Elsevier Editorial System(tm) for Vaccine
Manuscript Draft

Manuscript Number:

Title: ADVANCE system testing: estimating the incidence of adverse events following pertussis vaccination in healthcare databases with incomplete exposure data

Article Type: SI: ADVANCE

Keywords: missing exposure data; incidence rate derivation; adverse events following vaccination; database heterogeneity

Corresponding Author: Ms. Caitlin Nora Dodd, MS

Corresponding Author's Institution: Erasmus University Medical Center

First Author: Caitlin Nora Dodd, MS

Order of Authors: Caitlin Nora Dodd, MS; Kaatje Bollaerts; Daniel Weibel; Maria de Ridder; Olivia Mahaux; Francois Haguinet; Tom Desmedt; Simon de Lusignan; Chris McGee; Talita Duarte-Salles; Hanne-Dorthe Emborg; Consuelo Huerta; Elisa Martin-Merino; Gino Picelli; Klara Berencsi; Giorgia Danieli; Miriam Sturkenboom

Abstract: 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 vaccine benefit-risk monitoring using existing European healthcare databases. Incidence rate (IR) estimates of vaccination-associated adverse events that are needed to model vaccination risks can be calculated from existing healthcare databases when vaccination (exposure) data are available. We assessed different methods to derive IRs when data are missing in one database, using estimated IRs from other databases for febrile seizures, fever and persistent crying. IRs were estimated for children aged 0-5 years in outcome-specific risk and non-risk periods following the first dose of acellular pertussis (aP) vaccination in four primary care databases and one hospital database. We compared derived and observed IRs in each database using three methods: 1) multiplication of non-risk period IR for database i by IR ratio (IRR) obtained from meta-analysis of IRRs estimated using the self-controlled case-series method, from databases other than i; 2) same method as 1, but multiplying with background IR; and 3) meta-analyses of observed IRs from databases other than i. IRs for febrile seizures were lower in primary care databases than the hospital database. The derived IR for febrile seizures using data from primary care databases was lower than that observed in the hospital database, and using data from the hospital database gave a higher derived IR than that observed in the primary care database. For fever and persistent crying the opposite was observed. We demonstrated that missing IRs for a post-vaccination period can be derived but that the type of database and the method of event data capture can have an impact on potential bias. We recommend IRs are derived using data from similar database types (hospital or primary care) with caution as even this can give heterogeneous results.

Dr Gregory A Poland
Editor-in-Chief, Vaccine

14 November 2018

Dear Dr Poland

We are pleased to submit our paper ‘ADVANCE system testing: estimating the incidence of adverse events following pertussis vaccination in healthcare databases with incomplete exposure data’ to your Journal, Vaccine for the ADVANCE supplement. This paper describes results from methods that can be used to estimate adverse event rates in the absence of full data on vaccine exposure. It is the 7th of the series of 10 papers that will be included in the supplement.

On behalf of all co-authors

Ms Caitlin Dodd

***Author Agreement**

I, Ms Caitlin Dodd, declare that all authors have seen and approved the final version of the manuscript being submitted. We warrant that the article is our original work that has not been previously published and is not under consideration for publication elsewhere

***Suggested Reviewers**

Name	Institute	email
Heather Whitaker	Open University	heather.whitaker@open.ac.uk
Bruce Fireman	Kaiser Permanente	bruce.fireman@kp.org
Jason Glanz	Kaiser Permanente	jason.m.glanz@kp.org
Wei Hua	US FDA	Wei.Hua@fda.hhs.gov

Highlights

- Incidence rates of events following vaccination can be derived when exposure data is missing
- Incidence rate ratios from databases with complete exposure data can be used to derive IRs
- Type of database and event reporting methods determines bias for derived IRs
- Background incidence can be used to derive risk period incidence if the risk period is short

Abstract

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 vaccine benefit-risk monitoring using existing European healthcare databases. Incidence rate (IR) estimates of vaccination-associated adverse events that are needed to model vaccination risks can be calculated from existing healthcare databases when vaccination (exposure) data are available. We assessed different methods to derive IRs when data are missing in one database, using estimated IRs from other databases for febrile seizures, fever and persistent crying. IRs were estimated for children aged 0-5 years in outcome-specific risk and non-risk periods following the first dose of acellular pertussis (aP) vaccination in four primary care databases and one hospital database. We compared derived and observed IRs in each database using three methods: 1) multiplication of non-risk period IR for database *i* by IR ratio (IRR) obtained from meta-analysis of IRRs estimated using the self-controlled case-series method, from databases other than *i*; 2) same method as 1, but multiplying with background IR; and 3) meta-analyses of observed IRs from databases other than *i*. IRs for febrile seizures were lower in primary care databases than the hospital database. The derived IR for febrile seizures using data from primary care databases was lower than that observed in the hospital database, and using data from the hospital database gave a higher derived IR than that observed in the primary care database. For fever and persistent crying the opposite was observed. We demonstrated that missing IRs for a post-vaccination period can be derived but that the type of database and the method of event data capture can have an impact on potential bias. We recommend IRs are derived using data from similar database types (hospital or primary care) with caution as even this can give heterogeneous results.

ADVANCE system testing: estimating the incidence of adverse events following pertussis vaccination in healthcare databases with incomplete exposure data

Caitlin Dodd[a,b*], Kaatje Bollaerts[c], Daniel Weibel[a], Maria de Ridder[a], Olivia Mahaux[d], Francois Haguinet[d], Tom de Smedt[c], Simon de Lusignan[e,f], Chris McGee[e,f], Talita Duarte-Salles[g], Hanne-Dorthe Emborg[h], Consuelo Huerta[i], Elisa Martín-Merino[i], Gino Picelli[j], Klara Berencsi[k1], Giorgia Danieli[j] Miriam Sturkenboom[b,c,l]

[a]Erasmus University Medical Centre, Postbox 2040, 3000 CA Rotterdam, The Netherlands (caitlinndodd@gmail.com; d.weibel@erasmusmc.nl; m.deridder@erasmusmc.nl)

[b]J ulius Global Health, University Medical Center, Utrecht, Heidelberglaan 100, The Netherlands (caitlinndodd@gmail.com; m.c.j.sturkenboom@umcutrecht.nl)

[c]P95 Epidemiology and Pharmacovigilance, Koning Leopold III laan 1 3001 Heverlee, Belgium (kaatje.bollaerts@p-95.com; tom.desmedt@p-95.com; m.c.j.sturkenboom@umcutrecht.nl)

[d]GSK, Building W23, Avenue Flemming 20, 1300 Wavre, Belgium (olivia.x.mahaux@gsk.com; francois.f.haguinet@gsk.com)

