Accelerated Development of VAccine beNefit-risk Collaboration in Europe

IMI JU Grant Agreement nº115557

POC 1.2 study protocol

Testing a system for near real-time monitoring of vaccination coverage, benefits and risks in Europe with acellular pertussis vaccines as test case.

WP5 – Proof-of-concept of a framework
to perform near-real time vaccine benefit-risk monitoring

Disclaimer: *The study described in this protocol is conducted as part of the IMI ADVANCE project with the aim to test methodological aspects of the design, conduct and reporting of studies for vaccine benefit-risk monitoring activities.*

The protocol presented herein relates solely to the testing of these methodologies and is not intended to inform regulatory or clinical decisions on the benefits and risks of the exposures under investigation. Therefore any use of information from these studies should carry over this warning and be used accordingly.

V 2.2

DOCUMENT HISTORY

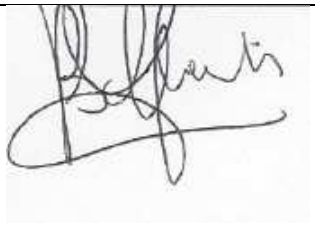
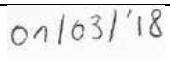
NAME	DATE	VERSION	DESCRIPTION
Vincent Bauchau, Kaat Bollaerts, Miriam Sturkenboom	7-01-2017		This document is based on the POC1.2 outline, final version
Kaatje Bollaerts, Vincent Bauchau, Lina Titievsky, Miriam Sturkenboom	4-08-2017		The POC outline was further worked out in the POC1.2 synopsis, final version
Kaatje Bollaerts	21-09-2017	0.1	First draft current document
Vincent Bauchau, Lina Titievsky, Daniel Weibel, Hanne-Dorthe Emborg	23-10-2017		Comments
Miriam Sturkenboom	30-10-2017		Comments
Kaatje Bollaerts	02-11-2017	0.2	Second draft
Hanne-Dorthe Emborg	8-11-2017		Comments
Miriam Sturkenboom	30-11-2017		Comments, Databases added
Kaatje Bollaerts	6-12-2017	0.3	Addressing comments
TinTin Htar, Vincent Bauchau, Hanne-Dorthe Emborg, Daniel Weibel	8-12-2017		Review, providing requested input
Kaatje Bollaerts	13-12-2017	1.0	Finalization
Patrick Mahy, Xavier Kurz, Jorgen Bauwens, Lieke van der Aa, Lina Titievsky, Linda Levesque, Sonja Banga	5-01-2018		SC review
Rosa Gini, Miriam Sturkenboom, Vincent Bauchau, Lina Titievsky, Kaatje Bollaerts	16-1-2018	2.0	Specifications success measures and ENCePP compliance
Kaatje Bollaerts	21-02-2018	2.1	Minor modifications after institutional review
Miriam Sturkenboom, Kaat Bollaerts	18-03-2018	2.2	Amendment to reflect changes in data management

SYNOPSIS

Title	Testing a system for near real-time monitoring on vaccination coverage, benefits and risks in Europe with acellular pertussis-containing vaccines as a test case.
Medicinal product	Any acellular pertussis-containing vaccine that was available during the study period and used in the study population in the 4 countries participating to the current study (Italy, UK, Spain and Denmark).
Product reference	Acellular pertussis-containing vaccines.
Research question and objectives	The overall objective of the ADVANCE Proof of Concept (POC) studies is to build and test a system (including testing data availability) for benefit-risk monitoring of vaccines in Europe. Specifically, the objective of POC1.2 is to establish the feasibility of continuously and rapidly updating the information on coverage, benefits and risks using electronic healthcare and surveillance databases, while visualizing these data using a dashboard. The overall objective is: Building the capacity for near real-time monitoring in Europe. The specific objectives are to: 1. Evaluate the ability of databases to generate periodic data with limited latency to secure near real-time monitoring on a preferably weekly basis (or bi-weekly/monthly if weekly would not be feasible) 2. Visually monitor vaccination coverage, benefits and risks using an interactive dashboard
Data sources	The potential data sources will include, Asl Cremona (ASLCR), Regional Database Tuscany (ARS) and PEDIANET from Italy; SIDIAP from Spain; RCGP from UK and SSI/AUH from Denmark.
Investigators and main authors	**POC1.2 study team**: Principal investigators (PIs) of POC1 will also act as PIs of POC1.2: Kaatje Bollaerts (benefit-risk), Daniel Weibel (Risk), Hanne-Dorthe Emborg (Coverage), Tin Tin Htar Myint (Benefit) Tom De Smedt (study statistician), Olivia Mahaux (study statistician), Lieke van der Aa (study manager), Vincent Bauchau (expert pharmacoepidemiologist), Lina Titievsky (WP5 lead, epidemiologist), Miriam Sturkenboom (WP5 lead, pharmaco-epidemiologist).
Disclaimer	The study described in this protocol is conducted as part of the IMI ADVANCE project with the aim to test methodological aspects of the

	design, conduct and reporting of studies for vaccine benefit-risk monitoring activities. The protocol presented herein relates solely to the testing of these methodologies and is not intended to inform regulatory or clinical decisions on the benefits and risks of the exposures under investigation. Therefore any use of information from these studies should carry over this warning and be used accordingly.

SIGNATURE PAGE

Study protocol version	Read and approved by (name)	Role	Signature	Date
V2.1	Kaatje Bollaerts	Principal investigator		
V2.1	Hanne-Dorthe Emborg	Principal investigator		
V2.1	Myint Tin Tin Htar	Principal investigator		13 March 2018
V2.1	Daniel Weibel	Principal investigator		1 March 2018
V2.2	Kaatje Bollaerts	Principal investigator		24/09/2018
V2.2	Hanne-Dorthe Emborg	Principal investigator		

V2.2	Myint Tin Tin Htar	Principal investigator		26 July 2018
V2.2	Daniel Weibel	Principal investigator		31 July 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADVANCE	Accelerated Development of VAccine beNefit-risk Collaboration in Europe
aP	acellular (pertussis vaccine)
B/R	Benefit/ Risk
CDM	common data model
DLP	data lock point
EU	European Union
GPP	good participatory practice
GVP	good pharmacovigilance practices
HHE	hypotonic-hyporesponsive episode
ICD	International Classification of Diseases
ICSR	individual case safety reports
IEC	independent ethics committee
IMI	Innovative Medicines Initiative
IRB	institutional review board
MAH	marketing authorisation holder
PAS	post-authorization study
PASS	post-authorization safety study
POC	proof-of-concept
REB	research ethics board
SAP	statistical analysis plan
SC	steering committee
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology
wP	whole-cell pertussis (vaccine)
WP	work package (i.e. WP4 or WP5)

1. RESPONSIBLE PARTIES

1.1. Main Author(s) of the Protocol

Name	***Institution***	***Role***	***Contributions***
Kaatje Bollaerts	P95	Principal investigator BR-study	First draft v0.1 and updates until last version
Hanne-Dorthe Emborg Daniel Weibel Tin Tin Htar	SSI EMC Pfizer	Principal investigators coverage, risk, benefit	Comments, adding exposure and outcome related information
Vincent Bauchau	GSK	Additional member, expert role	Review, methodological design
Lieke van der Aa	WIV-ISP	Study manager	Review
Miriam Sturkenboom Lina Titievsky	P95 Pfizer	WP5 co-leads	Review, methodological edits, design, coordination and adding data source descriptions

1.2. Principal Investigators

This study is a continuation of the first proof of concept (POC) study on pertussis vaccination for which different principal investigators led the study teams on the coverage, benefit and risk studies and evidence synthesis study (benefit-risk). For the current POC study (POC1.2), the same processes as in the POC 1 will be followed, however, this time we will do near real-time monitoring, whereas POC1 was a retrospective cohort study. As such, the same PIs will work together on the POC1.2 coordination team.

