Vaccine xxx (xxxx) xxx

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations

Miriam Sturkenboom [a,b,c,*], Toon Braeye [d], Lieke van der Aa [d], Giorgia Danieli [e,f], Caitlin Dodd [a,i], Talita Duarte-Salles [g], Hanne- Dorthe Emborg [h], Marius Gheorghe [i,1], Johnny Kahlert [j], Rosa Gini [k], Consuelo huerta [l], Elisa Martín-merino [l], Chris McGee [m,n], Simon de Lusignan [m,n], Gino Picelli [f], Giuseppe Roberto [k], Lara Tramontan [e,f], Marco Villa [o], Daniel Weibel [b,i,2], Lina Titievsky [p]

[a] *Julius Global Health, University Medical Center Utrecht, Heidelberglaan 100, the Netherlands*
[b] *VACCINE.GRID, Basel, Switzerland VACCINE.GRID, Spitalstrasse 33, Basel, Switzerland*
[c] *P-95, Leuven, Belgium Koning Leopold III laan, 1, 3001 Heverlee, Belgium*
[d] *Sciensano, Rue Juliette Wytsmanstraat 14, 1050 Brussels, Belgium*
[e] *Consorzio Arsenal.IT, Veneto Region, Italy*
[f] *Epidemiological Information for Clinical Research from an Italian Network of Family Paediatricians (PEDIANET), Padova, Italy*
[g] *Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain*
[h] *Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Artillerivej 5, DK-2300, Denmark*
[i] *Erasmus University Medical Center, PO Box 2014, 3000 CA Rotterdam, the Netherlands*
[j] *Aarhus University Hospital, Olof Palmes Alle 43–45, DK-8200 Aarhus, Denmark*
[k] *Agenzia regionale di sanità della Toscana, Osservatorio di epidemiologia, Florence, Italy*
[l] *Spanish Agency of Medicines and Medical Devices, Madrid, Spain*
[m] *University of Surrey, Guildford, Surrey GU2 7XH, UK*
[n] *Royal College of General Practitioners, Research and Surveillance Centre, 30 Euston Square, London NW1 2FB, UK*
[o] *ATS della Val Padana, Cremona, Italy*
[p] *219 East 42nd St, NY, NY 10017, USA*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Electronic health data
Big data
Fit-for-purpose
Vaccination
Real world evidence

ABSTRACT

Introduction: The public-private ADVANCE consortium (Accelerated development of vaccine benefit-risk collaboration in Europe) aimed to assess if electronic healthcare databases can provide fit-for purpose data for collaborative, distributed studies and monitoring of vaccine coverage, benefits and risks of vaccines.
Objective: To evaluate if European healthcare databases can be used to estimate vaccine coverage, benefit and/or risk using pertussis-containing vaccines as an example.
Methods: Characterisation was conducted using open-source Java-based (Jerboa) software and R scripts. We obtained: (i) The general characteristics of the database and data source (meta-data) and (ii) a detailed description of the database population (size, representatively of age/sex of national population, rounding of birth dates, delay between birth and database entry), vaccinations (number of vaccine doses,

Abbreviations: ADVANCE, Accelerated development of vaccine benefit-risk collaboration in Europe; AUH, Aarhus University Hospital datasources; ARS, Agenzia regionale di sanità; ATSVP, ATS (Local Health Agency) della Val Padana; BIFAP, Base de Datos para la Investigación Farmacoepidemiológica en Atencion Primaria; CDM, common data model; DAP, Data access provider; EHDEN, European Health Data & Evidence Network; ENCePP, European Network of Centres of excellence in Pharmacoepidemiology & Pharmacovigilance; HHE, hypotonic hyporesponsive episode; ICD-9/ICD-10, International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9) and tenth revision (ICD-10); ICPC, International Classification of Primary Care; MMR, measles, mumps, rubella; OMOP, Observational Medical Outcomes Partnership; POC, proof of concept; PRISM, Post-licensure Rapid Immunization Safety Monitoring program; RCGP RSC, Royal College of General Practitioners Research and Surveillance Centre; RRE, remote research environment; SIDIAP, The Information System for Research in Primary Care; SSI, Staten Serum Institute; THIN, Health Improvement Network; USA, United States of America; VAC4EU, Vaccine monitoring collaboration for Europe; VSD, Vaccine Safety Datalink; WHO, World Health Organization.

* Corresponding author at: University Medical Center Utrecht, Heidelberglaan 100, the Netherlands.
E-mail addresses: m.c.j.sturkenboom@umcutrecht.nl, m.sturkenboom@vaccinegrid.org, miriam.sturkenboom@p-95.com (M. Sturkenboom), toon.braeye@sciensano.be (T. Braeye), lieke.vanderaa@sciensano.be (L. van der Aa), gdanieli@consorzioarsenal.it (G. Danieli), c.n.dodd@umcutrecht.nl (C. Dodd), tduarte@idiapjgol.org (T. Duarte-Salles), hde@ssi.dk (Hanne- Dorthe Emborg), m.gheorge@erasmusmc.nl (M. Gheorghe), jok@clin.au.dk, kber@clin.au.dk (J. Kahlert), rosa.gini@ars.toscana.it (R. Gini), chuerta@aemps.es (C. huerta), emartinm@aemps.es (E. Martín-merino), c.mcgee@surrey.ac.uk (C. McGee), s.lusignan@surrey.ac.uk (S. de Lusignan), g.picelli@virgilio.it (G. Picelli), giuseppe.roberto@ars.toscana.it (G. Roberto), ltramontan@consorzioarsenal.it (L. Tramontan), Marco.Villa@ats-valpadana.it (M. Villa), lina.titievsky@pfizer.com (L. Titievsky).
[1] Current address: University of Oslo, Norway.
[2] Current address: Weibel Consulting, Den Haag, the Netherlands.

https://doi.org/10.1016/j.vaccine.2020.01.100

recording of doses, pattern of doses by age and coverage) and events of interest (diagnosis codes, incidence rates). A total of nine databases (primary care, regional/national record linkage) provided data on events (pertussis, pneumonia, death, fever, convulsions, injection site reactions, hypotonic hyporesponsive episode, persistent crying) and vaccines (acellular pertussis and whole cell pertussis) related to the pertussis proof of concept studies.

Results: The databases contained data for a total population of 44 million individuals. Seven databases had recorded doses of vaccines. The pertussis coverage estimates were similar to those reported by the World Health Organisation (WHO). Incidence rates of events were comparable in magnitude and age-distribution between databases with the same characteristics. Several conditions (persistent crying and somnolence) were not captured by the databases for which outcomes were restricted to hospital discharge diagnoses.

Conclusion: The database characterisation programs and workflows allowed for an efficient, transparent and standardised description and verification of electronic healthcare databases which may participate in pertussis vaccine coverage, benefit and risk studies. This approach is ready to be used for other vaccines/events to create readiness for participation in other vaccine related studies.

1. Introduction

The Vioxx scandal and the subsequent pharmacovigilance reforms in the United States of America (USA), resulted in the creation of a system that can monitor safety on more than 100 million subjects in the USA, currently the Sentinel system. This led to the creation of other distributed networks that conduct collaborative multisite studies throughout the world [1–3]. The creation of these collaborations and the need for regulators to use real-world evidence has also resulted in enhanced efforts to improve transparency about the quality of data sources, methods, definitions, common data models, analytics and reporting [4–7].

The need to collaborate and use healthcare data to assess vaccine safety was recognised in the 90s in the USA and the Vaccine Safety Datalink (VSD) was the first multi-site network set up to use health data [8]. This was then extended by the Post-licensure Rapid Immunization Safety Monitoring program (PRISM)/FDA Sentinel prior to the 2009 H1N1 influenza pandemic [9]. The VSD conducted extensive data quality checks at its inception, focusing on validation of completeness of data on vaccinations [10]. Recently, the FDA assessed the surveillance capabilities of PRISM by characterising the underlying data and concluded that the data on vaccine exposures and health outcomes appeared complete enough to support robust safety monitoring [9].

Large healthcare databases have been used to study the use and outcomes of therapeutics for several decades [1]. The size of the populations they cover allows the study of infrequent events and their representativeness of routine clinical care makes it possible to study real-world safety, effectiveness and drug utilisation patterns [11,12]. In addition, they are accessible to many researchers since they can be available at relatively low cost and without long delays. A recent review of post-authorisation safety studies registered in the European Post-Authorisation Study register showed 30–50% of studies used healthcare databases to perform post-authorisation mandatory studies [13]. While the use of these databases offers potential significant advantages, in the past the quality of routine healthcare data was often questioned [14,15]. Data in routine healthcare databases are recorded for purposes other than research and may not be suitable for use as such. The use of real-world evidence for regulatory decision-making requires assurance about the quality of the data.

The Accelerated Development of VAcciNe benefit-risk Collaboration in Europe project (ADVANCE) is a public-private consortium of 47 organisations [16]. The ADVANCE consortium vision was to deliver best evidence at the right time to support decision-making on vaccination in Europe. Its mission was to establish a prototype of a sustainable and suitable for purpose system that can rapidly provide the best-available scientific evidence on post-marketing vaccination benefits and risks for informed decision-making and effective post-marketing surveillance [2]. The ADVANCE consortium has designed and tested a system to enable the generation of evidence on background rates of events of interest, vaccine coverage, benefits and risks following vaccination from existing distributed healthcare databases [17]. To assess if European healthcare databases that could contribute to ADVANCE are able to generate evidence suitable for decision making, the ADVANCE project performed an assessment of their fitness for purpose, based on detailed database characterisation. This paper describes the parameters, methods and results of the healthcare database characterisation that were used to evaluate if the databases were fit to participate in the pertussis-related proof of concept (POC) studies performed by ADVANCE and described in other papers in this supplement [18–24].

2. Methods

2.1. Setting

Database characterisation on population, vaccines and events was performed in nine European healthcare databases from four countries: Denmark, Spain, Italy and the United Kingdom (Table 1; Table 2). Detailed descriptions of these databases can be found in the Supplementary File.

2.2. Data in the ADVANCE common data model for database characterisation and assessment

As described in the paper on the ADVANCE system the common data model comprises three data files: population, events and vaccinations [17].

