Innovative Medicines Initiative

Accelerated Development of VAccine beNefit-risk Collaboration in Europe

Grant Agreement nº115557

D4.4 Report on tested methods for accelerated assessment of vaccination coverage, vaccine benefits, risks and benefit-risk (compilation of the D4.4, D4.5, D4.6, D4.7, D4.8)

WP4 Methods for burden of disease, vaccination coverage, vaccine safety and effectiveness, impact and benefit-risk monitoring

V1 Final
28 February 2017

Lead beneficiary: WP4
Date: 28/02/2017
Nature: Report
Dissemination level: PU

TABLE OF CONTENTS

TABLE OF CONTENTS **2**

DOCUMENT INFORMATION **6**

DOCUMENT HISTORY* **13**

DEFINITIONS **14**

EXECUTIVE SUMMARY **17**

INTRODUCTION 17
PROJECT1. POPULATION DIFFERENCES AS SOURCE OF HETEROGENEITY 19
PROJECT2. METHODS TO ESTIMATE VACCINATION COVERAGE FROM DYNAMIC POPULATIONS 19
PROJECT 3. IMPACT OF DISEASE- AND EXPOSURE-MISCLASSIFICATION ON ESTIMATIONS OF VACCINE EFFECTIVENESS 19
PROJECT 4. VALIDATION OF CASE-FINDING ALGORITHMS IN HEALTHCARE RESEARCH: ANALYTICAL INTERRELATIONS BETWEEN VALIDITY INDICES 20
PROJECT 5.HETEROGENEITY IN DISEASE MISCLASSIFICATION: THE COMPONENT ANALYSIS 20
PROJECT 6. LATENT CLASS MODELS TO ESTIMATE VALIDITY OF CASE-FINDING ALGORITHMS WHEN THERE IS NO REFERENCE STANDARD 21
PROJECT 7. BENEFIT-RISK MONITORING OF VACCINES: A DASHBOARD 21
PROJECT 8. COMPOSITE BURDEN OF DISEASE MEASURES FOR ADVERSE EVENTS FOLLOWING IMMUNIZATION 22
PROJECT 9. CODEMAPPER: SEMI-AUTOMATIC CODING OF CASE DEFINITIONS 22

1. SELECTION OF PROJECTS **24**

1.1. RESEARCH PROPOSALS 24
1.2. SELECTED RESEARCH PROJECTS 26

2. PROJECT 1: POPULATION DIFFERENCES AS SOURCE OF HETEROGENEITY **27**

DOCUMENT HISTORY **27**

2.1. INTRODUCTION 28
2.2. METHODS 29
2.3. RESULTS 30
2.3.1. National population characteristics *30*
2.3.2. Databases population characteristics *33*
2.3.3. Databases comparison: entry in database and duration of follow-up *37*
2.4. DISCUSSION 38
2.5. REFERENCES 39

3. PROJECT 2: METHODS TO ESTIMATE VACCINATION COVERAGE FROM DYNAMIC POPULATIONS **40**

DOCUMENT HISTORY **40**

3.1. INTRODUCTION 41
3.2. METHODS 41

3.2.1. Cumulative distribution function method *41*
3.2.2. Inverse probability weighting method *43*
3.3. SIMULATION STUDY 44
3.3.1. Simulation model *44*
3.3.2. Results *46*
3.4. DISCUSSION 47
3.5. REFERENCES 48

4. PROJECT 3: IMPACT OF DISEASE- AND EXPOSURE MISCLASSIFICATION ON ESTIMATION OF VACCINE EFFECTIVENESS 49

DOCUMENT HISTORY 49

4.1. INTRODUCTION 50
4.2. METHODS 50
4.2.1. Notation *51*
4.2.2. Impact of misclassification at population level *51*
4.2.3. Simulation tool *53*
4.2.4. Scenarios *56*
4.3. RESULTS 56
4.3.1. Paediatric seasonal influenza *56*
4.3.2. Pertussis primary series *58*
4.3.3. Simulation tool *60*
4.4. DISCUSSION 61
4.5. REFERENCES 62

5. PROJECT 4: VALIDATION OF CASE-FINDING ALGORITHMS IN HEALTHCARE RESEARCH: ANALYTICAL INTERRELATIONS BETWEEN VALIDITY INDICES 64

DOCUMENT HISTORY 64

5.1. INTRODUCTION 65
5.2. METHODS 65
5.2.1. Definitions *65*
5.2.2. Interrelationships between validity indices *66*
5.2.3. Web-application *70*
5.2.4. Sensitivity analyses *70*
5.3. RESULTS 71
5.3.1. Illustrations *71*
5.3.2. Sensitivity analyses *72*
5.4. DISCUSSION 74
5.5. REFERENCES 76

6. PROJECT 5: HETEROGENEITY IN DISEASE MISCLASSIFICATION: THE COMPONENT ANALYSIS 78

DOCUMENT HISTORY 78

6.1. INTRODUCTION 79
6.2. METHODS 80
6.2.1. Specification of a component *80*

6.2.2. Selection of components *80*
6.2.3. Collection of a-priori knowledge *81*
6.2.4. Data-extraction *82*
6.2.5. Local data processing *82*
6.2.6. Analysis *83*
6.3. RESULTS 83
6.3.1. Selected components: results from literature review *83*
6.3.2. Selected components for pertussis *83*
6.3.3. Collecting a-priori knowledge: settings and semantics in databases *87*
6.3.4. Collecting a-priori knowledge: evidence on background rates for pertussis *87*
6.3.5. Data extraction and data processing *87*
6.3.6. Analysis *87*
6.4. DISCUSSION 88
6.5. CONCLUSION 91
6.6. REFERENCES 91

7. PROJECT 6: LATENT CLASS MODELS TO ESTIMATE VALIDITY OF CASE-FINDING ALGORITHMS WHEN THERE IS NO REFERENCE STANDARD 93

DOCUMENT HISTORY 93

7.1. INTRODUCTION 94
7.2. METHODS 94
7.3. ILLUSTRATION 95
7.3.1. Date generation *95*
Table 7.3:3 tests-1 population (conditional independence): example data. *97*
7.3.2. Estimation *97*
7.3.3. Results *97*
7.4. DISCUSSION 98
7.5. REFERENCES 99

8. PROJECT 7: BENEFIT-RISK MONITORING OF VACCINES: A DASHBOARD 101

DOCUMENT HISTORY 101

8.1. INTRODUCTION 102
8.2. METHODS 103
8.2.1. Data simulation *103*
8.2.2. Near real-time benefit-risk monitoring *105*
8.2.3. Visualizations *105*
8.2.4. Demonstration with fictitious data *107*
8.3. DISCUSSION 116
8.4. REFERENCES 117

9. PROJECT 8: COMPOSITE BURDEN OF DISEASE MEASURES FOR ADVERSE EVENTS FOLLOWING IMMUNIZATION 119

DOCUMENT HISTORY 119

9.1. INTRODUCTION 120
9.2. METHODS 121

9.2.1. Selection of candidate adverse events *121*
Literature search for relative risks *122*
9.2.2. Burden of disease methods and required parameters *122*
9.2.3. Selection of parameters for the example burden calculation *123*
9.2.4. Computational details *125*
9.2.5. Regulatory status *127*
9.3. RESULTS 127
9.3.1. Selection of adverse events *127*
9.3.2. Example YLD computation *127*
9.4. DISCUSSION 128
9.4.1. Challenges and limitations *129*
9.4.2. Recommendations *131*
9.4.3. Conclusions *131*
9.5. REFERENCES 131
10. PROJECT 9: CODEMAPPER: SEMI-AUTOMATIC CODING OF CASE DEFINITIONS 134
DOCUMENT HISTORY 134
10.1. INTRODUCTION 135
10.1.1. Coding heterogeneity in the EU *135*
10.1.2. Prior workflows and pathways to bridge the heterogeneity *135*
10.2. METHODS 135
10.2.1. Mapping approach *136*
10.2.2. User application *137*
10.2.3. Evaluation *139*
10.2.4. Error analysis *142*
10.3. RESULTS 143
10.3.1. Baseline *143*
10.3.2. Concept expansion *143*
10.3.3. Error analysis *143*
10.4. DISCUSSION 144
10.5. CODEMAPPER IN ADVANCEWP5 145
10.5.1. The intended mapping process *145*
10.5.2. Observations and recommendations *148*
10.5.3. Medical concepts selected in terminology mapping for the POC study in ADVANCE WP5 ..... *155*
10.6. REFERENCES 157
11. APPENDICES 159
11.1. APPENDIX P1 (COVERAGE) 159
11.2. APPENDIX P6 (BR DASHBOARD) 171
11.3. APPENDIX P5 (COMPONENTS) 174
11.4. APPENDIX P7 (BOD) 179

DOCUMENT INFORMATION

Grant Agreement Number	115557	**Acronym**	ADVANCE
Full title	Accelerated Development of VAccine beNefit-risk Collaboration in Europe		
Project URL	http://www.advance-vaccines.eu		
IMI Project officer	Angela Wittelsberger (angela.wittelsberger@imi.europa.eu)		

Deliverable	**Number**	4.5	**Title**	Methods development–final
Work package	**Number**	4	**Title**	Methods for burden of disease, vaccination coverage, vaccine safety and effectiveness, impact and benefit-risk monitoring

Delivery date	**Contractual**	Month	**Actual**	Month
Status	V1		Final	
Nature	Report **X** Prototype ☐ Other ☐			
Dissemination Level	Public ☐ Confidential**X**			

Co-Authors (WP4 leads and task leads)

Name	**Partner**	**Section(s)**
Kaatje Bollaerts (WP4 lead)	P95	Project 3, 4, 6
Hanne-Dorthe Emborg	SSI	Project 2
Scott McDonald	RIVM	Project 8
Caitlin Dodd	EMC	Project 1 and 5
Lina Titievsky	Pfizer	Project 7
Benedikt Becker	EMC	Project 9
John Weil (WP4 lead)	Takeda	

Project management

Name	**Partner**	**Section(s)**
Antonella Chiucchiuini	Takeda	

Contributors

Name	**Partner**	**Section(s)**
Maria de Ridder	EMC	Project 1
Caitlin Dodd	EMC	
Swabra Nakato	EMC	
Klara Berensci	AUH	
Talita Duartes-Salles	SIDIAP	
Elisa Martin, Consuelo Huerta	BIFAP	
Gino Picelli, Lara Tramontan, Giorgia Danieli	PEDIANET	
Ana Correa	RCGP	
Daniel Weibel	EMC	
Hanne-Dorthe Emborg	SSI	
Miriam Sturkenboom	EMC	

Name	**Partner**	**Section(s)**
Kaatje Bollaerts	P95	Project 2
Toon Braeye	WIV	
Tom Cattaert	P95	
Jose van Boxmeer	Seqirus (under SAPA of Novartis)	
Vincent Bauchau	GSK	
Miriam Sturkenbooom	EMC	
Hanne-Dorthe Emborg	SSI	

Name	**Partner**	**Section(s)**
Tom De Smedt	P95	Project 3
Elizabeth Merrall	GSK	
Denis Macina	SP	
Silvia Perez-Vilar	FISABIO	
Nick Andrews	PHE	
Kaatje Bollaerts	P95	

Name	Partner	Section(s)
Kaatje Bollaerts	P95	Project 4
Alexandros Rekkas	P95	
Tom De Smedt	P95	
Caitlin Dodd	EMC	
Nick Andrews	PHE	
Rosa Gini	ARS	

Name	Partner	Section(s)
Rosa Gini	ARS	Project 5
Caitlin Dodd	EMC	
Kaatje Bollaerts	P95	
Claudia Bartolini	ARS	
Giuseppe Roberto	ARS	
Consuelo Huerta	BIFAP	
Elisa Martin	BIFAP	
Talita Duarte Salles	SIDIAP	
Gino Picelli	PEDIANET	
Lara Tramontan	PEDIANET	
Benedikt Becker	EMC	
Charlotte Switzer	SP	
Sonja Banga	SP	
Jorgen Bauwens	UKBB/UNIBAS	
Daniel Weibel	EMC	
Miriam Sturkenboom	EMC	

Name	Partner	Section(s)
Kaatje Bollaerts	P95	Project 6
Alexandros Rekkas	P95	
Tom De Smedt	P95	
Rosa Gini	ARS	
Nicolas Praet	GSK	

Name	Partner	Section(s)

	Kaatje Bollaerts	P95	Project 7
	Tom de Smedt	P95	
	Katherine Donegan	MHRA	
	Vincent Bauchau	GSK	
	Lina Titievsky	Pfizer	

	Name	**Partner**	**Section(s)**
	Scott McDonald	RIVM	Project 8
	Susan Hahné	RIVM	
	Danielle Nijsten	RIVM	
	Kaatje Bollaerts	P95	
	Tom De Smedt	P95	
	Cassandra Nan	GSK	
	Nicolas Praet	GSK	
	Jorgen Bauwens	UKBB/UNIBAS	
	Vincent Bauchau	GSK	
	Miriam Sturkenboom	EMC	

	Name	**Partner**	**Section(s)**
	Benedikt Becker	EMC	Project 9
	Paul Avillach	EMC, Harvard Medical School	
	Silvana Romio	EMC, University of Milano-Bicocca	
	Erik van Mulligen	EMC	
	Miriam Sturkenboom	EMC	
	Jan Kors	EMC	
	Jorgen Bauwens	UKBB/UNIBAS	

Reviewers	**Name**	**Partner**	**Section(s)**
	Kaatje Bollaerts	P95	Project 1
	Miriam Sturkenboom	EMC	
	Vincent Bauchau	GSK	
	Linda Levesque	SP	

Mendel Haag	Seqirus (under SAPA of Novartis)	
Sonja Banja	SP	
Alena Khromava	SP	

Name	**Partner**	**Section(s)**
Sonja Banja	SP	Project 2
Alena Khromava	SP	
Miriam Sturkenboom	EMC	
Mendel Haag	Seqirus (under SAPA of Novartis)	
Linda Levesque	SP	

Name	**Partner**	**Section(s)**
Mendel Haag	Seqirus(under SAPA of Novartis)	Project 3
Marianne van der Sande	RIVM	
Sonja Banda	SP	
Charlotte Switzer	SP	
Linda Levesque	SP	

Name	**Partner**	**Section(s)**
Linda Levesque	SP	Project 4
Marianne van der Sande	RIVM	
Alena Khromava	SP	

Name	**Partner**	**Section(s)**
Kaatje Bollaerts	P95	Project 5
Miriam Sturkenboom	EMC	
Jorgen Bauwens	UKBB/UNIBAS	
Mendel Haag	Seqirus (under SAPA of Novartis)	
Marianne van der Sande	RIVM	

Name	**Partner**	**Section(s)**
Marianne van der Sande	RIVM	Project 6

Name	**Partner**	**Section(s)**
Nick Andrews*	PHE	Project 7
Julia Stowe*	PHE	
Alena Khromava*	SP	
Steven Bailey*	Pfizer	
Raphaele Roten*	Crucell	
Katharine Hartmann*	Takeda	
Mirjam Knol*	RIVM	
Robert Maroko*	Pfizer	
Paola Pirrotta*	GSK	
Mendel Haag	Seqirus (under SAPA of Novartis)	
Marianne van der Sande	RIVM	
Sonja Banja	SP	

*attended user-testing: the dashboard has been demonstrated and feedback on
usefulness was requested.

Name	**Partner**	**Section(s)**
Vincent Bauchau	GSK	Project 8
Nicolas Praet	GSK	
Jorgen Bauwens	UKBB/UNIBAS	
Miriam Sturkenboom	EMC	
Sonja Banja	SP	

Name	**Partner**	**Section(s)**
Simon de Lusignan	SURREY	Project 9
Patrick R. Mahy	WIV-ISP	
Marianne van der Sande	RIVM	
Vincent Bauchau	GSK	
Jorgen Bauwens	UKBB/UNIBAS	

Responsible Author	Kaatje Bollaerts	**Email**	kaatje.bollaerts@p-95.com
	Partner	P95	

Description of the deliverable	This deliverable describes methods developed with the objective of accelerating the assessment of vaccination coverage, benefits, risks and benefit-risk of vaccines to support decision-making on vaccines in Europe.
Key words	Vaccine burden of disease, vaccine effectiveness, impact, safety, benefit-risk.

DOCUMENT HISTORY*

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts (P95)	1 Oct 2016	-	Outline
Kaat Bollaerts(P95)	09 Nov2016	-	Add project 3
Benedikt Becker (EMC)	11 Nov 2016	-	Add project 9
Scott McDonald (RIVM)	6 Dec 2016	-	Add project 8
Kaat Bollaerts(P95)	20 Jan 2017	-	Add project 4
Scott McDonald (RIVM)	2 Feb 2017	-	Revised project 8
Kaat Bollaerts(P95)	5 Feb 2017	-	Add project 7
Kaat Bollaerts(P95)	8 Feb 2017		Add project 2
Kaat Bollaerts(P95)	10 Feb 2017		Add project 6
Kaat Bollaerts(P95)	13 Feb 2017		Add projects 1 and 5, editing
Vincent Bauchau (GSK), Miriam Sturkenboom (EMC), Jorgen Bauwens (UKBB/UNIBAS), Marianne van der Sande (RIVM), Alena Khromava (SP), Linda Levesque (SP), Sonja Banja (SP), Mendel Haag (Seqirus)			SC review
Keith Veitch (Takeda)	26 Feb 2017		Editing
Kaat Bollaerts (P95), Maria de Ridder (EMC), Scott McDonald (RIVM), Rosa Gini (ARS)	27 Feb 2017		Addressing comments
Kaat Bollaerts (P95)	28 Feb 2017		Final version

*Document history from the compiled document. The detailed document histories of the separate projects are given at the beginning of each project.

DEFINITIONS

Participants of the ADVANCE Consortium are referred to herein according to the following codes:

- **AEMPS.** Agencia Española de Medicamentos y Productos Sanitarios (Spain)
- **ASLCR.** Azienda Sanitaria Locale della Provincia di Cremona (Italy)
- **AUH.** Aarhus Universitetshospital (Denmark)
- **BC.** Brighton Collaboration
- **CRX**. Crucell Holland BV (Netherlands)
- **ECDC.** European Centre for Disease Prevention and Control (Sweden)
- **EMA.** European Medicines Agency (United Kingdom)
- **EMC**. Erasmus Universitair Medisch Centrum Rotterdam (Netherlands) - Coordinator
- **GSK.** GlaxoSmithKline Biologicals, S.A. (Belgium) – EFPIA Coordinator
- **KI.** Karolinska Institutet (Sweden)
- **LSHTM.** London School of Hygiene and Tropical Medicine (United Kingdom)
- **MHRA.** Medicines and Healthcare products Regulatory Agency (United Kingdom)
- **NOVARTIS.** Novartis Pharma AG (Switzerland)
- **OU.** The Open University (United Kingdom)
- **P95.** P95 (Belgium)
- **PEDIANET.** Società Servizi Telematici SRL (Italy)
- **PFIZER**. Pfizer Limited (United Kingdom)
- **RCGP.** Royal College of General Practitioners (United Kingdom)
- **RIVM.** Rijksinstituut voor Volksgezondheid en Milieu * National Institute for Public Health and the Environment (Netherlands)
- **SP MSD.** Sanofi Pasteur MSD (France)
- **SP.** Sanofi Pasteur (France)
- **SSI.** Statens Serum Institut (Denmark)
- **SURREY.** The University of Surrey (United Kingdom)
- **SYNAPSE.** Synapse Research Management Partners, S.L. (Spain)
- **TAKEDA**. Takeda Pharmaceuticals International AG(Switzerland)
- **UNIBAS**. Universitaet Basel (Switzerland) - **Managing entity of the IMI JU funding**
- **UTA.** Tampereen Yliopisto (Finland)
- **WIV-ISP.** Institut Scientifique de Santé Publique (Belgium)

- **Analytical dataset**. The minimum set of data required to perform the statistical analyses leading to the results for the primary objective(s) of the study.

- **Consortium.** The ADVANCE Consortium, comprising the above-mentioned legal entities.
- **Grant Agreement.** The agreement signed between the beneficiaries and the IMI JU for the undertaking of the ADVANCE project (115557).
- **Primary data collection.** Data collection directly from healthcare professionals or consumers (e.g. prospective observational studies and registries in which the data collected derive from routine clinical care.
- **Project Agreement.** Agreement concluded amongst ADVANCE participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.
- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.
- **Secondary data collection**: Secondary use of data previously collected from consumers or healthcare professionals for other purposes and where all the events of interest have already happened. Examples include medical chart reviews (including following-up on data with healthcare professionals), analysis of electronic healthcare records, systematic reviews, meta-analyses. Study designs may include case-control, cross-sectional, cohort or other study designs making secondary use of data.
- **Start of data collection.** The date from which information on the first study subject is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.

List of abbreviations:

- AE: adverse event
- AESI: adverse event of special interest
- AF: attributable fraction
- B/R: Benefit-risks
- CDF: cumulative distribution function
- CFA: case-finding algorithm
- CPRD: Clinical Practice Research Database
- CUI: concept unique identifier
- DALY: disability-adjusted life-year
- EHR: Electronic health record
- IBRR: incremental benefit-risk ratio
- ICD-9 CM: International Classification of Diseases version 9 Clinical Modifications
- ICD-10: International Classification of Diseases version 10
- ICPC-2: International Classification of Primary Care Version 2
- INHB: incremental net health benefit
- IS: intussusception
- IPW: inverse probability weighting
- LCM: Latent Class Model

- MeSH: Medical Subject Headings
- MedDRA: Medical Dictionary for Regulatory Activities
- MSE: mean squared error
- NPV: negative predictive value
- POC: proof of concept
- PPV: positive predictive value
- Read-2: Read codes version 2
- Read-CTv3: Read Clinical Terms version 3
- RI: Relative incidence
- SAFEGUARD: Safety Evaluation of Adverse Reactions in Diabetes
- SE: sensitivity
- SP: specificity
- SNOMED-CT: Systematized Nomenclature of Medicine - Clinical Terms
- UMLS: Unified Medical Language System
- VE: vaccine effectiveness
- VPD: vaccine preventable disease
- RVGE: rotavirus gastroenteritis
- YLD: years lived with disability
- YLL: years of life lost
- WHO: World Health Organization

EXECUTIVE SUMMARY

Introduction

The ADVANCE vision is to deliver the "best evidence at the right time to support decision-making on vaccination in Europe". The mission is to establish a prototype of a sustainable and compelling system that rapidly provides best available scientific evidence on vaccination benefits and risks post-marketing for well-informed decisions. In light of this goal, the ADVANCE work package on methods (WP4) aims to develop methods to rapidly estimate burden of disease, vaccination coverage, safety, benefits and benefit-risk to support decision-making by all stakeholders.

This report describes the methods that have been developed to support ADVANCE mission and vision.

As it is crucial to the ADVANCE vision and mission to provide timely evidence, much of the methodological research is dedicated to the secondary use of electronic health record (EHR) data (e.g. hospital admissions, GP-records) for epidemiological research. Indeed, the main advantage of using such secondary data sources is that they already exist so the time spent on the study is likely to be less than the time spent on studies that use primary data collection. However, as EHR data are collected primarily for clinical and administrative use rather than for research, concerns regarding completeness, data validity and limited ability to control for confounding exist. The project 'Population differences as source of heterogeneity' (**project 1**) looks at the differences in populations in countries participating in ADVANCE and differences incompleteness of follow-up of between databases. The project 'Methods on estimating vaccination coverage from dynamic populations' (**project 2**) is on estimating vaccination coverage when subjects are incompletely followed-up as failing to account for incomplete follow-up would lead to an underestimation of the vaccination coverage. The projects 'Impact of disease- and exposure-misclassification on estimation of vaccine effectiveness' (**project 3**), 'Validation of case-finding algorithms in healthcare research: analytical interrelations between validity indices' (**project 4**), 'Heterogeneity in disease misclassification: the component analysis' (**project 5**) and 'Latent Class Models to estimate validity when there is no reference standard' (**project 6**) all addressed the issue of (exposure- and) disease misclassification, which might result in strongly biased disease/exposure occurrence and risk estimates. In project 3, we quantified the potential impact of disease- and exposure misclassification on the estimates of vaccine effectiveness in order to assess study feasibility, and possibly, the need to correct for misclassification. If corrections for misclassification are deemed necessary, validation studies will be needed to obtain estimates of validity. The typically obtained estimates of validity are sensitivity, specificity, positive and negative predictive value. In project 4, we explored the analytical interrelations between the different validity indices. The analytical interrelations allow the conversion of validity indices and the obtention of estimates of the true prevalence corrected for

misclassification. For both projects 3 and 4, we developed web-applications to allow easy use of our results. Project 5 introduced the idea of building component algorithms to identify diseased subjects from EHR, which can then be combined with the techniques of Latent Class Modelling as introduced in Project 6 to obtain estimates of validity in the absence of a reference standard.

Furthermore, EHR data are reflective of healthcare use and reporting practices, which are different from country to country. As such, substantial heterogeneity is expected among estimates from EHR data across databases and countries. The projects 'Population differences as source of heterogeneity' (project 2) and 'Heterogeneity in disease misclassification: the component analysis' (project 5) are aimed at exploring and understanding the differences in the European databases. Any differences might stem from differences in databases populations (selection bias), differences in exposure- and event misclassification (misclassification bias), differences in the ability to control for confounding or real differences (effect modification).

Two projects were specifically aimed at developing benefit-risk methodology for vaccines. In the project 'benefit-risk monitoring of vaccines: a dashboard' (**project 7**) we developed methodology to allow for monitoring of the vaccine's benefit-risk and its components; vaccination coverage, safety, vaccine effectiveness and impact. Specifically, we developed an interactive dashboard web-application allowing the visual monitoring of the vaccine's benefit-risk profile over time. The dashboard was made interactive to allow users to set their own benefit-risk preference weights, to select age groups and time windows of interest and to easily conduct sensitivity analyses. Finally, the project 'Burden of disease of adverse events following immunization' (**project 8**) assessed the feasibility and usefulness of adapting the composite measure of burden of disease, disability-adjusted life-years (DALY), for estimating the burden of adverse events. The DALY approach has been widely used to quantify population health impact of disease or injury, but has not been fully explored to estimate the burden of adverse events following immunization. A measure of benefit-risk can then be obtained by quantifying the burden of disease prevented through vaccination and the burden of adverse events following immunization using the DALY-approach and comparing both the prevented and induced burden of disease.

Finally, different EHR databases use different coding vocabularies to report the medical information (e.g. ICD-9, ICD-10, ICPC-2, Read). The harmonization of codes across databases and countries can pose an important bottleneck to the rapid implementation of collaborative epidemiological studies, within and between countries. In the project 'CodeMapper, semi-automatic coding of case definitions' (**project 9**), a web application was developed, which assists in the mapping of case definitions to codes from different vocabularies, while keeping a transparent record of the complete mapping process.

More detailed summaries of the different projects are given below.

Project1. Population differences as source of heterogeneity

Differences between the results of observational studies intended to establish the association between an exposure and outcome by looking at differences in disease occurrence may have various origins, including chance, information bias (e.g. disease or exposure misclassification), selection bias, confounding or effect modification. The pharmaco-epidemiological literature is filled with conflicting findings and it has been shown that results from observational database studies can be sensitive to the choice of database. In this project we aim to investigate the differences in the populations of countries participating in ADVANCE, the source populations in the databases in those countries, and the differences between follow-up of the population in the databases. This should give us a better understanding of potential sources of heterogeneity and provide insights on ways to extrapolate the results beyond the data used to generate the results.

Project2. Methods to estimate vaccination coverage from dynamic populations

The introduction of new vaccines and the evaluation of vaccination programs requires tools that can closely monitor the vaccination coverage. Vaccination coverage and adherence with the recommended vaccination schedules are widely used indicators of vaccination program performance. These performance indicators are typically measured using registries, routine administrative reports or household surveys. Electronic healthcare records (EHRs) are an alternative source to monitor vaccination uptake. EHRs allow in principle timely monitoring at a relatively low cost and often cover large geographical areas. This could also provide coverage information needed for rapid assessment of new safety or vaccine effectiveness concerns. However, the populations captured in EHR are generally dynamic, with members moving in and out the population over time (i.e. transient membership) for example due to relocation or switch between general practices. This often results in incomplete follow-up, hampering the accurate estimation of vaccination coverage from EHRs. Incomplete follow-up would lead to an underestimation of the vaccination coverage as vaccines administered outside the HER's follow-up period will not always be recorded. We explored two methods (the so-called inverse probability weighting method and the cumulative distribution method) to estimate vaccination coverage for dynamic populations and assessed their performance through simulation.

Project 3. Impact of disease- and exposure-misclassification on estimations of vaccine effectiveness

Studies of vaccine safety and vaccine effectiveness (VE) rely on accurate identification of vaccination and cases of vaccine-preventable disease. In practice, diagnostic tests, clinical case definitions and vaccination records often present inaccuracies, leading to biased effect estimates. Misclassification

of these variables is a particular concern for observational database research. Previous simulation studies assessing the impact of misclassification on VE assumed non-differential misclassification and did not account for exposure misclassification. We developed a web-application to assess, through simulation, the potential (joint) impact of disease- and exposure-misclassification when estimating VE using cohort, case-control, test-negative and case-coverage designs. The misclassification can be differential or non-differential. The impact of misclassification on the estimated VE is presented graphically. We demonstrated the application on childhood seasonal influenza and pertussis vaccinations. Depending on the scenario, the misclassification parameters had differing impacts. Decreased exposure specificity (poorer identification of non-vaccinees) had greatest impact for influenza VE estimation when vaccination coverage was low. Decreased exposure sensitivity (poorer identification of vaccinees) had greatest impact for pertussis VE estimation for which high vaccination coverage is typically achieved. The impact of the misclassification parameters was found to be more noticeable than that of the different study designs. To conclude, misclassification can lead to significant bias in VE estimates and its impact strongly depends on the scenario. Therefore, our developed web-application for assessing the potential impact of misclassification can be modified by users to accommodate their own study. If the potential bias is believed to be unacceptable, corrective measures might have to be taken.

Project 4. Validation of case-finding algorithms in healthcare research: analytical interrelations between validity indices

Validation is recognized as an important component of research using healthcare databases. The typically evaluated validity indices of case-finding algorithms include sensitivity, specificity, positive and negative predictive values. These validity indices, as well as the observed and true disease prevalence are interrelated. For every combination of the observed prevalence and two other parameters, we derived the analytical expressions to obtain the remaining three parameters. We developed a web-application that calculates validity indices given user-defined values of the observed prevalence and any other two parameters, with the 95% uncertainty intervals of the derived parameters obtained through Monte Carlo simulation. We also conducted sensitivity analyses investigating the impact of estimation error in the input parameters on the derived parameters. This tool allows users to easily convert validity indices facilitating the comparison of results from validation studies and to derive estimates of the true prevalence for any combination of the observed prevalence and any two validity indices.

Project 5.Heterogeneity in disease misclassification: the component analysis

Detecting the occurrence of diseases in persons belonging to database populations is a key step in electronic healthcare record (HER) research, with wrongly classifying persons as diseased/non-

diseased being a major concern. A common strategy to identify disease in an EHR database is to extract the records with a diagnosis code or mentioning of that disease. Since EHR data capture more components than codes alone, e.g. drug prescriptions, procedures and tests, different combinations can be made to increase the sensitivity and specificity of the case-finding algorithms. In this project, we explore a different strategy to estimate disease misclassification in databases, called *component analysis*. Using sources of data other than diagnoses alone might alter the sensitivity and specificity of the event. The interplay between the common strategy of using all information and the component analyses strategy sheds light on the validity of the different strategies. This allows the design of sensitivity analyses allowing the investigation of whether differences in validity across databases explains or partially explains heterogeneity in the study results.

In this project, we show the preliminary results of the strategy applied to two health outcomes relevant for the work package 5 Proof-of-Concept study on pertussis vaccination: pertussis and convulsions.

Project 6. Latent Class Models to estimate validity of case-finding algorithms when there is no reference standard

Validation is recognized as an important component of research using electronic health care records (EHRs). Validation of case-finding algorithms (CFA) is typically obtained by comparing the CFA results with those from a reference standard. In electronic healthcare research a reference standard is often obtained through chart review or through asking healthcare professionals to complete a questionnaire. However, these approaches are not always feasible and often very expensive and time consuming. An alternative approach using Latent Class Modelling for which no reference standard is needed has been applied frequently to validate diagnostic tests, particularly in veterinary science. A Latent Class Model (LCM) treats the true disease status as an unmeasured (aka latent) categorical (typically binary) variable whereas the observed measurements of disease are treated as imperfect classifiers of the true disease status. To our knowledge, LCMs have not yet been applied to estimate the validity of CFAs from EHRs. In this work, we give an overview of commonly used LCM methodology and illustrate their performance through simulation. In the next step, this work will be integrated with the component analyses work (see Project 5) to explore whether LCMs can be successfully used to estimate validity of CFAs obtained via component analyses from EHR.

Project 7. Benefit-risk monitoring of vaccines: a dashboard

There are several, typically not integrated, parts of post-licensure or post-marketing vaccine surveillance: the surveillance of vaccination coverage and adherence with the recommended vaccination schedule, vaccine safety, effectiveness and impact. An increasing interest exists in near real-time surveillance using electronic healthcare databases. Although quantitative benefit-risk

assessments, for which the benefits of a medical intervention are offset by its risks at one point-in-time, are increasingly being performed, integrated post-marketing monitoring of coverage, benefits, risks, and benefit-risk of vaccines is – to the best of our knowledge - not yet implemented in practice in a quantitative manner. With this work, we explore methodology for near real-time benefit-risk monitoring/surveillance of vaccines using electronic healthcare databases. We visualize key data for monitoring vaccination coverage, benefits, and safety. These are then combined into composite measures of the vaccine's benefit-risk profile as it evolves over time. To facilitate the monitoring, we developed an interactive dashboard. We illustrated the dashboard using simulated data reflective of the introduction of the rotavirus vaccination in the UK.

Project 8. Composite Burden of Disease measures for adverse events following immunization

Composite measures of disease burden such as disability-adjusted life-years (DALY) have been widely used to quantify the population-level health impact of disease or injury, but have not yet been applied to estimate the burden of adverse events following immunization. Our objective was to assess the feasibility and usefulness of adapting the DALY methodology to estimate adverse event burden related to vaccination. To this end, we developed a practical methodological framework, explicitly describing all steps involved, from criteria for selection of events, through retrieval of parameters and background incidence rates from the literature, to computation of the years lived with disability (YLD) measure, with estimations of uncertainty. We present a working example in which we estimate YLD for four adverse events following three childhood vaccines, based on published background incidence rates and relative and absolute risks. YLD provided extra insight into the health impact of an adverse event over presentation of incidence rates only, as the severity and duration of the adverse events are additionally incorporated. We conclude that burden of disease methodology can be usefully applied to estimate the health burden of adverse events associated with vaccination, but the interpretation of the findings must consider the quality and accuracy of the data sources involved in the computation of the DALY.

Project 9. CodeMapper: semi-automatic coding of case definitions

Assessment of drug and vaccine effects by combining information from different healthcare databases in the European Union requires extensive efforts in the harmonization of codes as different vocabularies are being used across countries. We present a web application called CodeMapper, which assists in the mapping of case definitions to codes from different vocabularies, while keeping a transparent record of the complete mapping process. CodeMapper builds upon coding vocabularies contained in the Metathesaurus of the Unified Medical Language System. The mapping approach consists of three phases. First, medical concepts are automatically identified in a free-text case definition. Second, the user revises the set of medical concepts by adding or removing concepts, or

expanding them to related concepts that are more general or more specific. Finally, the selected concepts are projected to codes from the targeted coding vocabularies. The application was evaluated by comparing codes that were automatically generated from case definitions by applying CodeMapper's concept identification and successive concept expansion, with reference codes that were manually created in a previous epidemiological study. The web application is available under https://euadr.erasmusmc.nl/CodeMapper and was used for the creation of codes in the proof of concept studies. Several lessons were learned related to the granularity of the different coding schemes, and the level of expertise that is required to effectively map the codes.

1. SELECTION OF PROJECTS

In this chapter, we briefly describe how the different research projects were selected.

1.1. Research proposals

During the first year of the ADVANCE project the existing methodologies for burden of disease, vaccination coverage, vaccine safety, effectiveness, impact and benefit-risk were assessed and research gaps were identified (see deliverables D4.1, D4.2 and D4.3). After this initial phase, the ADVANCE consortium members were asked to prepare concrete research proposals. The proposals are available at the ADVANCE WP4 Sharepoint. Then, during a workshop in Brussels (March 6, 2015), the research proposals were discussed, prioritized and revised. The list of research proposals is given in Table 1.1. In addition, as the work package 5 Proof-of-Concept study on the benefit-risk of Pertussis vaccination (WP5 POC1) raised concerns regarding the validity of the Pertussis outcomes, methodological work on validation and the potential impact of disease- and exposure-misclassification was initiated (see research projects 3, 5 and 6 in Section 1.2).

Table 1.1: ADVANCE WP4 research proposals, (March 6, 2016)

Title	Name	Decision
Burden of disease associated with MMR vaccination	Scott McDonald	Develop methodology generically, not focussing on a specific vaccine (research project 8*)
Estimating the years of life lost due to vaccine preventable diseases	Maarten van Wijhe, Scott McDonald, Jacco Wallinga	not selected(although good to have, it was preferred to further investigate burden of disease methodology to quantify the burden of adverse events)
Explore the possibility to comparably estimate vaccination coverage across databases and countries: childhood and HPV vaccination	Hanne-Dorthe Emborg, Tyra Grove Krause, Palle Valentiner-Branth	Develop methodology generically (research project 1*)
Spatial methods to support benefit-risk assessment of vaccines	Daniel Weibel, Peter Rijnbeek, Maria de Ridder, Caitlin Dodd, Miriam Sturkenboom	not selected (was not considered a basic need)

Costs of repeated looks at data	Caitlin Dodd, Daniel Weibel, Peter Rijnbeek, Maria de Ridder, Miriam Sturkenboom	not selected (a NICE-TO-HAVE selected for further research if resources would allow)
Instrumental variable analysis for missing data	Caitlin Dodd, Daniel Weibel, Peter Rijnbeek, Maria de Ridder, Miriam Sturkenboom	not selected (this is a general pharmaco-epidemiological topic, and not specific to benefit-risk of vaccines
Calibration of empirical p-values to correct for bias	Caitlin Dodd, Daniel Weibel, Peter Rijnbeek, Maria de Ridder, Miriam Sturkenboom	Integrate in proposal to study data database heterogeneity (research project 4*)
Frequency of benefit-risk monitoring	Miriam Sturkenboom, Kartini Gadroen, Daniel Weibel, Benus Becker, Caitlin Dodd	Integrate on proposal for benefit-risk monitoring (research project 6*)
Study heterogeneity of vaccine risk between databases	Paddy Farrington	Selected (research project 4*)
Comparative methods evaluation in a benefit-risk framework	Paddy Farrington	Not selected (a NICE-TO-HAVE selected for further research if resources would allow)
Sequential evaluation of risk for benefit-risk assessment	Paddy Farrington	Integrate on proposal for benefit-risk monitoring (research project 6*)
Quantitative benefit-risk assessment of HPV vaccination	Kaatje Bollaerts, Thomas Verstraeten, Marc Baay, Lisen Dählstrom, Matti Lehtinen	Not selected (not methods development, rather an application)
Multi-country, multi-stakeholder preferences for the benefit and risks of HPV vaccination	Kaatje Bollaerts, Thomas Verstraeten	Not selected (not methods development, rather an application)
Use of Bayesian (max) SPRT-like for B:R surveillance	Ed Ledent, Vincent Bauchau	Integrate on proposal for benefit-risk monitoring (research project 6*)

*research projects as in Section 1.2

1.2. Selected research projects

The most important criterion for selecting a research project was being in line with the ADVANCE vision and mission. The ADVANCE vision is to deliver the "best evidence at the right time to support decision-making on vaccination in Europe". The mission is to establish a prototype of a sustainable and compelling system that rapidly provides best available scientific evidence on vaccination benefits and risks post-marketing for well-informed decisions.

In addition, the combined sets of selected research projects should cover all aspects of vaccine post-marketing surveillance, including burden of disease, coverage, safety, benefit and benefit-risk. An overview of the selected research proposals is given in Table 1.2.

Table 1.2: ADVANCE WP4 research projects

Project	Title	ADVANCE vision/mission
Project 1	Population differences: a source of heterogeneity	Multi-database studies, improved use of EHR data, timely evidence
Project 2	Methods to estimate vaccination coverage from dynamic populations	Improved use of EHR data, timely evidence
Project 3	Impact of disease- and exposure misclassification on estimation of vaccine effectiveness	Improved use of EHR data, timely evidence
Project 4	Validation of case-finding algorithms in healthcare research: analytical interrelations between validity indices	Improved use of EHR data, timely evidence
Project 5	Heterogeneity in disease misclassification: the component analysis	Improved use of EHR data, imely evidence
Project 6	Latent Class Models to estimate validity when there is no reference standard	Improved use of EHR data, timely evidence
Project 7	Benefit-risk monitoring of vaccines: a dashboard	Benefit-risk integration, timely evidence
Project 8	Burden of Disease of adverse events following immunization	Benefit-risk integration
Project 9	CodeMapper, semi-automatic coding of case definitions	Facilitating multi-database studies, timely evidence, enlarging scale

2. Project 1: Population differences as source of heterogeneity

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Maria de Ridder (EMC), Jose van Boxmeer (SEQ), Caitlin Dodd, Swabra Nakato (EMC)	13.04.2016	0.0	Adaptation of heterogeneity protocol to specific needs of sub-project
Maria de Ridder	15.09.2016	0.1	Outline shared with Jose and Caitlin
Klara Berensci (AUH), Talita Duartes (SIDIAP), Elisa Martin, Consuelo Huerta (BIFAP), Silvia Lucchi (ASLCR), Gino Picelli, Lara Tramontan, Giorgia Daniele (PEDIANET), Ana Correa (RCGP), Daniel Weibel (EMC/THIN), Hanne Dorthe Emborg (SSI), Vincent Bauchau (GSK), Miriam Sturkenboom (EMC)			WP 5: conduct and coordination of population fingerprint and creation of data going into the graphics
Maria de Ridder	11.11.2016	0.2	Draft graphics shared with Caitlin
Maria de Ridder	24.01.2017	1.0	Draft including graphics and text
Maria de Ridder	03.02.2017	1.1	Add discussion
Caitlin Dodd (coordinator)	06.02.2017	1.2	Minor edits and formatting
Kaat Bollaerts (P95), Miriam Sturkenboom	07.02.2017		Comments and conclusions/impact
Maria de Ridder	10.02.2017	1.3	New figures (fewer countries), several adjustments in text.
Miriam Sturkenboom	12.02.2017		Comments and adjustments
Maria de Ridder	13.02.2017		Adjustments
Linda Levesque Mendel Haag	24.02.2017		Revision
Maria de Ridder	28.20.2017		Addressed comments

2.1. Introduction

Differences in results between observational studies that aim to establish the association between an exposure and outcome by looking at differences in disease occurrence may have various origins:

1) **Chance**: The study population is only a sample of the target population.
2) **Information Bias** (exposure and/or outcome assessment): Measurement error in exposure measurement and/or disease occurrence due to imperfect measurement tools and incompleteness of datasources.
3) **Selection Bias**: Distortions that result from procedures used to select subjects and from factors that influence participation in the study.
4) **Confounding**: A third factor that is associated both with disease occurrence as well as exposure distorts the relationship between exposure and outcome.
5) **Effect modification**: Effect modification occurs when the magnitude of the effect of the primary exposure on an outcome (i.e., the association) differs truly depending on the level of a third variable. In this situation, computing an overall estimate of association is misleading.

The literature is full of controversial findings in pharmacoepidemiology, often arising from differences in design, measurement, definitions and analysis. In the Observational Medical Outcomes Partnership (OMOP) Madigan et al. (1) took a systematic approach to isolate the effect of data source on the effect estimates by holding all other aspects of the study design constant. They investigated 53 positive and negative drug-outcome pairs representing a range of typical epidemiologic scenarios and, for each pair, applied 2 different study designs with the exact same implementation in 10 different databases. When holding the study design constant, it was shown that estimated effects ranged from a statistically significant decreased risk to a statistically significant increased risk in 11 of 53 (21%) drug-outcome pairs that used a cohort design and 19 of 53 (36%) drug-outcome pairs that used a self-controlled case series design. This exceeded the proportion of pairs that were consistent across databases in both directions. The authors concluded that observational studies that use databases can be sensitive to the choice of the database, and as such, it should be investigated how the choice of data source may affect the results [1].

In distributed networks, such as ADVANCE, we aim to pool data from different countries to collect evidence on vaccine effects more rapidly and to benefit from the European scale in size and diversity in exposure to vaccination. Although most steps in this evidence generation are standardized: i.e. protocol, definitions, data transformation and data-analysis, we cannot eliminate differences due to healthcare systems and structures and the way data is captured.

In this project we aimed to investigate the differences in populations in countries participating in ADVANCE, the populations registered in the databases in those countries and the differences between follow-up of the population in the various databases. This should provide better insight into ways to extrapolate and into potential sources of selection bias.

2.2. Methods

We investigated heterogeneity in populations at three different levels: 1) between different countries; 2) between the database source population and the national population; 3) between databases (see Figure 2.1).

Data on the age and gender distribution in the countries who had submitted data to the ADVANCE population characterization were obtained from the United Nations (2). Data on the age and gender distribution in each database were collected from the output of the population characterization conducted as part of the fingerprinting of databases in ADVANCE WP5 as described in ADVANCE Deliverable 5.2. Further characterizations of the populations, regarding age at entry in a database and follow-up duration, were obtained during the characterization of vaccines dispensed.

Per country (United Nations data), the percentage of individuals in each age group was calculated as well as cumulative percentage in increasing age groups. To visualize differences between countries, the same characteristics were obtained for the data of all countries pooled together. These pooled data do not represent a well-defined reference population but were used as anchor to investigate the heterogeneity in the data sources. The (cumulative) percentages within each country were compared with the percentages in these pooled data. Per database, in each cohort defined by year of birth, the percentage of persons registered within 6 months from birth was calculated.

For comparison of databases with their corresponding country, only databases including the full range of age groups were used, i.e. the Aarhus University database (AUH, Denmark), the Danish Civil and Health Registration System (SSI, Denmark), ASL Cremona (ASLCR, Italy), Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria(BIFAP, Spain), The Health Improvement Network database (THIN, United Kingdom) and the Royal College of General Practitioners database (RCGP, UK). Pedianet was not compared since this database only contains children. Again, percentages and cumulative percentages for age groups were calculated, for both the database population and the country population, and compared.

