Contents lists available at ScienceDirect

Vaccine

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

Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project

Rosa Gini[a,*], Caitlin N. Dodd[b,c], Kaatje Bollaerts[d], Claudia Bartolini[a], Giuseppe Roberto[a], Consuelo Huerta-Alvarez[e], Elisa Martín-Merino[e], Talita Duarte-Salles[f], Gino Picelli[g], Lara Tramontan[g,h], Giorgia Danieli[g,h], Ana Correa[i], Chris McGee[i,j], Benedikt F.H. Becker[b], Charlotte Switzer[k,1], Sonja Gandhi-Banga[k], Jorgen Bauwens[l,m,n], Nicoline A.T. van der Maas[m,n], Gianfranco Spiteri[o], Emmanouela Sdona[o,2], Daniel Weibel[b], Miriam Sturkenboom[c,d,p]

[a] *Agenzia regionale di sanità della Toscana, Osservatorio di epidemiologia, Florence, Italy*
[b] *Erasmus University Medical Center, Post Box 2040, 3000 CA Rotterdam, Netherlands*
[c] *Julius Global Health, University Medical Center, Utrecht, Heidelberglaan 100, the Netherlands*
[d] *P95 Epidemiology and Pharmacovigilance, Koning Leopold III laan 1, 3001 Heverlee, Belgium*
[e] *BIFAP Database, Spanish Agency of Medicines and Medical Devices, Madrid, Spain*
[f] *Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain*
[g] *Epidemiological Information for Clinical Research from an Italian Network of Family Paediatricians (PEDIANET), Padova, Italy*
[h] *Consorzio Arsenal.IT, Veneto Region, Italy*
[i] *University of Surrey, Guildford, Surrey GU2 7XH, UK*
[j] *Royal College of General Practitioners, Research and Surveillance Centre, 30 Euston Square, London NW1 2FB, UK*
[k] *Sanofi Pasteur, 1755 Steeles Ave W, North York, ON M2R 3T4, Canada*
[l] *University Children's Hospital, Basel, Switzerland*
[m] *University of Basel, Switzerland*
[n] *Brighton Collaboration Foundation, Switzerland*
[o] *European Centre for Disease Prevention and Control, Gustav III's Boulevard 40, 16973 Solna, Sweden*
[p] *VACCINE.GRID Foundation, Spitalstrasse 33, Basel, Switzerland*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Incidence of pertussis
Event-finding algorithms
Event misclassification
Positive predictive value

ABSTRACT

Background: The Accelerated Development of VAccine beNefit-risk Collaboration in Europe (ADVANCE) is a public-private collaboration aiming to develop and test a system for rapid benefit-risk (B/R) monitoring of vaccines using European healthcare databases. Event misclassification can result in biased estimates. Using different algorithms for identifying cases of *Bordetella pertussis* (BorPer) infection as a test case, we aimed to describe a strategy to quantify event misclassification, when manual chart review is not feasible.
Methods: Four participating databases retrieved data from primary care (PC) setting: BIFAP: (Spain), THIN and RCGP RSC (UK) and PEDIANET (Italy); SIDIAP (Spain) retrieved data from both PC and hospital settings. BorPer algorithms were defined by healthcare setting, data domain (diagnoses, drugs, or laboratory tests) and concept sets (specific or unspecified pertussis). Algorithm- and database-specific BorPer incidence rates (IRs) were estimated in children aged 0–14 years enrolled in 2012 and 2014 and followed up until the end of each calendar year and compared with IRs of confirmed pertussis from the ECDC surveillance system (TESSy). Novel formulas were used to approximate validity indices, based on a small set of assumptions. They were applied to approximately estimate positive predictive value (PPV) and sensitivity in SIDIAP.

Abbreviations: BorPer, Bordetella pertussis; B/R, benefit-risk; ECDC, European Centre for Disease Prevention and Control; IR, incidence rate; PC, primary care; PPV, positive predictive value.

[*] Corresponding author.

E-mail addresses: rosa.gini@ars.toscana.it (R. Gini), kaatje.bollaerts@p-95.com (K. Bollaerts), claudia.bartolini@ars.toscana.it (C. Bartolini), giuseppe.roberto@ars.toscana.it (G. Roberto), chuerta@aemps.es (C. Huerta-Alvarez), emartinm@aemps.es (E. Martín-Merino), tduarte@idiapjgol.org (T. Duarte-Salles), g.picelli@virgilio.it (G. Picelli), ltramontan@consorzioarsenal.it (L. Tramontan), s.lusignan@surrey.ac.uk (A. Correa), c.mcgee@surrey.ac.uk (C. McGee), benedikt.becker@posteo.de (B.F.H. Becker), sonja.banga@sanofi.com (S. Gandhi-Banga), j.bauwens@brightoncollaboration.org (J. Bauwens), nicoline.van.der.maas@rivm.nl (N.A.T. van der Maas), Gianfranco.Spiteri@ecdc.europa.eu (G. Spiteri), d.weibel@erasmusmc.nl (D. Weibel), m.c.j.sturkenboom@umcutrecht.nl (M. Sturkenboom).

[1] Current address: McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada.

[2] Current address: Karolinska Institutet, Solnavägen 1, 171 77 Solna, Sweden.

https://doi.org/10.1016/j.vaccine.2019.07.045
0264-410X/© 2019 The Authors. Published by Elsevier Ltd.

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045

Results: The number of cases and the estimated BorPer IRs per 100,000 person-years in PC, using data representing 3,173,268 person-years, were 0 (IR = 0.0), 21 (IR = 4.3), 21 (IR = 5.1), 79 (IR = 5.7), and 2 (IR = 2.3) in BIFAP, SIDIAP, THIN, RCGP RSC and PEDIANET respectively. The IRs for combined specific/unspecified pertussis were higher than TESSy, suggesting that some false positives had been included. In SIDIAP the estimated IR was 45.0 when discharge diagnoses were included. The sensitivity and PPV of combined PC specific and unspecific diagnoses for BorPer cases in SIDIAP were approximately 85% and 72%, respectively.

Conclusion: Retrieving BorPer cases using only specific concepts has low sensitivity in PC databases, while including cases retrieved by unspecified concepts introduces false positives, which were approximately estimated to be 28% in one database. The share of cases that cannot be retrieved from a PC database because they are only seen in hospital was approximately estimated to be 15% in one database. This study demonstrated that quantifying the impact of different event-finding algorithms across databases and benchmarking with disease surveillance data can provide approximate estimates of algorithm validity.

1. Introduction

ADVANCE is a public-private collaboration aiming to develop and test a system for rapid benefit-risk (B/R) monitoring of vaccines using existing healthcare databases in Europe [1] (see Appendix for list of consortium members). These databases have proven very useful for studying drug effects and are commonly used in pharmacoepidemiology [2].

