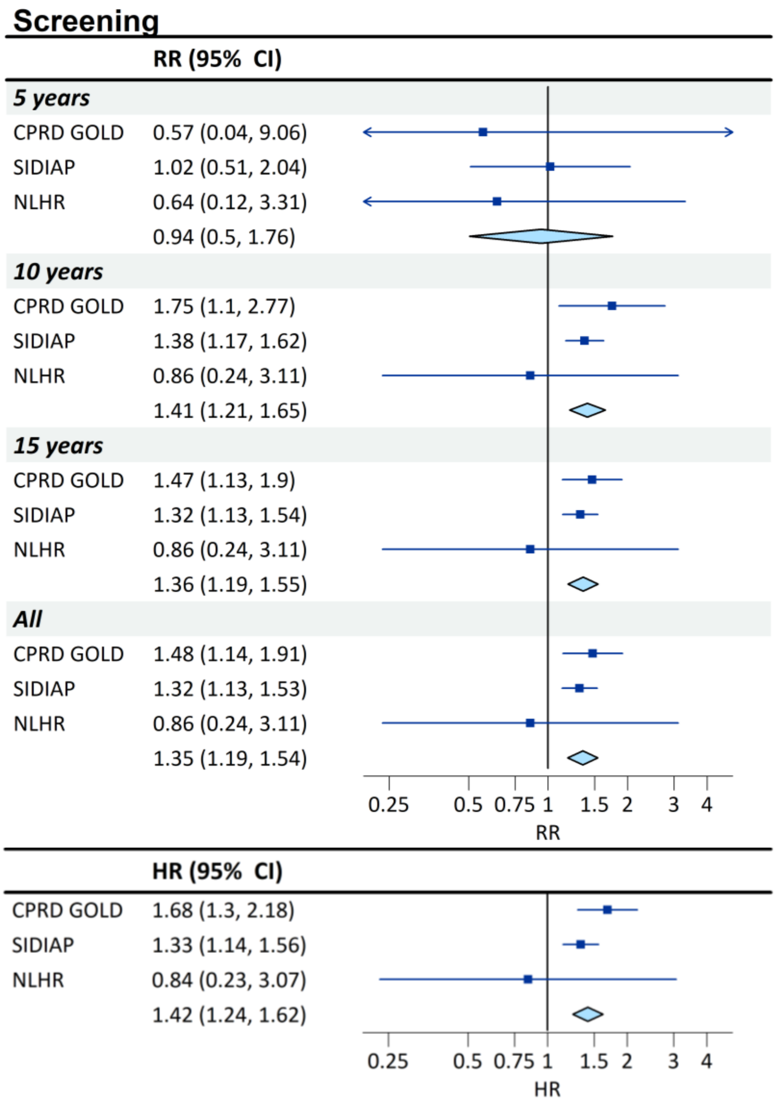
As for the chance of screening, we observed an increased uptake of cervical screening in vaccinated compared to the matched unvaccinated pairs, with an overall increase of 45% (12% to 87%) in CPRD-GOLD and 14% (5% to 23%) in SIDIAP, with no differential screening in NLHR (RR 0.83 (0.27 to 2.57)). The resulting meta-analytic RR for screening was 1.16 (1.08 to 1.26).

By brand, for Cervarix there was an increased risk of 48% (14% to 91%) in CPRD-GOLD and 32% (13% to 53%) in SIDIAP, with no differential screening in NLHR. The meta-analytic RR was 1.42 (1.24 to 1.62).