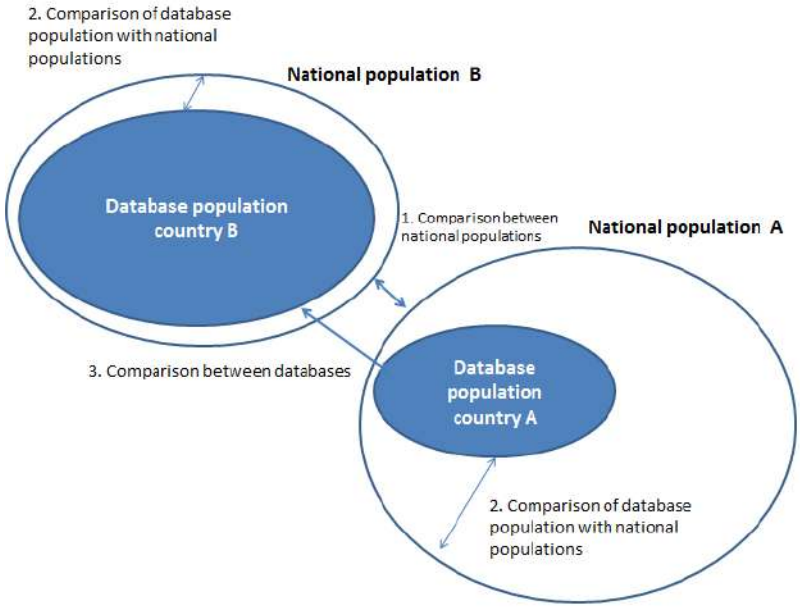


Figure 2.1: Graphical display of the population comparisons made in this deliverable.

2.3. Results

2.3.1. National population characteristics

Table 2.1 describes population based health-statistics for the countries that participated in the population characterization in ADVANCE. Differences in health related parameters were present, particularly with regards to health expenditures and physician density. Of note, Spain and Italy had more physicians per 10,000 inhabitants, lower health expenditure per capita and the lowest rate of adult mortality. Although these parameters do not have a direct effect on the work ADVANCE, it is important to realize that healthcare systems differ between the countries as does the 'health' of the various populations.

Figure 2.1a shows the percentages of the population in each age group by each country. Figure 2.1b shows the differences in cumulative percentages compared to the overall cumulative percentage (adding up all the population). In this plot, the line of a country with a similar age distribution as the overall population will fluctuate around the line with difference zero, as differences in one age group will be compensated for in subsequent age groups. Underrepresentation in consecutive young age groups will show a decreasing line, which will bend up later, because by definition at the end the difference will be zero. For overrepresentation of young ages, the opposite will occur. For example, the plot shows that in the UK the percentage in age group 0-4 years was 1 percentage point higher compared with the countries pooled together. The increasing line for UK up to age group 30-35 years reflects the fact that the UK had higher proportions in each of the younger age groups compared with the countries pooled. Also Denmark had a relatively young population. Italy had the oldest population.

Table 2.1: Health-related statistics by country (WHO Global Health Observatory data) (3) Year is 2015, unless stated otherwise.

Country	Physicians density (per 10,000 inhabitants)[1]	Neonatal mortality rate (per 1,000 live births)[2]	Under-five mortality rate (per 1,000 inhabitants)	Adult (≥ 20 years) mortality rate (per 1,000 inhabitants)	Disability-adjusted life years (DALY's)[3]	Per capita total expenditure on health in 2014 (US $)[4]
United Kingdom	27.41	2.4	1.75	68.82	17,856,300	3,935
Denmark	32.24	2.5	1.50	82.44	1,665,200	6,463
Italy	34.86	2.1	1.50	63.47	16,579,400	3,258
Spain	39.57	2.8	1.75	71.74	11,841,400	2,658

[1] DK and IT: 2009, UK and ES: 2010
[2] Deaths among live births during the first 28 days of life
[3] Source: WHO Global Health Estimates (4)
[4] Source: The World Bank, Health expenditure per capita (5)

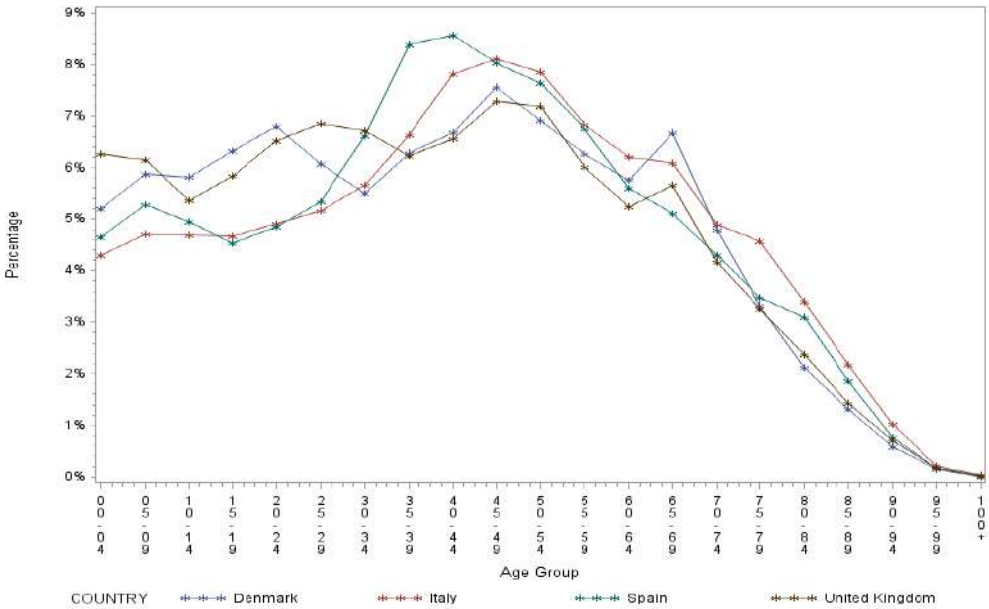


Figure 2.1a: Age distribution by country

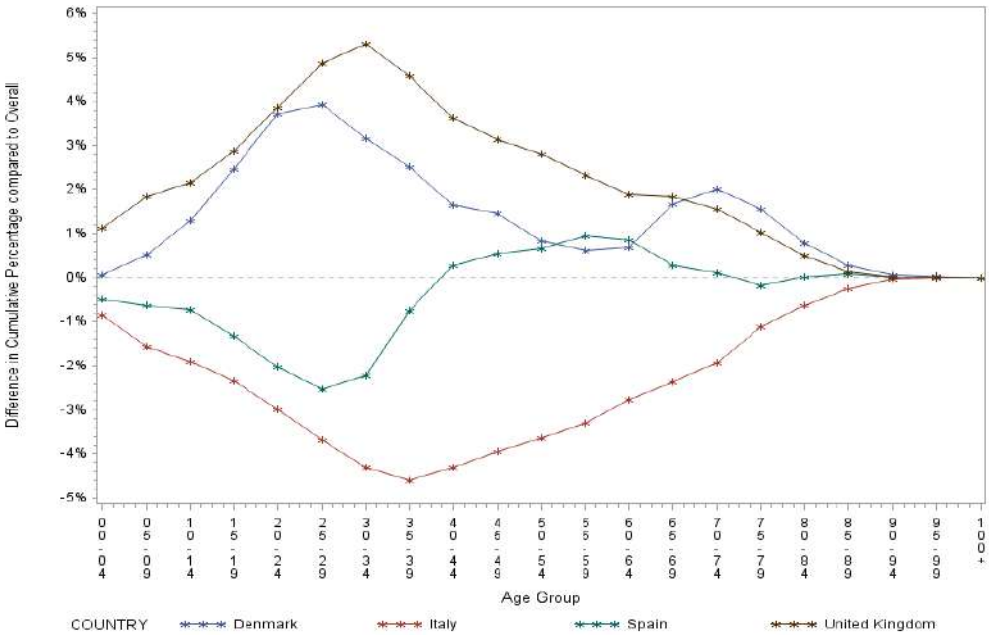


Figure 2.1b: Difference in cumulative percentage compared to overall

Figure 2.2 shows the percentages females by age group in each country. As expected, percentages were slightly below 50% in children and adolescents and increased in adults. The UK had lower percentages of females in older age groups than the other countries. However, all differences between countries were less than 4%.

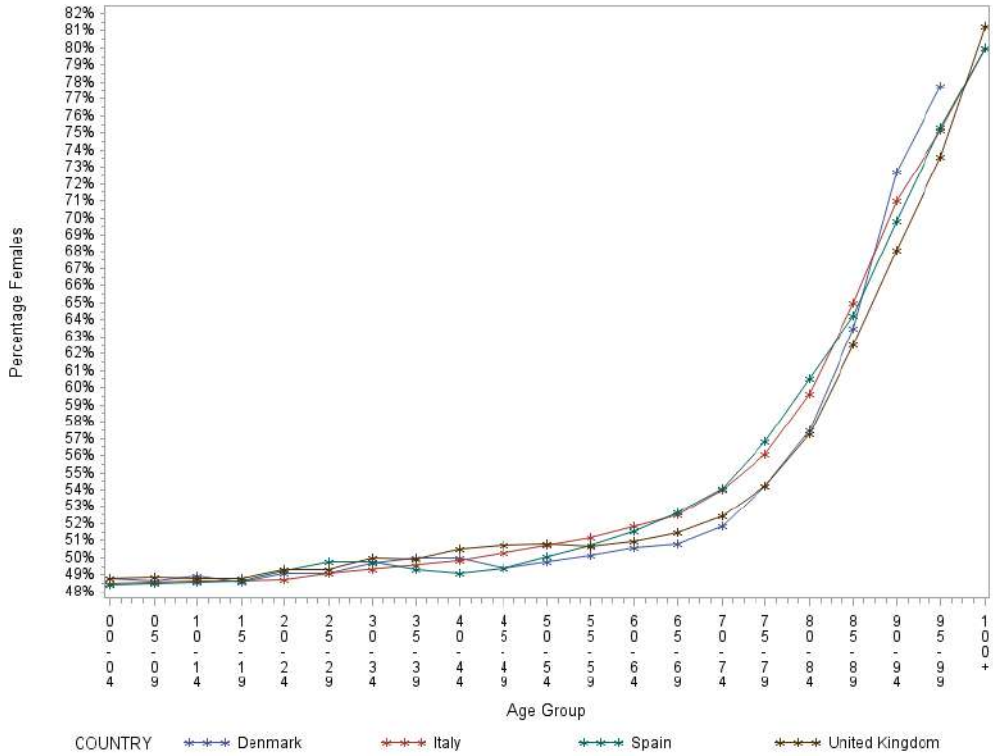


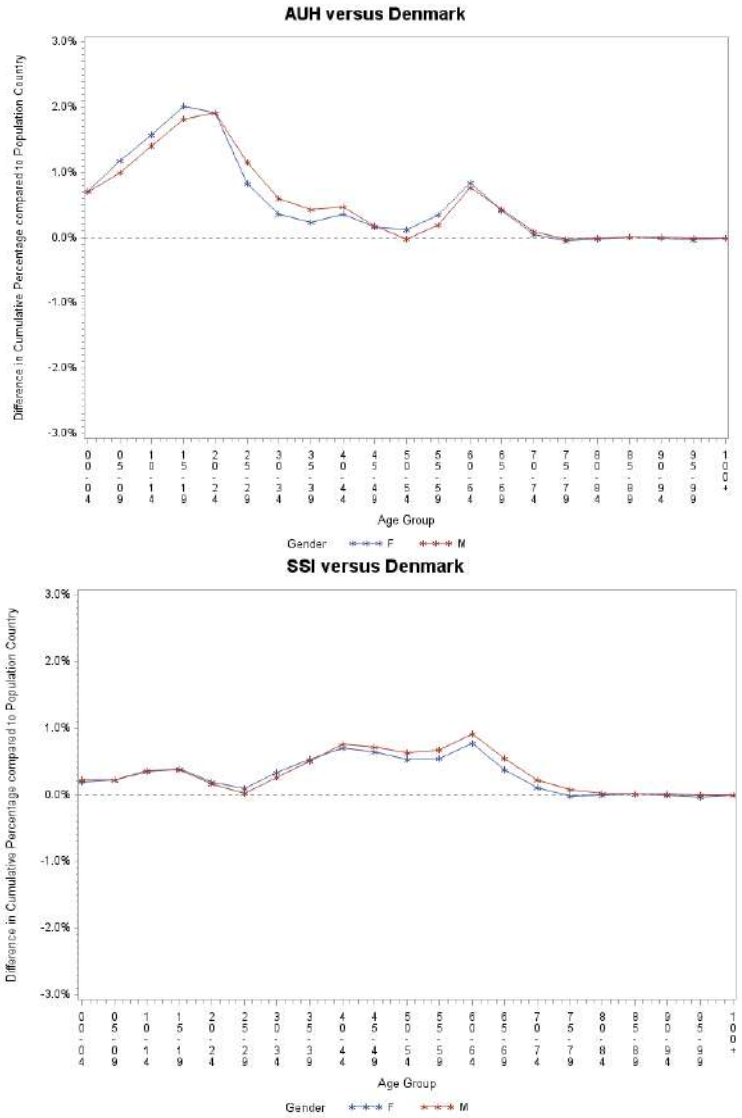
Figure 2.2: Percentage females by age group by country

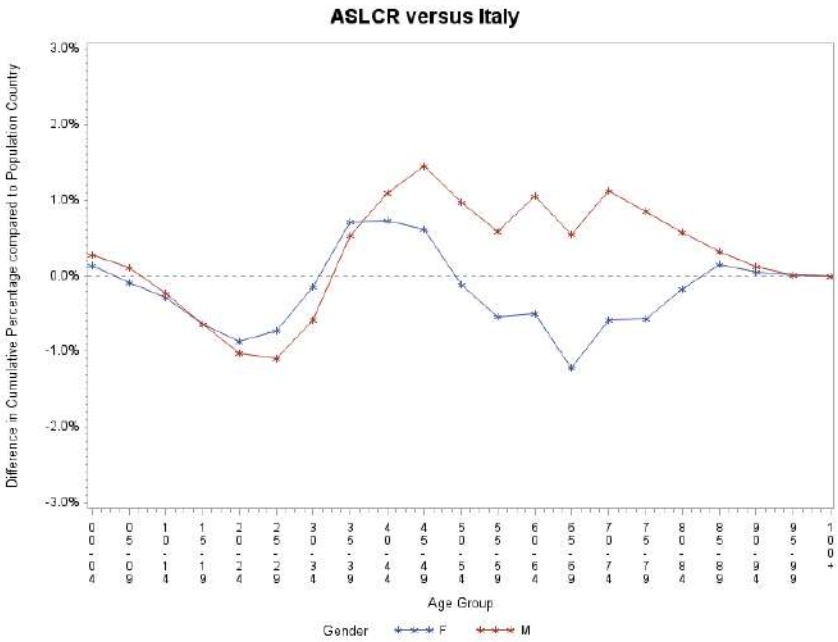
2.3.2. Databases population characteristics

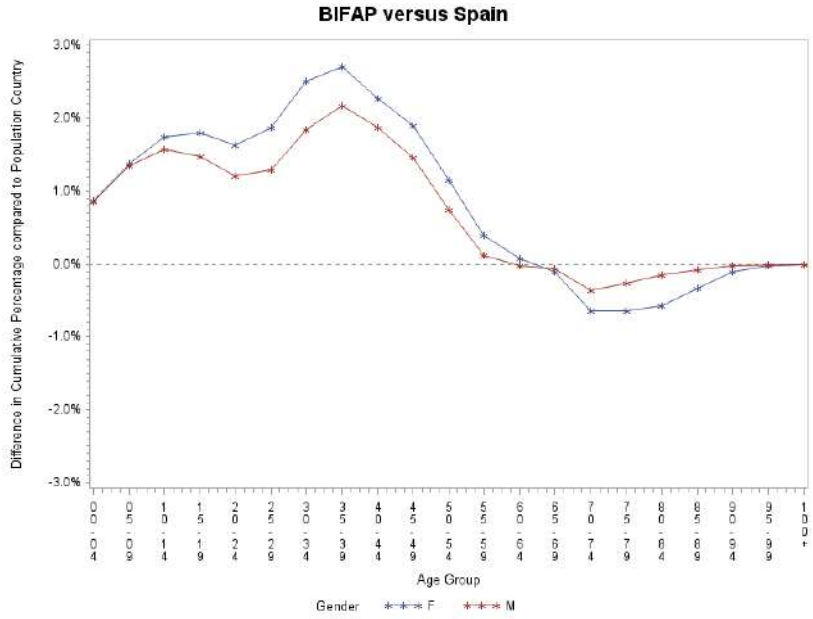
We were interested to see whether the ADVANCE databases that participated in the POC feasibility assessment have gender and age distributions that are representative of their country. Since studies carried out using these database populations, lack of representativeness would have implications with regards to the generalizability of the results to the population of the country.

Figure 2.3 shows the differences between database and country for the cumulative percentages in age groups, for males and females separately. In general, differences between a database and its country were small, less than 3%. In Denmark,in AUH the younger ages were slightly overrepresented, for both males and females (due to large influx of students in that area), but

in the SSI the age distribution was closer to that of the country, which is to be expected as the SSI covers the entire population. In ASLCR in Italy young adults were underrepresented, whereas males from age 35-39 years upwards were overrepresented in comparison with the entire Italian population. In BIFAP in Spain the pattern of a young database population was more pronounced. Both of the UK databases had an underrepresentation of younger age groups. In addition, since these are both general practitioners databases, the elderly may not be fully represented as some individuals in the group move to long term care facilities later in life.







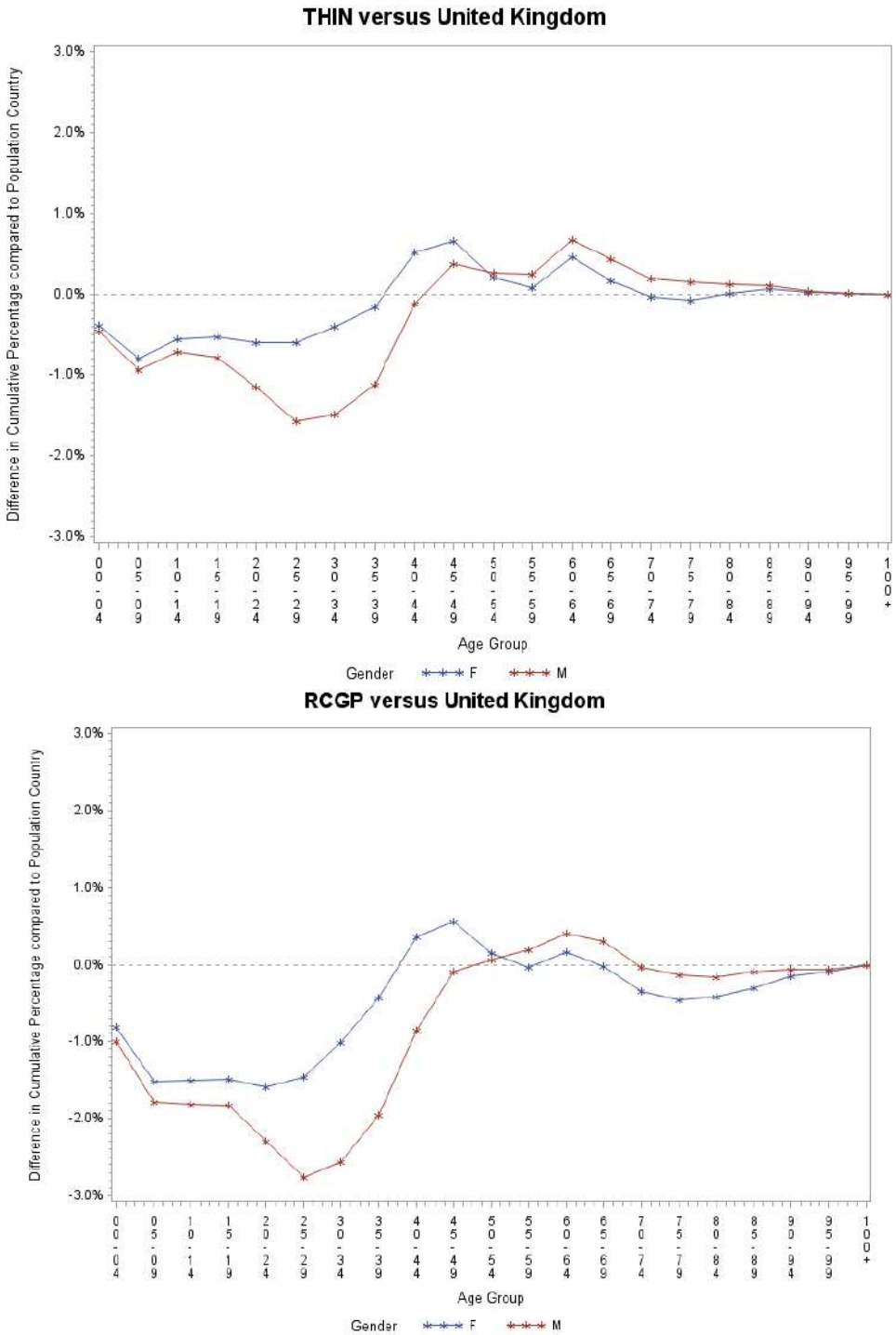


Figure 2.3: Differences in cumulative percentages in age groups between database and country population, males and females separately

2.3.3. Databases comparison: entry in database and duration of follow-up

Figure 2.4 shows the percentages of persons registered within 6 months after birth among all persons born that year and recorded in the database. We observed substantial differences depending on the type of database. In the general practitioners databases, which have dynamic populations, the percentages of persons registered within 6 months after birth were low (e.g. Sistema d'Informació per al desenvolupament de la Investigació en Atenció Primària (SIDIAP), BIFAP, RCGP, THIN). Percentages in SSI were high since this database captures the entire population and is very stable (see deliverable D5.2).

It should be noted that Figure 2.4 shows the situation in 2016 and percentages of persons registered within 6 months after birth will change in future. Each new registration of a person born in the years shown (up to 2015) will by definition lower the percentage, as all these will be 'late' registrations. Most likely, most added registrations will be in the relatively recent years (Figure 2.4).

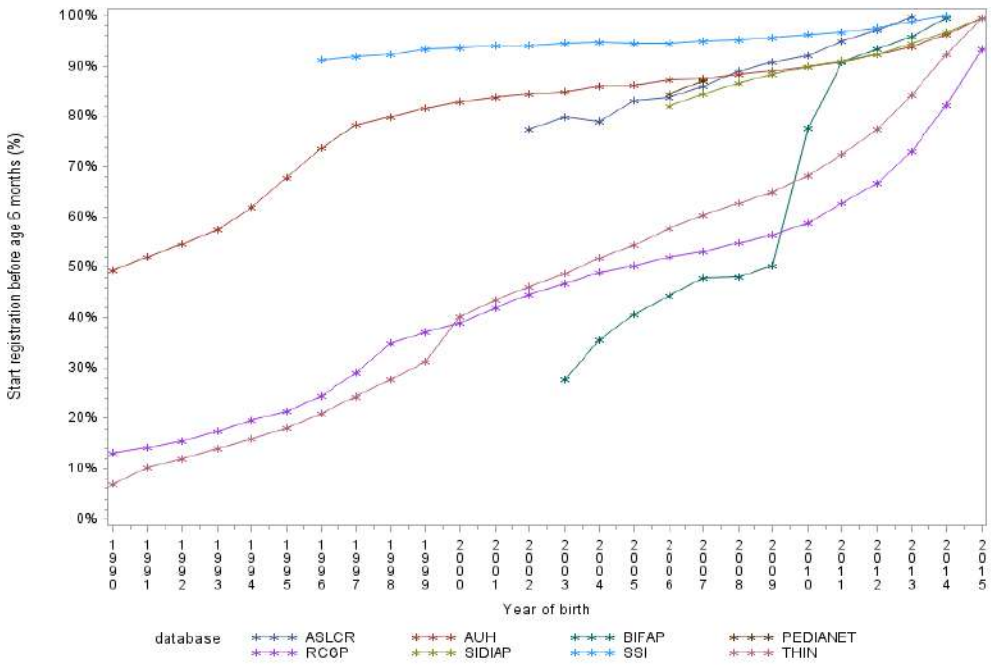


Figure 2.4: Percentage registered before age of 6 months

For persons registered within 6 months after birth, the length of follow-up within the database was determined. The difference between databases could be seen most clearly when examining one birth year at a time. Figure 2.5 shows the decline in percentages of follow-up of persons born in 2006.

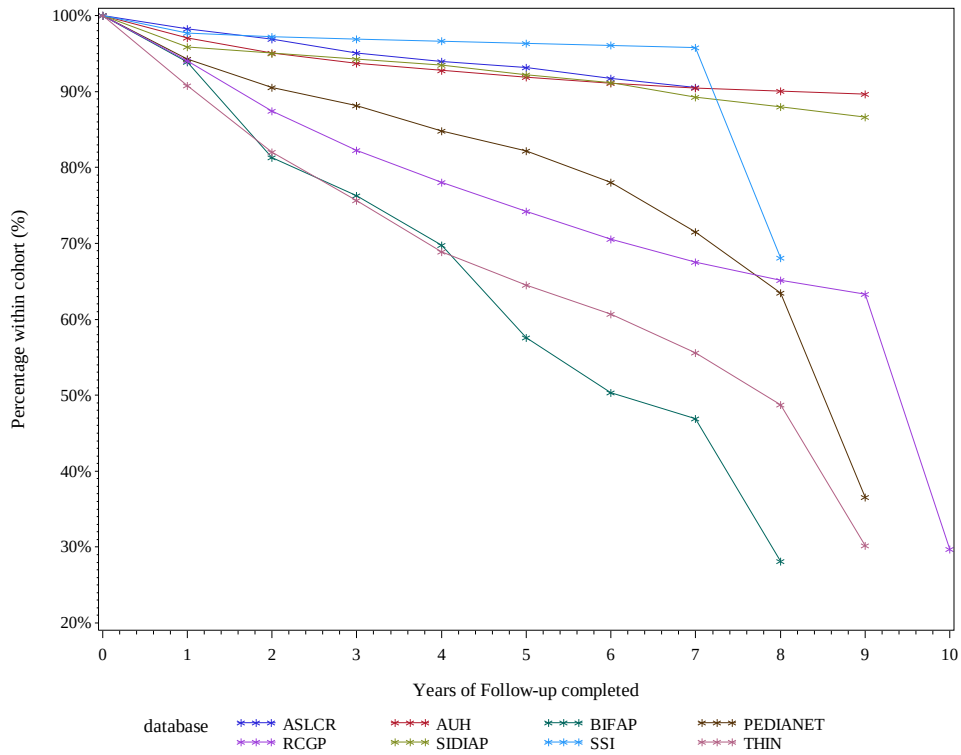


Figure 2.5: Percentages of persons with at least a certain follow-up per database(selection of persons born in 2006).

In ASLCR, AUH, SIDIAP and SSI the percentages of individuals with long term follow-up were considerable, more than 90% of the persons having a follow-up of at least 5 years. In the other databases, the loss to follow-up was much higher with the percentages with at least 5 years of follow-up being 82% in Pedianet, 74% in RCGP, 64% in THIN and 58% in BIFAP.

2.4. Discussion

Although from an overall perspective the European countries involved in the ADVANCE project may be quite similar, health-related statistics do show some differences. The differences in physician density and health expenditure mainly reflect differences in health-care systems, health-care seeking behaviour and resource use. For definition of diagnoses based on treatment, these differences could have impact. With regards to age distribution, countries with relatively young or old populations were observed, however, these differences are small and classical methods (e.g. standardization) to account for this exist.

Differences exist between the database populations and the national populations, depending on the size and the origin of the database. Although this impacts the generalizability of results

from database research, it does not necessarily impact the internal validity of studies. Again, there are classical methods to account for these differences (e.g. standardization). Of note, differences between the database populations and the national populations other than age and gender might be of more importance, but were outside the scope of this project.

Our analysis of the database populations showed large differences in the timing of registration and the duration of follow-up. As in dynamic databases, like general practitioners databases, age of registration and follow-up duration might be related to persons characteristics (like moving pattern, health care seeking behaviour), this could potentially cause bias in studies that use fixed cohorts (from birth). We therefore recommend using dynamic cohort approaches with adjustments for the incomplete follow-up that have been developed in WP4.

2.5. References

1. Madigan D, et al. Evaluating the impact of database heterogeneity on observational study results. Am J Epidemiol 2013;**178**(4):645-51
2. http://esa.un.org/unpd/wpp/Excel-Data/population.htm
3. http://www.who.int/gho/en/
4. http://www.who.int/healthinfo/global_burden_disease/en/
5. http://data.worldbank.org/indicator/SH.XPD.PCAP

3. Project 2: Methods to estimate vaccination coverage from dynamic populations

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts, Miriam Sturkenboom, Vincent Bauchau	11-12 Dec 2015	-	Initial discussion
Kaat Bollaerts	09 Nov2016	-	CDF method, outline shared with Toon Braeye
Toon Braeye	11 Nov 2016	-	IPW method
Kaat Bollaerts, Toon Braey, Tom Cattaert	6 Dec 2016	-	Discussions on R-code, methodology
Kaat Bollaerts	15Feb 2016	-	First draft
Kaat Bollaerts	27-Aug 2016		Presented at ISPE
Jose van Boxmeer	30 Aug 2016		Benchmarking data
Kaat Bollaerts	5 Feb 2017	-	Second draft
Tom Cattaert, Toon Braeye	7 Feb 2017	-	Review
Kaat Bollaerts	11 Feb 2017		Comments incorporated
Linda Levesque, Sonja Banja, Mendel Haag, Miriam Sturkenboom	25 Feb 2017		SC review
Kaat Bollaerts	28 Feb 2017		Addressed comments

3.1. Introduction

The introduction of new vaccines and the evaluation of vaccination programs requires tools to closely monitor the vaccination uptake. Vaccination coverage and compliance with the recommended vaccination schedules are widely used indicators of vaccination program performance. These performance indicators are typically measured by registries, routine administrative reports or household surveys. The methods used vary widely across European countries and the need for harmonization is generally acknowledged [1].

Electronic healthcare records (EHRs) are an alternative source of information to monitor vaccination uptake. They allow timely monitoring at a relatively low cost and often cover large geographical areas or sizeable populations. However, the populations captured in EHRs are generally dynamic, with members moving in and out of the population over time (i.e. transient membership) for example due to relocation or switch between general practices. This often results in incomplete follow-up, hampering the accurate estimation of vaccination coverage from EHRs. Incomplete follow-up would lead to an underestimation of the vaccination coverage as vaccines administered outside the follow-up period will not always be recorded. We explored two methods to estimate vaccination coverage for dynamic populations and assessed their performance through simulation.

3.2. Methods

3.2.1. Cumulative distribution function method

Vaccinations are often recommended to be given at specific ages (e.g. childhood vaccines) or at a given time of the year (e.g. seasonal influenza vaccination). This information can be used to correct for unrecorded vaccinations. For example, consider a childhood vaccine for which vaccination is recommended at 6 months, the distribution (or density function) of the age at vaccination is f_A.

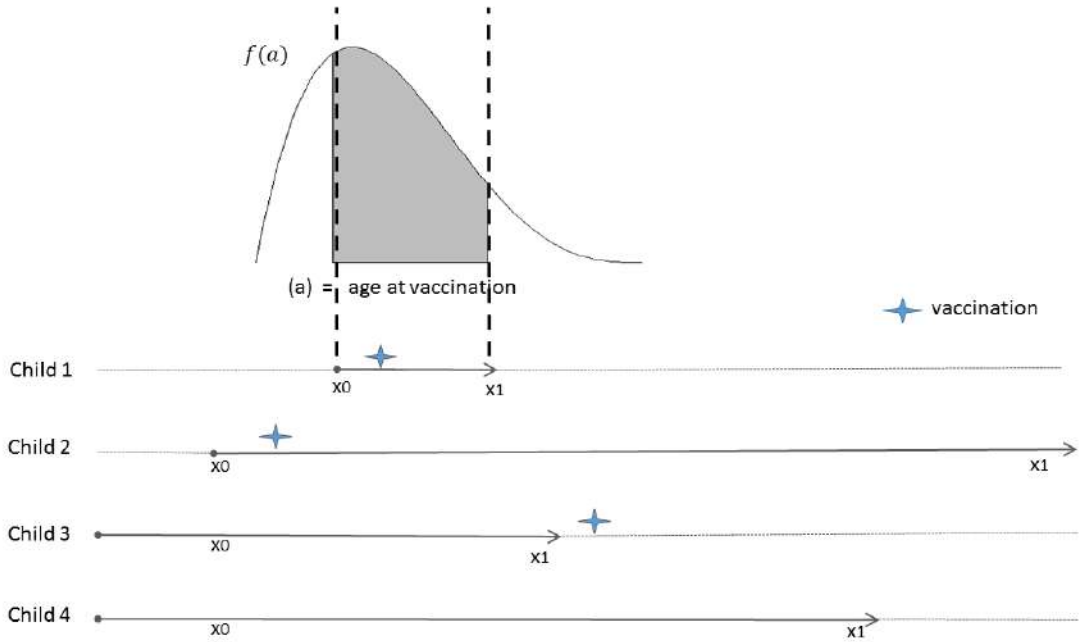


Figure 3.1: Hypothetical distribution for age at vaccination and follow-up time for four children (x0 indicates the start of follow-up and x1 the end of follow-up).

Figure 3.1 represents such a hypothetical distribution as well as the follow-up for four children. As can be seen, incomplete follow-up at neonatal age (child 2) or at toddler age (child 4) will not affect the coverage estimates because children are simply not vaccinated with the vaccine of interest at that age. On the other hand, incomplete follow-up around the age of 6 months will affect the estimation of the vaccination coverage (children 1 and 3). To quantify the amount of relevant follow-up time for each child (d_i), we use the cumulative distribution function Φ_A, which evaluates the probability of vaccinating a child before age a conditional on the child being vaccinated. Then the amount of relevant follow-up time for child i can be quantified as

$$d_i = \Phi_A(t_{1i}) - \Phi_A(t_{0i}), \quad (1.1)$$

with t_{0i} being the age of child i at the start of its follow-up and with t_{1i} being the age of the child at the end of its follow-up. The d_i's can be interpreted as weighted follow-up time. Observe that the d_i's are bounded between 0 and 1 because d_i is calculated as a difference in probabilities with $\Phi_A(t_{0i}) < \Phi_A(t_{1i})$. To illustrate, we represent the relevant follow-up time for child 1 by the shaded area of f_A, which corresponds to $d_1 = 0.6$. For the other children, the relevant follow-up time is $d_2 = 1$, $d_3 = 0.9$ and $d_4 = 1$ (Figure 3.1). The distribution f_A is typically not known, but can be estimated using the data from children with complete follow-up. To this end, a wide range of density estimation techniques can be used, either parametrically or non-parametrically [2]. By doing so, it is assumed that children with incomplete follow-up time have the same distribution of age at vaccination as children with complete follow-up time. Then, assuming that

incompleteness is independent of vaccination status and age at vaccination, the coverage π within a population of size N can be estimated as

$$\hat{\pi}_{CDF} = \frac{n}{\sum_{i=1}^{N} \hat{d}_i}, \quad (1.2)$$

with n being the number of recorded vaccinations and with $\hat{d}_i$ being the weighted follow-up time for child i, derived from the estimated density function $\hat{f}_A$. One way to obtain confidence intervals (CIs) is to first apply the normalizing logit transformation or $logit(\hat{\pi}_{CDF}) = g(\hat{\pi}_{CDF}) = \log(\hat{\pi}_{CDF}/1 - \hat{\pi}_{CDF})$ and then obtain the 95% Wald CIs as:

$$g(\hat{\pi}_{CDF}) \pm 1.96 s.e.(g(\hat{\pi}_{CDF})), \text{ with } s.e.(g(\hat{\pi}_{CDF})) = \sqrt{\frac{\widehat{Var}(\hat{\pi}_{CDF})}{\hat{\pi}_{CDF}(1-\hat{\pi}_{CDF})}}., \quad (1.3)$$

with $\widehat{Var}(\hat{\pi}_{CDF}) = \frac{n}{\sum_{i=1}^{N} \hat{d}_i}$ and finally transform the CI limits back to the prevalence scale using $g^{-1}(x) = (1/1 + \exp(-x))$. However, this approach overestimates precision as it assumes fixed values $\hat{d}_i$. Bootstrapping while each time refitting $\hat{f}_A$ will yield more accurate CIs [3] but might be too computationally intensive for use with EHR data. The optimal way of obtaining CIs, trading off accuracy of the CIs with computational time, is worthy of further investigation.

3.2.2. Inverse probability weighting method

Alternatively, we can correct for unrecorded vaccinations by comparing, within strata of interest, the observed person time within a population with the hypothetical person time if all subjects within that population would have been completely followed-up. This approach is graphically presented by means of a Lexis diagram (Figure 3.2). In this illustration, there are two stratification variables of interest (age and calendar time); the solid lines represent the actual follow-up and the dashed lines the theoretical follow-up if all subjects were completely followed-up.

Let pt_{jk} be the observed person time for subjects belonging to age group *j* (in weeks, months, years…) at time period *k* (in weeks, months, years…) and let pt^*_{jk} be the corresponding hypothetical person time if all subjects were completely followed-up. Then the ratiopt_{jk}/pt^*_{jk} can be interpreted as the probability of follow-up of a subject of age group *j* in time period *k*. In analogy with the use of inverse probability weighting (IPW) to correct for selection bias [4, 5] we can use $\left(pt_{jk}/pt^*_{jk}\right)^{-1}$to correct coverage estimates for incomplete follow-up. Assuming that within a given stratum *jk*, incompleteness does not depend on vaccination status, the coverage can be estimated as;

$$\hat{\pi}_{IPW} = N^{-1} \sum_{j=1}^{J} \sum_{k=1}^{J} \left(pt_{jk}/pt^*_{jk}\right)^{-1} n_{jk}, \quad (1.4)$$

with n_{jk} being the number of recorded vaccinations for subjects of age group *j* at time period *k* and N being the total population size.

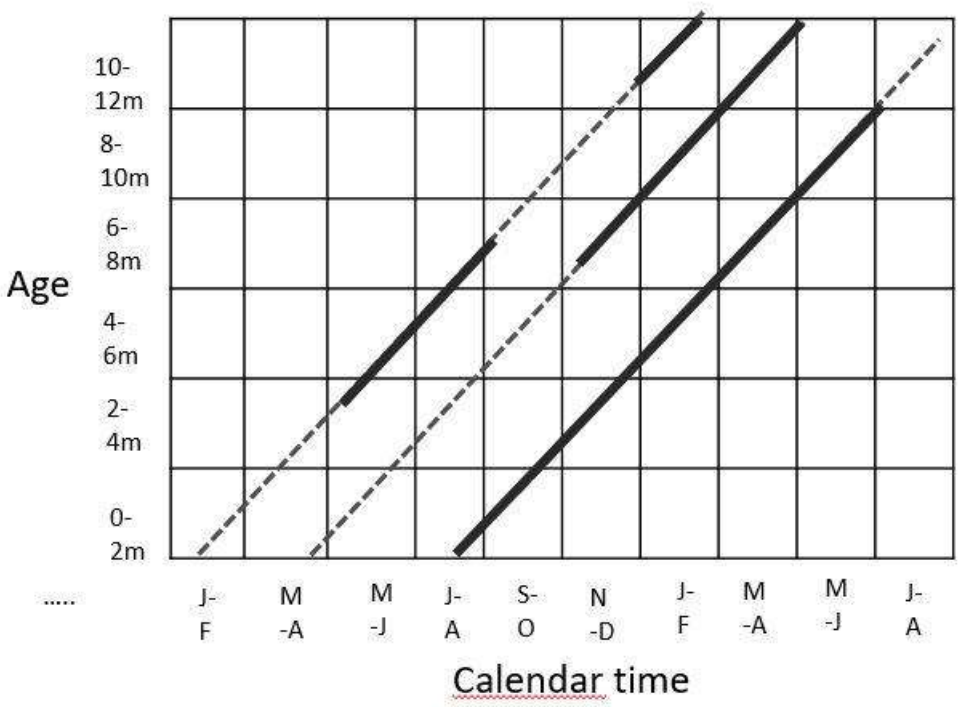


Figure 3.2: Lexis diagram, representing the actual (solid lines) and theoretical (dashed lines) of follow-up for three subjects.

The ordinary 95% CI for inverse probability weighted estimates may not provide the correct coverage. Instead, robust "sandwich" variance estimators or non-parametric bootstrapping would provide valid confidence intervals. [6]

We introduced the methodology using age and calendar time as the two stratification variables of interest. Evidently, the number of stratification variables can be less or more and the stratification variables should be chosen because they are related to both coverage and completeness of follow-up. Logistic regression models can be used to identify the relevant strata.

3.3. Simulation study

3.3.1. Simulation model

We simulated a single year birth cohort of N = 10,000 subjects with random dates of birth. Of these subjects, a percentage p = {30% , 50%, 85%} were randomly vaccinated. Their age of vaccination (in days) was randomly sampled from a 3-parameter Weibull distribution (location=120, scale=100, shape=1.1), which is a continuous right-tailed probability distribution often used to describe waiting times. This distribution had a median of 192days or

27 weeks, with the 95% inter-percentile ranges of [124- 456] days or [18-65] weeks. Subjects were followed starting from their date of birth (T_0) until one year of age (T_1). Then, we randomly introduced incompleteness in follow-up to varying degrees from 10% to 90% (scenario 1: random incompleteness). For a random selection of *k*% of the subjects, the start of the follow-up T_0 was delayed to any random time between the date of birth and three years of age; for another random selection of *k*% of the subjects, the end of follow-up $T_1$was any random time between the (delayed) start of follow up and one year of age. Observe that we expectedk^2% of the subjects to have a delayed start as well as an early end of follow-up. Secondly, we introduced time-dependent coverage and incompleteness. (scenario 2: non-random incompleteness). Specifically, 40% of children born in the first half of the year were randomly vaccinated whereas only 20% of the children born in the second half of the year were randomly vaccinated. In addition, children born in the second half of the year had a 20% higher probability of a delayed start/early end of follow-up compared with children born in the first half of the year.

For each scenario, i.e. for every vaccination coverage and degree in incompleteness *k,* we simulated 1000 datasets. Then, for every dataset, we estimated the vaccination coverage using (i) the cumulative distribution method with Kernel density estimation to estimate the distribution of the age at vaccination and (ii) the inverse probability weighting method with age and time in weeks. We opted for Kernel density estimation as this is a non-parametric method for which no distributional assumptions need to be made. For comparison, we also estimated the coverage (iii) by ignoring incomplete follow-up or

$$\hat{\pi}_{IGN} = \frac{n}{N},$$

with n being the number of observed vaccinations and N being the total population size. In addition, we estimated coverage using (iv) only children with complete follow-up or

$$\hat{\pi}_{CC} = \frac{n_{CC}}{N_{CC}},$$

with n_{CC}being the number of observed vaccinations and N_{CC} being the total number of children in the population with complete follow-up. Finally, we also estimated coverage using (v) scaled person time or

$$\hat{\pi}_{PY} = \frac{n}{\sum_{i=1}^{N} pt_i / \max(t)},$$

with pt_ibeing the person time for child i and with $\max(t)$ being the maximum person time allowed for the study (being 1 year for this simulation study). For each method and each scenario with varying levels of incompleteness, we calculated the bias and the mean squared error (MSE).

3.3.2. Results

The results when incompleteness is random are presented in Figure 3.3 (left panel: bias; right panel: MSE). Clearly, incompleteness cannot be ignored as it leads to serious underestimation of vaccination coverage. The coverage estimation using complete cases only was unbiased. However, as expected, this came at the cost of precision with the MSE increasing as the amount of incompleteness increased. Calculating coverage as a simple incidence rate was too naïve as well and resulted in overestimation. The CDF and IPW method outperformed the naïve methods, both in terms of bias and MSE. Even when the amount of incompleteness was as high as 90%, these methods were virtually unbiased with a very low MSE. The results were similar for coverage rates of 30%, 50% or 85% (results not shown). The results for non-random incompleteness are presented in Figure 3.4 (left panel: bias; right panel: MSE). Only the IPW method performed well in this case.

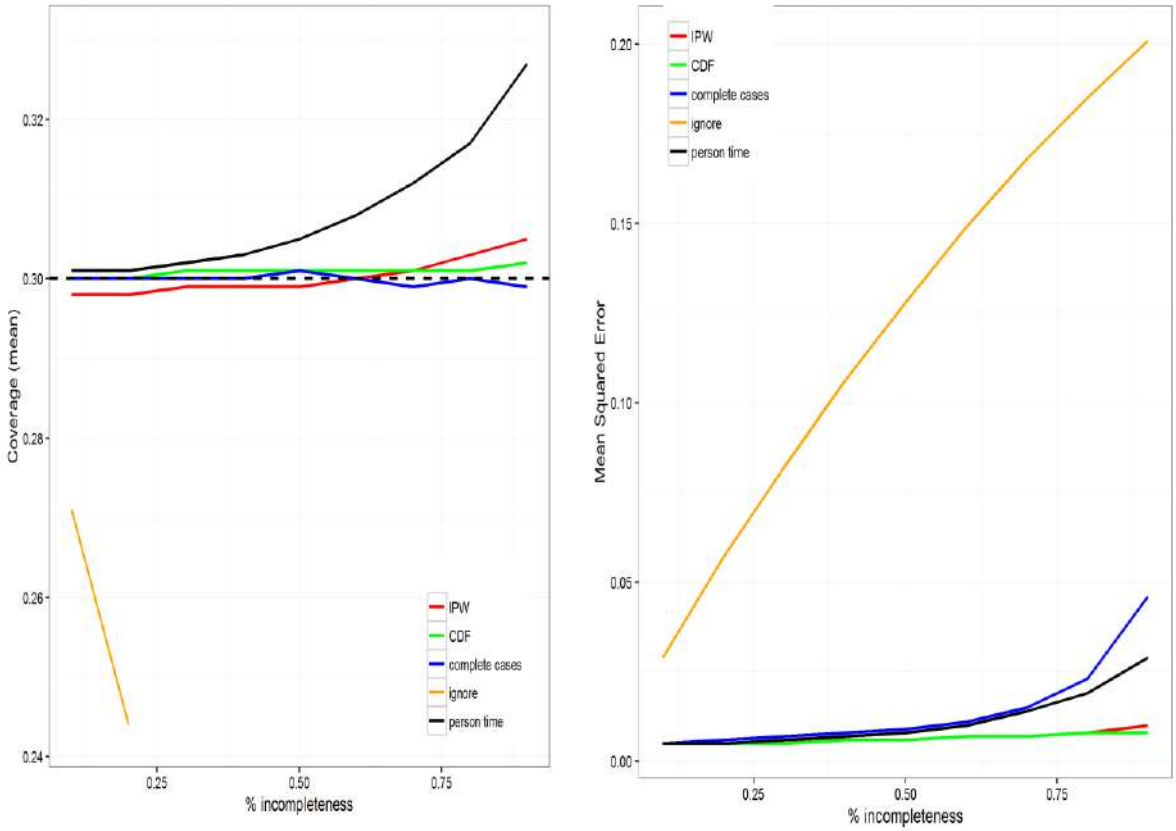


Figure 3.3: Random incompleteness: bias (left) and Mean Squared error (right) for a true coverage of 30% using different methods: (a) using the cumulative distribution function method, (b) using the inverse probability weighting method, (c) ignoring incomplete follow-up, (d) using only children with complete follow-up, and (e) using scaled person time.