2.2.1. Population

The source population comprised all persons registered in the database and having at least one day of follow-up during the study period. Data for all individuals recorded in each database from the start of follow-up, defined birth or first data availability, whichever was latest, until the end of follow-up, defined as the date at last data retrieval, leaving the database (moving out of area/other provider), or death whichever date was earliest, were used to define the follow-up for database characterisation. The only eligibility criteria were that the date of birth, start and end of follow-up dates, and gender needed to be present. The study period varied between

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

Table 1
Characteristics of databases included in the fit-for-purpose assessments.

Country	Denmark		Spain		Italy			United Kingdom	
Name	AUH	SSI	BIFAP	SIDIAP	PEDIANET	Val Padana	Tuscany	THIN	RCGP
Type of organisation providing access	Different public data holders		Spanish Agency of Medicines and Medical Devices	Public research organisation	Private organisation; vaccines from public health	Local public health agency	Regional public health agency	Academic License holder (Erasmus MC)	Charity
Origin of data	Hospital discharge diagnoses linked to population and vaccination registries. National health care		Family paediatricians and general practitioners medical records	Family paediatricians and general practitioners medical records	Family paediatricians medical records linked to Veneto vaccine register	Hospitalisation discharge diagnoses linked to population and vaccination registries	Hospitalisation discharge diagnoses linked to population and vaccination registries	General practitioners medical records	General practitioners medical records
Geographic spread	Regional	National	Multiregional 9 out of 17	Catalunya Region	Sample from Veneto Region	Regional, province	Tuscany Region	National sample	National sample
Data governance	Approval Danish Data protection Agency	Approval Danish Data Protection Agency posterior check	Protocol-based approval	Protocol-based approval	Generic consent from parents collected once	Generic approval	Generic approval (monthly meeting, posterior check)	Protocol-based approval	Protocol-based approval
Age range covered	All		All	All	0–14 years	All	All	All	All
Disease diagnosis coding	ICD-10 Danish version		ICD-9, ICPC & text	ICD-10	ICD-9 and text	ICD-9	ICD-9	READv2	READCTV3 & READv2
Type of outcomes covered	Emergency visits, hospitalisation, death		Primary care, incomplete specialist & hospitalisations only if GP enters	Primary care, specialist & hospitalisations	Primary care, incomplete specialist & hospitalisations only if FP enters	Only hospitalisations	Hospitalisations, emergency visits, death	Primary care, specialist & hospitalisations	Primary care, incomplete specialist & hospitalisations only if GP enters
Local coding of vaccines	ATC, local code & reimbursement database	ATC & local code, before 2013 reimbursement database	Local code		Regional code mapped to CVX code		Italian marketing authorisation code ('AIC')	READ codes and prevention records	READ codes
POC studies participation	FP POC1	FP POC1POC1.2	FP POC1	FP POC1	FP POC1	FP POC1.2	FP POC1.2	FP POC1	FP POC1 POC1.2

FP = fingerprint (feasibility assessment), POC1: coverage, benefit, risk and benefit-risk proof of concept study 1, POC1.2: Near real-time monitoring of pertussis vaccine coverage, benefits and risks proof of concept study 1.2

Table 2
Characteristics of the population (with one year of follow-up) used as input for event incidence calculations for databases participating in the ADVANCE proof of concept studies (POC-1/POC1.2).

Country	Denmark		Spain		Italy			United Kingdom		Total
Name	AUH	SSI	BIFAP	SIDIAP	PEDIANET	Val Padana	Tuscany	THIN	RCGP RSC	
Calendar years covered	1900–2016	1995–2014	2003–2014	2005–2015	2003–2015	2001–2016	2002–2016	1985–2015	1989–2016	
Birth date rounding rule	None	None	None	Day rounded to 1st of month	Day rounded to 15th of month	None	None	Children rounded to 1st of month, adults to year	Day rounded to 1st of month	
Total person/years (% of total)	81,400,848* (22%)	98,026,412 (27%)	29,997,580 (8%)	51,024,109 (14%)	414,765 (0%)	4,435,896 (1%)	48,423,256 (13%)	28,468,872 (8%)	22,291,050 (6%)	364,482,788 (100%)
Person/years from 1995 onwards	20,865,425	98,026,412	29,997,580	51,024,109	414,765	4,435,896	48,423,256	26,658,434	21,159,284	301,005,161
0–1 years	622,864	2,471,899	711,039	1,66,827	82,864	83,363	802,313	681,813	541,281	5,997,436
2–4 years	934,265	3,583,064	898,961	1,648,703	110,115	116,853	1,160,487	904,224	747,564	10,104,236
5–14 years	3,054,841	11,989,508	2,745,172	5,048,430	221,786	404,588	3,942,729	3,028,660	2,434,910	32,870,624
15–24 years	3,866,239	11,530,630	4,972,367	4,895,042		382,962	4,000,839	2,868,736	2,371,168	34,887,983
25–44 years	4,980,525	272,710,60	9,423,903	16,699,777		1,178,090	13,173,859	7,467,315	6,056,917	58,980,386
45–64 years	4,903,891	25,800,887	6,920,756	12,673,461		1,247,479	13,279,552	7,042,736	5,662,642	77,531,404
65+ years	2,502,800	15,379,365	4,325,382	8,891,869		1,022,561	11,2063,476	4,664,949	3,344,803	149,692,405

databases, depending on when the database collection started and ended in 2017. Data access providers created a population file in the format of the common data model (CDM) (patient ID, date start follow-up, date end follow-up, birthdate, gender), which was then analysed to estimate the population size, its gender, age and follow-up distribution and representativeness in comparison with the national population statistics.

2.2.2. Events

The coverage, benefit and risk POC studies for pertussis containing vaccine required data for several vaccine risk and effectiveness outcomes. The outcomes included the following events: convulsions/seizures (any type), death, febrile convulsions, fever, generalised convulsions, hypotonic hyporesponsive episode (HHE), injection site reactions, persistent crying, pertussis, pneumonia, and somnolence. The outcomes were defined using definitions from the Brighton Collaboration, learned societies, the World Health Organization (WHO) or the European Centre for Disease prevention and Control [19,20]. The case definitions were mapped to an initial list of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9), and tenth revision (ICD-10), Read and the International Classification of Primary Care (ICPC) codes using the ADVANCE Code mapper tool that maps between dictionaries using the Unified Medical Language System [25]. Data access providers (DAPs) for each database were asked to modify and validate the proposed codes based on local coding habits and prior experience. Each DAP extracted the final list of codes for the specific events in their local terminology and transformed the data into the event file of the common data model containing the following fields: patient ID, event type, date, original code (ICD-9/10, Read, ICPC, text) [17]. This event file was linked to the population file to calculate event incidence rates and to assess whether these rates were as expected, as a demonstration of suitability. The extracted event codes for each database are listed in supplementary Table S1.

2.2.3. Vaccination

The exposure of interest for the initial round of POC studies was vaccination with whole cell or acellular pertussis (wP or aP, respectively) containing vaccines. The DAPs extracted data for pertussis vaccination based on local coding and transformed the data into the vaccine input file which contained the fields: patient ID, vaccine type, ATC, brand, date, dose-recorded, dose-derived [17]. The vaccine type (vactype) was identified as aP or wP, when possible, and when not, it was unknown type (uP). The vaccine dose was recorded as 1, 2, 3 or booster dose. When the dose was not recorded in the data source it was derived by the DAP based on imputations from the chronological order of the individual's vaccination dates and the national recommended vaccination schedule.

2.2.4. Data management and analyses

The DAPs extracted data from their database using the local data format and software, which were transformed into the ADVANCE CDM (CSV format) [17]. We used Jerboa data processing software, which is JAVA-based, for event code counting and incidence calculations and R scripts for vaccine descriptions and coverage estimations [26]. The Jerboa software has been used for multiple studies and is freely available [27]. The R script for vaccination fingerprinting was created specifically for ADVANCE and is described in Table S2. The scripts and instructions were sent to the DAPs, who ran the scripts against their input files and the outputs were sent through a secure file transfer protocol (File Zilla or HighTail) to a remote research environment (RRE) [17].

To assess the population input file we investigated the following indicators from the Jerboa output: the size of the population (number of persons and person-time) by data source, age, calendar year, exit years and entry years and the time from birth till first registration, rounding of birthdates and persontime by birth year using the relevant R-script.

The event characterisation included code counts by type of event and database and event incidence rates in the population by calendar year, gender and age. Age was categorised per year until 17 years old, from 18 to 24 years, and then in 5- year categories. We subsequently categorised age in 0–1, 2–4, 5–14, 15–24, 25–64, 65 and older for description, as this coincides with age of routine vaccination in general and because this categorisation was compatible with the PRISM database age categories and therefore age-specific comparisons of incidence rates were possible [9]. For the population and incidence estimates calculated with Jerboa, there was one-year run-in period, except for individuals with an entry date within 6 months of birth, those persons all started follow-up at birth. Events recorded in the one year run-in prior to start of follow-up were not considered to be incident and only first events were considered to be incident. To have a

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

comparable period of calendar time across databases, incidence rates were limited to calendar years 1995–2016, if these were available.

Vaccine coverage was estimated to allow comparisons of 1-year coverage estimates with external coverage estimates as a marker for completeness. Vaccination coverage was estimated by birth-year using various methods; here we report the cumulative distribution function-based estimates that correct for left and right censoring [28]. To estimate pertussis coverage, only individuals born between 1990 and 2015 were selected, and to limit impact of left censoring, only those with a follow-up start date within 30 days of birth date were included.

The results of the data characterisation outputs were compared with external benchmarks and discussed with the DAPs who could then update the input files and repeat the assessment until they considered it was appropriate. The DAPs completed a feasibility assessment sheet to summarise their data and then the DAPs and study teams for the different pertussis POC studies decided on the suitability of the database to participate in the POC studies. All DAPs that had completed this process by November 2016 were eligible to participate in the first series of POC studies on pertussis vaccine coverage, benefits, risks and benefit/risk [18–21,23]. The DAPs that completed the assessment by January 2018 could also participate in a second POC that focused also on the same pertussis containing vaccines and benefit and risk outcomes, but tested prospective near real time monitoring (POC1.2) [24].