Identifying events, such as vaccine-preventable diseases, adverse events of interest, co-morbidities and exposure to vaccination, is a pivotal first step in vaccine B/R studies. Since there is limited or no control over the primary data collection when using existing healthcare databases, event retrieval is usually not perfect. Individuals who experienced the event might not be retrieved, for example if an individual is admitted to hospital for the event but no primary care (PC) diagnosis is recorded, the event will not be retrieved from PC databases: those individuals will be false negatives. Conversely, some individuals might be identified as having the event when in fact they did not: those individuals will be false positives. In a PC database, this typically happens when the physician had only a suspicion, or if it was a ruled-out diagnosis, or if a prevalent condition was recorded in a way that may be retrieved as a new diagnosis, or if a diagnosis was miscoded, or if the diagnosis was incorrect.

It is well established that misclassification of events (false positives or false negatives) can introduce bias in epidemiological studies, which can be corrected, to some extent, using statistical methods [5–7]. However, to correct this bias, some validity parameters such as sensitivity and positive predictive value (PPV) are required [8]. For this a gold standard, such as chart reviews, is required, which generally makes it costly and time-consuming.

Researchers who access these databases usually develop their own methods to identify events of interest, which are not always fully transparent [3,4]. Events may be retrieved by combining information from different settings (e.g., PC and hospital) and data domains, for example diagnostic codes, drugs as proxies (e.g. in the case of diabetes), or laboratory measurements. Use of information from more than one data domain, compared with using diagnoses information only, or from more than one setting, such as PC and hospital versus PC alone, can alter the sensitivity and PPV of the event-finding algorithm: indeed broadening the inclusion criteria can reduce the number of false negatives, at the price of possibly increasing the number of false positives. When a gold standard is not accessible to estimate the exact amount of false positives and false negatives associated with the chosen algorithms, a systematic approach to analyse them has the potential to support an approximate estimation of validity.

In an attempt to develop such a systematic approach, the *component algorithm* strategy was introduced and tested [9]: such strategy builds on the design a set of standardized algorithms for the same event, called *components*, are defined and applied in each database. The impact of different algorithms on the resulting estimates of disease occurrence is subsequently measured and compared across sites [9], thus providing qualitative evidence on the amount of false positives and false negatives associated with each algorithm. In this study we aimed to refine this strategy, by further standardizing the process, by using benchmark data from an external reference and by developing and applying novel formulae to turn the observed differences into approximate estimates of validity. Since the proof-of-concept studies of ADVANCE focused on pertussis, we used this event as case study.

2. Methods

2.1. Bordetella pertussis disease information

Bordetella pertussis causes pertussis, a vaccine-preventable infectious disease of the respiratory tract. Symptoms include paroxysms of cough typically lasting from 1 to 6 weeks or more and these may be milder in adolescents or immunised children [10,11]. Several tests are available to confirm *Bordetella pertussis* infection, including culture (which takes up to 14 days), serology and nucleic acid amplification tests. Pertussis is a notifiable infectious disease and cases should be reported to the national surveillance system in all the countries involved in ADVANCE. European Union member states are required to report available data on pertussis cases to the European Centre for Disease Prevention and Control (ECDC). A standardised case definition is used which classifies cases based on clinical, epidemiological and laboratory criteria [12]. All national reports are submitted to the European Surveillance System database (TESSy) managed by the ECDC [13].

2.2. Data sources

We assessed the impact of different event-finding algorithms using five databases that participated in the ADVANCE proof-of-concept studies: BIFAP and SIDIAP (Spain), PEDIANET (Italy) and RCGP RSC and THIN (United Kingdom). All databases were population-based with data from electronic medical records in the PC setting. In SIDIAP, the analyses were restricted to the population in this PC database that could be linked to hospital discharge records. Surveillance data on pertussis were obtained from the TESSy surveillance system through ECDC, a partner of ADVANCE.

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045

2.3. Study population and study design

We used a dynamic cohort study design to study the impact of different event-finding algorithms on the estimated pertussis IRs. Due to the methodological nature of this study, to enable us to explore in more detail a number of strategies, we included a larger cohort in the study population than that in the other ADVANCE studies. Therefore, children aged 0 to 14 years who were registered in the participating databases entered the study cohorts on 1st January 2012 and 1st January 2014, and were followed up during 2012 and 2014, respectively. Children who were born during 2012 or 2014 were followed up from birth until the end of the calendar year. Children who were older than 14 years at any point in the follow-up were excluded. To exclude any previous cases that had been notified before the start of the study period, children who had a record of one of the components of pertussis during the two years prior to one of the cohort entry dates were excluded, unless the component referred to the data domain of drugs (see below for more details on the component definition).

2.4. Selection of component algorithms

A component algorithm is a standardised event-finding algorithm specified by three characteristics: the *setting* of primary data collection (PC or hospital), the *data domain* (diagnosis, drugs, or laboratory tests) involved in the algorithm, and the *set of concepts* used to find the codes used to query the database [9]. In order to create the sets of concepts we built on the process described elsewhere: an initial list was created from the pertussis clinical definition, completed with a literature review [2,13] and was discussed with local experts, who in some cases included free text strings that were deemed to be pertinent; the process was supported by the tool CodeMapper [14]. Labelling and classification of identified concepts, as well as the construction of the components, were conducted by one of the authors who is a pertussis expert (NvdM). As a result, seven concept sets were created (Table 1) [15,16]. In particular, two sets of concepts belonged to the diagnoses data domain: the set labelled '(*Bordetella pertussis*)' included three concepts which specifically indicated *Bordetella pertussis* as the causative agent of the infection, while the set labelled '(pertussis unspecified)' included five concepts indicating unspecified pertussis. The corresponding codes and free text keywords are given in Supplementary Table 1.

The primary components associating concepts with settings (PC and hospital) are described in Table 2. Some secondary components, combining primary components in pre-defined temporal relations (e.g., symptoms in the presence of a drug prescription in the previous 30 days) were also created.

2.5. Analysis

Each database manager received an R-coded programme (quality checked by double-coding against Stata) which was programmed using the pre-specified common data model [1]. These programmes produced aggregated outputs, which were then transferred to the remote research environment. Event-finding algorithms were created as logical combinations of individual components using Boolean operators OR and AND: the combination of two components with the 'OR' operator allows to select those subjects who are positive for *either components*; the combination of two components with the 'AND' operator allows to select those subjects who are positive for *both components*. For example, the two components 'PC diagnosis, specific' and 'PC diagnosis, unspecified' were combined in one component: 'PC specific OR unspecified diagnoses', which detected all individuals that had any PC diagnosis, regardless of specificity. Based on the different event-finding algorithms, incidence rates (IRs) were estimated using the number of persons retrieved with the respective events as numerator and the follow-up person-time as denominator (see Supplementary File 1). Exact Poisson confidence intervals (95% CI) were calculated [18].