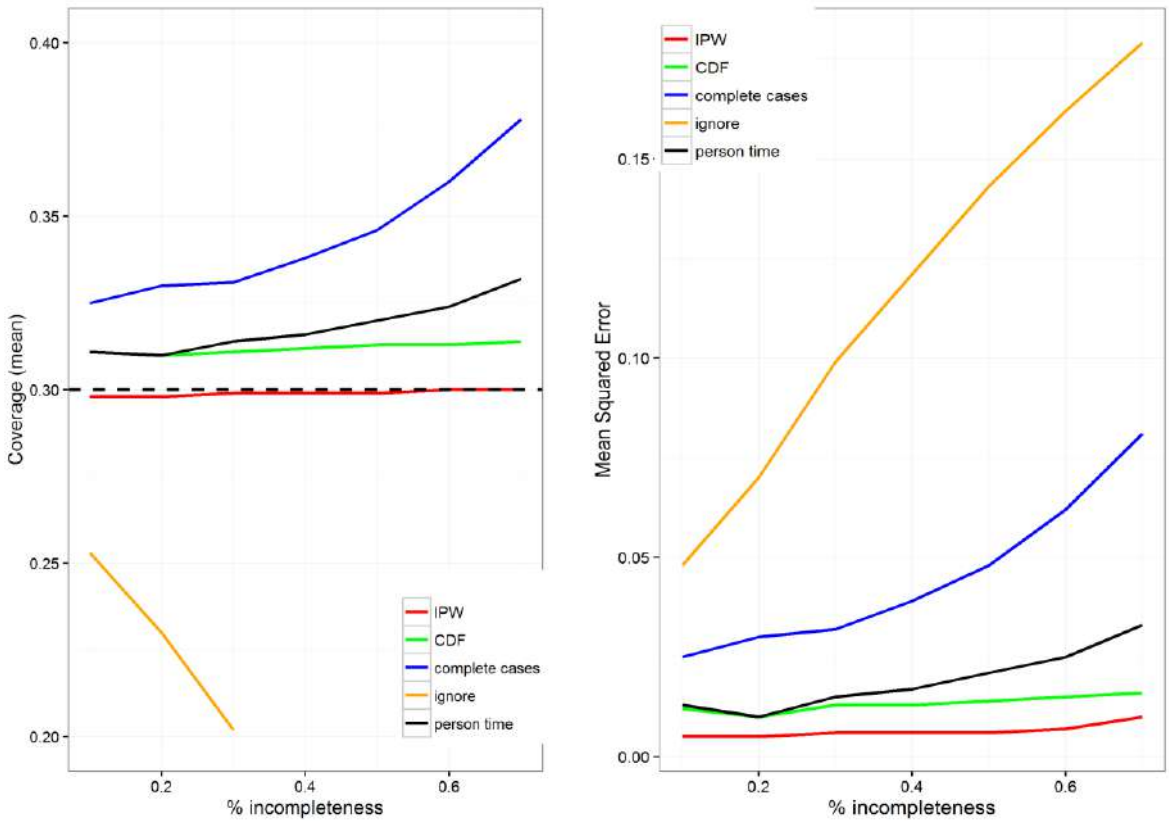


Figure 3.4: Non-random incompleteness: bias (left) and Mean Squared error (right) for a true coverage of 30% using different methods: (a) using the cumulative distribution function method, (b) using the inverse probability weighting method, (c) ignoring incomplete follow-up, (d) using only children with complete follow-up, and (e) using scaled person time.

3.4. Discussion

In this work, we investigated ways to estimate coverage in the case of incomplete follow-up, an often-encountered phenomenon with dynamic populations in healthcare databases. The cumulative distribution function (CDF) method exploits the fact that vaccines are often given at specific ages or at a given time of the year, resulting in a distinct distribution function for age or time at vaccination. The inverse probability weighting (IPW) method is similar in spirit as methods used to correct for selection bias.

Through simulation, we showed that both methods accurately and efficiently correct coverage estimates for incomplete follow-up when incompleteness is random, i.e. when incompleteness does not depend on a third variable, C, that is also related to coverage. When both coverage and incompleteness depend on a third variable C, the CDF method fails whereas the IPW method works well provided that the variable C is used as a stratification variable. The CDF method fails

when coverage and incompleteness depend on a third variable because the higher the incompleteness, the lower the effective sample size; hence, the overall coverage estimates are biased towards the coverage within the group for which incompleteness is smallest. The CDF method additionally assumes that the distribution for age/time at vaccination is the same for subjects with and without complete follow-up.

Based on theoretical considerations and on the initial simulation-based performance assessments, both the CDF and IPW methods are recommended for further testing. When incompleteness is random and given that it is reasonable to assume that the age/time at vaccination is the same for subjects with and without complete follow-up, the CDF method might be preferred as it outperforms the IPW method, both in terms of bias and MSE. When there is evidence that the incompleteness depends on a third variable, C, that is also related to coverage, the IPW method should be used whereby the C variable is a stratification variable. Logistic regression can be employed to identify potential stratification variables by building regression models to predict complete follow-up, and subsequently checking whether these variables are also related to exposure among the subjects with complete follow-up.

In future work, these methods will be used to estimate coverage from real databases to determine whether these estimates improve the naïve coverage estimates that ignore incomplete follow-up. To this end, benchmarking data on pertussis and influenza vaccination have already been obtained (Appendix P1).

3.5. References

1. Haverkate M, et al. Assessing vaccination coverage in the European Union: is it still a challenge? Expert Rev Vacc 2011; **10**(8): 1195-1205.
2. Silverman BW. Density estimation for statistics and data analyses. 1986, New York: Chapman & Hall/CRC.
3. Efron B, Tibshirani RJ. An introduction to the Bootstrap. 1993: Chapman & Hall/CRC.
4. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 1952; **47**: 663-685.
5. Mansournia MA, Altman DG. Inverse probability weighting. BMJ 2016; **352**: i189.
6. Lin DY, Wei LJ. The robust inference for proportional hazards model. J Am Stat Assoc 1989; **84**(408): 40.

4. Project 3: impact of disease- and exposure misclassification on estimation of vaccine effectiveness

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts	05.07.16	0.0	outline
Kaat Bollaerts	14.07.16	0.4	First draft, analytical derivations, discussion
Tom De Smedt	15.07.16	0.6	First draft, updated with figures
			First draft shared with co-authors
Kaat Bollaerts, Tom De Smedt, Denis Macina, Elizabeth Merrall	27.07.16		TC, discussing paper, web application and poster
Nick Andrews, Elizabeth Merrall, Silvia Perez, Denis Macina			Review, providing comments
Kaat Bollaerts	17.08.16		Incorporate comments co-authors + abstract
Kaat Bollaerts	15.09.16		Send to SC for review
Marianne van der Sande, Mendel Haag, Elizabeth Merrall, Charlotte Switzer, Sonja Banja	30.09.16		Comments received
Kaat Bollaerts	09.11.16		Adjustments, including reply to comments
Linda Levesque	25.02.17		SC review

4.1. Introduction

When studying vaccine effectiveness (VE) accurate identification of cases of the vaccine preventable disease and information on vaccination status are essential. Indeed, assuming misclassification is non-differential and independent of other errors, both disease and exposure misclassifications tend to bias the VE estimates toward the null [1]. Disease and exposure statuses may reciprocally affect each other's ascertainment (i.e. differential misclassification), and lead to biased estimates in either direction [2]. Therefore, laboratory confirmation is typically required when assessing VE [3]. However, laboratory test results are not always available or not totally accurate and, especially in healthcare database-based analyses, case definitions often have to rely on clinical criteria with the resulting risks to disease ascertainment accuracy. Likewise, the vaccination exposure information might be subject to coding or data entry errors or omissions that also potentially bias estimates of VE [4].

Concerns regarding disease and exposure misclassifications are particularly relevant when conducting epidemiological studies using healthcare databases [5]. Nonetheless large healthcare databases are increasingly being used to study vaccine use and the outcomes of vaccination. Indeed, the size of these databases allows for the study of rare events and, as they are embedded within clinical practice, they offer the potential to study the real-world effects of vaccines relatively efficiently from both cost and time perspectives.

When conducting VE studies it is important to quantify the potential impact of misclassification on the VE estimates in order to assess study feasibility, and possibly, the need to correct for the misclassification. In earlier work, the impact of disease misclassification on influenza VE has been quantified for cohort, case-control and test-negative designs using simulation studies [6,7]. These simulation studies assumed non-differential misclassification and did not account for misclassification of vaccination status. We extended these simulation studies to account for both disease- and exposure-misclassification and allow for both differential and non-differential misclassification. Furthermore, as we show that the impact of misclassification on the estimated VE depends both on the epidemiology of the vaccine-preventable disease and the expected vaccination coverage, we developed a web-application allowing simulations to be run with user-defined parameters. We illustrate the impact of misclassification on VE estimates using two examples with clearly different disease attack rates and expected vaccination coverage; a) childhood pertussis and b) pediatric seasonal influenza VE estimations.

4.2. Methods

In this section, we first present analytical derivations illustrating the impact of misclassification on VE estimates at the population level – hence ignoring estimation error – when considering

misclassification in its simplest form, that is single source non-differential misclassification. Although estimation error is ignored, such analytical derivations provide meaningful insights. However, the derivations become tedious in situations where misclassification is more complex, especially when considering the joint impact of disease and exposure misclassification. Therefore, we introduce a simulation tool that graphically displays the single and joint impact of differential and non-differential disease- and exposure-misclassification when estimating VE using the cohort, case-control, test-negative and case-cohort (screening method) designs.

4.2.1. Notation

First, let $\pi_{VPD.0}$ be the unobserved 'true' risk of disease due to the pathogen targeted by the vaccine (vaccine-preventable disease, VPD) in unvaccinated subjects, $\pi_{Other}$the corresponding risk of similar disease due to other pathogens than those targeted by the vaccine, and let γ be the 'true' vaccination coverage. Vaccination affects the VPD risk, with the risk among the vaccinated $\pi_{VPD.1}$=$(1-VE)\pi_{VPD.0}$,but does not affect the other disease risk. Furthermore, let p_0 be the observed disease prevalence among the subjects indicated as unvaccinated and $p_1$the observed prevalence among the subjects indicated as vaccinated. Finally, let SE_d be the disease sensitivity (probability of being classified as diseased if truly diseased) and SP_d the disease specificity (probability of being classified as not diseased if truly not diseased) of the case definition. Similarly let SE_e be the exposure sensitivity (probability of being classified as exposed if truly exposed) and SP_e the exposure specificity (probability of being classified as unexposed if truly unexposed) of the exposure ascertainment definition. In the case of differential misclassification, the disease misclassification parameters depend on exposure status and vice versa, yielding four disease misclassification parameters; $SE_{d,E=0}$, $SE_{d,E=1}$, $SP_{d,E=0}$, $SP_{d,E=1}$ (with E=0 indicating unvaccinated subjects and E=1 vaccinated subjects) and four exposure misclassification parameters; $SE_{e,D=0}$, $SE_{e,D=1}$, $SP_{e,D=0}$, $SP_{e,D=1}$ (with D=0 indicating not diseased subjects and D=1 diseased subjects.)

4.2.2. Impact of misclassification at population level

Non-differential disease misclassification

In the absence of exposure misclassification, the observed disease risk among the unvaccinated is the sum of the probability of having the VPD and being correctly classified as such (true positive for disease) and the probability of having the non-VPD and being incorrectly classified as having the VPD (false positive for disease) or

$$p_0 = SE_d\pi_{VPD.0} + (1-SP_d)\pi_{Other}. \quad (1)$$

Similarly, for the vaccinated, the observed prevalence equals

$$p_1 = SE_d\pi_{VPD.1} + (1-SP_d)\pi_{Other}, \quad (2)$$

with $\pi_{VPD.1} = (1-VE)\pi_{VPD.0}$.

The population-level bias due to disease misclassification is defined as the difference in VE for a population with and without misclassification or

$$\Delta = \left(1 - \frac{p_1}{p_0}\right) - \left(1 - \frac{\pi_{VPD.1}}{\pi_{VPD.0}}\right) = \frac{\pi_{VPD.1}}{\pi_{VPD.0}} - \frac{SE_d\pi_{VPD.1} + (1-SP_d)\pi_{Other}}{SE_d\pi_{VPD.0} + (1-SP_d)\pi_{Other}}. \quad (3)$$

This expression can be rewritten as

$$\Delta = \frac{(\pi_{VPD.1} - \pi_{VPD.0})(1-SP_d)\pi_{Other}}{\pi_{VPD.0}(SE_d\pi_{VPD.0} + (1-SP_d)\pi_{Other})}, \quad (4)$$

showing that if the disease specificity equals one, the bias equals zero irrespective of the disease sensitivity.

Now, solving (1) for $\pi_{VPD.0}$ and (2) for $\pi_{VPD.1}$, we have

$$\pi_{VPD.0} = (p_0 - (1 - SP_d)\pi_{Other})/SE_d. \quad (5)$$

$$\pi_{VPD.1} = (p_1 - (1 - SP_d)\pi_{Other})/SE_d,. \quad (6)$$

based on which, and given accurate estimates of disease misclassification parameters, an estimate of the 'true' VE corrected for disease misclassification can be obtained as

$$VE_\pi = 1 - \frac{p_1 - (1-SP_d)\pi_{Other}}{p_0 - (1-SP_d)\pi_{Other}}. \quad (7)$$

Interestingly, the correction equation requires an estimate of disease specificity but not of disease sensitivity. Obviously, the latter only holds if the disease misclassification is non-differential or independent by vaccination status.

Non-differential exposure misclassification

In the absence of disease misclassification, the disease risk among subjects classified as unvaccinated is the sum of the probability of having the VPD and being incorrectly classified as unvaccinated (false negative for vaccination), and the probability of having the VPD and being correctly classified as unvaccinated (true negative for vaccination) or

$$p_0 = (1 - SE_e)\,\gamma\,\pi_{VPD.1} + SP_e(1-\gamma)\pi_{VPD.0}, \quad (8)$$

with true vaccination coverage γ. Similarly, the true positives and false positives for vaccination determine the disease risk among the subjects classified as vaccinated or

$$p_1 = SE_e\,\gamma\,\pi_{VPD.1} + (1 - SP_e)(1-\gamma)\pi_{VPD.0}. \quad (9)$$

The population-level bias due to exposure misclassification is now defined as

$$\Delta = \left(1 - \frac{p_1}{p_0}\right) - \left(1 - \frac{\pi_{VPD.1}}{\pi_{VPD.0}}\right) = \frac{\pi_{VPD.1}}{\pi_{VPD.0}} - \frac{SE_e\,\gamma\,\pi_{VPD.1} + (1-SP_e)(1-\gamma)\pi_{VPD.0}}{(1-SE_e)\,\gamma\,\pi_{VPD.1} + SP_e(1-\gamma)\pi_{VPD.0}}. \quad (10)$$

This expression shows that the impact of sensitivity will be largest when coverage is high whereas the impact of specificity will be largest when coverage is low.

Solving (8) and (9) for $\pi_{VPD.0}$ and for $\pi_{VPD.1}$, we obtain

$$\pi_{VPD.0} = \big(p_0 SE_e - p_1(1 - SE_e)\big)/((1-\gamma)(SE_e + SP_e - 1)), \quad (11)$$

$$\pi_{VPD.1} = \big(p_1 SP_e - p_0(1 - SP_e)\big)/(\gamma(SP_e + SE_e - 1)). \quad (12)$$

Then, an expression of the 'true' VE corrected for exposure misclassification corresponds to

$$VE_{\pi} = 1 - \left(\frac{1-\gamma}{\gamma}\right)\frac{p_1 SP_e - p_0(1-SP_e)}{p_0 SE_e - p_1(1-SE_e)}. \quad (13)$$

This correction equation depends—next to the observed disease prevalences—on both exposure sensitivity and specificity as well as on the 'true' vaccination coverage.

4.2.3. Simulation tool

Similar to work published by Orenstein [7] and Jackson [6] we simulated populations at risk for two outcomes; the VPD and a comparable outcome due to infection with one or more pathogen(s) not targeted by the respective vaccination. We assumed that a number of subjects are vaccinated with coverage of γ. Unvaccinated subjects could develop the VPD (only once) with a risk equal to $\pi_{VPD.0}$ and the health outcome due to infection with other pathogens (only once) with a risk equal to π_{Other}. For vaccinated subjects, the risk of developing the VPD is reduced to $\pi_{VPD.1} = (1 - VE)\pi_{VPD.0}$, whereas the risk due to other pathogens is unaffected by vaccination. We furthermore assumed that the risks of developing both outcomes are independent. After having allocated the 'true' disease and exposure status, we randomly allowed these events to be misclassified. In particular, for the disease events, diseased cases were misclassified as not diseased with a probability of $1 - SE_d$ and non diseased cases were misclassified as diseased with a probability of $1 - SP_d$. The same holds for the exposure events, but using the exposure sensitivity SE_e and specificity SP_e parameters to simulate misclassification. In the case of differential misclassification, the disease misclassification parameters depended on exposure status and vice versa, yielding eight misclassification parameters in total; four disease misclassification parameters; $SE_{d,E=0}$, $SE_{d,E=1}$, $SP_{d,E=0}$, $SP_{d,E=1}$ and four exposure misclassification parameters; $SE_{e,D=0}$, $SE_{e,D=1}$, $SP_{e,D=0}$, $SP_{e,D=1}$.

Then, for a given parameter setting, a large number of simulated populations $(k = 1,2, \ldots K)$ of a predefined population size *N* are generated. Based on the observed exposure and disease statuses in each population *k*, VE is estimated using the cohort, case-control, test-negative and case-coverage (or screening method) designs as described in Tables3.1 and 3.2. Then, these estimates are compared with the true VE used to generate the simulated populations. The biases are compared graphically.

The simulation model is developed using R 3.3.1[8]. To allow modifications to the simulations for other parameter settings/diseases while maximizing user-friendliness, we encapsulated the source code of the simulation model in a web application created using the Shiny package [9]. Through the web application, the user can set all the necessary input parameters and the output files can be downloaded. The application will be made available to the ADVANCE consortium (user guide is provided as supplementary file).

	Cohort	**Case-control**
For each simulated population	We calculate the risk of the vaccine-preventable disease in the exposed (vaccinated) individuals versus in the unexposed (unvaccinated) individuals.	We identify cases of VPD and controls without VPD; and for these two groups compare the odds of exposure as an odds ratio. We used cumulative sampling as in [7].
Estimate VE as	$\widehat{VE}_{Co\,k} = 1 - \widehat{RR}_{Co\,k} = 1 - \frac{\widehat{p_v}}{\widehat{p_u}}$, with $\widehat{RR}_{Co\,k}$ the estimated ratio of the risk of VPD in the vaccinated vs. unvaccinated, and estimated risks $\widehat{p_v}$ and $\widehat{p_u}$ based on observed proportions of VPD in the vaccinated and unvaccinated respectively.	$\widehat{VE}_{CC\,k} = 1 - \widehat{OR}_{CC\,k} = 1 - \frac{\widehat{p_d}/(1-\widehat{p_d})}{\widehat{p_n}/(1-\widehat{p_n})}$, with $\widehat{OR}_{CC\,k}$ the estimated ratio of odds of exposure in cases vs. controls, which is equivalent to the odds of VPD in the exposed versus unexposed; and $\widehat{p_d}$ and $\widehat{p_n}$ observed proportions of exposure in the cases (diseased) and controls (non-diseased).

Table 3.1: Estimation of vaccine effectiveness (VE) for the cohort and case-control design.

	Test-negative	**Case-coverage (screening method)**
For each simulated population	Here, the cases are the outcome events due to the VPD pathogen, or test-positives, and the controls are the outcome events due to other pathogens, i.e. test-negatives.	We use only the exposure statuses of the observed cases and compare the odds of exposure in these cases with the odds of exposure in the overall population; the latter odds is typically derived from external sources.
Estimate VE as	$\widehat{VE}_{TN\,k}$ $= 1 - \widehat{OR}_{TN\,k}$ $= 1 - \frac{\widehat{p_{tp}}/(1-\widehat{p_{tp}})}{\widehat{p_{tn}}/(1-\widehat{p_{tn}})}$, With $\widehat{OR}_{TN\,k}$ being the estimated ratio of the odds of exposure in the cases versus controls by test-negative study design; $\widehat{p_{tp}}$ and $\widehat{p_{tn}}$ observed proportions of exposure in test-postitive and test-negative individuals respectively.	$\widehat{VE}_{SCREEN\,k}$ $= 1 - \widehat{OR}_{SCREEN\,k}$ $= 1 - \frac{\widehat{p_d}/(1-\widehat{p_d})}{\hat{X}/(1-\hat{X})}$, With $\widehat{OR}_{SCREEN\,k}$ being the estimated ratio of the odds of exposure in the cases versus controls by screening study design; $\widehat{p_d}$ as defined for the case-control design and $\hat{X}$ an, often externally-derived, estimate of the vaccine coverage for the study population. For the simulation model, for each simulation run, $\hat{X}$ is based on the proportion of individuals with observed exposures.

Table 3.2: Estimation of vaccine effectiveness (VE) for the test- negative and case-coverage design.

4.2.4. Scenarios

General settings

In this section, we present two specific vaccination scenarios, paediatric seasonal influenza and childhood pertussis vaccination. For each subsequent simulation scenario, we set K = 1000 and N = 50000 whereas VE, vaccination coverage and the respective attack rates depend on the specific scenarios detailed below. We vary one-by-one the disease- or exposure-misclassification from: $\{0.5, 0.6, \ldots 1\}$ while fixing the remaining misclassification parameters to 1.

Paediatric Seasonal Influenza

For consistency with Orenstein [7] and Jackson [6], we assumed a VE of 70%, an attack rate (AR) of influenza in the unvaccinated of 15% and an AR of influenza-like illness not caused by influenza of 30%. The paediatric seasonal influenza vaccination coverage was assumed to be 10% [10].

Pertussis primary series

We assumed a VE of 80%, derived as a conservative value from a Cochrane systematic review of vaccine efficacy estimates obtained in randomized controlled trials that found the efficacy of acellular pertussis vaccines in paediatric primary series to range between 71% and 85% [11]. We furthermore assumed that the AR of pertussis in the unvaccinated was 15% [12] and the AR of the non-vaccine preventable pathogens with similar clinical pattern was 10.5% [13]. For the vaccination coverage, we assumed a value of 95%, which reflects a coverage rate commonly reported for the paediatric primary series in high-income countries[14].

4.3. Results

4.3.1. Paediatric seasonal influenza

In the seasonal influenza scenario and assuming non-differential exposure and disease misclassification (Figure 3.1, top), the exposure specificity had the largest impact on the VE estimates followed by disease specificity. In the case of differential misclassification, the bias could go in either direction, with the estimated VE showing very large deviations from the true VE. In case of differential exposure misclassification (Figure 3.1, middle), the exposure specificity for the diseased had the strongest impact among all four exposure misclassification parameters and biases the VE estimates downwards. Finally, in case of differential disease misclassification (Figure 3.1, bottom), the disease specificity for the exposed had the largest impact among the four disease misclassification parameters. The impact of misclassification was virtually the same across designs, with the exception that low disease specificity among

the exposed was more strongly biasing the VE estimates when using the test-negative design compared with the other designs (Figure 3.1, bottom).

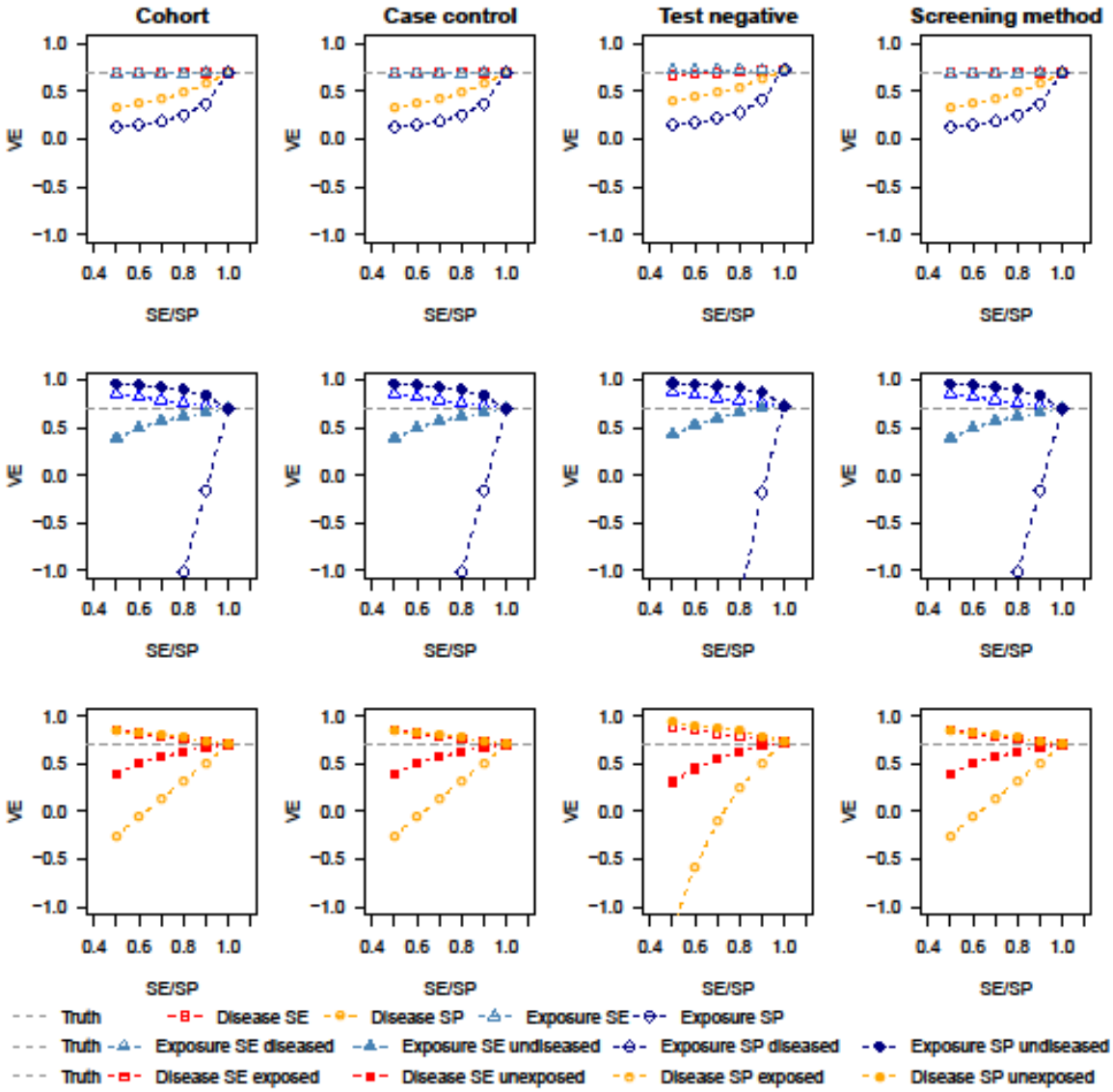


Figure 3.1: Influenza scenario: Vaccine effectiveness by design for varying levels of exposure- and disease misclassification (top: non-differential exposure and disease misclassification; middle: differential exposure misclassification; bottom: differential disease misclassification).

4.3.2. Pertussis primary series

In the pertussis scenario and assuming non-differential exposure and disease misclassification (Figure 3.2, top), the exposure sensitivity had the largest impact followed by disease specificity. In case of differential exposure misclassification (Figure 3.2, middle), the exposure sensitivity for the non-diseased had the strongest impact among all four exposure misclassification parameters and biased the VE estimates downwards. Finally, in case of differential disease misclassification (Figure 3.2, bottom), the disease specificity for the exposed had the largest impact among the four disease misclassification parameters. The impact of the misclassification parameters was comparable across designs, with the exception of disease specificity among the exposed leading to stronger bias when using the test-negative design compared to other designs. As with paediatric influenza, the bias due to differential misclassification could go in either direction and lead to very large deviations from the true VE.

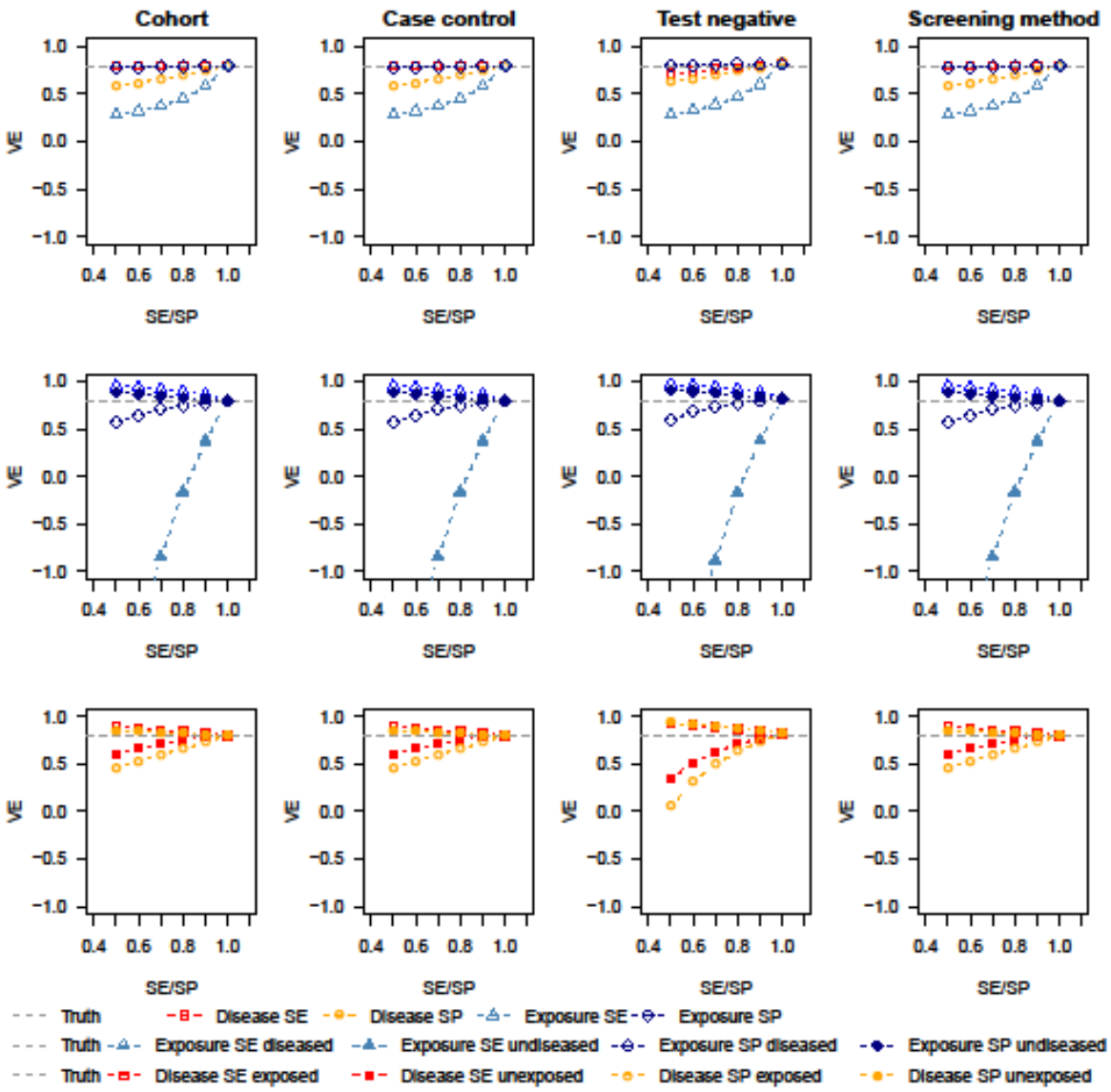


Figure 3.2: Pertussis scenario: Vaccine effectiveness by design for varying levels of exposure- and disease misclassification (top: non-differential exposure and disease misclassification; middle: differential exposure misclassification; bottom: differential disease misclassification).

4.3.3. Simulation tool

As the outcomes of the above scenarios will differ widely depending on the chosen parameters, a web application was created where the user can set the parameters for any given scenario. The web application will subsequently calculate the estimated VE graphs for the chosen parameters. Simulations can be performed assuming either non-differential or differential misclassification.

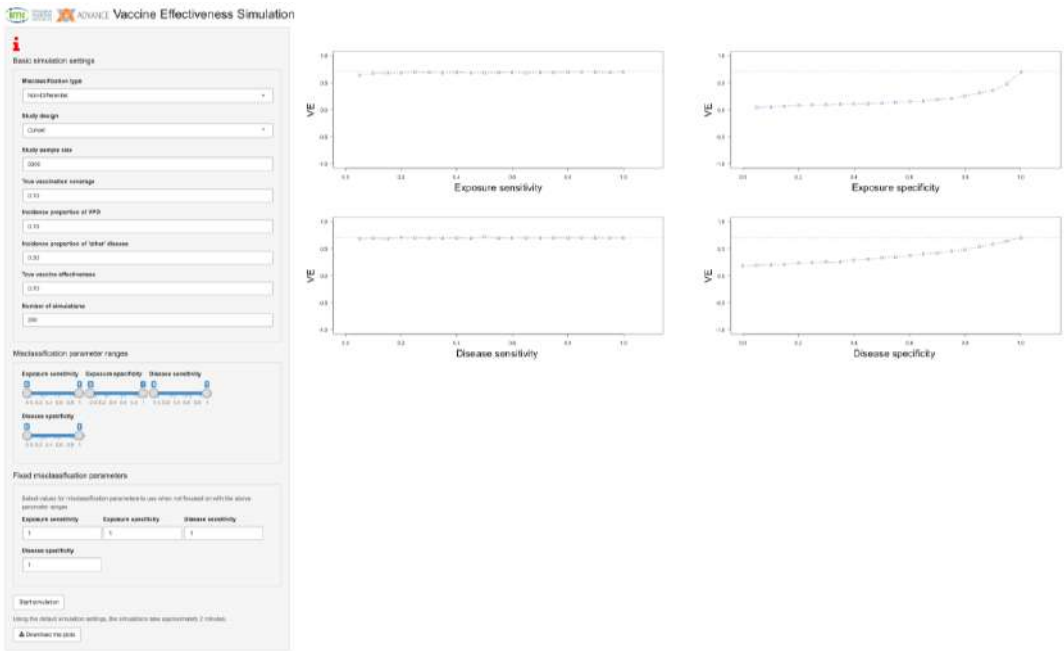


Figure 3.3: Web application for the simulation of vaccine effectiveness using different parameters the end-user can select. This application is made using R 3.3.1 and the Shiny package.

4.4. Discussion

The development of the simulation tool presented an opportunity to explore the interplay of disease- and exposure-misclassification in VE estimations for different study designs. In this study, we explored the single impact of non-differential and differential disease and exposure misclassification on the estimation of VE for childhood seasonal influenza and pertussis. Depending on the scenario, the misclassification parameters had differing impacts. Decreased exposure specificity (poorer identification of non-vaccinees) had the greatest impact on influenza VE estimation. Conversely decreased exposure sensitivity (poorer identification of vaccinees) had the greatest impact on pertussis VE estimation. These differing impacts correspond to the respective low and high vaccine coverages in the two scenarios, which is supported by the analytical derivations on the impact of non-differential exposure misclassification at population-level.

The dependence of the impact of misclassification on the scenario stimulated us to develop a user-friendly simulation tool that can be modified by users to their own study scenario. The tool allows users to assess the single and joint impact of both differential and non-differential disease- and exposure misclassification on VE estimates from cohort, case-control, case-coverage and test-negative studies.

The impact of the misclassification parameters was found to be more pronounced than that of the different study designs. This suggests that misclassification as such is not a compelling argument for or against a certain design. However, other sources of bias such a confounding and selection bias might be of concern for a particular VE study, making some designs better suited than others. For instance, observational studies on influenza VE might be strongly confounded by differences in healthcare seeking behaviour between vaccinated and unvaccinated persons, therefore the test-negative design could be an appropriate choice in this case [15].

This simulation tool helps to anticipate the magnitude and direction of the bias when estimating VE based on potentially misclassified data. The tool can guide the selection of the exposure and disease definitions that will minimize bias due to misclassification or set acceptable levels of sensitivity and specificity. In addition, if the potential impact of misclassification is found to be unacceptable, several methods to adjust estimates for misclassification exist, although they are not yet commonly used in pharmacoepidemiology [16]. We provided the correction equations for VE estimates in case of non-differential single source (either exposure or disease) misclassification (Section 3.2.2). Other correction methods include, amongst others, probabilistic bias analyses [17], Bayesian bias analyses[18, 19], modified maximum likelihood

methods [20] and imputation-like methods [21-24]. All these methods require assumptions about estimates of the disease- and exposure misclassification parameters, which -if deemed required- can be obtained using internal or external validation studies. Projects 5, 6 and 7 describe methods for validation.

Several limitations or areas of further development are worth considering. The simulation tool singles out the impact of disease- and exposure-misclassification and ignores other sources of bias. Specifically, it is assumed that there is no confounding and no selection bias. In addition, the tool does not account for dependent misclassification. For binary variables, misclassification is dependent when the probability of misclassification of one variable depends on the correctness of classification on the other variable [25]. For example, dependent measurement errors might arise if data on both exposure and outcome were obtained from medical records with paucity of data for some but not all subjects. Furthermore, the tool assumes binary disease- and exposure variables, whereas particularly the exposure variable might be polytomous (no vaccination, partial or complete vaccination).

Nonetheless, we believe the simulation tool is a useful tool to guide researchers to better design, conduct and interpret future VE studies when data are subject to misclassification. We advocate the use of such a simulation tool and the modifications of the parameters according to the study specifics since we have shown that the impact of misclassification strongly depends on the study scenario.

4.5. References

1. Jurek AM, Greenland S, Maldonado G How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? Int J Epidemiol 2008;**37**(2): 382-5.
2. Copeland KT, et al. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol 1977;**105**(5): 488-95.
3. Ozasa K. The effect of misclassification on evaluating the effectiveness of influenza vaccines. Vaccine 2008;**26**(50): 6462-6465.
4. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. Am J Epidemiol; 1986 **123**(4): 736-51.
5. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;**58**(4): 323-37.
6. Jackson ML, Rothman KJ. Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness. Vaccine 2015;**33**(11): 1313-6.
7. Orenstein EW, et al. Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. Int J Epidemiol 2007;**36**(3): 623-31.
8. R Development Core Team, R: a language and environment for statistical computing, ed. R Foundation for Statistical Computing. 2013, Vienna, Austria.
9. Chang WJC, Allaire JJ, Yihui Xie Y, McPherson J. Shiny: Web Application Framework for R. 2016.

10. Mereckiene J, et al. Seasonal influenza immunisation in Europe. Overview of recommendations and vaccination coverage for three seasons: pre-pandemic (2008/09), pandemic (2009/10) and post-pandemic (2010/11). Euro Surveill 2014;**19**(16): 20780.
11. Zhang L, et al. Acellular vaccines for preventing whooping cough in children. Cochrane Database Syst Rev 2014(9): p. CD001478.
12. Gustafsson L, et al., A controlled trial of a two-component acellular, a five-component acellular, and a whole-cell pertussis vaccine. N Engl J Med 1996;**334**(6): 349-55.
13. Stehr K, et al. A comparative efficacy trial in Germany in infants who received either the Lederle/Takeda acellular pertussis component DTP (DTaP) vaccine, the Lederle whole-cell component DTP vaccine, or DT vaccine. Pediatrics 1998;**101**(1 Pt 1): 1-11.
14. WHO/UNICEF. Estimates of national immunization coverage. DTP3 & DTP4. . 07/06/2016]; Available from: http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragebcg.htm.
15. De Serres G, et al. The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomised placebo-controlled clinical trials. Euro Surveill 2013;**18**(37): pii. 20585.
16. Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. Curr Epidemiol Rep 2014;**1**(4): 175-185.
17. Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data. Statistics for Biology and Health. 2009, New York: Springer.
18. Gustafson P. Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments, ed. C.a. Hall. 2003.
19. MacLehose RF, et al. Bayesian methods for correcting misclassification: an example from birth defects epidemiology. Epidemiology 2009;**20**(1): 27-35.
20. Lyles RH, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. Epidemiology 2011;**22**(4): 589-97.
21. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. Int J Epidemiol 2006;**35**(4): 1074-81.
22. Edwards JK, et al. Accounting for misclassified outcomes in binary regression models using multiple imputation with internal validation data. Am J Epidemiol 2013;**177**(9): 904-12.
23. Kuchenhoff H, Mwalili SM, Lesaffre E.A general method for dealing with misclassification in regression: the misclassification SIMEX. Biometrics 2006 **62**(1): 85-96.
24. Messer K, Natarajan L.Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. Stat Med 2008;**27**(30): 6332-50.
25. Hernan MA, Cole SR. Invited Commentary: Causal diagrams and measurement bias. Am J Epidemiol 2009;**170**(8): 959-62; discussion 963-4.

5. Project 4: validation of case-finding algorithms in healthcare research: analytical interrelations between validity indices

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts	19.07.16	0.0	Methods section -
Kaat Bollaerts	27.07.16	0.1	Shared with Rosa and Caitlin
Nick Andrews	27.07.16	0.1	Review
Rosa Gini	30.08	0.2	Draft of introduction, draft of validation subsection of the methods, draft of first subsection of applications, draft of discussion, overall structure
Kaat Bollaerts	21.09.16	0.5	New version
Caitlin Dodd	04.10.16		Add applications
Tom De Smedt	24.10.16		Web application
Kaat Bollaerts	25.10.16		Sensitivity analyses
Rosa Gini, Nick Andrews	30.10.16		Review
Kaat Bollaerts, Rosa Gina, Nick Andrews, Tom De Smedt, Caitlin Dodd	15.11.16		TC to discuss report
Kaat Bollaerts	05.12.16		Rewriting, methods – discussion – including results SA, abstract
Kaat Bollaerts	18.01.17		Rewrite…again after internal discussions at P95
Linda Levesque	25.02.17		

5.1. Introduction

Large computerised healthcare databases have become a prominent source of information in pharmacoepidemiology. They often cover large geographical areas, their size permit the study of rare events and their establishment within clinical practices provides the potential to study real-world effects of pharmaceutical products in a timely and cost-efficient manner. However, although healthcare databases provide a valuable source of data for pharmacoepidemiological research, these data are collected primarily for clinical and administrative use rather than for research and concerns regarding data quality exist, potentially resulting in misclassification bias [1, 2].

Validation is recognized as an important component of research using healthcare databases and studies reporting on the validity of exposure- or disease-case-finding algorithms (CFAs) are being increasingly performed [3-5]. A CFA is a combination of values of routinely collected variables that allow identification of cases of disease (or exposure) in a population captured by a database, without having to contact the patient. The validity indices of CFAs typically evaluated include sensitivity (SE_{CFA}), specificity (SP_{CFA}), positive predictive value (PPV_{CFA}) and negative predictive value (NPV_{CFA}). Once the value of these validity indices are known, the observed prevalence or risk estimates can be corrected for misclassification [6, 7].

In this report, we show how the true disease prevalence, the observed prevalence (as estimated from the misclassified data) and the four validity indices SE, SP, PPV and NPV are analytically interrelated. Specifically, for every combination of the observed prevalence and two other parameters, we derived the analytical expressions to obtain the remaining three parameters. We developed a user-friendly web-application that calculates validity indices and Monte Carlo 95% uncertainty intervals given the observed prevalence and any two other parameters. In addition, we conducted sensitivity analyses investigating the impact of estimation error in the input parameters on the derived parameters. We demonstrate the web-application using published results from studies validating intussusception and pneumonia CFAs.

5.2. Methods

5.2.1. Definitions

A CFA is typically validated by comparing its outcomes with that of a 'gold standard', which is assumed to perfectly represent the true dichotomous disease status. This concept is conventionally represented using a 2 x 2-table representing the joint probability distribution of the CFA-derived classification and the true disease status or 'gold standard' measure (Table 4.1). Given this representation, SE_{CFA} is the proportion of patients with the disease of interest

who is CFA-positive, SP_{CFA} is the proportion of persons without the disease who is CFA-negative, PPV_{CFA} is the proportion of CFA-positive patients who truly have the disease of interest and NPV_{CFA} is the proportion of CFA-negative persons without the disease of interest. These four validity indices are all conditional probabilities, where SE_{CFA}, SP_{CFA}, PPV_{CFA} and NPV_{CFA} are conditioned on the number of true positives, true negatives, test positives and test negatives, respectively (Table 1). The observed prevalence (P_{CFA}) is then the proportion of test positives and the true prevalence (π) the proportion of truly diseased among all N subjects. Observe that the observed prevalence and the four validity indices are CFA-dependent, which is indicated by the subscript. From now onwards, for reasons of parsimony, we omit this subscript.

Table 5.1: Validity indices for dichotomous data: sensitivity (SE_{CFA}), specificity (SP_{CFA}), positive (PPV_{CFA}) and negative predictive value (NPV_{CFA}), the observed (P_{CFA}) and true prevalence (π).

		'Gold' standard		
		Positive	Negative	Validity index
Case Finding Algorithm	Positive	Nr. of True positives TP	Nr. of False positives FP	PPV_{CFA} =TP/(TP+FP)
	Negative	Nr. of False negatives FN	Nr. of True negatives TN	NPV_{CFA} =TN/(FN+TN)
Validity index		SE_{CFA}=TP/(TP + FN)	SP_{CFA}=TN/(FP + TN)	N =TP+FP+FN+TN P_{CFA}= (TP+FP)/N π = (TP+FN)/N

5.2.2. Interrelationships between validity indices

The 2 x 2-table representation (Table 4.1) shows how the true prevalence, observed prevalence and the validity indices SE, SP, PPV and NPV are interrelated. Alternatively, these interrelations can be expressed in terms of the actual parameters themselves (and not the cell counts of the 2x2-table). Indeed, starting from the expression relating the observed prevalence to the true prevalence[7, 8] and from the definitions of PPV and NPV [9], we have the following system of algebraic equations with six unknown parameters;

$$P = SE\,\pi + (1 - SP)(1 - \pi), \qquad (1)$$

$$PPV = SE\,\pi/\big(SE\,\pi + (1-SP)(1-\pi)\big), \quad (2)$$

$$NPV = SP(1-\pi)/\big((1-SE)\pi + SP(1-\pi)\big). \quad (3)$$

Hence, if we know three parameters, we can derive the others. The observed prevalence P is easily obtained by applying the CFA to the population in the database. Then, once we can estimate two other parameters, the remaining parameters can be analytically derived by solving the system of algebraic equations above. For all combinations of P and any two other parameters, the analytical solutions for the remaining three parameters are given in Table 4.2.

The true prevalence, observed prevalence and the four validity indices are all (conditional) probabilities, and hence are bounded between zero and one. This imposes constraints on the input parameters without which the analytically derived parameters might be outside the zero-to-one range (constraints in Table 4.3). More restrictive constraints result if we impose that the CFA should detect disease better than chance alone [7] (constraints in Table 4.4). A CFA performs better than chance if it selects diseased persons with a higher probability than it does non-diseased persons. Note that the issue of a CFA performing worse than chance is easily alleviated through swapping the CFA-results, i.e. by re-labeling the CFA-positive results as negative and vice versa.

Finally, if we know the uncertainty associated with some of the input parameters, we can propagate the uncertainty to the derived parameters through Monte Carlo sampling. The uncertainty of the derived parameters can then be presented by means of uncertainty intervals (UIs) [10]. As the true prevalence, observed prevalence and the validity indices are typically all correlated, the sampling should ideally reflect this. Not accounting for correlation among the parameters might result in too wide UIs and in sampling unlikely parameter combinations. However, the correlations among the parameters are typically unknown. Therefore, we used independent sampling but rejected the invalid parameter combinations as defined by the constraints in Table 4.3 or 4.4.