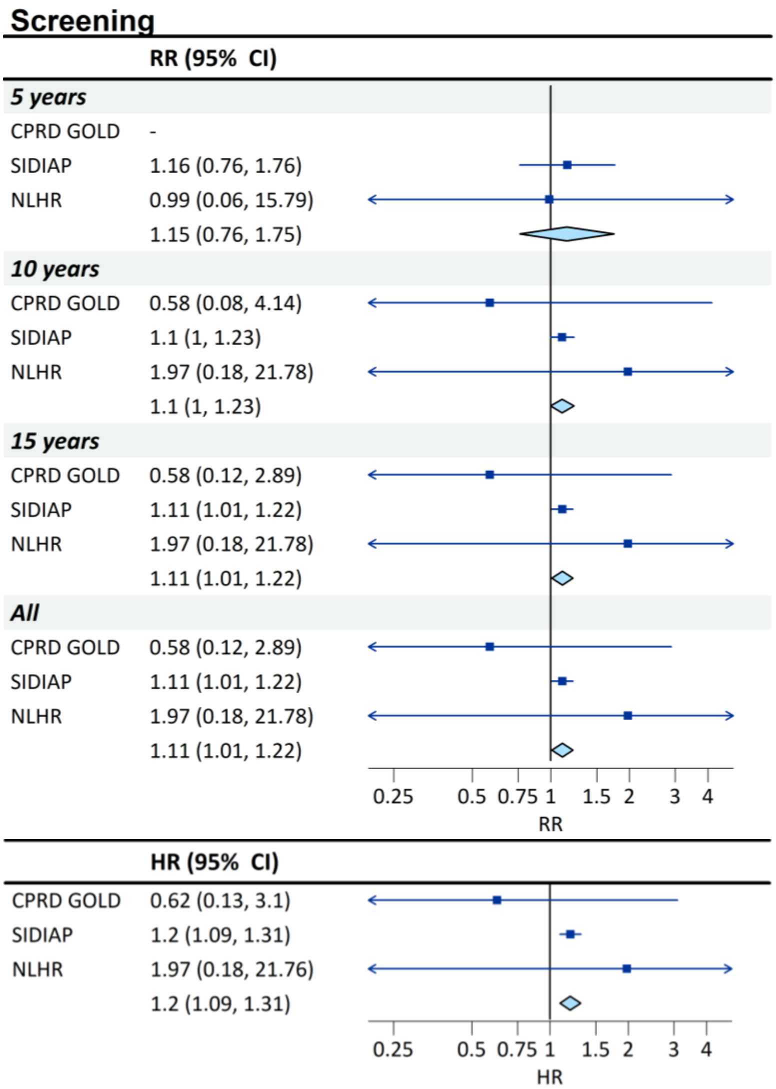
As for Gardasil, there was a decrease of 42% (-189% to 88%) in CPRD-GOLD and an increase of 11% (1% to 22%) in SIDIAP, and no differential screening in NLHR. The resulting meta-analytic RR was 1.11 (1.01 to 1.22).



Any brand, vaccinated vs unvaccinated



Cervarix, vaccinated vs unvaccinated



Gardasil, vaccinated vs unvaccinated

Figure 12.6. 5, 10, 15y RR, and Total RR & HR of getting a cervical cancer screening for vaccinated to unvaccinated for each brand separately, all vaccine brands together

12.5 Sensitivity analyses

We performed two sensitivity analyses, one restricting the analyses for CPRD-GOLD to those GP practices – year with a coverage of over 60%, and another one not censoring unvaccinated once they become vaccinated.

12.5.1 Restriction to 60% coverage (CPRD-GOLD)

After restricting women in CPRD-GOLD to those years and GP practices where the coverage was over 60%, we had a much lower number of events, and we could only estimate a VE of -5% CI95(-131 to 52%) for Cervarix. All results for this analysis are shown in **Table 12.6.**

Table 12.6. 5, 10, 15 RR, and Total RR & HR - for matched and unmatched (Vaccinated vs unvaccinated, minimum coverage 60% for CPRD-GOLD database)

		Matched				
Database name	**Brand - Outcome**	**5y IRR (95% CI)**	**10y IRR (95% CI)**	**15y IRR (95% CI)**	**All IRR (95% CI)**	**All HR (95% CI)**
CPRD-GOLD	**Cervarix - CIN2+**	0 (0, Inf)	0.94 (0.23, 3.74)	1.05 (0.48, 2.31)	1.05 (0.48, 2.31)	1.24 (0.56, 2.72)
	Gardasil - CIN2+	-	-	-	-	-

12.5.2 Not censoring unvaccinated if they become vaccinated

In this analysis, we did not censor the unvaccinated if they got a further vaccination after we assess vaccination status at 15 years old. **Table 12.7** summarises the results. The results show similar but slightly higher VE than in the censored analyses.

Table 12.7. 5, 10, 15 RR, and Total RR & HR - for matched and unmatched (Vaccinated vs unvaccinated).