[e]University of Surrey, Guildford, Surrey GU2 7XH, UK (s.lusignan@surrey.ac.uk; c.mcgee@surrey.ac.uk)

[f]Royal College of General Practitioners, Research and Surveillance Centre, 30 Euston Square, London NW1 2FB, UK (s.lusignan@surrey.ac.uk; c.mcgee@surrey.ac.uk)

[g]Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain (tduarte@idiapjgol.org)

[h]Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Artillerivej 5, DK-23002300, Denmark (hde@ssi.dk)

[1] Present address: Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[i]BIFAP database, Spanish Agency of Medicines and Medical Devices, Madrid, Spain (chuerta@aemps.es; emartinm@aemps.es)

[j]Epidemiological Information for Clinical Research from an Italian Network of Family Paediatricians (PEDIANET), Padova, Italy (g.picelli@virgilio.it; gdanieliconsorzioarsenal@gmail.com)

[k]Aarhus University Hospital, Olof Palmes Alle 43-45, Aarhus, DK-8200, Denmark (klara.berencsi@ndorms.ox.ac.uk)

[l]VACCINE.GRID Foundation, Spitalstrasse 33, Basel, Switzerland (m.c.j.sturkenboom@umcutrecht.nl)

* Corresponding author: Caitlin Dodd, Erasmus University Medical Centre, Rotterdam, The Netherlands; email: caitlinndodd@gmail.com

ADVANCE system testing: estimating the incidence of adverse events following pertussis vaccination in healthcare databases with incomplete exposure data

Caitlin Dodd[a,b*], Kaatje Bollaerts[c], Daniel Weibel[a], Maria de Ridder[a], Olivia Mahaux[d], Francois Haguinet[d], Tom de Smedt[c], Simon de Lusignan[e,f], Chris McGee[e,f], Talita Duarte-Salles[g], Hanne-Dorthe Emborg[h], Consuelo Huerta[i], Elisa Martín-Merino[i], Gino Picelli[j], Klara Berencsi[k1], Giorgia Danieli[j] Miriam Sturkenboom[b,c,l]

[a]Erasmus University Medical Centre, Postbox 2040, 3000 CA Rotterdam, The Netherlands (caitlinndodd@gmail.com; d.weibel@erasmusmc.nl; m.deridder@erasmusmc.nl)

[b]J ulius Global Health, University Medical Center, Utrecht, Heidelberglaan 100, The Netherlands (caitlinndodd@gmail.com; m.c.j.sturkenboom@umcutrecht.nl)

[c]P95 Epidemiology and Pharmacovigilance, Koning Leopold III laan 1 3001 Heverlee, Belgium (kaatje.bollaerts@p-95.com; tom.desmedt@p-95.com; m.c.j.sturkenboom@umcutrecht.nl)

[d]GSK, Building W23, Avenue Flemming 20, 1300 Wavre, Belgium (olivia.x.mahaux@gsk.com; francois.f.haguinet@gsk.com)

[e]University of Surrey, Guildford, Surrey GU2 7XH, UK (s.lusignan@surrey.ac.uk; c.mcgee@surrey.ac.uk)

[f]Royal College of General Practitioners, Research and Surveillance Centre, 30 Euston Square, London NW1 2FB, UK (s.lusignan@surrey.ac.uk; c.mcgee@surrey.ac.uk)

[g]Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain (tduarte@idiapjgol.org)

[h]Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Artillerivej 5, DK-23002300, Denmark (hde@ssi.dk)

[1] Present address: Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

[i]BIFAP database, Spanish Agency of Medicines and Medical Devices, Madrid, Spain (chuerta@aemps.es; emartinm@aemps.es)

[j]Epidemiological Information for Clinical Research from an Italian Network of Family Paediatricians (PEDIANET), Padova, Italy (g.picelli@virgilio.it; gdanieliconsorzioarsenal@gmail.com)

[k]Aarhus University Hospital, Olof Palmes Alle 43-45, Aarhus, DK-8200, Denmark (klara.berencsi@ndorms.ox.ac.uk)

[l]VACCINE.GRID Foundation, Spitalstrasse 33, Basel, Switzerland (m.c.j.sturkenboom@umcutrecht.nl)

* Corresponding author: Caitlin Dodd, Erasmus University Medical Centre, Rotterdam, The Netherlands; email: caitlinndodd@gmail.com

Abbreviations used

ADVANCE: Accelerated Development of VAccine beNefit-risk Collaboration in Europe

aP: acellular pertussis

B/R: benefit-risk

IR: incidence rate

IRR: incidence rate ratio

L-O-O: leave-one-out

MA: meta-analysis

POC: proof of concept

wP: whole-cell pertussis

Abstract

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 vaccine benefit-risk monitoring using existing European healthcare databases. Incidence rate (IR) estimates of vaccination-associated adverse events that are needed to model vaccination risks can be calculated from existing healthcare databases when vaccination (exposure) data are available. We assessed different methods to derive IRs when data are missing in one database, using estimated IRs from other databases for febrile seizures, fever and persistent crying. IRs were estimated for children aged 0-5 years in outcome-specific risk and non-risk periods following the first dose of acellular pertussis (aP) vaccination in four primary care databases and one hospital database. We compared derived and observed IRs in each database using three methods: 1) multiplication of non-risk period IR for database *i* by IR ratio (IRR) obtained from meta-analysis of IRRs estimated using the self-controlled case-series method, from databases other than *i*; 2) same method as 1, but multiplying with background IR; and 3) meta-analyses of observed IRs from databases other than *i*. IRs for febrile seizures were lower in primary care databases than the hospital database. The derived IR for febrile seizures using data from primary care databases was lower than that observed in the hospital database, and using data from the hospital database gave a higher derived IR than that observed in the primary care database. For fever and persistent crying the opposite was observed. We demonstrated that missing IRs for a post-vaccination period can be derived but that the type of database and the method of event data capture can have an impact on potential bias. We recommend IRs are derived using data from similar database types (hospital or primary care) with caution as even this can give heterogeneous results.

Keywords: Missing exposure data; incidence rate derivation; adverse events following vaccination; database heterogeneity

1. Introduction

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 existing healthcare databases in Europe (see Appendix for consortium members). A series of proof of concept (POC) studies were designed to assess the processes and system proposed for generating the required data to generate evidence on coverage, risks and benefits of vaccines as well as benefit-risk analyses. Modelling is one method that is widely used to analyse vaccine benefit-risk, to understand the impacts of diseases, interventions, and environmental exposures deterministically or in simulated populations [1]. Valid estimates of incidence rates (IRs) for vaccine-preventable disease and adverse event following immunisation, and vaccination coverage are needed to model the benefit-risk of vaccination. These input parameters may be obtained from the literature or by using data from available healthcare databases. However, when using healthcare databases, their heterogeneity and potentially important missing information on vaccinations need to be taken into consideration [2].