Benefit-Risk:
Kaatje Bollaerts, P95 Pharmacovigilance and Epidemiology Services, Leuven, Belgium
Tel: + 32 485 789657
Email: kaatje.bollaerts@p-95.com

Coverage:
Hanne-Dorthe Emborg, Statens Serum Institut (SSI), Copenhagen, Denmark
Email: HDE@ssi.dk

Benefit
Tin Tin Htar Myint, Clinical Epidemiology department, Pfizer
Email: myint.tintinhtar@pfizer.com

Risk:
Daniel Weibel, Erasmus Medical Center (EMC), Rotterdam, The Netherlands
Email: d.weibel@erasmusmc.nl

1.3. Study Team

PRINCIPAL INVESTIGATORS

- Kaatje Bollaerts (benefit-risk), P95
- Hanne-Dorthe Emborg (coverage study), SSI
- Tin Tin Htar Myint (benefit study), Pfizer
- Daniel Weibel (risk study), EMC

STATISTICIANS/DATA SCIENTISTS

- Tom De Smedt, P95
- Olivia Mahaux, GSK
- Kaatje Bollaerts, P95
- Marco Villa, ASLCR
- Rosa Gini, ARS
- Johnny Kahlert, AUH
- Claudia Bartolini, ARS

- Chris McGee, RCGP

ADDITIONAL STUDY MEMBERS, EPIDEMIOLOGISTS

- Vincent Bauchau, GSK
- Elisa Martin, BIFAP
- Consuelo Huerta, BIFAP
- Gino Picelli, PEDIANET
- Lara Tramontan, PEDIANET
- Giorgia Danieli, PEDIANET
- Talita Duarte Salles, SIDIAP
- Giuseppe Roberto, ARS

STUDY MANAGER

- Lieke van der Aa, WIV-ISP

WP5 CO-LEADS

- Miriam Sturkenboom, P95
- Lina Titievsky, Pfizer

DATABASE LIAISONS/CUSTODIANS

Databases will participate upon demonstration of data quality in the quality assessment module (see also Appendix 1) and when having the ability to generate periodic data, with a target of weekly refresh of data less than a week old (bi-weekly/monthly will also be accepted). Previous quality assessments conducted as part of POC1 remain valid. Custodians of contributing databases will be members of the study team if they meet the requirements and conditions from the ADVANCE Code-of-Conduct.

2. ABSTRACT

Date of protocol abstract: December 11, 2017

Title:

Testing a system for near real-time monitoring of vaccination coverage, benefits and risks in Europe with pertussis vaccines as a test case.

Rationale and Background:

POC1.2 aims to test near real-time and visual monitoring of coverage, benefits and risks of acellular pertussis (aP) vaccination. POC1.2 is designed as a continuation of POC1 to leverage previously completed work. POC1.2 will use near real-time visual monitoring of aP vaccines in Europe as a test case, aiming to mimic the introduction of a new vaccine. The interactive dashboard developed as an additional analysis to POC1 will be regularly updated for monitoring.

Research Question and Objectives:

The overall objective of the ADVANCE POC studies is to build and test a system (including data availability) for the benefit-risk monitoring of vaccines in Europe. The objective of POC1.2 is to determine the feasibility of periodic and rapid assessments of vaccine coverage, benefits and risks using electronic healthcare databases, while displaying these data using a dashboard.

The specific objectives are to:

1. Evaluate the ability of databases to generate periodic data with limited latency to secure near real-time monitoring on a preferably weekly basis (or bi-weekly/monthly if weekly would not be feasible)
2. Visually monitor vaccination coverage, benefits and risks using an interactive dashboard

Target Population:

All children from their start of follow-up in the database until school-entry pertussis booster, 6 years of age or any periodic data lock point within the eligible ADVANCE databases.

Observation Period:

The total observation period consists of two distinct periods. The first period starts from 01 January 2014 until the start of the near real-time monitoring upon approval of the protocol and has as objective to establish a baseline. The start of the first period (01 January 2014) was chosen to balance data accuracy and computational burden. The second period has the objective of near real-time monitoring and will cover a few months. This period will start upon protocol approval until last periodic data lock point (DLP) (i.e. the time the last periodic database extract is produced) and will cover a few months.

Study design:

A dynamic cohort study

Exposure

The exposure of interest is vaccination with any acellular pertussis-containing (aP) vaccine recorded in the study population covered by the ADVANCE databases participating in POC1.2

Outcomes

The risk outcomes are hypotonic hypo-responsive episodes (HHE), febrile convulsions/seizures, fever, somnolence and persistent crying. The benefit outcome is the vaccine preventable disease (here: pertussis, confirmed or probable).

Data analyses

The test case is near real-time (weekly or bi-weekly/monthly, depending on the database) visual monitoring of vaccination coverage, benefits and risks of pertussis vaccination. The monitoring will be facilitated through the use of an interactive dashboard developed based on the POC1 data. The dashboard contains three monitoring tabs with visualizations:

- Coverage: number of administered doses per week over calendar time (extrapolated to the population of interest) and vaccination coverage (%) within specific age groups by calendar time
- Benefits: observed pertussis incidence in the total population by calendar time
- Risks: incidence rates in event specific risk windows and in control periods (out of risk windows) at vaccination eligible ages, separately for each risk outcome by dose, estimated cumulatively over calendar time.

Data Sources:

The potential data sources were selected from all databases available from ADVANCE partners and associated partners, based on the following eligibility criteria:

- Successful database quality assessment based on database fingerprinting and benchmarking for the events and vaccine of interest
- Theoretical ability to generate periodic data, with a target of weekly refresh of data, as recovered based on a detailed survey

Based on these two criteria the following six data sources remained: ASLCR, ARS, PEDIANET, RCGP, SIDIAP and SSI/AUH.

Sample size:

Entire eligible study population from the six eligible databases.

Informed Consent and Ethical Approval:

This analysis will be conducted on the basis of secondary use of data from ADVANCE partners and associate partners. Each contributing database will follow local governance and privacy rules prior to sharing anonymized data.

Disclaimer

The study described in this protocol is conducted as part of the IMI ADVANCE project with the aim to test methodological aspects of the design, conduct and reporting of studies for vaccine benefit-risk monitoring activities. The protocol presented herein relates solely to the testing of these methodologies and is not intended to inform regulatory or clinical decisions on the benefits and risks of the exposures under investigation. Therefore any use of information from these studies should carry over this warning and be used accordingly.

3. AMENDMENTS AND UPDATES

Protocol amendments following IRB approval of the local databases:

Table 1 Overview of Protocol Amendments and Updates

	Date (DDMMMYY)	***Section of the study protocol***	***Amendment or update***	***Reason***
1	**16/03/2018**	**Section 7.2.9.3**	**Aggregated data will not be sent to RRE anymore for post-processing, but transformed to its most aggregate level locally and then immediately transferred to the P-95 dashboard server. This step now avoids use of the RREm with full backend (local) data transformation. Figure 2 shows the process.**	**Due to a recent move of office, P95 lost its fixed IP-address, which is needed to access the RRE. Getting a fixed IP-address from the internet provider takes much longer than anticipated. Avoiding the RRE will reduce delays**
2				
….				