Cervarix

Database name	Outcome	Matched				
		5y IRR (95% CI)	10y IRR (95% CI)	15y IRR (95% CI)	All IRR (95% CI)	All HR (95% CI)
1) CPRD-GOLD	CIN (grades 2-3)	-	0.98 (0.23, 4.09)	0.55 (0.27, 1.15)	0.55 (0.27, 1.15)	0.63 (0.3, 1.3)
	Conisation (broad)	-	-	0.59 (0.04, 9.47)	0.59 (0.04, 9.47)	0.67 (0.04, 10.69)
2) SIDIAP	CIN (grades 2-3)	-	1.07 (0.12, 9.6)	0.67 (0.13, 3.47)	0.67 (0.13, 3.47)	0.68 (0.13, 3.51)
	Conisation (broad)	-	-	-	-	-
3) NLHR	CIN (grades 2-3)	-	-	-	-	-
	Conisation (broad)	0.56 (0.19, 1.6)	0.47 (0.19, 1.17)	0.47 (0.19, 1.17)	0.47 (0.19, 1.17)	0.47 (0.2, 1.16)

Gardasil

Database name	Outcome	Matched				
		5y IRR (95% CI)	10y IRR (95% CI)	15y IRR (95% CI)	All IRR (95% CI)	All HR (95% CI)
1) CPRD-GOLD	CIN (grades 2-3)	-	-	0.29 (0.03, 3.19)	0.29 (0.03, 3.19)	0.32 (0.03, 3.49)
	Conisation (broad)	-	-	-	-	-
2) SIDIAP	CIN (grades 2-3)	-	0.52 (0.17, 1.55)	0.53 (0.29, 0.98)	0.53 (0.29, 0.98)	0.59 (0.32, 1.09)
	Conisation (broad)	-	-	-	-	-
3) NLHR	CIN (grades 2-3)	-	-	-	-	-
	Conisation (broad)	0.49 (0.04, 5.45)	0.49 (0.04, 5.45)	0.49 (0.04, 5.45)	0.49 (0.04, 5.45)	0.49 (0.04, 5.44)

13. DEVIATIONS FROM ORIGINAL PROTOCOL

Since the publication of the protocol, updated analyses have been conducted, as shown below together with a description of the impact on the overall findings:

a. Calculation of vaccine coverage and comparison with public health data as a diagnostic for vaccine records completeness.

b. Replacement of the IQVIA DA Germany by the NLHR data source after diagnostics, due to the high risk of misspecification of index dates of CIN2+ and cancer, and the potential misclassification on vaccination.

c. Comparative analyses for cervical cancer were not performed due to the low number of cases.

d. Comparative analyses of dose schedules by brand were mostly not possible due to the low number of cases.

e. Addition of HPV screening outcome to measure the potential differential uptake of screening between groups that could impact outcomes, where a group may be less likely to get screened and thus less likely to have a CIN2+ diagnosed.

f. Further analyses with unmatched cohorts and only age, year, with region matched cohorts presented in the Shiny app only.

g. In the updated version of this report, two ASMD values of SIDIAP have been updated (Table 12.1), as they were rounded incorrectly in the first version.

    - Ibuprofen 20MG/ML Oral Suspension: Original ASMD PS-Matched = 0.10; Updated ASMD PS-Matched = 0.11.
    - Albuterol: Original ASMD PS-Matched = 0.06; Updated ASMD PS-Matched = 0.05.

    This change has no impact on the results.

h. NLHR data in this report have been revised to incorporate emigration data, which was not available at the time of the initial analyses. This update impacts the end-of-observation date for some participants, which in turn affects the study's inclusion and exclusion criteria (specifically the matching process) and, consequently, its findings. Updates in this report include:

    8.6 Section 12.1 and Figure 12.1: Update of NLHR study participants. Final sample for NLHR decreased from 17,900 vaccinated and 4,574 unvaccinated to 14,855 vaccinated and 4,073 unvaccinated.

    8.7 Additionally, while at the time of the initial analyses only Cervarix was recorded in NLHR, the updated dataset includes Gardasil and Gardasil-9 records.

    8.8 Section 12.1, Figure 12.2, and Table 12.1: Update of the absolute standardised mean difference before and after ps-matching, according to the new NLHR study population. While estimates for NLHR only changed, the overall findings did not.

    8.9 Section 12.2.1 and Table 12.2. Update of the number of doses for each vaccine brand for the new NLHR study population. Same as 8.6.

    8.10 Section 12.2.2, Table 12.3: Update of the baseline characteristics based on the new NLHR study population. Estimates slightly changed for NLHR but not significantly, except for an increase in follow up time available.

