www.advance-vaccines.eu

Accelerated Development of VAccine beNefit-risk Collaboration in Europe

IMI JU Grant Agreement nº115557

POC feasibility study Protocol

(RISK PILLAR)

Testing new approaches to monitoring benefit/risk with pertussis vaccines as test case: Incidence rates of safety outcomes of whole-cell pertussis and acellular pertussis vaccines in pre-school children.

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

Disclaimer: *The study described in this protocol are 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 1.4

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Daniel Weibel	02-07-2015	0.1	First draft outline
Miriam Sturkenboom	09-07-2015	0.1	Comments
Daniel Weibel	16-07-2015	0.2	Revised draft outline
Vincent Bauchau, Jan Bonhoeffer, Sandrine Gilhet-Mailfait, Raphael Roten, Victoria Abbing-Karahagopian, Caitlin Dodd, Kaat Bollaerts, Suzie Seabrooke, Nicoline Van der Maas, Jan Cleerbout, Benedikt Becker, Edouard Ledent, va570273	21-07-2015	0.2	Comments and contributions
Daniel Weibel	23-07-2015	0.3	Revised draft
Miriam Sturkenboom, Vincent Bauchau, Edouard Ledent	07-08-2015	0.3	Comments
Daniel Weibel	08-08-2015	0.4	Revised draft
Miriam Sturkenboom	09-08-2015	0.4	Final edits prior to SC
Miriam Sturkenboom, Vincent Bauchau, Edouard Ledent, John Weil, Marianne van der Sande, ECDC, NVD	31-08-2015	0.4	SC comments
Daniel Weibel, Caitlin Dodd, Miriam Sturkenboom	16-09-2015	0.5	Revised draft
Vincent Bauchau/ Miriam Sturkenboom	17-09-2015	0.6	Revised draft
Elisa Martín Merino, Consuelo Huerta, Raphaele Roten	09-10-2015	0.6	ADVANCE consortium comments
Daniel Weibel	16-10-2015	1.0	1. Version for further medical writer treatment.
Denis Macina, Sandrine Gilhet-Mailfait, Vincent,	23-10-2015 – 4-11-2015	1.0	Comments
Maxwell Gough	2-11-2015	1.0	Treated by medical writer

NAME	DATE	VERSION	DESCRIPTION
Daniel Weibel, Miriam Sturkenboom	23-11-2015	1.1	Cleaned from Medical Writer, Revised version
Miriam Sturkenboom	12-01-2016	1.1	Study protocols distributed to databases and SC
ECDC, GSK, SPSMSD	12-02-2016 – 15-03-2016	1.1	Comments on Version 1.1
Daniel Weibel	24-03-2016	1.2	Revised version
Miriam Sturkenboom, Vincent Bauchau, Daniel Weibel, Lisen Arnheim, Rosa Gini, Caitlin Dodd	12-4-2016	1.3	Additions of system testing parts from POC evaluation team and feasibility testing from fingerprint documents in response to comments from SPMSD, SP etc. Insertion of quality criteria (in collaboration with WP4).
Daniel Weibel	04-5-2016	1.4	Updated section "8.1 Regulatory and Ethical Compliance"

Title	Testing new approaches to monitoring benefit/risk with pertussis vaccines as test case: Incidence rates of safety outcomes of whole-cell pertussis and acellular pertussis vaccines in pre-school children.
Medicinal product	All available whole-cell pertussis- and acellular pertussis-containing vaccines
Product reference	Any acellular pertussis- and whole-cell pertussis-containing vaccines
Research question and objectives	The overall ADVANCE proof-of-concept (POC) question is to test the system for benefit-risk monitoring of vaccines in Europe. This will first be done by using test cases. For this POC feasibility study, the following research question is used: "Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from whole-cell pertussis vaccines to acellular pertussis vaccines"? The objectives of this specific study, which focuses on the incidence rates of safety outcomes of pertussis-containing vaccines, are: 1. To evaluate participating databases on quality criteria for inclusion in the study 2. To feed safety information into the benefit/risk analysis. The safety information required for the analysis is: To estimate incidence rates of specific events (i.e. injection site reactions, fever, somnolence, persistent crying, irritability, febrile or afebrile seizure/convulsion, hypotonic-hyporesponsive episode, extensive limb swelling) a) within specific risk windows after each dose of whole-cell pertussis or acellular pertussis vaccines in pre-school children b) in the time period outside the risk windows before and after each dose of whole-cell pertussis or acellular pertussis vaccines in pre-school children c) Over calendar time, to allow for analysis that will focus on sequential monitoring of B/R
Countries of study	Electronic health care databases (record linkage, surveillance and GP-based databases) currently available in the ADVANCE consortium and eligible are located in Denmark, Spain, Italy, The Netherlands and UK. Short descriptions of databases and their full names will be included in this protocol upon final decisions of inclusions.
Authors	**Protocol Team**: Daniel Weibel, Caitlin Dodd, Miriam Sturkenboom **Authors from POC feasibility study outline document** (Parts of this study protocol have been copied from the POC feasibility study outline document) **Main Authors**: Nicoline van der Maas, Kaat Bollaerts, Denis Macina, Miriam Sturkenboom, Vincent Bauchau **Reviewers**: Simon de Lusignan, Hanne Dorthe Emborg, Mendel Haag, Michael Greenberg, Ulrich Heininger, Alena Khromava, Piotr Kramarz, Xavier Kurz, Harshana Liyanage, Patrick Mahy, Laurence Pagnon, Tin Tin Htar Myint, Marianne van der Sande, John Weil, Eddy Ziani

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADVANCE	Accelerated Development of VAccine beNefit-risk Collaboration in Europe
AEFI	adverse event following immunization
aP	acellular (pertussis vaccine)
ATC	Anatomical Therapeutic Chemical
CDM	common data model
CI	confidence interval
D1, …, D5	1st dose, …, 5th dose
DTwP	diphtheria tetanus whole-cell pertussis (vaccine)
DT	diphtheria tetanus (vaccine)
EMC	Erasmus Medical Center
HHE	hypotonic-hyporesponsive episode
ICD	International Classification of Diseases
IEC	independent ethics committee
IMI	Innovative Medicines Initiative
IRB	institutional review board
IR	incidence rate
OR	odds ratio
POC	proof-of-concept
REB	research ethics board
RRE	remote research environment
SCCS	self-controlled case series
WHO	World Health Organization
WP	work package (i.e. WP5)
wP	whole-cell (pertussis vaccine)

1. RESPONSIBLE PARTIES

1.1. Main Author(s) of the Protocol

Name	***Institution***	***Role***	***Contribution***
Daniel Weibel	Erasmus University Medical Center, Rotterdam, The Netherlands	Principal investigator	First draft v0.1 and updates until last version
Miriam Sturkenboom & Vincent Bauchau	Erasmus University Medical Center, Rotterdam, The Netherlands	WP5 leaders	Comments to v0.1-1.2 Insertion of fingerprinting information, system testing, quality
Caitlin Dodd	Erasmus University Medical Center, Rotterdam, The Netherlands	Statistician	Section 7.4
Jan Cleerbout	GSK	Safety Physician	Risk factors and risk windows
Main Authors: Nicoline van der Maas, Kaat Bollaerts, Denis Macina, Miriam Sturkenboom, Vincent Bauchau Reviewers: Simon de Lusignan, Hanne Dorthe Emborg, Mendel Haag, Michael Greenberg, Ulrich Heininger, Alena Khromava, Piotr Kramarz, Xavier Kurz, Harshana Liyanage, Patrick Mahy, Laurence Pagnon, Tin Tin Htar Myint, Marianne van der Sande, John Weil, Eddy Ziani		POC Outline document authors and reviewers	Sections 5 and 7.5 were largely extracted from the POC outline
Germano Ferreira, Lampros Stergiopoulos	SURREY, P95	Section on POC evaluation	
Caitlin Dodd, Rosa Gini	EMC, ARS	Section on quality of database	In collaboration with WP 4

1.2. Principal Investigator

- Daniel Weibel, Erasmus University Medical Center, Rotterdam, The Netherlands: d.weibel@erasmusmc.nl

1.3. Study Team

STATISTICIANS

- Caitlin Dodd: EMC
- Olivia Mahaux, Elizabeth Merall: GSK

VACCINOLOGISTS

- Nicoline van der Maas: RIVM

ADDITIONAL MEMBERS

- Jan Cleerbout, Victoria Abbing, Catherine Cohet: GSK
- Silvia Pérez-Vilar: FISABIO & Erasmus MC
- Sandrine Gilhet-Mailfait: SPMSD
- Benedikt Becker: EMC
- Raphaele Roten: Crucell
- Danitza Tomianovic: UNIBAS
- Emil Cochino: EMA
- Suzie Seabroke: MHRA

DATABASE LIAISONS/CUSTODIANS

Databases will participate upon demonstration of data quality in the quality assessment module (fingerprinting) that will take place prior to start of the POC feasibility study. Custodians of contributing databases will be members of the study team.

Table 1: List of potential databases for feasibility testing

Datasource	Contact
BE_network of sentinel GPs	Viviane.VanCasteren@wiv-isp.be
BE_Pedisurv	martine.sabbe@wiv-isp.be
DK_SSI	Hanne-Dorthe Emborg (hde@ssi.dk)
DK-AUH	Lars Pedersen (lap@clin.au.dk)
ES_BIFAP	Elisa Martin Merino (emartin_fcsai@bifap.org), Consuelo Hueta (chuerta_fcsai@bifap.aemps.es)
ES_FISABIO	Silvia Pérez Vilar (perez_silvil@gva.es)
ES_IDIAP	Talita Duarte Salles (tduarte@idiapjgol.org)
FI_HPVCHRT	Matti Lehtinen (matti.lehtinen@uta.fi
IT_ARS	Rosa Gini (rosa.gini@ars.toscana.it)
IT_ASLCR	Salvatore Mannino, Silvia Lucchi (silvia.lucchi@aslcremona.it)
IT_Arianna	Gianluca Trifiro (trifirog@unime.it)
IT_PEDIANET	Gino Picelli (g.picelli@virgilio.it), Anna Cantarutti (anna.cantarutti@gmail.com)
NL_IPCI	Miriam Sturkenboom (m.sturkenboom@erasmusmc.nl)
NL_RIVM	Hester de Melker (hester.de.melker@rivm.nl)
SE_KI	Lisen Arnheim-Dahlstrom (lisen.arnheim.dahlstrom@ki.se)
UK_RCGP	David Mullett (d.mullett@surrey.ac.uk)
UK_THIN	Daniel Weibel (d.weibel@erasmusmc.nl)

2. ABSTRACT

Date of Protocol Abstract:
April 16, 2016

Title of Study: Testing new approaches to monitoring benefit/risk with pertussis vaccines as test case: Incidence rates of safety outcomes of whole-cell pertussis and acellular pertussis vaccines in pre-school children.

Study Period: 01 January 1990 – 31 December 2015

Rationale and Background: The overall ADVANCE proof-of-concept (POC) question is to test the system for benefit-risk monitoring of vaccines in Europe. This will first be done by using test cases. For this POC feasibility study, the following research question is used: "Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from whole-cell pertussis (wP) vaccines to acellular pertussis (aP) vaccines?"

This protocol aims to create the safety data for a benefit-risk analysis of aP versus wP vaccines.

Research Question and Objectives:

The objectives of this specific POC feasibility study focusing on the incidence rates of safety outcomes of pertussis vaccines are:

1. To evaluate participating databases on quality criteria for inclusion in the study (i.e. vaccination data on pertussis vaccine available, at least one of the outcomes available, data access and clearance of protocol possible within timelines of POC feasibility study).
2. To provide safety information for a benefit/risk analysis model.

   The safety information required for the model is: incidence rates of specific events (i.e. injection site reactions, fever, somnolence, persistent crying, irritability, febrile or afebrile seizure/convulsion, hypotonic-hyporesponsive episode [HHE], extensive limb swelling) in risk and baseline periods. Incidence rates will be estimated within specific risk windows after each dose of wP or aP vaccine in pre-school children and within the periods outside the risk windows (baseline).
3. To provide calendar time specific incidence data as test for methods development in ADVANCE WP4.

Study Design: The main study design is a retrospective dynamic cohort study.

The study will be conducted utilizing electronic health care data from ADVANCE partners in different countries (i.e., Short descriptions of databases in respective coutries and their full names will be included in this protocol upon final decisions of inclusions).

In databases that cover the period in which wP was still provided, the incidence rate (IR) ratio for wP versus baseline risk will be calculated using a self-controlled case series (SCCS) design. This IR ratio will be applied to baseline rates for the aP period in databases that do not capture the wP period, to estimate the rate of events during wP in each specific database.

Population: The study population will comprise all children registered in any of the participating databases during the study period and for whom an adequate start and end of follow-up and date of birth can be defined.

Children will be followed from start of the study period, one month after date of birth (i.e. to allow for pre-vaccination time for the SCCS design and to avoid pre-term related or birth-induced increase in incidence rates), or date of valid data in the database (whichever is the latest) until the end of study period (31-12-2015, the school-entry pertussis booster, transferring out of the database, death, reaching age 6 years: whichever is the earliest).