4. MILESTONES

Table 2: Overview of Study Milestones

Activity	*Due date*
Submission protocol to ADVANCE steering committee (SC)	Dec 15, 2017
Comments from SC	January 5, 2017
Submission to Ethics Committee/Institutional Review Board	January 16, 2017
Protocol approval Ethics Committee/Institutional Review Board	February 28, 2018
Statistical Analysis Plan (SAP)	January 24, 2018
Review SAP	January 31, 2018
Final SAP	February 5, 2018
Near real-time monitoring (3 months)	March – April, 2018
Draft report version 1	May 18, 2018
Study team review report version 1	May 25, 2018
Report version 2	May 30, 2018
SC and ADVANCE consortium review report version 2	June 15, 2018
Rebuttal and final report submitted to IMI	June 30, 2018

5. RATIONALE AND BACKGROUND

At the core of the mission of ADVANCE and many of its stakeholders is the concept of vaccine Benefit/Risk (B/R) monitoring. Monitoring should be understood as a periodic assessment of several key parameters including, coverage, incidence of adverse events, and incidence of the vaccine preventable disease to trigger an alert if and when there is an indication that the B/R profile in the population is different from what is expected (based on clinical trials, observational studies or similar products). This alert would generate a subsequent and possibly more formal assessment and analysis of the vaccine. Monitoring should, in principle, start as soon as a new vaccine is introduced in a given country and continue throughout the vaccine's lifecycle, and would also be applied to vaccines with established B/R profiles.

B/R monitoring requires information that is available in a timely fashion. Hence, the goal is to have access to near real-time information, which is defined as either weekly or monthly refresh of data that is only a few days old. Weekly periodicity is preferred and has been previously successfully implemented for influenza vaccine monitoring (H1N1 pandemic). It is also the approach used for the VSD rapid cycle analyses and used in some published studies [1, 2]. Weekly monitoring has also been recommended in a recent EMA guidance for seasonal influenza vaccine safety monitoring.

Visual monitoring of the component parameters of B/R as well as two composite B/R measures has been recently prototyped in a dashboard using simulated data on rotavirus (http://apps.p-95.com/BRMonitor/). This approach was found useful and informative. Subsequently, the dashboard was further developed to incorporate the retrospective real world data on pertussis from the first proof of concept study (POC1), demonstrating good acceptability.

POC1 is the first ADVANCE POC study. The overall objective of POC1 is to develop and test the ability of the currently available ADVANCE system to perform post-marketing studies of the benefits and risks of vaccination in Europe. Pertussis vaccination, notably comparing the benefits and risks of whole-cell pertussis (wP) vaccinations with aP vaccinations, was chosen as the first test case. The POC1 study is completed and is extensively described in ADVANCE deliverable D5.6. (Accessible at the ADVANCE Members area: http://www.advance-vaccines.eu/).

The current POC, POC1.2, will be a real-world test case of near real-time visual monitoring of coverage, benefits and risks using electronic healthcare data. POC1.2 will be designed as a continuation of POC1 to maximally leverage previous work. The dashboard developed as additional analyses to POC1 and described above will be used for monitoring.

The expected learnings from POC1.2 include:

1. Identifying databases with proven fast and frequent data refresh that can be used for near real time monitoring of vaccination coverage, benefits and risks rates (on top of the theoretical ability as identified based on a survey), and understanding the reasons, if any, why some databases cannot be used for monitoring.
2. Testing the requirements for data flow and data processing to ensure near real-time monitoring;
3. Using a dashboard for near real-time monitoring.

6. RESEARCH QUESTION AND OBJECTIVES

The overall objective of POC1.2 is to determine the feasibility of periodic and rapid assessments of vaccine coverage, benefits and risks using electronic healthcare databases.

The specific objectives include:

1. Evaluate the ability of databases to generate periodic data with limited latency to secure near real-time monitoring on a preferably weekly basis (or bi-weekly/monthly if weekly would not be feasible)
2. Visually monitor vaccination coverage, benefits and risks using an interactive dashboard

The test case will be the monitoring of acellular pertussis vaccination in Europe, mimicking the launch of a new vaccine.

7. RESEARCH METHODS

7.1. System testing: measures of success

The following criteria will guide the measurement of the level of success of POC1.2:

- 1. Feasibility: at least 3 databases (i.e. 50% of the pre-identified databases) can provide data on coverage, benefits and risks for near real-time monitoring (meeting criteria 2 and 3 below) to demonstrate the feasibility of vaccine benefit-risk monitoring in Europe.
- 2. Frequent refresh rates: at least 3 databases can provide weekly (most successful), bi-weekly (highly successful) or monthly (successful) data on coverage, benefits and risks.
- 3. Timeliness: at least 3 databases can provide data of 1 week (most successful), 2 weeks (highly successful), 1 month (moderately successful) or 2 months (minimally successful) old. The timeliness of the monitoring will be based on the distribution of the 'reporting delays': time between event date

and dashboard system date (i.e. date when new events that happen during the monitoring phase are displayed for the first time on the dashboard).

- 4. Acceptability of the approach by all ADVANCE stakeholders: the link to the monitoring dashboard will be shared with all ADVANCE stakeholders. A semi-structured survey and discussions will be organised to collect feedback. The ADVANCE consortium and the scientific advisory board (see section 7.2.10.2) will discuss and judge the successfulness of the POC1.2 study, discussing strengths, limitations and areas for further improvement.

7.2. Illustration: pertussis dashboard

Several sections below are taken directly from the POC1 protocol because POC1.2 will follow similar processes as POC1. To create a stand-alone document, but to avoid repetition for those familiar with POC1, we indicated the sections from POC1 with a *. Readers familiar with POC1 might skip these sections.

7.2.1. Design

This will be a dynamic cohort study. In a dynamic cohort subjects possibly enter or leave the cohort at any time during the study period.

7.2.2. Setting

This POC 1.2 study will be conducted in six population-based healthcare databases from four European countries.

7.2.3. Databases/Data Sources

The eligible data sources will be the eight European electronic health care databases from ADVANCE partners and associated partners that passed the POC1 feasibility assessment (see Appendix 1). In addition, these databases need to have the theoretical ability to generate periodic data. This information was collected through a face to face discussion and survey. This additional criterion reduced the number of databases that will participate to POC1.2 to six; ASLCR (Italy), ARS (Italy), PEDIANET (Italy), RCGP (UK), SIDIAP (Spain), SSI/AUH (Denmark). A description of the individual POC1.2 databases is given below.

Italy: ASL della provinca di Cremona (ASLCR)*

ASCLR is a record linkage database. It contains all the mortality data (with cause of death), hospitalizations (with diagnosis), outpatient visits, drug prescriptions of the citizens. Moreover, it contains the registry of all the vaccinations administered by (or notified to) the Local Health Authority and the registry of infectious diseases to be notified by law.

The Local Health Authority (ASL) of Cremona is the institution in charge of the health of the citizens living in the province of Cremona. The ASL is responsible for the provision of all health-related services (prevention, treatment, residential care, etc.).

Italy: Agenzia Regionale della Sanita Tuscany (ARS)*

The ARS database contains health care data from the Tuscany region of Italy, including pharmacy, outpatient and hospital data along with linked data on death, and birth and malformation registries. Since 2017 the data is also linked to vaccination registries. The database is described in the ENCePP registry of datasources http://www.encepp.eu/encepp/viewResource.htm?id=8133

United Kingdom RCGP*

The RCGP's programme of influenza and respiratory disease surveillance has been supported by the Department of Health (DH) since 1967.

Between 1967 and 1993 the registered population within the RCGP RSC grew to 200,000 and it comprised 40 practices. By its 30th anniversary the population had risen to 570,000, with electronic links to general practices and laboratory data. The network exceeded 1million in 2015, and has now expanded to 237 general care practices and 2.4 million. From the start the network wanted clinicians to record a diagnosis in each consultation based on their clinical judgment; and only change that diagnosis if it adds to patient care. Diagnoses were originally recorded on paper written with "f", "n", or "o" written after– to designate first, new or an ongoing problem. This enabled incident and prevalent problems to be differentiated in routine data, vital for surveillance, but also for the Fourth National Morbidity Survey (MSGP4).[i] High quality data recording was previously encouraged by visits to practices and quarterly feedback; now we apply principals of audit-based education to give practices personalized feedback on their data quality; we also provide use of on-line learning.