8.11 Section 12.2 and Table 12.4: Update of the incidence rates based on the new NLHR study population. In the new NLHR dataset, there is an increased number of conisations in both vaccinated and unvaccinated. Incidence of conisation point estimates have become 3-4 times higher in the unvaccinated.

8.12 Section 12.3 and Figure 12.4, Table 12.5: Update of conisation results based on the new NLHR study population. With the new data from NLHR, the VE against conisation for NLHR has increased significantly, and the confidence interval no longer includes the null. This new estimate has also moved the metanalytic estimate to a higher VE and its confidence interval no longer includes the null. This reinterprets the overall VE estimate as 58% instead of 27%, although both estimates fall within each other's confidence intervals.

8.13 Section 12.4, Figure 12.5: Update of the negative control outcomes results based on the new NLHR study population. Point estimates for negative control outcomes in NLHR changed slightly with no implications for the interpretation.

8.14 Section 12.4.5: Update of the screening results based on the new NLHR study population. Point estimates for screening risk in NLHR changed slightly with no implications for the interpretation.

8.15 Section 12.5, Table 12.7. Update of the “not censoring unvaccinated if they become vaccinated” sensitivity analysis results based on the new NLHR study population. Estimates for conisation in NLHR changed similarly to 8.11, with no changes for interpretation.

i. In the updated version of this report, the total follow-up values in Section 12.2/Invasive cervical cancer values have been revised. The previous version reported the values from the non-matched cohorts; these have now been updated to reflect the values in the matched cohorts, for consistency and clarity.

14. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions are not collected or analysed as part of this evaluation. The nature of this non-interventional evaluation, through the use of secondary data, does not fulfil the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf).

Only in the case of prospective data collection, there is a need to describe the procedures for the collection, management, and reporting of individual cases of adverse events/adverse reactions.

15. DISCUSSION

15.1 Key results

Given the low number of invasive cervical cancer cases identified, with less than 5 events per database, we were not able to assess the effectiveness of HPV vaccination in the prevention of invasive cervical cancer. This was likely due to a lower-than-expected number of participants and a shorter than expected follow-up duration in all three databases, but particularly in NLHR, due to restrictions with the available data, which spans only from 2018.

Against CIN2+, we estimated a pooled VE of 42% overall, of 38% for Cervarix and of 41% for Gardasil. These estimates are in line with previous randomised controlled trials and meta-analyses of RCTs, (33) as detailed below. Regarding conisation, used as a proxy for CIN2+, we obtained a pooled VE of 58%.

For the interpretation of all these results, it should be taken into consideration that our analyses suggest that the uptake of cervical screening in the vaccinated was higher than the matched unvaccinated participants, with an increased chance of screening of 45% in CPRD-GOLD, and 14% in SIDIAP. This could result in an increased probability of diagnosis of CIN2+ (and subsequent conisation) in the vaccinated, therefore driving the VE estimates to the null.

We were also unable to assess the effect of different dose schedules (1 vs 2+ doses, 2 vs 3+ doses) due to high uncertainty in the results related to the low number of participants, limited follow-up and very low number of cases mentioned above.

15.2 Limitations of the research methods

This study is informed by routinely collected health care data and so data quality issues, and adequate capture of the variables of interest in primary care data, must be considered.

Overall, the most important limitation is the number of women and the length of follow up available for analysis. Due to the long latency to develop invasive cervical cancer, the mean 7 years of follow up in CPRD-GOLD, 10 in SIDIAP, and 5 in NLHR proved insufficient to capture a sufficient number of cases to accurately estimate VE. This could be further impacted by the differential uptake of cervical screening, with screening programmes starting at 25 years old in the 3 participant countries, which would impact the likelihood of diagnosis of CIN2+ and subsequent conisation.

Additionally, we found evidence of exposure misclassification with incomplete data on HPV vaccination in CPRD-GOLD. However, SIDIAP and NLHR seemed to have complete information on vaccination. This could lead to misclassification of exposure, and, as in this case it might be related to health seeking behaviour, to bias in the CPRD-GOLD results. We conducted a sensitivity analysis including only those GP practices and years where we deemed there was complete information, but this restricted the study population and led to imprecise estimates.

Comparing vaccine brands was not feasible, as most vaccination programmes only had one schedule active at each point in time.

There may have been also potentially incomplete outcomes in the 3 databases. In SIDIAP, for example, sensitivity of recording cervical cancer in primary care using ICD-10 codes has been reported to be very low (34). Adding information from sexual and reproductive health clinics may have identified more CIN2+ and most of the conisations but was not yet available. This could also have led to not finding all the in-situ carcinomas not coded as ICD-10 cancers.