Variables:

Exposure of interest

Any wP vaccines and aP pertussis-containing vaccines and their doses in the vaccine schedule (D1, D2, D3, D4, D5)

Outcomes

- Injection site reactions: erythema, edema, induration/nodule/sterile abscess, pain/tenderness
- Fever
- Somnolence
- Persistent crying, irritability
- Generalized convulsive seizures
- HHE
- Extensive limb swelling

Data Sources:

- Electronic health care databases (record linkage, surveillance and GP-based databases) currently available in the ADVANCE consortium and eligible are located in Denmark, Spain, Italy, The Netherlands and UK. Short descriptions of databases and their full names will be included in this protocol upon final decisions of inclusions.

Study Size: Total population (0-6 year of age) of all eligible ADVANCE databases

Data Analysis: The purpose of this study is to provide incidence rates (i.e. baseline and risk-window specific) of known adverse reactions following vaccination with pertussis-containing vaccines for use in a multi-criteria decision analysis (MCDA) model of benefits and risks of wP versus aP pertussis vaccines (models are described in a separate benefit-risk study protocol). In some more recent databases, wP information will not be captured. To generate risk-window specific incidence rates for the wP period in these databases, the IR ratio originating from an SCCS analysis of wP versus baseline in other databases will be multiplied by the baseline IR.

Informed Consent and Ethical Approval: This study will be conducted on the basis of secondary use of electronic healthcare records. Each database will apply local governance and privacy rules prior to aggregating and sharing anonymized data.

Milestones:

Draft protocol: July 31 2015

Submission to SC: August 6, 2015

Comments from SC: August 31, 2015

Submission for consortium review: September 2015

Submission to in house clearances/ governance boards: January 2016

Updated protocol after review: April 15, 2016

Final data extraction to CDM: June 15, 2016

Running scripts and submission to RRE: June 30, 2016

Data analysis: July 2016

Data interpretation and reporting: August 2016

Final report of study results: September 2016

3. AMENDMENTS AND UPDATES

Protocol amendments following IRB approval:

Table 2: Overview of Protocol Amendments and Updates

Number	*Date (DDMMMYY)*	*Section of the study protocol*	*Amendment or update*	*Reason*
1				
2				
….				

4. MILESTONES

Table 3: Overview of Study Milestones

Draft protocol: July 31 2015

Submission to SC: August 6, 2015

Comments from SC: August 31, 2015

Submission for consortium review: September 2015

Submission to in house clearances/ governance boards: January 2016

Updated protocol after review: April 15, 2016

Final data extraction to CDM: June 15, 2016

Running scripts and submission to RRE: June 30, 2016

Data analysis: July 2016

Data interpretation and reporting: August 2016

Final report of study results: September 2016

5. RATIONALE AND BACKGROUND[1]

The ADVANCE vision is to deliver "best evidence at the right time to support decision-making on vaccination in Europe". The mission is to establish a prototype of a sustainable and compelling system that rapidly provides best available scientific evidence on vaccination benefits and risks post-marketing for well-informed decisions. In light of this goal, the ADVANCE platform aims to provide evidence on the benefits and risks of vaccines to support decision-making by all stakeholders in a wide range of contexts. Examples of scenarios are the inclusion of a new vaccine in a vaccination program, and the occurrence of a new safety issue, e.g. when the benefits of the vaccine are questioned or when a new population is targeted (see Pertussis POC Outline).

The concept this POC feasibility study aims to demonstrate is as follows: in the event that an important decision regarding a health intervention is to be made, a benefit-risk assessment will be carried out. Upon a favorable benefit-risk assessment, the health intervention is implemented and the benefits and risks are monitored to investigate whether the benefit-risk balance is changing over time. The benefit-risk monitoring may focus primarily on the benefits and risks that could potentially modify the benefit-risk balance. If there is a strong indication that the benefit-risk has changed over time, a full re-assessment of the benefit-risk balance of the health intervention may be triggered using all accumulated evidence available at that point in time. To inform the benefit-risk assessment and monitoring, electronic health care databases available within Europe will be used.

To be able to prove this concept of benefit-risk monitoring in ADVANCE without waiting for the evidence to accumulate prospectively, we will start from a historical decision and simulate monitoring through a retrospective analysis. Pertussis vaccination, particularly comparing wP and aP vaccination, was chosen by the ADVANCE Steering Committee as the subject of the first POC feasibility study. Therefore, the starting point of the current POC feasibility study study is the historical decision to switch from wP to aP vaccination in children in the pioneering countries.

PERTUSSIS DISEASE

Pertussis, also referred to as whooping cough, is a highly contagious respiratory disease caused by bacteria of the *Bordetella* genus, mainly *B. pertussis*, although other *Bordetella*-species also occur [1,2]. Pertussis is acquired through transmission of large respiratory droplets generated by coughing or sneezing from infected persons [3].

Transmission by the indirect route occurs extremely rarely if ever [4]. *B. pertussis* causes respiratory symptoms, along with systemic effects, presumably mediated by secreted toxins [4]. Infections range in clinical presentation from asymptomatic to severe. They are most severe, even life-threatening, in young infants before they are immunized [5]. It is less frequently severe in older children, adolescents, and adults.

[1] This section is obtained from the POC outline: https://www.dropbox.com/s/ioru753h9h8cy44/240315_POC%20pertussis%20outline_version%201.5_tobedistributed.docx?dl=0

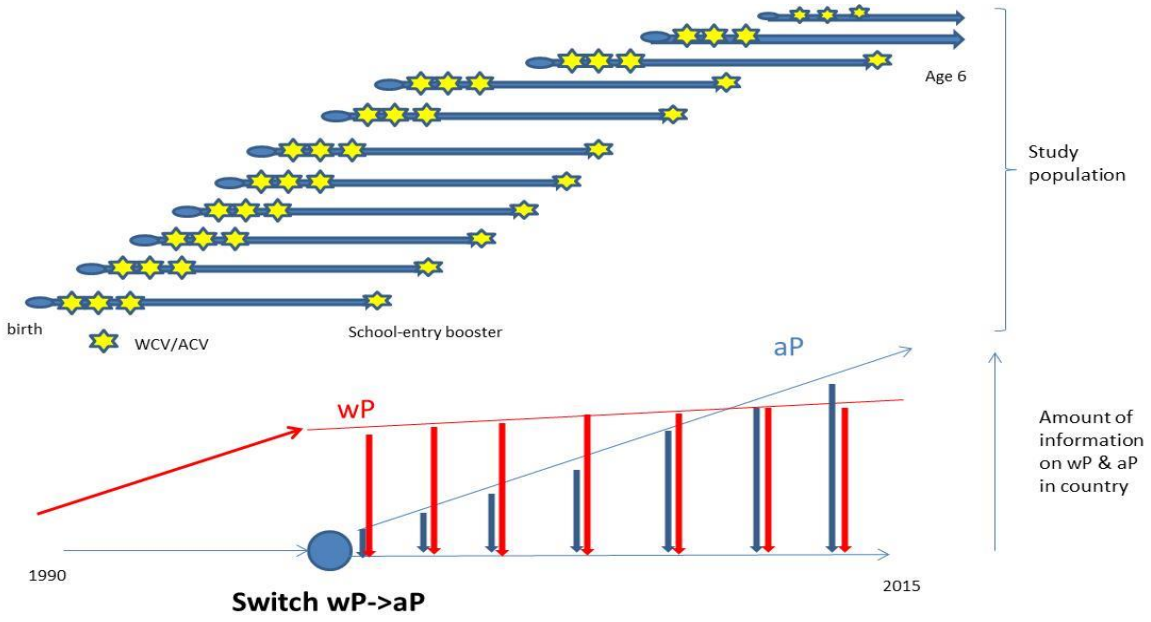


Figure 1:Graphic Representation of Time Axes/Horizons

For the study, a retrospective approach is taken (i.e. all benefits and risks to be measured have already occurred). For the decision analysis, benefits and risks will be compared between wP and aP. The evidence for wP vaccines has been accumulated mostly prior to the switch, with some data coming afterwards from the last vaccinated persons. Evidence from aP vaccines at the time of the switch comes mostly from clinical data and is complemented by post-marketing data after the switch. The arrows in Figure 1 show that the increasing amount of evidence on benefits and risks for wP and aP will be cumulatively assessed from the time of the switch. The upper part represents the time frame for the study subjects; these will be followed from start of the study period, one month after date of birth (i.e. to allow for pre-vaccination time for the SCCS design and to avoid pre-term related or birth-induced increase in incidence rates), or date of valid data in the database (whichever is the latest) until the end of study period (31-12-2015, the school-entry pertussis booster, transference out of the database, death, attainment of age 6 years: whichever is the earliest). Rates and risks of benefits and risks will be assessed by type and dose of aP or wP. Note: the primary vaccination series in infancy follows a '2+1' or '3+1' schedule, depending on country.

Typical clinical disease is characterized by three phases. After 7-28 days of incubation, the catarrhal phase (1-2 weeks) is largely nonspecific with coryza, eye redness, frequent coughing and sneezing. It is followed by a 1-6 week-long paroxysmal phase during which intense paroxysms of cough may lead to choking, emesis and the characteristic inspiratory whoop [3]. In very young infants, cough is often absent and apnea seems more characteristic of the disease [6]. Fever is rare in pertussis. The convalescent phase sees declining symptoms over widely variable duration [3].

Clinical criteria for the diagnosis of pertussis include a cough lasting at least two weeks and at least one of the following three: paroxysms of coughing, inspiratory 'whooping', and/or post-tussive vomiting; or any person diagnosed as pertussis by a physician, or apnoeic episodes in infants.

Pertussis infection may be followed by common but usually self-limiting complications such as apnea, seizures, vomiting, gastroesophageal reflux, rib fracture, subconjunctival hemorrhages, epistaxis or

syncope secondary to the paroxysms [4,5]. According to the Institute of Medicine report[2], apnea and respiratory arrest are the most common complication of pertussis followed by pneumonia and gastroesophageal reflux. Encephalopathy is a rare complication and occurs most often in younger patients. Other complications include seizures, ataxia, aphasia, blindness, deafness, subconjunctival hemorrhages, syncope, and rib fractures. Pertussis is most serious in infants less than 12 months of age, and the risk of death is highest among infants less than 6 months old.

TYPE OF PERTUSSIS VACCINES

Vaccines against pertussis were developed in the 1920s and have been used more widely since the 1940s [7]. The first vaccines were wP vaccines containing suspensions of killed *B. pertussis* organisms [4]. The production process varied between different wP vaccines, resulting in differences in antibody responses. Furthermore, due to the use of the whole bacterium, composition and thus immunogenicity, efficacy and reactogenicity of a specific wP vaccine could change over time, from lot to lot within one wP vaccine, and from one wP vaccine to another. Reactogenicity of the wP vaccine is probably due to their endotoxin lipopolysaccharide (LPS) content [8]. In an attempt to reduce reactogenicity, aP vaccines were developed. They were used for the first time in 1981 during mass vaccination campaigns in Japan [9], and more widely during and since the 1990s. aP vaccines contain purified secreted and surface components of *B. pertussis* which, based on animal models, are thought to play an important role in pathogenesis and induction of immunity [4]. Later on, several aP vaccines were manufactured, containing between one and five different pertussis components [2].

All aP vaccines contain at least a detoxified pertussis toxin (PT); the second antigen added in all formulations with two or more components is filamentous haemagglutinin (FHA); three-component vaccines contain also pertactin (PRN); finally, four- and five-component vaccines contain one or more fimbrial agglutinogens or fimbriae (FIM). Almost all aP vaccines are adjuvanted with aluminium salts and combined with diphtheria and tetanus toxoids, and possibly also additional vaccine valences such as inactivated poliovirus, *Haemophilus influenzae* type b and/or hepatitis B [4].

PERTUSSIS VACCINATION SCHEDULES IN EUROPE

Between 2004 and 2015 several countries switched from wP to aP vaccines for infants and children. All other countries in Europe switched to aP vaccines prior to 2006 (starting in the 1990s for some countries). As of March 2015, all countries except Poland use exclusively aP vaccines. In most cases, the switch to aP vaccines was conducted over a narrow age-cohort, while only in very few cases, such as Poland, the switch occurred, or is occurring progressively over years. Different types of vaccines are being used. As part of the coverage pillar the way participating countries switched should be described.

Since the start of the introduction of pertussis vaccine in the 1940s, many countries have tended to adapt and customize the schedules of their vaccination programs, adding and removing doses, changing ages of primary and booster schedules, with or without catch-up campaigns, and transitioning from wP to aP vaccines for all doses, for one or more booster doses only, or not yet at all. As a result, pertussis vaccine schedules vary largely across Europe.

For detailed information on the schedules currently used in Europe, see Figure 2 and for more details see European Centre for Disease Prevention and Control (ECDC)-website (http://vaccine-schedule.ecdc.europa.eu/Pages/Scheduler.aspx).