Denmark: SSI/AUH*

In Denmark national registry data are assembled and linked for specific studies according to the scientific question. Overall, the Danish Civil Registration System (CRS) database comprises individual level information on birth date, gender, vital status and sequential dates of migration of all residents in Denmark, i.e. approx. 5.6 mill. persons in a year, and an accumulated number of approx. 7.5 mill. persons over the period 1996-2014. Hence, using a unique personal identifier assigned to all Danish residents upon birth or immigration makes it possible to link the CRS with health care databases such as the Danish Vaccination Registry (DVR) and the Danish National Patient Registry (DNPR).

The national childhood vaccination database was established in 2000. This comprises information on all vaccinations administered to children below the age of 18 years, including vaccinations recommended by the Danish childhood vaccination programme, such as HPV, influenza etc. From November 2015, GPs and

others authorized to vaccinate have an obligation to record all vaccinations, and this information is now merged with data that were compiled before 2015 into the Danish Vaccination Registry.

The Danish National Patient Registry has recorded every inpatient hospitalization since 1977, and every outpatient and emergency room visit since 1995. For each admission, one primary and potentially several secondary diagnoses are registered, classified according to the *International Classification of Diseases, Eighth Revision* (ICD-8) until the end of 1993, and *Tenth Revision* (ICD-10) thereafter. In Advance this database is used to identify events.

While CRS and DVR are updated on a daily basis, DNPR have a delay of up to 3 months. The data are kept at the Research Services at the SSI from where access can be granted provided that the analysis has been reported and approved by the Danish Data Protection Agency. There is a long tradition of epidemiological studies in Denmark based on the health care databases that were proven valid for research.

Spain: SIDIAP*

The Information System for Research in Primary Care (SIDIAP; www.sidiap.org) contains information recorded in anonymized patients' electronic health records for nearly six million people (approximately 75% of the Catalan population) since 2005 by 286 primary care practices of the Catalan Health Institute throughout Catalonia. It was designed to provide a valid and reliable database of selected information from clinical records of patients registered in primary care centres for use in biomedical research. SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions. It includes data collected by health professionals in Catalonia during routine visits in primary care, including clinical diagnoses (International Classification of Diseases 10th revision [ICD-10]), laboratory test results, prescribed and dispensed drugs, vaccination, hospital referrals, anthropometric and other measurements performed during routine visits, demographic and lifestyle information. Quality checks to identify duplicate patient IDs are performed centrally at each SIDIAP data base update, which is done annually. Checks for logical values and data harmonisation are performed. For biochemistry data, consistency for measurements taken in different laboratories is assessed, and unit conversion is undertaken when needed. The high quality of these data has been previously documented[1], and SIDIAP has been successfully applied to epidemiological studies of key exposures and outcomes [2]. SIDIAP is listed in the ENCePP registry of datasources: http://www.encepp.eu/encepp/viewResource.htm?id=4646

[1] Bolíbar, B. et al. [SIDIAP database: electronic clinical records in primary care as a source of information for epidemiologic research]. Med. Clínica 138, 617–621 (2012).
García-Gil, M. D. M. et al. Construction and validation of a scoring system for the selection of high-quality data in a Spanish population primary care database (SIDIAP). Inform. Prim. Care 19, 135–145 (2011).
Ramos, R. et al. Validity for use in research on vascular diseases of the SIDIAP (Information System for the Development of Research in Primary Care): the EMMA study. Rev. Esp. Cardiol. Engl. Ed 65, 29–37 (2012).

[2] Premaor, M. O. et al. The association between fracture site and obesity in men: a population-based cohort study. J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res. 28, 1771–1777 (2013).

Italy: Pedianet*

Pedianet (www.pedianet.it) is a longitudinal observational DB that collects epidemiological clinical data for clinical research from family paediatricians involved in the Pedianet network in Italy. This system is based on the transmission of specific data from computerised clinical files, which the paediatricians in the network fill out during their daily professional activities. Informed consent is required from the parents. Such data is collected anonymously by a central server in Padua, where it is validated and elaborated.

Pedianet is an independent network. The coordination of the projects and data analysis is carried out by a scientific committee that include internationally renowned paediatricians, epidemiologists and researchers. Approximately 400 paediatricians throughout the country have taken part in Pedianet projects.

The paediatric population involves infants, toddlers, children and adolescents up to 14 years. The DB contains information on about 400 000 patients. This is the cumulative amount of patients who have ever been part of the dynamic cohort of patients who have been registered.

Pedianet includes primary care diagnoses, outpatient specialist diagnoses, hospital discharge diagnoses, prescribed drugs, vaccinations and laboratory test results.

In order to provide more comprehensive and validated data, linkage with external Regional databases, including hospitalizations, vaccinations and specialist visits/exams, is ongoing. PEDIANET is involved in several EU funded projects including TEDDY, GRIP, EU-ADR, EMIF and SOS. For the ADVANCE project, PEDIANET has linked the birth cohorts of 2006 and 2007 to the Regional Vaccination Register of Veneto on 9000 children. Linkage of susbequent birth cohorts is ongoing once consent is obtained from the subjects. Pedianet is listed in the ENCePP registry of datasources (http://www.encepp.eu/encepp/viewResource.htm?id=20131)

7.2.4. Study Population

All children from start of follow-up in their database until end of follow-up within the eligible ADVANCE databases. Start of follow-up will be the latest of: start of the study (i.e. 01 January 2014), first date of registration in the database, date of birth. End of follow-up will be the earliest of: death, end of registration in the database or moving out, receipt of pertussis booster dose, reaching 6 years of age, or data lock point.

Prieto-Alhambra, D. et al. Relationship between mortality and BMI after fracture: a population-based study of men and women aged ≥40 years. J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res. 29, 1737–1744 (2014).
Garcia-Gil, M. et al. Linking of primary care records to census data to study the association between socioeconomic status and cancer incidence in Southern Europe: a nation-wide ecological study. PloS One 9, e109706 (2014).
sIncidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. Ann. Rheum. Dis. 73, 1659–1664 (2014).

Children will be excluded if at least one of the following variables is missing: date of birth, date of start of follow-up (date the patient is entered into the registration system) and date of end of follow-up (e.g. end of registration, moving out, death, last data draw down - whichever is earliest).

7.2.5. Study Period

The study period will consist of two distinct periods. The first period will start from 01 January 2014 until the start of the near real-time monitoring and has as objective to establish a baseline. The second period is the actual near real-time monitoring which will start upon approval of the protocol by the institutional review board/independent ethics committee/research ethics board (IRB/IEC/REB). The period of near real-time monitoring will cover a minimum of 3 data loads and be maximum 3 months long.

It should be noted that the start of follow-up for the POC1.2 study period is arbitrary and was selected to ensure an observation period long enough to obtain stable rates while also minimizing the computational burden. This arbitrary date does not correspond in any way to a date where there would be a need or a trigger to start monitoring (as in normal use of a monitoring system). The same principles apply to the date that will be used for the start of the short period of near real time monitoring.

7.2.6. Variables

7.2.6.1. Common Data Model (CDM)*

Each study database will create three input files in the pre-defined study-specific ADVANCE common data model (CDM): a population, event and vaccination input file (see Appendix 2). All three data files will use the same unique, encrypted person identifier to allow linkage across the database-specific CDM input files.

7.2.6.2. Exposure of Interest, Operationalization and Validation

EXPOSURE OF INTEREST

The exposure of interest is any acellular pertussis-containing (aP) vaccine used by the study population covered by the ADVANCE databases participating in POC1.2.

OPERATIONALIZATION

Vaccinations will be obtained from the databases using vaccine names and database specific codes (Appendix 3). Only data on aP vaccines will be retrieved, and there will be no stratification by vaccine brand or product name.

VALIDATION

Validation will be done according to fingerprinting methodology developed in ADVANCE (ADVANCE sharepoint: WP5 deliverables D5.2) The vaccine fingerprinting process (characterizing the data in a systematic manner) comprises the following components: 1) assessment of rounding of birth dates, 2) assessment of the concordance between doses and recommended country specific schedules; 3) Assessment of coverage against country specific benchmarks (WHO or national data).