The first vaccines developed against *Bordetella pertussis* contained whole killed organisms [3]. Due to the reactogenicity of this vaccine, Between 2004 and 2015 several countries switched from whole-cell pertussis (wP) to acellular pertussis (aP) vaccines for infants and children due to the reactogenicity of the wP vaccine [4]. In the ADVANCE POC studies the benefits and risks of wP and aP vaccines in children were compared as an example. For this, IRs of known benefits and adverse events in outcome-specific risk periods following each dose of wP and aP vaccine were required. Since we used existing healthcare databases that collected data for purposes other than for research, we were faced with the problem of comparing the effects of exposure which occurred in distinct time periods, often with missing exposure data for the period before the switch from wP to aP. To compare the B/R for the wP

and aP vaccines, we attempted to estimate IRs for various outcomes following wP vaccination in some databases that were established too recently to contain wP exposure data. In this paper we compared different methods for deriving IRs in the risk period following vaccination. To test these methods we limited the analysis to aP exposure, assuming that the aP exposure data were missing, which allowed us to compare the observed and derived IRs.

2. Methods

2.1 Data sources and population

This study was conducted with data generated for the ADVANCE proof of concept risk study that included seven population-based healthcare databases from Denmark, Spain, UK and Italy (**Table 1**) [5, 6]. Two databases were excluded in this methods study: AUH because it is a subset of the national SSI database in Denmark, and PEDIANET from Italy, in which vaccination data was linked only for the 2006 and 2007 birth cohorts. We excluded data from the SSI database, which is a hospital database, in sensitivity analyses to study the impact of hospital data on the results.

The study population comprised all children aged <6 years registered in any of the participating databases during the study period, who had received at least the first dose of aP vaccine. For the calculation of background rates, children were followed from start of the study period (1 January 1990), one month after their date of birth (to allow for pre-vaccination person time and to avoid pre-term related or birth-induced increase in IR), or date of valid data in the database, whichever occurred the latest. For the calculation of baseline rates and incidence rate ratios, children were followed from 31 days before their first dose of aP vaccine. All children were followed until the end of study period (31 December 2015), until they received their pertussis booster dose, transferring out of the database, death, reaching age 6 years, or end of data availability in the database, whichever occurred first. Children with missing date of birth or sex were excluded.

Data from each participating database was extracted locally and transformed into a common data model, comprising vaccination, event, and population files [7].

2.2 Outcomes

To test the methodology we selected three outcomes from the risk study that have different patterns of care: febrile convulsions, fever, and persistent crying. Febrile convulsions are rare and are usually considered to be serious clinical events requiring presentation to the emergency room. Fever is common but does not often require hospitalisation. Persistent crying is non-specific and often lacks a specific diagnosis code even in primary care. Definitions, codes and methods for data extraction and harmonisation can be found in other papers in this supplement [6, 7].

2.3 Definition of exposure

Data on aP vaccination were obtained from the healthcare databases [6]. Although our study was driven by the need to estimate IRs during the wP risk period, we limited our methodological study to aP risk period since the IRs could be estimated in all participating databases, therefore we could compare the IRs derived using different methods with the estimated IRs.

Outcome-specific risk windows were defined as day 0 to 3 for febrile convulsions and fever and day 0 to 1 for persistent crying, with day 0 being the day of vaccination. Baseline periods were defined as 31 to 8 days before dose one and the interval from the last day of the risk window to 31 days after the dose. The week prior to vaccination was excluded from the baseline period to avoid the 'healthy vaccinee effect', i.e. vaccine avoidance by subjects experiencing an illness (**Figure 1**) [8]. The pertussis vaccination schedules were 3, 5 and 12 months, 2, 4 and 11 months and 2, 3 and 4 months for Denmark, Spain and UK, respectively.

146 *2.4 Statistical methods*

147 IRs were calculated by age in months and in the aP vaccination risk and non-risk period for
148 each outcome. We conducted self-controlled case series (SCCS) analyses for each of the
149 outcomes to obtain IRRs, comparing risk to non-risk periods for the first dose of aP
150 vaccination [9]. The study population for each outcome-specific SCCS analysis included
151 children who experienced the event at least once during their follow-up.

152 For each database *i* and event, a leave-one-out (L-O-O) random effects IRR was estimated
153 using a meta-analysis of the IRRs from all databases other than database *i*, independent of the
154 type of data source [10]. The result is referred to as L-O-O_IRR _ma. IRs in the risk period
155 following vaccination were derived using three methods (**Box**). In the first method, we
156 multiplied the baseline IR calculated in non-risk periods around aP vaccination in database *i*
157 by the L-O-O_IRR_ma that excluded database *i* (IR_bl) (**Figure 2**). In the second method, we
158 multiplied the background IR that was calculated in the month of age at the recommended
159 first dose by the L-O-O_IRR_ma that excluded database *i* (IR-bg). In the third method, we
160 derived a pooled risk period IR using a meta-analysis of the IRs for the observed risk period
161 for all databases other than *i* (IR_ma). We then assessed the agreement between observed and
162 derived risk period IRs.

163 **3. Results**

164 The study population comprised 2.6 million children aged <6 years who had received at least
165 one dose of aP-containing vaccine. The database-specific sample sizes varied from 152,784
166 (RCGP RSC) to 980,843 (SSI) (**Table 1**). Over 400,000 children experienced at least one of
167 the three events of interest during the study period.

168 The overall background IR (per 1,000 person-years) in this paediatric population for febrile
169 convulsion ranged between 3 (BIFAP) to 11 (SSI; hospitalization). The age-stratified IRs
170 peaked between 1 and 2 years of age in all databases (**Figure 3**). For fever, the overall IR (per

1,000 person-years) varied between the databases from 8 (SSI) to 184 (BIFAP). The age-stratified IRs for fever were high up to 18 months of age in most of the databases (**Figure 3**). The overall IRs (per 1,000 person-years) of persistent crying ranged from 2 (THIN) to 22 (BIFAP). The age-stratified IRs peaked in the first months of life and then declined rapidly (**Figure 3**). No data for persistent crying were available in the SIDIAP and SSI databases since there are no specific ICD-9 or ICD-10 codes for this event. The event was identified using BIFAP-specific-ICPC or ICD-9 codes as well as free-text in the BIFAP database. .