3. Results

Nine databases participated in the event and vaccine fit-for-purpose assessment. Five of these databases collected longitudinal medical record data from multiple health professionals and the other four were regional or national record linkage systems (Table 1). Access to the databases was provided by public health organisations or regulatory agencies (SSI, BIFAP, ARS, ATSVP), public research organisations (AUH, RCGP RSC, SIDIAP). PEDIANET is a private research organisation that agreed to provide access, and for THIN, access was via a license held by Erasmus Medical Centre (an academic institute and partner in ADVANCE). Two databases (ATSVP and ARS) could not participate in the first POC studies as they did not meet the timelines but they contributed to the POC1.2 pertussis near real-time monitoring study (Table 1) [24].

3.1. Population characterisation

The period of data availability varied between the databases, with AUH having data available for the longest period (Table 2). Table 2 and Fig. 1 show the results of the population characterisation based on the Jerboa based event fingerprinting output (limiting the population to persons with at least one year of history to be able to estimate incidence). The total follow-up time for the 34 million individuals with at least one year of data in the nine databases was 364,482,788 person-years. Limiting the person time to calendar years from 1995 onwards reduced this to 301,005,161 person-years. The largest overall contribution of person-years of follow-up (27%) was from the SSI database (Fig. 1; Table 2). All databases were representative in gender and age distributions with their national statistics, with a slight underrepresentation of very old persons (>80 year) (data reported as deliverable 5.2 in ADVANCE project).

3.2. Event characterisation

The incidence rates of some of the events of interest by calendar year showed peaks in the earlier years of follow-up but were more stable in the later years (Fig. 2). Age-specific incidence rates were comparable between similar types of databases (record linkage, GP type) in the same country (Table 3, Fig. 2), most rates were highest in the youngest age categories, pneumonia was high in the youngest and olders categories, whereas death rates were highest in the 65+ category. Temporal patterns were observed for some databases (Fig. 2). In the ARS database the incidence of all events increased after 2009 when records of emergency visits became available. Persistent crying could not be extracted from the AUH, SSI or SIDIAP databases using ICD-10, due to the absence of codes specific for this event. Rates of the other events in the general practice based databases were comparable within countries, except for fever. The rates for fever showed the largest variation because in the BIFAP and PEDIANET databases data for fever could also be extracted using free text, which increased sensitivity, and in the BIFAP database fever associated with other illnesses, signs or symptoms was also included. In the RCGP RSC and SIDIAP databases, the incidence rates for most events increased over time (Fig. 2).

3.3. Vaccine characterisation

The total population without limitations in duration in follow-up that was used for vaccine fingerprinting comprised 44,398,858 persons (Table 4). The percentage of individuals registered within 6 months of birth was high for recent birth years, compared with earlier years (Fig. 3). The percentage of persons registered from birth in older birthyears was lower in the primary care databases, suggesting censoring and incompleteness of follow-up from birth.

Data for more than 14 million dose of aP were captured in the nine databases and dates for almost 7 million doses of wP vaccinations were captured in four databases (RCGP RSC, THIN, SSI and BIFAP) (Table 4). Data on the recorded vaccine doses were available in seven of the databases but not for the ATSVP and RCGP RSC databases. Pertussis vaccine coverage estimates based on the cumulative distribution function (CDF) were calculated for specific birth years and were similar to those reported to WHO by the countries, except for early birth years in ATSVP (Fig. 4) [29]. The ARS database did not have vaccination data available for birth cohorts until 2012, at the moment of the database characterization execution.

4. Discussion

This paper describes the characterisation of study populations from the nine databases participating in the ADVANCE project located in Denmark, Spain, Italy and the United Kingdom. These databases included data for more than 44 million individuals and more than 360 million years of follow-up. Our analyses showed that the system works, i.e., that the data on the population, vaccinations and events from the databases could be transformed locally and analysed using a simple common data model. The current CDM was mapped to concepts locally by the DAP according to specific supplied rules, which may be adapted for subsequent studies (adaptive rule system), but for the future we recommend the use of more organizing models, which will provide more transparency and centralized options for the mapping to concepts. Standardised Jerboa and R programs were sent to the DAPs who used them to generate output that was used to create fit-for-purpose assessment indicators, such as population distributions, dynamics, incidence rates of events and vaccination exposure and coverage that were reviewed and approved by the DAPs. Using this model we were able to perform a fit-for-purpose assessment of the databases prior to their participation in specific protocol-based studies on pertussis vaccination as described in this supplement.

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

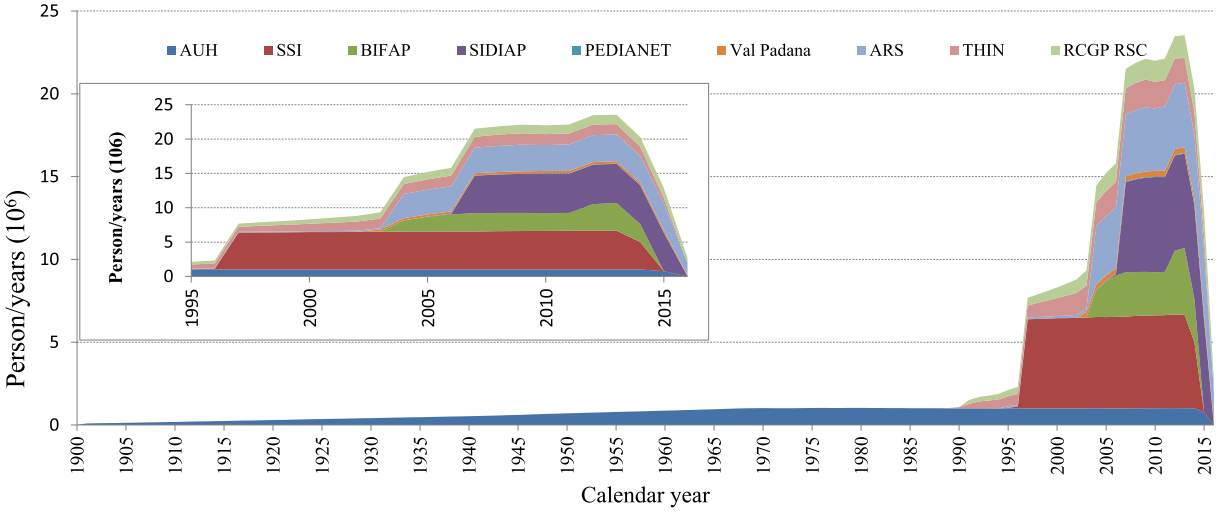

Fig. 1. Follow-up time in person/yearsx10^6 by calendar year.

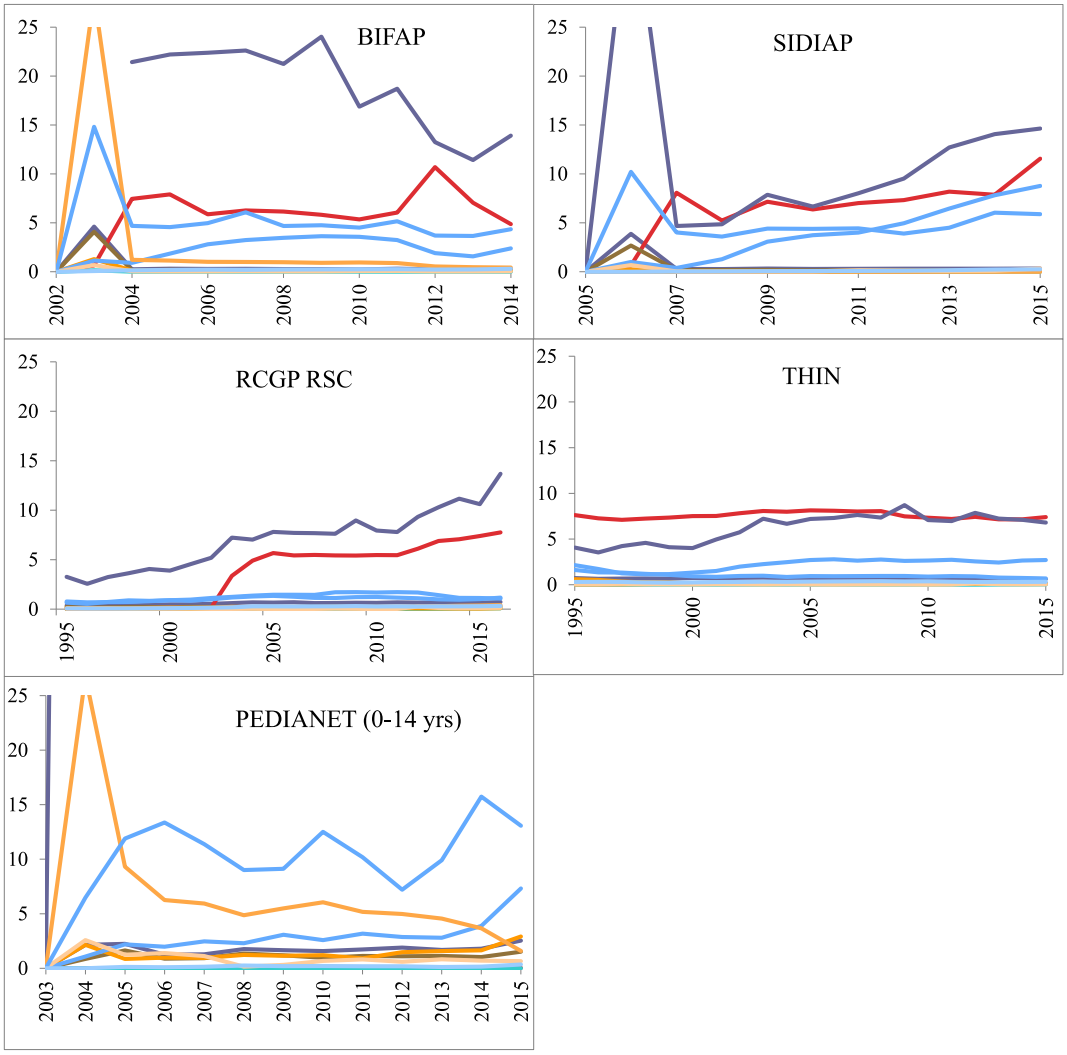

Fig. 2. Incidence rates (per 1000 person/years on Y-axis) of first occurrence of the indicated events of interest (per 1000 person/years) (all ages) by calendar year, to assess any changes in recording of diagnoses over time (A) General practice-based databases; (B) Hospital-based databases.