7.2.6.3. Outcomes, Operationalization and Validation

OUTCOMES

The risk outcomes of interest will be the same as in POC 1. However, injection site reactions and generalized convulsions were excluded. The exclusion of injection site reactions is based on the experience of POC 1, which demonstrated that many participating databases were able to extract general skin and muscle related events, signs and symptoms that potentially were not occurring at an injection site. The absence of "injection site" related information in some databases resulted therefore in a broad spectrum of extracted skin and muscle events, signs and symptoms that have been considered to be too sensitive and not specific enough to be further included as potential adverse events of interest in the POC 1.2. Generalized convulsions have been excluded because the experience from POC 1 showed that febrile convulsions/seizures dominated the entity of the convulsions/seizures. As most of the databases provided highly specific codes for the identification of febrile convulsions/seizures, we decided to continue with febrile convulsions/seizures only. The outcomes of interest (and associated risk windows) are listed in the Table 1. The benefit outcome is confirmed or probable pertussis.

Table 3: Outcomes of interest and risk windows

	Outcome	*Risk Window After Vaccination/*start of protection*
Risks		
	Fever	0-72 hours
	Somnolence	0-48 hours
	Persistent crying, irritability	0-24 hours
	Febrile convulsions/seizures	0-72 hours
	HHE	0-48 hours
Benefits		
	Pertussis	*2 wks after vaccination

OPERATIONALIZATION*

Outcome events will be obtained from the databases using diagnosis codes or free-text proxies (code lists as in POC1 listed in Appendix 4).

VALIDATION

Semantic harmonization of events into different coding dictionaries was conducted using the ADVANCE Codemapper [3]. All events are benchmarked by assessing age-specific incidence rates and comparison of these rates between the databases in the consortium, as well as against published data. For the POC1.2 eligible databases, this process was conducted successfully as part of their quality assessment.

7.2.6.4. Other Variables and Operationalizations

The incidence rates of the benefit and risk outcomes will be stratified by calendar time and country. The unit of stratification by calendar time will be weeks to allow for timely monitoring.

7.2.7. Data Analysis

7.2.7.1. Demographics

To better understand the potential differences in results between the databases, the database-specific study populations will be described using frequency tables and proportions by age group, sex, calendar time. Attrition diagrams summarizing the deletion of records and reasons for deletion will be created.

7.2.7.2. Reporting delays

Timeliness is essential for monitoring. Therefore, we will study delays in reporting by comparing event dates with the date the event is first used in the dashboard for monitoring. The analysis will be restricted to events for which it is known when they were first used (i.e. events that are present in the updated dashboard at time *I*, but were not in the previous update at time *I-1*). The distribution of reporting delays will be obtained by event type or group of events (events recorded at hospital vs. primary care; exposure events, health outcome events) and will be summarized using measures of central tendency (average or median, whichever is most appropriate) and measures of spread (standard deviation or inter-percentile ranges, whichever is most appropriate).

As the total reporting delay is an accumulation of time lost due the processes at the level of the healthcare system (e.g. he time it takes for hospitals to report back to the GP - healthcare level) and processes at the level of the databases (database level) we will also look at the time between the most recent end of follow-up dates in the population CDM files and the use of this data in the dashboard. This will inform us regarding the amount of delays at the database level.

This way, we will obtain the total reporting delays as well as the delays at database level, allowing us to derive the delays at healthcare level.

7.2.7.3. Statistical Hypothesis

The objective of the current study is to test a system for near real-time monitoring of vaccination coverage, benefits and risks. As such, no formal statistical hypothesis testing will be done.

7.2.7.4. Sample size considerations

We derived the margin of error at the 5% significance level (or the half width of the 95% confidence intervals) based on the Poisson distribution relative to the assumed incidence, expressed in %. We assumed incidence rates of 2/10.000py, 50/10.000py and 100/10.000py (which correspond to the range of incidence estimates obtained in POC1). The margin of error is presented as a function of person-time with person-time varying from 100,000 to 800,000 person-years (Figure 1). For reference, the expected annual number of person-years for the population of interest is +/- 37,000 in RCGP (one of the smaller POC1 databases) and is +/- 220,000 for SSI (one of the larger POC1 databases).

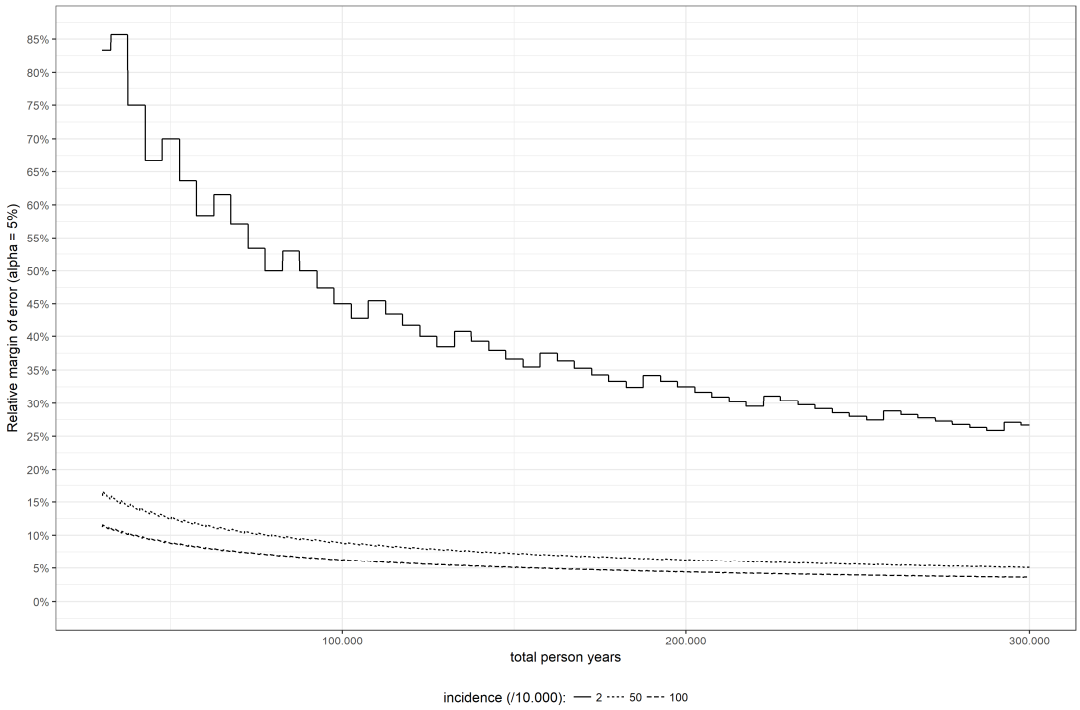


Figure 1: Relative margin of error (5% significance level) assuming incidence rates of 2/10.000py, 50/10.000py and 100/10.000py by database size (person yrs)

7.2.7.5. Statistical Methods

In POC1.2, the objective is near real-time visual monitoring of vaccination coverage, benefits and risks using aP vaccination as the test case. As such, we will monitor epidemiological measures including, vaccination coverage and incidence rates as they evolve over time. The monitoring will be facilitated through the use of a "Shiny" web-application [4] with a user-friendly interactive dashboard containing three monitoring tabs for visualization of coverage, benefits and risks. This dashboard will be interactive and allow for inter-country comparisons.

Coverage: bar charts with weekly number of administered doses extrapolated to the whole population

The total number of doses n_{ij} (for dose 1, dose 2 and dose 3) given during week i in age group j will be calculated from each database. Then, the total number of doses will be extrapolated to the whole population as follows:

$$Ntot_{ij} = (N_{ij}/pop_j)^{-1} n_{ij} = w_{ij}\, n_{ij},$$

where N_{ij} is the number of active subjects in the database at week i and age group j and where pop_j is the number of subjects in the total population of same age group j. The number of subjects in the total population will be obtained from the National Offices of Statistics.