Overall, using conisation as a proxy for CIN2+ may not be reliable in some settings where clinicians may decide not to treat CIN2+, especially in younger females. In addition, conisation practice would vary by institution and health care system and can have an impact on outcomes. In this study, conisation was initially chosen as a proxy due to uncertainty about the availability and completeness of CIN2+ data at the time of study design, therefore, this data should be considered as complementary.

Although every effort was made to minimise confounding, there may still be confounding due to unmeasured confounders or effect modification. Our analysis of negative control outcomes indeed suggests some residual confounding, particularly for the SIDIAP database. The main confounders that we were unable to measure are those related to sexual behaviour and socio-economic factors.

15.3 Interpretation and generalisability

The effectiveness of HPV vaccination in preventing invasive cervical cancer could not be assessed due to low outcome numbers in both vaccinated and unvaccinated matched participants, with less than 5 events in all databases. This could be due to the relatively short follow up in most databases, although longer than most studies published to date. Similarly, the low number of participants eligible and short follow-up in the data available limited our ability to estimate the impact of dose and brand on study outcomes.

The observed effectiveness against CIN2+ is similar to the one yielded by studies in similar settings, against CIN2+ regardless of HPV type, between 26% and 67% as a Cochrane review shows (33). Our findings using a target trial emulation framework therefore replicate those from previous RCTs, despite all the limitations mentioned above. Additionally, our results provide reassurance of a large protective effect based on European settings and populations and routine healthcare conditions, which differ from those in previous studies, performed mostly in non-European countries (33).

It is worth noting that our study observed a differential uptake of cervical screening, with an increased probability of screening in the vaccinated compared to matched unvaccinated women. This is likely due to a healthy vaccinee effect and a higher use of health services amongst those vaccinated. This, together with the evidence of an increased risk of negative control outcomes observed in CPRD-GOLD and SIDIAP, points to an underestimation of vaccine effectiveness in our study. This aligns with our estimates being in the lower range of the results observed in previous trials (33).

Given what we know about the aetiology and pathophysiology of invasive cervical cancer, our findings of a reduction in the risk of CIN2+ could lead to reductions in the risk of developing invasive cervical cancer in the longer term.

Our results contrast with some recent observational data on the effectiveness of HPV vaccination (12, 13, 35, 36), but these focus on CIN3+ and calculate effectiveness at the population level, and it is in line with a recent systematic review for CIN2+ regardless of HPV type.(37) Our unmatched results (accessible on the Shiny app) yield similarly high vaccine effectiveness. These differences show that the target trial emulation framework produces results closer to the ones of clinical trials, while focussing on individual level efficacy. Conversely, unmatched cohort and ecological analyses like those recently published provide different estimates, potentially being more influenced by other factors, like herd immunity. However, they are also more likely to be confounded, limited by ecologic fallacy, and less likely to provide a valid causal effect estimate, as they do not align time zero or follow-up, hence potentially comparing unvaccinated older women with younger vaccinated ones.

For future work on this topic, care should be put in selecting data partners with complete vaccination and cancer/cytology data coverage and with a complete and long enough follow up (10y+). This could increase the number of events detected and the number years of follow up that are currently limiting the conclusions. Additional methods for the correction of residual confounding, like negative control outcome empirical calibration or accounting for the differences in screening, could be considered to provide a more accurate estimate of vaccine effectiveness.

16. CONCLUSION

We were unable to assess the causal effect of HPV vaccines against cervical cancer using a target trial emulation design due to limited number of outcomes and limited available follow-up to account for the long cancer latency period post-vaccination. For CIN2+ and conisation, vaccine effectiveness seems in the lower range of what is known with the evidence from clinical trials but is potentially underestimated by differences in screening rates between vaccinated and unvaccinated groups.

17. REFERENCES

1. Office of National Statistics. Cancer registration statistics, England: 2017 2017 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017#the-three-most-common-cancers-vary-by-sex-and-age-group.

2. European Centre for Disease Prevention and Control. Factsheet about human papillomavirus 2018 [Available from: https://www.ecdc.europa.eu/en/human-papillomavirus/factsheet.

3. World Health Organization. Cervical Cancer Elimination Initiative 2020 [Available from: https://www.who.int/initiatives/cervical-cancer-elimination-initiative.