Table 5.2: Overview of the interrelations between validity indices and the true prevalence, given the observed prevalence P and two other parameters.

	Known		Expressions	
1.	Π, P, SE	$SP = 1 - \frac{P\ (SE \times \Pi)}{1\text{-}\Pi}$	$PPV = \frac{SE \times \Pi}{P}$	$NPV = 1 - \frac{\Pi(1-SE)}{1-P}$
2.	Π, P, SP	$SE = \frac{P-(1-\Pi)(1-SP)}{\Pi}$	$PPV = 1 - \frac{(1-\Pi)(1-SP)}{P}$	$NPV = \frac{SP\ (1-\Pi)}{1-P}$
3.	Π, P, PPV	$SE = \frac{P \times PPV}{\Pi}$	$SP = 1 - \frac{P(1-PPV)}{1-\Pi}$	$NPV = 1 - \frac{\Pi - P \times PPV}{1-P}$
4.	Π, P, NPV	$SE = 1 - \frac{1-\Pi-NPV(1-P)}{\Pi}$	$SP = \frac{NPV(1-P)}{1-\Pi}$	$PPV = 1 - \frac{1-\Pi-NPV(1-P)}{P}$
5.	P, SE, SP	$\Pi = \frac{P+SP-1}{SE+SP-1}$	$PPV = 1 - \frac{(P-SE)(1-SP)}{P\ (1-SP-SE)}$	$NPV = \frac{(P-SE)\ SP}{(1-P)(1-SP-SE)}$
6.	P, SE, PPV	$\Pi = \frac{P \times PPV}{SE}$	$SP = 1 - \frac{P\ (1-PPV)SE}{SE - P \times PPV}$	$NPV = 1 - \frac{(1-SE)\ (P \times PPV)}{SE\ (1-P)}$
7.	P, SE, NPV	$\Pi = \frac{(1-P)(1-NPV)}{1-SE}$	$SP = \frac{(1-P)(1-SE)\ NPV}{(1-SE)-(1-P)(1-NPV)}$	$PPV = \frac{SE \times (1-P)(1-NPV)}{P(1-SE)}$
8.	P, SP, PPV	$\Pi = 1 - \frac{P \times (1-PPV)}{1-SP}$	$SE = \frac{P \times PPV(1-SP)}{1-SP-P(1-PPV)}$	$NPV = \frac{P \times SP \times (1-PPV)}{(1-P)(1-SP)}$
9.	P, SP, NPV	$\Pi = 1 - \frac{(1-P) \times NPV}{SP}$	$SE = \frac{P \times SP - (1-SP)(1-P) \times NPV}{SP - (1-P) \times NPV}$	$PPV = \frac{P \times SP - (1-SP)(1-P)NPV}{P \times SP}$
10.	P, PPV, NPV	$\Pi = (1-P)(1-NPV) + P \times PPV$	$SE = \frac{P \times PPV}{(1-P)(1-NPV) + P \times PPV}$	$SP = \frac{(1-P) \times NPV}{1-(P \times PPV + (1-P)(1-NPV))}$

Table 5.3: Constraints on the input parameters ensuring that the derived parameters belong to the interval [0,1].

	Known	Constraints		
1.	Π, P, SE	$1 - P > \Pi(1 - SE)$	$P > SE\ \Pi$	
2.	Π, P, SP	$P > (1 - \Pi)(1 - SP)$	$P < 1 - SP(1 - \Pi)$	
3.	Π, P, PPV	$\Pi > P\ \ PPV$	$\Pi < 1 - (1 - PPV)P$	
4.	Π, P, NPV	$\Pi > (1 - NPV)(1 - P)$	$\Pi < 1 - NPV(1 - P)$	
5.	P, SE, SP	$\frac{P + SP - 1}{SE + SP - 1} > 0$	$\frac{P + SP + 1}{SE + SP - 1} > 0$	$\frac{SE(P + SP) - P}{SE + SP - 1} < 1$
6.	P, SE, PPV	$P\ (PPV(1\text{-}SE)/SE + 1) < 1$		
7.	P, SE, NPV	$(1 - P)\frac{1 - SE\ NPV}{1 - SE} < 1$		
8.	P, SP, PPV	$P + SP\frac{P(1 - PPV)}{1 - SP} < 1$		
9.	P, SP, NPV	$P\ SP > (1 - SP)(1 - P)NPV$		
10.	P, PPV, NPV			

Table 5.4: Parameter constraints corresponding to a case-finding algorithm that performs better than chance.

Constraints	
$P > \Pi \times SE$	$P < SE$
$\Pi \times (1 - SE) <$	$SP > (1 - P)$
$(1 - \Pi) \times SP <$	$\Pi <$
$(1 - \Pi) \times (1 - SP) <$	$1 - NPV < \Pi$

5.2.3. Web-application

To allow users to easily explore the interrelations between the true prevalence, observed prevalence and the validity indices SE, SP, PPV and NPV, we developed a web application using R [11] and the Shiny package [12]. The application will be made available to the ADVANCE consortium. The application calculates the validity indices given user-defined mean values of the observed prevalence and any other two parameters. Optionally, the 95% percentile UIs of the derived parameters are calculated through Monte Carlo simulation if the 95% confidence intervals (CIs) of the known parameters are provided. More specifically, we assign beta distributions—typically used to model uncertainty in probabilities [13]—to all known parameters for which CIs are provided, with the shape parameters of the beta distribution derived from the provided mean values and CIs based on the method of moments [14]. Invalid combinations of parameter values are discarded from the Monte Carlo analysis and the percentages of constraint violations are reported. We provide two types of UIs, one with the 'bounded between 0 and 1' constraints applied (Table S.1) and one with the more restrictive 'better than chance' constraints applied (Table S.2.)

The web-application is demonstrated using published results on the validation of two CFAs: one for intussusception and one for pneumonia.

5.2.4. Sensitivity analyses

We additionally conducted sensitivity analyses to investigate the impact of estimation error in the input parameters on the derived parameters. For every combination of the observed prevalence and any two other parameters, we varied the input parameters one-at-the-time (OAT) while keeping the remaining input parameters at their baseline values [15]. Specifically, the input parameters p are varied between an under- and an overestimation with one standard error s.e. (i.e. between $p - s.e.$ and $p + s.e.$) with s.e. calculated for the binomial proportion p from a sample of size 1000. We investigated three baseline scenarios for varying levels of $\pi = \{0.01, 0.05, 0.2\}$ while keeping SE and SP fixed at 0.95 and 0.75, respectively. The corresponding baseline values for the observed prevalence and the predictive values were $P = \{0.26, 0.28, 0.39\}$, $PPV = \{0.04, 0.17, 0.49\}$ and $NPV = \{1.0, 1.0, 0.98\}$. The biases of the derived indices are expressed relative to their standard errors as well. For the sensitivity analyses, we applied the less restrictive 'bounded between 0 and 1' constraints.

5.3. Results

5.3.1. Illustrations

Ducharme et al conducted a validation study of the diagnostic, procedural, and billing codes for the identification of intussusception in children [16]. The authors calculated SE, SP, PPV, and NPV using manual validation of hospital records using the Brighton Collaboration diagnostic criteria as a gold standard. Case finding algorithms were based on a single or combination of ICD-9 diagnosis codes, procedure codes, and billing codes. In total, 185 out of the 417,997 patients (0.04%) met the case criteria according to the CFA chosen by the authors. The CFA's PPV was 72.4% (95%CI: 65.4-78.7) while the NPV was >99.9% (95% CI: >99.9-100.0). The estimated SE was 89.3% (95% CI: 83.3-93.8) and the SP was >99.9% (95% CI: >99,9-100). Starting from the observed prevalence, SE and PPV, we derived the NPV and SP (Figure 5.1). The derived validity indices are, as expected, the same as those reported by Ducharme et al [16]. The derived true prevalence was 0.032% (95% UI: 0.031 – 0.034).

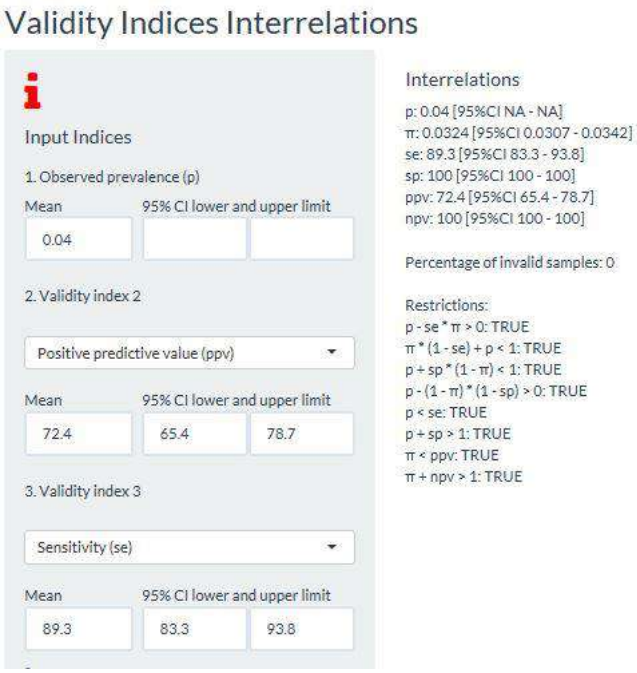


Figure 5.1: Intussusception; deriving true prevalence, specificity and negative predictive value from the observed prevalence, sensitivity and positive predictive value.

A second example was the validation study of claims-based pneumonia CFA. In a cross-sectional study of patients visiting the emergency department (ED) of a hospital in Salt Lake City, Utah during a 5-month period, Aronsky et al assessed the validity of five different claims-based pneumonia CFA

against a 'gold standard' of manual review of each patient encounter [17]. Their selected algorithm was positive for 219 of 10828 ED encounters (2.02%). For this algorithm, the authors reported SE of 65.1% (95% CI: 59.2-70.5), SP of 99.6% (95% CI: 99.5- 99.7), PPV of 80.8% (95% CI: 75.1-85.5), and NPV of 99.1% (95% CI: 98.9- 99.3). As expected, the derived validity indices were the same as the published values. The estimated true prevalence was 2.51% (95% UI:2.4 – 2.6) (Figure 5.2).

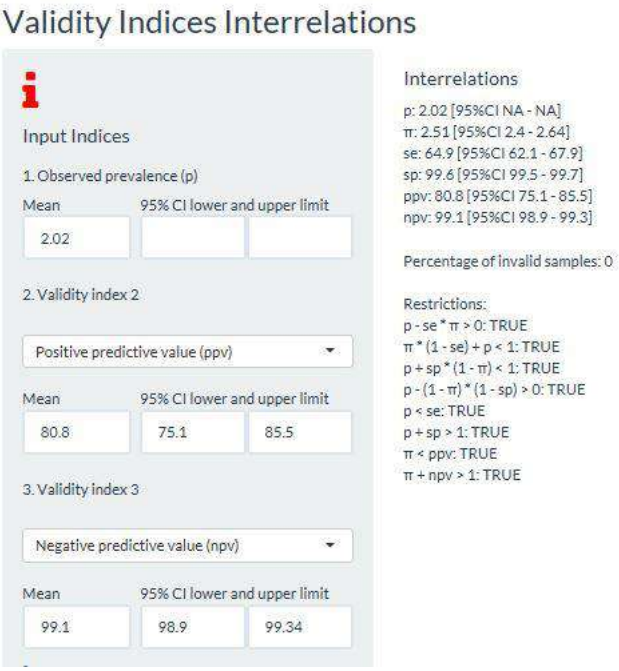


Figure 5.2: Pneumonia; deriving true prevalence, sensitivity and specificity from the observed prevalence, positive and negative predictive value.

5.3.2. Sensitivity analyses

The impact of changing the input parameters (from -1 s.e. to + 1 s.e.) on the output parameters is depicted by the vertical bars in Figures 5.3 and 4. The biases of the derived indices are expressed relative to their standard errors as well and are truncated at ±3 s.e. For example, for the input parameter combination $\pi - P - SE$ and when $\pi = 0.01$ (Figure 5.3: upper left panel), varying π from -1 s.e. to + 1 s.e, has a small impact on SE and NPV (< 1 s.e. change in both directions), but a more substantial impact on PPV (~2 s.e. change in both directions). The combined results indicate that for the scenarios investigated the estimation error of the derived parameters is smallest when using the parameter combination P – SE – PPV.

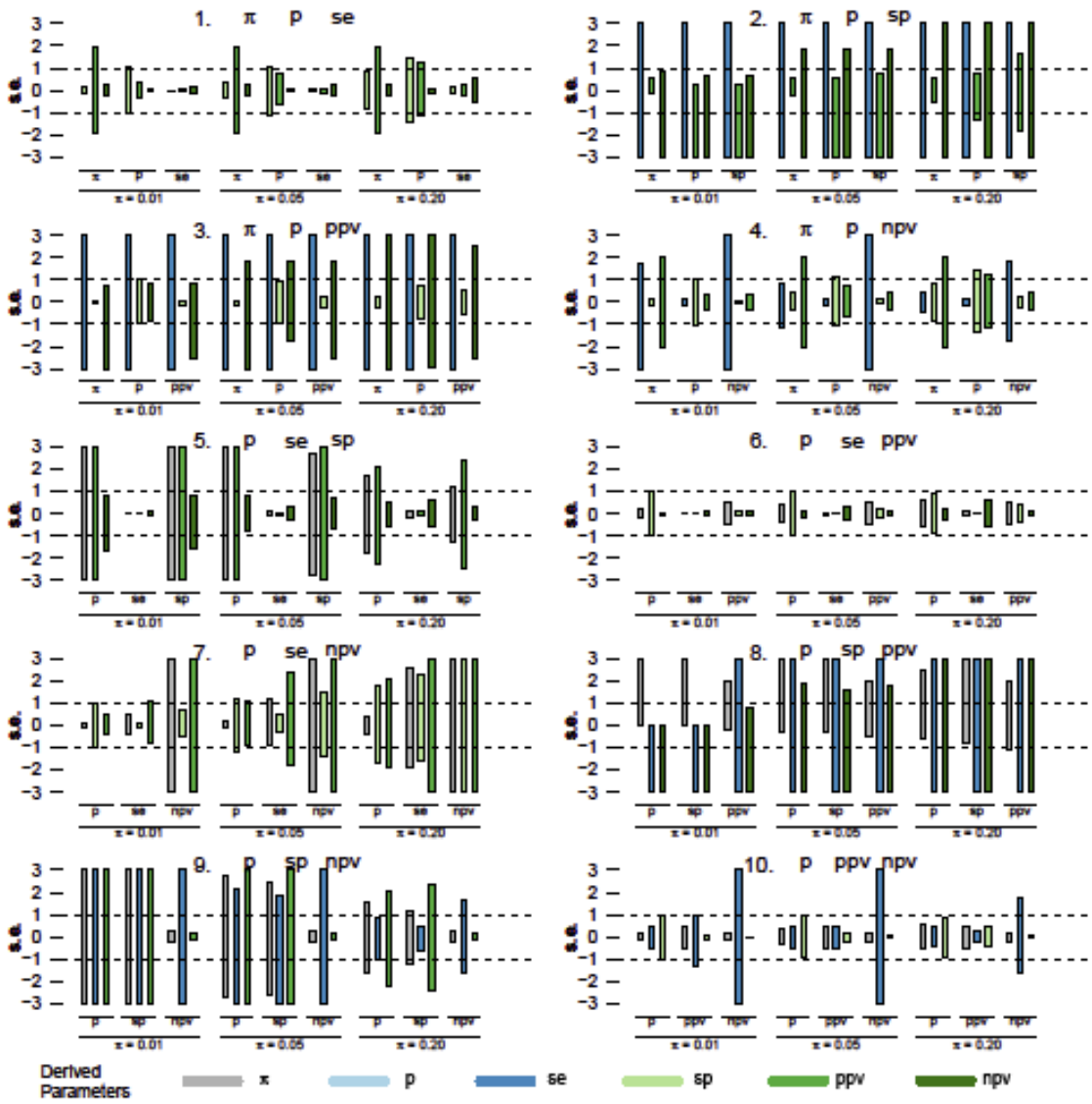


Figure 5.3: Results of the sensitivity analyses: investigating the impact of changing the input parameters from -1 to +1 standard error (s.e.) on the derived parameters for varying levels of true prevalence, $\pi = \{0.01, 0.05, 0.2\}$, SE = 0.95 and SP= 0.75. The bias of the derived indices are truncated at ±3 s.e.

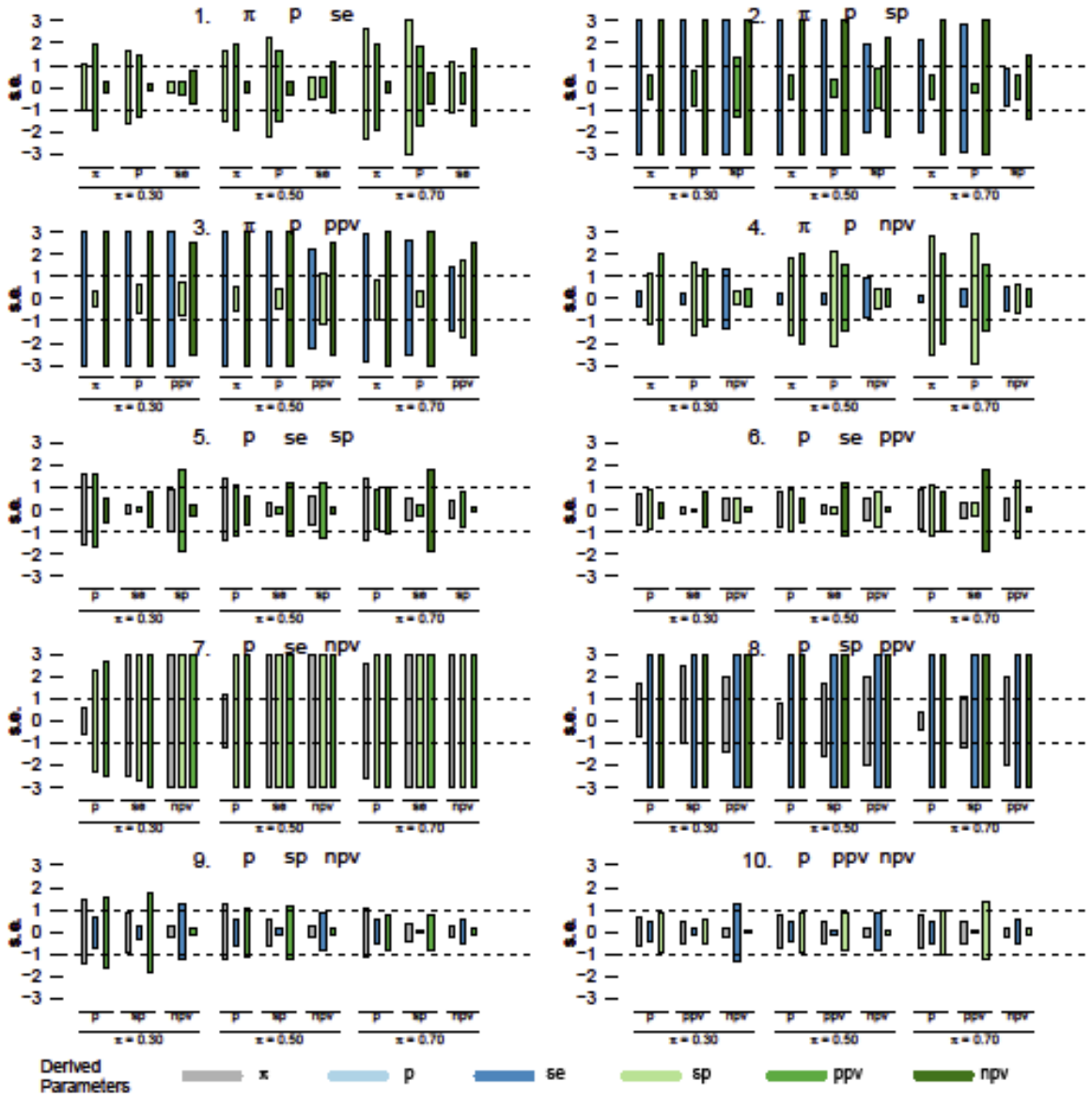


Figure 5.4: Results of the sensitivity analyses: investigating the impact of changing the input parameters from -1 to +1 standard error (s.e.) on the derived parameters for varying levels of true prevalence, $\pi = \{0.3, 0.5, 0.7\}$, SE = 0.95 and SP= 0.75. The bias of the derived indices are truncated at ± 3 s.e.

5.4. Discussion

In this work, we derived the analytical expressions to obtain validity indices or the true prevalence given an estimate of the observed prevalence and any two other parameters. To allow users to easily explore these interrelations between the observed prevalence, the true prevalence and the validity

indices, we developed a web-application, optionally allowing for the calculation of Monte Carlo 95% UIs of the derived parameters.

Our tool allows the user to easily obtain an estimate of the true prevalence given an estimate of the observed prevalence and any two other validity indices. Deriving the true prevalence from the observed prevalence, SE and SP is the well-known Rogan-Gladen estimator [7]. To our knowledge, the Rogan-Gladen estimator is the only estimator available to obtain the true prevalence from the observed prevalence using estimates of validity. We provided the other estimators as well.

The ability to convert validity indices facilitates the comparison of validation studies, which is often hampered by the use of different validity indices. Benchimol et al [4] conducted a systematic review of validation studies of CFAs and found that only 36.9% of the studies reported four or more validity indices. They found that the most common validity indices used to report the diagnostic accuracy of CFAs are SE (67.2%) and PPV (63.8%) and to a lesser extent SP (49.8%) and NPV (32.1%). Most studies that validate diagnoses in the Clinical Practice Research Database (CPRD) are restricted to assessing the proportion of CFA-positive cases that were confirmed by medical record review or responses to questionnaires, thus providing an estimate of PPV [18,19]. However, the SE, SP and NPV values were not assessed. In such cases, our tool allows one to easily derive the remaining validity indices. In the case where only one validity index is reported, the remaining validity indices can be derived from the observed prevalence only if an estimate of the true prevalence is available. The estimated true prevalence might be obtained from external data sources, such as recently performed epidemiological studies or national surveillance systems. Obviously, in this case, it is important to ensure that the external estimate applies to the database population under study.

The ability to analytically derive other validity indices from those that are reported allows the researcher to use independent validation samples (i.e. using different samples/subjects to estimate each validity index) provided the same CFA and gold standard measure were used. However, the presence of sampling error or selection bias might result in invalid parameter combinations (i.e. resulting in derived parameters outside the [0,1] range or corresponding to a CFA that performs worse than chance). To investigate the impact of estimation error in the input parameters on the derived parameters we conducted sensitivity analyses. The results show that, for the scenarios we investigated, the parameter combinations P – SE – PPV resulted in the smallest estimation errors in the derived parameters.

The assumptions applying to our analytical derivations are the same as those underlying the conventional 2 x 2-table representation of validity indices (Table 1). These assumptions are that the true disease status is truly dichotomous and the dichotomous gold standard measure reflects the true disease status without error. However, disease is not always absent or present and there might

be an underlying continuous condition (i.e. spectrum of severity) on which classification of disease status is based, varying from the clear absence to the clear presence of disease. In such cases, the SE and SP depend on the distribution of the underlying condition, and hence on the true disease prevalence [20,21]. In addition, if the gold standard measure is erroneous, the validity indices will be biased [22]. Finally, the validity of CFAs might depend on other factors such as population characteristics, access to healthcare and the completeness of the medical information contained in the database, thereby hampering the generalizability of the validity indices to populations others than those for which the validity of the CFA was initially assessed [2,18]

As many others [2,4,5], we believe that some validation of CFAs is essential to permit proper interpretation of the results obtained from healthcare database studies. The estimated validity indices might ultimately be used to obtain estimates of disease occurrence [7] or risk [6] corrected for misclassification or to adjust power calculations [23]. By providing the ability to easily obtain estimates of the true prevalence and to convert validity indices we aim to support the conduct and comparison of validation studies. In this way, we hope to help pave the way to the more widespread use of validation studies and their results.

5.5. References

1. Schneeweiss S, Avorn J.A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol 2005;**58**(4): 323-37.
2. Ehrenstein V, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. Clin Epidemiol 2016;**8**: 49-51.
3. Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. BMJ 2010;**341**: c4226.
4. Benchimol EI, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol 2011;**64**(8): 821-9.
5. Benchimol EI, et al., The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015;**12**(10): e1001885.
6. Brenner H, Gefeller O.Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. Am J Epidemiol 1993;**138**(11): 1007-15.
7. Rogan WJ, Gladen B.Estimating prevalence from the results of a screening test. Am J Epidemiol 1978;**107**(1): 71-6.
8. Altman DG, Bland JM.Diagnostic tests. 1: Sensitivity and specificity. BMJ 1994;**308**(6943): 1552.
9. Altman DG, Bland JM.Diagnostic tests 2: Predictive values. BMJ 1994;**309**(6947): 102.
10. Buckland ST. Monte Carlo confidence intervals. Biometrics 1984;**40**(3): 7.
11. R Development Core Team, R: a language and environment for statistical computing, ed. R Foundation for Statistical Computing. 2013, Vienna, Austria.
12. Chang WJC, Allaire JJ, Yihui Xie Y, McPherson J. Shiny: Web Application Framework for R. 2016.

13. Vose D. Risk Analysis, A Quantitative Guide (Third edition). 2008: John Wiley & Sons.
14. Bowman KO, Shenton LR.Estimator: Method of Moments, in Encyclopedia of statistical sciences. 1998, Wiley.
15. Saltelli A, Chan K, Scott EM, ed. Sensitivity analysis. Wiley Series in Probability and Statistics. 2000, John Wiley and Sons: New York.
16. Ducharme R, et al., Validation of diagnostic codes for intussusception and quantification of childhood intussusception incidence in Ontario, Canada: a population-based study. J Pediatr 2013;**163**(4): 1073-9 e3.
17. Aronsky D, et al. Accuracy of administrative data for identifying patients with pneumonia. Am J Med Qual 2005;**20**(6): 319-28.
18. Herrett E, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**(1): 4-14.
19. Khan, N.F., S.E. Harrison, and P.W. Rose, Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;**60**(572): e128-36.
20. Ransohoff, D.F. and A.R. Feinstein, Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;**299**(17): 926-30.
21. Brenner, H. and O. Gefeller, Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med 1997;**16**(9): 981-91.
22. Staquet, M., et al., Methodology for the assessment of new dichotomous diagnostic tests. J Chronic Dis 1981;**34**(12): 599-610.
23. Mullooly, J.P., et al., Predictive value of ICD-9-CM codes used in vaccine safety research. Methods Inf Med 2008. **47**(4): 328-35.

6. Project 5: Heterogeneity in disease misclassification: the component analysis

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Rosa Gini, Giuseppe Roberto (ARS), Kaat Bollaerts (P95), Caitlin Dodd (EMC)	23.03.2016		Research plan upon which the deliverable was based
Charlotte Switzer (SP), Sonja Banga, Daniel Weibel (EMC)	July 2016		Literature review for components on convulsions and pertussis
Jorgen Bauwens, Daniel Weibel	July 2016		Elicitation of expert opinion from clinicians
Giuseppe Roberto, Rosa Gini, Caitlin Dodd, Benus Becker (EMC)	19.07.2016		List of components & mapping
Rosa Gini	16.11.2016	0.1	First draft of deliverable
Rosa Gini,Caitlin Dodd	23.11.2016	0.2	Revision
Rosa Gini, Caitlin Dodd	18.12.2016		Instructions for databases
Rosa Gini, Caitlin Dodd	30.01.2017		Stata and R script shared with databases
Talita Duartes (SIDIAP), Elisa Martin, Consuelo Huerta (BIFAP), Gino Picelli, Lara Tramontan, Giorgia Daniele (PEDIANET), Caitlin Dodd (EMC/THIN), Miriam Sturkenboom, Daniel Weibel	1.02-08.02.2017		Coordination feedback from databases (WP5)
Rosa Gini	06.02.2017		Version submitted to scientific committee with introduction and methods complete, and results and discussion incomplete
Kaat Bollaerts (P95), Miriam Sturkenboom	07.02.2017		Comments draft deliverable
Rosa Gini	08.02.2017		Complete draft to databases and scientific committee
Talita Duartes (SIDIAP), Elisa Martin, Consuelo Huerta (BIFAP), Gino Picelli, Lara Tramontan, Giorgia Danieli (PEDIANET), Caitlin Dodd (EMC/THIN), Vincent Bauchau (GSK), Miriam Sturkenboom (EMC)	09.02.2017		coordination of data extraction (WP5)
Rosa Gini	12.02.2017		Final version
Rosa Gini	28.02.2017		Addressed SC comments

6.1. Introduction

Detecting the occurrence of diseases in persons belonging to the population documented in an ADVANCE database is a key step both in benefit studies (when the disease is the infectious disease that the vaccine is expected to prevent) and in safety studies (when the disease is an event that may occur as an unwanted consequence of the immunization).

Detection in a database may happen in an imperfect way, in both directions: persons with the disease may go unnoticed, and persons may be labeled as having the disease when in fact they do not. Misclassification introduces bias in the study results, as illustrated in the chapter (see also Project 3). As misclassification may happen differently across databases, this is one of the sources of heterogeneity across study results. Reducing misclassification has therefore a double impact: reducing bias, and hence reducing heterogeneity.

The simplest strategy to detect a disease in a database is to collect all the records of a diagnosis of that disease. This is the principal strategy adopted in electronic healthcare research. In ADVANCE, however, we are exploring a different strategy to estimate disease misclassification in databases, called *component analysis*. Using different sources of data other than diagnoses alone might alter the sensitivity and specificity of the event. For example using drugs or procedures that are used to treat/diagnose a disease, might change the validity parameters.

In the component analysis the potential for each database to detect the disease is explored by extracting new algorithms, called *component algorithms*, and by specifying a priori expectations on validity of each of them (for instance whether the component is expected to be more sensitive than the original strategy of collecting all records for a diagnosis). The interplay between the original strategy and each component, per database, sheds light on the validity of the original strategy, and allows the design of sensitivity analyses to test to which extent the observed heterogeneity may be due to differences in validity in case finding algorithms across databases. The strategy builds on previous research conducted in another project funded by the Innovative Medicines Initiative, the European Medical Information Framework, where the concept of component algorithm was developed and tested in 8 European databases for type 2 diabetes [Roberto 2016, Gini2016].

Study objective

The aim was to identify a meaningful set of components for pertussis. Moreover, in a subset of databases, we estimated the effect of using different case-finding algorithms on the resulting incidence rate of events, in children aged 0-14.

6.2. Methods

6.2.1. Specification of a component

A component is a case-finding algorithm which follows this structure: select records containing a data item belonging to a specified *data domain* and collected in a specified *setting,* provided the data item matches one of the codes or free text terms whose meaning belongs to a specified *set of concepts*. For instance a component may be: select records containing a diagnosis (*data domain*) collected in primary care (*setting*), provided the diagnosis matches one of the codes/free text terms meaning specifically 'pertussis from Bordetella', mapped in the Concept Unique Identifiers 'Bordetella pertussis infection' or 'pertussis' or 'pertussis with pneumonia' (*set of concepts*).

In this structure, the terms data domain, setting and set of concepts have the following meaning.

- The **data domain** to which the data items belong is one among diagnosis, symptoms, drug utilization, execution of a diagnostic test, results from a diagnostic test.
- The **setting** where the data items were collected is one among primary care, secondary care in outpatient setting, inpatient care, emergency care, death.
- The **set of concepts** identifies a meaning, for instance 'Symptoms of pertussis'; each concept is associated with a Concept Unique Identifiers (CUI) in the Unified Medical Language System (UMLS), that is projected to codes in the coding systems used by the databases for that data domain (for instance, ICD10, ICD9, READ,… for diagnoses), and to free text strings in the language used in the databases, if any.[1] In the case of some data domains the CUIs may not exist (for instance, for drugs); in this case, the projection to local coding systems is done manually.

6.2.2. Selection of components

Three sources of information were leveraged in order to choose pertinent components.

- A literature search, aiming at identifying algorithms used for pertussis or convulsions in previous studies; the search started from the following query in the PubMed database, including the name of the disease, the term "validation OR phenotyping", and a list of

[1]In some primary care databases (in ADVANCE: BIFAP and PEDIANET) diagnostic codes are recorded by the GP with a short string of text, which may be used to refine the search. In particular it may contain notes such as 'suspect', or 'familiarity'… the records containing a correct code but which are accompanied by a text that modifies its meaning, or without code but with a pertinent text, are interpreted by the local expert during extraction, in such a way that the semantics is preserved. The free text strings used in this context are shared by the local expert in the assessment sheet.

keywords referring to electronic databases, or to countries that often produce validation studies in electronic databases bur may not use the same keywords to refer to it, like Taiwan; in the case of convulsions, the query was too strict, to the term "validation OR phenotyping" was dropped. The queriewas

  - "pertussis AND (validation OR phenotyping) AND ("Electronic Health Records"[Mesh] OR administrative database[Title/Abstract] OR claims database[Title/Abstract] OR electronic registry[Title/Abstract] OR EHR[Title/Abstract] OR administrative data[Title/Abstract] OR claims data[Title/Abstract] OR medical records[Title/Abstract] OR PHARMO[Title/Abstract] OR CPRD[Title/Abstract] OR GPRD[Title/Abstract] OR THIN[Title/Abstract] OR IPCI[Title/Abstract] OR Mondriaan[Title/Abstract] OR SIDIAP[Title/Abstract] OR Korea[Title/Abstract] OR Canada[Title/Abstract] OR Denmark[Title/Abstract] OR France[Title/Abstract] OR Italy[Title/Abstract] OR Japan[Title/Abstract] OR Taiwan[Title/Abstract] OR Australia[Title/Abstract])"

- A questionnaire to the databases, to ask for their previous experience or advice (see Appendix P5).
- The web page of the European Center for Disease Prevention and Control (ECDC) was searched for information the typical natural history of the diseases (symptoms, diagnostic tests, typical setting of diagnosis and treatment, pharmaceutical treatment, outcome).

During selection the CodeMapper tool (see Project 9) was used to associate CUIs to concepts, and to project CUIs to all the coding systems.

6.2.3. Collection of a-priori knowledge

Pertussis is a compulsory notifiable infectious disease in all the countries involved in ADVANCE, whose national surveillance systems provide estimates of background rates of this event in the corresponding populations. Those estimates are available from the ECDC web site [ECDC 2014], or from national/regional surveillance centres.

The evaluation questionnaire of the fingerprinting was used to collect a priori knowledge from databases, see Deliverable 5.2 and Deliverable 5.4. In the feasibility assessment sheet for databases conducted as part of the POC-1 in WP5, database owners were asked to describe the settings from which they collect diagnosis, database owners were asked to describe the settings from which they collect diagnosis, whether they collected tests, laboratory results, dispensings and/or prescriptions of drugs, whether they foresee having missing data, what is the background incidence rate of pertussis they can obtain from literature and/or surveillance systems in their geographic catchment

area. This data was compared with the fingerprint data of the POC and a feedback on revealed discrepancies was asked (see Deliverable 5.2 and Deliverable 5.4).

The literature review provided information about sensitivity and predictive value of some components in databases external to the ADVANCE consortium.

6.2.4. Data-extraction

Database owners received a document describing the extractions they were asked to perform (the document is available at the link indicated under the label [Instructions] at the end of this section). Dependent on the structure of the database, only some components were requested. They were asked to append the results of all the component algorithms to a single table with four columns: identifier of the patient, date of the record, label of the component, label of the event.

6.2.5. Local data processing

Databases were asked to save the dataset of components (Event file) in the same directory as the population file that was used for the POC-fingerprinting. The population file comprises for each person, the entry and exit date, birthdate and gender. Databases were asked to run a tailored component analysis script in Stata or in R, according to their preference.

Two index dates (1st September 2012 and 1st September 2014) were chosen to allow assessment of changes over time, each with a follow-up of 365 days and a look-back of 720 days. The dates were chosen at the beginning of autumn, rather than at New Year's day, to avoid splitting the winter season.

In each index date, each person in the population file was retained if they had all the look-back data available or were born during look-back. For each component two binary variables were computed: one was positive if the person was positive for the component during look-back, the other was positive if the person was positive for the component during follow-up. For instance if subject 1 and subject 2 are included in the Event file with the following components as given Table 6.1, the dataset of binary variables is as given in Table 6.2.

The dataset of binary variables was then linked to the gender and the age band at index date for the person, and aggregated by age band, gender, and per index date. The resulting aggregated dataset was shared with the investigators responsible for the next steps of the analysis.
The procedure had been double-coded in Stata and R by ARS, and had been tested on simulated data to ensure consistency. The scripts were parameterized, for future reuse, and are available at the link indicated in the references with [Procedure R] and [Procedure Stata].

Table 6.1: illustration of data processing, input file

PatientID	Date	Eventtype	Code
1	20110302	PERTUSSIS	PERTUSSIS_TEST
1	201103012	PERTUSSIS	PERTUSSIS_ DIAG_PC
2	20130501	PERTUSSIS	PERTUSSIS_TEST
2	20130507	PERTUSSIS	PERTUSSIS_DIAG_PC

Table 6.2: Illustration of data processing, output file

PatientID	IndexDate	PERTUSSIS_DIAG_PC_bef	PERTUSSIS_DIAG_PC_fup	PERTUSSIS_TEST_fup
1	20120901	1	0	0
2	20120901	0	1	1
1	20140901	0	0	0
2	20140901	1	0	0

6.2.6. Analysis

The study population comprised children aged 0-14 at index date.

The components were labelled as 'inclusion', 'exclusion', or 'refinement' criteria. According to a-priori knowledge, additional components were created as combination of inclusion and exclusion criteria, or of inclusion and refinement criteria. The resulting list of components entered the analysis.
In each database the incidence of each component was computed as number of persons positive during follow-up divided by the number of people in the population at index date.

6.3. Results

6.3.1. Selected components: results from literature review

After refinement of the Pubmed search, 24 papers contained information relevant to define a component. All settings were involved in at least one study. In some studies a broad algorithm, comprising symptoms such as cyanosis or vomiting after coughing, was refined using positive results from laboratory tests. The complete results from the literature review are contained in Appendix P5.

6.3.2. Selected components for pertussis

The components selected were

- Five from the diagnosis domain, one per setting (primary care, secondary outpatient care, inpatient care, emergency care, death), using a set of concepts (indicated by the label (Pertussis) in Table 6.4) which specifically refer to Bordetella pertussis
- Four from the diagnosis domain or symptoms domain, one per setting (death excluded), using a set of concepts concepts (indicated by the label (Symptoms of pertussis) in Table 6.4) which refer to diagnosis of pertussis with no indication of the organism causing the symptoms, or to symptoms that, according to the literature review, are predictive of pertussis; this set of concepts also can be extracted from the data domain of symptoms and signs
- One from the domain of execution of diagnostic tests, using both generic and specific concepts concepts (indicated by the label (Pertussis test) in Table 6.4)
- One from the domain of results from diagnostic tests (concepts are indicated by the label (Positive results from a pertussis test) in Table 6.4)
- Two from the domain of drug utilization, both referring to the same class of antibiotics (indicated by the label (macrolides) in Table 6.4), but one referring to dispensings and the other to prescriptions

The full list of component algorithms for pertussis is in Table 6.3 and the set of concepts referred to by the label in round parentheses can be found in Table 6.4.

Table 6.3: Components for pertussis.

Name	Description	Rule to identify records
PERTUSSIS_DIAG_PC	Diagnosis of pertussis recorded during primary care	Records recorded during primary care and containing one of the codes (Pertussis) in a field where diagnoses are collected
PERTUSSIS_DIAG_SC	Diagnosis of pertussis recorded during secondary care	Records recorded during secondary (specialist) outpatient care and containing one of the codes (Pertussis) in a field where diagnoses are collected
PERTUSSIS_DIAG_INP	Diagnosis of pertussis recorded during inpatient care	Records recorded during inpatient care and containing one of the codes (Pertussis) in a field where diagnoses are collected
PERTUSSIS_DIAG_EC	Diagnosis of pertussis recorded during emergency care	Records recorded during emergency care and containing one of the codes (Pertussis) in a field where diagnoses are collected
PERTUSSIS_DIAG_DEATH	Diagnosis of pertussis recorded as a cause of death	Records recorded at death and containing one of the codes (Pertussis) in a field where diagnoses are collected
PERTUSSIS_SYMPT_PC	Symptoms of pertussis recorded during primary care	Records recorded during primary care and containing one of the codes or strings (Symptoms of pertussis) in a field where diagnoses, symptoms or signs are collected
PERTUSSIS_SYMPT_SC	Symptoms of pertussis recorded during secondary care	Records recorded during secondary (specialist) outpatient and containing one of the codes or strings (Symptoms of pertussis) in a field

		where diagnoses, symptoms or signs are collected
PERTUSSIS_SYMPT_INP	Symptoms of pertussis recorded during inpatient care	Records recorded during inpatient and containing one of the codes or strings (Symptoms of pertussis) in a field where diagnoses, symptoms or signs are collected
PERTUSSIS_SYMPT_EC	Symptoms of pertussis recorded during emergency care	Records recorded during emergency care and containing one of the codes or strings (Symptoms of pertussis) in a field where diagnoses, symptoms or signs are collected
PERTUSSIS_LAB	Positive results from a laboratory analysis regarding pertussis	Records containing one of the codes (Pertussis test) in a field containing description of a laboratory test, and containing one of the codes or strings (Positive result from a pertussis test) in the corresponding result field
PERTUSSIS_TEST	Laboratory analysis regarding pertossi	Records containing one of the codes (Pertussis test) in a field containing description of a laboratory test,
PERTUSSIS_DRUG_DISP	Dispensings of macrolides	Records recorded in a facility dispensing medications in the community and containing one of the codes (Macrolides) in a field where drug codes are collected
PERTUSSIS_DRUG_PRESC	Prescriptions of macrolides	Records recorded during primary care and containing one of the codes (Macrolides) in a field where drug codes are collected
PERTUSSIS_COMMENT_PC	Diagnosis of pertussis recorded during primary care in the comments field	Records recorded during primary care and containing one of the free text keywords in (Pertussis) in a field where comments are collected

Table 6.4: Set of concepts used in the components of pertussis. The projection on the local coding systems and free text keyword is omitted in this table and included in the complete version in the Appendix P5.

Group of concepts	Description of each included concept	Concept Unique	Included in the POC out
(Pertussis)	Whooping cough due to Bordetella pertussis without pneumonia	C2887068	YES
	Whooping cough due to Bordetella pertussis with pneumonia	C2887069	YES
	Pneumonia in pertussis	C0155865	YES
	Pertussis	C0043167	YES
	Bordetella Infections	C0006015	YES
(Symptoms of pertussis)	Whooping cough - other specified organism	C0343484	YES
	paroxysms of coughing	C0231911	
	Apnea	C0003578	
	Whooping cough due to unspecified organism	C0043168	YES
	cyanosis	C0010520	
	Post-tussive vomiting	C1740793	
	Infection due to Bordetella parapertussis (disorder)	C0275742	YES
	Whooping cough due to other Bordetella species	C0348128	YES
	Whooping cough-like syndrome	C0343485	YES
(Macrolides)	Macrolides		
(Pertussis test)	polymerase chain reaction test		
	culture or serology of Bordetella pertussis		
	isolation of Bordetella pertussis from a clinical specimen		
(Positive result from a pertussis test)	positive polymerase chain reaction test		
	culture or serology of Bordetella pertussis		
	positive isolation of Bordetella pertussis from a clinical specimen		

6.3.3. Collecting a-priori knowledge: settings and semantics in databases

Assessment sheets were available from 7 databases: THIN, BIFAP, PEDIANET, RCGP, SSI, AUH and SIDIAP. For general information on the databases, see Deliverable 5.2.

The majority of databases (THIN, BIFAP, PEDIANET, RCGP) only collect diagnoses from the primary care setting. Even though referrals from secondary outpatient care, inpatient care, emergency care, and causes of death may be recorded by the general practitioner, no automatic mechanism is in place. In SSI and AUH diagnoses are available from secondary outpatient care, inpatient care or emergency care. In SIDIAP diagnoses from primary care are available in the whole population, but a specific subpopulation has diagnoses available from all the settings, except causes of death. Drugs are available in all databases (in the form of dispensing and /or prescription), except in SSI. Prescriptions of diagnostic tests are available in all the databases, but results are not available in AUH, and data is deemed to be fairly incomplete in all databases.

6.3.4. Collecting a-priori knowledge: evidence on background rates for pertussis

From the report [ECDC 2014] the background incidence rates of pertussis from surveillance systems in children 0-14 in 2012 in Europe was around 20 per 100,000.

6.3.5. Data extraction and data processing

Extraction was performed in time for inclusion in this deliverable by 4 databases: THIN, PEDIANET, BIFAP, and SIDIAP.

Components available were diagnosis and symptoms from primary care, and utilization of macrolides, in all databases. In THIN and PEDIANET also tests and their results were available. As assessed from a priori knowledge (see previous subsection "collecting a priori knowledge"), all the databases which could provide data are primary care medical records, and diagnosis from outpatient secondary care, inpatient care, emergency care, or death, were not available, In SIDIAP data from other settings would be available, in principle, on a subpopulation, but this data could not be extracted in time for this deliverable.

Data processing was performed using the Stata script in all databases, except in PEDIANET which used R.

6.3.6. Analysis

In total, 3,841,957 person years of children aged 0-14 at index date entered the analysis, 1,287,406 from THIN, 1,719,770 from SIDIAP, 759,674 from BIFAP, and 75,107 from PEDIANET. The IR of

events detected by the diagnostic component was 4, 5, 14, and 1 per 100,000 PY. Among events detected by the symptoms component 35.5%, 82.0%, 40.8% and 50.8% were also detected by one of the macrolides utilization components.

Adding to the events in the diagnostic component only the children who were detected both by the symptoms and macrolides utilization components would increase the IR to 11, 73, 20, and 43 per 100,000 PY.

Records of laboratory confirmations were very scarce in the study period: in PEDIANET there were none, in THIN there were just 2. Records of test prescriptions were 35 in PEDIANET and 235 in THIN. Among events detected by the symptoms component, 6.6% were detected also by the test component in PEDIANET and 5.7% in THIN; the single event detected as a diagnosis had a test in PEDIANET, and 25% of the diagnostic events in THIN had also a record of a test.

The results are represented in more detail in Table 6.6.

Table 6.6. Component analysis. The labels of the components refer to Table 1, and are all extracted in the setting of primary care (PC in Table 1). The component DRUG was computed as either a dispensing or a prescription.

	THIN	SIDIAP	BIFAP	PEDIANET
Person-years	1,287,406	1,719,770	759,674	75,107
DIAG	52	93	110	1
IR (per 100,000)	4.0	5.4	14.5	1.3
SYMPT	244	1426	98	61
IR (per 100,000)	19.0	82.9	12.9	81.2
DIAG OR SYMPT	295	1519	206	62
IR (per 100,000)	22.9	88.3	27.1	82.5
SYMPT AND DRUG	94	1169	40	31
% of SYMPT	38.5	82.0	40.8	50.8
DIAG OR (SYMPT AND DRUG)	143	1262	150	32
IR (per 100,000)	11.1	73.4	19.7	42.6
TEST	236			35
TEST AND DIAG	13			1
% of DIAG	25.0			100.0
TEST AND SYMP	14			4
% of SYMPT	5.7			6.6
LAB	2			0

6.4. Discussion

The component algorithms selected for pertussis from expert knowledge and literature review were 14. The majority of algorithms meant to be used as inclusion criteria (9) belonged to the domain of diagnosis and of symptoms, and differed one from the other either for semantics aspects, or for the setting where the information was collected. Test, laboratory values, drug utilization, and additional diagnostic components were meant to be used mainly as refinement or exclusion criteria.