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

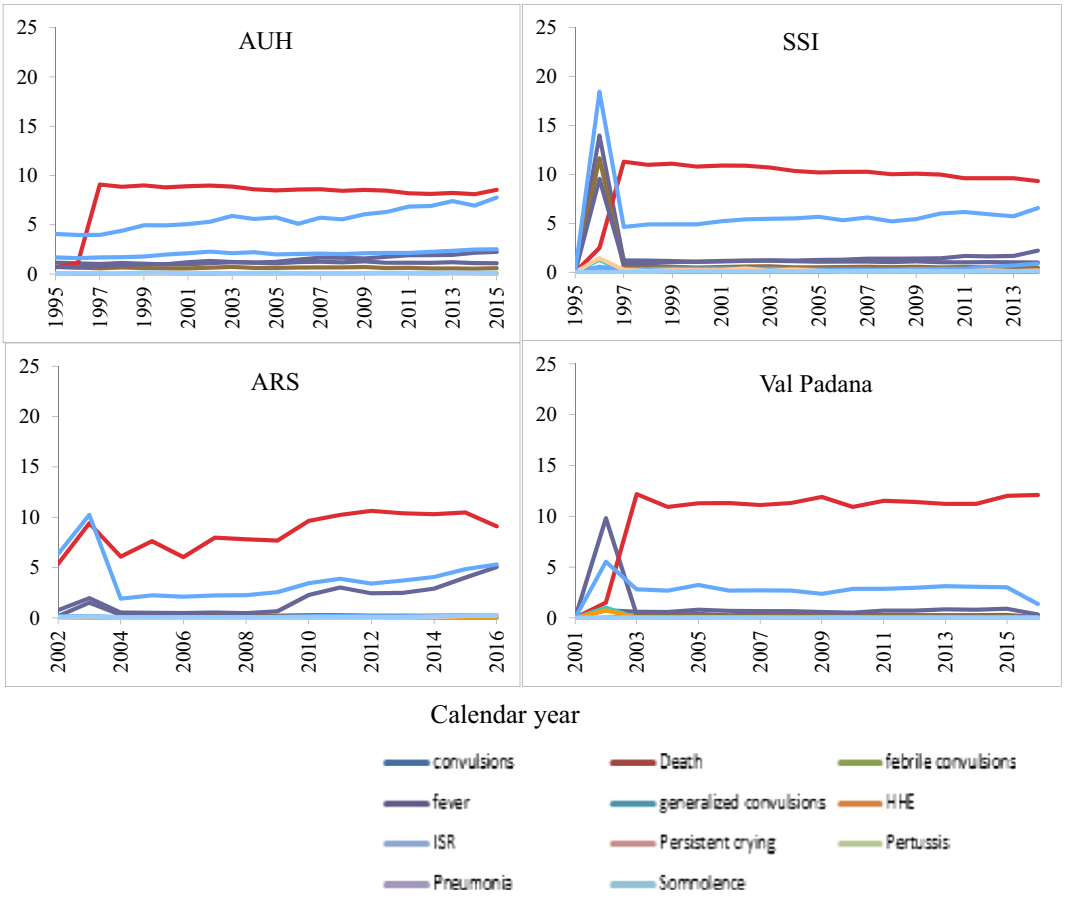

Fig. 2 (*continued*)

The key findings of this fit-for-purpose assessment showed, first, that information on more than 21 million doses of pertussis-containing vaccinations were available across the nine databases, providing combined power to study both benefits and risks of events following pertussis vaccination. Coverage estimates of the first dose of pertussis-containing vaccinations were comparable to the estimates that the countries report to the WHO, suggesting that, although, in some databases, the data were incomplete for the earlier years of data collection, they were complete in the later years on a population level.

Second, the periods covered by the databases differed and incidence rates of events in initial years were not always similar to those for more recent years, probably due to changes in recording and/or completeness. Third, certain events associated with vaccination reactogenicity (e.g. somnolence, fever and persistent crying) could be reliably extracted from primary care-based databases, and to a certain extent from databases with emergency visits (e.g. ARS), but not from databases capturing only hospital discharge diagnoses. This demonstrates that databases from different settings can provide complementary data when included in studies. Fourth, the incidence rates for the events of interest were similar between databases of the same type, demonstrating the success of our terminology mapping and harmonisation processes. Also, the data were similar to public or published benchmark data (Table S1). For example, the age-specific incidence rates for convulsions and seizures in the PRISM data in the USA were comparable with the incidence rates in the included European record linkage and general practice databases [9]. In addition, our incidence rates for febrile and generalised seizures were similar to previously published European data [30]. No publicly available benchmark data were available for other events, except for pneumonia, death and pertussis rates. An overview of studies on community-acquired pneumonia leading to hospitalisations, reported incidence rates of between 5 and 20 per 1000 person/years for patients aged 65 years and older, which is consistent with the rates we found in the record linkage databases [31]. Prior to introduction of pneumococcal vaccine, the incidence rate of clinical pneumonia in children aged 0–5 years old was reported to be 0.06 episodes per child/year in Europe [32]. The incidence rates in the primary care databases in our study were similar but the rates in hospital-based database were lower, which is to be expected as not all children with community-acquired pneumonia will be hospitalised. Our estimates for pertussis rates were based on the diagnoses recorded in the databases, and for Denmark the rates were similar to those for laboratory-confirmed diagnoses. For other countries, comparisons made with national surveillance data showed similarity [19,22]. Death rates were similar to those from national statistics. And lastly, the populations of the databases were representative of their national populations, in age and gender, even if the databases enrolled only a sample of the population.

Although our results from this database characterisation and fit-for-purpose assessment suggest the utility of the proposed system, we recognise that we used our own indicators. Selection of additional indicators to assess whether databases are fit-for-purpose for vaccine-related studies could be performed with other distributed vaccine-oriented data networks such as Vaccine Safety Datalink, and PRISM. Europe is challenging as many coding systems and practices are used in the diverse health care systems. Databases collect information from disparate provenances, hampering the use of a deep common data model, where all the raw

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

Table 3
Number of event and incidence rates (95% confidence intervals) per 1000 person-years aggregated across all calendar years.

	Denmark		Spain		Italy			United Kingdom	
	AUH	SSI	BIFAP	SIDIAP	PEDIANET	Val Padana	Tuscany	THIN	RCGP
Pertussis (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	671; 122; 188; 135; 237; 217; 74	3987; 2280; 5506; 744; 1628; 810; 260	467; 158; 415; 48; 145; 62; 25	1406; 730; 2045; 145; 508; 327; 114	65; 103; 146	10; 0; 6; 1; 1; 0; 0	523; 265; 261; 71; 183; 240; 258	321; 182; 269; 75; 178; 155; 55	218; 136; 321; 113; 351; 338; 84
IR_{0-1}	0.2 (0.2, 0.2)	1.6 (1.6, 1.7)	0.7 (0.6, 0.7)	1.2 (1.1, 1.3)	0.8 (0.6, 1.0)	0.1 (0.1, 0.2)	0.7 (0.6, 0.7)	0.4 (0.4, 0.5)	0.4 (0.3, 0.4)
IR_{2-4}	0 (0, 0)	0.6 (0.6, 0.7)	0.2 (0.1, 0.2)	0.4 (0.4, 0.5)	0.9 (0.8, 1.1)	0 (0, 0)	0.2 (0.2, 0.3)	0.2 (0.2, 0.2)	0.2 (0.1, 0.2)
IR_{5-14}	0 (0, 0)	0.5 (0.4, 0.5)	0.2 (0.1, 0.2)	0.4 (0.4, 0.4)	0.7 (0.6, 0.8)	0 (0, 0)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)
IR_{15-24}	0 (0, 0)	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0.1)
IR_{25-44}	0 (0, 0)	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0.1 (0, 0.1)
IR_{45-64}	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.1)
IR_{65+}	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Death (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	430, 62, 144, 565 2872, 20,297, 106,090	3884; 592; 1335; 5062; 28,210; 164,038; 808,138	363; 219; 516; 1868; 8918; 31,058; 161,527	523; 236; 557; 1211; 10,002; 46,171; 330,119	1; 2; 4	43; 17; 37; 155; 816; 5139; 44,446	221; 126; 332; 930; 6134; 35,800; 376,965	450; 128; 351; 1124; 5568; 28,563; 181,013	167; 58; 112; 482; 2522; 12,091; 83,162
IR_{0-1}	1.1(1.0, 1.2)	1.6 (1.5, 1.6)	0.5 (0.5, 0.6)	0.4 (0.4, 0.5)	0 (0, 0.1)	0.5 (0.4, 0.7)	0.3 (0.2, 0.3)	0.6 (0.5, 0.7)	0.3 (0.2, 0.3)
IR_{2-4}	0.1(0.1, 0.1)	0.2 (0.2, 0.2)	0.2 (0.2, 0.3)	0.1 (0.1, 0.2)	0 (0, 0.1)	0.1 (0.1, 0.2)	0.1 (0.1, 0.1)	0.1 (0.1, 0.2)	0.1 (0.1, 0.1)
IR_{5-14}	0.1(0.1, 0.1)	0.1 (0.1, 0.1)	0.2 (0.2, 0.2)	0.1 (0.1, 0.1)	0 (0, 0)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0 (0, 0.1)
IR_{15-24}	0.3(0.3, 0.4)	0.4 (0.4, 0.5)	0.4 (0.4, 0.4)	0.2 (0.2, 0.3)		0.4 (0.3, 0.5)	0.2 (0.2, 0.2)	0.4 (0.3, 0.4)	0.2 (0.2, 0.2)
IR_{25-44}	0.8(0.8, 0.8)	1.0 (1.0, 1.0)	0.9 (0.9, 1)	0.6 (0.6, 0.6)		0.7 (0.6, 0.7)	0.5 (0.5, 0.5)	0.7 (0.7, 0.7)	0.4 (0.4, 0.4)
IR_{45-64}	5.2(5.1, 5.3)	6.4 (6.3, 6.4)	4.5 (4.4, 4.5)	3.6 (3.6, 3.7)		4.1 (4, 4.2)	2.7 (2.7, 2.7)	3.8 (3.8, 3.9)	2 (2.0, 2.1)
IR_{65+}	45.3(45.0, 45.5)	52.5 (52.4, 52.7)	37.3 (37.2, 37.5)	37.1 (37, 37.3)		43.5 (43.1, 43.9)	31.3 (31, 31)	36.5 (36.3, 36.7)	24.5 (24.3, 24.7)
Pneumonia (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	7160, 1867, 1360, 939, 4485, 13,015, 36,910	53,261; 17,585; 11,550; 9026; 42,453; 99,749; 286,540	10,528; 17,239; 13,871; 6999; 23,597; 25,537; 36,522	19,543; 37,677; 29,617; 6098; 37,251; 39,140; 56,818	665; 2059; 1587	504; 640; 570; 120; 618; 1684; 8100	7184; 7431; 6177; 1697; 8936; 18,492; 102,845	1060; 973; 909; 616; 2920; 5141; 16,806	1155; 1008; 902; 581; 2956; 4835; 11,779
IR_{0-1}	19.3(18.8, 19.7)	22.0(21.9–22.2)	15 (14.7, 15.3)	16.9 (16.7, 17.2)	8.1 (7.5, 8.7)	6.1 (5.6, 6.6)	9.0 (8.8, 9.3)	1.4 (1.3, 1.5)	2.0 (1.9, 2.1)
IR_{2-4}	3.7(3.5, 3.8)	5.2(5.1–5.2)	20.4 (20.1, 20.7)	24.5 (24.3, 24.8)	19.5 (18.7, 20.4)	5.6 (5.0. 2, 6)	6.6 (6.4, 6.8)	1.0 (0.9, 1.1)	1.3 (1.2, 1.4)
IR_{5-14}	0.8(0.7, 0.8)	1.0(1.0–1.02)	5.3 (5.2, 5.4)	6.4 (6.3, 6.5)	7.7 (7.3, 8.1)	1.4 (1.3, 1.6)	1.6 (1.6, 1.7)	0.3 (0.3, 0.3)	0.4 (0.3, 0.4)
IR_{15-24}	0.6(0.5, 0.6)	0.8(0.8–0.8)	1.4 (1.4, 1.5)	1.3 (1.2, 1.3)		0.3 (0.3, 0.4)	0.4 (0.4, 0.5)	0.2 (0.2, 0.2)	0.2 (0.2, 0.3)
IR_{25-44}	1.3(1.2, 1.3)	1.6(1.5–1.6)	2.5 (2.5, 2.6)	2.3 (2.2, 2.3)		0.5 (0.5, 0.6)	0.7 (0.7, 0.7)	0.4 (0.4, 0.4)	0.5 (0.4, 0.5)
IR_{45-64}	3.4(3.4, 3.5)	3.9(3.9–4.0)	3.8 (3.7, 3.8)	3.2 (3.1, 3.2)		1.4 (1.3, 1.4)	1.4 (1.4, 1.4)	0.7 (0.7, 0.7)	0.8 (0.8, 0.8)
IR_{65+}	17.1(16.9, 17.3)	20.0(19.9–20.0)	8.7 (8.6, 8.8)	6.6 (6.6, 6.7)		8.1 (8, 8.3)	8.7 (8.7, 8.8)	3.5 (3.4, 3.5)	3.5 (3.5, 3.6)
Convulsions (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	11,264; 3759; 1440; 1827; 2190; 2735; 2142	42,501; 13,874; 5631; 6059; 12,085; 14,501; 11,239	3209; 1752; 586; 426; 726; 543; 822	6702; 3649; 1509; 557; 1159; 867; 1179	330; 276; 102	625; 350; 155; 27; 23; 62; 72	5138; 3115; 1315; 202; 349; 416; 1107	1638; 1168; 2022; 2219; 4239; 3951; 4118	1044; 809; 1302; 1597; 2903; 2706; 2894
IR_{0-1}	3.5 (3.4, 3.6)	17.4 (17.3, 17.6)	4.5 (4.4, 4.7)	5.8 (5.6, 5.9)	4 (3.6, 4.4)	7.5 (7.0, 8.2)	6.4 (6.3, 6.6)	2.2 (2.1, 2.3)	1.8 (1.7, 1.9)
IR_{2-4}	0.8 (0.8,	4.0 (4.0, 4.1)	2 (1.9, 2.1)	2.2 (2.2, 2.3)	2.5 (2.2,	3.1 (2.7,	2.7 (2.6,	1.2 (1.1, 1.3)	1.0 (1.0, 1.1)