Coverage: line plots with vaccination coverage (%) by year-month birth cohorts over time

For every birth-month cohort, we will evaluate the vaccination coverage by dose over time (i.e. at every start of the week/month depending on the refresh frequency of the database). The coverage at week i for birth cohort j will be calculated by dividing the number of vaccinated subjects n_{ij} by the total number of subjects still under follow-up at week i (N_{ij}), expressed as a percentage.

Benefits: line plots with pertussis incidence and 95% confidence intervals over time

The pertussis incidence (/100.000 person-years) will be calculated as the number of pertussis events divided by the total person-time at risk for week i multiplied with 100.000. Exact Poisson 95% confidence intervals will be calculated.

Risks: line plots with incidence of risk outcomes inside and outside the risk window and 95% confidence intervals over time using accrued data

The incidence rate (per unit person- years) for the risk outcomes within the pre-defined, outcome-specific risk windows will be estimated by week. In addition, the incidence rates will be estimated cumulatively over time, using all data accrued from the start of the study period until week i or

$$inc_I = \left({\sum_{i=1}^{I} n_i} \Big/ {\sum_{i=1}^{I} py_i} \right) \times unit\ person\ years$$

where n_i is the number risk events of interest that happened during week i and where $py_i$is the amount of person-time (in years) within that week. Incidence rates outside the risk windows in vaccination comparable age groups will be calculated as well. Each time, exact Poisson 95% confidence intervals will be calculated.

7.2.8. Limitations

There are several limitations to this POC study;

- POC1.2 is affected by the lack of chart validation of the events.
- Results might differ between databases because of differences introduced by the provenance of the data (hospital discharge *vs.* primary care).
- POC1.2 is perceived as a first step towards near real-time monitoring of vaccines in Europe, using simple visualizations only. More advanced monitoring methods (such as CUSUM or max-SPRT [5]) are not considered for reasons of simplicity.

- The duration of the actual near real-time monitoring will be a maximum of 3 months aiming for at least 3 data refreshes to meet report deliverables by assigned deadlines. However, this might be too short for comprehensive system testing.

7.2.9. Data Management

Processing of data from the different databases will be according to ADVANCE princioples (ADVANCE sharepoint: WP5 deliverables D5.1 and D5.2[3]), with the exception that an additional data processing and storage step is foreseen to operationalize the web-application with the interactive dashboard. A graphical representation is given in Figure 1 and the different data management steps are explained below.

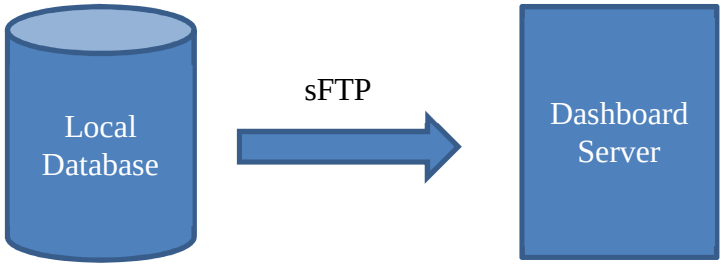


Figure 2: POC1.2 data management flow

7.2.9.1. Phase 1: Extraction & transformation of local data*

This entails the extraction of study specific individual-level data from the original databases into study specific common data model input files (CDM). This will be done locally by the database custodians. POC1.2 will use the same CDM input files as POC1, with the exception that the study period will be different and some events will be excluded (such as the pertussis complications). The CDM files will stay locally and will have the same structure across all databases.

The following three CDM input files will be created (see also Appendix 2):

1. ***Patients.txt:*** This input file will contain the information of all persons recorded in the study cohort, including start and end of follow-up, date of birth, sex and the unique, encrypted person identifier.

[3] https://publication.wiv-isp.be/workspaces/advance/wp5/Shared%20Documents/Forms/AllItems.aspx

2. ***Vaccinations.txt:*** This input file will contain information on acellular pertussis vaccinations including date of vaccine supply/prescription, type of vaccine (coded by ADVANCE dictionary), the recorded dose and/or derived dose and the unique, encrypted person identifier.
3. ***Events.txt:*** This input file contains information about the date and type of the events of interest (hypotonic hypo-responsive episodes (HHE), febrile convulsions (FCONVULS), fever (FEVER), injection site reactions (ISR), somnolence (SOMNOL) and persistent crying (PCRYING) and pertussis (PERT)) and the unique, encrypted person identifier.

7.2.9.2. Phase 2: Transformation of CDM data files into analytical datasets*

The study specific CDM data will be transformed and minimized into aggregated datasets suitable for further sharing and analysis. The analytical dataset will contain anonymized data only. The data transformation files will be programmed in R.

The POC1.2 programs for transforming the CDM data into the analytical datasets will be created by modifying the existing R data transformation programs created for the POC1 risk study. Cleaning of the vaccination files will follow the specifications reported in POC1 (ADVANCE sharepoint: WP5 D5.6[4]).

Once the codes have been tested, they will be shared with the local database teams, who will run these programs locally, and ensure all needed documentations (log files, recording of site-specific changes to the code, and all versions of the code if there are modifications) are saved and archived.

7.2.9.3. Phase 3: Web-application server

The data tables with summary statistics will be transferred by the data access providers to a "Shiny" server application on a password protected server controlled by P95 via secure file transfer protocol. On this server, a "Shiny" web application will transform the data tables – based on the user input- to the information to be plotted for the visual monitoring of vaccination coverage, benefits and risks. The additional data transformations on the Shiny server will allow the dashboard to be interactive.

The resulting visualizations will be accessible through a web page, with password protection and logging of user names.

[4] https://publication.wiv-isp.be/workspaces/advance/wp5/POC%20I%20Pertussis%20study%20working%20documents/Forms/AllItems.aspx

7.2.10. Quality Control

7.2.10.1. Record Retention*

Documents that individually and collectively permit evaluation of the conduct of a study and the quality of the data produced will be retained for a period of 5 years in accordance with Good Participatory Practice (GPP) guidelines. These documents could be retained for a longer period, however, if required by the applicable regulatory requirements or by an agreement between study partners. It will be the responsibility of the principal investigator to inform the other investigators/institutions as to when these documents no longer need to be retained.

Study records or documents may also include the analysis files, syntaxes (usually stored at the site of the database), and questionnaires.

7.2.10.2. Advisory Committee*

The members of the ADVANCE Scientific Advisory Board are Dr. Hector Izurieta (US Food and Drug Administration), Prof Martin Kuldorff (Harvard University Medical School and Brigham and Women's Hospital - Harvard Medical School), Dr. Frank de Stefano (Centers for Disease Control and Prevention), Prof. Barend Mons (Leiden University Medical Centre) and Dr. June Raine (Medicines and Healthcare products Regulatory Agency).

7.3. Use of the data generated in this study

The data generated in this study will be transferred to a web-based interactive dashboard, designed to facilitate the near real-time monitoring of vaccination coverage, benefits and risks. Data will only be used for the purposes of the POC1.2

8. PROTECTION OF HUMAN SUBJECTS

8.1. Regulatory and Ethical Compliance*

This protocol will be conducted following the ADVANCE code of conduct [6].

The European legislation describes the obligations to be fulfilled by marketing authorization holders (MAHs) and national competent authorities for medicinal products (including vaccines) authorized in the European

Union (EU). The European legislation does not apply to post-authorization studies conducted by organizations such as academia, medical research charities or research organizations in the public sector. These organizations should follow local requirements defined in the national legislation applicable in the countries where the research is conducted. In the context of the ADVANCE consortium, the proof-of-concept studies (POCs) will be conducted to test new approaches (data sources, methods) using a test case; within this framework, the POC studies will collect and use data on vaccines authorized in the EU. Vaccine MAHs are partners of the ADVANCE consortium and participate in the design, conduct and funding (through in-kind contribution) of the POCs concerning vaccines (at the product or substance level) for which they hold an authorization. They should therefore be considered as having input on the design of POCs, in which case requirements of the GVP applies. (21) The GVP requirements' will be addressed in the following ways:

1) The study protocol and study report will be posted on the EU PAS register.