The preliminary results from the pertussis exercise provided evidence that the sensitivity of the most specific component (DIAG) is heterogeneous across data sources, and that alternative, less specific algorithms are available, which identify a larger set of events, especially in the databases were DIAG appears to be less sensitive.

Expectations on components for pertussis

In the case of pertussis the components were meant to allow exploring strategies in the direction of higher sensitivity, while controlling positive predictive value (PPV).
There were two families of diagnostic components, each referring to a separate group of concepts. The concepts grouped under the label (Pertussis) contained concepts that were explicitly mentioning Bordetella pertussis.
The concepts grouped under the label (Symptoms of pertussis) referred both to diagnoses compatible with symptoms of Bordetella pertussis (such as "Whooping cough due to an unspecified organism") and to diagnosis that excluded Bordetella pertussis but implied pertussis symptoms (such as "Whooping cough due to other Bordetella species"), as well as to symptoms that may be due to pertussis (like "whooping cough" or "cyanosis"). The rationale behind this group of concepts was that when a physician recorded one of the diagnoses in the group, the fact that Bordetella was not explicitly mentioned implied that the physician was not taking responsibility for attributing the symptoms to a Bordetella infection. We therefore considered one of those diagnoses as the indication that symptoms were observed. Moreover, we included those set of diagnoses symptoms that were found in the literature to be highly predictive of Bordetella pertussis in young children. As a consequence, the group of concepts (Symptoms of pertussis) was expected to detect cases with lower PPV with respect to the group of concepts (Pertussis). However we expected this component to provide an added value, especially in primary care databases: some children are observed in primary care and transferred to an inpatient setting for further diagnostic investigations and treatment, and the specific diagnosis is then made in hospital. The (Pertussis) set of concepts alone is therefore expected to miss some cases in primary care databases. The extent of this loss is expected to be heterogeneous across databases, due to complementary characteristics of the databases. For instance in PEDIANET, a database where diagnoses were mostly recorded in free text during the study periods, the word ‘pertosse’ (pertussis, in Italian) was included in the (Symptoms of pertussis) set of concepts, unless the word 'Bordetella' was found as well. However, the word in natural language contains an intrinsic ambiguity: indeed it may be synonymous of ‘Bordetella pertussis’. Therefore the sensitivity of DIAG was expected to be lower, and the PPV of SYMPT to be higher, in PEDIANET than in the other databases.
In the case of databases collecting diagnoses from inpatient setting we had the expectation that the component based on (Symptoms of pertussis) captured only unspecific diagnoses, and not symptoms, because in inpatient setting symptoms are rarely coded in discharge records: so in

databases with inpatient diagnoses this component may capture those children who were first admitted to hospital and then obtain the result of their tests after discharge. It would be very valuable to observe the interaction of the components in a database which can access both inpatient and outpatient secondary or primary care diagnoses, to estimate the extent to which the settings contribute to a complete diagnosis of the case. Unfortunately, it was not possible to obtain this data in time for this deliverable.
To tame the loss of PPV in the components using (Symptoms of pertussis), we planned to use concurrent utilization of macrolides as a refinement criterion, because, due to the indications of those drugs, their utilization is expected to be associated with the fact that the origin of the symptoms was a Bordetella infection.

The component based on positive results from microbiological tests have the potential to be a very valuable source, because this result is direct evidence of a confirmed case, however its sensitivity was expected to be low, on the basis of previous experience from databases.
The presence of a test is in principle a confirmation for cases detected with the concepts (Pertussis) and an exclusion criterion for cases detected only with the (Symptoms of pertussis) concepts, because if a test was performed and a diagnosis of Bordetella pertussis was not recorded, this is evidence that the result of the test was negative. However it must be noted that by the time the result of the test is available, the setting of care may have changed (from inpatient to outpatient, or vice versa), so the result of the test may not have a chance inform a recording in the same database where the test prescription was recorded.

Evidence from preliminary component analysis on pertussis

The databases which could provide data in time for this deliverable were all primary care databases, so it was not possible to observe other settings. The component analysis supported the expectation that sensitivity and PPV of the DIAG strategy are imperfect, and heterogeneous across databases.

The more specific component, DIAG, as expected, yielded an IR which was lower than the background rate of 20 per 100,000. The background rate is an average of European countries, and some variation is expected in local populations. However the two Spanish databases show an IR of 14 in BIFAP and of 5 in SIDIAP: this supports the expectation that DIAG alone is losing many cases, at least in SIDIAP. Adding SYMPT may yield a more sensitive strategy, probably at the expenses of PPV. As we expected, PEDIASNET had a very low IR of DIAG, but the PPV of SYMPT is probably higher than in other databases, because of the ambiguity in the Italian term 'pertosse' included in the SYMPT component.
The planned compromise to increase sensitivity while containing the loss of PPV was to add to the diagnostic component only events detected *both* by symptoms and macrolides utilization. This

strategy resulted in an IR that is more compatible than DIAG with the expected background rate in THIN, BIFAP, and PEDIANET. In SIDIAP it yielded an IR which still exceeds the expected IR, but in a complementary way with respect to the observed IR id DIAG alone. This algorithm can be considered for a sensitivity analysis, possibly in a database-tailored fashion, in case heterogeneity of results from the study is observed, to test the possibility that it is due to heterogeneous sensitivity and PPV of the outcome definition.

Recordings of positive results from laboratory tests cannot be used to detect confirmed cases in the databases which entered this preliminary study, because, as expected from the previous experience collected, this information resulted to be very rarely recorded, even when it was in theory available in the database.

Developments

In future work, the results of the component analyses will be analysed using Latent Class Modelling allowing to estimate indices of validity. The work on Latent Class Modelling is initiated in Project 6.

6.5. Conclusion

The heterogeneity across databases in detecting events relevant to estimate benefit and risk for vaccine in multi-database studies can be summarized using the following categories: data domain, setting of data collection, and semantics of the recording.

A component analysis, taking into account such categories, is recommended before finalizing the protocol of the main study, to design sensitivity analyses aimed at assessing whether possible heterogeneity in results can be attributed to heterogeneity in outcome misclassification.

If quantitative estimation of misclassification is obtained, adjustment for validity indices must be included in the statistical analysis plan of the main study.

6.6. References

[Drive Pertussis] Spreadsheet for Pertussis. Available at this link https://docs.google.com/spreadsheets/d/1wI_gLExO0aBbxVcyPoJ-imuN3xW6-PqFSYXCOwdcYAM

ECDC. Surveillance report. Annual epidemiological report. Vaccine-preventable diseases. 2014. Available at this link http://ecdc.europa.eu/en/publications/Publications/AER-2014-VPD-FINAL.pdf

Gini R. A Validation Odyssey: from big data to local intelligence. Erasmus University Rotterdam; 2016. Available from: http://hdl.handle.net/1765/93461

[Instructions] ADVANCE - Instructions to extract components for component analysis of PERTUSSIS and CONVULSIONS. Available at this link: https://drive.google.com/open?id=0B59Sfie-3laCbHk2dExVd0g5VWs

[Procedure R] R procedure to obtain aggregated data from event and population file. Available with simulated data at this link. https://drive.google.com/open?id=0B59Sfie-3laCMnF6RDh2SDloWDA

[Procedure Stata] Stata procedure to obtain aggregated data from event and population file. Available with simulated data at this link. https://drive.google.com/open?id=0B59Sfie-3laCandrQ2x3OHV3NGM

Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, et al. Identifying cases of type 2 diabetes in heterogeneous data sources: strategy from the EMIF project. PLoS ONE 2016;11(8):e0160648.

7. Project 6: Latent Class models to estimate validity of case-finding algorithms when there is no reference standard

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts, Rosa Gini	23 March 2016	-	Research plan upon which this work is based
Kaat Bollaerts, Nicolas Praet, Tom De Smedt, Alexandros Rekkas	13 July 2016	-	Face2Face meeting at GSK to discuss LCM analyses and sharing of references
Alexandros Rekkas, Kaat Bollaerts	July/Aug 2016	-	Implementing analyses, developing R functions
Alexandros Rekkas	23 Aug 2016	-	Presenting results at P95, sharing with Rosa and Caitlin
Kaat Bollaerts	02 Feb 2016	-	First draft
Nicolas Praet, Rosa Gini	12-Feb 2016		Review
Kaat Bollaerts	13 Feb 2017	-	Comments incorporated

7.1. Introduction

Validation is recognized as an important component of research with electronic healthcare records (EHR). Studies reporting on the validity of exposure- or disease case-finding algorithms (CFAs) are being increasingly performed [1] and calls for more research on validation of CFAs in EHR research have been recently launched [2, 3].

Validation of CFA is typically obtained by comparing the CFA results with those from a reference standard. Ideally, the reference standard should be error-free ('gold' standard), but in reality this is often not the case and serious bias in estimated validity might results if the reference test is wrongly assumed to be error-free [4, 5]. By comparing the CFA with the reference standard, validity indices such as sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), likelihood ratios and raters agreement statistics are estimated [6-8] (See also Project 3) .
In electronic healthcare research, a reference standard is often obtained through chart review or through contacting healthcare professionals asking them to complete questionnaires [9]. However, this is not always feasible and often very expensive and time consuming.

An alternative approach using Latent Class Modelling [10, 11] for which no reference standard is needed has been frequently applied to validate diagnostic tests, particularly in veterinary science [12, 13]. A Latent Class Model (LCM) treats the true disease status as an unmeasured (aka latent) categorical variable (typically binary) whereas the observed measurements of disease are considered as imperfect classifiers of the true disease status. To our knowledge, LCMs have not yet been applied to estimate validity of CFAs from EHR. In this work, we give an overview of commonly used LCM methodology and illustrate their performance through simulation. In a next step, this work will be integrated with the component analyses work (see Project 5) to explore whether LCMs can be successfully used to estimate validity of CFAs from EHR.

7.2. Methods

To be able to use Latent Class Modelling, several disease classifiers (here: CFAs) need to be applied to the same subjects. The CFA results can then be aggregated as in Table 7.1. It is necessary to specify a LCM that is identifiable, i.e. the number of freely estimated parameters does not exceed the number of unique patterns of disease classification.

Table 7.1: Table shell: aggregated results from applying different case-finding algorithms to the same database.

	CFA 1	CFA 2	CFA 3	Nr of subjects
Pattern 1	-	-	-	N_{---}
Pattern 2	+	-	-	N_{+--}
Pattern 3	-	+	-	N_{-+-}
Pattern 4	-	-	+	N_{--+}
Pattern 5	+	+	-	N_{++-}
Pattern 6	+	-	+	N_{+-+}
Pattern 7	-	+	+	N_{-++}
Pattern 8	+	+	+	N_{+++}

Two commonly used LCMs are the Hui-Walter and the Walter-Irwig models. The Hui-Walter model (also the '2 tests-2 populations' model) allows the estimation of SE and SP of two tests as well as the true prevalences when the tests are applied to two different populations with a different true disease prevalence, under the assumption that the SE and SP are the same in both populations and under the assumption of conditional independence [10]. The extension to >2 tests and >2 populations is straightforward. The Walter-Irwig model (also the '3 tests-1 population' model) allows the estimation of the SE and SP of three tests applied to the same population [11].

The assumption of conditional independence means that, conditional on the true (binary) disease status, the test results are independent. In other words, the test results only depend on the true disease status and a misclassification error for one test does not increase or decrease the probability of misclassification for the other test. This assumption is easily falsified [14]. For instance, conditional dependence among test results could arise if individuals with less severe disease are more likely to be missed by different tests compared with individuals with severe disease. However, when relaxing the assumption of conditional independence, the LCMs are no longer identifiable. This issue of non-identifiability can be solved by imposing additional constraints, coming either from external sources or expert opinion. Bayesian approaches have been suggested to incorporate this external information through the specification of prior distributions on the parameters of the LCM [15]. Berkvens et al proposed a re-parametrization of the conditional dependence model to facilitate the expert opinion elicitation [16].

7.3. Illustration

7.3.1. Date generation

For both the '2 tests x 2populations' and '3 tests x 1 population' LCMs assuming both conditional independence and conditional dependence, we generated data with the specifications as summarized

in Table 7.2. For the purposes of illustration, we give the simulated data for '3 tests x 1 population' assuming conditional independence in Table 7.3. These data are then used to estimate the LCMs.

Table 7.2: Data generation specifications

Model	Specifications
3 tests-1 population,	N=10,000, $\pi = 0.30$,
conditional independence	SE_{test1} = 0.90, SP_{test1} = 0.99,
	SE_{test2} = 0.70, SP_{test2} = 0.85,
	SE_{test3} = 0.50, SP_{test3} = 0.85
3 tests-1 population,	N=10,000, $\pi = 0.30$,
conditional dependence	SE_{test1} = 0.60, SP_{test1} = 0.80,
	SE_{test2} = 0.60, SP_{test2} = 0.99,
	($SE_{test2\|test1=0}$ = 0.30, $SE_{test2\|test1=1}$ = 0.80)
	($SP_{test2\|test1=0}$ = 0.97, $SP_{test2\|test1=1}$ = 1)
	SE_{test3} = 0.70, SP_{test3} = 0.80
	($SE_{test3\|test1=0\&test2=0}$ = 0.77, $SP_{test3\|test1=0\&test2=0}$ = 0.80)
	($SE_{test3\|test1=1\&test2=0}$ = 0.60, $SP_{test3\|test1=1\&test2=0}$ = 0.86)
	($SE_{test3\|test1=0\&test2=1}$ = 0.61, $SP_{test3\|test1=0\&test2=1}$ = 0.80)
	($SE_{test3\|test1=0\&test2=1}$ = 0.66, $SP_{test3\|test1=0\&test2=1}$ = 0.81)
2 tests-2 populations,	N_{pop1}=10,000, N_{pop2}=8000,
conditional independence	π_{pop1}=0.70, π_{pop2}=0.30,
	SE_{test1} = 0.80, SP_{test1} = 0.80,
	SE_{test2} = 0.70, SP_{test2} = 0.99
2 tests-2 populations,	N_{pop1}=10,000, N_{pop2}=8000,
conditional dependence	π_{pop1}=0.70, π_{pop2}=0.20,
	SE_{test1} = 0.90, SP_{test1} = 0.99,
	SE_{test2} = 0.80, SP_{test2} = 0.90,
	($SE_{test2\|test1=0}$ = 0.79, $SP_{test2\|test1=0}$ = 0.91)
	($SE_{test2\|test1=1}$ = 0.87, $SP_{test2\|test1=1}$ = 0.35)

Table 7.3:3 tests-1 population (conditional independence): example data.

Pattern	Nr of subjects
+ + +	981
+ + -	910
+ - +	423
- + +	258
- - +	888
- + -	975
+ - -	437
- - -	5126
total	10,000

7.3.2. Estimation

We have implemented '2 tests x 2populations' and '3 tests x 1 population' LCMs assuming both conditional independence and conditional dependence with the re-parametrization proposed by Berkvens et al [16] using the statistical software R 3.3.1 [17]. To estimate the LCMs, we used a Bayesian approach with a multinomial likelihood [18]. For the conditional independent models, we used non-informative prior information as they are identifiable whereas for the conditional dependent models, we used informative prior information that was not mis-specified (i.e. in line with the parameters used to generate the data).

7.3.3. Results

Tables 7.4 to 7.7 show the estimated parameters from the four different LCMs. All estimates from all models are very accurate with the accuracy being slightly worse (i.e. larger deviations between simulated and estimated values as well as larger standard errors) when assuming conditional dependence compared to conditional independence. These results are all as expected.

Table 7.4:3 tests-1 population (conditional independence): parameter estimates

Parameter	Simulated	Estimate (s.d.)
π	0.30	0.29 (0.01)
SE_{test1}	0.90	0.93 (0.02)
SP_{test1}	0.99	0.99 (0.00)
SE_{test2}	0.70	0.68 (0.01)
SP_{test2}	0.85	0.84 (0.01)
SE_{test3}	0.50	0.50 (0.01)
SP_{test3}	0.85	0.84 (0.01)

Table 7.5:3 tests-1 population (conditional dependence): parameter estimates

Parameter	Simulated	Estimate (s.d.)
π	0.30	0.29 (0.04)
SE_{test1}	0.60	0.56 (0.06)
SP_{test1}	0.80	0.78 (0.03)
SE_{test2}	0.60	0.56 (0.06)
SP_{test2}	0.99	0.96 (0.02)
SE_{test3}	0.70	0.75 (0.05)
SP_{test3}	0.80	0.81 (0.03)

Table 7.7:2 tests-2 populations (conditional dependence): parameter estimates

Parameter	Simulated	Estimate (s.d.)
π_{pop1}	0.70	0.71 (0.01)
π_{pop2}	0.30	0.31 (0.01)
SE_{test1}	0.80	0.80 (0.01)
SP_{test1}	0.80	0.80 (0.01)
SE_{test2}	0.70	0.70 (0.01)
SP_{test2}	0.99	1.00 (0.01)

Table 7.6:2 tests-2 populations (conditional independence): parameter estimates

Parameter	Simulated	Estimate (s.d.)
π_{pop1}	0.70	0.67 (0.06)
π_{pop2}	0.20	0.14 (0.05)
SE_{test1}	0.90	0.91 (0.05)
SP_{test1}	0.99	0.94 (0.05)
SE_{test2}	0.80	0.80 (0.04)
SP_{test2}	0.90	0.86 (0.04)

7.4. Discussion

The need for validation of CFAs in EHR research is generally acknowledged and calls for more research in this area have been recently launched [2,3]. In this work, we introduced Latent Class Modelling as a possible alternative to estimate validity of case-finding algorithms used in EHR research. The obvious advantage of this approach is that there is no need to conduct a validation study, which is often expensive, time-consuming or even not feasible. To our knowledge, LCM have not yet been used to validate CFA although they have been widely used in other areas [13].

We have implemented '2 tests x 2populations' and '3 tests x 1 population' LCMs assuming both conditional independence and conditional dependence with the re-parametrization proposed by Berkvens [16] in R 3.3.1 [17]. To test our code and illustrate the performance of Latent Class Modelling, we generated data to which we subsequently fitted the LCM models. The estimated parameters were almost exactly the same as the parameters we used to generate the data.

In next steps, we will apply LCMs to EHR and investigate whether LCMs can be built with plausible assumptions accurately reflecting the nature of EHR. Currently, we believe the LCM with conditional dependence are the most appropriate for EHR and we will solicit expert opinion to inform the parameters of the model. Additionally, we are investigating the robustness of LCMs against assumption violations (i.e. the assumption of conditional independence and misspecification of the prior information).

7.5. References

1. Benchimol EI, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. J Clin Epidemiol 2011;**64**(8): 821-9.
2. Ehrenstein V, et al., Helping everyone do better: a call for validation studies of routinely recorded health data. Clin Epidemiol 2016;**8**: 49-51.
3. Lash TL, Olshan AF. EPIDEMIOLOGY Announces the "Validation Study" Submission Category. Epidemiology 2016;**27**(5): 613-4.
4. Gart JJ, Buck AA. Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. Am J Epidemiol, 1966;**83**(3): 593-602.
5. Staquet M, et al. Methodology for the assessment of new dichotomous diagnostic tests. J Chronic Dis 1981;**34**(12): 599-610.
6. Altman DG, Bland JM.Diagnostic tests 2: Predictive values. BMJ 1994;**309**(6947): 102.
7. Altman DG, Bland JM.Diagnostic tests. 1: Sensitivity and specificity. BMJ 1994;**308**(6943): 1552.
8. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004;**329**(7458): 168-9.
9. Herrett E, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**(1): 4-14.
10. Hui SL, Walter SD.Estimating the error rates of diagnostic tests. Biometrics 1980;**36**(1): 167-71.
11. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol 1988;**41**(9): 923-37.
12. Enoe C, Georgiadis MP, Johnson WO.Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev Vet Med 2000;**45**(1-2): 61-81.
13. van Smeden M, et al. Latent class models in diagnostic studies when there is no reference standard--a systematic review. Am J Epidemiol 2014;**179**(4): 423-31.
14. Toft N, Jorgensen E,Hojsgaard S.Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev Vet Med 2005;**68**(1):19-33.
15. Branscum AJ, Gardner IA, JohnsonWO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev Vet Med 2005;**68**(2-4): 145-63.

16. Berkvens D, et al. Estimating disease prevalence in a Bayesian framework using probabilistic constraints. Epidemiology 2006;**17**(2): 145-53.
17. R Development Core Team, R: a language and environment for statistical computing, ed. R Foundation for Statistical Computing. 2013, Vienna, Austria.
18. Johnson WO, Gastwirth JL, Pearson LM.Screening without a "gold standard": the Hui-Walter paradigm revisited. Am J Epidemiol 2001;**153**(9): 921-4.

8. Project 7: Benefit-Risk monitoring of vaccines: a dashboard

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Kaat Bollaerts	01.11.16	0.3	First draft, introduction + simulated data + plots
Tom De Smedt		0.3	architecture
Lina Titievsky		0.4	Major edits introduction
Katherine Donegan	07.11.16	0.7	Minor edits to introduction & first draft of issues to consider in discussion
Lina Titievsky	07.11.16	0.7b	Minor edits and comments to all sections
Kaat Bollaerts	08.11.16	0.8	Cleaning up intro + methods: data generation
Katherine Donegan	08.11.16	0.9	Additional comments on introduction
Kaat Bollaerts		0.91	Cleaning up + discussion, suppl material
Lina Titievsky	14.11.16	0.92	Minor edits and comments on abstract, background, methods and discussion
Katherine Doneagan	15.11.16	0.93	Minor edits and comments on background and methods
Kaat Bollaerts		0.94	Cleaning up
Katherine Donegan	21.11.16	0.95	Additional comments on discussion
Kaat Bollaerts	21.11.16	0.96	Cleaning up
Tom De Smedt	22.11.16	0.97	Add results
Vincent Bauchau	05.12.16		Review
Kaat Bollaerts	05.12.16	1.0	Revision
Vincent Bauchau	15.01.17		Review
Kaat Bollaerts	20.01.17	1.0	Revision
Nick Andrews, Julia Stowe, Katherina Hartmann, Vincent Bauchau, Kaatje Bollaerts, Raphaele Roten, Alena Khromava, Steven Bailey, Mirjam Knol, Linda Levesque, Robert Maroko	08.02.17		User feedback
	24.02.17		SC review
Kaat Bollaerts	27.02.17		Addressing comments

8.1. Introduction

At the core of the mission of ADVANCE and many of its stakeholders is the concept of Benefit/Risk (B/R) monitoring. Monitoring should be understood as periodic check on several key parameters (such as coverage, incidence of adverse events or of the preventable disease) at the local (country, region) level to trigger an alert if and when there would be suspicion that the B/R profile in the population may be different from the expected profile. This alert would generate a further, subsequent and possibly more formal analysis and assessment. Monitoring should start as soon as a new vaccine is used in a given country or region, based on information from the clinical development. The target should be near real-time information, by which we mean possibly weekly refresh of data that would only be a few days old.

There are several, typically not integrated, aspects of post-licensure or post-marketing vaccine surveillance; the surveillance of vaccination uptake and compliance, safety, vaccine effectiveness and impact. Vaccination uptake and compliance to the recommended vaccination schedules are typically measured by registries, routine administrative reports or household surveys. Vaccine safety monitoring is normally implemented upon introduction of a new vaccine, and may be further enhanced in the event of switching vaccine brand or expansion of the targeted population. Timely safety monitoring of adverse events (AEs) in subpopulations excluded from pre-licensure studies is critical in the evaluation of pre-identified AEs of special interest (AESI), of rare AEs and of those with a long latency period, all of which are less detectable in pre-licensure studies due to the lack of power and limited follow-up time [1]. For the purposes of this work, the definition of post-licensure vaccine safety monitoring is restricted to the ongoing evaluation of AESI such as those safety signals identified during clinical development or from experience in previous vaccine campaigns [2]. A wide variety of methods are available to carry out such safety monitoring of vaccines during the post-licensure period [3], with an enhanced interest in near real-time surveillance using electronic healthcare databases [4-6]. Vaccine effectiveness and impact are also considered following the start of a vaccine programme. The emphasis here is placed on monitoring incidence rates of the vaccine-preventable disease (e.g. using laboratory confirmed cases or hospital admissions) while the vaccine effectiveness is usually estimated through epidemiological analyses conducted at one point in time.

Although quantitative benefit-risk assessments, by which the benefits of a medical intervention are offset against its risks at one point-in-time, are increasingly performed [7, 8], post-marketing (integrated) monitoring of coverage, benefits, risks, and benefit-risk is – to our knowledge - not yet implemented in practice. Recently, Gagne et al [9] were the first to explore the feasibility of near real-time monitoring of the comparative safety and benefits of drugs using electronic healthcare databases. For vaccines, several one point-in time benefit-risk assessments have been carried out (e.g. [10-12]), but none of them considered/involved ongoing monitoring.

With this work, we explore methodology for near real-time benefit-risk monitoring of vaccines. We visualize key data for monitoring vaccination coverage, benefits, and safety; that are then combined into composite measures of the vaccine benefit-risk profile as it evolves over time. To facilitate the monitoring, we developed an interactive dashboard. We illustrate the dashboard using simulated fictitious data reflective of the introduction of rotavirus vaccination in the UK. We chose this test case because the benefits are expected to be immediate, there is at least one serious identified AE and many publications on the safety and benefits of rotavirus vaccination in the UK exist [10,13-15], including a benefit-risk analysis [10] which we used to inform the key parameters of our data simulation model.

8.2. Methods

8.2.1. Data simulation

The national immunisation programme in the UK includes RV1, a vaccine for the prevention of severe gastroenteritis caused by rotavirus infection in young children. It was introduced into the schedule in July 2013. The vaccine is administered orally in primary care, in 2doses at two and three months of age. To illustrate the monitoring, we chose to include two benefit outcomes and one risk associated with RV1 vaccination only. The benefits are reductions in rotavirus gastroenteritis-related primary care visits and hospital admissions. The risks are events of intussusception (IS), a rare but also naturally occurring serious condition where part of the intestine prolapses into itself. Intussusception was shown to be temporarily associated with administration of a previous rotavirus vaccine that was withdrawn from the market [16] and was later found to be also associated with the newer RV vaccines in use today.

To simulate data reflecting the introduction of rotavirus vaccination in the UK, we closely follow the model as detailed in Clark et al [10]. All assumptions and parameters, except the coverage and age at vaccination, are obtained from that publication. Specifically, we simulate data on 5 consecutive birth cohorts of an arbitrary size of 300,000 children each. All children are followed from date of birth until 12 months of age. The first two birth cohorts are from prior to the introduction of the vaccination programme to allow estimation of baseline rates and detection of changes over time unrelated to the vaccine. We generate dates of birth, dates of vaccination with the first and second RV1 dose, dates of onset of rotavirus gastroenteritis (RVGE) resulting in a primary care visit (GP) or a hospital admission (HOSP) and dates of onset of intussusception (IS) (see Table 8.1 for parameters and assumptions). The coverage and age at vaccination for the first and second doses reflect the actual RV1 uptake in the UK, for which a 2-dose coverage of 88% at 12 months of age was reported for the year of vaccine introduction [15]. For vaccinated subjects, in order to estimate the number of RVGE prevented events, the likelihood of prevention is simulated as a function of the dose- and outcome-specific vaccine effectiveness (VE) and the time of the prevented event since last dose

accounting for waning of protection [10]. For vaccinated subjects, the risk of IS is simulated for two risk windows (1-7 days and 8-21 days post-vaccination, where day 0 is the day of vaccination) following both the first and second dose. The statistical package R. 3.3.1 is used to simulate the data [17].**Table 8.1:** Parameters and probability distributions used to generate the simulated data

Parameter	Value /distribution
RV1 vaccination*	
Coverage (at 12 months) – dose 1	93%
Age at vaccination (in weeks) – dose 1	~Gamma(rate =1.42, shape =12.16) | (0, 52.14)
Coverage (at 12 months) – dose 2	88%
Age at vaccination (in weeks) – dose 2	~Gamma(rate =0.68, shape =3.05, shift = 8) | (0, 52.14)
Rotavirus gastroenteritis (RVGE)**	
Annual baseline incidence per 1000 births (< 5 years)	
RVGE GP visits	28.4
RVGE hospitalizations	4.5
Age at RVGE in weeks (mean=70.7 , sd =36.6)	~Gamma(rate =0.053, shape =3.73, shift = 8.26) | (0, 52.14)
Vaccine effectiveness (VE)***	
RVGE GP visits – dose 1	87.4%, decay curve: $VE\ x\ (1-\Phi_t(\mu=3.2,\sigma=0.55)$
RVGE GP visits – dose 2	95.2%, decay curve: $VE\ x\ (1-\Phi_t(\mu=3.11,\sigma=0.96)$
RVGE hospitalisations – dose 1	96.04%, decay curve: $VE\ x\ (1-\Phi_t(\mu=3.17,\sigma=0.42)$
RVGE hospitalisations – dose 2	99.4%, decay curve: $VE\ x\ (1-\Phi_t(\mu=3.43,\sigma=0.77)$
Intussusception (IS)**	
Annual baseline incidence per 100.000 births (< 12 months)	28.1
Age at IS in weeks (mean = 30.8, sd = 14.2)	~Gamma(rate =0.15, shape =4.7, shift = -0.36) | (0, 52.14)
Relative risk of vaccine-related intussusception versus background rate	
Risk period (1-7 days) – dose 1	6.76
Risk period (8-21 days) – dose 1	3.45
Risk period (1-7 days) – dose 2	2.84
Risk period (8-21 days) – dose 2	2.11

*from Public Health England [15]** from Clark et al [10]
*** Sigmoid lognormal decay curve: $VE\ x\ (1-\Phi_t(\mu,\sigma)$ with VE at the time of vaccination and with time t in months.

8.2.2. Near real-time benefit-risk monitoring

Two composite benefit-risk measures are used to visually monitor benefit-risk; the incremental net health benefit (INHB) and the incremental benefit-risk ratio (IBRR) [18,19] (see also section 2.2.4). They are the only two trade-off indices recommended by IMI PROTECT [20] and are valued for their simplicity, which makes them suited for monitoring.

The components for vaccine benefit-risk monitoring are measures of vaccination coverage, benefits and risks. These components are then combined into the composite INHB and IBRR measures. The components are also monitored to allow interpretation of potential changes in the composite measures. We built a web-application with an interactive dashboard to facilitate the monitoring. The dashboard is interactive as it allows users to define certain options. The architecture of the dashboard is described in S1. All analyses are carried out using R 3.3.1 [17] and the web-application is build using the Shiny package [21]. Details on the calculations are given in S2.

8.2.3. Visualizations

Coverage

The weekly number of RV1 doses by (user-defined) age groups for doses 1 and 2 are monitored. The number of recorded doses given are extrapolated to the whole UK population, accounting for the age-structure of the active population captured in the database, in order to provide exposure data for potential safety signal evaluation analysis such as observed-to-expected [22].

The compliance to the recommended vaccination schedule is assessed through monitoring the weekly vaccination coverage (%). Specifically, the weekly coverage is calculated by birth cohort (defined by year and month of birth) and as the proportion of children who have been vaccinated of those who have reached a certain (user-defined) age that week.

Risk

The intussusception incidence rate (per 10,000 person years) and pointwise exact Poisson 95% CIs are estimated for two risk windows (1-7 and 8-21 days) after each dose. The rates were estimated cumulatively over time, i.e. using the accruing data. We opted to do so to maximize sample size, and hence accuracy, as the expected absolute vaccine-associated AE rate is small and unlikely to change over time.

The baseline intussusception incidences rates are estimated for children of vaccination eligible age from the two pre-vaccination birth cohorts. Specifically, the baseline risks are estimated for children aged 8-10, 9-12, 12-14 and 13-15 weeks for comparison with the incidence within 1-7 days post dose 1, 8-21 days post dose 1, 1-7 days post dose 2 and 8-21 days post dose 2, respectively. These

age groups were chosen given the age specific recommendations for each dose (8 weeks at dose 1 and 12 weeks at dose 2) with time after those ages included to account for the variability in the age at vaccination as well as the length of the risk windows.

Benefits

The incidence rate (per 10.000 py) and pointwise exact Poisson 95% CIs of RVGE GP visits and of hospitalizations in the total population of infants aged 0 to 1 year are calculated starting two years prior to the vaccine introduction within 'moving windows' of data. To allow balancing freshness of data (i.e. only using the most recent data) and accuracy of the rate estimates, the length of the look-back period can be chosen by the end-user.

A reference line is obtained through calculating the expected benefits given assumed values for the baseline incidence and VE and accounting for the observed age-specific vaccination coverage and age-distribution of the database population. For simplicity, and as most of the vaccinated children received two doses, a conservative estimate of expected benefits is obtained ignoring the one-dose VE and the potential indirect effects.

Benefit-Risk

The composite benefit-risk measures INHB and IBRR are used for monitoring the vaccine benefit-risk profile. The INHB is essentially a difference between a sum of weighted incremental benefits and a sum of weighted incremental risks or,

$$INHB = \sum_{k=1}^{K} w_k \times (E_{0\,k} - E_{v\,k}) + \sum_{k=1}^{K'} w_k \times (R_{0\,k} - R_{v\,k}) = \boldsymbol{E} + \boldsymbol{R}\,, \qquad (1)$$

where K and K' refer to the number of benefit and risk outcomes, where the incremental benefits are the difference between the benefits in the absence of vaccination or baseline benefits ($E_{0.}$) and the benefits after vaccination ($E_{v.}$), and similarly for the incremental risks ($R_{0.}$ and $R_{v.}$). The weights w_k are all positive and reflect the severity of the health outcomes. Note that, because $E_{v.}$ and $R_{v.}$ are subtracted from their baseline values, the incremental benefits ($\boldsymbol{E}$) are positive and the incremental risks ($\boldsymbol{R}$) negative. The IBRR is the ratio of the incremental benefits to the incremental risks or,

$$\text{IBRR} = \boldsymbol{E}/_{-\boldsymbol{R}}, \qquad (2)$$

with positive terms for both numerator and denominator.

First, the benefit-risk of RV1 vaccination is monitored using the observed benefits and observed risks in the total population of infants aged 0 to 1 years. The INHB and IBRR are calculated with the incremental benefits being estimated by comparing the 'moving window' incidence rates (per 10,000 py) of RVGE GP visits and of hospital admissions after vaccine introduction with the incidences from the pre-vaccination birth cohorts (as calculated in 2.3.3). The incremental (or excess)

intussusception risk for each risk window after each dose is obtained by estimating the attributable number of cases (per 10,000 py) within the population for which the benefits were calculated. First the attributable fractions (AF) are calculated from the relative incidences (RI) as AF = (RI−1)/RI with the RI estimated using the accrued data (see 2.2.2). Then, the observed number of cases within each risk window is multiplied with the AR to obtain the attributable number of cases. The pointwise 95% Wald CIs of the INHB and IBRR are obtained as well (see S2 for the derivations).

As the benefits are often observed late compared with the short term risks, the INHB and IBRR are also calculated based on theoretical benefits. The theoretical benefits are obtained by multiplying assumed levels of baseline incidence with assumed levels of VE for both RVGE GP visits and hospitalizations while accounting for the observed coverage. The incremental risks are calculated as before.

8.2.4. Demonstration with fictitious data

The weekly number of administered doses within user-defined age groups is depicted using stacked area charts (Figure 8.1). Dose 1 is mostly given to 8-10 week olds whereas dose 2 is mostly given to 12-14 week olds. The weekly vaccination coverage (%) by user-defined age groups and by month-year birth cohort show a rapid uptake of the vaccine upon introduction (Figure 8.2 and Figure 8.3). The intussusception incidence rates (/10.000 py) are displayed using line charts with the shaded areas representing the 95% pointwise CIs (Figure 8.4). Immediately after the introduction of the vaccine, the CIs of the intussusception incidences within the post-vaccination risk windows are wide. For accumulating data, the CIs narrow with the largest increased risk observed 1-8 days after dose 1. The observed incidence rates of AGE GP visits and hospitalisations (/10.000 py) for a look-back period of 26 weeks are lower compared with the expected incidences assuming baseline incidences as observed pre-vaccination and a VE of 60% (Figure 8.5). Given the assumed preference weights for IS, RVGE GP and RGE HOSP of 100, 50 and 1, respectively, the composite benefit-risk measure with observed benefits was initially negative and turned positive at ±27 weeks after its introduction. The INHB was -13.7 (95% CI -82.5, -52.2) at week 25 and 989 (95% CI: 964.9, 1013.5) at week 65 for a population of 10.000 children followed from birth until 1 year of age. The IBRR showed similar trends. The initial negative benefit-risk is explained by comparing immediate risks (excess risk of intussusception within the first 3 weeks after vaccination recommended at 8 and 12 weeks of age) with long-term benefits (with the peak age of AGE infection between six months and two years).The INHB and IBRR with theoretical benefits assuming baseline incidences as observed pre-vaccination and a VE of 60% are always positive with a INHB of 688.4 (95%CI: 680.6, 696.2) and IBRR of 91.6 (95%CI: 33.0, 254.0) at week 25.

Figure 8.1: Number of administered doses (left: dose 1, right: dose 2) in the UK population by user-specified age groups, by calendar time (in weeks).

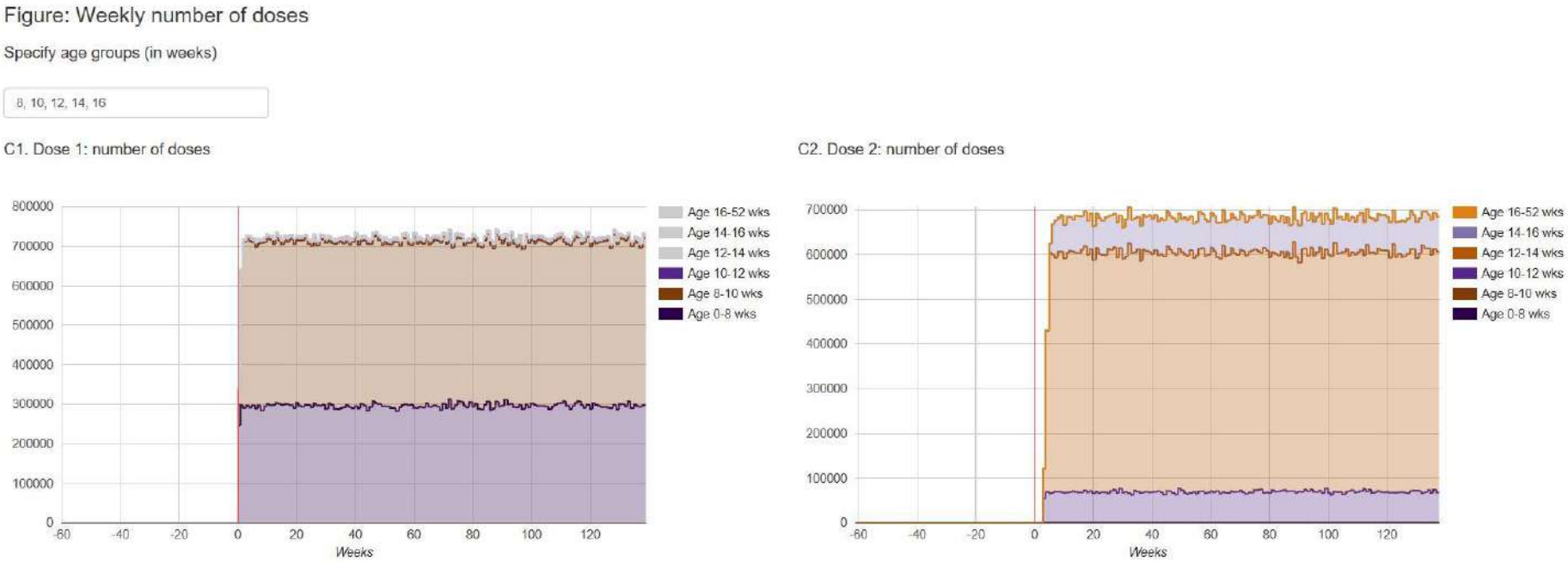


Figure 8.2: Coverage (%) (left: dose 1, right: dose 2) in children who reached a certain-specified age, by calendar time (in weeks).

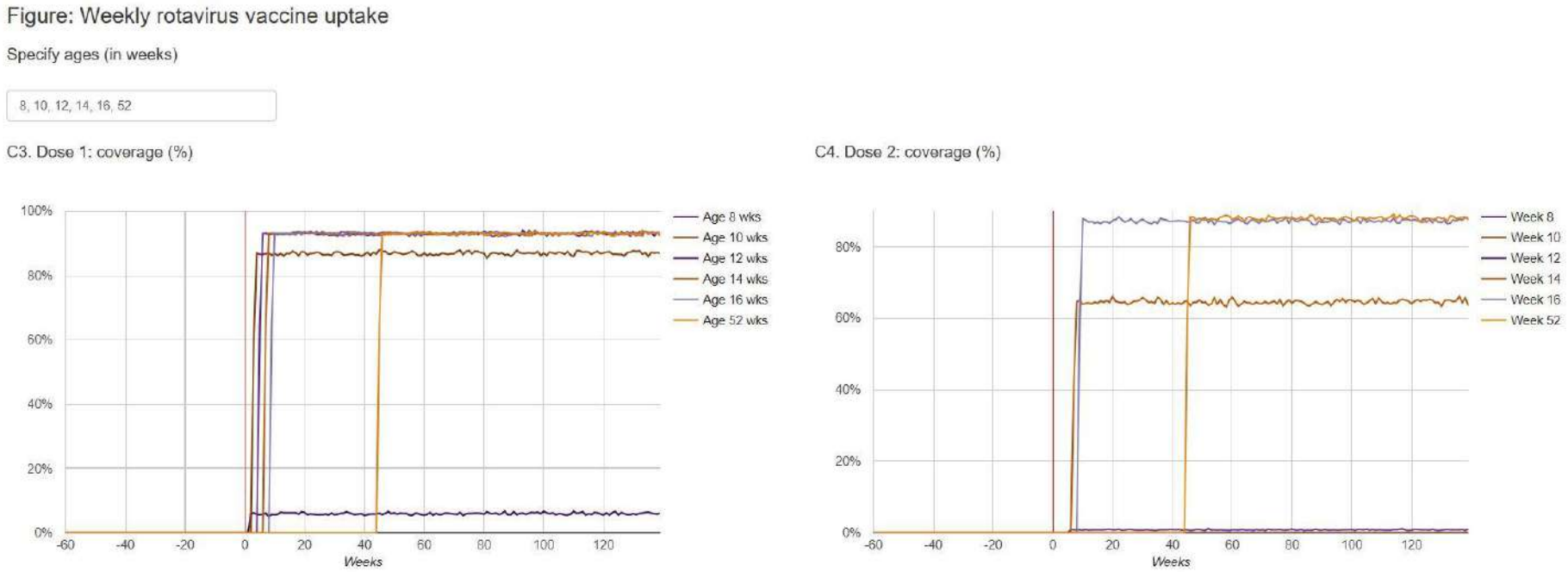


Figure 8.3: Coverage (%) (left: dose 1, right: dose 2) by birth cohort defined by year and month.

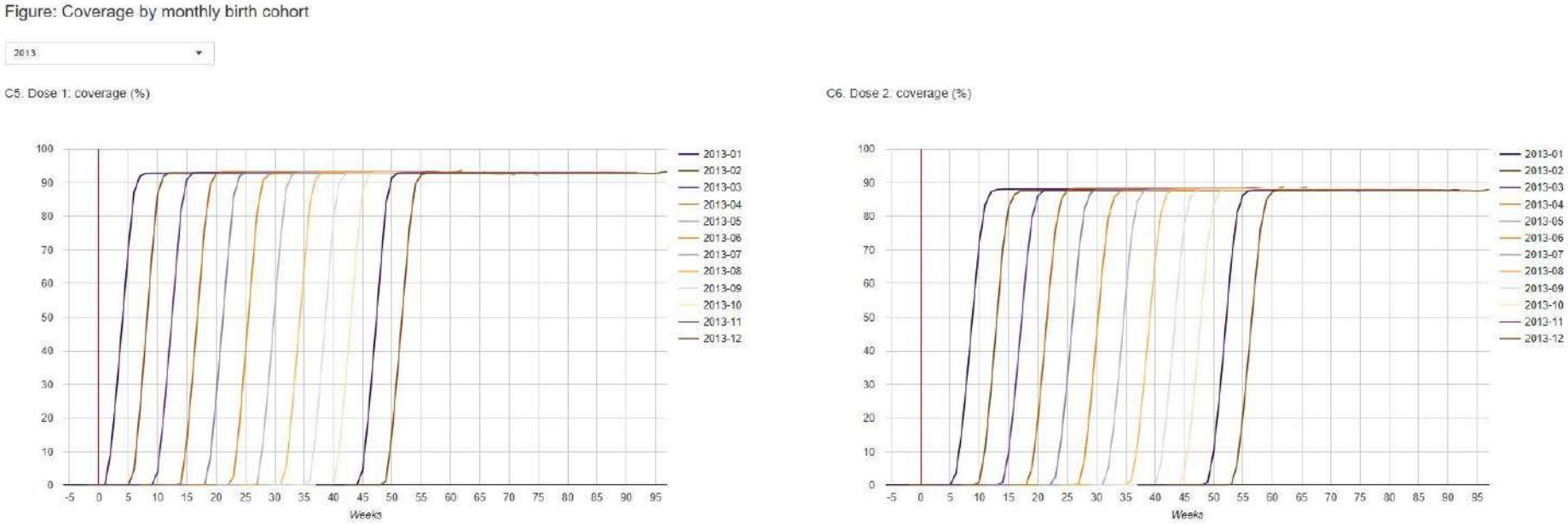


Figure 8.4: Incidence rate per 10.000 person years [95% CI] of intussusception estimated cumulatively over time, incidence prior to vaccination (7 to 12 weeks and 11 to 16 weeks) and in risk windows (1-7 and 8-21 days post vaccination) by dose.