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

Table 3 (*continued*)

	Denmark		Spain		Italy			United Kingdom	
	AUH	SSI	BIFAP	SIDIAP	PEDIANET	Val Padana	Tuscany	THIN	RCGP
	0.8)				2.9)	3.4)	2.8)		
IR $_{5\text{-}14}$	0.1 (0.1, 0.1)	0.5 (0.5, 0.5)	0.2 (0.2, 0.2)	0.3 (0.3, 0.3)	0.5 (0.4, 0.6)	0.4 (0.3, 0.5)	0.3 (0.3, 0.4)	0.6 (0.6, 0.7)	0.5 (0.5, 0.5)
IR $_{15\text{-}24}$	0.1 (0.1, 0.1)	0.5 (0.5, 0.5)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)		0.1 (0, 0.1)	0.1 (0, 0.1)	0.7 (0.7, 0.8)	0.6 (0.6, 0.7)
IR $_{25\text{-}44}$	0.1 (0.1, 0.1)	0.4 (0.4, 0.5)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)		0 (0, 0)	0 (0, 0)	0.5 (0.5, 0.5)	0.5 (0.4, 0.5)
IR $_{45\text{-}64}$	0.2 (0.2, 0.2)	0.6 (0.6, 0.6)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)		0 (0, 0.1)	0 (0, 0)	0.5 (0.5, 0.6)	0.5 (0.4, 0.5)
IR $_{65+}$	0.4 (0.4, 0.4)	0.7 (0.7, 0.7)	0.2 (0.2, 0.2)	0.1 (0.1, 0.1)		0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.8 (0.8, 0.9)	0.9 (0.8, 0.9)
Generalised convulsions									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	560; <5	2493; 1; 4; 0; 11; 3; 2	47; 4; 22; 15; 42; 26; 44	55; 4; 1; 1; 0; 0; 0	4; 0; 0	40; 0; 1; 0; 0; 0; 0	84, 4, 0, 0, 0, 0	106; 16; 8; 10; 15; 12; 9	75; 14; 26; 41; 59; 60; 87
IR$_{0\text{-}1}$	0.2 (0.2, 0.2)	1 (1, 1.1)	0.1 (0, 0.1)	0 (0, 0.1)	0 (0, 0.1)	0.5 (0.3, 0.7)	0.1 (0.1, 0.1)	0.1 (0.1, 0.2)	0.1 (0.1, 0.2)
IR$_{2\text{-}4}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0.01)	0 (0, 0)	0 (0, 0)
IR $_{5\text{-}14}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{15\text{-}24}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{25\text{-}44}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{45\text{-}64}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{65+}$	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Febrile convulsions									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	10,075; 3529; 1116; 1255; 1523; 1460; 964	36,518; 12,308; 1867; 88; 159; 188; 142	2747; 1531; 206; 32; 23; 10; 29	5233; 2971; 373; 35; 43; 11; 18	243; 201; 24	464; 289; 22; 1; 0; 2; 3	3859, 2574, 520, 22, 23, 34, 53	3388; 1755; 223; 11; 23; 19; 19	2960; 1519; 235; 31; 26; 16; 14
IR$_{0\text{-}1}$	3.1 (3.1, 3.2)	14.9 (14.8, 15.1)	3.9 (3.7, 4)	4.5 (4.4, 4.6)	2.9 (2.6, 3.3)	5.6 (5.1, 6.1)	4.8 (4.7, 5)	4.5 (4.4, 4.7)	5.1 (4.9, 5.3)
IR$_{2\text{-}4}$	0.8 (0.7, 0.8)	3.6 (3.5, 3.6)	1.7 (1.6, 1.8)	1.8 (1.8, 1.9)	1.8 (1.6, 2.1)	2.5 (2.2, 2.8)	2.2 (2.2, 2.3)	1.8 (1.7, 1.9)	1.9 (1.8, 2)
IR $_{5\text{-}14}$	0.1 (0.1, 0.1)	0.2 (0.2, 0.2)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.2)	0.1 (0, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)
IR $_{15\text{-}24}$	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{25\text{-}44}$	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{45\text{-}64}$	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR $_{65+}$	0.2 (0.2, 0.2)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Hypotonic hyporesponsive episode (HHE) (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	394; 88; 482; 1342; 1916; 3069; 3454	1449; 231; 105; 19; 39; 64; 77	661; 375; 669; 216; 258; 277; 752	594; 312; 540; 117; 168; 223; 542	189; 164; 191	25; 3; 3; 3; 9; 4; 38	290, 43 56, 18, 45, 76, 156	510; 390; 595; 281; 578; 852; 2576	242; 210; 225; 91; 187; 234; 607
IR$_{0\text{-}1}$	0.1 (0.1, 0.1)	0.6 (0.6, 0.6)	0.9 (0.9, 1)	0.5 (0.5, 0.6)	2.3 (2, 2.6)	0.3 (0.2, 0.4)	0.4 (0.3, 0.4)	0.7 (0.6, 0.7)	0.4 (0.4, 0.5)
IR$_{2\text{-}4}$	0 (0, 0)	0.1 (0.1, 0.1)	0.4 (0.4, 0.5)	0.2 (0.2, 0.2)	1.5 (1.3, 1.7)	0 (0, 0.1)	0.0 (0.0, 0.0)	0.4 (0.4, 0.4)	0.3 (0.2, 0.3)
IR $_{5\text{-}14}$	0 (0, 0)	0 (0, 0)	0.2 (0.2, 0.3)	0.1 (0.1, 0.1)	0.9 (0.7, 1)	0 (0, 0)	0.0 (0.0, 0.0)	0.2 (0.2, 0.2)	0.1 (0.1, 0.1)
IR $_{15\text{-}24}$	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0.0)	0.1 (0.1, 0.1)	0 (0, 0)
IR $_{25\text{-}44}$	0.1 (0.1, 0.1)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.1)	0 (0, 0)
IR $_{45\text{-}64}$	0.2 (0.2, 0.2)	0 (0, 0)	0 (0, 0)	0 (0, 0)		0 (0, 0)	0.0 (0, 0.0)	0.1 (0.1, 0.1)	0 (0, 0)
IR $_{65+}$	0.6 (0.6, 0.6)	0 (0, 0)	0.2 (0.2, 0.2)	0.1 (0.1, 0.1)		0 (0, 0.1)	0.0 (0.0, 0.0)	0.5 (0.5, 0.5)	0.2 (0.2, 0.2)
Fever (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	10,124; 7641; 2573; 2640; 4320; 5472;	32,524; 11,401; 9265; 5759; 15,253; 21,899; 31,595	144,568; 61,333; 80,732; 52,934; 74,669; 46,983;	117,893; 74,961; 82,523; 24,069; 72,888; 40,726;	19,010; 7859; 6206	136; 48; 105; 122; 393; 647; 1566	18,502; 11,866; 10,103; 4037; 9915; 9265;	60,896; 34,036; 23,417; 6166; 17,461; 15,998;	45,615; 26,309; 19,591; 7639; 21,212; 21,439;