IRRs for adverse events following vaccination which compared the IRs in risk periods after aP vaccination with those at baseline, as estimated via SCCS analyses, varied between databases. For febrile convulsions, no significant association after dose one of aP vaccine was seen in the BIFAP and RCGP RSC databases, while the risk was significantly lower in the SSI and THIN databases. L-O-O IRR_ma estimates were closer to 1 than those estimated in the SCCS in all databases. Statistically significant protective effects observed in the SSI and THIN databases were no longer present in the L-O-O_IRR_ma estimates. When the estimates from the SSI database were excluded, the L-O-O_IRR_ma estimates increased slightly closer 1 due to removal of the significantly protective IRR in the SSI database (**Table 2**).

IRRs for fever showed a significant protective effect in the BIFAP and SIDIAP databases whereas the risk was increased in the SSI database and no association was observed in the THIN and RCGP RSC databases. Again, L-O-O meta-analysis removed much of the heterogeneity in these results. All L-O-O_IRR_ma estimates had confidence intervals including one (**Table 2**).

Persistent crying was significantly elevated in all databases that provided data for this event. L-O-O_IRR_ma results were consistent across databases and remained significantly greater than one. Because SSI did not contribute persistent crying cases, removal of SSI had no impact on L-O-O_IRR_ma estimates (**Table 2**).

The IR_bl and IR_bg methods performed similarly for febrile convulsions, tending to underestimate observed risk period IRs. In the primary care databases, with the exception of RCGP RSC, the derived IR_ma tended to be higher than the observed IR, because of the impact of the elevated incidence from the hospital database, SSI. For the SSI database, the observed risk period IR was higher than the derived IR_ma as this was based on the risk period IRs of the primary care databases. In analyses excluding SSI, IR_bl and IR_bg performed similarly and were in agreement with the observed risk period IR except in the RCGP RSC database. The IR_ma method produced higher estimates with wider confidence intervals than IR_bl and IR_bg in all scenarios (**Figure 4**).

For fever the IR_bl and IR_bg methods gave similar results, i.e., derived IR estimates that were generally lower than the observed estimates. The derived IR_ma estimates were similar across databases. In the BIFAP database where the background IR for fever was highest, the IR_ma underestimated the observed IR for the risk period while in the SSI database, where the background rate of fever was the lowest, the IR_ma overestimated the IR for the risk period compared with the observed IR. (**Figure 3, Figure 4**). In analyses excluding SSI, IR_bl and IR_bg significantly underestimated the observed risk period IRs in all databases except for the BIFAP and SIDIAP databases, while the IRs from IR_ma were similar across databases and produced an underestimation of observed risk period IR in BIFAP.

For persistent crying, the results from the IR_bl and IR_bg approaches were similar. In the UK databases, the IRs derived by both methods were slightly lower than the observed risk period IRs, but not statistically significantly lower, whereas the IRs derived by both IR_bl and IR_bg were higher than those observed for the BIFAP database. The risk period IRs derived by the IR_ma method were similar across databases but they were underestimated compared with the observed risk period IRs in the BIFAP and RCGP RSC databases, and overestimated compared with the observed risk period IRs in the THIN database. Since no data for persistent

crying events were available from the SSI database, its removal had no impact on the estimated IRs.

4. Discussion

The results from this study demonstrate that it is possible to obtain estimates for event-specific IRs occurring during risk windows after vaccination in a certain database using incidence rate ratios and incidence rates from other data sources, even if the data on the type of vaccination (for the IR_bl method) or the occurrence of vaccination (for the IR_bg and IR_ma methods) are not available in that database. The results also demonstrate that use of IR estimates from other data sources may not always be valid, since the type of data source (e.g. primary care setting versus hospital setting) has a major impact, which differs by type of event and the care pattern for that event.

Febrile convulsions are acute and can lead to emergency room visits and, therefore, primarily appear in hospital records [11, 12]. Since the SSI database contains only hospital-derived data, this might explain why the background, baseline and risk period incidence rates are higher in the SSI database than in the other databases which contain primary care-derived data (SIDIAP, BIFAP, THIN, and RCGP RSC). The observed IRs for febrile convulsions and their peak at around 15-16 months of age, especially in the SSI database, are consistent with those in the literature that reports a peak incidence at around 18 months old [13, 14]. The derived estimates for febrile convulsions IRs were in much better agreement with observed risk period IRs when the SSI hospital-based database was removed because of the difference in background incidence between primary care and hospital databases.

The post-vaccination IRs for fever derived using baseline or background rates produced estimates that were lower than the observed IRs in the risk window. Fever had a very low background incidence in the SSI database because it is a symptom and is unlikely be recorded as a hospital discharge diagnosis. The IRs derived using meta-analysis also tended to be lower

that the observed risk period IRs except in the SSI database where the observed risk period IR was low. Removal of SSI did not improve the agreement between the derived IRs and observed risk period IRs due to its small contribution and therefore minimal changes to the L-O-O estimates.

Persistent crying is a non-specific condition that is not easy to record using medical coding systems and only the BIFAP database had specific codes for this event. Agreement was good for the methods in all databases, except BIFAP where the derived IRs using baseline and background rates were over-estimates compared with the observed risk period IRs, due to the higher baseline and background rates of persistent crying in BIFAP. The usefulness of the IR_ma estimates for the BIFAP, RCGP RSC and THIN databases is uncertain as they are derived from the meta-analysis of data from the other two databases while for the SIDIAP and SSI databases the IR_ma is the only estimate available due to the absence of persistent crying events in these databases.

In general, IR_ma estimates produced wider CIs due to our use of a random-effects meta-analysis and therefore, the 95% CIs for the IR_ma estimates were more likely to contain the observed IR. The L-O-O_IRR_ma estimates were similar across databases for each event, irrespective of which database was left out, suggesting that any differences in the resulting IR_bl or IR_bg estimates were due to difference in underlying baseline or background rates.

The aim of this study was to assess methods to fill gaps in information in one database using estimates from other databases. We demonstrated that this is possible, but that how data for each event are captured should be taken into consideration, as this may have a greater impact on the absolute IRs than on the IRRs. If an event, such as fever or persistent crying, is not captured in a database, we recommend that the pooled IRs (IR_ma) from databases which were able to capture the event of interest in similar settings are used. For example, the incidence of febrile convulsions was lower in the primary care databases than in the hospital

database, but the IR_ma method produced derived IRs that were more in line with those observed in the hospital database. This method may be preferable if observed IRs in primary care databases are assumed to be underestimated.

Although the type of event type may have an important impact on the performance of methods for derivation, we demonstrated that the IR_bl and IR_bg methods provided very similar results for the events we used, which means that the approach using the background IRs (which does not require vaccine exposure time) can be used. This may be because the risk periods represent a very small period in comparison with the total follow-up period, and the risk increase was small during the risk period. These methods may be preferable if background and baseline IRs are assumed to be accurate, and the IR_bg method may be preferable if the risk period is short or cannot be observed due to missing exposure data.