Age and country-specific incidences per 100,000 person-years of confirmed *Bordetella Pertussis* cases notified to the TESSy surveillance system in both 2012 and 2014 were calculated for children

Table 1
Sets of concepts selected for the component algorithms. Each set of concepts has a description and can contain one or more concepts. Each concept has a description and, if available, a Concept Unique Identifier (CUI) of the Unified Medical Language System.

Concept set label	Concept set description	Concept	CUI
(*Bordetella pertussis*)	Concepts referring to diagnoses specifically mentioning pertussis induced by an infection of *Bordetella pertussis*	*Bordetella pertussis*	C0043167
		Whooping cough due to *Bordetella pertussis* without pneumonia	C2887068
		Whooping cough due to *Bordetella pertussis* with pneumonia	C2887069
(Pertussis unspecified)	Concepts referring to diagnoses which refer to pertussis, but without a specific indication that *Bordetella pertussis* is responsible for the infection	Whooping cough due to unspecified organism	C0043168
		Bordetella infections	C0006015
		Whooping cough-like syndrome	C0343485
		Notification of whooping cough	
		Pneumonia in pertussis	C0155865
(Symptoms compatible with pertussis)	This set of concepts was introduced because the Spanish translation of 'whooping cough' was found to be considered by physicians as a symptom, not as a diagnosis	Concept of 'tos pertusoide' in Spanish general practice	
(Symptoms in infants)	Concepts referring to symptoms that were found to be predictive of pertussis in infants [13,14]	Apnea	C0003578
		Cyanosis	C0010520
		Post-tussive vomiting	C1740793
		Paroxysms of coughing	C0231911
(Macrolides)	Use of macrolides	Macrolides	
(*Bordetella pertussis* test)	The concepts listed in this set indicate the prescription of tests that are considered to be confirmatory of a *Bordetella pertussis* infection	Polymerase chain reaction test	
		Culture or serology	
		Isolation of *Bordetella pertussis* from a clinical specimen	
(Positive result from a *Bordetella pertussis* test)	The concepts listed in this set indicate a positive result from a tests confirmatory of a *Bordetella pertussis* infection	Positive polymerase chain reaction test	
		Positive culture or serology	
		Positive isolation of *Bordetella pertussis* from a clinical specimen	

Table 2
Components for pertussis. The concept sets referred to by the words in round parentheses can be found in Table 1.

Name	Setting	Data domain	Concept set
PC diagnosis, specific	Primary care practice	Diagnosis	(*Bordetella pertussis*)
Inpatient diagnosis, specific	Hospital	Diagnosis	(*Bordetella pertussis*)
PC diagnosis, unspecified	Primary care practice	Diagnosis	(Pertussis unspecified)
Inpatient diagnosis, unspecified	Hospital	Diagnosis	(Pertussis unspecified)
Symptoms	Primary care practice	Diagnosis or signs/ symptoms	(Symptoms compatible with pertussis)
Symptoms in infants	Primary care practice	Diagnosis or signs/ symptoms	(Symptoms in infants)
Lab test	Any setting where a laboratory test can be prescribed, or facility where the test is administered	Laboratory test	(*Bordetella pertussis* test)
Positive laboratory results	Any setting where a health professional records the results of a laboratory test, or facility where the results of the test are generated	Results from laboratory test	(Positive result from a *Bordetella pertussis* test)
Drug use	Facility dispensing medications or primary care practice issuing prescriptions	Drug	(Macrolides)
Secondary components			
Symptoms and drugs within 30 days	A patient is positive if they have both a record of symptoms and of drug use, and the interval between the dates is less than 30 days		
Symptoms in infants and drugs within 30 days	A patient is positive if they are 0 or 1 and has both a record of symptoms in infants and of drug use, and the interval between the dates is less than 30 days		

aged 0–14 years. The calculations used the reported confirmed cases in the TESSy surveillance system in 2012 and 2014 as the numerator, and person-time from population distributions in EUROSTAT for 2012 and 2014 as the denominator [17]. Exact Poisson confidence intervals (95% CI) were calculated [18].

Some formulae link the true proportion of BorPer and/or validity indices with each other and with the observed proportion of the component algorithms (Table 3). These formulas are explained in Supplementary File 2.

In this study we considered Π = IR (see Supplementary File 1) and we assumed that for all algorithms A and B, the proportion of true positives among those detected by both algorithms (PPV of A **AND** B), was the same as the PPV of A or of B, whichever was the highest: this may be considered the most conservative assumption, since being retrieved by two algorithms is at least as reliable a criterion as the better of the two.

We expected that the component 'PC diagnosis, specific' would not be not sensitive enough to retrieve all the cases of *Bordetella pertussis* occurred in the database study population, but that the composed algorithm 'PC specific OR unspecified diagnoses' would retrieve, alongside with some true cases, also some false positives. Moreover we were aware that some cases diagnosed and treated in hospital rather than in PC, would not be retrieved by 'PC specific OR unspecified diagnoses', either. We therefore aimed to use the formulae in Table 3 to provide an approximate estimation of the amount of false positives and false negatives of the algorithm 'PC specific OR unspecified diagnoses'. To this aim, we used SIDIAP, because it was the only database where we could observe cases retrieved in hospital but not in PC. In order to apply our formulae we had to make assumptions on three parameters: (a) the PPV of the components associated with the '(*Bordetella pertussis)*' concept set: since this is composed of codes explicitly mentioning the bacterium, we considered that its associated components had a high likelihood of extracting true cases; we chose 90% as a conservative assumption for this PPV; (b) the PPV of the components associated with the '(pertussis unspecified)' concept set: we explored two scenarios, i.e. 70% or 50%; and (c) the sensitivity of the composite of the four diagnosis and laboratory components: we assumed that all true cases in SIDIAP were recorded in at least one of the diagnosis or laboratory-based components. Note that, based on our clinical definition of true case the component 'positive laboratory results' only retrieves true cases, hence its PPV is 100%.

Based on the approximated sensitivity and PPV estimates for the algorithm 'PC specific OR unspecified diagnosis' in SIDIAP we computed the adjusted IR of BorPer in the relevant study population.

3. Results

3.1. Study population

We followed 3,173,268 person-years of children during the study period: 488,847 from the SIDIAP database, 796,324 from BIFAP, 88,754 from PEDIANET, 1,387,939 from THIN and 411,404 from RCGP RSC (Table 4). The percentages of children aged 0 or 1 years in the population aged 0–14 years in Spain were 12.1%

Table 3
Analytic formulae linking the true proportion of pertussis and validity indices of one or two algorithms. In the formulas, Π is the true proportion of cases of pertussis, P is the proportion of cases detected by the algorithm, SE is the sensitivity and PPV is the positive predictive value of the algorithm.