4. Garland SM, Hernandez-Avila M, Wheeler CM, Perez G, Harper DM, Leodolter S, et al. Quadrivalent vaccine against human papillomavirus to prevent anogenital diseases. N Engl J Med. 2007;356(19):1928-43.

5. Paavonen J, Naud P, Salmerón J, Wheeler CM, Chow SN, Apter D, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. Lancet. 2009;374(9686):301-14.

6. Joura EA, Giuliano AR, Iversen O-E, Bouchard C, Mao C, Mehlsen J, et al. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. N Engl J Med. 2015;372(8):711-23.

7. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 2010;11(11):1048-56.

8. European Centre for Disease Prevention and Control. Guidance on HPV vaccination in EU countries: focus on boys, people living with HIV and 9-valent HPV vaccine introduction 2023 [Available from: https://www.ecdc.europa.eu/en/publications-data/guidance-hpv-vaccination-eu-focus-boys-people-living-hiv-9vHPV-vaccine.

9. European Medicines Agency. Gardasil 2022 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/gardasil.

10. European Medicines Agency. Cervarix 2023 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/cervarix.

11. European Medicines Agency. Gardasil 9 2023 [Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/gardasil-9.

12. Falcaro M, Castanon A, Ndlela B, Checchi M, Soldan K, Lopez-Bernal J, et al. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. Lancet. 2021;398(10316):2084-92.

13. Lei J, Ploner A, Elfstrom KM, Wang J, Roth A, Fang F, et al. HPV Vaccination and the Risk of Invasive Cervical Cancer. N Engl J Med. 2020;383(14):1340-8.

14. Mesher D, Panwar K, Thomas SL, Edmundson C, Choi YH, Beddows S, et al. The impact of the national HPV vaccination program in England using the bivalent HPV vaccine: surveillance of type-specific HPV in young females, 2010–2016. The Journal of infectious diseases. 2018;218(6):911-21.

15. Palmer T, Wallace L, Pollock KG, Cuschieri K, Robertson C, Kavanagh K, et al. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12-13 in Scotland: retrospective population study. BMJ. 2019;365:l1161.

16. Drolet M, Bénard É, Pérez N, Brisson M, Ali H, Boily M-C, et al. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. The Lancet. 2019;394(10197):497-509.

17. Stier EA, Chigurupati NL, Fung L. Prophylactic HPV vaccination and anal cancer. Hum Vaccines Immunother. 2016;12(6):1348-51.

18. Tsentemeidou A, Fyrmpas G, Stavrakas M, Vlachtsis K, Sotiriou E, Poutoglidis A, et al. Human papillomavirus vaccine to end oropharyngeal cancer. A systematic review and meta-analysis. J Sex Transm Dis. 2021;48(9):700-7.

19. Abbas KM, van Zandvoort K, Brisson M, Jit M. Effects of updated demography, disability weights, and cervical cancer burden on estimates of human papillomavirus vaccination impact at the global, regional, and national levels: a PRIME modelling study. Lancet Glob Health. 2020;8(4):e536-e44.

20. Brotherton JML, Sundström K. More evidence suggesting that 1-dose human papillomavirus vaccination may be effective. Cancer. 2020;126(8):1602-4.

21. Kreimer AR, Sampson JN, Porras C, Schiller JT, Kemp T, Herrero R, et al. Evaluation of Durability of a Single Dose of the Bivalent HPV Vaccine: The CVT Trial. J Natl Cancer Inst. 2020;112(10):1038-46.

22. Sankaranarayanan R, Prabhu PR, Pawlita M, Gheit T, Bhatla N, Muwonge R, et al. Immunogenicity and HPV infection after one, two, and three doses of quadrivalent HPV vaccine in girls in India: a multicentre prospective cohort study. Lancet Oncol. 2016;17(1):67-77.

23. Baisley K, Kemp TJ, Kreimer AR, Basu P, Changalucha J, Hildesheim A, et al. Comparing one dose of HPV vaccine in girls aged 9–14 years in Tanzania (DoRIS) with one dose of HPV vaccine in historical cohorts: an immunobridging analysis of a randomised controlled trial. Lancet Glob Health. 2022;10(10):e1485-e93.

24. Man I, Georges D, de Carvalho TM, Ray Saraswati L, Bhandari P, Kataria I, et al. Evidence-based impact projections of single-dose human papillomavirus vaccination in India: a modelling study. Lancet Oncol. 2022;23(11):1419-29.