(*continued on next page*)

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

Table 3 (*continued*)

	Denmark		Spain		Italy			United Kingdom	
	AUH	SSI	BIFAP	SIDIAP	PEDIANET	Val Padana	Tuscany	THIN	RCGP
	9184		42,959	35,704			23,459	15,484	16,275
IR_{0-1}	2.4 (2.3, 2.4)	13.3 (13.2, 13.5)	252.3 (251, 253.6)	111.2 (110.6, 111.8)	296.6 (292.4, 300.8)	1.6 (1.4, 1.9)	23.6 (23.2, 24)	87.9 (87.2, 88.6)	84.7 (83.9, 85.4)
IR_{2-4}	0.5 (0.5, 0.6)	3.3 (3.2, 3.3)	126.6 (125.6, 127.6)	57.5 (57.1, 57.9)	157.5 (154.1, 161)	0.4 (0.3, 0.5)	10.7 (10.5, 10.9)	43.9 (43.5, 44.4)	41.2 (40.7, 41.7)
IR_{5-14}	0.2 (0.2, 0.2)	0.8 (0.8, 0.8)	43.5 (43.2, 43.8)	19.2 (19, 19.3)	65.9 (64.3, 67.6)	0.3 (0.2, 0.3)	2.6 (2.6, 2.7)	8.7 (8.6, 8.9)	9.2 (9.1, 9.3)
IR_{15-24}	0.2 (0.2, 0.2)	0.5 (0.5, 0.5)	11.5 (11.4, 11.6)	5.1 (5.1, 5.2)		0.3 (0.3, 0.4)	1 (1, 1)	2.1 (2.1, 2.2)	3.3 (3.2, 3.4)
IR_{25-44}	0.3 (0.3, 0.3)	0.6 (0.6, 0.6)	8.3 (8.2, 8.4)	4.5 (4.5, 4.5)		0.3 (0.3, 0.4)	0.8 (0.7, 0.8)	2.2 (2.2, 2.2)	3.3 (3.3, 3.4)
IR_{45-64}	0.7 (0.7, 0.7)	0.9 (0.8, 0.9)	7 (7, 7.1)	3.3 (3.3, 3.3)		0.5 (0.5, 0.6)	0.7 (0.7, 0.7)	2.2 (2.1, 2.2)	3.7 (3.6, 3.7)
IR_{65+}	1.8 (1.8, 1.9)	2.1 (2, 2.1)	10.3 (10.2, 10.4)	4.1 (4, 4.1)		1.5 (1.5, 1.6)	1.9 (1.9, 2.0)	3.2 (3.1, 3.2)	4.9 (4.9, 5)
Injection site reactions (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	1213; 990; 2360; 6790; 11,949; 13,295; 9230	1606; 883; 1866; 2375; 7991; 10,977; 9669	1938; 1709; 4704; 11,999; 23,025; 20,537; 12,792	3231; 2750; 9456; 14,522; 54,857; 59,408; 80,976	231; 376; 622	3; 1; 15; 12; 18; 22; 40	426, 124, 349, 243, 788,1191, 2926	2086; 1113; 3777; 7120; 19,597; 18,361; 13,158	808; 468; 1698; 2846; 8429; 8540; 5803
IR_{0-1}	0.4 (0.4, 0.4)	0.7 (0.6, 0.7)	2.7 (2.6, 2.9)	2.8 (2.7, 2.9)	2.8 (2.4, 3.2)	0 (0, 0.1)	0.5 (0.5, 0.6)	2.8 (2.7, 2.9)	1.4 (1.3, 1.5)
IR_{2-4}	0.2 (0.2, 0.2)	0.2 (0.2, 0.3)	1.9 (1.8, 2)	1.7 (1.6, 1.7)	3.4 (3.1, 3.8)	0 (0, 0)	0.1 (0.1, 0.1)	1.2 (1.1, 1.2)	0.6 (0.5, 0.6)
IR_{5-14}	0.2 (0.1, 0.2)	0.2 (0.1, 0.2)	1.7 (1.7, 1.8)	1.9 (1.8, 1.9)	2.8 (2.6, 3.1)	0 (0, 0.1)	0.1 (0.1, 0.1)	1.2 (1.1, 1.2)	0.7 (0.6, 0.7)
IR_{15-24}	0.4 (0.3, 0.4)	0.2 (0.2, 0.2)	2.4 (2.4, 2.5)	3 (2.9, 3)		0 (0, 0.1)	0.1 (0.1, 0.1)	2.4 (2.3, 2.4)	1.2 (1.1, 1.2)
IR_{25-44}	0.6 (0.6, 0.6)	0.3 (0.3, 0.3)	2.5 (2.4, 2.5)	3.3 (3.3, 3.4)		0 (0, 0)	0.1 (0.1, 0.1)	2.5 (2.4, 2.5)	1.3 (1.3, 1.3)
IR_{45-64}	1 (1, 1)	0.4 (0.4, 0.4)	3 (3, 3)	4.8 (4.7, 4.8)		0 (0, 0)	0.1 (0.1, 0.1)	2.5 (2.5, 2.5)	1.4 (1.4, 1.5)
IR_{65+}	1.7 (1.6, 1.7)	0.6 (0.6, 0.6)	3 (2.9, 3)	9.4 (9.3, 9.4)		0 (0, 0.1)	0.2 (0.2, 0.2)	2.7 (2.7, 2.8)	1.7 (1.7, 1.8)
Somnolence (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)	22; 23; 89; 551; 1497; 2145; 2839	31; 33; 105; 560; 2080; 2552; 1559	82; 92; 93; 660; 1481; 1428; 2989	79; 69; 117; 256; 1033; 1539; 2606	31; 19; 23	0; 0; 0; 1; 4; 4; 1	130; 102; 171; 79; 199; 412; 2853	449; 214; 184; 280; 1107; 1771; 5316	268; 102; 115; 244; 719; 896; 2819
IR_{0-1}	0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.4 (0.3, 0.5)	0 (0, 0)	0.2 (0.1, 0.2)	0.6 (0.5, 0.7)	0.5 (0.4, 0.5)
IR_{2-4}	0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.1)	0 (0, 0.1)	0.2 (0.1, 0.3)	0 (0, 0)	0.1 (0.1, 0.1)	0.2 (0.2, 0.3)	0.1 (0.1, 0.2)
IR_{5-14}	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.2)	0 (0, 0)	0 (0, 0.1)	0.1 (0, 0.1)	0 (0, 0.1)
IR_{15-24}	0 (0, 0)	0 (0, 0.1)	0.1 (0.1, 0.1)	0.1 (0, 0.1)		0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)
IR_{25-44}	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)	0.2 (0.1, 0.2)	0.1 (0.1, 0.1)		0 (0, 0)	0 (0, 0)	0.1 (0.1, 0.1)	0.1 (0.1, 0.1)
IR_{45-64}	0.2 (0.1, 0.2)	0.1 (0.1, 0.1)	0.2 (0.2, 0.2)	0.1 (0.1, 0.1)		0 (0, 0)	0 (0, 0)	0.2 (0.2, 0.2)	0.1 (0.1, 0.2)
IR_{65+}	0.5 (0.5, 0.5)	0.1 (0.1, 0.1)	0.7 (0.7, 0.7)	0.3 (0.3, 0.3)		0 (0, 0)	0.2 (0.2, 0.2)	1.1 (1, 1.1)	0.8 (0.8, 0.9)
Persistent crying (codes)									
Number of events by age group (0–1; 2–4; 5–14, 15–24, 45–64 65+)			17,852; 1004; 199; 893; 2197; 1685; 1276		1985; 203; 15	2; 0; 0; 0; 0; 0; 0	2098, 143, 43, 16, 20, 21, 22	3366; 272; 58; 74; 199; 144; 76	4299; 262; 55; 44; 103; 83; 40
IR_{0-1}			25.9 (25.5, 26.3)		24.6 (23.6, 25.7)	0 (0, 0.1)	2.6 (2.5, 2.7)	4.5 (4.4, 4.7)	7.5 (7.2, 7.7)
IR_{2-4}			1.2 (1.1, 1.2)		1.9 (1.7, 2.2)	0 (0, 0)	0.1 (0.1, 0.1)	0.3 (0.2, 0.3)	0.3 (0.3, 0.4)
IR_{5-14}			0.1 (0.1, 0.1)		0.1 (0, 0.1)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR_{15-24}			0.2 (0.2, 0.2)			0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR_{25-44}			0.2 (0.2, 0.2)			0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR_{45-64}			0.2 (0.2, 0.3)			0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
IR_{65+}			0.3 (0.3, 0.3)			0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

Table 4
Pertussis vaccine coverage estimated during database fingerprint (all ages).