Known parameters	Formula to derive another parameter
One algorithm	
PPV and SE	$\pi = \frac{P \times PPV}{SE}$
PPV and Π	$SE = \frac{P \times PPV}{\pi}$
SE and Π	$PPV = \frac{SE \times \pi}{P}$
Two algorithms A and B	
SE of A, of B, and of A **AND** B	$SE_{A\mathbf{OR}B} = SE_A + SE_B - SE_{A\mathbf{AND}B}$
Π and PPV of A, of B, and of A **AND** B	$SE_{A\mathbf{OR}B} = \frac{P_A \times PPV_A}{\pi} + \frac{P_B \times PPV_B}{\pi} - \frac{P_{A\mathbf{AND}B} \times PPV_{A\mathbf{AND}B}}{\pi}$
SE of A **OR** B, and PPV of A, of B, and of A **AND** B	$\pi = \frac{P_A \times PPV_A + P_B \times PPV_B - P_{A\mathbf{AND}B} \times PPV_{A\mathbf{AND}B}}{SE_{A\mathbf{OR}B}}$
PPV of A, of B, and of A **AND** B	$PPV_{A\mathbf{AND}B} = \frac{P_A \times PPV_A + P_B \times PPV_B - P_{A\mathbf{AND}B} \times PPV_{A\mathbf{AND}B}}{PPV_{A\mathbf{OR}B}}$

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045

Table 4
Study results. Number of person-years (PYs) entering the study in each database. Incidence rates of pertussis per 100,000 children aged 0–14, with 95% confidence interval (CI), from the TESSy surveillance system in the corresponding country and the estimate incidence rate per 100,000 for each component algorithm are shown. In composite algorithms, the incidence rates were stratified per type of case: cases detected only by the left-hand component (indicated in the label before the keyword 'OR'), cases detected by both components, and cases detected by the right-hand component (indicated in the label after the keyword 'OR'). Data for years 2012 and 2014 were pooled.

DB	SIDIAP (Spain)				BIFAP (Spain)				PEDIANET (Italy)				THIN (United Kingdom)				RCGP (United Kingdom)			
Person-years	488,847				796,324				88, 754				1,387,939				411,404			
TESSy (IR and 95% CI)	21.2 (20.5–22.0)				21.2 (20.5–22.0)				5.4 (5.1–5.8)				13.4 (13.0–13.9)				13.4 (13.0–13.9)			
Component algorithms (N and IR per 100,000 PYs)																				
PC diagnosis, specific	21 (4.3)				0 (0.0)				2 (2.3)				79 (5.7)				21 (5.1)			
PC diagnosis, unspecified	173 (35.4)				135 (17.0)				37 (41.7)				178 (12.8)				77 (18.7)			
Inpatient diagnosis, specific	27 (5.5)																			
Inpatient diagnosis, unspecified	26 (5.3)																			
Symptoms					166 (20.8)															
Symptoms and drug within 30 days					122 (15.3)															
Symptoms in infants	27 (5.5)								6 (6.8)				172 (12.4)				30 (7.3)			
Symptoms in infants and drug within 30 days	1 (0.2)												8 (0.6)							
Lab test	96 (19.6)				38 (4.8)				38 (42.8)				209 (15.1)				32 (7.8)			
Positive laboratory results	19 (3.9)				0 (0.0)								3 (0.2)							
Composite algorithms																				
	N (IR)	N (IR) in left-hand component only	N (IR) In both compo nets	N (IR) In right-hand component only	N (IR)	N (IR) in left-hand component only	N (IR) In both compo nets	N (IR) In right-hand component only	N (IR)	N (IR) in left-hand component only	N (IR) In both compo nets	N (IR) In right-hand component only	N (IR)	N (IR) in left-hand component only	N (IR) In both compo nets	N (IR) In right-hand component only	N (IR)	N (IR) in left-hand component only	N (IR) In both componets	N (IR) In right-hand component only
PC specific OR unspecified diagnosis	194 (39.6)	21 (4.3)	0 (0.0)	173 (35.4)	135 (17.0)	0 (0.0)	0 (0.0)	135 (17.0)	39 (43.9)	2 (2.2)	0 (0.0)	37 (41.7)	246 (17.7)	68 (4.9)	11 (0.8)	167 (12.0)	91 (22.1)	14 (3.4)	7 (1.7)	70 (17.0)
Inpatient specific OR unspecified diagnosis	52 (10.6)	25 (5.1)	1 (0.2)	26 (5.3)																
PC OR inpatient diagnosis	220 (45.0)	168 (34.3)	26 (5.3)	26 (5.3)																
PC diagnosis OR lab test	271 (55.4)	77 (15.8)	19 (3.9)	175 (35.8)	168 (21.1)	33 (4.1)	5 (0.6)	130 (16.3)	69 (77.7)	30 (33.8)	8 (9.0)	31 (34.9)	426 (30.7)	181 (13.0)	29 (2.1)	217 (15.6)	115 (28.0)	24 (5.8)	8 (1.9)	83 (20.2)
Positive lab results OR PC diagnosis	197 (40.3)	3 (0.6)	16 (3.3)	178 (36.4)	135 (17.0)	0 (0.0)	0 (0.0)	135 (17.0)					247 (17.8)	1 (0.1)	2 (0.1)	244 (17.6)				
PC diagnosis OR symptoms and drugs					255 (32.0)	133 (16.7)	2 (0.3)	120 (15.1)												
Any diagnosis OR positive lab results	223 (45.6)	204 (41.7)	16 (3.3)	3 (0.6)																

and 16.1% in SIDIAP in BIFAP, respectively, compared with 13.5% in the EUROSTAT Spanish population. In the UK the percentages were 15.1% and 14.8% in RCGP RSC and THIN 13.0%, respectively, compared with 14.3% in the EUROSTAT UK population. The percentage was 13.0% in PEDIANET, compared with 12.9% in the EUROSTAT Italian population.