25. Prem K, Choi YH, Bénard É, Burger EA, Hadley L, Laprise J-F, et al. Global impact and cost-effectiveness of one-dose versus two-dose human papillomavirus vaccination schedules: a comparative modelling analysis. BMC medicine. 2023;21(1):313.

26. Brotherton JM, Budd A, Rompotis C, Bartlett N, Malloy MJ, Andersen RL, et al. Is one dose of human papillomavirus vaccine as effective as three?: A national cohort analysis. Papillomavirus Res. 2019;8:100177.

27. Rodriguez AM, Zeybek B, Vaughn M, Westra J, Kaul S, Montealegre JR, et al. Comparison of the long-term impact and clinical outcomes of fewer doses and standard doses of human papillomavirus vaccine in the United States: a database study. Cancer. 2020;126(8):1656-67.

28. Immunisation TJCoVa. JCVI statement on a one-dose schedule for the routine HPV immunisation programme. 2022.

29. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Van Staa T, et al. Data resource profile: clinical practice research datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.

30. García Gil MdM, Hermosilla E, Prieto-Alhambra D, Fina F, Rosell M, Ramos Blanes R, et al. Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). Informatics in Primary Care, 2011, vol 19, p 135-145. 2011.

31. Ali MS, Groenwold RH, Pestman WR, Belitser SV, Roes KC, Hoes AW, et al. Propensity score balance measures in pharmacoepidemiology: a simulation study. Pharmacoepidemiol Drug Saf. 2014;23(8):802-11.

32. Burn E, Catala M. CodelistGenerator: Generate Code Lists for the OMOP Common Data Model 2023 [Available from: https://cran.r-project.org/web/packages/CodelistGenerator/index.html.

33. Arbyn M, Xu L, Simoens C, Martin-Hirsch PP. Prophylactic vaccination against human papillomaviruses to prevent cervical cancer and its precursors. Cochrane Database Syst Rev. 2018;5(5):CD009069.

34. Recalde M, Manzano-Salgado CB, Diaz Y, Puente D, Garcia-Gil MDM, Marcos-Gragera R, et al. Validation Of Cancer Diagnoses In Electronic Health Records: Results From The Information System For Research In Primary Care (SIDIAP) In Northeast Spain. Clin Epidemiol. 2019;11:1015-24.

35. Falcaro M, Soldan K, Ndlela B, Sasieni P. Effect of the HPV vaccination programme on incidence of cervical cancer and grade 3 cervical intraepithelial neoplasia by socioeconomic deprivation in England: population based observational study. BMJ. 2024;385:e077341.

36. Palmer TJ, Kavanagh K, Cuschieri K, Cameron R, Graham C, Wilson A, et al. Invasive cervical cancer incidence following bivalent human papillomavirus vaccination: a population-based observational study of age at immunization, dose, and deprivation. J Natl Cancer Inst. 2024;116(6):857-65.

37. Christine Schmucker PK, Timo Brugger, Lea Gorenflo, Eberhard Thörel, Waldemar Siemens, Marianne, Röbl-Mathieu MA, Thomas Harder, Joerg J Meerpohl, EU/EEA NITAG Collaboration HPV Working Group*, representatives* aE. Efficacy, effectiveness and safety of HPV vaccination in women with conisation: a systematic review and meta-analyses Stockholm: ECDC; 2024.

18. ANNEXES

Appendix I: Table 1. Codes used to identify vaccines in each of the data partners.

Only those present in the datasets, full list can be found here.