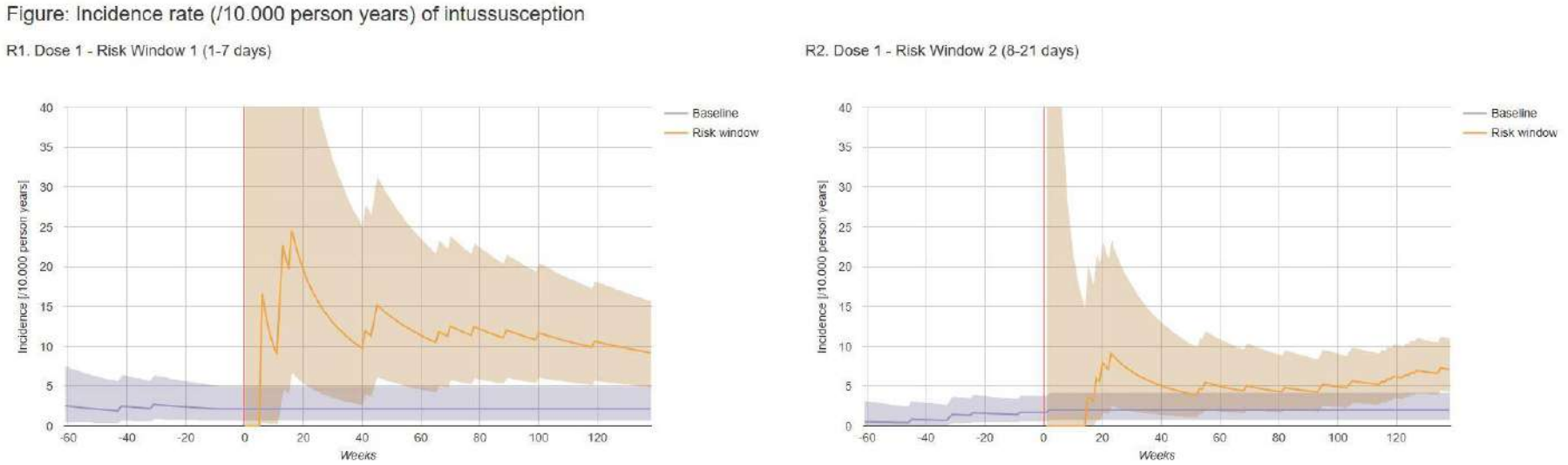
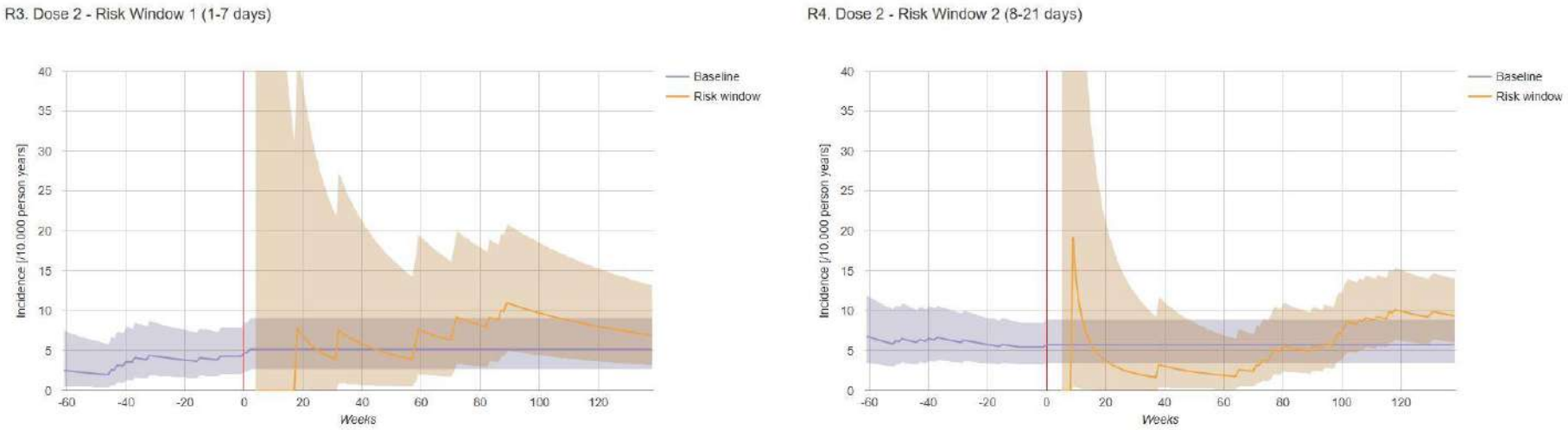


Figure 8.5: Running incidence rate per 10.000 person years [95% CI] of AGE GP Visits (left) and Hospital Admissions (right) visits in total population within a user-defined look-back period. The expected incidence is calculated for user-defined levels of baseline incidence and vaccine effectiveness.



Lookback period (in weeks)

26

Baseline incidences (/10.000 person years)

AGE GP Visits

280

AGE Hospital Admissions

44

Vaccine Effectiveness

AGE GP Visits

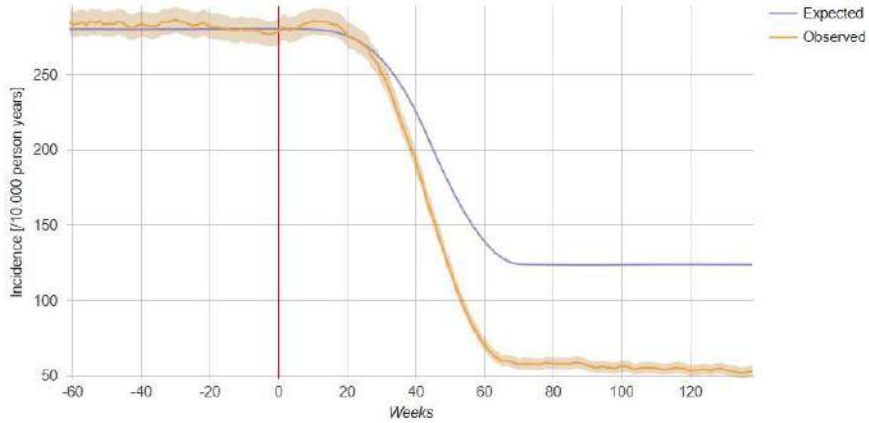
0.6

AGE Hospital Admissions

0.6

Figure: Running incidence rate (/ 10.000 person years) of acute gastroenteritis (AGE), total population

B1. GP Visits



B2. Hospital Admissions

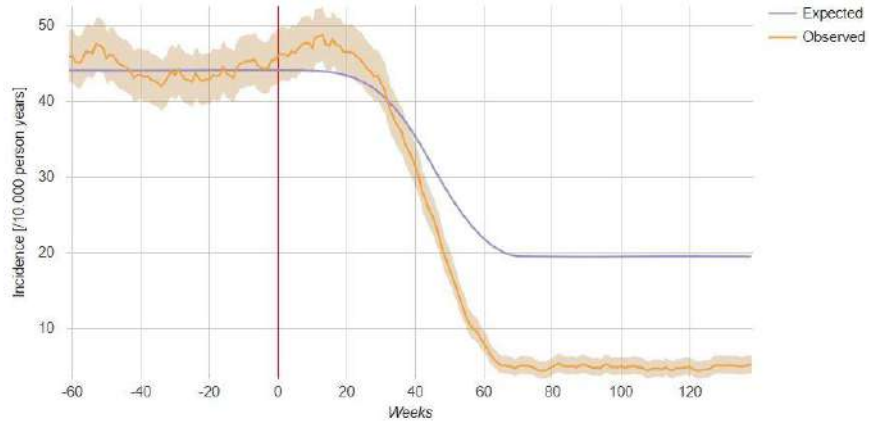


Figure 8.6: INHB (left) and the BRR (right) [95% CI] with observed benefits for a population of 10.000 children with the vaccination coverage as observed followed from birth till 1 year of age. For the INHB, the weighted components are also displayed. The user-defined settings are as in Figure 5.

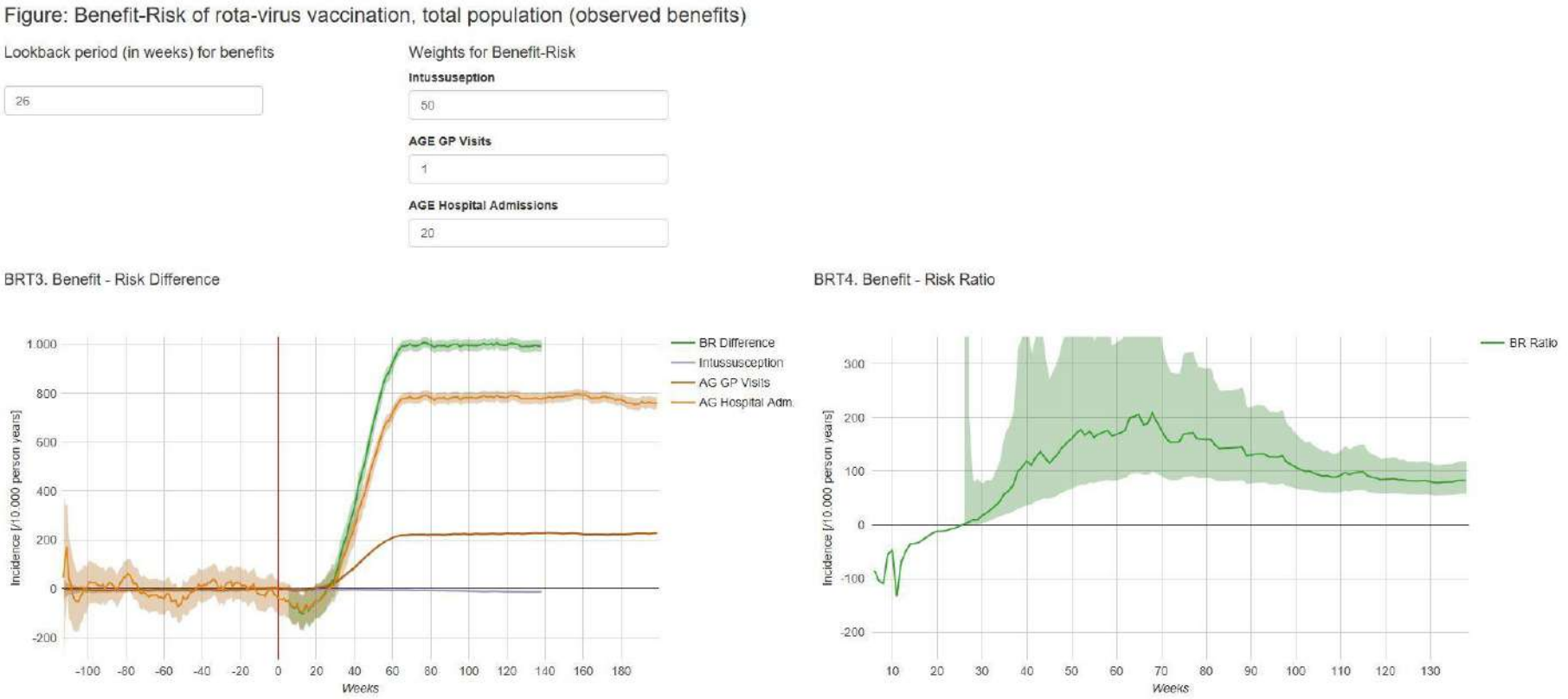
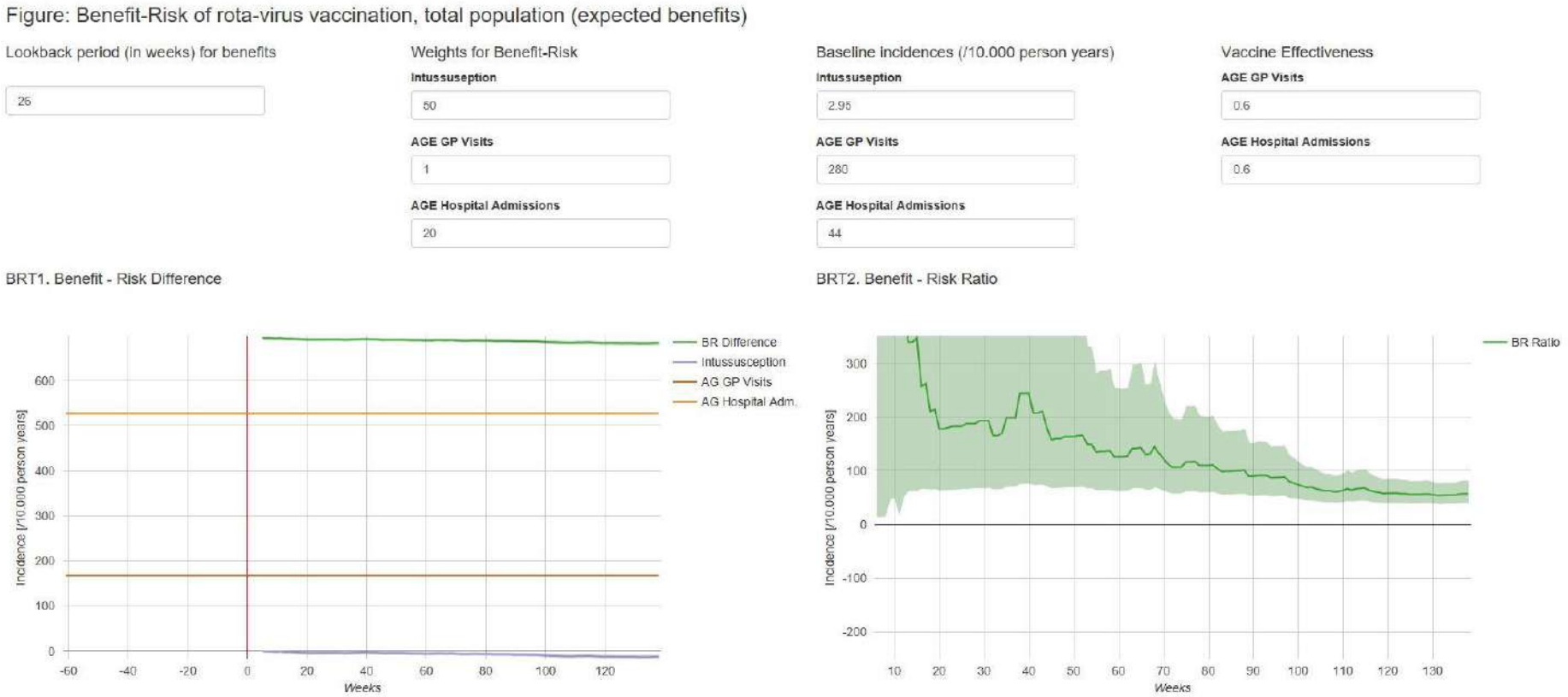


Figure 8.7: INHB (left) and the BRR (right) [95% CI] with expected benefits per 10.000 fully vaccinated children followed from birth till 1 year of age. For the INHB, the weighted components are displayed as well.



8.3. Discussion

With this work we propose a methodology for the near real-time benefit-risk monitoring of vaccines. Particularly, we visually monitor the composite benefit-risk measures INHB [18] and IBRR [19] as well as their components: vaccination coverage, benefits and risks. To facilitate the monitoring, we built a web-application with an interactive dashboard.

The INHB, which underlies many benefit-risk assessment methods [23], as well as the IBBR, are the two simple and intuitive trade-off indices [20], making them suited for monitoring. Both measures are also commonly used in cost-effectiveness research [24]. In the context of immunization, the INHB is an absolute measure indicating how much the total disease burden in the population changes due to vaccination whereas the IBRR is a relative measure, indicating how much disease burden is prevented relative to the disease burden induced through vaccination. This also implies that the INHB requires preference weights on the absolute scale (with a defined zero point), whereas the IBRR only requires weights on a relative scale.

To the best of our knowledge, this is the first time that a web-application with an interactive dashboard has been developed in the context of post-marketing monitoring of vaccines. Dashboards are used in many fields and are well suited for monitoring as changes over time can be visualized and the underlying data can be seamlessly updated. The dashboard is made interactive, allowing end-users to select age groups or time windows, to calculate the benefit-risk measures for different sets of preference weights and to conduct sensitivity analyses.

For developmental and illustrative purposes, we use simulated data reflective of the introduction of rotavirus vaccination in the UK. Several simplifications were made when simulating the data. We ignored herd immunity and seasonal trends in gastroenteritis. Although the dashboard was developed with the ultimate objective of near real-time benefit-risk monitoring using electronic healthcare databases, our simulated data assume no disease misclassification (e.g. RVGE events are mostly not recorded as such in healthcare databases but rather as unspecified gastroenteritis), no exposure misclassification, no confounding and no incomplete follow-up, which are all commonly present in healthcare databases.

Obviously a successful near real-time benefit-risk monitoring depends on data being available in a timely fashion, i.e. both frequent refresh and small time-lag between occurrence of the event and recording in the database. In order to assess the actual feasibility and added value of the proposed methodology, it should be tested in a real-life scenario of near real-time benefit-risk monitoring using one or several electronic healthcare databases, possibly accounting for confounding,

misclassification, incomplete follow-up, and, when combining data, heterogeneity across databases. The visualisations and underlying calculations will probably have to be modified depending on the vaccine, health outcomes of interest and data sources. Relevant future methodological work might include the development of sequential hypothesis testing for composite benefit-risk measures. While cost-effectiveness is beyond the scope of this work some of the outputs from the benefit-risk monitoring could be used to monitor key variables that impact on cost-effectiveness as well.

Taking all together, we believe near real-time benefit-risk monitoring of vaccines using interactive dashboards is worth further exploration. It will complement, but not replace, other activities. For example, signal detection will still run, in parallel, and if a new safety concern is identified, then the event can be easily added to the monitoring. The dashboard will only monitor the pre-identified benefits and risks.

Finally, we would like to stress that the results presented in this paper should not be used to support any conclusions with regard to the actual benefit-risk of rotavirus vaccination in the UK as this work used simulated data and several simplifying assumptions were made with the sole purpose of methodology development.

8.4. References

1. Lopalco PL, et al. Monitoring and assessing vaccine safety: a European perspective. Exp Rev Vaccines 2010;**9**(4): 371-80.
2. Nelson JC, et al. Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. Pharmacoepidemiol Drug Saf 2012;**21** Suppl 1: 62-71.
3. Nelson JC, et al. Methods for observational post-licensure medical product safety surveillance. Stat Methods Med Res 2015;**24**(2): 177-93.
4. Davis RL, et al. Active surveillance of vaccine safety: a system to detect early signs of adverse events. Epidemiology 2005;**16**(3): 336-41.
5. Leite A, Andrews NJ,Thomas SL.Near real-time vaccine safety surveillance using electronic health records-a systematic review of the application of statistical methods. Pharmacoepidemiol Drug Saf 2016;**25**(3): 225-37.
6. Glanz JM, et al. White paper on studying the safety of the childhood immunization schedule: for the Vaccine Safety Datalink.
7. FDA. Enhancing Benefit-Risk Assessment in Regulatory Decision-Making. Available from: http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm326192.htm.
8. IMI-PROTECT Benefit-Risk Group: Recommendations for the methodology and visualisation techniques to be used in the assessment of benefit and risk of medicines. 2014; Available a: http://protectbenefitrisk.eu/documents/HughesetalRecommendationsforthemethodologyandvisualisationtechniquestobeusedintheassessmento.pdf.
9. Gagne JJ, et al. Prospective benefit-risk monitoring of new drugs for rapid assessment of net favorability in electronic health vare data. Value Health 2015;**18**(8): 1063-9.
10. Clark A, et al. Evaluating the potential risks and benefits of infant rotavirus vaccination in England. Vaccine 2014;**32**(29): 3604-10.

11. Ledent E, et al., Post-marketing benefit-risk assessment of rotavirus vaccination in Japan: asimulation and mModelling analysis. Drug Saf2016;**39**(3): 219-30.
12. Marcelon L, et al. Quantitative benefit-risk assessment by MCDA of the quadrivalent HPV vaccine for preventing anal cancer in males. Expert Rev Vaccines 2016;**15**(1): 139-48.
13. Atchison CJ, et al. Rapid declines in age group-specific rotavirus infection and acute gastroenteritis among vaccinated and unvaccinated individuals within 1 year of rotavirus vaccine introduction in England and Wales. J Infect Dis 2016;**213**(2): 243-9.
14. Stowe J, et al. The risk of intussusception following monovalent rotavirus vaccination in England: A self-controlled case-series evaluation. Vaccine 2016;**34**(32): 3684-9.
15. Public Health England (PHE). National rotavirus immunisation programme: preliminary data for England, October 2013 to September 2014. 2014; Available from: https://www.gov.uk/government/publications/rotavirusimmunisation-programme-vaccine-coverage-estimates.
16. FDA. Withdrawal of rota-virus vaccine recommendation. MMWR Morb Mortal Wkly Rep, 1999;**48**: 1.
17. R Development Core Team, R: a language and environment for statistical computing, ed. R Foundation for Statistical Computing. 2013, Vienna, Austria.
18. Stinnett AA, Mullahy J.Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making 1998;**18**(2 Suppl):S68-80.
19. Korting H, Schafer-Korting M. The benefit-risk ratio: a handbook for the rational use of potentially hazardous drugs. 1999, Boca Raton: CRC Press LLC.
20. Mt-Isa S, et al. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepidemiol Drug Saf 2014;**23**(7): 667-78.
21. Chang WJC, Allaire JJ, Yihui Xie Y, McPherson J. Shiny: Web Application Framework for R. 2016.
22. Najafzadeh M, et al. A unified framework for classification of methods for benefit-risk assessment. Value Health 2015;**18**(2): 250-9.
23. Craig BA, Black MA.Incremental cost-effectiveness ratio and incremental net-health benefit: two sides of the same coin. Expert Rev Pharmacoecon Outcomes Res 2001;**1**(1): 37-46.
24. Armitage P, Berry G. Statistical methods in medical research. 1987, Oxford: Blackwell Scientific.

9. Project 8: Composite Burden of Disease measures for adverse events following immunization

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Scott McDonald	15.11.16		First draft of main text and Appendix
Kaat Bollaerts	06.12.16		Minor edits to Appendix
Danielle Nijsten	15.12.16		Minor edits to main text
Susan Hahné	01.02.17		Minor edits, and comments
Scott McDonald	02.02.17		Edited to take Vincent Bauchau's and Patrick Mahy's comments into account, and additional minor edits for clarity/cleaning up
Miriam Sturkenboom, Vincent Bauchau, Jorgen Bauwens			SC review
Scott McDonald	27.02.17		Addressed SC comments

9.1. Introduction

Vaccination is recognised indisputably as one of the most effective public health interventions. Despite the drastic improvements in population health attributed to vaccination programmes, there has been public concern regarding possible negative consequences from being vaccinated: adverse events. Consistent with the WHO guidelines (WHO, 2013), adverse events following immunization (AEFI) are defined as "any untoward medical occurrence which follows immunization and which does not necessarily have a causal relationship with the usage of the vaccine". All AEFI represent departures from a state of full health.

To date, there has been limited assessment of the population-level health burden of key health outcomes associated with vaccination using composite burden of disease (BoD) measures, such as the disability-adjusted life-years (DALY) measure (Clark & Cameron, 2006; Cho et al., 2010). Although selected safety aspects of vaccination have a long history of investigation, a comprehensive estimate of the health burden of events associated with vaccination would be a useful contribution to current knowledge. In addition, quantitative estimates of adverse event burden fit well within a benefit-risk framework (Mt-Isa et al., 2013) for assessment of new or existing vaccines, if the (projected) averted disease burden due to vaccination can also be quantified (Halloran et al., 1997; van Wijhe et al., 2016). Therefore, the aim of this work is to describe a methodological framework for computing the population-level disease burden of vaccination-attributable adverse events.

A quantitative measure of health burden should ideally take into account the frequency of occurrence, severity and duration of illness, the risk of complications, and the risk of mortality. The most commonly used summary, or composite, health measures are the quality-adjusted life years (QALY) and the DALY measures. Summary measures of population health allow meaningful comparison between heterogeneous conditions and their effects on the full spectrum of health. The utility of summary measures goes beyond that of simple epidemiological indicators such as incidence or mortality rates, as they integrate mortality with morbidity in a single indicator – for the latter, taking into account both severity and duration of illness/disability – and therefore they may also prove suitable for making comparisons between events, vaccine types, age-groups, and national or regional populations.

The DALY is the most commonly-used summary measure of population health burden [5,6], and is typically applied to compare the relative impact of diseases on a population. The DALY combines the years lived with disability for a health state (i.e., living with a condition, disease, disability, or injury) with the years of life lost due to premature mortality; thus, time is the metric for both morbidity and mortality. One DALY is equivalent to one lost year of healthy life [7,8].

Our principal objective is therefore to assess the feasibility and usefulness of adapting current disease burden methodology for estimating the burden of adverse health outcomes associated with vaccination. An eventual follow-on goal, which is not addressed in this deliverable, is to integrate the developed Burden of Disease (BoD) methodology within a benefit-risk monitoring platform (i.e. the ADVANCE vision), in which the prevented burden of vaccine-preventable disease can be compared with the burden of adverse events, using a common metric such as the DALY.

9.2. Methods

We piloted our methodological approach for estimating the disease burden using a selected set of events recognised as adverse events following immunization (AEFI). Selection was based on the frequency of occurrence and potential severity of the event, and was carried out independent of knowledge regarding links to specific vaccination types. Typically, AEFIs that occur relatively frequently are mild, but can still be responsible for causing disability (i.e., vaccination recipient experiences a short period of life at less than full health), whilst those that occur extremely rarely can have serious consequences. Therefore, we estimated the disease burden for both broad categories of adverse events: (i) infrequent, but potentially serious events; and (ii) relatively frequent, but less serious events.

Originally, it was planned to obtain background event incidence rates from the ADVANCE databases (which would be generated as part of the WP5 fingerprinting task), and to conduct between-country and across-time comparisons of adverse event burden. To this end, semi-automatic mapping of case definitions (via the CodeMapper tool; see Project 9) for the selected events was carried out, but ultimately these were not used (see Appendix Methods 3). As the fingerprinting results could not be delivered within the time-frame for this report, we instead focus on describing the methodology and applying it to a worked example. The resulting BoD estimates reported here should be taken as highly provisional only, as they were produced solely for illustration purposes.

9.2.1. Selection of candidate adverse events

For the selection of adverse events we distinguished adverse events that are infrequent, but potentially serious and relatively frequent, but less serious. The set of investigated events in this study was based on the adverse events included in the Global Research in Paediatrics(GRiP) reference set (Brauchli Pernus et al. 2015) and the Postlicensure Rapid Immunization Safety Monitoring (PRISM) study (Baker et al, 2013), and for the frequent event selection we utilized the Eudravigilance (1995-2014) dataset. Events from the GRiP and PRISM studies were ranked according to frequency in Eudravigilance (see Appendix D). We restricted the set of candidate events to those for which an incidence rate could potentially be determined from an electronic health records (EHR)

database (thus excluding very mild events, such as injection site tenderness). From the GRiP reference set of 14 adverse events and 13 vaccines, only those vaccine-event pairs with a likelihood of a strong association were retained. This step was based on the GRiP evaluation of strength of evidence (i.e., vaccine-event pair either identified as a 'positive control', or present in the Reports of the Institute of Medicine (IOM): Adverse Effects of Vaccines: Evidence and Causality, 2011), and logically restricts adverse event burden computation to established, positive associations with vaccination (Table 8.1).

Table 8.1: Selected adverse events and sources for background incidence rates

Adverse event	Category [frequency/ severity]	Age-group	Background inc. rate (95% CI)	Period and setting	Reference
Idiopathic thrombo-cytopenic purpura	Infrequent/high	<2 yrs 2-5 yrs	6.8/100,000 (4.9-9.2) 7.2/100,000 (5.9-8.8)	1990-2005, UK	Yong et al., 2010
Anaphylaxis	Infrequent/high		*See Table 8.2*		
Febrile convulsions	Frequent/low	2-12 mos 13-24 mos 25-60 mos 61-120 mos 121-180 mos	556/100,000 (537-575) 1377/100,000 (1348-1407) 432/100,000 (413-433) 58/100,000 (54-61) 23/100,000 (18-28)	1999-2011, UK	Sammon et al., 2015

Literature search for relative risks

Publications providing estimates of the relative risk (or the absolute risk, defined terms of cases per vaccine dose) for the identified vaccine-event pairs were retrieved via PubMed searches, and setting, study population, pharmaceutical details, and effect estimates were abstracted.

9.2.2. Burden of disease methods and required parameters

DALY calculation. The vaccination-associated disease burden of each adverse event of interest was estimated using the composite DALY measure. The DALY is the sum of years of life lost to premature mortality (YLL) and years of life lived with disability (YLD), with YLL and YLD computed from a number of essential parameters [7,11]:

$$DALY = YLL + YLD$$

$$YLL = No.\ deaths \times life\ expectancy\ at\ age\ of\ death$$

$$YLD = No.\ events \times disability\ weight \times duration$$

Disability weights and durations. Disability weights encode the severity of the health outcome, and can be obtained from professional or lay populations using a variety of preference elicitation methods [12]; the current Global Burden of Disease (GBD) [7] approach is to use general public

survey respondents [13]. The disability weight is on a scale from 0 (perfect health) to 1 (death). If not available from existing databases or the relevant literature, then weights from proxy health outcomes need to be assigned, ideally through consultation with experts with appropriate medical knowledge. Disability durations are typically determined from literature review and/or clinical expert knowledge.

Outcome trees/subsequent sequelae. In cases where the occurrence of a certain adverse event can precipitate recurrence of the same event, or can increase the risk of severe sequelae later in life, an outcome tree (also known as disease progression pathway) can be specified to incorporate the risk, severity and duration of subsequent health outcomes [14].

Mortality. Distinguishing mortality as a causal reaction to vaccination from coincidental death is crucial, given the extreme rarity of vaccination-attributable death [15]. Causality assessment of all deaths potentially having a relation to the vaccine product, a vaccine quality defect, or contamination is highly recommended [16]. The BoD framework can easily include estimation of YLL for deaths confirmed as an immediate adverse outcome, or for premature mortality associated with development of a long-term sequela (via definition of an outcome tree, with a specified case-fatality ratio; e.g., [14]). For a comprehensive burden estimate, it is vital to compute YLL for any adverse event with a non-zero case-fatality rate. For YLL, life expectancies from standard life tables are additionally required.

Under-ascertainment and under-reporting. Determination of either background adverse event incidence or direct attribution of the number of events to vaccination is susceptible to under-reporting (failure to report or misclassification of cases seeking healthcare to an EHR database (or comparable system)) and to under-ascertainment (missing cases: who do not seek healthcare) [17]. In the presence of either, the BoD will be under-estimated. These factors can be problematic for comparison of burden between adverse events, if the extent of under-reporting/ascertainment differs between the type of event.

9.2.3. Selection of parameters for the example burden calculation

The single most important outcome required for computing the health burden of adverse events is vaccination-attributable event incidence. By 'vaccination-attributable', we do not make a strong assumption that the observed adverse event has a causal relationship with the vaccine itself, but merely that the event is associated with administration of the vaccine. By 'attributable', we refer to the extent to which the event incidence is associated with vaccination, adjusting for the expected, or background incidence in the population.

This outcome measure can be obtained by various means: for instance, via querying of an EHR database with linked date(s) of vaccination(s), through primary data collection via cohort or self-controlled case series designs [18], or from published reports of event incidence. If only background incidence rates are available (whether from databases or from published sources), then vaccination-attributable incidence can be inferred through application of appropriate relative risk estimates and risk window-size to the background incidence (see Appendix Methods 1). For the example burden computation, we made a number of choices based on availability of data and parameters (following subsections).

Setting. For reasons of tractability, our worked example computes the adverse event burden associated with routine vaccinations administered to young children (<4 years of age) only. Based on the availability of background incidence rates for overlapping time periods (below), we chose the UK as the population setting, and estimated burden for the arbitrarily chosen year 2005.

Background event incidence rates. For idiopathic thrombocytopenic purpura (ITP) and febrile convulsions, published background incidence rates were located for recent periods (1990-2005 and 1999-2014) from two studies using the UK General Practice Research Database (Table 8.1). For the latter two events, narrow age-groups were reported. We computed vaccination-attributable event incidence and YLD based on incidence rates within these narrow age-groups, and then later aggregated to two wider groups (2 to 12 months, and 13 months to <4 years) for reporting purposes. Note that age-groups can be fine-tuned to the target ages for vaccination within the routine vaccination schedule, if event incidence rate data are available at a suitable granularity, for instance by month of age.

Relative risks of vaccine-attributable event. The age-groups for which published relative risks or risks per dose were available did not necessarily match the relevant ages within the UK vaccination schedule (Table 8.2 and http://vaccine-schedule.ecdc.europa.eu/Pages/Scheduler.aspx). For instance, the selected study for the vaccine-event pair DTaP-ITP provided an estimated relative risk for the age group 4-6 years, but in the UK the DTaP booster is recommended to be given at 3 years 4 months of age. Therefore, we made the following assumptions regarding applicability of a given published RR (or risk) to a particular age-group. First, the first three DTaP (infant) doses were all assumed to have the same RR of ITP; the RR based on 12-19 month-old children [19] was used. Second, the published RR for DTaP–ITP based on 4-6 year-olds was applied to the age-group (3 years 4 months) receiving the fourth dose. Third, identical RRs/risks were used for both first and booster MMR doses, as separate estimates were not available from the selected studies.

Disability weights and durations. As first source for disability weights, we obtained weights from most recent Global Burden of Disease study (GBD 2013; [13]), which updates and expands the set of weights elicited for GBD 2010 [20]. Proxy weights were adopted for two of the three selected events. Disability durations were retrieved from a variety of published sources (Table 8.2); for convulsions and anaphylaxis, duration was assumed to correspond to average stay in hospital.

Table 8.2: Parameters for the example YLD computations for the selected vaccine-event pairs

Vaccine – adverse event pair	Age-group	RR or risk per 1M doses (95% CI)	Reference	DW	DD
DTaP – ITP	12-19 mos 4-6 yrs	1.00 (0.21-4.81) 2.57 (0.53-12.37) [6 week window]	[19]	0.159	5 weeks
MMR – ITP	<18 yrs	12.5/1M doses (11.8-13.2)	[21]	0.159	5 weeks
DTaP/wP – Anaphylaxis	0+ yrs	5.14/1M doses (1.06-15.01)	[22]	0.552	1 day
MMR– Anaphylaxis	<18 yrs	1.3/1M doses (0.03-7.1)	[23]	0.552	1 day
HBV – Anaphylaxis	<18 yrs	1.1/1M doses (0.1-3.9)	[23]	0.552	1 day
MenC – Anaphylaxis	0+ yrs	6.16/1M doses (1.68-15.78)	[22]	0.552	1 day
VZV – Anaphylaxis	0+ yrs	6.58/1M doses (0.80-23.77)	[22]	0.552	1 day
MMR – Febrile convulsions	3 mo – <10 yrs	2.75 (2.55-2.97) [14 day window]	[24]	0.263	1 day
MMR – Generalised convulsions	<7 yrs	1.11 (0.11-11.28) [8-14 day window]	[25]	0.263	1 day

Note. ITP = idiopathic thrombocytopenic purpura; DTaP = diphtheria/tetanus/acellular pertussis; MMR = measles/mumps/rubella; HBV=hepatitis B virus; MenC = meningococcal C; VZV = varicella-zoster virus; DW = disability weight. DD = disability duration

Consequent health outcomes. For certain events, risks of progressing to subsequent health outcomes following the initial event have been reported. For simplicity, we exclude the potential burden from additional health outcomes from our YLD estimates, as for our selected events these risks are either small, or there is insufficient evidence for progression. For instance, for febrile convulsions, we excluded the risk of suffering recurrent seizures (increased rate of recurrence has been estimated at 19% [24]). For ITP patients with a low platelet count, complications (severe bleeding) can rarely occur [26]; we excluded this sequela from our estimates. We also excluded mortality as a consequent health outcome following any of our selected adverse events, due to recognised rarity of occurrence [13].

9.2.4. Computational details

Disease burden measures can be computed using specialised software (e.g., the DALY package for R; [27]), custom software (Appendix Methods 2.2), or via a spreadsheet (Appendix Methods 2.1).

We estimated YLD for the two relevant age-groups: (i) the age-range encompassing the first set of vaccinations within the UK childhood scheme (2 to 12 months), and (ii) the age-group covering receipt of the third DTaP dose, the second MenC dose, and the first MMR dose (at between 12 and 13 months old), as well the DTaP and MMR boosters for three-year-olds (13 mos to <4 years). All estimates were for the year 2005, but given that background incidence rates were available aggregated over a period of 13 to 16 years (Table 8.1), YLD estimates for other years within the same period will be very similar, with variability due only to variation in population size and vaccination coverage.

The first step was to calculate the vaccination-attributable incidence rate for each event from either the background incidence rate (for ITP and febrile convulsions) or the risk per dose (anaphylaxis). See Appendix Methods 2.2 for a detailed description of the computation. As the UK schedule specifies three doses of DTaP (but only one of MMR, HPV, and Men C) during infancy, three at-risk periods were defined for the DTaP–ITP pair according to the relevant window period sizes (6 weeks each; see Table 8.2) within the first 24 months of life (because the background incidence rate for ITP was available for <2 years). The at-risk periods for the DTaP and MMR boosters were similarly defined according to the relevant window periods, for the background incidence age-groups 2-5 years (for ITP) and 13-24 months (for febrile convulsions).

For anaphylaxis, no published background incidence rates were located; therefore, vaccination-attributable incidence was estimated based on the risk per million doses. We equated the number of doses to the number of vaccinated children in 2005, in turn estimated from the size of the UK birth cohort in 2005 (for doses 1-3 of DTaP, doses 1-2 of MenC), 2004 (dose 1 of MMR, dose 3 of MenC), or 2002 (booster doses of DTaP and MMR), multiplied by the relevant vaccination coverage value (see Appendix Methods 1, Eq. 8).

The second step was to compute YLD for each vaccine-event pair, stratified by age-group, based on the vaccination-attributable incidence adjusted for vaccination coverage (see Appendix Methods 1, Eq. 7) and the disability weight and duration (Table 8.2). Point estimates and 95% uncertainty intervals (UIs) were computed using R statistical software ([28]; see Appendix Methods 2.2). Finally, we tabulated YLD vaccine-event pair and age-group, and additionally computed estimates aggregated over vaccine type and age-group.

As vaccination against either HBV or varicella is not part of the UK's routine childhood immunization programme (they are given to certain risk groups only), we did not estimate the associated YLD for the HBV–anaphylaxis and VZV–anaphylaxis pairs. Should suitable vaccination uptake data be available, YLD for these pairs can be straightforwardly computed.

9.2.5. Regulatory status

Because the current study uses only data from the literature on event incidence rates, independent of exposure, it is not considered to be a post-authorization study.

9.3. Results

9.3.1. Selection of adverse events

A preliminary set of nine adverse events was identified for further investigation (Appendix Table D); four were selected as examples of rare events with potentially serious consequences, and five were selected to represent more frequently occurring, but less serious events. For the present illustration of the BoD methodology, this set was reduced to a total three events (Table 8.1). Based on the GRiP evaluation of strength of evidence for occurrence following various vaccination types, we then searched PubMed for relevant studies providing estimates of the (relative) risks of event occurrence for the eight relevant vaccine-event pairs (Table 8.2). At least one suitable study was located for all vaccine-event pairs, except for HBV–ITP. Results of the literature search are provided in Appendix Tables A1-A4. A single RR/risk study for each vaccine-event pair was then chosen for the worked example through discussion among three of the co-authors (SAM, SH, DN).

9.3.2. Example YLD computation

In the following, we describe the results of our example study for the three selected adverse events. The reader should note that these results are provided to illustrate the burden computation procedure only, and because a number of parameters require confirmation through consultation with medical experts they should not be regarded as definitive. Our intention is to describe how adverse event burden estimates can be usefully reported.

Table 8.3: Results of example YLD computations (both absolute YLD and YLD per million population), for the selected vaccine-event pairs, UK 2005. Vaccination-attributable event incidence rate is per 100,000 person-years.

Vaccine – adverse event pair	Age-group	Vaccination-attrib. inc. rate (95% UI)	YLD (95% UI)	YLD/1,000,000 (95% UI)
DTaP – ITP	2 – 12 mos	2.12 (0.62-4.76)	0.19 (0.06-0.43)	0.32 (0.094-0.73)
	13m – <4 yrs	1.27 (0.32-3.18)	0.53 (0.13-1.32)	0.19 (0.048-0.49)
MMR – ITP	13m – <4 yrs	0.53 (0.51-0.55)	0.22 (0.21-0.23)	0.081 (0.078-0.084)
ITP (all vaccines)	*2m to <4 yrs*	*1.90 (1.00-3.50)*	*0.96 (0.51-1.77)*	*0.29 (0.15-0.54)*
DTaP/wP – Anaphylaxis	2 – 12 mos	1.61 (0.64-3.32)	0.014 (0.006-0.030)	0.024 (0.010-0.050)
	13m – <4 yrs	0.10 (0.02-0.32)	0.004 (0.001-0.013)	0.002 (0.000-0.005)
MMR– Anaphylaxis	13m – <4 yrs	0.15 (0.07-0.29)	0.006 (0.003-0.012)	0.002 (0.001-0.004)
MenC – Anaphylaxis	2 – 12 mos	1.35 (1.19-1.53)	0.012 (0.011-0.014)	0.020 (0.018-0.023)
	13m – <4 yrs	0.14 (0.12-0.17)	0.006 (0.005-0.007)	0.002 (0.002-0.003)
Anaphylaxis (all)	*2m to <4 yrs*	*0.88 (0.65-1.22)*	*0.044 (0.032-0.061)*	*0.013 (0.010-0.019)*
MMR – Febrile convulsions	13m – <4 yrs	58.3 (31.9-101)	1.14 (0.62-1.98)	0.42 (0.23-0.73)
	2m to <4 yrs	*47.9 (26.2-83.1)*	*1.14 (0.62-1.98)*	*0.35 (0.19-0.60)*
All vaccine-event pairs	***2m to <4 yrs***	***96.3 (63.5-142)***	***2.29 (1.51-3.38)***	***0.69 (0.46-1.03)***

The estimated vaccination-attributable event incidence rates for the UK in 2005, per vaccine-event pair and age-group (2–12 months and 13 months to <4 years), are shown in Table 8.3. The highest vaccination-attributable incidence rate was for febrile convulsions associated with MMR vaccination (55.3/100,000person-years, for the 2–12 months age-group), and the lowest vaccination-attributable incidence rate was estimated for anaphylaxis associated with DTaP vaccination (0.10/100,000, for the 13 months to <4 years age-group).

The morbidity burden in YLD, as well as in YLD per 1,000,000 persons (to facilitate comparisons between age-groups, and across time and/or between populations) are also shown in Table 8.3 (see also Fig. 8.1). The largest absolute morbidity burden was estimated for febrile convulsions (YLD of 1.14; 95% UI: 0.63-2.02), and the lowest YLD for anaphylaxis (0.044; 95% UI: 0.033-0.062). The YLD per 1,000,000 measure indicated a higher population-level burden for children under 13 months of age compared with the older age-group, for all applicable vaccine-adverse event pairs.

Aggregating over age-group and over all vaccine-event pairs for which YLD was computed, the overall absolute morbidity burden for these three events was estimated at 2.19 (95% UI: 1.42-3.31) DALYs.

9.4. Discussion

We have presented methodology for estimating the morbidity burden associated with adverse events following vaccination within the burden of disease framework, and as a proof of concept, we illustrate the methodology in the form of a worked example. Transparency of the computations involved, and

the expected ease of deployment beyond the three events we investigated are positive attributes of the developed method.

The retrieval of background incidence rate data, risks, and other necessary parameters from the literature, together with the computation of YLL for three adverse events occurring after routine UK childhood immunizations provided a real-world example for application of BoD methods to this area. The extra insight provided by YLD (over vaccination-attributable incidence only), which additionally takes into account the severity and duration (and possible longer-term consequences) of the event, is clear. In our worked example, YLD/1,000,000 distinguished the population-level health impact of vaccine-event pairs with very similar attributable incidence rates (e.g., DTaP–ITP and DTaP–Anaphylaxis for 2–12 month-year-old infants). Although the vaccination-attributable incidence rates for the two events we had *a priori* classified as frequent/less serious was higher than for the two events classified as infrequent/serious (Table 8.3), the pattern according to YLD differed, illustrating that such a two-way classification can alternatively be quantified by a measure that captures the population-level health burden.

9.4.1. Challenges and limitations

A literature search often did not yield 'ideal' relative risk estimates. It was difficult to find large studies (to provide sufficient statistical precision), or studies that were reasonably recent and/or geographically relevant. Conducting a meta-analysis of published risks for each vaccine-event pair might be a preferred approach. The granularity of the relative risks obtained from the literature was variable, with often very broad age-groups defined, and for vaccines given in multiple doses, separate estimates for each dose were not provided. Accordingly, we had to assume identical (relative) risks and risk periods for each dose. In addition, multiple vaccines administered at the same occasion – the norm for routine childhood immunisation – complicate estimation of vaccination-attributable incidence due to overlapping at-risk periods. Clearly, application of published relative risks to background incidence rates, or use of absolute risks, requires numerous assumptions to be made about generalisability across time, setting, dose, and age-group.

For one of our vaccine-event pairs (DTaP–ITP), the selected relative risks were not significantly different from 1.0. Of course, statistical significance of the published RR depends on study power, and the number of outcomes is often very small. Obtaining relative risks (or better, vaccination-attributable event incidence directly) from large EHR databases is a promising approach for improving precision. The purpose of our worked example was to illustrate all steps required to produce the burden estimates. We stress that decisions regarding the power of candidate studies (from which parameters such as RR are obtained), and/or if meta-analyses should be conducted, need to be made a priori

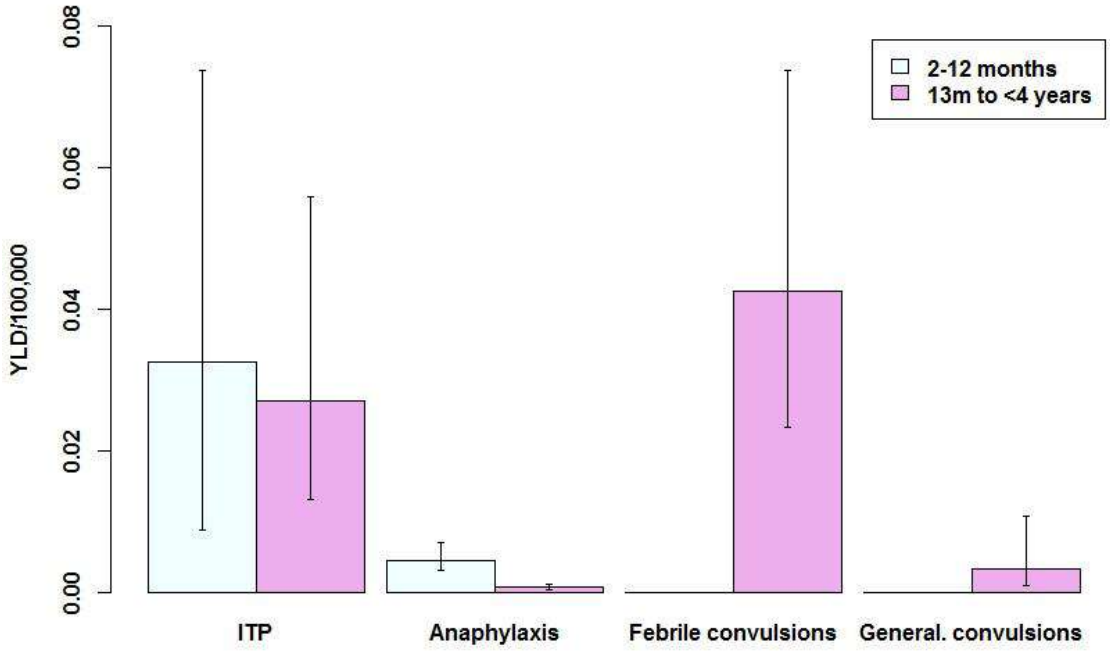


Figure 8.1: Estimated YLD per 100,000 persons, with 95% uncertainty intervals, by event and age-group.