5. Conclusions

Although we were able to compare derived and observed IRs for aP exposure, we did not have the estimates of the true incidence of each event in the post-wP vaccination risk period in all databases. We cannot draw general conclusions regarding which method provides the best estimates of the true incidence, but we can conclude that, in case of short risk windows and small increases in IRRs, the IR_bl and IR_bg methods provide similar estimates. Additionally, the IR_ma method may provide derived IRs that are closer to the observed IRs when these latter come from a similar type of database. However, it is important to note that this method is sensitive to heterogeneity in baseline incidence in each of the database as it uses absolute measures of incidence, [15, 16].

We demonstrated that the type of events and databases have a large impact and it is important to distinguish if the events are diagnosed in primary care, hospital or both, and perform stratified analyses for the type of events the databases capture. It is important to have a clear understanding of the external and internal validation of the databases as well as the

heterogeneity of the studied databases and those used for deriving the parameters before proceeding to parameter derivation. We conclude that derived IRs for events following vaccination in the absence of specific vaccine exposure data in a specific database is possible if the background IRs can be calculated and IRRs are available from a similar type of database.

Acknowledgements

The authors acknowledge input and guidance from Heather Whitaker (Open University, United Kingdom) and Paddy Farrington (Open University, United Kingdom). They acknowledge Bart Spiessens (Johnson & Johnson, Belgium) who authored the SAS code used to conduct the self-controlled case series. They also acknowledge 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 n° 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 potential conflicts of interest

Caitlin Dodd, Kaatje Bollaerts, Maria de Ridder, Tom de Smedt, Chris McGee, Talita Duarte-Salles, Hanne-Dorthe Emborg, Consuelo Huerta, Elisa Martín-Merino, Gino Picelli, Klara Berencsi, Giorgia Danieli declared that they have no potential conflicts of interest. Daniel Weibel declared that he has received personal fees from GSK for work unrelated to the submitted work. Olivia Mahaux and Francois Haguinet declared that they are employed by GSK and hold company shares. Simon de Lusignan declared that he has received grants from GSK, Takeda, and Seqirus / JSS, and also personal fees from Sequirus and Sanofi, 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.

References

[1] Funk S, Salathe M, Jansen VA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface. 2010;7:1247-56.

[2] Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health. 2003;6:9-17.

[3] Plotkin S OW, Offit PA. Vaccines. 6th edn. ed. Philadelphia: Elsevier Saunders; 2012.

[4] Zhang L, Prietsch SO, Axelsson I, Halperin SA. Acellular vaccines for preventing whooping cough in children. Cochrane Database Syst Rev. 2014:CD001478.

[5] Weibel D. ADVANCE POC I Risk pillar - Testing new approaches to monitoring benefit/risk with pertussis vaccines as test case: Incidence rates of safety outcomes of whole-cell pertussis and acellular pertussis vaccines in pre-school children: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; 2017 [Last accessed 28 May 2018]. Available from: http://www.encepp.eu/encepp/viewResource.htm?id=13813.

[6] 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. 2018;Paper 6 in Supplement.

[7] 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. 2018;Paper 2 in supplement.

[8] Virtanen M, Peltola H, Paunio M, Heinonen OP. Day-to-day reactogenicity and the healthy vaccinee effect of measles-mumps-rubella vaccination. Pediatrics. 2000;106:E62.

[9] Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med. 2006;25:1768-97.

[10] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177-88.

[11] Le Saux N, Barrowman NJ, Moore DL, Whiting S, Scheifele D, Halperin S. Decrease in hospital admissions for febrile seizures and reports of hypotonic-hyporesponsive episodes presenting to hospital emergency departments since switching to acellular pertussis vaccine in Canada: a report from IMPACT. Pediatrics. 2003;112:e348.

[12] Quantin C, Benzenine E, Velten M, Huet F, Farrington CP, Tubert-Bitter P. Self-controlled case series and misclassification bias induced by case selection from administrative hospital databases: application to febrile convulsions in pediatric vaccine pharmacoepidemiology. Am J Epidemiol. 2013;178:1731-9.

[13] Hauser WA. The prevalence and incidence of convulsive disorders in children. Epilepsia. 1994;35 Suppl 2:S1-6.

[14] 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.

[15] Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat Med. 2000;19:1707-28.

[16] Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med. 2002;21:1575-600.

Figure legends

Figure 1: Schematic representation of the timeline of a typical observation period for dose 1.

Figure 2: Approach for calculating risk window specific incidence rates in databases when wP exposure is missing or under the assumption of missing aP exposure [5]

Figure 3: Background incidence of events of interest per 1,000 person years by age in months and database (NB: the y-axes are not the same scale)

Figure 4: Comparison of results from the three methods for calculating incidence rates (IRs) for febrile convulsions, fever and persistent crying following aP vaccination (A) in all databases and (B) in primary care databases (excluding SSI)

BIFAP = Base de datos para la Investigación Famacoepidemiológica en Atención Primaria, RCGP RSC = Royal College of General Practitioners Research and Surveillance Centre, SIDIAP = Information System for Research in Primary Care, SSI = Statens Serum Institute, THIN = The Health Information Network, aPE = acellular pertussis vaccine

Table 1: Databases providing data for the ADVANCE POC safety study [6]

Country	Database	Geographic coverage	Type of data	Years with available data	Switch from wP to aP	Size (N persons)	Children exposed to aP	Primary care diagnoses	Hospital discharge diagnoses
Denmark	SSI	National	National claims data record linkage	2000 - 2014	1997	7.5 million	980,843	No	Yes (ICD-10)
Spain	BIFAP	Multi regional sample	GP medical records	2002 - 2013	2000-2004	4.8 million	320,638	Yes (ICPC-based codes + free text)	Limited to free text comments recorded by the GP
Spain	SIDIAP	Regional (Cataluña)	GP medical records & partial linkage to hospital	2005-2014	2000-2004	5.8 million	570,225	Yes (ICD-10)	Yes (ICD-9)
United Kingdom	RCGP RSC	National sample	GP medical records	2003 - 2014	2004	2.0 million	152,784	Yes (READ)	Yes (READ)
United Kingdom	THIN	National sample	GP medical records	1996-2013	2004	8.3 million	576,151	Yes (READ)	Yes (READ)

AUH = Aarhus University Hospital, SSI = Statens Serum Institute, BIFAP = Base de datos para la Investigación Famacoepidemiológica en Atención Primaria, SIDIAP = Information System for Research in Primary Care, RCGP RSC = Royal College of General Practitioners Research and Surveillance Centre, THIN = The Health Information Network, GP = General Practitioner, ICD = International Classification of Diseases