[2] http://www.nap.edu/openbook.php?record_id=13164&page=529

Schedule type	First year of life			Second year of life	Third year of life	Preschool booster	Adolescent booster	Country
'2p+1'	From 6 weeks to 6 months		Around first birthday					
	P1	P2	B1			B2	B3	F, IT, FI, NO, IS, SK
								SE*, DK, RO, AT
'3p+1'	P1	P2	P3	B1		B2	B3	BE, BG, CZ, EE, DE, GR, HU, LI, LU,
				B1		B2		HR, CY, LV, LI, MT, NL, PL, PT, SI, ES
					B1			UK
						B1	B2	IE
Vaccine combination generally used in EU/EEA	• DTaP-IPV-HepB/Hib ('hexavalent') • DTaP-IPV/Hib ('pentavalent') • DTaP-IPV ('tetravalent') • DTwP-IPV/Hib ('whole-cell pertussis combo') • Hep B (used in conjunction with pentavalent)				• Hib-MenC combo	• DTaP-IPV • TdaP • Tdap-IPV • Tdap	• Td • Tdap • Tdap-IPV	

P=primary dose; B=booster dose

Figure 2:Recommended Pertussis Schedules from ECDC Report on Pertussis Vaccine Shortage (October 2015)

The World Health Organization (WHO) standards of childhood pertussis vaccination recommend a three-dose primary series administered between 6 weeks and 6 months of age, followed by a booster dose, preferably in the second year of life. As previously mentioned, various countries have adapted and customized their vaccination schedules according to their policy considerations. As of March 2015, 19 countries use a three-dose schedule for the primary series, either at 2-3-4 months of age (n=9) or 2-4-6 months of age (n=10). Most of these countries give a toddler booster dose towards the end of the first year of life (n=3) or during the second year of life (n=13). In one country this booster dose can be administered before or after the first birthday; two countries do not recommend any toddler booster dose. Seven countries administer a two-dose primary series at three and five months of age, followed by a booster for 12 month-olds. Only France has a two-dose schedule at two and four months of age followed by a booster at 11 months of age.

All countries, except Malta, recommend one (n=11), two (n=15) or three (n=4) further aP booster doses between two and 18 years of age.

Six countries recommend one or more aP booster doses for all adults and/or elderly. Two countries (United Kingdom and Ireland) only target pregnant women in their last trimester of pregnancy (since late 2012).

VACCINE RISKS

Safety Studies

In a study including 15,752 doses of diphtheria, tetanus, whole cell pertussis (DTwP) vaccine and 784 doses of diphtheria, tetanus (DT) vaccine performed in the Los Angeles area in 1978-1979, local redness, swelling and pain occurred in 37.4%, 40.7% and 50.9% of DTwP recipients compared with 7.6%, 7.6% and 9.9% of DT recipients, respectively. Frequencies increased in subsequent doses. Frequency of fever (≥38 °C) was 46.5% and 9.3% after DTwP and DT, respectively. Furthermore, high-pitched unusual crying, convulsions and HHE following DTwP were reported in 0.1%, 0.06% and 0.06% recipients, respectively. Those events were not reported after DT, probably due to the low number of DT vaccinees [3].

A large study, performed in the USA, assessed common reactogenicity of 13 candidate aP vaccines compared to wP vaccine. For all adverse events following immunization (AEFIs), frequencies after aP vaccines were lower than after wP vaccines. For fever (≥37.8 °C) frequency was 4.2% after the first dose of aP vaccine compared with 27.3% following wP vaccination. Temperature increased after subsequent doses. Incidence of local redness was 13.5% and 49.4%, local swelling was 8.7% and 39.7%, and local pain was 3.8% and 27.3% after aP and wP vaccination, respectively [10].

For rarer and more severe AEFIs, frequencies also differed. Persistent crying, very high fever (≥40.5 °C), HHE and seizures were respectively 7, 7, 17 and 6 times more likely after vaccination with wP compared with aP [11].

In a large (n=28,796) questionnaire study carried out in the Netherlands between 2004 and 2007, frequencies of rarer, more severe AEFIs were higher after the Dutch wP vaccine (used until 2005) compared with the aP vaccine (introduced from January 1st 2005 onwards) [12]. In total, 1.5% of the infants reported prolonged crying ≥3 hours after first dose at two months after wP vaccination compared with 0.4% after aP. Frequencies of very high fever (≥40.5 °C) after the fourth dose of wP and aP vaccines at 11 months were 0.8% and 0.2%, respectively, whereas frequencies for febrile convulsions after the fourth dose were 0.06% and 0.02%, and for HHE after the first dose were 0.12% and 0.03%, respectively.

A review by Jefferson et al showed that wP vaccines were associated with a significantly higher incidence of swelling and induration (odds ratio [OR] 11.7; 95% confidence interval [CI] 8.8-15.4), fever >39 °C (OR 3.4; 95% CI 2.1-5.5) and crying for >2 hours (OR 4.7; 95% CI 2.9-7.6) than placebo or DT-vaccinations without a pertussis component. Differences in the incidence of HHE and convulsions were not statistically significant. aP vaccines did not cause a higher incidence of local signs, fever, convulsions, HHE or prolonged crying than placebo or DT vaccine [13].

Permanent damage to the central nervous system was repeatedly reported as 'causally related' to wP vaccines [14]. However, the National Childhood Encephalopathy Study and a consecutive huge body of scientific evidence eventually showed there was no association between neurological sequelae and wP vaccines [15,16]. However, these concerns were behind the motivation to develop aP vaccines.

In an observational study evaluating the safety of hexavalent vaccines (DTaP-IPV-Hib-HepB) involving 78 preterm infants, immunization triggered transient cardiorespiratory events in 47% of infants (15% apnoea, 21% bradycardia, 42% desaturations). In a retrospective study involving 53 infants, transient apnea or bradycardia was observed in 13% of infants following immunization with pentavalent or hexavalent vaccines. Apnea seemed less frequent and less severe following DTaP than wP vaccines [17].

6. RESEARCH QUESTION AND OBJECTIVES

The overall ADVANCE POC question is to test the currently available system for benefit-risk monitoring of vaccines in Europe. There will be several system testing methods, using the following question: Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from wP to aP pertussis vaccines?

The objectives of this specific POC feasibility study, which focuses on the incidence rates of safety outcomes of pertussis-containing vaccines, are:

1. To evaluate participating databases on quality criteria for inclusion in the study (i.e. vaccination data on pertussis vaccine available, at least one of the outcomes available, data access and clearance of protocol possible within timelines of the POC feasibility study)
2. To provide information that feeds into the benefit/risk model. The safety information required for the model is:

   Incidence rates of specific events (i.e. injection site reactions, fever, somnolence, persistent crying, irritability, febrile or afebrile seizure/convulsion, HHE, extensive limb swelling)

   a) within specific risk windows after each dose of wP and aP vaccines in pre-school children
   b) during the period outside the risk windows before and after each dose of wP and aP vaccines in pre-school children.
   c) Over calendar time, to allow for analysis that will focus on sequential monitoring of B/R

3. To provide calendar time specific incidence data as test for method development in ADVANCE WP4.

7. RESEARCH METHODS

The overall ADVANCE POC question is to test the currently available system for benefit-risk monitoring of vaccines in Europe. The system will be tested around the following B/R question: Has the initial benefit-risk profile in children prior to school-entry booster been maintained after the switch from wP to aP pertussis vaccines?

7.1. Process and methodology for system-testing

Although the system testing will occur largely outside of this study, it is summarised here, with details provided in annexes.

The system testing follows several steps which are visualized in the figure 3 and described in the following chronological order:

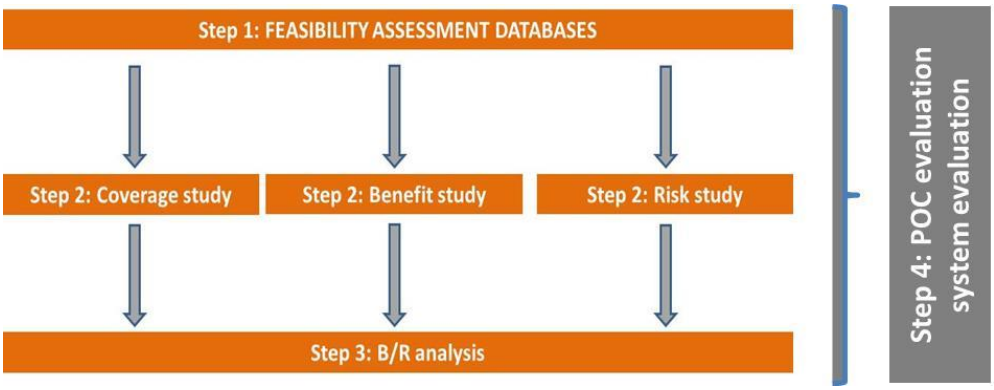


Figure 3 Visualization of the stepwise approach to the system testing

Step 1: Feasibility assessment of the databases: this step will assess whether the quality of the candidate database is sufficient for inclusion in the study. The focus will be on what type of data are available in the databases and whether population, events, and exposure may be misclassified. This step is largely based on the so-called fingerprinting which has been described in deliverable D5.2. A summary of the components and methods is provided in appendix 1. A quality assessment summary will be created per database, with decisions whether the data-sources can or cannot participate in the different pillar studies (see below under 7.2). Responsibility of fingerprinting lies with the workpackage leaders.

Step 2: Estimation and delivery of the rates for coverage, benefits and risks, this is described in the different 'pillar' protocols in the databases that may generate adequate results according to the feasibility assessment. Responsibilities are with the study teams that have generated the protocols

Step 3: B/R analysis: integration of the incidence rates (generated from step 2) with the utilities to generate a B/R model, as described in the B/R analysis protocol, responsibility with the B/R study team

Step 4: Evaluation of the studies and the systems used. This is conducted by a POC evaluation team which is separated on purposes from the POC study teams. A description of the framework for the POC evaluation is attached in appendix 2.

This purpose of this protocol is to describe in detail the methods for the risk study in step 2.

7.2. Methods for estimations in the scientific question

7.2.1. Design

This is a retrospective dynamic cohort study to estimate incidence rates of safety outcomes.

A SCCS design will be used to estimate IR ratios for wP vaccines versus baseline periods in databases that capture data on wP vaccinations and will be applied to databases that do not capture data on wP vaccinations to estimate IR in risk windows for those databases.

7.2.2. Setting

The POC feasibility study will be conducted in multiple population-based healthcare databases in various European countries.

7.2.3. Databases/Data Sources

The POC feasibility study will be conducted on data in electronic healthcare databases that reside with partners and associate partners of the ADVANCE consortium. Based on an assessment of the quality of information on exposure and outcomes, which will take place as part of the fingerprint process prior to study initiation for all the outcomes, the databases will be selected. The quality criteria for a database (i.e., inclusion criteria) are:

a) Vaccination data on pertussis vaccine available

b) At least one of the outcomes available; and

c) Data access and clearance of protocol possible within timelines of the POC feasibility study.

d) Comparison of the rate of events against country specific benchmarks

The list below provides an overview of potential databases, based on initial assessment of population sizes; this list will be updated with new associate partners in the consortium.

Based on the meta-data, fingerprint data, and discussions and information from the databases, parameters in table 5 will be collected and described. This will be the basis for a decision whether the databases will be eligible for the next step (estimation), eligibility may differ for the different POC protocols.

Table 4: Characteristics of the Databases of ADVANCE Partners

Datasource Country_name	Coverage, Region	Type of data	Years covered	Size (N persons)	Outpatient diagnoses	Inpatient diagnoses	Vaccines general	Prescribed/ dispensed drugs
BE_network of sentinel GPs	National	Surveillance network (paper based)	Since 1979	Based on patient contacts	Some (surveillance of specific diseases)	Specific diseases	No	No
BE_Pedisurv	National	Pediatric surveillance network of specific rare disease	Since 2002	National case based	Specific diseases	Specific diseases	no	No
DK_SSI	National	Record linkage	1996 - 2014	7.5 million	Yes (ICD-10)	Yes (ICD-10)	yes	yes
DK-AUH	Regional (Aarhus)	Record linkage	2004 - 2013	1.7 million	Yes (ICD-10)	Yes (ICD-10)	partial	yes
ES_BIFAP	National	GP	2002 - 2013	4.8 million	Yes (ICPC+free text)	yes (as text)	No (only influenza)	yes
ES_FISABIO	Regional (Valencia)	Record linkage	Since 2005	5 million	Yes (ICD-9)	Yes (ICD)	yes	yes
ES_IDIAP	Regional (Cataluña)	GP	Since 2005	5.8 million	Yes (ICD-9)	Yes	yes	yes
FI_HPVCHRT	Trial participants	HPV RCTs+ extension through record linkage	Prospective since 2007	Around 20,000	Yes (ICD-10)	Yes (ICD-10)	yes	yes
IT_ARS	Regional (Tuscany)	Record linkage			no	Yes (ICD-9)	Not yet	yes
IT_ASLCR	Regional (Cremona)	Record linkage	2002 - 2013	454,188	No	Yes (ICD9)	Yes	yes
IT_Arianna	Regional (Casserta)	GP + record linkage	Since 2000	1.1 million	ICD9	ICD9	Not yet	yes
IT_PEDIANET	Regional (Veneto)	Family pediatricians	2004 - 2014	77,021	Yes	Yes	yes	yes
NL_IPCI	National	GP linked to RIVM vaccine registries	1996 - 2014	1.8 million	Yes (ICPC)	Yes (from letters/ specialist)	no	yes
NL_RIVM	National	Case surveillance of infectious disease	?	16 million base population	Some (surveillance of specific diseases)	Specific diseases	no	No
SE_KI	National	Record linkage	1998 - 2010	9.4 million	Yes (ICD-10)	Yes (ICD-10)	partial	yes
UK_RCGP	National	GP	2003 - 2014	2.0 million	Yes (READ)	Yes (READ)	yes	yes

UK_THIN	National	GP	1996-2013	8.3 million	Yes (READ)	Yes (READ)	yes	yes

Table 5 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) • exiting from database (end of follow-up < last data availability for practice	N Median (5^{th}, 95^{th} percentile of lag time from date of delivery till	Percentage of total	Jerboa Event fingerprint

		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 studies in the database		Event team & database experience
	Estimates of frequency of the event in the population	Available estimates with		Event team & database

	represented by the database according to external data sources (e.g. literature)	source (and comments)		experience

7.2.4. Source Population

The source population in each of the eligible databases will be the pediatric population from birth to age six years or when receiving the booster dose at school entry that is registered in the participating databases.