3.2. Incidence rates estimated by the algorithms

The IRs for the component and composite algorithms, as well as the benchmark IRs from the TESSy surveillance system are illustrated in Fig. 1 and documented in Table 4, while 95% confidence intervals are included in Supplementary Table 2. The IRs estimated from the TESSy surveillance system in 2012 and 2014 for children aged 0–14 years were 21.2 (95% CI: 20.5; 22.0) for Spain, 13.4 (95% CI: 13.0; 13.9) for the United Kingdom, and 5.4 (95% CI: 5.1; 0.8) for Italy. The number of cases of 'PC diagnosis, specific' (and IRs per 100,000 PY) were 0 (IR = 0.0), 21 (IR = 4.3), 21 (IR = 5.1), 79 (IR = 5.7), and 2 (IR = 2.3) in the BIFAP, SIDIAP, RCGP RSC, THIN and PEDIANET databases, respectively. The component 'PC diagnosis, unspecified' had a higher IR in all databases, and combining the two components ('PC specific OR unspecified diagnosis') increased the number of cases detected and the IRs to 135 (IR = 17.0), 194 (IR = 39.6), 39 (IR = 43.9), 246 (IR = 17.7), and 91 (IR = 22.1), respectively. In BIFAP, SIDIAP, RCGP RSC and THIN, the IRs were similar or higher with respect to the corresponding IRs from the TESSy surveillance system (17.0 vs 21.2; 39.6 vs. 21.2; 22.1 vs. 13.4; 17.7 vs. 13.4, respectively): this is consistent with the expectation that the unspecified component captured, alongside with some true cases, also some false positives. In PEDIANET the composite IR was also higher than the IR from the TESSy database, but unlike in the other databases, where the IR was less than double, in PEDIANET a disproportionate difference was observed (43.9 vs 5.4).

SIDIAP was the only database in which data from both the PC and hospital settings could be linked. The total number of cases in 'PC OR inpatient diagnosis' in SIDIAP was 220 (IR = 45.0), including 26 (12%) that had not been identified in the PC setting. Unlike in the PC setting, where most of the diagnoses were unspecified, in the inpatient setting there were around half specific and half unspecified diagnoses.

In BIFAP, the 'symptoms and drugs within 30 days' component identified 122 cases with an IR of 15.3 per 100,000 PYs. When this component was combined with 'PC specific OR unspecified diagnosis', the IR increased to 32.0, which was higher than the reference IR which was 21.2. Almost none of the children aged 0 or 1 year old in 'symptoms in infants' in any database had a corresponding prescription or dispensing of macrolides in the 'symptoms in infants and drugs within 30 days' component.

The 'lab test' component was available in all databases and had a relatively high IR (from 4.8 in BIFAP to 42.8 in PEDIANET). 'Positive laboratory results' were only available in SIDIAP and THIN, with only 19 and 3 cases, respectively. In SIDIAP, 3 of the 19 cases were not captured by a diagnosis in either primary care or hospital settings.

In Supplementary Fig. 1 and Supplementary Table 3, the analysis was repeated for infants (children aged 0 or 1). The IRs in this subpopulation were around three times higher than the IRs in the overall study population. The findings confirmed the relationship between components observed in the general study population, with the exception of 'PC OR inpatient diagnosis' in SIDIAP (n = 98), where 25.5% (n = 25) were not retrieved from the PC setting, vs 11.8% in the overall study population.

In order to obtain an approximate estimate of algorithm validity, we explored two scenarios in SIDIAP, corresponding to different assumptions for PPV of 'PC diagnosis, unspecified' and of 'inpatient diagnosis, unspecified': in the first scenario, PPV was

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045

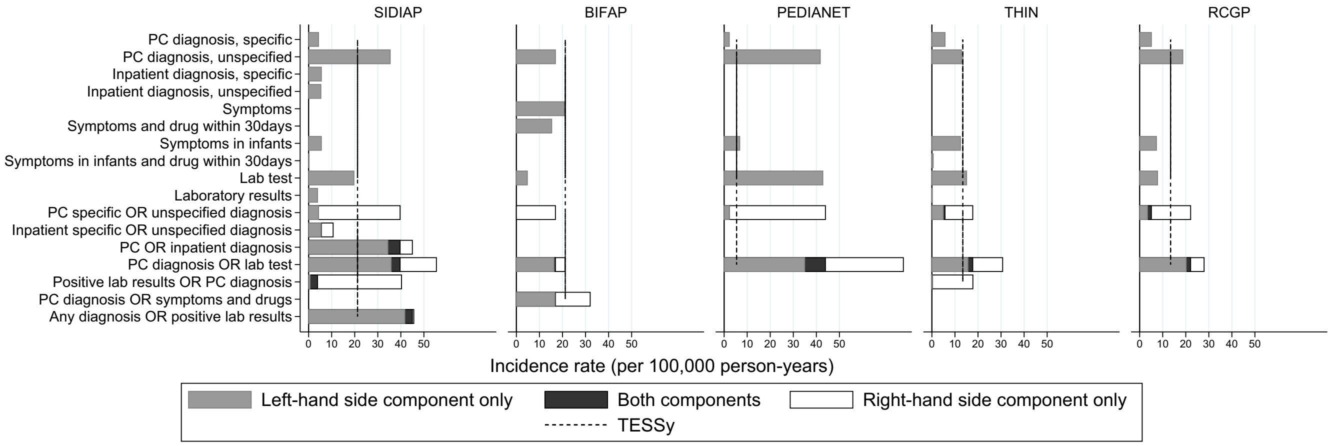

Fig. 1. Study results for the incidence of tested component and composite algorithms. For each component algorithm the incidence rate per 100,000 person-years is shown. For the composite algorithms, the incidence rates were stratified per type of case: the gray bar represents cases detected only by the left-hand component (indicated in the label before the key Boolean operator 'OR'); the black bar represents cases detected by both components; the white bar represents cases detected by the right-hand component (indicated in the label after the key Boolean operator word 'OR'). The dashed line represents the national incidence rate per 100,000 person-years based on data from the TESSy surveillance system. Data for years 2012 and 2014 were pooled.

70%, in the second scenario, PPV was 50%. As a consequence, in the first scenario 'PC specific OR unspecified diagnosis' had a PPV of 72% (or, in the second scenario: 54%) and a sensitivity of 85% (or, in the second scenario: 83%). Based on this estimate, the adjusted IR of BorPer in the SIDIAP study population was 35.5 per 100,000 PY (or, in the second scenario: 25.9) vs the TESSy surveillance system IR 21.2.

4. Discussion

We designed and applied several algorithms as potential strategies to detect cases of *Bordetella pertussis* and thus estimate the IR in five European healthcare databases. The IRs estimated by these algorithms were heterogeneous within and between databases. The algorithm retrieving specific *Bordetella pertussis* diagnostic codes from the PC setting was consistently underestimating IR with respect to the reference estimates from TESSy surveillance system. More sensitive algorithms mostly retrieved a higher number of cases than then number predicted by the TESSy surveillance system, which is compatible with the retrieval of some false positives. Among all the algorithms tested, the one retrieving either specific *Bordetella pertussis* or unspecified pertussis diagnostic codes from the PC setting, despite including probably some false positives, still was not perfectly sensitive, as it was probably missing some of the cases diagnosed and treated in hospital setting. In SIDIAP, in which both hospital and PC diagnoses are collected and can be linked when they refer to the same patient, this expectation was confirmed, since hospital cases were found to be partially missing in PC. In fact, based on a few assumptions and some novel formulae, we could obtain, in SIDIAP, a range of approximate estimates of PPV and sensitivity of this algorithm from the observed IRs: PPV ranged from 54% to 72% and from 83% to 85%, respectively. Based on such approximate values of PPV and sensitivity, the estimate of adjusted IRs of *Bordetella pertussis* in the corresponding population ranged from 25.9 to 35.5 per 100,000 person-years, against the TESSy surveillance system estimate of 21.2.