	Denmark		Spain		Italy			United Kingdom	
	AUH	SSI	BIFAP	SIDIAP	PEDIANET	Val Padana	Tuscany	THIN	RCGP
Total persons (all ages) no selection in input file	2,563,188	7,512,032	7,550,262	7,096,695	9708	454,188	4,837,755	11,696,261	2,678,749
Recommended acellular pertussis vaccine schedule	3,5,12 months, 5 years		2,4, 11 months, 6 years		3,5,11 months, 6 years, 12-18 years, pregnancy (Tdap)			2,3,4 months, 3 years,15-45 pregnancy	
Exposure									
Number of aP doses recorded (all ages)	751,629	4,089,505	1,750,171	2,846,877	36,124	90,713	403,793	3,295,699	990,135
Number of wP doses of (all ages)	0	22,565	33,588	0	0	0	0	5,986,936	909,606
aP-containing vaccine combinations (all population)	Not specified	1-valent: 573 3-valent: 26,344 4-valent: 1,701,864 5-valent: 2,179,619 6-valent: 116,913	3-valent: 368,345 4-valent: 643 5-valent: 828,067 6-valent: 548,166	1-valent: 28,383 2-valent: 587,032 5-valent: 705,940 6-valent: 1,525,522	Not supplied	3-valent: 10,023 4-valent: 12,404 5-valent: 500 6-valent: 67,786	Not submitted	Not submitted	1-valent: 140,911 3-valent: 594 4-valent: 224,245 5-valent: 624,385
Dose recorded or derived	Doses recorded	Doses recorded	Doses recorded	Doses recorded	Doses recorded	All doses derived	Doses recorded	Dose recorded	All doses derived
Coverage estimation selected population									
Number of persons eligible for coverage analysis	188,553	1,219,732	293,170	467,960	9,708	18,759	335,062	468,897	46,240
Percentage coverage for D1 aP (CDF) age 12 months by selected birth-year									
2006	86	90	92	98	94	71		97	95
2007	90	91	93	97	94	92		97	93
2008	92	92	95	96		98		97	94
2009	92	93	95	96		98		97	94
2010	95	94	90	97		97		97	94
2011	96	94	95	96				98	94
2012	95	94	96	96				98	95
Percentage coverage for D2 aP (CDF) age 12 months									
2006	89	88	94	97	93	59		96	94
2007	91	89	94	96	94	87		96	92
2008	92	91	94	95		94		96	93
2009	92	91	94	95		93		96	94
2010	93	92	94	95		94		97	92
2011	94	92	94	95				97	93
2012	94	92	95	95				97	94

D1: dose 1; D2: dose 2, CDF: cumulative distribution function; aP: acellular pertussis vaccine; wP: whole cell pertussis vaccine.

data is transformed into a CDM, independently of the study. Although the use of a deep common data model, such as that used in Sentinel/PRISM, VSD or the OMOP common data models would enable a broader standardised assessment of quality, its cost makes it currently not feasible in Europe, due to the absence of sustainable public funding for distributed healthcare data networks [33,34]. This might change in the future with the European Health Data & Evidence Network (EHDEN) project funded by the Innovative Medicines Initiative, which aims to convert European data for 100 million individuals into the OMOP common data model [35]. The initial quality assessment in VSD involved checking the data against the individual paper records, which was not feasible in ADVANCE given the resources provided. We, therefore, compared the data between the databases and with external benchmarks, similar to the approach that was taken in PRISM [9]. In addition to the events relevant for pertussis studies, autoimmune disorders of interest and vaccine preventable diseases were characterised to create further preparedness, these will be described separately. We currently cannot provide traditional estimates of validity (e.g. sensitivity, positive predictive value) because we benchmarked our data on a population level. Chart validation studies on the investigated events had been done in other contexts in some of the databases (Table S1), showing that in Denmark positive predictive values were high for ICD10 pertussis code A37 against laboratory confirmed cases as well as in Spain for ICD9 (037) and ICPC (R71.1) codes. The PPV for ICD-10 code febrile convulsions (R560) was more than 92% in Denmark in prior hospitalization validation studies. Pneumonia hospitalization code ICD10 J12-J18 had PPV of more than 92% in prior validation studies of hospitalizations in Denmark, and were above 70% in the BIFAP for ICPC and ICD-9 codes (see Table S1). Medical chart validation studies may be done in each of the participating data sources, but this requires considerable resources and ethical approval, and there was no budget available for this task within the IMI-ADVANCE funding. Future vaccine effect studies should assure funding to assess the validity of the event codes on an individual case level. Additional limitations are the lack of information on demographic variables other than age and gender to assess representativeness of the populations, ethnicity and social economic status is not generally recorded due its sensitivity. Moreover, we only assessed the population, vaccine and event data from four countries of relatively high income countries in Europe. With the increasing computerization of health care and vaccination registries we recommend that the follow-up of ADVANCE: the VAccine monitoring Collaboration for EUrope (VAC4EU) includes additional countries (preferable eastern) to become more representative of Europe and cover brands/vaccines provided in those countries as well.

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

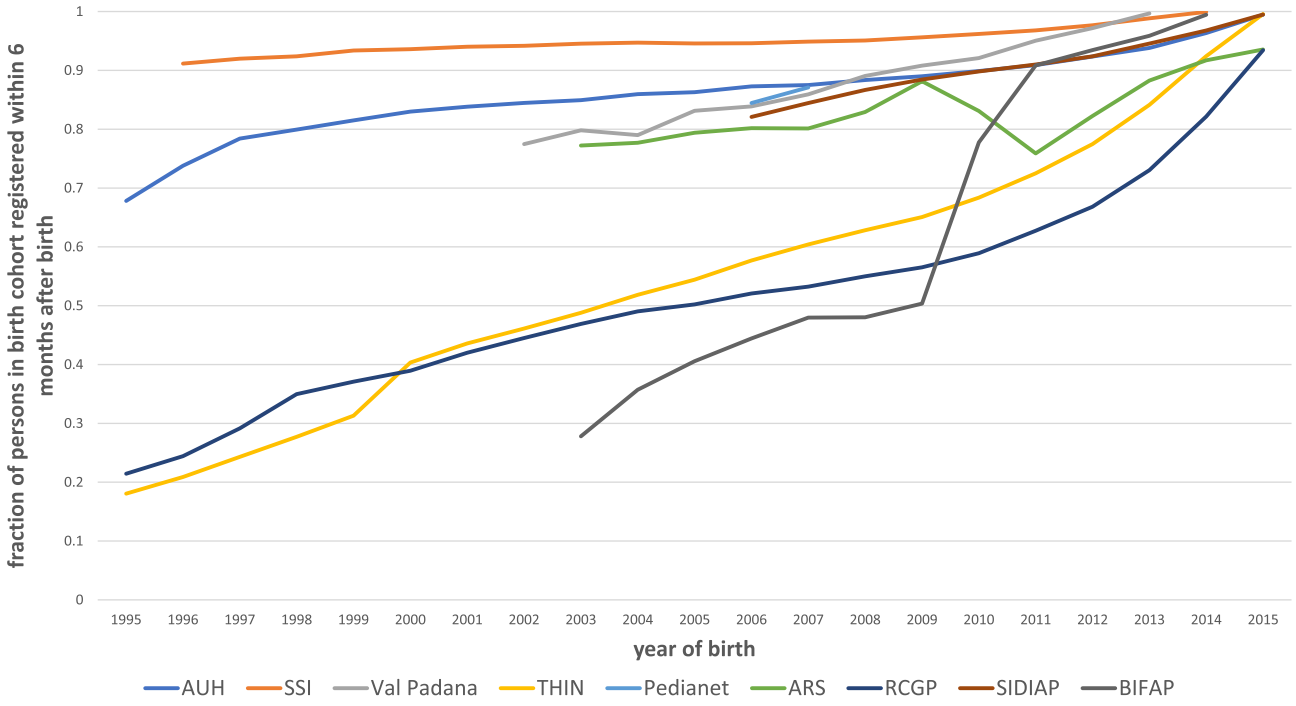

Fig. 3. Percentage of children registered within six months after birth by year of birth to verify completeness of follow-up per birth cohort.

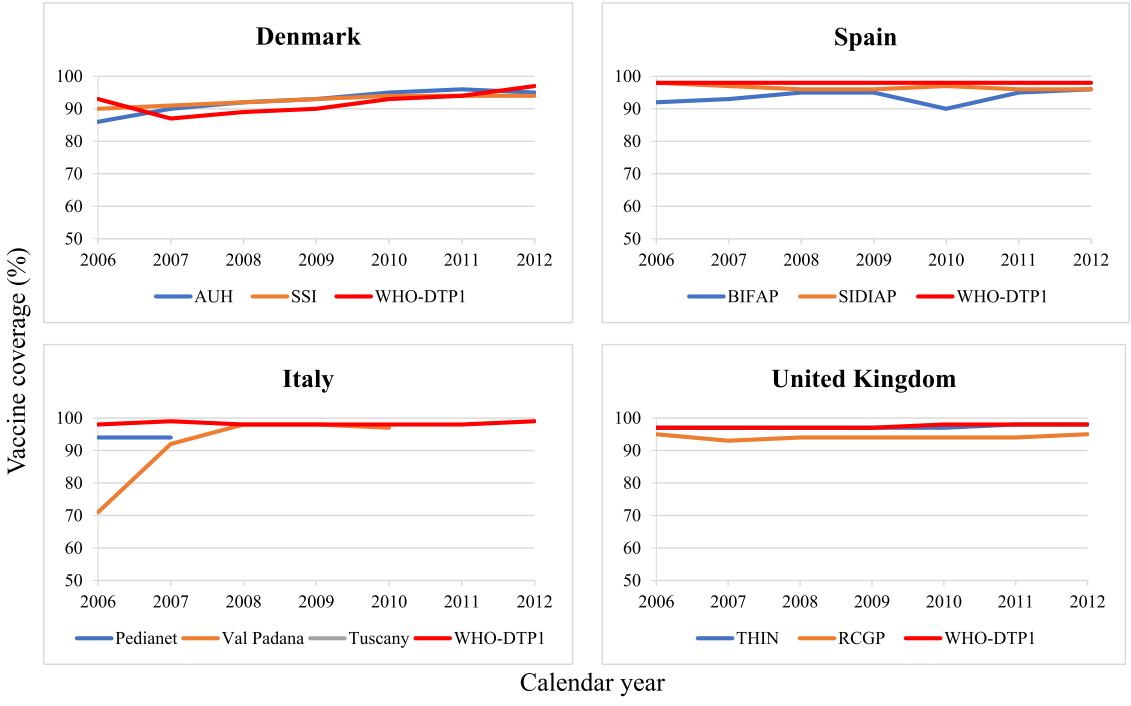

Fig. 4. Percentage of pertussis vaccine coverage estimate for dose 1 of acellular pertussis vaccine by calendar year, compared with national estimates for DTaP vaccine reported to the World Health Organisation.