7.2.5. Population Selection

The study population will comprise all children registered in any of the participating databases during the study period and for whom an adequate start and end of follow-up and date of birth can be defined. Children will be followed from start of the study period, one month after date of birth (i.e. to allow for pre-vaccination time for the SCCS design and to avoid pre-term related or birth-induced increase in incidence rates), or date of valid data in the database (whichever is the latest) until the end of study period (31-12-2015, the school-entry pertussis booster, transferring out of the database, death, reaching age 6 years: whichever is the earliest). Persons with missing dates of birth or sex will not be included.

For the incidence rate part of the study, we will retain all subjects meeting age and calendar time inclusion criteria with person time greater than zero.

The vaccination day and outcome-specific risk-windows after vaccination will be classified as risk periods, and the period between the end of the risk-window after vaccination until the next pertussis-containing vaccination or end of follow-up as the baseline period. To keep track of dose these periods will also be labeled by the number of previous pertussis-containing vaccinations. For all eligible subjects, CohortStart will be equal to PopulationStart and CohortEnd equal to PopulationEnd.

Databases will be asked to provide dates during which both wP and aP pertussis vaccines were both being used in the population. Patient time during the period of the switch from wP to aP will be removed from the cohort and SCCS analyses to avoid misclassification of exposures.

In order to avoid confounding by infections or other vaccinations, the study population for the incidence rate cohort design will NOT be the same as for the SCCS design.

For the self-controlled case series, analysis sets containing cases of each of the outcomes under study will be defined. For each of these cases, exposure will be defined as: exposed to wP and exposed to aP.

The study population for each outcome-specific SCCS analysis will be those subjects who experienced the event at least once during follow-up. Eligible person time for these subjects will be one month before the first recorded pertussis vaccine exposure until one month after the last pertussis vaccine exposure. Risk periods will be those specified for each outcome after each dose. Baseline periods will be days -31 to -8 before exposure and the interval from the last day of the risk window + 1 to day 31.

7.2.6. Study Period

The study period is from 01 January 1990 to 31 December 2015, and also be dependent on availability of data from specific databases.

7.2.7. Variables

All study variables will be obtained from the participating electronic health care databases. These will be transformed their local datasets into a pre-defined ADVANCE common data model (CDM) to allow for the running of common scripts across all the databases.

7.2.7.1. Exposure of Interest, Operationalization and Validation

EXPOSURE OF INTEREST

The exposure of interest in this study will be all types of pertussis-containing vaccines that were available during the study period and used in the participating study population in the participating countries.

OPERATIONALIZATION

Vaccinations will be obtained from the databases by using names of vaccines and database specific codes. Vaccines will be categorized as wP, aP or unknown. There will be no further stratification in brand or product name (see Appendix 4 for categorization).

EXPOSURE WINDOWS

Person time of follow up in the cohort will be categorized as risk and baseline periods. Persons will be followed up over the whole period independent of occurrence of an event or vaccination. Person time will not be stopped at the occurrence of an event or vaccination. The vaccination day and outcome-specific risk-windows after vaccination will be classified as risk periods, and the period between the end of the risk-window after vaccination until the next pertussis-containing vaccination or end of follow-up as the baseline period. To keep track of dose, these periods will also be labelled by the number of previous pertussis-containing vaccinations.

Outcome	***Risk Window After Vaccination***
Injection site reactions	0-7 days
Fever	0-72 hours
Somnolence	0-48 hours
Persistent crying, irritability	0-24 hours
Generalized convulsions	0-72 hours
HHE	0-48 hours
Extensive limb swelling	0-7 days

VALIDATION

Information on the quality of recording of vaccinations will be obtained from the ADVANCE fingerprint (quality assessment) data by comparing coverage estimates from the databases with WHO/national coverage rates in the same age range.

7.2.7.2. Outcomes, Operationalization and Validation

OUTCOMES

The safety outcomes of interest are known reactions:

- Injection site reactions: erythema, edema, induration/ nodule/sterile abscess, pain/tenderness
- Fever
- Somnolence

- Persistent crying, irritability
- Generalized convulsions
- HHE
- Extensive limb swelling

A definition for somnolence is provided in Appendix 3. The Brighton Collaboration has case definitions available for injection site reaction [18], fever [19], persistent crying [20], generalized convulsions [21], HHE [22], and swelling (i.e. swelling at or near injection site as an increase in size or volume at the injection site that may extend to the entire limb according to severity) [23].

Events will be extracted from the electronic health care databases as diagnosis codes or text or by means of proxies. Mapping to ICD-9/10, READ and ICPC codes will be done semi-automatically using the ADVANCE code mapper. Algorithms for identification will be created with the databases and benchmarked against external sources and within consortium as described in section 7.1.

VALIDATION

Verification of the extracted codes and proxies for extraction of events will be done as part of the quality assessment and harmonization process (fingerprinting) prior to the start of the POC feasibility study. The verification will be done by comparing age-specific incidence rates between databases and comparison against external sources such as literature or other publicly available data. If possible, positive predictive values and sensitivity from prior studies will be retrieved. No chart validation at the patient level is foreseen for this POC feasibility study.

7.2.7.3. Other Variables and Operationalizations

The incidence rates will be stratified by age, sex, calendar time and country which will also allow for rolling back the benefit-risk model to look at the impact of sequential monitoring.

7.2.8. Data Analysis

7.2.8.1. Statistical Hypothesis

The purpose of the current POC feasibility study is to provide incidence rates of safety events observed after administration of wP and aP vaccines. There is no null hypothesis being tested in this study.

Since there is no formal hypothesis testing, there is no need for power and sample size calculation. The databases currently involved in ADVANCE capture 34 million persons, the population aged 0-6 may be between 5-8 million, even if event rates are low, enough cases should occur to estimate incidence rates. If stratification leads to zero numerators, higher level aggregation will be conducted.

7.2.8.2. Analysis of Demographics and Baseline Characteristics

The population of subjects meeting the study population (i.e. cohort) inclusion criteria will be described in terms of the following variables. Continuous variables will be described as means, medians (i.e. follow-up time) and standard deviations, while categorical variables will be described as frequencies and percentages:

- age and sex at study population (i.e. cohort) entry, age at D1, D2, D3, D4, D5 by country, follow-up time in the cohort by calendar year and month, dose and age (year of birth), number of pertussis vaccine doses per child.

7.2.8.3. Statistical Methods

COHORT STUDY

The incidence rate of each event will be calculated using a dynamic cohort approach. For each person in the cohort, follow-up time will be classified by calendar year, sex, age in year of birth and months and the different risk window (see exposure section). This persontime will be the denominator for the incidence rate calculations. Incidence rates will be calculated stratified by:

1. type of vaccine: wP/aP
2. database, country
2. dose of wP/aP (actually received dose)
3. risk window
4. age in months and year of birth
5. calendar year

Cases will be counted in the numerator for the rate calculation if the date of onset/occurrence of an event falls into the specific window of that vaccine, and will be further stratified to count in the correct age/sex and calendar year category at the date of onset.

A Poisson distribution will be used to calculate 95% CIs.

SCCS

A pre-vaccine baseline risk period will be applied to control for the healthy-vaccinee effect. The SCCS will include age in days, and season (classified from the month) as time-varying covariates, as well as all available doses of pertussis vaccine as exposures. Conditional Poisson regression analyses will be conducted for the estimations.

10 describes the availability of wP information of two different databases (i.e. in database [Db] A, wP information is available and in and Db B it is not) and the approach how IR within the risk period will be calculated.

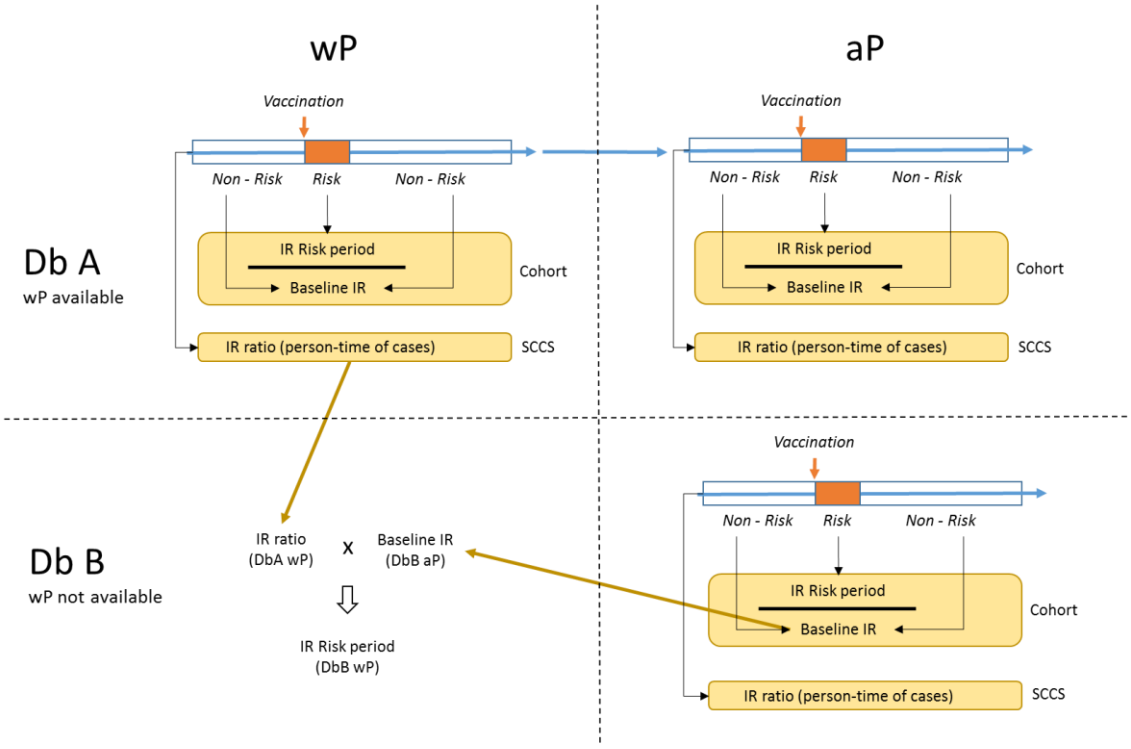


Figure 4: Approach How to Calculate Risk Window Specfic IR in Databases where wP Information is Either Available or Lacking

7.2.9. Limitations

In this POC feasibility study that tests and evaluates the ADVANCE system, tailored to generate rapid evidence on vaccine effects, there are several methodological limitations. These limitations will be assessed, and are an integral part of the system evaluation process. The following points are considered to be the expected limitations:

- Given that this POC feasibility study evaluates a system and its different processes, the epidemiological results produced (i.e. incidence rates and risk ratios) should not be considered as evidence that can be directly translated to inform epidemiological and pharmacovigilance decision-making processes. The epidemiological results produced in this POC feasibility study are part of a set of evaluation indicators to test the ADVANCE system and may not be the best or true estimates
- It could be possible that within the process of database selection it will become evident that proposed case definition(s) for the events are either different, or are differently interpreted, in a specific database or country. Different countries and databases could be using different case definitions. To address this limitation, the study will produce country-and database-specific analyses and results.
- Some event outcomes are signs and symptoms and not disease entities. The coding in databases might not be applicable to signs and symptoms (e.g. crying). Therefore, database extraction algorithms and strategies need to be adapted (e.g. include also free text search).
- Event codes in databases might contain the whole spectrum of potential, probable and confirmed cases. Given the POC feasibility study nature to test the system and to save costs and time, diagnostic codes will not be validated on a patient level using medical histories and charts.