Retrieving all required disability weights from published elicitation studies is a principled approach; however, two of our three selected events were not included in the most comprehensive and contemporary source available (GBD 2013: [13]), and consequently disability weights for proxy health outcomes needed to be chosen. Comparable data sources for disability durations do not exist, and although values can be located from diverse published sources (as we have done), a systematic review approach is clearly preferable, coupled with medical experts' opinions and review of the selected durations (also applicable to the selection of proxy disability weights).

To illustrate the BoD computation, we focussed on events associated with routine (early) childhood vaccinations only. The adverse event burden associated with vaccinations received in adolescence and adulthood (e.g., travel vaccinations, annual influenza jabs) is also of substantial interest, but the challenges in estimating BoD are even greater, especially when estimating vaccination-attributable event incidence from background incidence rates and relative risks. This is because one-time or ongoing medication use that may also cause the event of interest needs to be distinguished from vaccination; without information about the temporal relationship between either intervention and the event, correct attribution or adjustment is very difficult.

Vaccination of infants and children can influence the health state of the parent, for example as anxiety due to the occurrence of an adverse event and/or uncertainty of prognosis. We followed the conventional BoD approach and ascribed burden only to the individual who received the vaccination.

9.4.2. Recommendations

Selection of all DALY parameters should be guided by clinical knowledge of the adverse events of interest, and/or undergo review by medical experts (safety physicians), as the resulting disease burden estimates are crucially dependent on the appropriateness and correctness of the parameters chosen. A process of expert consultation is recommended at the data collection stage as well as for interpretation of the findings, for the dual purposes of ensuring scientific accuracy and credibility [29].

9.4.3. Conclusions

BoD methodology can feasibly be applied to estimate the health burden of adverse events following immunization, but interpretation of the findings must consider the quality, appropriateness, and accuracy of all data sources contributing to the DALY computation. DALYs are most usefully evaluated in context, for instance to create a ranking of diseases in terms of burden. For the burden of adverse events following vaccination, estimates are meaningfully interpreted when they are compared with the burden of the disease(s) the vaccination prevents. If the population-level disease burden averted by a vaccination programme is similarly quantified using the DALY – which importantly allows benefits and risks to be expressed in a common currency – then the current methodology may find useful application within a benefit-risk monitoring platform.

9.5. References

Baker, M. A., Nguyen, M., Cole, D. V., Lee, G. M., & Lieu, T. A. (2013). Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. Vaccine, 31, K98-K112.

Black, S., Eskola, J., Siegrist, C. A., Halsey, N., MacDonald, N., Law, B. & Vellozzi, C. (2010). Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. The Lancet, 374(9707), 2115-2122.

Brauchli Pernus Y. et al. (2015). Reference set for performance testing of pediatric vaccine safety signal detection methods and systems. Vaccine. doi: 10.1016/j.vaccine.2015.10.013.

Cho, B. H., Clark, T. A., Messonnier, N. E., Ortega-Sanchez, I. R., Weintraub, E., & Messonnier, M. L. (2010). MCV vaccination in the presence of vaccine-associated Guillain–Barré Syndrome risk: A decision analysis approach. Vaccine, 28(3), 817-822.

Clark, M., & Cameron, D. W. (2006). The benefits and risks of bacille Calmette-Guérin vaccination among infants at high risk for both tuberculosis and severe combined immunodeficiency: assessment by Markov model. BMC pediatrics, 6(1), 5.

Cohen, Y. C., Djulbegovic, B., Shamai-Lubovitz, O., & Mozes, B. (2000). The bleeding risk and natural history of idiopathic thrombocytopenic purpura in patients with persistent low platelet counts. Archives of internal medicine, 160(11), 1630-1638.

Devleesschauwer, B., Havelaar, A. H., De Noordhout, C. M., Haagsma, J. A., Praet, N., Dorny, P., et al. (2014). Calculating disability-adjusted life years to quantify burden of disease. International journal of public health, 59(3), 565-569.

Devleesschauwer, B., McDonald, S., Haagsma, J., Praet, N., Havelaar, A., & Speybroeck, N. (2013). DALY: The DALY Calculator-A GUI for stochastic DALY calculation in R. R package version 1.2.0.

Gibbons, C. L., Mangen, M. J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., ... & Kretzschmar, M. E. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health, 14(1), 1.

Gold, M. S., Balakrishnan, M. R., Amarasinghe, A., & MacDonald, N. E. (2016). An approach to death as an adverse event following immunization. Vaccine, 34(2), 212-217.

Halloran, M. E., Struchiner, C. J., & Longini, I. M. (1997). Study designs for evaluating different efficacy and effectiveness aspects of vaccines. American journal of epidemiology, 146(10), 789-803.

Havelaar, A. H., Kirk, M. D., Torgerson, P. R., Gibb, H. J., Hald, T., Lake, R. J., ... & Speybroeck, N. (2015). World Health Organization Global estimates and regional comparisons of the burden of foodborne disease in 2010. PLoS Med, 12(12), e1001923.

Mangen MJ, Plass D, Havelaar AH, Gibbons CL, Cassini A, Muhlberger N, et al. (2013). The pathogen- and incidence-based DALY approach: an appropriate methodology for estimating the burden of infectious diseases. PLoS One 8(11):e79740.

Mathers CD, Vos T, Lopez AD, Ezzati M. (2001). National burden of disease studies: a practical guide. Edition 2.0. Geneva: World Health Organization.

Miller, E. R., Moro, P. L., Cano, M., & Shimabukuro, T. T. (2015). Deaths following vaccination: what does the evidence show?. Vaccine, 33(29), 3288-3292.

Murray CJL (1994) Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Org. 72:429–445.

Murray CJ, Lopez AD. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet 349(9063):1436–42.

Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2013 Jan 4;380(9859):2197-223.

R Development Core Team (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Salomon, J. A., Haagsma, J. A., Davis, A., de Noordhout, C. M., Polinder, S., Havelaar, A. H., ... & Murray, C. J. (2015). Disability weights for the Global Burden of Disease 2013 study. The Lancet Global Health 3(11): e712-e723.

Salomon JA, Murray CJ, Ustun TB, Chatterji S. (2003). Health state valuations in summary measures of population health. In: Murray CJ, Evans DB, eds. Health systems performance assessment: debate, methods, and empiricism. Geneva: World Health Organization. pp. 409–436.

van Wijhe M, McDonald SA, de Melker HE, Postma MJ, Wallinga J. Effect of vaccination programmes on mortality burden among children and young adults in the Netherlands during the 20th century: a historical analysis. Lancet Infect Dis. 2016; 16(5): 592-598.

Vestergaard, M., Hviid, A., Madsen, K. M., Wohlfahrt, J., Thorsen, P., Schendel, D., ... & Olsen, J. (2004). MMR vaccination and febrile seizures: evaluation of susceptible subgroups and long-term prognosis. Jama, 292(3), 351-35.

Weldeselassie, Y. G., Whitaker, H. J., & Farrington, C. P. (2011). Use of the self-controlled case-series method in vaccine safety studies: review and recommendations for best practice. Epidemiology and infection, 139(12), 1805-1817.

World Health Organization (WHO). (2013). Causality assessment of an adverse event following immunization (AEFI). User manual for the WHO revised classification. Available from http://apps.who.int/iris/bitstream/10665/80670/1/9789241505338_eng.pdf (accessed Nov 2016).

.

10. Project 9: CodeMapper: semi-automatic coding of case definitions

DOCUMENT HISTORY

NAME	DATE	VERSION	DESCRIPTION
Benus Becker	05-01-2016	0.1	First draft
Miriam Sturkenboom	07-02-2016		Comments and revision
Benus Becker	17-02-2016	0.2	Second draft
Paul Avillach	29-02-2016		Review and comments
Jan Kors	09-03-2013		Review and comments
Benus Becker	03-04-2016	0.3	Third draft
Paul Avillach	10-02-2016		Review and comments
Benus Becker	01-07-2016	0.31	Addressing comments
Jan Kors	05-07-2016		Review and comments
Benus Becker	12-07-16	0.32	
Jan Kors	04-08-2016		Final remarks and changes
Benus Becker	12-08-2016	1	
Jan Kors, Miriam Sturkenboom	15-08-2016		Comments and review
Simon de Lusignan, Patrick Mahy, Marianne van der Sande, Vincent Bauchau	16-08-2016		Review and comments
Benus Becker	12-08-2016	1.1	
Silvana Romio, Miriam Sturkenboom	01-09-2016		Review and comments
Benus Becker		1.2	
Jan Kors, Benus Becker	21-09-2016	10.3	
Silvana Romio, Benedikt Becker	22-09-2016	2	Final version

10.1. Introduction

10.1.1. Coding heterogeneity in the EU

In order to increase the scale of pharmacoepidemiological studies, information from multiple EHR databases should be combined in a distributed, collaborative fashion [1]. However, EHR databases use different coding vocabularies to record medical information [2,3], such as the ICD-9 CM [4] and ICD-10 [5], ICPC-2 [6], Read-2[7] and Read-CTv3[8]. In multi-database studies, the extraction of an event typically requires several steps to achieve consistency between databases. A case definition that describes the event in the study protocol is translated into an operational definition, which is then mapped for each vocabulary into a set of codes that represents the event. The code sets are combined into queries for case identification and harmonized between databases by comparison with benchmarks from the literature and by feedback from the database custodians.

The creation of code sets for each vocabulary from the textual case definitions has been largely a manual process. Given the number and complexity of the targeted vocabularies, the mapping and harmonization process can pose an important bottleneck in the rapid implementation of collaborative epidemiological studies [9,10]. Furthermore, the rationale for including or excluding individual codes is not consistently documented, which hampers the possible reuse of code sets and queries in subsequent studies.

10.1.2. Prior workflows and pathways to bridge the heterogeneity

A previous attempt to accelerate the creation of code sets from multiple vocabularies was made in the EU-ADR project [7,9,10]. Medical concepts such as diseases, symptoms, laboratory procedures, or tests were automatically identified in a case definition using the MetaMap program [11]. Code sets representing the concepts in the targeted vocabularies were then generated using the UMLS [14], a biomedical terminology system that integrates many vocabularies including coding vocabularies commonly used in EHR databases. Whereas the identification of concepts and their projection to codes was automated, the overall workflow was not integrated or recorded to facilitate the later reuse of the mapping. We present a web application called CodeMapper, which was developed to assist in mapping case definitions to code sets from different vocabularies while keeping a record of the complete mapping process. We evaluate the application by comparing code sets that were automatically generated by CodeMapper with reference code sets that were manually created in a previous epidemiological study.

10.2. Methods

CodeMapper's mapping approach consists of three phases (Figure 9.1, top). First, medical concepts are automatically identified from a free-text case definition. The user can then revise the set of

medical concepts by adding or removing concepts, or expanding a concept to more general or more specific concepts. For example, the concept *Coughing* can be expanded to more general concepts such as *Respiratory disorders* and *Abnormal breathing*. Expanding it to concepts that are more specific results in subtypes of coughing such as *Paroxysmal cough* and *Evening cough*. Finally, each concept is represented by (possibly several) codes in the targeted vocabularies, and the projection of the concepts to codes forms the result of the mapping process.

10.2.1. Mapping approach

CodeMapper builds upon information from the Metathesaurus of the UMLS. The Metathesaurus is a compendium of many medical vocabularies, which have been integrated by assigning equivalent codes and terms from different source vocabularies to the same concepts. Each concept in the UMLS is identified by a Concept Unique Identifier (CUI). For example, the concept Coughing (CUI: C0010200) is among others associated with the codes 786.2 (ICD-9 CM), R05 (ICD-10) and XC07I (Read-CTv3). The Metathesaurus contains more than one million concepts connected to codes from 201 vocabularies. Each concept is assigned to one or more of 127 semantic types, which define broad conceptual categories like *Disease or syndrome*, *Finding*, or *Substance*. To provide even broader structure, semantic types are combined into 15 semantic groups [15]. We used version 2016AA of the UMLS in this evaluation.

The automatic concept identification of CodeMapper is based on lexical information from the Metathesaurus. The lexical information of a concept consists of terms that can be used in free-text to refer to that concept (Figure 9.1, bottom left). We compiled a dictionary for the concepts in the semantic groups *Anatomy*, *Chemicals & Drugs*, *Di*sorders, *Genes & Molecular Sequences*, *Living Beings*, *Phenomena*, *Physiology*, and *Procedures* of non-suppressible, English terms from the following vocabularies: MeSH [16], MedDRA [17] SNOMED-CT [18], ICD-9 CM, ICD-10 CM, ICPC-2, and Read-CTv3. Our text-indexing engine, Peregrine, uses this dictionary to identify medical concepts in the case definition [19].

CodeMapper provides two operations to improve the sensitivity of the mapping by expanding a concept to more general or more specific concepts, based on the hierarchical relationships in the Metathesaurus. Hierarchical relationships connect concepts that are more general or more specific in meaning (Figure 9.1, bottom centre). For example, the concept for Coughing is connected to the more general concept *Respiratory Disorders*, and to the more specific concept *Paroxysmal cough*. To expand a concept in CodeMapper, all concepts that have a more general or more specific relationship with it are identified and displayed in the application for selection by the user. Hierarchical relationships in the Metathesaurus are inherited from the source vocabularies (called parent and child), or defined in the Metathesaurus (called broader and narrower) [20]. Both types of hierarchical relationships are taken into account for concept expansion.

The projection of concepts to code sets from the targeted vocabularies follows the identification of equivalent codes in the Metathesaurus (Figure 9.1, bottom right).

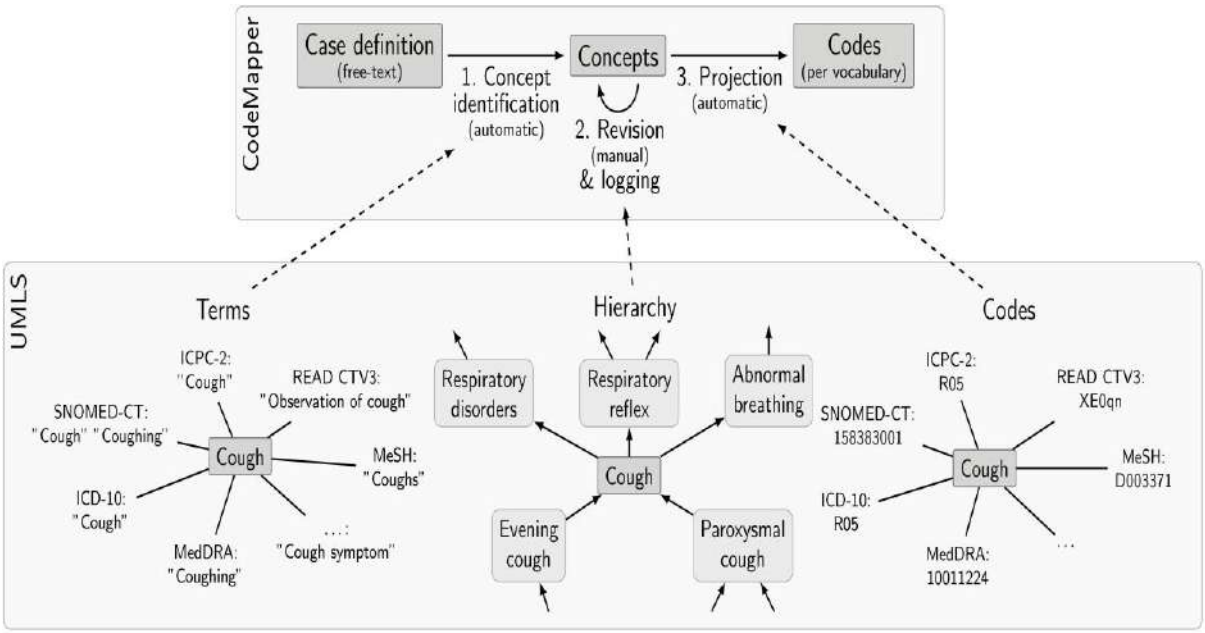


Figure 9.1: Key phases of CodeMapper (top) and the usage of information from the UMLS Metathesaurus, exemplified by the concept for *Cough* with CUI C0010200 (bottom). Terms from the Metathesaurus drive the automatic identification of concepts in the free-text case definition. Hierarchical information about concepts in the Metathesaurus is used to retrieve related concepts during revision of the mapping. Information in the Metathesaurus is used to project the selected concepts to codes from the targeted vocabularies.

10.2.2. User application

The CodeMapper application is implemented as a web application[2]. CodeMapper has three screens. On the first screen, the user enters a clinical case definition of an event as free-text. Medical concepts are automatically identified in the text and highlighted inline. By default, only concepts that belong to the semantic group of *Disorders* are preselected for further processing in the application, but the user can select and deselect any identified concept depending on their relevance for the described event.

The second screen displays the mapping as a table with one row for each medical concept, and one column for each targeted vocabulary (Figure 9.2). Each cell contains the names of the codes that

[2] https://euadr.erasmusmc.nl/CodeMapper

are used to represent the medical concept of the row in the targeted vocabulary of the column. The codes are displayed when the names are hovered over with the mouse. Several user operations are available for revising the mapping. The user can remove concepts from the mapping, search and add concepts, or retrieve more general and more specific concepts. The retrieved concepts are shown in a list and can be selected by the user for inclusion in the mapping. The user can also add or remove vocabularies that should be targeted by the mapping. After every operation, the codes are automatically updated and displayed in the table.

The third screen shows a list of all operations that have been made, for later traceability of the mapping process. When the user saves the mapping, theyhave to provide a summary of the modifications, which is incorporated into the mapping history. After saving, the mapping and history list are available to other users of the application. Comments can be attached to concepts to capture the discussion about the mapping. Concepts can be categorized by tags. Finally, the user can download the mapping as a spreadsheet file, for example to incorporate the codes into extraction queries. The spreadsheet file comprises the original free-text case definition, the concepts of the mapping, the codes for the targeted vocabulary, and the full history of the mapping process.

The source code and additional documentation about the application including a walk-through of the functionality and the recording of a webinar about CodeMapper held in December 2015 are available on the ADVANCE SharePoint (WP4 documents | Methods testing (phase2) | 4_CODEMAP).

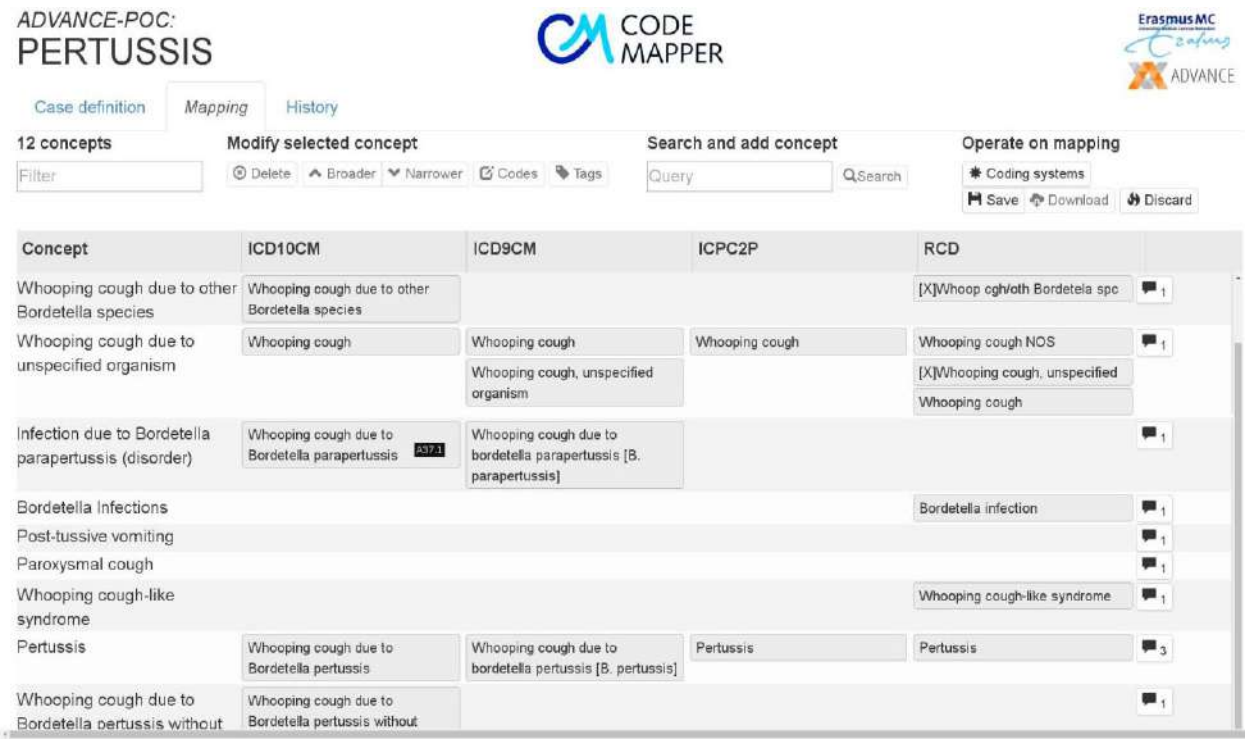


Figure 9.2: The second screen of the CodeMapper application provides operations to revise the concepts of a mapping. The mapping is displayed as a table. The cells show the code names from the vocabulary stated in the column that correspond to the concept of the row. Individual codes are shown when hovering the terms. The balloons in the last column indicate the number of comments attached to a concept.

10.2.3. Evaluation

We initially evaluated the effectiveness of the CodeMapper approach for creating realistic code sets for a number of case definitions, by comparing code sets that were generated with CodeMapper with manually created reference code sets. We used case definitions and reference code sets from the FP-7 funded SAFEGUARD project[3][21], which was conducted in nine different EHR databases in the EU and USA. The protocol can be found at in the EU-PAS registry[4]. This project was selected for the variety of mapped events and the range of targeted vocabularies. The manual mapping process consisted of deriving operational definition from the textual case definition, choosing codes from the targeted vocabularies without the use of the Metathesaurus, and refining the code set based on feedback from database custodians. The reference mappings also contained exclusion codes, which were not considered in the evaluation because they were not generally derived from the case definitions.

[3] http://www.safeguard-diabetes.org
[4] http://www.encepp.eu/encepp/viewResource.htm?id=8323

SAFEGUARD studied nine events: acute pancreatitis, bladder cancer, haemorrhagic stroke, heart failure, ischemic stroke, myocardial Infarction, pancreatic cancer, sudden cardiac death, and ventricular arrhythmia. One event (sudden cardiac death) was excluded from the evaluation because of several missing code sets, and another (heart failure) because the case definition contained only a short symptomatic description of the event, unrelated to the codes representing the event. The events were mapped for nine EHR databases with four vocabularies: Medicare, PHARMO, HSD and regional EHR databases from Lombardy and Puglia (all these databases use ICD-9 CM), GePaRD (ICD-10, German modifications), IPCI and BIFAP (both ICPC-2 and keywords), and CPRD (Read-2). We selected the code sets for Medicare for ICD-9 CM as the reference since it contained less database-specific additions than the other code sets using ICD-9 CM. The codes for GePaRD are contained by the ICD-10 and ICD-10 CM vocabularies in the UMLS, so we combined the codes generated by CodeMapper for these vocabularies. The Metathesaurus covers only Read-CTv3 and not Read-2. To generate codes for Read-2, a translation table between Read-2 and Read-CTv3 was integrated into CodeMapper (available at the Health & Social Care Information Centre (HSCIC), https://isd.hscic.gov.uk). Codes from the IPCI mapping were trimmed to three digits to adjust for the database-specific codes in IPCI.

Overall, the reference code sets contained 420 codes (Table 9.1).The size of the reference code sets vary widely between vocabularies: on average, the code sets for Read-2 contain 48.3 codes, whereas the code sets for ICPC-2 contain 1.1 codes. This discrepancy is firstly due to the differences of granularity of the vocabularies (Read-2 has 77290 codes in the Metathesaurus, ICPC-2 only 1397). Secondly, the queries to the IPCI database (to which the ICPC-2 code sets are targeted) are supported by keyword searches on the free-text portion of the IPCI medical records and additional exclusion criteria.

Table 9.1: Number of words in case definitions and number of codes in the reference set. The numbers of exclusion codes are given in brackets.

Event	Case definition (word count)	Codes ICD-9	ICD-10	ICPC-2[a)]	READ-2
Acute Pancreatitis	49	1 (0)	6 (0)	1 (0)	7 (0)
Bladder cancer	87	12 (0)	12 (0)	1 (3)	91 (0)
Hemorrhagic stroke	48	3 (2)	22 (2)	1 (2)	36 (0)
Ischemic stroke	53	10 (0)	11 (0)	2 (1)	20 (0)
Myocardial Infarction	39	11 (1)	7 (0)	1 (6)	-[b)]
Pancreatic Cancer	19	8 (0)	9 (0)	1 (1)	109 (0)
Ventricular Arrhythmia	234	5 (0)	5 (0)	1 (1)	27 (0)
Sum	529	50 (3)	72 (2)	8 (14)	290 (0)
Average	75.57	7.14 (0.43)	10.29 (0.29)	1.14 (2.0)	48.33 (0.0)

[a)] Additional text-based queries for IPCI database
[b)] Text-based query only for GePaRD database

Different code sets were generated by CodeMapper for the events of the reference project based on the same case definitions. The baseline code sets resulted from the concepts identified automatically in the case definition (Figure 9.3). We then simulated the actions of an "informed user" who wants to improve the sensitivity of the mapping. We assumed that this user would expand the concepts and, from all possible concepts that are more general or more specific, would only retain those that map to codes present in the reference set. The resultant set of concepts defined a new code set. We simulated four of these expansion steps on successive concept sets.

For each target vocabulary and event, the generated code set was compared with the reference code set. We determined the number of true-positive codes (TP), false-positive codes (FP), and false-negative codes (FN), and computed sensitivity (TP / (TP + FN)) and PPV (TP / (TP + FP)). We report for each vocabulary the sensitivity and PPV averaged over all events in the reference set.

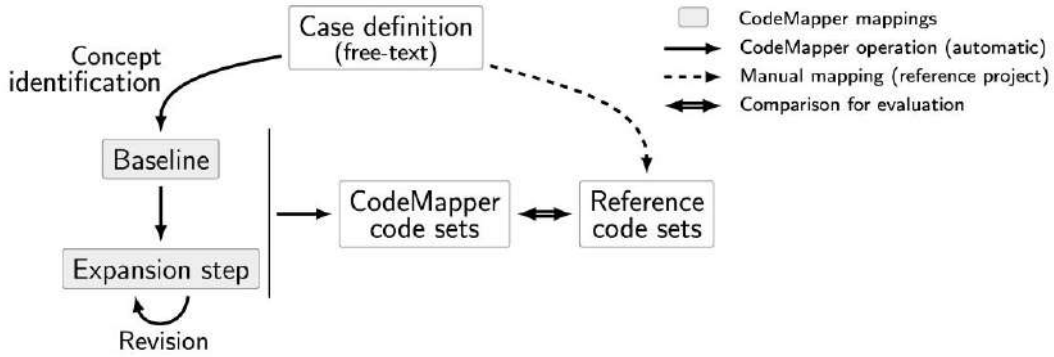


Figure 9.3: Automatic evaluation of CodeMapper. Reference code sets were created manually for each targeted vocabulary from the free-text case definition of an event. The baseline mappings and expansion steps were generated automatically from the same case definition using the operations available in CodeMapper.

10.2.4. Error analysis

We then carried out an automatic error analysis of the false-positive and false-negative codes after the third expansion step (Figure 9.4). Error categories were defined based on the notion of sibling codes: two codes are siblings if they are associated with the same concept. For false negatives, we distinguished between codes that are not contained in the Metathesaurus and codes whose siblings are not in the reference sets. False positive codes were categorized as having or not having a true-positive sibling code.

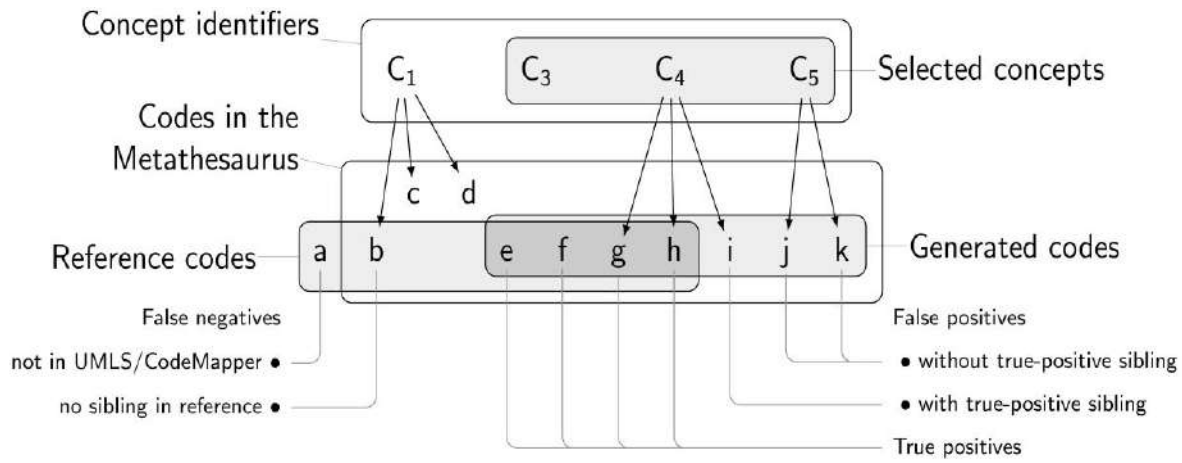


Figure 9.4: Categories of false negatives and false positives in the error analysis. Two codes are siblings if they are associated with the same concept.

10.3. Results

10.3.1. Baseline

The baseline mapping created by CodeMapper had an average sensitivity of 0.246 for reproducing the reference code sets (Table 9.2). The sensitivity in each vocabulary was inversely proportional to the number of reference codes in the vocabulary. The average PPV of the baseline mapping was 0.420. Without filtering by the semantic group of *Disorders*, the number of concepts would increase from 46 to 77 without affecting the sensitivity of the codes sets.

Table 9.2: Number of concepts and performance measures of the mappings in the evaluation. Numbers per vocabularies are macro-averages over all events.

Revision (concepts)		ICD-9	ICD-10	ICPC-2	READ-2	Average
Baseline (46)	Sensitivity	0.300	0.195	0.357	0.131	0.246
	PPV	0.387	0.380	0.500	0.411	0.420
Expansion step 1 (183)	Sensitivity	0.858	0.848	1.000	0.568	0.818
	PPV	0.483	0.558	0.762	0.729	0.633
Expansion step 2 (297)	Sensitivity	0.914	1.000	1.000	0.846	0.940
	PPV	0.463	0.509	0.762	0.749	0.621
Expansion step 3 (335)	Sensitivity	0.929	1.000	1.000	0.882	0.953
	PPV	0.462	0.498	0.762	0.742	0.616

10.3.2. Concept expansion

The sensitivity of the baseline mapping greatly improved in the first expansion step, to 0.818. Sensitivity further increased in the second (0.940) and third (0.953) expansion steps. All ICPC-2 codes were produced after the first expansion step and all ICD-10 codes were produced after the second step. The sensitivity increased incrementally for Read-2 and ICD-9 CM. The PPV improved after one expansion step (0.633) and decreased slightly after two (0.621) or three (0.616) expansion steps. The performance did not improve further in a fourth expansion step. The sensitivity was lower after three expansion steps when using only hierarchical relationships that were inherited from the source vocabularies (0.928) or defined in the Metathesaurus (0.879).

10.3.3. Error analysis

False-positive codes were generated in all vocabularies after the third expansion step (N=234, table 9.3). Most false-positive codes had true-positive siblings (N=164; 70.1%). False-positive codes without true-positive siblings (N=70; 29.9%) resulted from the initial concept identification step because the concept expansion steps (simulating the informed user) added only concepts with true-positive codes.

False-negative codes occurred only for Read-2 and ICD-9 CM (table 9.4). Most false negative codes did not have any sibling in the reference set (N=24; 68.6%), suggesting that the code was added to the reference set due to database specific needs. Other false-negative Read-2 codes were not contained in the conversion table from Read-CTv3 codes to Read-2 codes, or the Read-CTv3 codes corresponding with the Read-2 codes were not in the Metathesaurus (N=11; 31.4%).

A mapping constructed to maximize sensitivity by selecting concepts to generate all available codes from the reference sets had a sensitivity of 0.991 and PPV of 0.733.

Table 9.3: **Number of false-positive codes after three expansion steps by vocabulary and error category, and their percentage of all false-positive codes.**

Vocabulary	FP category	Count	Percentage
ICD-9 CM	With TP sibling	52	22.2%
	No TP sibling	22	9.4%
ICD-10	With TP sibling	66	28.2%
	No TP sibling	30	12.8%
ICPC-2	With TP sibling	3	1.3%
	No TP sibling	1	0.4%
READ-2	With TP sibling	43	18.4%
	No TP sibling	17	7.3%
Overall	With TP sibling	164	70.1%
	No TP sibling	70	29.9%

Table 9.4: Number of false-negative codes after three expansion steps by vocabulary and error category, and their percentage of all false-negative codes.

Vocabulary	FN category	Count	Percentage
READ-2	No sibling in reference	19	54.3%
	Not in UMLS	11	31.4%
ICD-9 CM	No sibling in reference	5	14.3%
Overall	No sibling in reference	24	68.6%
	Not in UMLS	11	31.4%

10.4. Discussion

In this report, we presented the CodeMapper web application that assists in the mapping of textual case definitions to code sets from multiple vocabularies, which is often a bottleneck in the implementation of epidemiological multi-database studies. We showed the effectiveness of CodeMapper's approach by simulating an informed usage of the application.

Creating a mapping only by the automatic identification of medical concepts in the case definition was insufficient to reproduce the reference code sets (sensitivity 0.246). The mapping process cannot be replaced by a simple indexing step. However, the goal of CodeMapper is to support an informed user to create such mappings, and CodeMapper's operations for concept expansion provide an effective and efficient way to do so. The reference code sets were regenerated with a sensitivity of 0.953 and PPV of 0.616 after only three expansion steps. Indeed, the reference codes for ICPC-2 were completely regenerated after the first expansion step and the reference codes for ICD-10 after only two expansion steps.

The mapping that simulates maximal sensitivity (0.991 with associated PPV of 0.733) forms an upper bound of CodeMapper's performance in regenerating the reference code sets. The imperfect sensitivity is due to reference codes that are missing in the UMLS or in the mapping between Read-2 and Read-CTv3. The moderate PPV may be due to inconsistencies in the reference code sets or the Metathesaurus. The reference code sets may be inconsistent between vocabularies for two reasons. First, the inclusion of one code in the reference mapping did not always imply the inclusion of all sibling codes in the targeted vocabularies, which is reflected by the large number of false positives with true-positive siblings. Second, different code sets were created for databases with the same vocabularies, which can be necessary to compensate for characteristics of the databases. For example, when an event is only available as an inpatient diagnosis in one database, a drug that is usually prescribed in case of the event in outpatient setting can be included in the query as a proxy. Such database-specific additions can also explain some false-negative codes without siblings in the reference set. Inconsistencies in the Metathesaurus such as missing identification of equivalent codes and incomplete coverage of vocabularies have been discussed before [22–25].

10.5. CodeMapper in ADVANCEWP5

CodeMapper was used in preparation for the event fingerprinting of the first proof-of-concept (POC) study of ADVANCE work package 5. Here we give a short description of the mapping process as an example of how CODEMAPPER is used in practice, the problems that we encountered, and possible improvements and recommendations about future use of CodeMapper in terminology mapping.

10.5.1. The intended mapping process

Events in the fingerprinting

The fingerprinting around events for the first POC study covered thirteen events: acute disseminated encephalomyelitis, convulsions, generalized convulsions, febrile convulsions, death, fever, hypotonic-hyporesponsive episodes, injection site reactions, persistent crying, pertussis, pneumonia, and somnolence.

Members of the ADVANCE consortium compiled documents with case definitions for these events in preparation of the protocols for the POC studies. The documents contained clinical case definitions, synonyms and lay terms, epidemiological criteria, descriptions of the laboratory and diagnosis tests, and the associated drugs and procedures. The information was based on publications of public health organizations, vaccine-safety organizations, disease-specific special-interest groups, and academic research. The documents are available on the ADVANCE SharePoint (WP5 | WP5.2 Fingerprinting | Fingerprinting events | Events).

Mapping process

The workflow for the terminology mapping in WP5 was designed as an iterative process (Figure 9.5). Four event teams were formed and assigned to create the terminology mapping of between two and four events each. Each event team comprised of at least the principal investigator of the study, a clinical expert for the event, a database custodian, and a person knowledgeable of CodeMapper. The event teams created the initial terminology mappings of the first iteration using CodeMapper's approach, i.e., by automatic identification of medical concepts in the clinical case definition of the event, followed by a discussion in the team of the identified concepts and corresponding codes, where deletions, additions, and expansions of concepts were decided. The decisions were based on the member's expertise, and on comparisons with existing code sets for the event that have been created in previous projects.

The codes derived from the medical concepts of the first iteration were then distributed to the databases for extraction, resulting, among other measures, in the incidence of each event in defined age groups and in the counts of individual codes in the databases. A comparison of the extracted information with data reported in the literature and between databases built the basis for adaptions to the selected medical concepts, and re-iteration of the process. The medical concepts that were selected from the first and second iteration for some example events are shown in Table 9.7.

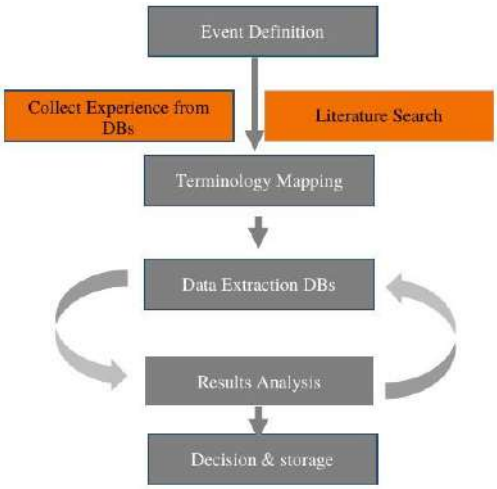


Figure 9.5: Workflow fingerprinting of events and component analysis (WP5)

Databases, vocabularies, vocabulary sets

The POC study targeted eight databases in four countries: The Health Improvement Network database (THIN) and Royal College of General Practitioners database (RCGP) in the UK; Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP) and Sistema d'Informació per al desenvolupament de la Investigació en Atenció Primària (SIDIAP) in Spain; the Aarhus university databases (AUH) and The Danish Civil and Health Registration System (SSI) in Denmark; the Pedianet database and ASL Cremona (ASLCR) in Italy (see Table 9.5).

The target databases use four different medical vocabularies to represent medical information: Read-v2 (THIN, RCGP), ICD-9 CM (Pedianet, ASLCR, BIFAP), ICD-10 CM (AUH, SSI, ASLCR), and ICPC-2 with local modifications (BIFAP). The following related UMLS vocabularies were targeted with CodeMapper: ICD9, ICD9CM, MTHICD9, ICD10, ICD10CM, ICPC, ICPC2EENG, and RCD. The codes from the UMLS vocabularies ICD9, ICD9CM, and MTHICD9 were combined for the distribution to the ICD-9 CM databases, the codes from the UMLS vocabularies ICD10 and ICD10CM were combined for the ICD-10 CM databases, and codes from the UMLS vocabularies ICPC and ICPC2EENG were combined for ICPC-2 databases. The RCD (Read-CTv3) codes were mapped to Read-v2 codes for THIN and RCGP using the mappings provided by HSCIC.

Table 9.5: Electronic health record databases targeted in the first proof of concept study in WP5.

Database	Country	Vocabulary	UMLS Vocabularies
THIN	United Kingdom	Read-v2	RCD + mapping

RCGP	United Kingdom	Read-v2	RCD + mapping
BIFAP	Spain	ICD-9 CM, ICPC-2[a,b)]	ICD9, ICD9CM, MTHICD9, ICPC, ICPC2EENG
SIDIAP	Spain	ICD-10 CM	ICD10, ICD10CM
SSI	Denmark	ICD-10 CM	ICD10, ICD10CM
AUH	Denmark	ICD-10 CM	ICD10, ICD10CM
Pedianet	Italy	ICD-9 CM[b)]	ICD9, ICD9CM, MTHICD9
ASLCR	Italy	ICD-9 CM	ICD9, ICD9CM, MTHICD9

[a)] Use of local adaptions to ICPC-2
[b)] Additional free-text queries for case identification

10.5.2. Observations and recommendations

A number of observations about the practicality of CodeMapper could be made in the process of terminology mapping for the ADVANCE event fingerprinting.

Prospective and operational case definitions

The initial terminology mapping in the preparation of the POC study was based on clinical case definitions of the events, such as for example the Brighton Collaboration case definitions which are often used in Vaccine Safety research to classify the certainty of an event. A clinical or Brighton Collaboration case definition states whether the conditions of person qualify as a case of a medical event. It is based on diagnostic criteria (signs & symptoms), personal criteria, and contextual criteria that are often combined by logical operators like AND or OR. A clinical case definition is best used for prospective research in that its application requires the availability or accessibility of the relevant criteria.

In the event teams we found that clinical case definitions do not form the best basis for extracting medical events from EHR databases for three reasons:

- First, information in most HER is recorded for other purposes than clinical research. It is a health record of a general practitioner, or a claims record that is collected in regular care. These data can be used for epidemiological studies, but this is always retrospective and information that is captured differs between the different databases (see WP 3 AIRR survey). What often can be retrieved is the diagnosis or a clear sign or symptom. However information about the diagnostic criteria that were used for the diagnosis is often unavailable, especially in claims databases. Therefore not all EHR databases may provide sufficient information to

replicate a diagnosis retrospectively, since the test results that are often required in Brighton/clinical definitions may not be captured as such.

- Second,for the POC studies initial instructions were to extract the codes supplied by the event teams and these events were considered to be cases in the initial harmonization runs. Since the clinical case definitions were used as the basis for terminology mappings, the initial mappings resulted in the inclusion of non-specific signs and symptoms. Clearly a symptom that is associated with a disease/event in solitude does not mean that the patient had the event. Sign/symptoms may be used for confirmation of a diagnosis code, but as standalone cannot replace the diagnosis. In the subsequent harmonization runs these non-specific symptoms were removed. It became clear that these symptoms/signs coming from the clinical case definitions can be used as components but this requires more complex algorithms that are currently being developed and tested in the Component algorithm group as part of the heterogeneity task force of WP4.

- Third, each clinical case definition that was used in the terminology mapping process focused on one medical event, but depending on the granularity of the medical terminologies, an event(e.g., ICD-10: A37 Whooping cough) can be represented by several codes that differentiate, for example by the causative agent of the condition (e.g., A37.0 Whooping cough due to Bordetella pertussis, A37.1 Whooping cough due to Bordetella parapertussis), by the presence or absence of accessory symptoms (A37.00 Whooping cough due to Bordetella pertussis without pneumonia, A37.01 with pneumonia), or by contextual criteria. However, the clinical case definition does not make a statement about which subtypes of the event should be extracted, or the circumstances of the event that may be represented in the diagnostic code.

Example of impact of inclusion of symptoms

The effect of the inclusion of diagnostic criteria in the event extraction is exemplified by the initial code counts for *Injection site reactions* (ISR) from SIDIAP (Table 9.6) as supplied by the event teams. The mapping of the first iteration contained concepts that are part of the clinical case definition but lack the localization of the condition (e.g., cutaneous abscess, cellulitis, or edema). Codes associated with these concepts account for 80% of the counts of the code set in SIDAP, which lowered the positive predictive value of the case identification algorithm considerably. These concepts were removed in the second iteration.

Table 9.6: Number of occurrences of codes for Injection site reactions in SIDIAP database using the code set from the first and second iteration.

Concept name	Code	Code count	
		Iteration 1	Iteration 2
Cutaneous abscess, furuncle and carbuncle, unspecified	L02.9	159387	-
Localised swelling, mass and lump, unspecified	R22.9	144334	-
Localized edema	R60.0	92352	85942
Cellulitis	L03.9	73819	-
Cutaneous abscess, furuncle and carbuncle, unspecified	L02	63198	-
Cellulitis	L03	29448	-
Localized swelling, mass and lump of skin and subcutaneous tissue	R22	3919	2256
Localized swelling, mass and lump of skin and subcutaneous tissue	R22	3919	795
Localized swelling, mass and lump of skin and subcutaneous tissue	R22	3919	265
Localized swelling, mass and lump of skin and subcutaneous tissue	R22	3919	209
Localized swelling, mass and lump of skin and subcutaneous tissue	R22	3919	138
Edema	R60.9	2978	-
Changes in skin texture	R23.4	1983	1374
Granulomatous disorder of the skin and subcutaneous tissue	L92.9	612	238
Localized swelling, mass and lump, lower limb	R22.4	472	316
Localized swelling, mass and lump, upper limb	R22.3	423	265
Granulomatous disorder of the skin and subcutaneous tissue	L92	417	2318
Granulomatous disorder of the skin and subcutaneous tissue	L92	417	896
Granulomatous disorder of the skin and subcutaneous tissue	L92	417	509
Granulomatous disorder of the skin and subcutaneous tissue	L92	417	265
Granulomatous disorder of the skin and subcutaneous tissue	L92	417	101
Granulomatous disorder of the skin and subcutaneous tissue	L92	417	56
Intra-abdominal abscess following a procedure	T81.4	1983	-
Cellulitis	L03.90	0	-
Localized swelling, mass and lump, unspecified upper limb	R22.30	0	0
Localized swelling, mass and lump, right upper limb	R22.31	0	0
Localized swelling, mass and lump, left upper limb	R22.32	0	0
Localized swelling, mass and lump, unspecified lower limb	R22.40	0	0
Localized swelling, mass and lump, right lower limb	R22.41	0	0
Localized swelling, mass and lump, left lower limb	R22.42	0	0
Localized swelling, mass and lump, unspecified	R22.9	-	139508

Example of impact of different granularity/diagnosis types

The effect of the differentiation between diagnosis types is can be seen in the extraction results for the event *Fever* from the SSI database (Figure 9.6). The concept "Pyrexia during labor" conforms to the case definition of fever that was used and was included the first output of the terminology mapping by the event team. In the SSI database the associated ICD-10 CM code (O75.2) was accountable for 6.7% of the identified cases of fever which appear as a peak of the incidence rate within women in their thirties after the first iteration (Figure 9.6, left). Upon review of the codes by

vaccine safety epidemiologists it was clear that this subtype should not be considered a potential vaccine safety outcome, as there is a clear other (necessary) cause, where vaccination cannot contribute as component and was therefore excluded in the second iteration. The incidence rates resulting from the second extraction (Figure 9.6, right) conformed to the incidence rates reported in the literature.