We conclude that we have developed a database characterisation workflow model, based on the ADVANCE common data model, which enables transparent fit-for-purpose assessment of databases for specific studies, which was tested in the pertussis vaccination POC studies described in this supplement [18–24]. The database characterisation provided a detailed description of the structure of the population covered by the health care databases, the coding and incidence rates of events, as well as exposure to vaccines and vaccination coverage. Our system enabled comparisons between database and comparisons with external benchmarks, providing indicators of completeness of data on a population level. We showed that the older data in databases may be less reliable than more recent data and that both age and calendar years patterns need to be inspected prior to deciding which period is fit-for-purpose. The ADVANCE project, which has evolved into the Vaccine monitoring collaboration for Europe (VAC4EU) will aim to make

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100

the data for event rates and coverage estimates available in an interactive tool which would enable rapid and transparent inspection of the feasibility of databases to participate in specific studies [36]. The workflow that was created can easily be extended to other vaccines and events, to increase readiness and speed in conducting protocol-based vaccine coverage, benefit and risk studies.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Toon Braeye, Lieke van der Aa, Giorgia Danieli, Caitlin Dodd, Talita Duarte-Salles, Hanne- Dorthe Emborg, Marius Gheorge, Consuelo Huerta, Elisa Martín-Merino, Chris McGee, Gino Picelli Giuseppe Roberto, Lara Tramontan, Marco Villa and Daniel Weibel declared no potential conflicts of interest. Miriam Sturkenboom, declared that she has received grants from Novartis, CDC and Bill & Melinda Gates Foundation, for work unrelated to the submitted work. Johnny Kahlert declared that he is employed at Department of Clinical Epidemiology, Aarhus University Hospital, which is involved in studies with research grants from various companies that are administered by Aarhus University; he does not receive fees, honoraria, grants or consultancies personally. Simon de Lusignan declared that he is Director of RCGP RSC that he has received funding through his University for vaccine-related studies from GSK and Seqirus and he is also a member of advisory groups for Seqirus and Sanofi. Lina Titievsky declared that she is employed by Pfizer and holds company stocks/shares. Rosa Gini's institution is involved in studies funded by Novartis, Eli Lilly, and Daiichi, compliant with the ENCePP CoC".

Acknowledgements

The authors would like to acknowledge contributions from Maria de Ridder in creating SAS programs for the pooling of data and calculation of follow-up, Klara Berensci for initial data extraction from the AUH database, and Benedikt Becker for assistance with the Codemapper tool. They would also like to acknowledge editorial services from Margaret Haugh, MediCom Consult, Villeurbanne, France.

Funding source

The Innovative Medicines Initiative Joint Undertaking funded this project under ADVANCE grant agreement n° 115557, composed of a financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and in kind contributions from EFPIA member companies.

Disclaimer

The views expressed in this article are the personal views of the authors and should not be understood or quoted as being made on behalf of or reflecting the position of the agencies or organisations with which the authors are affiliated.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2020.01.100.

References

[1] Andrews EB, Margulis AV, Tennis P, West SL. Opportunities and challenges in using epidemiologic methods to monitor drug safety in the era of large automated health databases. Curr Epidemiol Rep 2014;1:194–205.

[2] Sturkenboom Miriam, Bahri Priya, Chiucchiuini Antonella, Grove Krause Tyra, Hahné Susan, Khromava Alena, Kokki Maarit, Kramarz Piotr, Kurz Xavier, Larson Heidi J, de Lusignan Simon, Mahy Patrick, Torcel-Pagnon Laurence, Titievsky Lina, Bauchau Vincent. Why we need more collaboration in Europe to enhance post-marketing surveillance of vaccines. Vaccine 2019. https://doi.org/10.1016/j.vaccine.2019.07.081.

[3] Platt R, Brown JS, et al. The FDA sentinel initiative — an evolving national resource. NEJM 2018;379:2091–3.

[4] Pacurariu A, Plueschke K, McGettigan P, Morales DR, Slattery J, Vogl D, et al. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. BMJ Open 2018;8:e023090.

[5] Plueschke K, McGettigan P, Pacurariu A, Kurz X, Cave A. EU-funded initiatives for real world evidence: descriptive analysis of their characteristics and relevance for regulatory decision-making. BMJ Open 2018;8:e021864.

[6] Schneeweiss S, Glynn RJ. Real-world data analytics fit for regulatory decision-making. Am J Law Med 2018;44:197–217.

[7] Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. Pharmacoepidemiol Drug Saf 2017;26:1018–32.

[8] DeStefano F. The Vaccine Safety Datalink project. Pharmacoepidemiol Drug Saf 2001;10:403–6.

[9] Baker MA, Nguyen M, Cole DV, Lee GM, Lieu TA. Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. Vaccine 2013;31:K98–K112.

[10] Mullooly J, Drew L, DeStefano F, Chen R, Okoro K, Swint E, et al. Quality of HMO vaccination databases used to monitor childhood vaccine safety. Vaccine Safety DataLink Team. Am J Epidemiol 1999;149:186–94.

[11] Weng C, Appelbaum P, Hripcsak G, Kronish I, Busacca L, Davidson KW, et al. Using EHRs to integrate research with patient care: promises and challenges. J Am Med Inform Assoc 2012;19:684–7.

[12] Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;58:323–37.

[13] Carroll R, Ramagopalan SV, Cid-Ruzafa J, Lambrelli D, McDonald L. An analysis of characteristics of post-authorisation studies registered on the ENCePP EU PAS Register. F1000Research 2017;6:1447.

[14] Burnum JF. The misinformation era: the fall of the medical record. Ann Intern Med 1989;110:482–4.

[15] Shapiro S. The role of automated record linkage in the postmarketing surveillance of drug safety: a critique. Clin Pharmacol Ther 1989;46:371–86.

[16] ADVANCE; 2018 [last accessed 20 June 2018]. Available from: http://www.advance-vaccines.eu/.

[17] Sturkenboom M, van der Aa L, Bollaerts K, Emborg HD, Ferreira G, Gino R, et al. The ADVANCE distributed network system for evidence generation on vaccines coverage, benefits and risks based on electronic health care data. Vaccine 2020. Submitted to Vaccine.

[18] Emborg HD, Kahlert J, Braeye T, Bauwens J, Bollaerts K, Danieli G, et al. ADVANCE system testing: can coverage of pertussis vaccination be estimated in EU countries using electronic health data: an example. Vaccine 2019. https://doi.org/10.1016/j.vaccine.2019.07.039 (in press).

[19] Tin Tin Htar M, de Ridder M, Braeye T, Correa A, de Lusignan S, Duarte-Salles T, et al. ADVANCE system testing: vaccine benefit studies by using multi-country electronic health data: An example on pertussis vaccination. Vaccine 2020. https://doi.org/10.1016/j.vaccine.2019.08.078 (in press).

[20] Weibel D, Dodd C, Mahaux O, Haguinet F, de Smedt T, Duarte-Salles T, et al. ADVANCE system testing: can safety studies be conducted using electronic health data: an example with pertussis vaccination?. Vaccine 2020. https://doi.org/10.1016/j.vaccine.2019.06.040 (in press).

[21] Dodd C, Bollaerts K, Weibel D, de Ridder M, Mahaux O, Haguinet F, et al. ADVANCE system testing: estimating the incidence of adverse events following pertussis vaccination in healthcare databases with incomplete exposure data. Vaccine 2020. Submitted to Vaccine.

[22] Gini R, Dodd C, Bollaerts K, Bartolini C, Roberto G, Huerta-Alvarez C, et al. Quantifying outcome misclassification in multi-database studies: the case study of pertussis in the ADVANCE project. Vaccine 2020. https://doi.org/10.1016/j.vaccine.2019.07.045 (in press).

[23] Bollaerts K, Ledent E, de Smedt T, Weibel D, Emborg HD, Correa A, et al. ADVANCE system testing: benefit-risk analysis of a marketed vaccine using multi-criteria decision analysis and cohort modelling. Vaccine 2020. https://doi.org/10.1016/j.vaccine.2019.09.034 (in press).

[24] Bollaerts K, de Smedt T, McGee C, Emborg HD, Villa M, Alexandridou M, et al. ADVANCE: towards near real-time monitoring of vaccination coverage, benefits and risk using European electronic health record databases. Vaccine 2020. https://doi.org/10.1016/j.vaccine.2019.08.012 (in press).

[25] Becker BFH, Avillach P, Romio S, van Mulligen EM, Weibel D, Sturkenboom M, et al. CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. Pharmacoepidemiol Drug Saf 2017;26(8):998–1005.

[26] Trifiro G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how?. J Intern Med 2014;275:551–61.

[27] Jerboa home page; 2018 [last accessed 28 March 2019]. Available from: https://github.com/mi-erasmusmc/Jerboa/wiki.

[28] Bollearts K, Weil J, WP4 Working Group. D4.4 Report on tested methods for accelerated assessment of vaccination coverage, vaccine benefits, risks and

benefit-risk; 2017 [last accessed 1 April 2019]. Available from: http://www.advance-vaccines.eu/app/archivos/publicacion/19/ADVANCE_WP4_D4.4_submitted_.pdf.

[29] World Healh Organisation. WHO-UNICEF estimates of DTP3 coverage; 2018 [last accessed 29 March 2019]. Available from: http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragedtp3.html.

[30] Sammon CJ, Charlton RA, Snowball J, Weil JG. The incidence of childhood and adolescent seizures in the UK from 1999 to 2011: a retrospective cohort study using the Clinical Practice Research Datalink. Vaccine 2015;33:7364–9.

[31] Shi T, Denouel A, Tietjen AK, Lee JW, Falsey AR, Demont C, et al. Global and tegional burden of hospital admissions for pneumonia in older adults: a systematic review and meta-analysis. J Infect Dis 2019.

[32] Rudan I. Epidemiology and etiology of childhood pneumonia. Bull World Health Organ 2008;86:408–16.

[33] Gini R, Schuemie M, Brown J, Ryan P, Vacchi E, Coppola M, et al. Data extraction and management in networks of observational health care databases for scientific research: a comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE strategies. EGEMS (Washington, DC) 2016;4:1189.

[34] Sturkenboom M. Advancing collaborative vaccine benefits and safety research in Europe via the ADVANCE code of conduct. Vaccine 2018;36:194–5.

[35] European Health Data & Evidence Network (EHDEN); 2019 [last accessed 29 March 2019]. Available from: https://www.ehden.eu/.

[36] Vac4Eu. Toolbox: benefits and risk; 2018 [last accessed 26 March 2019]. Available from: https://vac4eu.org/brrota/.

Please cite this article as: M. Sturkenboom, T. Braeye, L. van der Aa et al., ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations, Vaccine, https://doi.org/10.1016/j.vaccine.2020.01.100