- It is expected that only a limited number of databases will have information on wP vaccination available. Therefore, the suggested strategies in the methods sections 7.4.3 (i.e. multiplying baseline incidence rates with IR ratios across databases, time periods, and wP/ aP data) will be applied. The additional use of literature information will be carefully evaluated and interpreted.
- The type of database used (e.g. GP database, hospital database, claim database, etc) will influence incidence rates. To address this limitation, the study will produce country/ database-specific analyses and results.
- The current methodology will not account for differences in vaccination schedules: different countries use different vaccines and schedules that may have changed over time. ATC codes will be used to identify which types of vaccines were used; this will allow stratification on vaccine types. Further, risks on schedule level will not be evaluated but will be evaluated on a dose level as each case is followed after each dose given.
- Misclassification or missing information on vaccinations and events is a common limitation of observational studies conducted in electronic healthcare databases. Missingness and misclassification will be carefully followed and attempts will be made to implement data cleaning procedures (e.g. probability of wP or aP vaccine given the time of administration, dose according to timing in vaccination schedule, etc). Records with missing dates (i.e. birthdates, data of vaccination, date of event occurring) will be excluded. ATC code is a mandatory data field. A missing dose information will be derived (i.e., dose will be 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 purpose of the study).
- aP and wP vaccine administration could have been overlapping. The period of potential overlap in a database will be excluded if aP or wP vaccine type information is missing (i.e. database entry is only referring to 'pertussis vaccine' and aP or wP is not specified).
- The smallest possible time unit in electronic healthcare databases is 'days'. This could potentially make it difficult to judge whether vaccination happened before the event or vice versa when recorded on the same day. In this study, it is assumed a priori that vaccination occurs before the event whenever they are recorded on the same day, so the vaccination day is classified as risk period and starting at Day 0.

7.2.10. Study Size

There is no target sample size for this study. The sample size is determined by the study population in the ADVANCE databases, which currently accumulates to more than 34 million subjects and 314 million person-years. The actual size available to estimate incidence rates of safety outcomes after vaccination may diminish if databases do not have good quality data on the vaccinations, which is part of the study assessment.

7.2.11. Data Management

7.2.11.1. Data Processing

This section is taken directly from POC feasibility study outline document.

Processing of data from the different databases will be done in two steps as per ADVANCE policy:

Extraction of study specific de-identified data from the original databases into study specific common input files. This will be done according to the specifications in the POC feasibility study protocol. There is currently no common IT component supporting this, except from the tools that data processors use regularly on their local data.

Transformation of the study specific data into analytical datasets suitable for statistical analysis. This will be done according to the specifications in the POC feasibility study protocol with a common script.

7.2.11.2. Data Extraction

Following approval of the study protocols, data processors locally will be asked to extract study-specific data into a simple CDM. The data in this CDM could be used by the POC feasibility study teams on coverage, safety and benefit. Before it can be used, the data will be harmonized and checked under quality control procedures; this will be done in the fingerprinting in Work Package 5 (WP5). A description of the required data is needed from the POC feasibility study teams so one specification can be made for the data owners.

7.2.11.3. Data Transformation

Data transformation is the step from having data locally in the CDM to creation of analytical datasets locally that can be shared for further analysis on the remote research environment (RRE). It is important that the analytical datasets are 'stripped' from variables that may create 'identification' issues, e.g. dates.

A central scripting approach will be used for the data transformation steps from CDM files to analytical datasets. This will be done by the statisticians in the POC feasibility study team, using R and SAS. The SAS and R scripts will be compared and serve as double-coded scripts for quality assurance.

7.2.12. Software and Hardware

Data analysis will be conducted by the statisticians on the remote research environment (RRE) that is called OCTOPUS. This will be done as outlined in Section 7.2.8 and specified in the statistical analysis plan. The RRE has R, SAS, and other programs.

The OCTOPUS RRE is a framework that has already proven its value in various European Commission and European Medicines Agency funded projects. It stimulates geographically dispersed research groups to collaborate and has resulted in consortia that were engaged in all the phases of drug safety research.

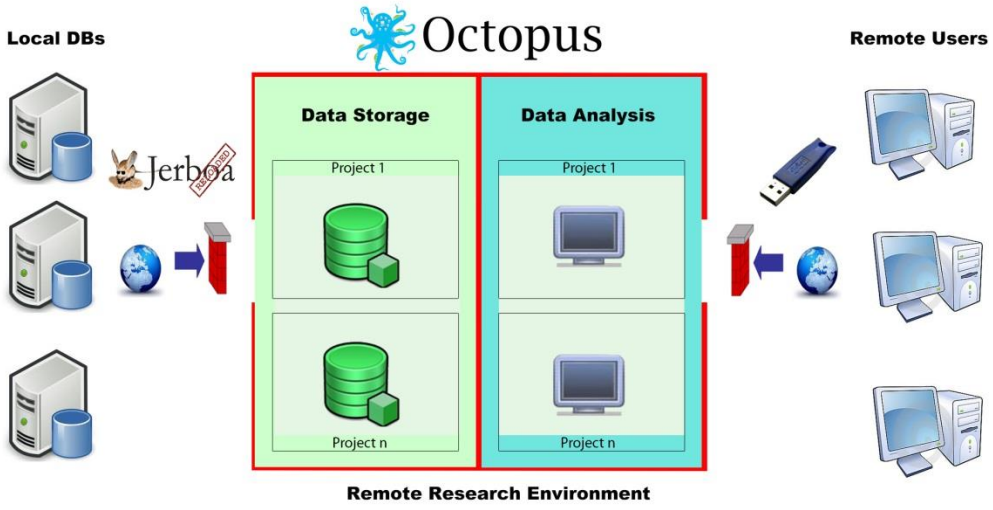


Figure 5: Schematic Representation of the OCTOPUS Remote Research Environment

OCTOPUS ARCHITECTURE

Octopus is hosted on an application server (Windows Server 2008R2) located in the data center of the Erasmus University Medical Center (EMC). The data center is a Tier level III data center which means it has multiple independent distribution paths serving the IT equipment and has an expected availability of 99.9%. The server is secured by the EMC firewall and will not have any direct connections to the LAN of the hosting institute. Access to application server is only allowed from a restricted set of IP addresses using two-factor authentication with a password and token. The infrastructure is monitored by the Erasmus MC Computer Emergency Response Team (CERT).

Procedures have been developed to ensure data protection and secure file transfer from and to the collaborating partners. The following paragraphs describe these procedures in more detail.

DATA SECURITY PROCEDURE

For the OCTOPUS infrastructure many policies for data security have been put in place, for example:

- To acquire access to the RRE, each user has to fill in a request form and sign a confidentiality agreement. WP5 leaders (or Steering Committee) need to formally approve each request.
- Users will only have access to the RRE using a remote desktop session.
- Authentication of users consists of two factors: in addition to the basic authentication procedure (with username and password), an authentication with a personal token is performed (SafeNet eToken Pro, www.safenet-inc.com).
- All log on/log off operations are automatically logged (registered) by the RRE.
- The authentication of users is performed by asking, at each login attempt, the username and password (i.e. saved credentials are not allowed).
- Users only gain access to folders/files that are part of the project in which they collaborate. The system administrators can grant permissions to users based on their role in the project.
- Users will not have access to the control panel, internet, and administrative tools.
- Any attempt to copy and paste files between the remote session and local PCs of partners will be disabled.
- All devices on local PCs of partners (i.e. printers, storage...) will be disabled in the remote session.
- A complete log of all requests for files and copies of these files sent outside the RRE will be kept and can be inspected upon request.
- A screensaver will be activated on the remote desktop if the user is not active for a predefined time interval.

Any misconduct or violation of RRE security principles will be notified to the data manager and project manager immediately. Standard operating procedures for access and file transfer rights will be developed within the consortium.

DATA TRANSFER PROCEDURE

The procedure is illustrated in Figure 5. The user will upload new data, e.g. Jerboa encrypted file, to a personal upload directory using the secure sftp protocol in FileZilla (step 1,2), and after approval, the

administrator will decrypt the file and will import the data in the data folder of the project (step 3). The administrator will confirm the dataset preparation and the user can view and work on the data using the token (step 4,5). To avoid data manipulation, the data folder is read-only for all users. Upload of other files, e.g. SAS scripts, will follow the same procedure.

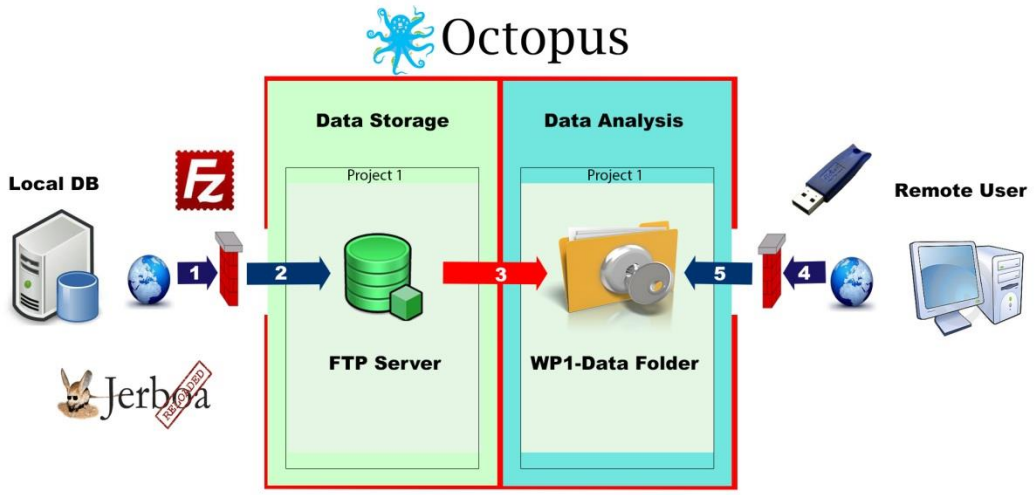


Figure 6: Upload of Data to OCTOPUS

To download results from the RRE (see figure 12), e.g. a PowerPoint presentation, the user has to login to the RRE (step 1) and place these files in a personal export directory (step 2). In order to receive those files, a request must be sent to the RRE system administrator having the corresponding WP leader in carbon copy (cc). The system administrator will verify that the files do not contain any restricted data, and then will put the files in the download directory of the user (step 3). Subsequently, the user can download the files from the server using the FileZilla sftp client (step 4,5).

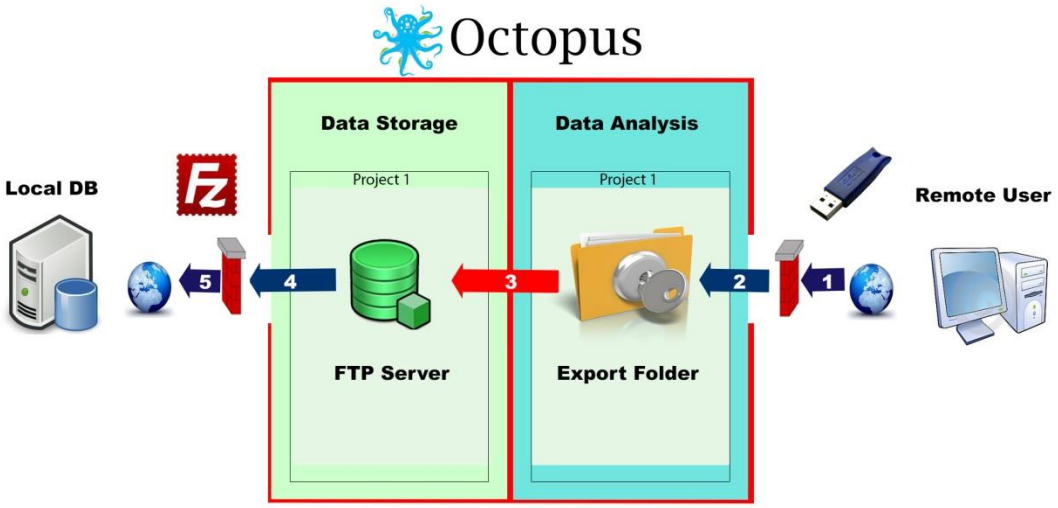


Figure 7:Download of Data from the RRE

DATA ANALYSIS PROCEDURE

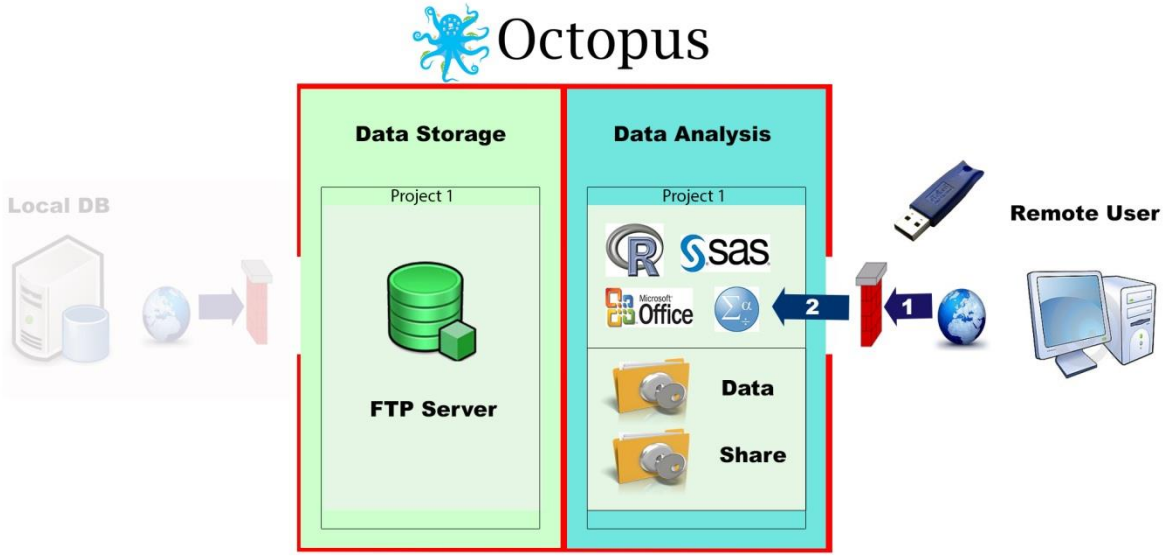


Figure 8: Data Analysis in the RRE

Data analysis will be done inside the RRE, i.e. the user logins in the server (step 1) and has access to a number of analysis and word processing tools (step 2). All users in the project or WP have access to the data folders (read only) or can share files with others using their personal Share. In the Share, only the owner has write-access; all others only have read permissions. The folder structures can easily be customized to address specific project needs.