4.1. General comments

Three components were expected to have a high PPV: PC and inpatient specific diagnoses, and positive laboratory results. Two were expected to have lower PPV (PC and inpatient unspecified diagnoses). One was expected to be sensitive (prescription of a laboratory test), two were very unspecific (symptoms and symptoms in infants) and were planned to be used only in combination with the last component (prescription or use of macrolides) in a 30-days window of time.

In all the databases, the combination of the components which was expected to have a high PPV (specific diagnoses and laboratory tests) identified less cases than the number expected from the TESSy surveillance system,. We concluded that some of the cases recorded with an unspecified diagnosis were actually true *Bordetella pertussis* cases, that had not been updated when the diagnosis was confirmed. One possible explanation could be that it takes several days to confirm the diagnosis of pertussis after the disease is suspected, and there may be no opportunity for the specific diagnosis to be recorded if the patient does not return to the healthcare facility. Another possible explanation may be that the medical personnel may not see the need to update the record for the purposes of clinical care, or that they rather prefer recording informal terms in the event of sharing their records with the patients themselves. This attitude may be influenced by the level of awareness of possible reuse of electronic records for research purposes. These potential explanations may have varying levels of impact in the different databases. For example, in some databases we observed that among the cases detected by a diagnostic component (unspecified or specific), the specific diagnosis was more frequent.

Based on the results of this study, in all the databases it is now possible to design sensitivity analysis using a more specific (but less sensitive) definition of pertussis. In case of heterogeneity the results of a study concerning pertussis, either as an outcome or a study cohort characteristic, designing such sensitivity analyses should be considered as a valid option. On the other hand, in all the databases there is now a possible choice among with different sensitivity: we explored several of them, among which 'unspecified diagnoses' (the most conservative) and 'test' (the least conservative). Even though these algorithms are likely to have lower PPVs, they may still be useful for sensitivity analyses, especially if there are reasons to think that a specific algorithm could be affected by differential misclassification. For example, pertussis may be more readily suspected and tested for in unvaccinated children, and therefore would be recorded in a more accurate manner. In the event that validation with a gold standard is available, validating cases selected via a sensitive algorithm has the chance to identify cases that were not recorded with a diagnostic code.

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045

We developed a component for infants that we though would be sensitive and, although it was likely to have a low PPV, it was less prone to differential misclassification, because it captured symptoms that physicians may not think of as being related with pertussis. However this component proved to be unusable; in reality, when we added a secondary component for concurrent macrolide use there were very few cases that would have been expected to be found in infants with an infection. In contrast, we developed a component specifically for the symptom 'pertussis-like cough' *(tos pertusoide* in Spanish language) that was apparently specific for pertussis cases that were only found in the BIFAP database. Not only did the majority of cases have a concurrent record of prescription of macrolides, but a manual review of a sample of 100 records including physician free text comments, found 2 cases of unspecified pertussis and 2 cases of suspected pertussis. Therefore, this component may be considered for sensitivity analysis or as source for cases to be validated.

4.2. Compatibility with TESSy and seroprevalence surveys

In this study we were able to compare the IRs estimated for paediatric cohorts in five databases using the various algorithms with the national IR estimates from ECDC's TESSy surveillance database. The cases captured by the two types of systems were expected to be slightly different, for various reasons. First, TESSy provides estimates at the national level using census denominators, while three of the databases participating in this study had a regional/multiregional scope (SIDIAP, BIFAP and PEDIANET) and two were based on a representative sample of the national population (THIN, RCGP RSC). Therefore it is possible that some clusters of the infectious disease might be under or over-represented in these database. Second, we collected only confirmed cases from TESSy, while some true cases captured by a PC database with a sensitive algorithm may never be confirmed (under ascertainment), or may never be notified (underreporting) [19,20]. Thus the databases may be a complementary source of true cases which are not notified, while adding potentially false positive cases. Finally, the TESSy data for pertussis may also be affected by under ascertainment and underreporting.

The IR found for PEDIANET, which was much higher than the IR estimate from TESSy for Italy (43.9 vs 5.4), may be explained by a combination of both phenomena discussed above. PEDIANET collects data from PC physicians working in the Italian region Veneto, in the North East of the country. The Regional Office for Infectious Diseases of the Veneto Region provided an estimated IR of 10.0 to the data custodians of PEDIANET. This shows that the region had a higher pertussis notification rate than at the national level for 2012 and 2014, although almost all the diagnoses in PEDIANET were unspecified. However, the regional estimate could be underestimated because of under ascertainment. Finally, as in the other databases, many cases in PEDIANET could be false positives. In general, if estimates of the PPV of the diagnoses are available, the estimated IR from databases can provide a quantitative estimate of under ascertainment and under notification in TESSy. Vice versa, if under notification to TESSy is known to be small, estimates of the PPV for the algorithm can be obtained.

In SIDIAP we assumed a range of plausible values for PPV of unspecified diagnoses (50 and 70%) and that PPV of specific diagnoses was 90%. Under those assumptions the adjusted IR of *Bordetella pertussis* in the database study population was higher than the IR from TESSy. This suggests either that the assumptions we made on PPVs were too generous, or that the number of cases predicted by the TESSy rates was lower than the number truly occurred in the database study population.

Results from seroprevalence surveys have provided estimates for the incidence of *Bordetella pertussis* infection [21–23]. These have provided prevalence estimates beyond those of the surveillance systems, partly as they also capture asymptomatic or mildly symptomatic infections. On the contrary, in this study, we observed that estimates of incidence obtained from databases are roughly comparable with those of TESSy.

4.3. Scope of the component strategy

The scope of this component strategy goes beyond ADVANCE and has the potential of being a comprehensive tool to address heterogeneity and disease misclassification in databases, particularly in multi-database pharmacoepidemiology studies. Components should be designed by individuals who have expertise in the databases involved, and can be used to approach their heterogeneous characteristics in a standardized, transparent and systematic fashion.

Inspection of components can provide knowledge that can inform the design of validation studies: for instance, cases to be validated could be chosen among those positive for a component with high sensitivity and low PPV, such as the component 'lab test' in our study. In many European databases, estimating the PPV of simple algorithms such as components is feasible in a relatively timely and inexpensive way [24–26]. If this is not feasible, scenarios for possible PPVs of the components can be exhibited, based on their characteristics: for instance in our case, some components were expected to have higher sensitivity and lower PPVs. Regardless of their source (validation, external sources or assumptions) the PPVs of components can finally be incorporated in our formulas in Table 3 and provide a picture of the overall validity.