2) The POC studies will be monitored by MAHs as PAS.

3) This proof-of-concept study aims to test components of the ADVANCE system for the benefit-risk monitoring of vaccines in Europe, more specifically:

- to evaluate participating databases on ability to provide data suited for near real-time monitoring of vaccination coverage, benefits and risks

- to test the methodology of near real-time monitoring of vaccination coverage, benefits and risks, facilitated through the development of a web-used interactive dashboard

The objectives of this study are on methodological aspects of the monitoring system and not intended to provide any information on the safety of the concerned pertussis-containing vaccines. Therefore, this study is not considered as a PASS.

4) Management and reporting of adverse events/adverse reactions:

This study is observational, based on secondary use of data in large healthcare databases and will provide only incidence rates of events by vaccines type (aP). Thus, the reporting of suspected adverse reactions in the form of individual case safety reports (ICSRs) is not required and no individual adverse events/reactions will be summarized in the final study report.

ADVANCE is following the ADVANCE CoC, which differs from the ENCePP Code of conduct regarding involvement of study team members who are employed by organizations with a financial interest in one of the outcomes of the study. This specific study is about system testing using pertussis vaccines as an

example and has personnel from vaccine manufacturers as study team members and is compliant with ADVANCE CoC. The coverage, risk and risk-benefit parts of this study are also fully compliant with the ENCePP CoC because all the PIof these parts are from ENCePP Inventory of Research Centres.

8.2. Informed Consent*

No informed consent is necessary as this is a retrospective study using de-identified/anonymized data for secondary purposes.

8.3. Responsibilities of the Investigator and IRB/IEC/REB*

The protocol and waiver of informed consent will be reviewed and approved by a properly constituted institutional review board/independent ethics committee/research ethics board (IRB/IEC/REB) before study start. A signed and dated statement that the protocol has been approved by the IRB/IEC/REB and waiver of informed consent will be given to the principal investigator before study initiation.

8.4. Protocol Adherence*

Investigators will apply due diligence to avoid protocol deviations. If the investigator feels a change to the protocol would improve the conduct of the study this will be considered a protocol amendment and, unless such an amendment is agreed upon by all partners involved and approved by the IRB/IEC/REB, it will not be implemented. All significant protocol deviations will be recorded and reported in the Study Report. Specifically, reportable Protocol Deviations are those which directly or indirectly have a significant impact on any 1 or more of the following:

1. Subject's rights, safety, or well-being
2. Data integrity, i.e. completeness, accuracy, and reliability of safety, efficacy, and immunogenicity outcomes of the clinical study, and
3. Regulatory compliance.

9. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ ADVERSE REACTIONS

N/A

10. PLANS FOR DISSEMINATING AND COMMUNICATING RESULTS

10.1. Registration in Public Database(s)*

Principal investigator assures that the key design elements of this protocol will be posted in a publicly accessible database, where applicable and in compliance with current recommendations.

Principal investigator also assures that key results of this study will be posted in a publicly accessible database within the required time-frame from completion of the data collection, where applicable and in compliance with current recommendations.

10.2. Publications*

Further to legislated data disclosure, the results of this study will be published as a scientific paper in a peer-reviewed journal. Such a manuscript will be prepared independently by the investigators and in accordance with the current guidelines of STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) and properly reported to improve reproducibility and facilitate validity assessment of healthcare database studies [7]. The ADVANCE Steering Committee will be entitled to view the results and interpretations included in the manuscript and provide comments prior to submission of the manuscript for publication

11. DISCLAIMER

The study described in this protocol is conducted as part of the IMI ADVANCE project with the aim to test methodological aspects of the design, conduct and reporting of studies for vaccine benefit-risk monitoring activities. The protocol presented herein relates solely to the testing of these methodologies and is not intended to inform regulatory or clinical decisions on the benefits and risks of the exposures under investigation. Therefore any use of information from these studies should carry over this warning and be used accordingly..

12. REFERENCES

(1) **Donegan K, et al.** Bivalent human papillomavirus vaccine and the risk of fatigue syndromes in girls in the UK. *Vaccine* 2013; **31**(43): 4961-4967.
(2) **Yih WK, et al.** Prospective influenza vaccine safety surveillance using fresh data in the Sentinel System. *Pharmacoepidemiol Drug Saf* 2016; **25**(5): 481-492.
(3) **Becker BFH, et al.** CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. *Pharmacoepidemiol Drug Saf* 2017; **26**(8): 998-1005.
(4) Winston Chang JC, JJ Allaire, Yihui Xie and Jonathan McPherson. Shiny: Web Application Framework for R. In, 2016.
(5) **Leite A, Andrews NJ, Thomas SL.** Near real-time vaccine safety surveillance using electronic health records-a systematic review of the application of statistical methods. *Pharmacoepidemiol Drug Saf* 2016; **25**(3): 225-237.
(6) **Kurz X, et al.** The ADVANCE Code of Conduct for collaborative vaccine studies. *Vaccine* 2017; **35**(15): 1844-1855.
(7) **Wang SV, et al.** Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf* 2017; **26**(9): 1018-1032.

13. APPENDIX 1: QUALITY ASSESSMENT

Table: Assessment sheet for the quality /feasibility of the database to participate in the POC studies

Quality of information				
META-DATA				
Category	**Data**	**Measure(s)**		**Origin of information**
Provenance of information	Sources for diagnoses (as codes or text) • primary care • outpatient specialist • hospital discharge • emergency admission • causes of death	per type yes/no	Certain missingness?	Data provenance questionnaire to databases
	Vaccinations • routine childhood • HPV • travel • influenza • voluntary Drugs • prescribed/dispensed primary by GP • prescribed/dispensed by specialist • prescribed/dispensed during hospitalization	per type yes/no	Certain missingness?	AIRR survey
	Diagnostic tests • primary care outpatient specialist during hospitalization	yes/no	Certain missingness? Results?	AIRR survey
POPULATION				
Size	Number of lives (at any point in time) in population	N		Population fingerprint
	Number of subjects active at 1/1/2015	N		Population fingerprint
Dates	Missing Birthdate (no valid date entry (to be supplied by database owner)	N	Percentage on total number of lives	Attrition diagrams DBs
	Birth dates (day of birth independent of month)	Frequency of each day of the month of the DOB (1-31)	Percentage on total number of lives	Vaccine fingerprint (R)
Observation Time & lag time	Origin for the start of follow-up • birth (start of follow-up = birth) • registration with database (start of follow-up > 1 month after date of birth)	N	Percentage of total	Jerboa Event fingerprint
	Origin for the end of follow-up • death (end of follow-up = date of death in event file)	N	Percentage of total	Jerboa Event fingerprint