7.2.13. Quality Control

7.2.13.1. 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 should be retained for a longer period, however, if required by the applicable regulatory requirements or by an agreement between study partners. It is the responsibility of the principal investigator to inform the other investigators/institutions as to when these documents no longer need to be retained.

These principles of record retention will also be applied to the storage of laboratory samples, if applicable, providing that the integrity of the stored sample permits testing.

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

7.2.14. Advisory Committee

The ADVANCE Scientific Advisory Board

7.3. Use of the data generated in this study

The data generated in this study will be used into a B/R model, as describe in the B/R analysis protocol. The data will also be used for further test of statistical methods within the ADVANCE WP4 working groups.

7.3.1. Benefit/risk analysis

This POC study will yield incidence rates (during exposure window and after exposure window) plus relative risk estimates that will be used in the B/R analysis.

Table 6:Evidence Required for Safety Outcomes to Build the Effects Table by Country, Year of Birth and for Each Outcome in the B/R analysis

Incidence in risk period after wP or aP vaccination						
	Dose 1*		Dose 2		Dose 3	
Time since dose**	Inc	95% CI	Inc	95% CI	Inc	95% CI
Risk window 1						
Risk window 2						
Risk window 3						
Baseline incidence						
2-3 months						
3-4 months						
4-5 months						
…						

* number of doses might depend on the country and year
** risk windows might be different depending on the safety outcome

In addition, the relative risk estimate from the SCCS will be used as described in the analysis

7.3.2 Re-use of data from the POC study for methods development

The table below details how the data generated in the feasibility step and the scientific step will be re-used /produce in the methods development proposals.

	Data from fingerprint for these protocols	**Data from Rate and risk estimations in POC**
Methods development topics research topic		
Burden of adverse event	Disease rates of events, quality of databases	Disease rates of events
Effectiveness	Rates of disease, PPV, quality of databases	Differentiality
Monitoring of B/R	Lag times to get data	Monthly rates of events, coverage, outcomes
Heterogeneity	Population, event, vaccine misclassification	None
Ontology	Vaccine	Will provide information to POC
Coverage	Vaccine & dates distributions	Will provide information to POC
Codemapper	Event	Will be used in POC

8. PROTECTION OF HUMAN SUBJECTS

8.1. Regulatory and Ethical Compliance

The European legislation describes obligations to be fulfilled by marketing authorisation holders (MAHs) and national competent authorities for medicinal products (including vaccines) authorised in the European Union (EU). The European legislation does not apply to post-authorisation studies conducted by organisations such as academia, medical research charities or research organisations in the public sector. These organisations should follow local requirements defined in the national legislation applicable in the countries where research is conducted. In the context of the ADVANCE consortium, proof-of-concept studies (POCs) will be conducted to test new approaches (data sources, methods) by using a test cases; within this framework the POC studies will collect and use data on vaccines authorised 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 authorisation. They should therefore be considered as having a control 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 quality criteria for inclusion in the study

- To test integration of incidence rates of specific safety events into the benefit-risk analysis.

The objectives of this study are on methodological aspects 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/wP). 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 summarised in the final study report.

8.2. Informed Consent

Data bases with an internal review board approval indicating that informed consent is waived and the rational for this decision will be included in the analyses.

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

The protocol and waiver of informed consent must 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 must 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 must 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 cannot be implemented. All significant protocol deviations will be recorded and reported in the Study Report. Specifically, observational reportable Protocol Deviations are those Protocol Deviations 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 regulations.

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 regulations.

10.2. Publications

Further to legislated data disclosure, the results of this study will be published as scientific papers in peer-reviewed journals. Preparation of such manuscripts will be prepared independently by the investigators and in accordance with the current guidelines of STrengthening the Reporting of OBservational studies in Epidemiology (STROBE). 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. REFERENCES

1. Pittet LF, Emonet S, Francois P, et al. Diagnosis of whooping cough in Switzerland: differentiating Bordetella pertussis from Bordetella holmesii by polymerase chain reaction. PLoS ONE 2014;9:e88936.
2. Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to Bordetella pertussis and other Bordetella subspecies. Clin Microbiol Rev 2005;18:326-82.
3. Waters V, Halperin BA. Bordetella pertussis. Mandell, GL, Bennett JE, Dolin R, editors. In: Principles and Practice of Infectious Diseases, Philadelphia: Churchill Livingstone Elsevier; 2010. p. 2955-2964.
4. Vaccines, 6th edn. Plotkin S, Orenstein WA, Offit PA. Philadelphia: Elsevier Saunders; 2012.
5. Cortese MM, Baughman AL, Zhang R, Srivastava PU, Wallace GS. Pertussis hospitalizations among infants in the United States, 1993 to 2004. Pediatrics 2008;121:484-92.
6. Cherry JD, Tan T, Wirsing von Konig CH, et al. Clinical definitions of pertussis: Summary of a Global Pertussis Initiative roundtable meeting, February 2011. Clin Infect Dis 2012;54:1756-64.
7. Plotkin SA, Plotkin SL. The development of vaccines: how the past led to the future. Nat Rev Microbiol 2011;9:889-93.
8. Sato Y, Kimura M, Fukumi H. Development of a pertussis component vaccine in Japan. Lancet 1984;1:122-6.
9. Noble GR, Bernier RH, Esber EC, et al. Acellular and whole-cell pertussis vaccines in Japan. Report of a visit by US scientists. JAMA 1987;257:1351-6.
10. Decker MD, Edwards KM, Steinhoff MC, et al. Comparison of 13 acellular pertussis vaccines: adverse reactions. Pediatrics 1995;96:557-66.
11. Ciofi degli Atti ML, Olin P. Severe adverse events in the Italian and Stockholm I pertussis vaccine clinical trials. Dev Biol Stand 1997;89:77-81.
12. David S, Vermeer-de Bondt PE, van der Maas NA. Reactogenicity of infant whole cell pertussis combination vaccine compared with acellular pertussis vaccines with or without simultaneous pneumococcal vaccine in the Netherlands. Vaccine 2008;26:5883-7.
13. Jefferson T, Rudin M, DiPietrantonj C. Systematic review of the effects of pertussis vaccines in children. Vaccine 2003;21:2003-14.
14. Byers RK, Moll FC. Encephalopathies following prophylactic pertussis vaccine. Pediatrics 1948;1:437-57.
15. Cherry J, Heininger U. Pertussis and Other *Bordetella* Infections. Feigin, R, Cherry J, Demmler-Harrison G, Kaplan S, editors. In: Textbook of Pediatric Infectious Diseases, 6th edn, Philadelphia: WB Saunders; 2009. p. 1683-1706.
16. Miller D, Wadsworth J, Diamond J, Ross E. Pertussis vaccine and whooping cough as risk factors in acute neurological illness and death in young children. Dev Biol Stand 1985;61:389-94.
17. Bonhoeffer J, Siegrist CA, Heath PT. Immunisation of premature infants. Arch Dis Child 2006;91:929-35.
18. Gidudu J, Kohl KS, Halperin S, et al. A local reaction at or near injection site: case definition and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine 2008;26:6800-13.
19. Marcy SM, Kohl KS, Dagan R, et al. Fever as an adverse event following immunization: case definition and guidelines of data collection, analysis, and presentation. Vaccine 2004;22:551-6.

20. Bonhoeffer J, Vermeer P, Halperin S, et al. Persistent crying in infants and children as an adverse event following immunization: case definition and guidelines for data collection, analysis, and presentation. Vaccine 2004;22:586-91.

21. Bonhoeffer J, Menkes J, Gold MS, et al. Generalized convulsive seizure as an adverse event following immunization: case definition and guidelines for data collection, analysis, and presentation. Vaccine 2004;22:557-62.

22. Buettcher M, Heininger U, Braun M, et al. Hypotonic-hyporesponsive episode (HHE) as an adverse event following immunization in early childhood: case definition and guidelines for data collection, analysis, and presentation. Vaccine 2007;25:5875-81.

23. Kohl KS, Walop W, Gidudu J, et al. Swelling at or near injection site: case definition and guidelines for collection, analysis and presentation of immunization safety data. Vaccine 2007;25:5858-74.

APPENDIX 1: FEASIBILITY TESTING (FINGERPRINTING)

Step 1: feasibility assessment of databases

The concept of bringing data together within and across countries with the purpose of addressing vaccine benefit/risk questions in a collaborative and integrated approach can be addressed in several ways with respect to:

1) Standardization of protocols to conduct studies on multiple data sources
2) Local data extraction
3) Transformation of the data into analytical datasets
4) Pooled analyses of data

In ADVANCE steps 1, 3 and 4 are harmonized and centrally coordinated.

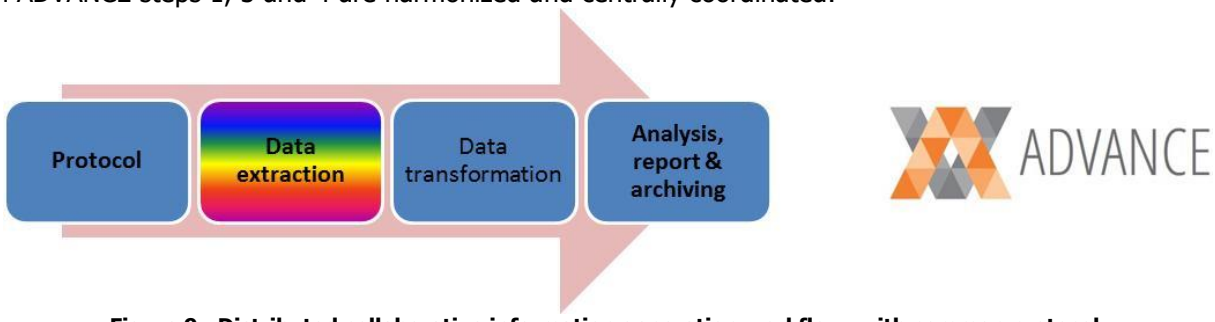


Figure 9 : Distributed collaborative information generation workflow, with common protocol, standardized transformation and shared analyses while data extraction and original data remain local.

Step 2 cannot be harmonized for the following reasons:

1) Different structures of health care systems across EU member states
2) Different types of databases within a country and across EU member states (i.e. health care databases, claims databases, inpatient databases, surveillance networks, laboratory data, microbiology data, vaccination registries, medical record databases), if possible all these databases will be fingerprinted
3) Different content of similar types of databases across EU member states
4) Different coding/terminologies and language of similar information between databases in different EU member states

To gain insight into the underlying determinants and mechanisms of data generation, and to address these differences in a consistent and informed way, such that we can actually use the data for the purpose of vaccine benefit/risk monitoring we will use the following approaches:

1) Use of local source data knowledge: Full involvement of the database custodian in data extraction processes and interpretation of the data to appreciate differences, and filling out the survey on the database characteristics (AIRR survey) as well as database experience forms
2) Semantic harmonization: mapping of terminologies and variables for population, events (outcomes and covariates), vaccines and drugs & creation of ontologies and mappings of codes and terms to allow for specific data to be integrated into a common data model
3) Fingerprinting: (i.e characterizing) of what data is actually available in the databases by real data extraction (transparency)
    a. Stepwise conversion of specific required study data into a simple common data model
    b. Describing the data quantitatively using a common script and visualization

c. Iterative harmonization and verification of data extraction steps under item 2 across the databases
d. Benchmarking of data extracted against available external sources of information.