Finally, comparing the distribution of components across exposure strata can indicate if differential misclassification is to be suspected. Differential misclassification can be an important source of bias, even if validity is high [5, 6, [27,28]. If components are unevenly distributed across exposure strata, sensitivity analyses of the study results must be conducted to check whether they are robust to differential misclassification, and components with different validity can be used to this aim.

4.4. Strengths and limitations

In this study, we used standardised component algorithms as a transparent way of documenting the data extraction process across multiple databases. At the same time, we could also perform a qualitative evaluation of the expected validity of each component of Bordetella pertussis, based on its specified semantics and setting. Quantitative scenarios for the validity of each component can also be made using the same approach. We showed that estimates of the validity of various composite algorithms can then be derived in a purely algebraic manner. We could use the incidence estimates based on data from the TESSy surveillance system, which is where European Union member states are required to report pertussis cases, as a reference value, although we cannot exclude the possibility that they may also be subject to under ascertainment and underreporting.

The estimates of sensitivity that we obtained for SIDIAP cannot be generalised to the other PC databases. The sensitivity of the PC databases depends on how often a person with the disease symptoms would seek attention in a PC practice. Although in all the databases, the PC physicians have a gatekeeper role, emergency care can be sought without PC referral, and PC practices may not be accessible at night or weekends. Referrals from other settings may be recorded in the PC practice, but no automatic mechanism is in place. In the absence of a database-specific estimate, however, estimates from another database are a realistic alternative to the assumption that sensitivity is 100%.

5. Conclusions

Retrieving BorPer cases using only specific concepts has low sensitivity in PC databases, while including cases retrieved by unspecified concepts introduces false positives, which were approximately estimated to be 28% in one database. The share of cases that cannot be retrieved from a PC database because they are only seen in hospital was approximately estimated to be 15% in one database. This study demonstrated that quantifying the impact of different event-finding algorithms across databases and benchmarking with disease surveillance data can provide approximate estimates of algorithm validity.

Acknowledgements

The authors medical writing and editorial assistance from Margaret Haugh (MediCom Consult, Villeurbanne, France).

Disclaimer

The results described in this publication are from the proof of concept studies conducted as part of the IMI ADVANCE project with the aim of testing the methodological aspects of the design, conduct and reporting of studies for vaccine benefit-risk monitoring activities. The results presented relate solely to the methodological testing and are not intended to inform regulatory or clinical decisions on the benefits and risks of the exposures under investigation. This warning should accompany any use of the results from these studies and they should be used accordingly. 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.

Funding source

The Innovative Medicines Initiative Joint Undertaking funded this project under ADVANCE grant agreement no. 115557, resources of which were composed of a financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and in kind contributions from EFPIA member companies.

Declaration of Competing Interest

Caitlin Dodd, Kaatje Bollaerts, Claudia Bartolini, Giuseppe Roberto, Consuelo Huerta-Alvarez, Elisa Martín-Merino, Talita Duarte-Salles, Gino Picelli, Lara Tramontan, Giorgia Danieli, Ana Correa, Chris McGee, Benedikt Becker, Charlotte Switzer, Jorgen Bauwens, Nicoline van der Maas, Gianfranco Spiteri, Emmanouela Sdona declared no conflicts of interest. Rosa Gini declared that her institution participates in studies funded by Novartis, Eli Lilly, Daiichi Sankyo, compliant with the ENCePP Code of Conduct. Sonja Gandhi-Banga declared that she works for Sanofi Pasteur and holds company shares. Daniel Weibel declared that he has received personal fees from GSK for work unrelated to the submitted work. Miriam Sturkenboom declared that she has received grants from Novartis, CDC and Bill & Melinda Gates Foundation for work unrelated to the submitted work.

Appendix A. Members of ADVANCE consortium (October 2018)

Full partners

AEMPS: Agencia Española de Medicamentos y Productos Sanitarios (www.aemps.es)
ARS-Toscana: Agenzia regionale di sanità della Toscana (https://www.ars.toscana.it/it/)
ASLCR: Azienda Sanitaria Locale della Provincia di Cremona (www.aslcremona.it)
AUH: Aarhus Universitetshospital (kea.au.dk/en/home)
ECDC: European Centre of Disease Prevention and Control (www.ecdc.europa.eu)
EMA: European Medicines Agency (www.ema.europa.eu)
EMC: Erasmus Universitair Medisch Centrum Rotterdam (www.erasmusmc.nl)
GSK: GlaxoSmithKline Biologicals (www.gsk.com)
IDIAP: Jordi Gol Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (http://www.idiapjordigol.com)
JANSSEN: Janssen Vaccines - Prevention B.V. (http://www.janssen.com/infectious-diseases-and-vaccines/crucell)
KI: Karolinska Institutet (ki.se/meb)
LSHTM: London School of Hygiene & Tropical Medicine (www.lshtm.ac.uk)
MHRA: Medicines and Healthcare products Regulatory Agency (www.mhra.gov.uk/)
MSD: Merck Sharp & Dohme Corp. (www.merck.com)
NOVARTIS: Novartis Pharma AG (www.novartisvaccines.com)
OU: The Open University (www.open.ac.uk)
P95: P95 (www.p-95.com)
PEDIANET: Società Servizi Telematici SRL (www.pedianet.it)
PFIZER: Pfizer Limited (www.pfizer.co.uk)
RCGP: Royal College of General Practitioners (www.rcgp.org.uk)
RIVM: Rijksinstituut voor Volksgezondheid en Milieu (www.rivm.nl)
SCIENSANO: Sciensano (https://www.sciensano.be)
SP: Sanofi Pasteur (www.sanofipasteur.com)
SSI: Statens Serum Institut (www.ssi.dk)
SURREY: The University of Surrey (www.surrey.ac.uk)
SYNAPSE: Synapse Research Management Partners, S.L. (www.synapse-managers.com)
TAKEDA: Takeda Pharmaceuticals International GmbH (www.tpi.takeda.com)
UNIBAS-UKBB: Universitaet Basel – Children's Hospital Basel (www.unibas.ch)
UTA: Tampereen Yliopisto (www.uta.fi)

Associate partners

AIFA: Italian Medicines Agency (www.agenziafarmaco.it)
ANSM: French National Agency for Medicines and Health Products Safety (ansm.sante.fr)
BCF: Brighton Collaboration Foundation (brightoncollaboration.org)
EOF: Helenic Medicines Agency, National Organisation for Medicines (www.eof.gr)
FISABIO: Foundation for the Promotion of Health and Biomedical Research (www.fisabio.es)
HCDCP: Hellenic Centre for Disease Control and Prevention (www.keelpno.gr)
ICL: Imperial College London (www.imperial.ac.uk)
IMB/HPRA: Irish Medicines Board (www.hpra.ie)
IRD: Institut de Recherche et Développement (www.ird.fr)
NCE: National Center for Epidemiology (www.oek.hu)
NSPH: Hellenic National School of Public Health (www.nsph.gr)
PHE: Public Health England (www.gov.uk/government/organisations/public-health-england)
THL: National Institute for Health and Welfare (www.thl.fi)
UMCU: Universitair Medisch Centrum Utrecht (www.umcu.nl)
UOA: University of Athens (www.uoa.gr)
UNIME: University of Messina (www.unime.it)

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045

Vaccine.Grid: Vaccine.Grid (http://www.vaccinegrid.org/)
VVKT: State Medicines Control Agency (www.vvkt.lt)
WUM: Polish Medicines Agency - Warszawski Uniwersytet Medyczny (https://wld.wum.edu.pl/)

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2019.07.045.