Table 2: Self-controlled case series (SCCS) and leave-one-out (L-O-O) incidence rate ratios (IRRs) following dose one of acellular pertussis vaccine

Event	Database	SCCS IRR (95% CI)	L-O-O IRR (95% CI)	L-O-O IRR without SSI (95% CI)
Febrile convulsions	SSI	0.24 (0.18; 0.31)	0.88 (0.32; 2.39)	NA
	BIFAP	2.23 (0.77; 6.47)	0.46 (0.18; 1.18)	0.63 (0.20; 1.98)
	SIDIAP	0.40 (0.13; 1.27)	0.72 (0.20; 2.57)	1.12 (0.33; 3.77)
	RCGP RSC	1.93 (0.66; 5.65)	0.48 (0.18; 1.32)	0.67 (0.19; 2.30)
	THIN	0.31 (0.10; 0.98)	0.76 (0.21; 2.74)	1.23 (0.43; 3.50)
Fever	SSI	1.33 (1.21; 1.47)	0.83 (0.62; 1.11)	NA
	BIFAP	0.72 (0.67; 0.78)	0.96 (0.65; 1.43)	0.87 (0.56; 1.33)
	SIDIAP	0.58 (0.54; 0.62)	1.02 (0.78; 1.33)	0.93 (0.72; 1.21)
	RCGP RSC	1.12 (0.96; 1.30)	0.87 (0.61; 1.22)	0.75 (0.54; 1.04)
	THIN	1.01 (0.94; 1.08)	0.89 (0.60; 1.31)	0.77 (0.57; 1.04)
Persistent crying	SSI	NA	2.38 (1.55; 3.64)	NA
	BIFAP	1.60 (1.34; 1.91)	2.95 (2.56; 3.39)	2.95 (2.56; 3.39)
	SIDIAP	NA	2.38 (1.55; 3.64)	2.38 (1.55; 3.64)
	RCGP RSC	2.83 (2.18; 3.66)	2.19 (1.18; 4.06)	2.19 (1.18; 4.06)
	THIN	3.00 (2.54; 3.54)	2.11 (1.20; 3.68)	2.11 (1.20; 3.68)

BIFAP = Base de datos para la Investigación Famacoepidemiológica en Atención Primaria, RCGP RSC = Royal College of General Practitioners Research and Surveillance Centre, SIDIAP = Information System for Research in Primary Care, SSI = Statens Serum Institute, THIN = The Health Information Network, SCCS = Self Controlled Case Series. L-O-O = Leave-one-out

Box: Methods used to derive incidence rates in risk period following vaccination

(1) Derived from baseline IR (IR_bl):

The baseline IR in database *i* was multiplied by the L-O-O_IRR_ma calculated excluding database *i*. Confidence intervals (CIs) were obtained by calculating the standard error of the log IR_bl as follows:

The standard error of the sum of the log IR and the log L-O-O_IRR_ma was calculated as:

$$\sqrt{se(\log(IR))^2 + se(\log(L-O-O\_IRR\_ma))^2} \qquad (1)$$

where

$$se(\log(IR)) = \frac{1}{\sqrt{N\_events}} \qquad (2)$$

(2) Derived from background IR (IR_bg):

The background IR of each outcome in the month of age when the first dose was recommended in the country of database, *i*, was multiplied by the L-O-O_IRR_ma calculated excluding database *i*. CIs were obtained by calculating the standard error of the log IR_bg as in equations (1) and (2).

(3) Derived via meta-analysis of risk period IRs (IR_ma):

The log-transformed risk period IRs of all databases except database *i* were meta-analysed, providing IR_ma. CIs were obtained using the DerSimonian and Laird method for random effects meta-analysis [10].

Appendix: 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)

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/)