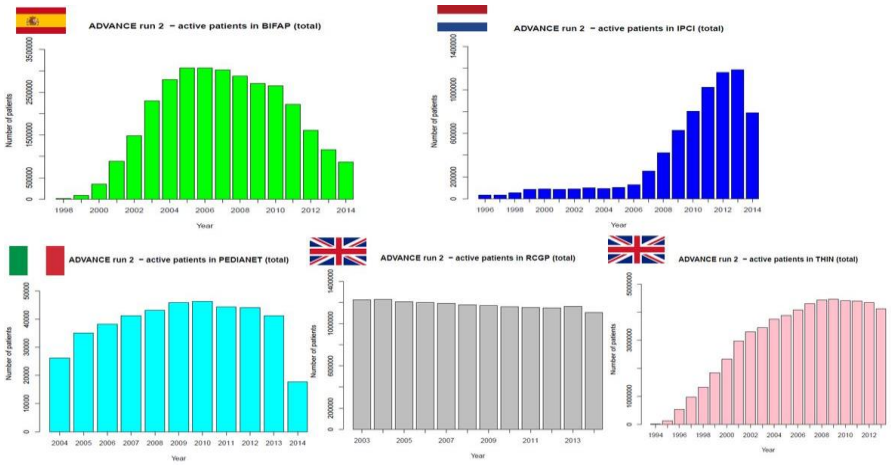
4) Knowledge & information management: Reporting of generated evidence and knowledge and making it available and accessible

In the fingerprinting task which informs the feasibility assessment, we describe the databases based on the data that are locally extracted into the ADVANCE common data model. Database custodians will have to transform their local data into these common input files. These input files will be processed locally by a common tool that is either R, Jerboa or SAS and these scripts will generate aggregated fingerprinting data that will allow us to assess the quality of the database for specific vaccines/events. The fingerprinting is not the responsibility of the POC PI but of the WP 5 leaders.

Population fingerprint

Based on the common input files that have been agreed in ADVANCE, Jerboa generates standard statistics for the population per gender, calendar year and age group.

Outputs of this fingerprint allow for assessment of the representativeness of the population and many other features. An example from the D5.2 is provided below.



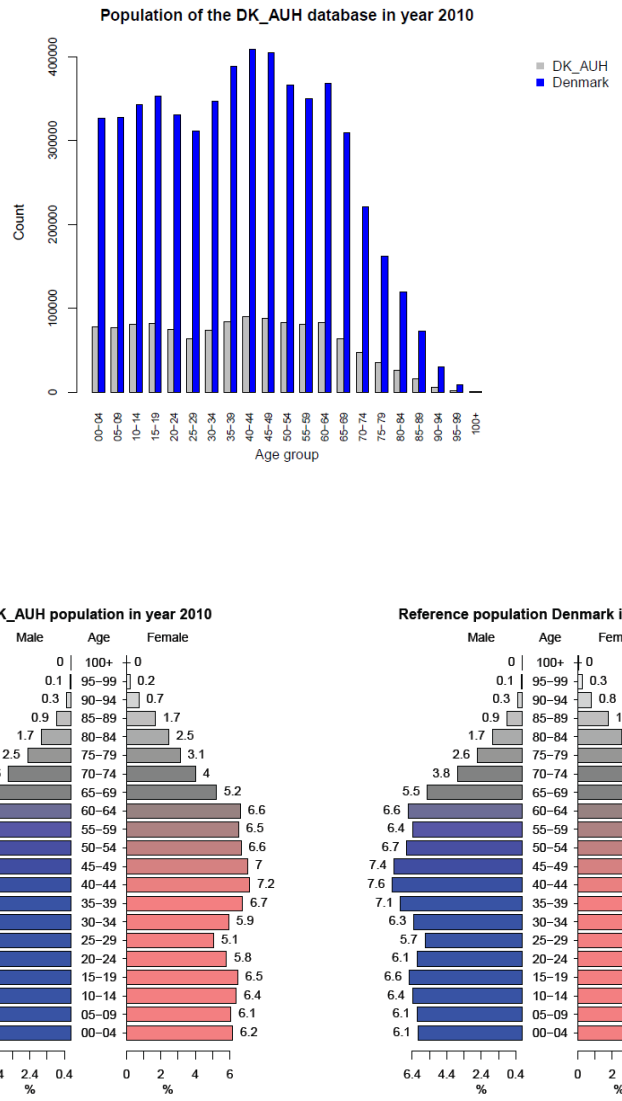


Figure 10: Example of population fingerprint output

Event fingerprinting

The overall aim of the event fingerprinting workflow (or also called data derivation workflow) is to obtain for each data source the best algorithm to extract an event/condition, and document this in a structured fashion. The full data extraction algorithm is a logical combination (AND, OR, or AND NOT) of components that could contain for example the following information:

- a diagnosis, recorded in a primary, secondary, inpatient care or other settings
- diagnostic evidence, for example laboratory measurements
- utilization of healthcare services specifically indicated to diagnose or treat that condition, such as a drug, a diagnostic test, a procedure or other health service

For instance, for some data sources a diagnostic code could be available for pertussis. In other data sources, results from blood tests may be available or surveillance data. In Figure 5 a high level graphical representation of the workflow is presented. In short the following questions are answered in each of the steps:

- Definition: How do we define the specific condition and its context?
- Collect experience: collection of how these conditions have been collected in the past. How can we leverage the valuable domain knowledge of the data custodians?
- Literature Search: what is the incidence/prevalence of this conditions in the countries of interest, this information will serve as an external benchmark to see whether the data retrieved have external validity
- Terminology Mapping: How can we translate the case definition into different terminology systems?
- Component algorithms: collection of information on how each database extracts the data which algorithms are used? Which is the list of unique component algorithms that each database should be invited to extract?
- Results Analysis: comparison of incidence rates between databases and with literature, and if possible component analysis, what is the influence of certain components on the incidence/prevalence?
- Final Decision: what is the best algorithm for each database, and archive this?

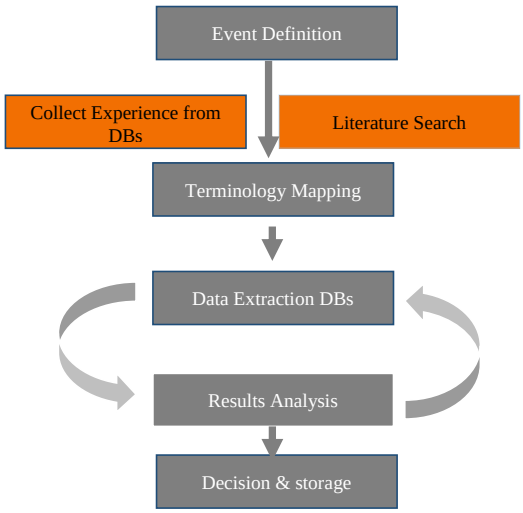


Figure 11: Workflow fingerprinting of events and component analysis

The event fingerprinting is led by an event team comprising the event team leader, the person who defined the event, the PIs of the study teams that need the event and the CODEMAPPER developer.

Event definition forms

Event definition forms are filled for each event. These forms contain the following information:

- Event definition & validity classifications
- Synonyms / lay terms used
- Laboratory tests done specific for event
- Diagnostic tests done specific for event
- Drugs used specific for event treatment
- Procedures used specific for event treatment
- References
- Codes (ICD-9 codes, or ICD 10 codes)

Event routing questionnaire

A survey will be launched to physicians to understand the pattern of care for each of these events of interest. This will inform us which components are important in algorithm constructions.

Database experience sheet

Data base experts will be asked to fill an excel sheet that will provide information on how they have extracted the events in the past. The following information is collected:

- experience with extraction of this event: yes/no/other
- Extraction algorithms used (ie was a logic applied in terms of at least two codes, prescription)
- Was validation performed?
- Do you have papers on the validation of this event in your database

Literature

In order to validate the extractions against an external benchmark, literature on the incidence of the event will be searched by the eventteam for the country of interest and if this is absent from other countries. For ADVANCE we will need to search of Spain, Italy, Netherlands, UK, Sweden, Finland and Denmark.

Terminology mapping

Terminology mapping is done with the ADVANCE Codemapper . The ADVANCE Codemapper is mapping the codes of different coding systems to concepts and terminology found in the clinical definition forms according to the flow described in figure 5.

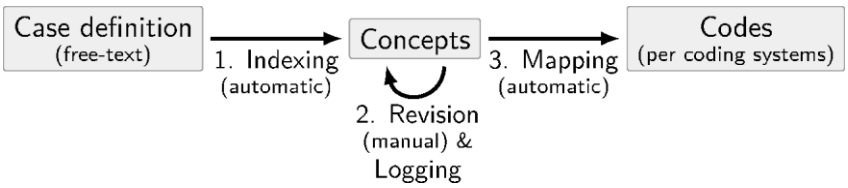


Figure 12: workflow of the ADVANCE Codemapper

The workflow of CodeMapper has three phases as shown in figure 6 which comprise multiple components. It starts from the case definition forms that are copied in the Codemapper. The Codemapper calls the Peregrine text indexing engine to identify medical concepts in the case definition. These are highlighted in the copied text. Concepts that belong to the semantic group of disorders in UMLS are pre-selected. After this automated selection the user can walk up and down the ontology to select further specify individually which concepts should be additionally included/excluded for further processing. Second, these concepts are related to concepts of the Unified Medical Language System (UMLS). The UMLS is then used to automatically retrieve codes that correspond to the selected UMLS concepts in a set of user defined coding systems. The concepts are displayed in a table alongside the associated codes in each coding system. Third, the user can revise codes of the mapping by applying concept-level operations. The user can add concepts, remove concepts, and retrieve more general or more specific concepts according to the hierarchical information in the UMLS. The set of targeted coding systems can be changed on-the-fly. After every operation, the code sets associated with the concepts are automatically updated. Feedback about the mapping is captured in comments that can be attached to the concepts.

Every operation is recorded in a history for later traceability. When saving her or his work, the user has a machine generated to summary of the modifications. The summary is added to the history. The mapping, comprised of the concepts and code sets is stored online together with the case definition, the initial mapping and history. All data can be downloaded as an Excel document to support incorporation into extraction scripts.

Availability: The CodeMapper application is freely available for non-commercial projects at https://euadr.erasmusmc.nl/CoMap.

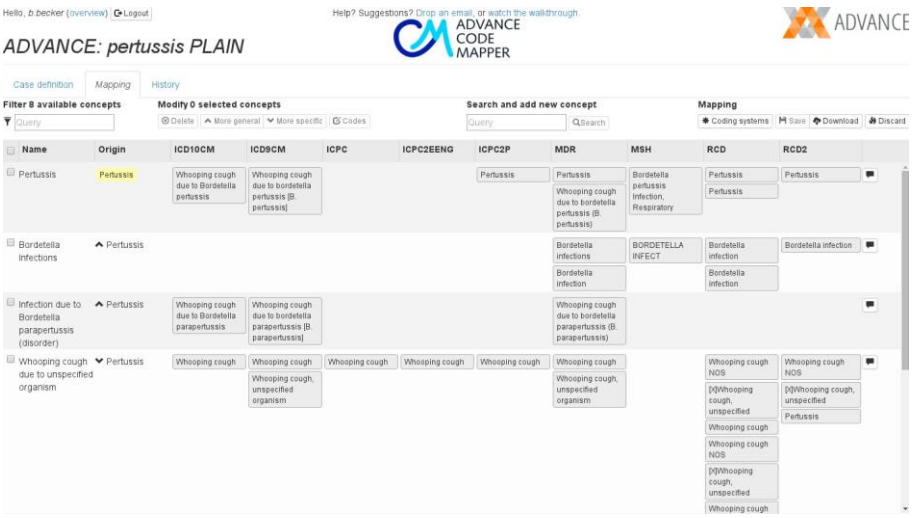


Figure 13: Screen-shot of the CodeMapper application

Extraction of outcomes

The Codemapper lists will be sent to the databases who will be requested to extract the events and transform them in the common data model format. Codemapper will provide only disease codes, thus database experience should be used to define alternative algorithms. (e.g. using drugs as proxy, or using a combination of codes and drugs.

The case definition comprises information on the drugs and procedures that are used for the event. The event teams will collate the experience from the databases, the codes and drugs/procedures in the case definitions to suggest recommended 'component' algorithms is created (combination of disease codes,

text, procedures/measurements and drugs), and each database is invited to extract as many components as apply.

Local database contacts provide feedback on component algorithms and/or propose new ones. The final list of components results from an iterative process of refinement involving local experts and the data derivation leader.

Databases are making available different component algorithms per event.

Database experts extract the component algorithms and run the Algorithm Comparison Module of Jerboa. The result is a dataset of aggregated data which must be uploaded on the Remote Research Environment.

Extraction of codes for the POC have a time priority over the WP 4 events

Results analysis

For all events the following statistics will be calculated

1) age specific incidence rates and standardized incidence rates. These data will be compared
    a. Across databases
    b. Against the literature
2) Overview of code counts (e.g. ICD 9 codes, READ codes)

Data needs to be submitted to RRE by the databases and the data will be post-processed and produce graphics that can be used for discussion.