References

[1] Sturkenboom M, van der Aa L, Bollaerts K, Emborg HD, Ferreira G, Gini R, et al. The ADVANCE distributed network system for evidence generation on vaccines coverage, benefits and risks based on electronic health care data. Vaccine. 2018; Paper 2 in supplement.

[2] Sturkenboom M, Weibel D, van der Aa L, Braeye T, Gheorge M, Becker B, et al. ADVANCE database characterization and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of vaccinations. Vaccine. 2018;Paper 3 in Supplement.

[3] 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.

[4] Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, et al. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. J Am Med Inform Assoc 2013;20:184–92.

[5] Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. Curr Epidemiol Rep 2014;1:175–85.

[6] Hofler M. The effect of misclassification on the estimation of association: a review. Int J Methods Psychiatr Res 2005;14:92–101.

[7] De Smedt T, Merrall E, Macina D, Perez-Vilar S, Andrews N, Bollaerts K. Bias due to differential and non-differential disease- and exposure misclassification in studies of vaccine effectiveness. PLoS One 2018;13:e0199180.

[8] Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. Am J Epidemiol 1993;138:1007–15.

[9] Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: strategy from the EMIF Project. PLoS One 2016;11:e0160648.

[10] Barlow RS, Reynolds LE, Cieslak PR, Sullivan AD. Vaccinated children and adolescents with pertussis infections experience reduced illness severity and duration, Oregon, 2010–2012. Clin Infect Dis 2014;58:1523–9.

[11] McNamara LA, Skoff T, Faulkner A, Miller L, Kudish K, Kenyon C, et al. Reduced severity of pertussis in persons with age-appropriate pertussis vaccination-United States, 2010–2012. Clin Infect Dis 2017;65:811–8.

[12] European Parliament and of the Council. (Decision EU 2012) Commission implementing decision of 8 August 2012 amending Decision 2002/253/EC laying down case definitions for reporting communicable diseases to the Community network under Decision No 2119/98/EC of the European Parliament and of the Council. Annex to L 262. Official Journal of the European Union 27/9/2012. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012D0506&qid=1428573336660&from=EN#page=22. [accessed on: 9 November 2018].

[13] European Centre for Disease Prevention and Control. The European Surveillance System (TESSy). Available at: https://ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy. [accessed on: 9 November 2018].

[14] 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.

[15] Bellettini CV, de Oliveira AW, Tusset C, Baethgen LF, Amantea SL, Motta F, et al. laboratorial and radiographic predictors of Bordetella pertussis infection]. Revista paulista de pediatria : orgao oficial da. Sociedade de Pediatria de Sao Paulo. 2014;32:292–8.

[16] Hurtado-Mingo A, Mayoral-Cortes JM, Falcon-Neyra D, Merino-Diaz L, Sanchez-Aguera M. Obando I [Clinical and epidemiological features of pertussis among hospitalized infants in Seville during 2007–2011]. Enferm Infecc Microbiol Clin 2013;31:437–41.

[17] Eurostat. Population data. Available at: https://ec.europa.eu/eurostat/web/population-demography-migration-projections/population-data/database. [accessed on: 12 November 2018].

[18] Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). Am J Epidemiol 1990;131:373–5.

[19] McDonald SA, Teunis P, van der Maas N, de Greeff S, de Melker H, Kretzschmar ME. An evidence synthesis approach to estimating the incidence of symptomatic pertussis infection in the Netherlands, 2005–2011. BMC Infect Dis 2015;15:588.

[20] Schielke A, Takla A, von Kries R, Wichmann O, Hellenbrand W. Marked underreporting of pertussis requiring hospitalization in infants as estimated by capture-recapture methodology, Germany, 2013–2015. Pediatr Infect Dis J 2018;37:119–25.

[21] Barkoff AM, Grondahl-Yli-Hannuksela K, He Q. Seroprevalence studies of pertussis: what have we learned from different immunized populations. Pathog Dis 2015;73.

[22] de Greeff SC, de Melker HE, van Gageldonk PG, Schellekens JF, van der Klis FR, Mollema L, et al. Seroprevalence of pertussis in The Netherlands: evidence for increased circulation of Bordetella pertussis. PLoS One 2010;5:e14183.

[23] de Melker HE, Versteegh FG, Schellekens JF, Teunis PF, Kretzschmar M. The incidence of Bordetella pertussis infections estimated in the population from a combination of serological surveys. J Infect 2006;53:106–13.

[24] Coloma PM, Valkhoff VE, Mazzaglia G, Nielsson MS, Pedersen L, Molokhia M, et al. Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. BMJ Open 2013;3.

[25] Valkhoff VE, Coloma PM, Masclee GM, Gini R, Innocenti F, Lapi F, et al. Validation study in four health-care databases: upper gastrointestinal bleeding misclassification affects precision but not magnitude of drug-related upper gastrointestinal bleeding risk. J Clin Epidemiol 2014;67:921–31.

[26] Gini R, Schuemie MJ, Mazzaglia G, Lapi F, Francesconi P, Pasqua A, et al. Automatic identification of type 2 diabetes, hypertension, ischaemic heart disease, heart failure and their levels of severity from Italian General Practitioners' electronic medical records: a validation study. BMJ Open 2016;6:e012413.

[27] De Smedt T, Merrall E, Macina D, Perez-Vilar S, Andrews N, Bollaerts K. Bias due to differential and non-differential disease- and exposure misclassification in studies of vaccine effectiveness. PLoS One 2018;13(6):e0199180.

[28] Newcomer SR, Kulldorff M, Xu S, Daley MF, Fireman B, Lewis E, et al. Bias from outcome misclassification in immunization schedule safety research. Pharmacoepidemiol Drug Saf 2018 Feb 1;27(2):221–8.

Please cite this article as: R. Gini, C. N. Dodd, K. Bollaerts et al., Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project, Vaccine, https://doi.org/10.1016/j.vaccine.2019.07.045