Figure 1

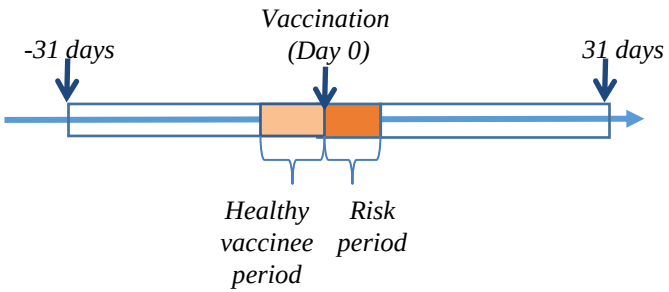

Vaccination
(Day 0)
-31 days
31 days
Healthy
vaccinee
period
Risk
period

Figure 2

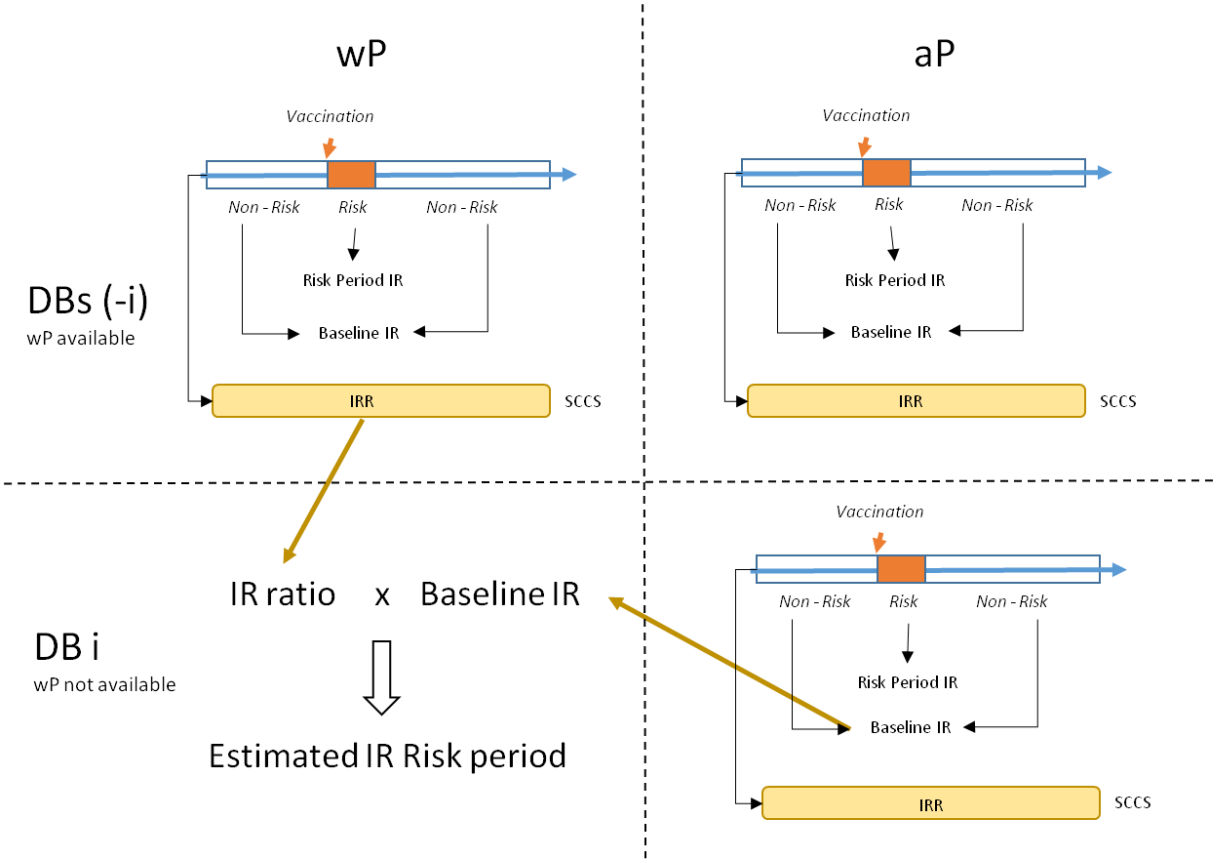
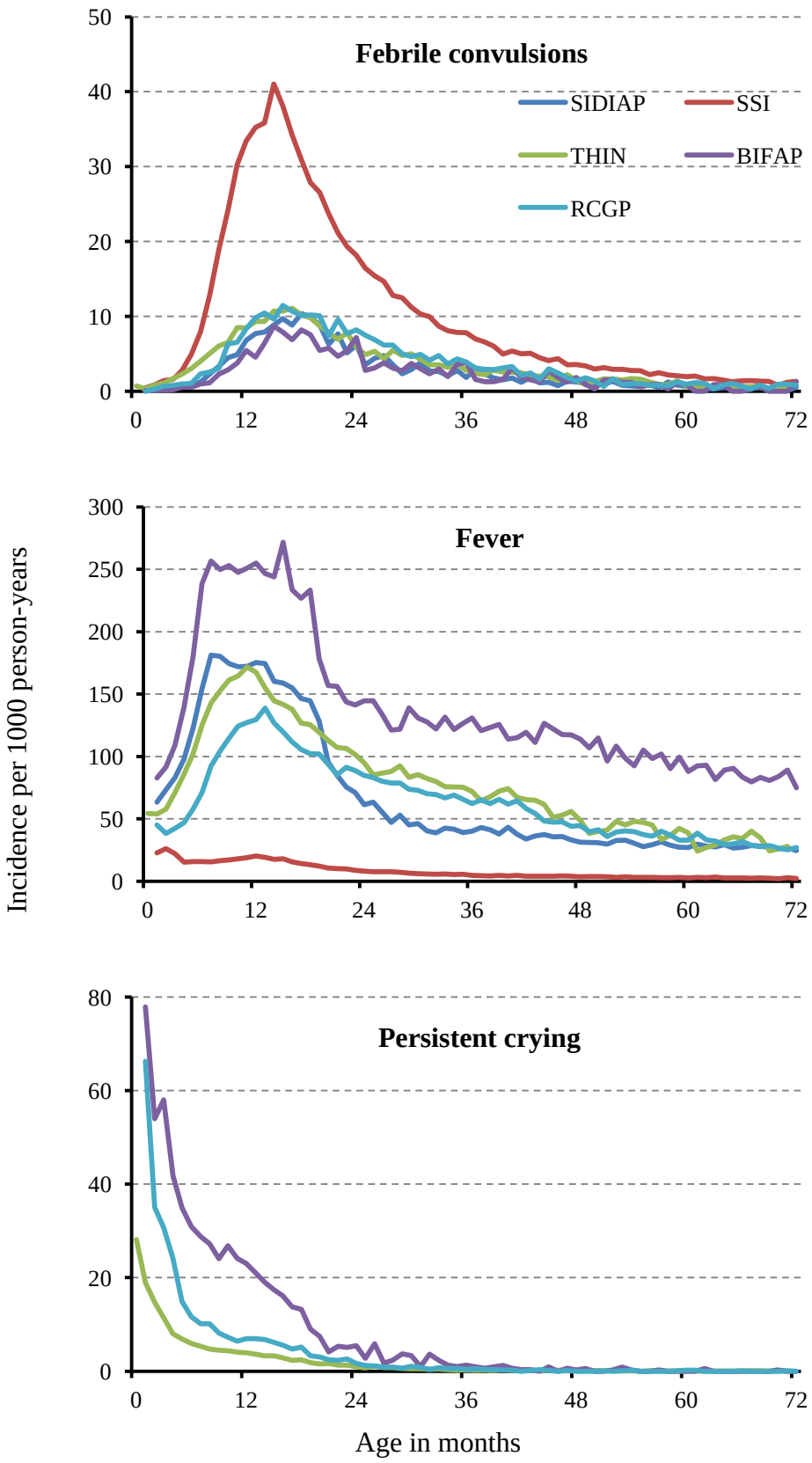

wP
aP
Vaccination
Non - Risk
Risk
Non - Risk
Risk Period IR
Baseline IR
IRR
SCCS
DBs (-i)
wP available
IR ratio x Baseline IR
DB i
wP not available
Estimated IR Risk period

Figure 3



SSI = Statens Serum Institute, BIFAP = Base de datos para la Investigación Famacoepidemiológica en Atención Primaria, SIDIAP = Information System for Research in Primary Care, RCGP RSC = Royal College of General Practitioners Research and Surveillance Centre, THIN = The Health Information Network