All the information will be uploaded to the ADVANCE sharepoint

Component analysis

Different algorithms using specific components will be compared and analyzed. Conclusive decision will be taken on algorithms to be used for final extractions

Using the Analysis Tool developed within the EMIF project, local experts are allowed to test the extracted component algorithms in different logical combinations using Boolean operators (AND, OR, AND NOT) in order to build more complex extraction strategies, referred to as *composite algorithms (e.g. ≥1 primary care diagnosis AND ≥1 test result positive)*.

The event team leader makes a proposal for each database. Local experts choose the composite algorithm that they recommend for the identification of the event in their data source, following or challenging the recommendation of the event team leader. Each recommended composite algorithm is stored together with a comment of a data source expert explaining its choice.

An estimate of the sensitivity and PPV is also provided based on previous validation studies, local expert's expectations, information from other data sources and procedures developed in WP4.

Archiving of final algorithms and fingerprint results

Each database will submit the final algorithm that was used both in code as well as in narrative (pseudo code). All this information will be stored in the Codemapper/Sharepoint.

Vaccine fingerprinting

Initial discussions with the databases showed that most databases will have information on the vaccinetype and the ATC code or at least part of it. In the Anatomical Therapeutic Chemical (ATC)[3] classification system, the active substances are divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. Vaccines are coded in different Anatomical groups although the majority are part of the therapeutic subgroup J07.

The divisions are made between bacterial (J07A), viral (J07B), bacterial & viral (J07C), other vaccines (J07D) and cancer vaccines (in L03). Bacterial vaccines are divided in 14 subgroups, viral vaccines are divided in 13 subgroups, grouping is based on the vaccine preventable disease. Within the groupings the most detailed level finishes with the type of antigen.

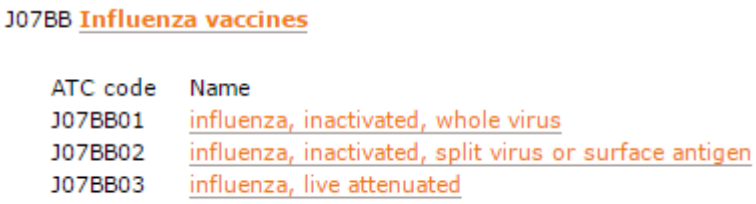


Figure 14: Most detailed ATC codes for vaccines

For future benefit risk analyses we may need additional information on the vaccines (e.g. valence, excipients). This information needs to be retrieved from other sources, therefore we will need a vaccine ontology that can provide additional information on the vaccines which may be useful for analysis as well as to enable the fingerprinting across multiple databases.

Analysis of vaccination fingerprint

The vaccine fingerprint will be described by using vaccination coverage estimates as well as the number of doses per person. To fingerprint the datasources in terms of vaccine coverage /uptake data we will:

Estimate coverage by age, gender, calendar year of the following vaccinations and compare these to the monitoring data from WHO[4], the ECDC funded VENICE II consortium[5] and available national statistics: Bacille Calmette-Guérin (BCG) vaccine, the third dose of diphtheria and tetanus toxoid and pertussis vaccine (DTP3), the third dose of polio vaccine — either oral polio vaccine or inactivated polio vaccine, the first dose diphtheria and tetanus toxoid and pertussis vaccine (DTP1) and the third dose of haemophilus influenza type b (Hib3), seasonal Influenza (compared to VENICE)[6] and the first dose and third dose of human papillomavirus vaccinations (HPV)[7]. These vaccines are also reported to WHO and provide for benchmark

[3] http://www.whocc.no/atc_ddd_index/
[4] http://www.who.int/immunization/en/
[5] http://venice.cineca.org/the_project.html
[6] VENICE II: Go on combining our efforts towards a European common vaccination policy! F D'Ancona on behalf of VENICE II group. Eurosurveillance, 2009, vol. 14 n.12: Seasonal influenza immunisation in Europe. Overview of recommendations and vaccination coverage for three seasons: pre-pandemic (2008/09), pandemic (2009/10) and post-pandemic (2010/11). J Mereckiene, S Cotter, A Nicoll, P Lopalco, T Noori, J T Weber, F D'Ancona, D Lévy-Bruhl, L Dematte, C Giambi, P Valentiner-Branth, I Stankiewicz, E Appelgren, D O'Flanagan, the VENICE project gatekeepers group. Eurosurveillance, 19 (16) 2014.
[7] Health technology assessments on human papillomavirus vaccinations in Europe: a survey from Venice network, Frédérique Dorléans, Daniel Lévy-Bruhl, Cristina Giambi, Fortunato D'Ancona, Giuseppe La Torre, Suzanne Cotter, Jolita Mereckiene, Pawel Stefanoff, Eva Appelgren and the Vaccine European New Integrated Collaboration Effort (VENICE II) project gatekeepers, Italian Journal of Public Health, Volume 9, N. 1 (2012)

Assess timing of childhood immunizations in databases (by age) and compare these with local recommended schedule.

For vaccination fingerprint we will look at describing coverage at 12, 24, 48 and 120 months (BCG, DTP, polio, Hib), and a cumulative approach (Kaplan Meier) for birth cohorts. For HPV, we will assess coverage at age 16. For seasonal influenza vaccination, we will assess coverage by year of age. Timing of vaccinations will be described by plots for age of vaccination by type and dose of vaccine and this will be compared with the information in the vaccine schedules.

APPENDIX 2: POC EVALUATION (SYSTEM TESTING FRAMEWORK)

Methods for the evaluation of the system: POC Evaluation Framework

The POC evaluation will be conducted by the POC evaluation team which is independent from the study teams. The team is coordinated by prof. L Stergioulos (SURREY) and dr. G Ferreira (P-95)

Purpose of the POC evaluation: POC evaluation focuses on combining, analysing and reporting on the performance and knowledge generated during the performance of the POC experiments, to inform the reliability and sustainability of a post-ADVANCE platform, as defined in the Vision and Mission. Conceptually, POC evaluation aims to evaluate the "whole system", including the technology, the framework, and the process used in the POC to perform vaccine B/R assessment.

The POC evaluation is therefore based on a systematic assessment whether the concept designed and tested through conducting the POC is *acceptable and good enough to be recommended for "release into production"* in the ultimate ADVANCE blueprint. Thus, the focus will be on whether ADVANCE adds value in terms of: *(1) Speed to obtain results; (2) Fostering productive collaboration; (3) Enabling good science*.

OBJECTIVES:

1. **Feasibility and effectiveness:** Demonstrate that the selected POC cases can be implemented (more) effectively using the ADVANCE Framework
   "assess the level of attainment of the ADVANCE mission (and vision) statements, through collecting, analysing and reporting on the outputs of the POCs"
   Questions to be answered by the evaluation:
   I: Is it doable?
   Are *the basic conditions and processes* sufficient/enabling for ADVANCE to operate?
   II: Is the output meaningful?
   Is *the evidence produced* meaningful for the purpose of B/R decision making?
   III: Adds-value, cost-effective?
   Is the value/cost ratio in maintaining and performing a study improved *(i.e. greater cost effectiveness),* as perceived by the partners in the collaboration?
2. **Scaling:** Derive general guidelines to guide the implementation of the evidence creation process for any Research Question (RQ) in the future (i.e.: developing the Blueprint: how to design and conduct any vaccine B/R study based on the ADVANCE Framework**)**

   Questions to be answered by the evaluation:

   IV: Is it generalizable and scalable?
   Can ADVANCE be used and perform equally when addressing other B/R questions in other vaccines, and disease areas?

In a nutshell, the ADVANCE Evaluation Framework offers quality indicators, methods, and a timeplan. It spans five evaluation areas, which follow loosely the POC work progression * (i.e. concurrent with the POC timeplan), a number of dimensions with the corresponding sets of indicators and related datasets, and the methods with which these will be measured **:

AREA 1: ADVANCE Process performance and IT infrastructure

The *Process & IT evaluation* covers both **Technical infrastructure** and **Process Workflows** at the same time, and follows the *4 stages* of the overall process: protocol writing, data extraction, transformation, analysis.

Evaluating the Research protocol formation process. The ultimate goal of a process evaluation is to illuminate the pathways linking what starts as a Research Question (theoretical/scientific protocol), and its underlying causal assumptions, to the outcomes produced. In order to achieve this, it is necessary to understand:

- The implementation, both in terms of how the RQ was implemented (new or "tech-enhanced/ IT-enabled" protocol) and the quality of what was delivered;
- the mechanisms of impact linking RQ implementation activities to outcomes;
- how the context in which the RQ is investigated (e.g. external factors) affects both what is implemented and how outcomes are achieved.

For this, the **ISO/IEC 15504 (SPICE)** standard can be followed, which is a framework for the process assessment that defines a **process dimension** and a **capability dimension**.

Evaluating the IT infrastructure. The evaluation measures are built on provisions included in the ISO/IEC 25010 System and Software Product Quality standard (part of ISO/IEC 25000 - SQuaRE). This standard defines internal metrics (static, do not rely on software execution) and external metrics (applicable to running software). It comprises 8 quality characteristics:

IT infrastructure metrics: a. User satisfaction; b. Processing capacity and speed; c. Flexibility; d. Resources and effort; e. CPU times for analyses, loading; f. Failure rates, errors, black outs, off line time; g. Gaps in IT tools and functionality (e.g. document review platform, archiving and version control)

A1 Evaluation dimensions: Time, Cost, Data Access, IT infrastructure, Data sharing, Privacy and ethics, Data processing, Data protection, Data privacy, Qualified study personnel, Accessibility of data (study results and resources)

AREA 2: Scientific validity and innovation

This area covers **Data sources**; **Methods;** and **Innovation potential.**

Approach: Consultation with Experts, Scientific advisory committee

Flexibility/adaptability: how flexible to address the new guidance and requirements WP3 use cases, flu guidance, new vaccines.

Scientific validity: SAB review and feed-back

- Quality of research questions
- Early scientific input in the formulation of relevant B/R framework prior to the protocol development

Innovation aspects:

What has been done that was not already done before?

What has been developed that would not be possible to do without ADVANCE?

What was developed in the POC that can become a tool or asset in the real world?

What was not tested in the POC, but should had been?

Is there potential to implement or support continuous B/R monitoring

A2 Evaluation dimensions: Science / Scientific quality, Data access, Flexibility / generalizeability, Scientific validity, Documentation, Reproducability, Innovation (multiple aspects)

AREA 3: Quality standards, regulatory compliance and legal robustness

Compliance: with legislation, standards, approvals to run the study

Quality of process, data management, data integrity, privacy and security, validation of the writing, validation of the programming, number of amendments to the protocols, number of errors

A3 Evaluation dimensions: Compliance, Ethics, Quality, Quality control, Confidentiality, Data protection, Privacy

AREA 4: Stakeholder satisfaction

Does the POC answer the needs of the different stakeholders and perspectives in terms of decision B/R focus, and the satisfaction and added value of working collaboratively?

Acceptability of study team, study proposal, workflow and report by stakeholders (stakeholder feedback survey), including decision process.

Before and after survey: Needs satisfied

Transparency as perceived by all stakeholders: What information can be made public (protocol, authors), how much time after it was done, is the information understandable, the decision-making processes (minutes, agendas), whose interests are involved / who benefits.

A4 Evaluation dimensions: Acceptability, Transparency, Satisfaction, Public trust

AREA 5: Code of conduct and Collaboration

This area covers the **Code of conduct, Collaboration** and **Rules of governance.**

An important aspect of the evaluation will be the **European network** (regulators and standards organisations, as well as industry) and the enablement/facilitation or strengthening of cross-discipline/cross-sector international collaboration – including sustainable collaboration, and collaboration opportunities.

Before and after survey: Perceived added value of collaboration

A5 Evaluation dimensions: Satisfaction, Interoperability, Network building, Trust

Methods of data collection for the evaluation activities include:

- Surveys, collecting mass feedback from users and stakeholders
- Review and analysis of meeting minutes
- Interviews with experts and stakeholders
- Quantitative data analysis of existing data sets

A detailed description of the specific methods to be used for each dimension/indicator is provided in the *ADVANCE Evaluation Indicators* table**.

**The evaluation process should be as observational as possible (non-interventional) and PIs and study teams should operate as per process rather than towards fulfilling the indicators requirements.*
***A detailed description of all the **indicators, together with specific methods and requirements**, is provided in a separate spreadsheet document (ADVANCE Evaluation Indicators)*

APPENDIX 3: CLINICAL DEFINITIONS

SOMNOLENCE

Drowsiness

Reduced interest in surroundings or increased sleeping:

Grade 1: Sleepier than usual or less interested in surroundings

Grade 2: Not interested in surroundings or did not wake up for a feed/meal

Grade 3: Sleeping most of the time or difficult to wake up

APPENDIX 4: PERTUSSIS VACCINES

ATC code	ATC name	Whole cell or a-cell
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
J07CA05	Diphtheria-hepatitis B-pertussis-tetanus	wP
J07AJ01	Pertussis	wP
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	?
J07AG	Hemophilus influenzae B, combinations with pertussis and toxoids	wP