	• exiting from database (end of follow-up < last data availability for practice	Median (5^{th}, 95^{th} percentile of lag time from date of delivery till last data)		
Gender/age	Population age Distribution (Overall and by sex) * at 1/1/2015 (representativeness of population)	N	Compared to national statistics (see D5.2)	Population fingerprint
Per type of vaccination				
Vaccinations: BCG, DTaP, DTwP, polio, Hib, HPV, Seasonal Influenza	Granularity of vaccine exposure data • vaccine type • ATC code • brand	N	Percent of total (vaccinetype) for these levels	Vaccine fingerprint
	Recorded dose vs. Derived dose vs. sequence (all possible combinations)	Cross-tabulation		Vaccine fingerprint, R
	Vaccination records without dose	N		Vaccine fingerprint, R
	Coverage in birth cohorts at age ….	Estimated Coverage (per methodology as developed in ADVANCE)	Comparison against WHO data, VENICE and local information	Vaccine fingerprint, R
	Coverage by dose	histogram of doses		Vaccine fingerprint, R
Per database and event				
Events	Name of event			
	Availability of codes	List of available codes per data domain	Frequency of each code in input files	Event fingerprint, Jerboa
	List of components	Name and description of query		Event team
	Frequency of events as detected by each component algorithm	Table of frequency of possible combinations		Algorithm comparison module of Jerboa
	Frequency of event as detected according to chosen algorithm(s)	Frequency by year		Component analysis
	Chosen algorithm and reason			Component analysis
Validity	PPV of chosen algorithm(s)	%	confidence measure	Output of the validity workflow
	Sensitivity of chosen algorithm(s)	%	confidence measure	Output of the validity workflow
	Specificity of chosen algorithm(s)	%	confidence measure	Output of the validity workflow
	Procedure to obtain the above estimates			Output of the validity workflow
External benchmarks	Validation Studies	Summaries of previously conducted validation		Event team & database experience

		studies in the database		
	Estimates of frequency of the event in the population represented by the database according to external data sources (e.g. literature)	Available estimates with source (and comments)		Event team & database experience

14. APPENDIX 2: COMMON DATA MODEL

The database partners will do study-specific extractions from their database structure and transform these data into the common data model structure (see below). Data will be formatted in the following input files and will stay local. The data format is identical for all the data-sites.

14.1. Patients.txt

For each patient, there is one record (in the case of patients' reentering the population either a suffixed ID number shall be used or the most relevant follow-up period shall be chosen) in the *Patients.txt* input file containing the following variables:

Table S14.1: Patients.txt

Name of the Variable	Description of the Variable	Format / Possible Values	Description of the Values
PatientID	Patient Identifier	any string maximum length: 32 characters	
Birthdate	Date of birth	YYYYMMDD.	
Gender	Gender	F M	Female Male
Startdate	Date from which the patient is eligible to be included in the study. This is typically the date the patient is entered into the registration system (date of registration with insurance/region, date GP started to collaborate). Run in periods should not be applied when defining StartDate	YYYYMMDD.	
Enddate	Date after which the patient is no longer eligible for inclusion in the study (e.g. end of registration with GP/database, insurance, moving out, death, last data draw down (whichever is earliest)).	YYYYMMDD.	

14.2. Vaccinations.txt

For each vaccination, there is one record in the *Vaccinations.txt* input file containing the following variables:

Table S14.2: Vaccinations.txt

Name of the Variable	Description of the Variable	Format / Possible Values	Description of the Values
PatientID	Patient Identifier	any string maximum length: 32 characters	
Date	Date of administration	YYYYMMDD.	
Brand	Product name of the vaccine	any string	

VacType	Type of vaccination	Will contain components of the vaccine separated by hyphens. The different components codes are listed in Annex 5.	Pertussis vaccines used in the POC I study will be any containing uPE (unclassifiable into acellular or whole-cell), aPE (acellular pertussis vacc) or wPE (whole-cell vacc) components.
ATC	ATC code of the vaccine	7 characters long string	May be shorter if not full ATC
DoseRecorded*	Dose received as specified in the database	P1, P2, etc. B1, B2, etc.	For priming doses For booster doses
DoseDerived*	Defined as dose per recommendation, as determined by the database custodian based on knowledge of the local immunization schedule and as documented in an analytic variable generated for the study	P1, P2, etc. B1, B2, etc.	For priming doses For booster doses

* At least one of them is mandatory. Databases may choose which dose variable to prioritize based upon local expertise and results of the vaccination fingerprint. Both can be filled in at the same time.

14.3. Events.txt

This input file contains information about comorbidity and diagnostic events (events) of interest for the persons in *Patients.txt*; events with an invalid Patient ID or Date variable cannot be included.

For every event of interest for each patient in *Patients.txt* there is one record in the *Events.txt* input file containing the following variables:

Table S14.3: Events.txt

Name of the Variable	Description of the Variable	Format / Possible Values	Description of the Values
PatientID	Patient Identifier	any string maximum length: 32 characters	
Date	Date of event	YYYYMMDD.	
Eventtype *	Type of event	PERT	PERT
		FCONVULS	FCONVULS
		FEVER	FEVER
		HHE	HHE
		PCRYING	PCRYING
		SOMNOL	SOMNOL
Code	SEE APPENDIX 3		

15. APPENDIX 3: CODE LISTS OF EXPOSURE EVENTS

ATC code	ATC name	Acellular Pertussis
J07AJ52	Pertussis, purified antigen, combinations with toxoids	aP
J07CA02	Diphtheria-pertussis-poliomyelitis-tetanus	aP
J07CA06	Diphtheria-hemophilus influenzae B-pertussis-poliomyelitis-tetanus	aP
J07CA09	Diphtheria-hemophilus influenzae B-pertussis-poliomyelitis-tetanus-hepatitis B	aP
J07AG52	Hemophilus influenzae B, combinations with pertussis and toxoids	aP
J07AJ02	Pertussis	aP
J07AJ51	Pertussis	wP
J07CA11	Diphtheria-Hemophilus influenzae B-pertussis-tetanus-hepatitis B	?
J07CA12	Diphtheria-pertussis-poliomyelitis-tetanus-hepatitis B	?
J07CA13	Diphtheria-hemophilus influenzae B-pertussis-tetanus-hepatitis B-meningococcus A + C	?

16. APPENDIX 4: CODE LISTS OF BENEFIT AND RISK EVENTS

Variable	Role	Data source(s)	Operational definition*
Fever	Safety outcome	Participating healthcare databases	ICD9:078.2, 780.6 ICD10: P81.9, R50 READ-CTV3: 1653. 2E34. ,2EZ.., A782., R006.,R0060, R0061, R0062, R0063, R006z, X76Df, X76Di, X76Dk, X76Dl, X76EF,X76EI, X76EI, XM05S, XM09q, XM0yv, XM0yw, XM1AX, Xa9sd, READ-v2: 165.., 165..,2E...,A782.,R006. ICPC: A03
Somnolence	Safety outcome	Participating healthcare databases	ICD-9:780.09, 780.54 ICD-10: G47.1, R40.0 READ-CTV3: R0000, R0001, R0054,X007w, X007x, XM06R, Xa2bY, XaC0p READ-v2: 1B67., 1BX1.,2234.,E2743, R0000,R0001, R0054 ICPC: none (text extractions)
Persistent crying, irritability	Safety outcome	Participating healthcare databases (except SSI/AUH)	ICD9: 780.92, 780.95 ICD10: R45.83, R68.11 READ-CTV3: 1B1I0, Xa2lv, Xa9zv READ-v2: 1B1I0 1B1P. ICPC: A15
Febrile convulsions	Safety outcome	Participating healthcare databases	ICD-9: 780.31; 780.32 ICD10: R56.0 READ-CTV3: R0030; XM03I READ-v2: 1B6B.; R0030 ICPC: none
Hypotonic hypo-responsive episode (HHE)	Safety outcome	Participating healthcare databases	ICD-9: 770.88; 782.5; 782.61; 799.02 ICD10: P84; R09.02; R23.0; R23.1 READ-CTV3: R025.; R0260; R2y01; X76qQ; XM041; XM07N; XM09U; XM09U;Xa9E7 READ-v2: R025.; R0260; R2y01 ICPC: none

Variable	Role	Data source(s)	Operational definition*
Pertussis	Benefit outcome	Participating healthcare databases	ICD-9: 033.9; 484.3 ICD-10: A37 READ-CTv3: A33y. A33yz; A33z.; Ayu39; Ayu3A; H243.; X70I8; XE0Qw; XE0Qw; XM00D READ v2: A33.. Ayu39; Ayu3A; H243. ICPC: R71

*any of the codes would qualify