Figure 4A

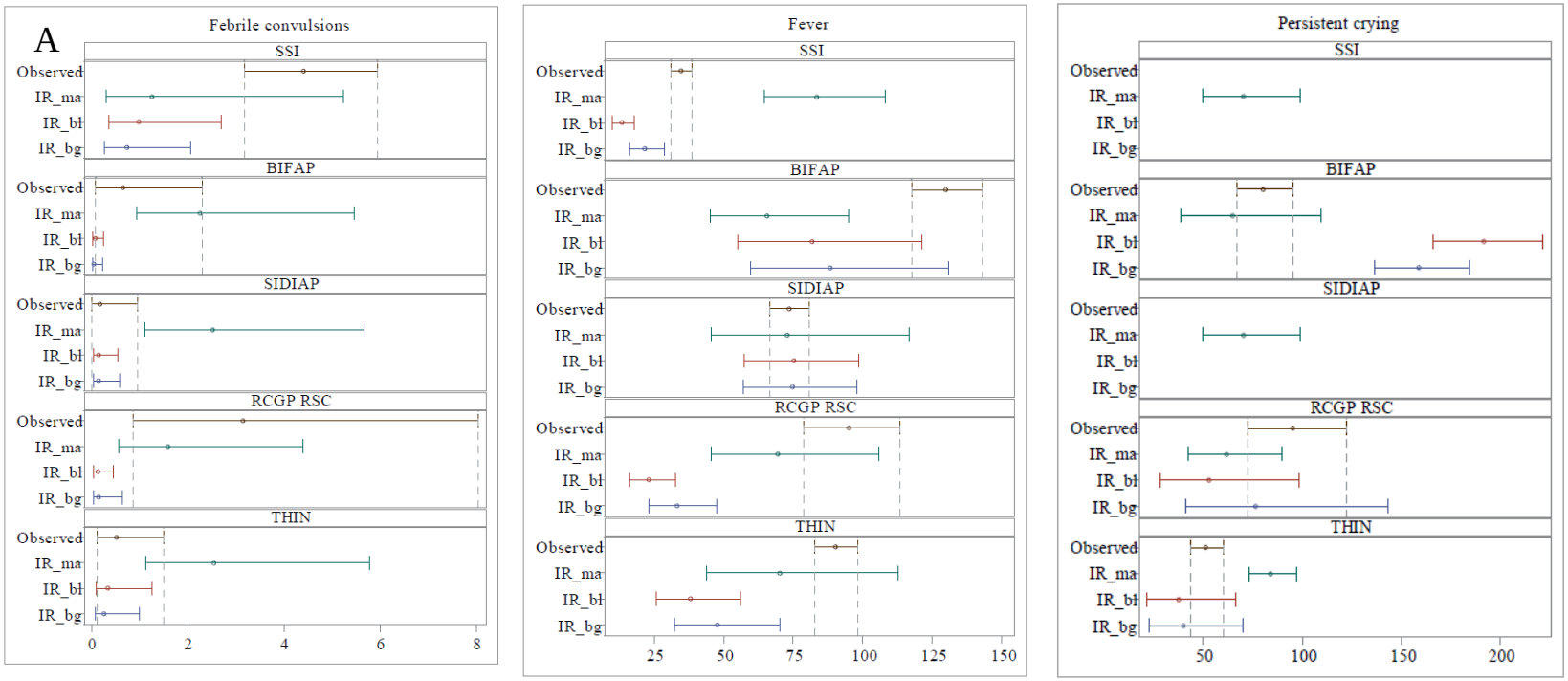
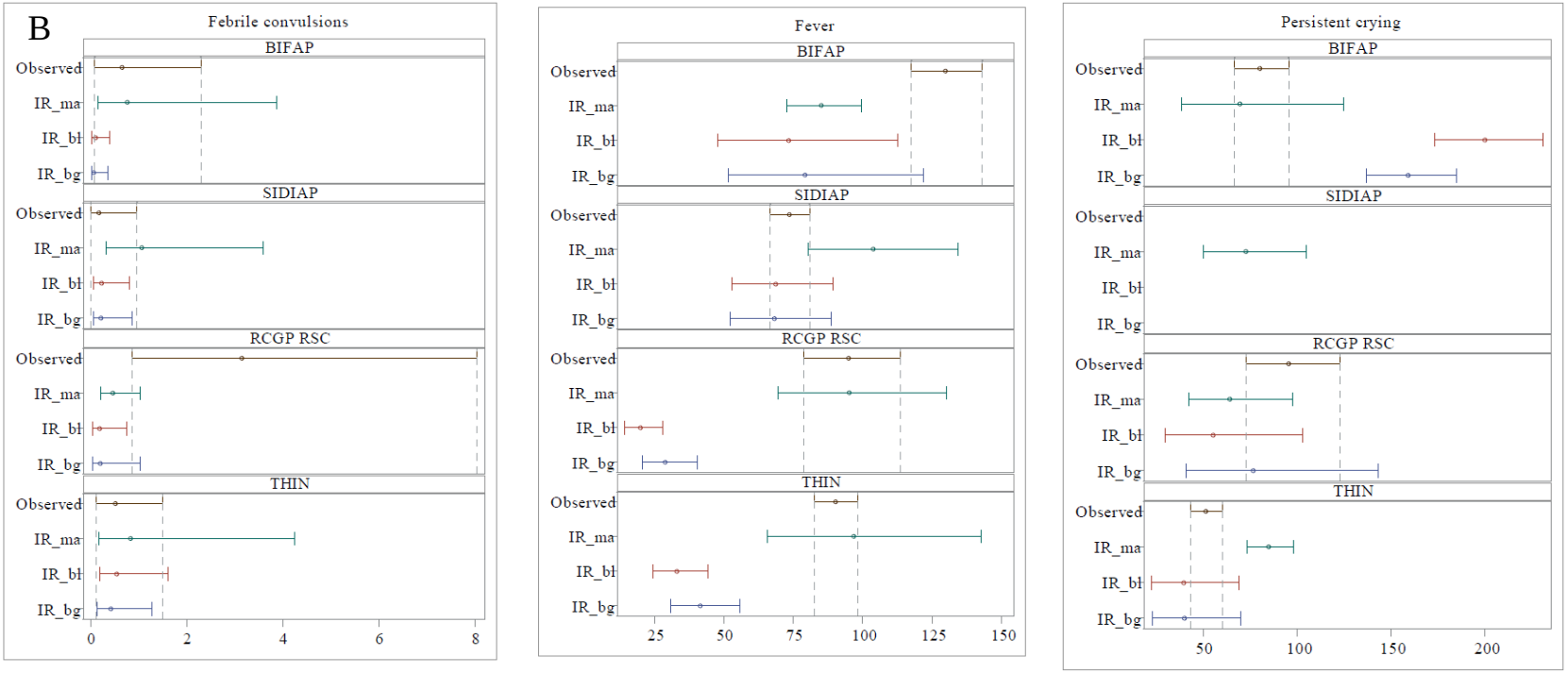

A
Febrile convulsions
Fever
Persistent crying
SSI
BIFAP
SIDIAP
RCGP RSC
THIN
Observed
IR_ma
IR_bl
IR_bg
0
2
4
6
8
25
50
75
100
125
150
50
100
150
200

Figure 4B


B
Febrile convulsions
BIFAP
SIDIAP
RCGP RSC
THIN
Observed
IR_ma
IR_bl
IR_bg
0
2
4
6
8
Fever
25
50
75
100
125
150
Persistent crying
50
100
150
200