Standard concept id	Standard concept name	Data Partner
36789910	L1 protein, Human papillomavirus type 11 Vaccine / L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 Vaccine / L1 protein, Human papillomavirus type 6 Vaccine Injectable Suspension [Gardasil]	CPRD-GOLD
40167170	L1 protein, human papillomavirus type 16 vaccine / L1 protein, human papillomavirus type 18 vaccine Injectable Suspension [Cervarix]	CPRD-GOLD
40213321	HPV, unspecified formulation	CPRD GOLD
40753446	L1 protein, Human papillomavirus type 11 Vaccine / L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 Vaccine / L1 protein, Human papillomavirus type 31 Vaccine / ... Injectable Suspension [Gardasil 9]	CPRD-GOLD
44025856	L1 protein, Human papillomavirus type 11 Vaccine / L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 / L1 protein, Human papillomavirus type 6 Injectable Suspension	CPRD-GOLD
44055725	L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 Injectable Suspension	CPRD GOLD
36789911	L1 protein, Human papillomavirus type 11 Vaccine / L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 Vaccine / L1 protein, Human papillomavirus type 6 Vaccine Injectable Suspension	SIDIAP
36893469	L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 Vaccine Injection	SIDIAP
40213322	Human Papillomavirus 9-valent vaccine	SIDIAP
40150715	L1 protein, human papillomavirus type 11 vaccine / L1 protein, human papillomavirus type 16 vaccine / L1 protein, human papillomavirus type 18 vaccine / L1 protein, human papillomavirus type 6 vaccine Prefilled Syringe	NLHR
41144528	L1 protein, Human papillomavirus type 11 Vaccine / L1 protein, Human papillomavirus type 16 Vaccine / L1 protein, Human papillomavirus type 18 Vaccine / L1 protein, Human papillomavirus type 6 Vaccine Prefilled Syringe [Silgard]	NLHR
35408900	Human Papillomavirus Injectable Suspension [Gardasil]	NLHR
35412768	Human Papillomavirus Injectable Suspension	NLHR
35753734	Human Papillomavirus Injectable Suspension [Cervarix]	NLHR
36267065	Human Papillomavirus Injectable Suspension [Gardasil 9]	NLHR

Appendix I: Table 2. Codes used to identify outcomes in each of the data partners.

Only those present in the datasets, full list can be found here.

CIN 2+

Standard concept id	Standard concept name
194611	Carcinoma in situ of uterine cervix
196165	Cervical intraepithelial neoplasia grade 2
4098948	Cervical intraepithelial neoplasia grade III with severe dysplasia
4243120	Carcinoma in situ of endocervix
4069557	Squamous intraepithelial neoplasia, high grade
4243874	Carcinoma in situ of exocervix
45757384	High grade squamous intraepithelial lesion on vaginal Papanicolaou smear
45763589	High grade squamous intraepithelial lesion on cervical Papanicolaou smear

Conisation

Standard concept id	Standard concept name
4003896	Cervix excision
4046830	Loop electrosurgical excision procedure of cervix
4074137	Loop diathermy cone biopsy of cervix uteri
4074291	Laser cone biopsy of cervix uteri
4127884	Diathermy of cervix
4181912	Cone biopsy of cervix
4213044	Cold knife cone biopsy of cervix

Invasive cervical cancer

Standard concept id	Standard concept name
198984	Malignant tumour of cervix
4092515	Malignant neoplasm, overlapping lesion of cervix uteri
4095156	Malignant neoplasm of endocervical canal
4095158	Malignant neoplasm of squamocolumnar junction of cervix
4157449	Malignant neoplasm of endocervix
4162876	Malignant neoplasm of exocervix
4069557	Squamous intraepithelial neoplasia, high grade
196359	Primary malignant neoplasm of uterine cervix
436358	Primary malignant neoplasm of exocervix
441805	Primary malignant neoplasm of endocervix
45770837	Cytological evidence of malignancy on cervical Papanicolaou smear

Screening

Standard concept id	Standard concept name	Data Partner
4235948	Sampling of cervix for Papanicolaou smear	SIDIAP
4064912	Cancer cervix screen-no result yet	CPRD-GOLD
4062484	Screening for malignant neoplasm of cervix	CPRD-GOLD
45763689	Human papilloma virus screening	CPRD-GOLD

Table 3. List of negative control outcomes**.**

Constipation	Glaucoma
Ulcer of lower extremity	Otitis externa
Cellulitis of lower limb	Osteopenia
Iron deficiency anaemia	Dry eyes
Wax in ear canal	Ulcer of foot
Actinic keratosis	Squamous cell carcinoma of skin
Cataract	Acquired hypothyroidism
Hearing loss	Age related macular degeneration
Hypothyroidism	Acid reflux
Rectal haemorrhage	Laceration of lower leg
Foot pain	Inguinal hernia
Urinary incontinence	Traumatic wound
Bilateral cataracts	Gallstone
Vitamin d deficiency	Pressure ulcer
Basal cell carcinoma of skin	Polyp of colon
Haemorrhoids	Impacted cerumen
Senile hyperkeratosis	Laceration injury
Intraocular pressure left eye	Open wound of lower leg
Hearing difficulty	Acute conjunctivitis