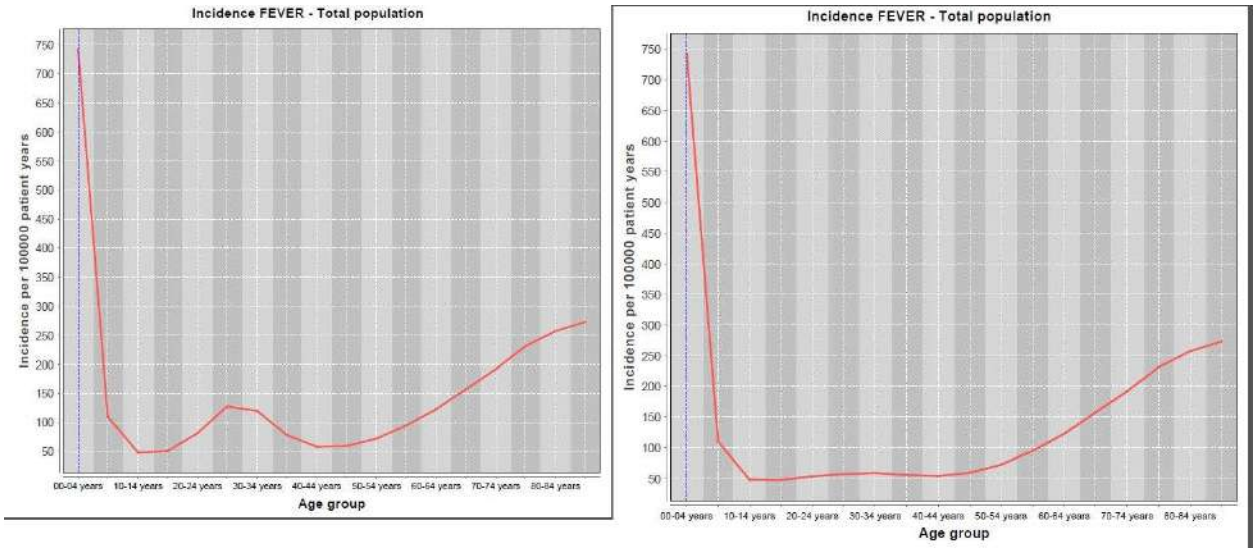


Figure 9.6: Incidence of fever in SSI database according to the codes from the first iteration (left) and from the second iteration (right) of the terminology mapping.

Unbalanced codes sets

UMLS concepts are an identity of equivalent codes in different medical vocabularies. CodeMapper's approach is based on the assumption that the projection of concepts to vocabularies generally results in balanced code sets, i.e. code sets with representative codes from each target vocabulary. However, when vocabularies differ in granularity or in the conceptualization of medical events, the projections of specific UMLS concepts can result in unbalanced code sets. During terminology mapping using CodeMapper, it is necessary to identify such concepts and include additional concepts that provide appropriate codes in vocabularies that previously had no codes in the projection.

CodeMapper was extended to identify concepts with unbalanced code sets automatically (Figure 9.7). The code sets of groups of concepts that have been defined by assigning a common tag are combined in the analysis. The user is requested to add corresponding concepts to the mapping, to ensure the generation of balanced code sets with CodeMapper.

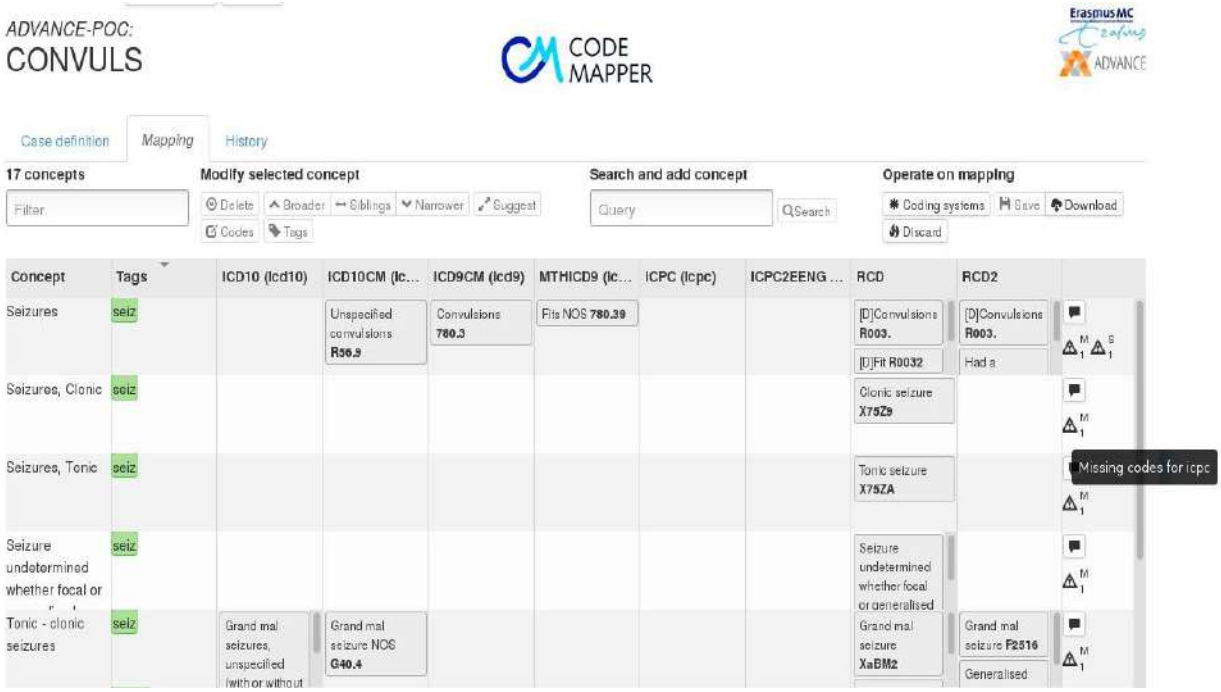


Figure 9.7: Automatic identification of unbalanced code sets within groups of concepts that are defined by shared tags, and groups of target vocabularies. The warning sign on the right indicates concepts with unbalanced code sets.

Code usage in databases

Efficient terminology mapping depends on information about the usage of individual codes in the databases for two reasons:

- First, information about the absence of codes can cut short a discussion about the inclusion or exclusion of some concepts in the mapping. A concept that is contested in the event team but whose codes are not used in the database scan safely be excluded from the mapping.
- Second, a measure of the incidence of codes in the databases can point to differences in the representation of events between databases. When the occurrence of one code is lower than the occurrence of an equivalent code in a different database, this can suggest that the event is represented by more general or more specific codes in the other database.Medical concepts that correspond to these codes should then be explored.

The retrieval of information about code usage in the databases currently requires a full cycle of data extraction, but the information is not specific to the terminology mapping of one event. Rather, an occurrence measure of all codes should be collected from all databases beforehand and displayed in the CodeMapper application to inform the mapping process.

Beyond UMLS

CodeMapper's current approach was insufficient to fully specify the codes for events in some databases, and required database-specific post-processing of the code sets, mainly due to specific characteristics of participating databases:

- First, a local extension of the ICPC-2 vocabulary is used in the BIFAP database. The local codes are not included in the UMLS, and could therefore not be generated with CodeMapper. Instead, the database custodians of BIFAP added local codes that correspond to the UMLS concepts of each event after the distribution of code sets.
- Second, some databases comprise unstructured information such as text; the generated code sets were complemented by the database custodians of BIFAP and Pedianet by keyword-based queries on the free-text portion of the EHRs.
- Third, some primary-care databases do not encode the event of death. The information is then only available through connected administrative databases that record the reason for exit of registration.

These changes for database-specific characteristics were not recorded by CodeMapper, which should be extended to include the storage and documentation of local codes, descriptions of special representation of events, and storage of keyword queries.

Event teams

The organization of the four event teams represented a bottleneck in the terminology mapping process. The scheduling of meetings (as telephone conferences) with all members of the event teams was difficult, which prolonged the mapping process unnecessarily. Additionally, previous experience in terminology mapping and experience in extraction of events from healthcare databases is required to estimate the decisions in the mapping process, but was lacking in some teams.

The process of terminology mapping requires clinical knowledge about the event, epidemiological knowledge including an understanding of the study goals, and knowledge about the representation of medical information in databases. Smaller teams where single members contribute expertise in several of these topics could perform the terminology mapping of the events more efficiently. Each team should be led by the principal investigator of the study (who needs to have expertise in extraction of events from healthcare databases) who has the objective of the study in view and takes final decisions about the exclusion or inclusion of medical concepts based on the input of the event team.

Knowledge management

The CodeMapper web application constitutes a single entry point for the different phases of the terminology mapping process. The application first assists in creating an initial mapping based on textual case definitions, and in adjusting it by exploring relevant codes. It then generates the codes for an event from the targeted medical vocabularies. Finally, its record of applied user operations and summaries constitute an essential and important documentation and reasoning of the terminology mapping process.

Tools for post-processing

Several additional tools have been developed during the terminology mapping for the POC study. These tools, including additional documentation, are available in the CodeMapper folder on the ADVANCE SharePoint.

The script called *compile.py* combines the code sets of multiple events into one XLS file for easy distribution to the databases. The name of the input files must be the event name (+ .xls extension). Each event is in a separate worksheet of the resulting Excel file. A vocabulary map is used to assign codes from a set of CodeMapper/UMLS vocabularies to databases. Codes that correspond to concepts with tags are grouped together. Only tags that start with an upper-case letter are taken into consideration.

The script called *code-counts.py* connects the code counts that resulted from the event fingerprinting with the code sets that were generated with CodeMapper. It runs on the code counts from all events of one database at once.

The script called *stack-mappings.py* combines a set of CodeMapper mappings in XLS files into one XLS file, in long-form, adding a column to indicate the event.

Recommendations for next steps and future developments based on learnings in the POC-1 study

We recommend that the terminology mapping process:

- Should be based on an operational definition of the event that specifies at a diagnostic level the types of medical events that are relevant in the context of the study, including possible proxy variables or exclusion criteria.
- Events that are coded to their cause, which is a necessary and sufficient cause, should not be included as codes for vaccine-related outcomes, since the vaccination could not contribute to that causation. However different databases have differences in coding granularities, whereas the cause for the event may exist in one dictionary but may not be in the other, leading to heterogeneity in event definitions. The impact can be studied by comparing the population based age specific incidence rates.

- Diagnostic criteria should be included only for the confirmation of a diagnosis, or to estimate the onset of an event. Concepts that represent diagnostic criteria should be tagged as such in CodeMapper, to inform the database custodians during extraction about the secondary role of these codes.
- The mapping process could be made more efficient if all database provide an output with the counts of each code. This could be displayed in the CodeMapper application to inform the mapping process.
- CodeMapper could be extended to the storage and documentation of local codes, descriptions of special representation of events, and storage of keyword queries.
- Eventteams should be led by database pharmacoepidemiologists and not clinicians, as the key knowledge that is required in the mapping process is familiarity with the representation of disease and concepts in electronic healthcare databases. While clinical expertise is crucial, one should be aware that clinicians think that data are collected prospectively and therefore would assume you can still decide whether a case is a case based on signs/symptoms. This may be a pitfall.
- If code sets are already available in one vocabulary or a Standard Medical Queries of MedDRA exists, CodeMapper could be used to identify UMLS concepts that correspond to the codes, and to project the concepts to all other target vocabularies. This approach still requires further exploration because of the differences in granularity as shown above.

10.5.3. Medical concepts selected in terminology mapping for the POC study in ADVANCE WP5

Table 9.7: Medical concepts used in the terminology mapping of Fever and Injection site reactions (ISR) for the POC study in ADVANCE WP5. Concepts that were removed between the first and second iterations are highlighted in red, concepts that were added between the first and second iterations are highlighted in green.

Event	Concept name
FEVER	Drug induced fever
FEVER	Drug-induced hyperpyrexia
FEVER	Fever
FEVER	Fever of the newborn
FEVER	Fever of Unknown Origin
FEVER	Fever presenting with conditions classified elsewhere
FEVER	Fever symptoms (finding)
FEVER	Fever with chills
FEVER	Fever with rigors
FEVER	Fever with sweating
FEVER	Hyperpyrexia

Event	Concept name
FEVER	Low grade fever
FEVER	O/E - fever
FEVER	O/E - temperature elevated
FEVER	Other specified fever
FEVER	Persistent fever
FEVER	Post vaccination fever
FEVER	Postimmunization fever
FEVER	Postprocedural fever
FEVER	Sweating fever
FEVER	Disturbance of temperature regulation of newborn, unspecified
FEVER	Febrile nonhaemolytic transfusion reaction
FEVER	Pyrexia during labor, not elsewhere classified in ICD10CM
FEVER	Pyrexia postprocedure
FEVER	Fever and other physiologic disturbances of temperature regulation
FEVER	Fever of other and unknown origin
ISR	Abnormally hard consistency
ISR	Administration site reaction
ISR	Changes in skin texture
ISR	Granulomatous disorder of the skin and subcutaneous tissue
ISR	Induration of skin
ISR	Injection site edema
ISR	Injection site erythema
ISR	Injection site reaction
ISR	Local reaction
ISR	Localized edema
ISR	Localized swelling, mass and lump of skin and subcutaneous tissue
ISR	Localized swelling, mass and lump, left lower limb
ISR	Localized swelling, mass and lump, left upper limb
ISR	Localized swelling, mass and lump, lower limb
ISR	Localized swelling, mass and lump, right lower limb
ISR	Localized swelling, mass and lump, right upper limb
ISR	Localized swelling, mass and lump, unspecified lower limb
ISR	Localized swelling, mass and lump, unspecified upper limb
ISR	Localized swelling, mass and lump, upper limb
ISR	Peeling of skin
ISR	Pyogenic granuloma of skin and subcutaneous tissue
ISR	Subcutaneous nodule
ISR	subcutaneous nodules (localized)(superficial)
ISR	Superficial swelling
ISR	Abscess
ISR	Cellulitis
ISR	complications; vascular
ISR	Cutaneous abscess, furuncle and carbuncle, unspecified

Event	Concept name
ISR	Edema
ISR	Granuloma
ISR	Intra-abdominal abscess following a procedure
ISR	Local superficial swelling, mass or lump NOS
ISR	Localised swelling, mass and lump, unspecified
ISR	Mass of body structure
ISR	Sepsis following a procedure
ISR	Stitch abscess following a procedure
ISR	Subphrenic abscess following a procedure
ISR	Swelling
ISR	Thickening of skin
ISR	Wound abscess following a procedure
ISR	Localized superficial swelling, mass, or lump
ISR	Localized swelling, mass and lump, unspecified
ISR	Thick skin

10.6. References

1. Trifirò G, Coloma P, Rijnbeek P, et al.Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: Why and how? J Intern Med 2014; **275**: 551–561.
2. de Lusignan S, Minmagh C, Kennedy J, Zeimet M, Bommezijn H, Bryant J. A survey to identify the clinical coding and classification systems currently in use across Europe. Stud Health Technol Inform 2001: 86–89.
3. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. J Innov Health Inform 2005; **13**: 65–69.
4. US Department of Health and Human Services and others. ICD 9 CM. The International Classification of Diseases. 9. Rev: Clinical Modification.; Vol. 1: Diseases: Tabular List. ; Vol. 2: Diseases: Alphabetic Index. ; Vol. 3: Procedures: Tabular List and Alphabetic Index. US Government Printing Office, 1980.
5. Pavillon G, Maguin P. The 10th revision of the International Classification of Diseases. Rev Epidemiol Sante Publique 1992; **41**: 253–255.
6. Lamberts H, Okkes I, others. ICPC-2, International Classification of Primary Care. 1998.
7. Read JD, Sanderson HF, Sutton YM. Terming, Encoding, Grouping, The Language of Health. In: Proceedings International Medical Information Association's 8th World Congress on Medical Informatics. Vancouver, 1995.
8. O'Neil M, Payne C, Read J. Read Codes Version 3: a user led terminology. Methods Inf Med 1995; **34**: 187–192.
9. Gini R, Schuemie M, Brown J, et al.Data Extraction and Management In Networks Of Observational Health Care Databases For Scientific Research: A Comparison Among EU-ADR, OMOP, Mini-Sentinel And MATRICE Strategies. EGEMs Gener Evid Methods Improve Patient Outcomes 2016; **4**.
10. Avillach P, Joubert M, Thiessard F, et al. Design and evaluation of a semantic approach for the homogeneous identification of events in eight patient databases: A contribution to the European EU-ADR project. Stud Health Technol Inform 2009; **160**: 1085–1089.

11. Avillach P, Coloma PM, Gini R, 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**: 184–192.
12. Coloma PM, Schuemie MJ, Trifirò G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf 2011; **20**: 1–11.
13. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: Proceedings of the AMIA Symposium. American Medical Informatics Association, 2001; 17.
14. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med 1993; **32**: 281–291.
15. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform 2001; **84**: 216.
16. Rogers F. Medical subject headings. Bull Med Libr Assoc 1963; **51**: 114.
17. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf 1999; **20**: 109–117.
18. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006; **121**: 279.
19. Schuemie MJ, Jelier R, Kors JA. Peregrine: Lightweight gene name normalization by dictionary lookup. In: Proc of the Second BioCreative Challenge Evaluation Workshop., 2007; 131–133.
20. Bean A, Green R. Relationships in the Organization of Knowledge. Springer Science & Business Media, 2013.
21. Schmedt N, Bezemer I, Berardis GD, et al.Benchmarking of Definitions and Algorithms to Extract Outcome Events in the EU-funded SAFEGUARD Project. 2013. Available at: http://www.safeguard-diabetes.org/?q=system/files/private/SAFEGUARD_DGEpi_2013_Benchmarking_FINAL.pdf. Accessed August 11, 2016.
22. Bodenreider O, Burgun A, Botti G, Fieschi M, Le Beux P, Kohler F. Evaluation of the Unified Medical Language System as a medical knowledge source. J Am Med Inform Assoc 1998; **5**: 76–87.
23. Cimino JJ, Min H, Perl Y. Consistency across the hierarchies of the UMLS Semantic Network and Metathesaurus. J Biomed Inform 2003; **36**: 450–461.
24. Erdogan H, Erdem E, Bodenreider O. Exploiting UMLS semantics for checking semantic consistency among UMLS concepts. Stud Health Technol Inform 2010; **160**: 749.
25. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. J Biomed Inform 2012; **45**: 689–696.

11. Appendices

11.1. Appendix P1 (coverage)

DATA COLLECTION AND COVERAGE ESTIMATION METHODOLOGIES

Netherlands

Pertussis

Since 2006 the pertussis vaccination status is determined on an individual level according to the vaccination schedule and no longer on a fixed date, independent of the exact age of the child. As an example, in the reporting year 2015 (coverage estimates up to December 2014), the coverage estimates are given for children from the 2004, 2009 and 2012 birth cohorts. These are children that by December 2014 have all reached 10, 5 and 2 years ofage, respectively. At the age of 2 children should have received 4 pertussis-containing vaccinations, at the age of 5 children should have received their 5thpertussis-containing vaccine and at the age of 10 children should have received their 6thpertussis-containing vaccine, according to the national vaccination schedule. The denominator includes all registered children in the respective birth cohorts (based on calendar year of birth).

Influenza

For influenza the coverage is estimated using a database that holds data from approximately 500 GP's in the Netherlands which represents approximately 1.5M registered people. For the GP practices in the database to be included in the coverage estimation a number or quality criteria needed to be met. As an example, at least 70% of morbidity records needed to be coded according to ICPC standards.

For the denominator all persons of **60 years or older** in the database pool are included. This included all people who were 60 years or older on 1 May, as determined by NHG (national society for family doctors). For calculating the age as a characteristic the cut-off date of 1 January is used, which means that a number of 59-year-olds were added to the target group of 60+. If date or month of birth is unknown, 1 June is used.

The vaccination status is determined based on the use of the ICPC code (R44) for an immunization or the ATC code (J07BB02) for a prescription of an influenza vaccine. If registration of either one of these codes is found between 1 September until 31 December, vaccination status was confirmed positive.

Denmark

Pertussis

For the pertussis-containing vaccines the numerator is set as the number of children within each birth cohort (based on the calendar year of birth) who received the vaccine and the denominator is the number of children within the birth cohort who at the time of the coverage estimation were living in Denmark. When a child is vaccinated the GP registers an administrative code. This code specifies the dose number. From 2012 onwards the **date** of vaccination is registered; before 2012 it was **week** of vaccination. The date the vaccination is registered in the system is also available. The vaccinations are entered into a separate system and although we ask for date of vaccination we cannot rule out that sometimes he date of registration is used instead of the date of vaccination.The single booster vaccination is recommended at 5 years of age.

Influenza

For influenza the numeratoris the number of vaccines administered to individuals aged 65 years and above (age registered at time of vaccination) and the denominator is the number of individuals aged 65 years and above who at the time of the coverage estimation were living in Denmark. Coverage is estimated on a monthly basis so each month it is determined how many individuals are 65 years and above and how many of those received the seasonal influenza vaccine. For administrative reasons, the flu season starts in week 40 one year and ends in week 20 the following year. Individuals 65 years and above are vaccinated free of charge from 1 October to 31 December.

United Kingdom

Pertussis

- The Cover of Vaccination Evaluated Rapidly (COVER) programme by Public Health England (PHE) is a quarterly data collection.
- The majority of COVER data is extracted from Child Health Information Systems(CHISs) using a standardised output.
- CHIS IT suppliers who provide services and maintenance for the majority of Child Health Information Departments use an Information Standard Notice (ISN) to build the specification for data outputs.

(https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/383125/SCCI_0089_COVER_User_Guide.pdf)

Definition for one year

Child received3 doses before their 1st birthday; if child received primary immunisations outside UK then 3 doses of each: DTP or DTaP, IPV or OPV, Hib before 1st birthday,

Numerator: Of those children who reached their 1st birthday within the evaluation dates, the number that completed a course of DTaP/IPV/Hib vaccine.

Definition for 5 years

Child received 3 doses of DTaP/IPV/Hib vaccine before their 5th birthday. If child received primary immunisations outside UK then 3 doses of each: DTP or DTaP, IPV or OPV before 5th birthday

Numerator: Of those children who reached their 5th birthday within the evaluation dates, the number that a completed a course of DTaP/IPV/Hib vaccine.

The reporting year runs from April to March and children are included in the denominator if they have reached their respective birthdays within that evaluation period (April-March). Therefore the data does not necessarily reflect a birth cohort per calendar year, but a birth cohort from April-March of each year. As data is collected per quarter, it would be possible to add the quarterly data from one single calender year.

Influenza

- The COVER programme by PHE
- Seasonal Influenza Frontline Healthcare Workers Vaccine Uptake Survey (through web-based system – ImmForm)

Numerator: Patients registered at all GP practice in the UK

Spain

Pertussis

For pertussis, vaccine coverage is calculated using the vaccine doses administered in the government community clinics. Private sector doses are not included.

In Spain, each autonomous community (region) reports its data to the Ministry of Health. Each region has its own database to identify and register this information.

The vaccine coverage percentage includes:

1) First dose vaccination: Percentage of children younger than one year of age that have received three doses of DTPa vaccine

2) Booster vaccination:

- A) Percentage of childrenaged1-2 years that have received a booster dose of DTPa
- B) Percentage of childrenaged4-6 years that have received a booster dose of dTpa

Influenza

For influenza, vaccine coverage is calculated using the vaccine doses administered in the government community clinics. Private sector doses are not included.

In Spain, each autonomous community (region) reports its data to the Ministry of Health. Each region has its own database to identify and register this information. Each region defines the risk groups that should be vaccinated for influenza. Vaccine coverage information available at the Ministry of Health is for elderly people (≥65 years).

The vaccine coverage percentage includes the percentage of adults aged 65 years or older who have received one dose of influenza vaccine in the corresponding influenza season.

World Health Organization (WHO)

Immunization coverage

Immunization coverage levels are presented as the percentage of a target population that has been vaccinated. Coverage is usually calculated for each vaccine and for the number of doses received. The target population varies depending on national policies, the specific vaccine and the dose for which coverage is being calculated. The estimates refer to immunizations given during routine immunization services to children less than 12 months of age where immunizations are recorded.

Methods

- The administrative method using reported routine immunization data, i.e. registry system of doses administered
- Immunization coverage surveys using survey methods recommended by WHO. Surveys should be conducted periodically (3-5 years)

Administrative method

In most countries "administrative coverage data" are the number of doses administered to the target population. In order to estimate percentage immunization coverage, this number is divided by the total estimated number of people in the target population.

The target population groups vary from country to country and are dependent on the national immunization schedules in place.

The coverage estimates calculated using the administrative method can be biased due to inaccurate numerators or denominators.

Numerators may be:

- underestimated (due to incomplete reporting from reporting units or non-inclusion of other vaccinationsources (e.g. private sector, non-governmental organizations), or
- overestimated (due to over-reporting from reporting units e.g. inclusion of other target groups).

Denominator inaccuracies may be due to issues such as:

- population movement
- inaccurate census estimations or projections and/or
- numerous sources of denominator data

Survey methods

Surveys aim to estimate the levels of immunization coverage at either national or regional levels. They aim to either establish baseline information and to provide a comparison with administrative estimates (to verify administrative coverage data), while efforts to improve routine reporting systems are ongoing, andthey can be used to respond to specific questions regarding factors associated withcoverage or to satisfy information demands of the partner agencies.

Although the primary objective of an immunization coverage survey is to provide a coverage estimate for selected vaccines or a set of vaccines (fully vaccinated for age) among infants, children, women of childbearing age, etc, other information, which is usually not available through routine monitoring systems, can be obtained.

Furthermore, surveys facilitate assessing equity in immunization, by allowing disaggregating coverage by factors such as place of residence, sex, maternal education, economic status or subnational region.

Vaccination coverage surveys can be complemented with serosurveys. Serosurveys can help establish baseline prevalence of a vaccine-preventable disease prior to the introduction of a particular vaccination policy, or to assess the impact of such vaccine programme. Serosurveys have been mainly used for Hepatitis B and measles.

The WHO vaccination coverage survey

Since the early 1990s, the WHO has provided guidance to Member States, partner agencies and institutions on methods for measuring immunization coverage through surveys and has provided

manuals and tools to conduct the EPI cluster survey. With the goal of improving survey precision, accuracy, and overall quality, an extensive review and revision of coverage survey methods and materials resulted in the release, in 2015, of the working draft of WHO Vaccination Coverage Cluster Survey Reference Manual.

While the statistical methods outlined in the new Survey Reference Manual are commonly used on large household health surveys, such as DHS and MICS, immunization programmes may be less familiar with them. Therefore, WHO is preparing tools to facilitate the management, analysis, presentation and interpretation of survey. One of these tools, "Vaccination Coverage Quality Indicators (VCQI)" is set of scripts in STATA and R, a program intended to be used by data managers to manage entering and cleaning survey data; statisticians and epidemiologist to analyse survey data; and for programmers to add further modifications and additional analysis. A beta version of VCQI can be obtained on request by emailing vpdata@who.int.

Other survey types used to monitor vaccination coverage include:

- international household survey initiatives, such as UNICEF Multiple Indicator Cluster Survey (MICS) and the Demographic and Health Surveys (DHS)Program.
- the Lot Quality Assurance (LQA) technique, mostly used to assess polio supplementary immunization activities (vaccination campaigns).

Data Quality Self-assessment

The Data Quality Self-assessment (DQS) is a flexible toolbox of methods used to evaluate different aspects of the immunization monitoring system at district and health unit (HU) levels.

It is designed by and for staff collecting and using immunization data at national, provincial or district levels. It aims to assist countries in diagnosing problems relatedto data collection and providing orientation to improve district monitoring as highlighted in the Reaching Every District (RED) approach.

The DQS aims to determine:

- the accuracy of reported figures for coverage (i.e. number of immunizations) and for any other immunization system indicator.
- the quality of any component of the immunization monitoring system

The assessment includes a review of data accuracy at different levels and a self-designed questionnaire reviewing monitoring quality issues (e.g. availability of vaccination cards, use of tally

sheets, directly-observed recording and reporting practices). Data are then analysed with a view to identifying strengths and weaknesses which need to be corrected

WHO/UNICEF estimates of national immunization coverage

WHO and UNICEF have reviewed data available on national immunization coverage and made country-specific estimates of immunization coverage

Estimates were made for BCG, the third dose of diphtheria-tetanus-pertussis vaccine (DTP3), the third dose of either oral polio vaccine or inactivated polio vaccine (Pol3), the first dose of measles vaccine (MCV), and the third dose of hepatitis B vaccine (HepB3). Estimates have also made of the proportion of live births protected (PAB) through maternal immunization with at least two doses of tetanus toxoid for countrieswhere the risk of neonatal tetanus is a significant public health problem for the year 2000 onward. In 2005 estimates of the first dose of diphtheria-tetanus-pertussis vaccine (DTP1) and the third dose of *Haemophilus influenzae* type b (Hib3)were added.

These estimates are based on data officially reported to WHO and UNICEF by Member States as well as data reported in the published and grey literature. Whenever possible local experts - primarily national EPI managers and WHO regional office staff - were consulted for additional information regarding the performance of specific local immunization services. Based on the data available, consideration of potential biases, and contributions from local experts attempts to determine the most likely true level of immunization coverage were made. Coverage data are reviewed and the estimates updated annually.

'Grade of confidence'

The GoC reflects the degree of empirical support upon which the estimates are based. It is not a judgment of the quality of data reported by national authorities.

• • • Estimate is supported by reported data [R+], coverage recalculated with an independent denominator from the World Population Prospects: 2012 revision from the UN Population Division (D+), and at least one supporting survey within 2 years [S+]. While well supported, the estimate still carries a risk of being wrong.

• • Estimate is supported by at least one data source; [R+], [S+], or [D+]; and no data source, [R-], [D-], or [S-], challenges the estimate.

•There are no directly supporting data; or data from at least one source; [R-], [D-], [S-]; challenge the estimate.

WHO	**Pertussis containing vaccine (children)**

	DPT1: First dose of diphtheria toxoid, tetanus toxoid and pertussis vaccine				
	UK*	**NL**	**ES**	**IT**	**DK**
References	http://www.who.int/immunization/monitoring_surveillance/data/gbr.pdf http://apps.who.int/immunization_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=GBR	http://www.who.int/immunization/monitoring_surveillance/data/nld.pdf http://apps.who.int/immunization_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=NLD	http://www.who.int/immunization/monitoring_surveillance/data/esp.pdf http://apps.who.int/immunization_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=ESP	http://www.who.int/immunization/monitoring_surveillance/data/ita.pdf http://apps.who.int/immunization_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=ITA	http://www.who.int/immunization/monitoring_surveillance/data/dnk.pdf http://apps.who.int/immunization_monitoring/globalsummary/countries?countrycriteria%5Bcountry%5D%5B%5D=DNK
Coverage rate per birth cohort	DTP1 figures per year (1980 to 2014) range from 66% in 1980 to 98% in 2014	DTP1 figures per year (1980 to 2014) range from 98% in 1980 to 99% in 2013	DTP1 figures per year (1984 to 2014) range from 91% in 1987 to 99% in 2014	DTP1 figures per year (1990 to 2014) range from 94% in 1990 to 99% in 2012	DTP1 figures per year (1980 to 2014) range from 94% in 1996 to 99% in 2002
a'Grade of confidence' GoC	From 2003 to 2014, GoC ● Estimates based on DTP3 coverages ; no accepted empirical data	From 2003 to 2014, GoC ● Estimates based on DTP3 coverages ; no accepted empirical data	From 2003 to 2014, GoC ● Estimates based on DTP3 coverages ; no accepted empirical data	From 2003 to 2014, GoC ● Estimates based on DTP3 coverages ; no accepted empirical data	From 2003 to 2014, GoC between ● and ●● Estimates based on DTP3 coverages in 2003; then estimate based on coverage reported by national government ; then survey for 2012-2014 estimates
Immunization schedule	http://vaccine-schedule.ecdc.europa.eu/Pages/Scheduler.aspx				

WHO	**Pertussis containing vaccine (children)** **DTP3: Third dose of diphtheria toxoid, tetanus toxoid and pertussis vaccine**				
	UK*	**NL**	**ES**	**IT**	**DK**
References	http://www.who.int/immuniza	http://www.who.int/immuniza	http://www.who.int/immuniza	http://www.who.int/immuniza	http://www.who.int/immuniza

	tion/monitoring_surveillance/data/gbr.pdf	tion/monitoring_surveillance/data/nld.pdf	tion/monitoring_surveillance/data/esp.pdf	tion/monitoring_surveillance/data/ita.pdf	tion/monitoring_surveillance/data/dnk.pdf
Coverage rate per birth cohort	DTP3 figures per year (1980 to 2014) range from 41% in 1980 to 95% in 2014	DTP3 figures per year (1980 to 2014) range from 96% in 1980 to 98% in 2005	DTP3 figures per year (1984 to 2014) range from 77% in 1987 to 98% in 2006	DTP3 figures per year (1990 to 2014) range from 83% in 1990 to 97% in 2007	DTP3 figures per year (1980 to 2014) range from 86% in 1984 to 99% in 1999
'Grade of confidence' GoC	From 2003 to 2014, GoC ●● Estimates based on coverage reported by national government	From 2003 to 2014, GoC ●● Estimates based on coverage reported by national government	From 2003 to 2014, GoC between ● and ●● Estimates based on coverage reported by national government	From 2003 to 2014, GoC between ● and ●● Estimates based on coverage reported by national government	From 2003 to 2014, GoC between ● and ●● Estimates based on coverage reported by national government; then survey for 2012-2014 estimates
Immunization schedule	http://vaccine-schedule.ecdc.europa.eu/Pages/Scheduler.aspx				

SUMMARIZED COVERAGE DATA

Pertussis containing vaccine(s)

2003 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	89%	98%		97%		99%		98%
2nd dose			88%							
3rd dose	94.3%	98%	88%	96%		91%	98.2%	98%	96.6%	96%
Booster	91.9%		84%				94.9%			
Booster 2	92.7%		n/a				92.3%			

2004 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	88%	95%		97%		99%		98%
2nd dose			88%							
3rd dose	94%	98%	87%	95%		92%	96.6%	97%	96.6%	94%
Booster	91.7%		83%				95%			
Booster 2	92.7%		n/a				88.3%			

2005 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	87%	93%		97%		98%		98%
2nd dose			87%							
3rd dose	94.5%	98%	86%	93%		91%	96.2%	96%	96.2%	95%
Booster	92%		83%				95.2%			
Booster 2			n/a				88.9%			

2006 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		98%	87%	93%		97%		99%		98%
2nd dose			87%							
3rd dose	95.2%	96%	87%	93%	93.2%[§]	92%	97.6%	98%	96.6%	96%
Booster	92.3%		84%		87.4%[±]		95.1%			
Booster 2			n/a				81.1%			

2007 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		98%	88%	87%		97%		98%		99%
2nd dose			88%							
3rd dose	95%	96%	88%	87%	93.4%§	92%	96,4%	96%	96.7%	97%
Booster	92.3%		84%		88.8%±		94,8%			
Booster 2			n/a				87.7%			

2008 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	90%	89%		97%		99%		98%
2nd dose			89%							
3rd dose	95.4%	97%	89%	88%	93.9%§	92%	96.7%	97%	96.7%	96%
Booster	92%		87%		89.6%		94.1%			
Booster 2			n/a				89%			

2009 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**Italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	91%	90%		97%		98%		98%
2nd dose			90%							
3rd dose	95.4%	97%	90%	89%	94.8%§	93%	95.9%	96%	96.2%	96%
Booster	91.9%		80%		89.3%		93.7%			
Booster 2			n/a				91.6%			

2010 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**Italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK*	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	92%	93%		98%		99%		98%
2nd dose			91%							
3rd dose	95.5%	97%	90%	90%	94.2%	94%	96.6%	97%	96.4%	96%
Booster			62%		88.3%		94.1%			
Booster 2			n/a				n/a			

2011 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**Italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK*	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	92%	94%		98%		99%		98%

2nd dose			91%							
3rd dose	95.4%	97%	90%	91%	94.7%	95%	97.1%	97%	96.3%	96%
Booster							93.1%			
Booster 2			n/a				n/a			

2012 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**Italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	92%	97%		98%		99%		99%
2nd dose			92%							
3rd dose	94.8%	97%	91%	94%	94.7%	95%	96.3%	97%	96.2%	97%
Booster							92.3%			
Booster 2			n/a				89%			

2013 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**Italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		99%	92%	97%		98%		98%		98%
2nd dose			91%							
3rd dose		97%	90%	94%	94.3%	95%	95.6%	96%	95.6%	96%
Booster							94.6%			
Booster 2			n/a				n/a			

2014 birth cohort	**Netherlands**		**Denmark**		**UK**		**Spain**		**Italy**	
	NL	NL (WHO)	DK	DK (WHO)	UK	UK (WHO)	ES	ES (WHO)	IT	IT (WHO)
1st dose		98%	93%	96%		98%		99%		98%
2nd dose			90%							
3rd dose		96%	77%	94%	94.5%	95%	96.6%	97%	94.6%	94%
Booster							n/a			
Booster 2			n/a				n/a			

*The UK reporting year runs from April to March. Therefore the 2006 birth cohort percentage will reflect a coverage estimate for the period April 2006-March 2007 for primary doses and April 2011-March 2012 for the booster and so on.
§ Scotland and Northern Ireland coverage estimates not included.
± Includes only data from England

Seasonal influenza vaccine (excluding pandemic vaccine) for 60+ or 65+

	Netherlands		Denmark		UK		Spain		italy	
Season	NL (60+)	NL (WHO)	DK (65+)	DK (WHO)	UK (65+)	UK (WHO)	ES (65+)	ES (WHO)	IT (65+)	IT (WHO)
2008/2009	76.9%						65.4%		66.2%	
2009/2010	76.3%		50%				65.7%		65.6%	
2010/2011	75.4%		47%				56.9%		60.2%	
2011/2012	71.3%		48%				57.7%		62.7%	
2012/2013	67.8%		45%				57.0%		54.2%	
2013/2014	65.7%		47%				56.4%		55.4%	
2014/2015	60.1%		45%				56.2%		49%	
2015/2016			44%		71%					

11.2. Appendix P6 (BR dashboard)

Benefit-risk dashboard: architecture

The architecture of the dashboard contains the following distinct steps (Figure S.1):

- Pre-processing 1: This step starts from the electronic healthcare data and transforms it into an individual level analytical dataset containing the exposure, outcome and covariate information of interest. For developmental purposes, we simulated the analytical dataset.
- Pre-processing 2: This step starts from the individual-level analytical data and transforms it into various data tables containing aggregated data needed to produce the charts displayed by the web-application. The data tables contain aggregated data such as weekly number of active patients and events by age groups, weekly number of active subjects by age group and person time information.
- Web application: The web-application is an interactive dashboard allowing end-users to visually explore benefit-risk measures and their components. The inputs of the chart generating functions are the data tables generated in the second pre-processing step as well as some user-defined settings (e.g. age groups, baseline incidences and preference weights).

The major advantage of using a web interface is its user-friendliness in accessibility and usage; the end-user can use a web browser of choice to access the dashboard without the need to install or understand R and the underlying electronic healthcare data can be seamlessly updated. The architecture allows for secure storage of the individual-level data as the web-application only uses as input the aggregated data generated by the second pre-processing step. The pre-processing

steps, which use the individual-level data, only need to be performed when the healthcare data is updated and can be done separately using a dedicated, secured server.

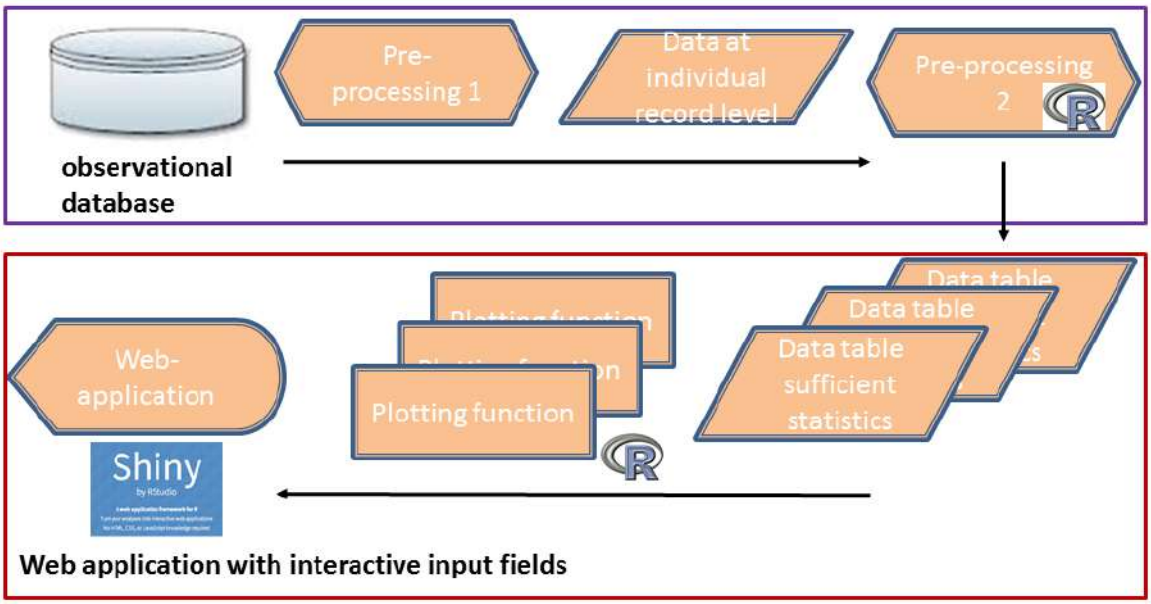


Figure S.1: Architecture of the interactive dashboard for benefit-risk monitoring of vaccines.

Benefit-risk dashboard: Detailed description of the visualizations for monitoring

Coverage

Weekly number of doses extrapolated to the whole UK population

Let n_{ij} denote the total number of doses given during week i by age group j as estimated from the database. Then, the total number of doses extrapolated to the whole UK population is calculated as

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

where N_{ij} is the number of active subjects within the database by week i and age group $j$andwhere pop_j is the number of subjects in the total UK population of age group j obtained from the National Office of Statistics, UK. As such, the weights w_{ij} can be interpreted as inverse sampling weights.

Coverage

Let n_{ij} denote the total number of vaccinated children at week i of age group j and let N_{ij} denote the total number of children. Then, coverage at week i for age group j is simply obtained as

$$cov_{ij} = n_{ij}/N_{ij} \quad (2)$$

Risks

Intussusception incidence rates; baseline incidences and within risk windows

The person-time incidence rate (per 10,000 person years) of intussusception by week I is estimated cumulatively over time, using all data accrued from the start of the study period/vaccination period till week I or

$$inc_I = \left({\sum_{i=1}^{I} n_i} \Big/ {\sum_{i=1}^{I} py_i} \right) \times 10{,}000, \quad (3)$$

where n_i is the number intussusception events that happened during week i and where py_i is the amount of person time (in years) within that week.

The baseline incidence rates of intussusception are estimated using data from the start of the study period until the start of the vaccination period. The incidence rates of intussusception following immunization during two consecutive risk windows (1-7 and 8-21 days post-vaccination) after dose 1 and dose 2 are estimated using data from the start of the vaccination period till the end of the study period.

Benefits

The person-time incidence rate (per 10,000 person years) of RVGE GP visits and of hospital admissions by week I is estimated using the most recent data within a look-period of length Δ or

$$inc_I = \left({\sum_{i=I-\Delta}^{I} n_j} \Big/ {\sum_{i=I-\Delta}^{I} py_j} \right) \times 10.000, \quad (4)$$

where n_i is the number events that happened during week i and where $py_i$is the amount of person time (in years) within that week.

For comparison, we also predicted the expected incidence rates of RVGE GP visits and of hospital admissions given assumed levels of VE and age-specific baseline incidence accounting for the vaccination coverage and age structure within the healthcare database. Specifically, we calculated the expected weekly number of cases as

$$E(n_i) = \sum_{j=1}^{J} py_{ij} \times \frac{E_{0\,j}}{10.000} \times \left(cov_{ij}(1 - VE) + \left(1 - cov_{ij}\right) \right), \quad (5)$$

where py_{ij} is the age- and week-specific person time (in years), cov_{ij} is the age- and week-specific coverage and where $E_{0\,j}$ refers to the assumed age-specific baseline incidences and VE to the assumed VE. Then, the expected incidence is calculated similarly as in (4) but using the expected counts instead or,

$$E(inc_I) = \left({\sum_{i=I-\Delta}^{I} E(n_j)} \Big/ {\sum_{i=I-\Delta}^{I} py_j} \right) \times 10.000. \quad (6)$$

Benefit-risk

Benefit-risk measures

The INHB is calculated as

$$\Delta'_i = \sum_{k=1}^{K} w_k \times (E_{0\,k} - E_{v\,ki}) + \sum_{k=1}^{K'} w_k \times p_k\,(R_{0\,k} - R_{v\,ki}) = \boldsymbol{E_i} + \boldsymbol{R_i}. \quad (7)$$

With its variance equal to

$$\sigma^2(\Delta'_i) = \sum_{k=1}^{K} (w_k \times p_k)^2 \left(E_{v\,k\,i} \times {}^{10.000}/_{pt_{v\,ki}}\right) + \sum_{k=1}^{K'} (w_k \times p_k)^2 \left(R_{v\,ki} \times {}^{10.000}/_{pt_{v\,ki}}\right)$$
$$= \sigma^2(\boldsymbol{E_i}) + \sigma^2(\boldsymbol{R_i}) \quad (8)$$

assuming $w_k, p_k, E_{0\,k}$ and $R_{0\,k}$ in (11) to be known and using the Poisson approximation to the binomial variance of the incidences [24]. The 95% Wald confidence intervals are then obtained as $CI = \Delta'_i \pm 1.96 \times \sqrt{\sigma^2(\Delta'_i)}$.

The IBRR uses the same terms $\boldsymbol{E_i}$ and $\boldsymbol{R_i}$ as in (11), but uses the ratio instead, or

$$\Omega'_i = {}^{\boldsymbol{E_i}}/_{-\boldsymbol{R_i}}. \quad (9)$$

The variance is then expressed as

$$\sigma^2(\ln(\Omega_i')) = {}^{\sigma^2(\boldsymbol{E_i})}/_{\boldsymbol{E_i}^2} + {}^{\sigma^2(\boldsymbol{R_i})}/_{\boldsymbol{R_i}^2}, \quad (10)$$

with the 95% Wald confidence intervals equal to $CI = e^{\ln(\Omega'_i) \pm 1.96\sqrt{\sigma^2(\ln(\Omega'_i))}}$. In case of theoretical benefits, $\sigma^2(\boldsymbol{E_i}) = 0$.

11.3. Appendix P5 (components)

Questionnaire on previous experience of databases

1. Do you have previous experience in extracting pertussis from your database? (yes/no/partial)
2. If yes:
   a. which codes were used, in which algorithm?
   b. Did you ever validate this event in your database? If so what is was PPV? What was the sensitivity?
   c. Did you publish papers in peer-reviewed journals, or write reports?
3. Can you estimate the background incidence of pertussis in your country in children? (please provide estimate or reference or link)

Literature review

Manuscript Author	Manuscript Title	Component	Component Description	Is there any hint or data about validity?
Acosta et al.	Tdap Vaccine Effectiveness in Adolescents During the 2012 Washington State Pertussis Epidemic	Pertussis	reports of cough illness of any duration plus isolation of Bordetella pertussis from a clinical specimen with either a positive polymerase chain reaction test result or contact with a laboratory-confirmed case (epidemiologic link)	

Boehmer et al.	Use of Hospital Discharge Data to Evaluate Notifiable Disease Reporting to Colorado's Electronic Disease Reporting System	Pertussis	Inpatient hospital data: IHD dataset includes demographic, diagnostic, procedural, payment, and length-of-stay variables. Used discharge diagnosis and discharge diagnosis followed by medical record review	59% sensitivity with discharge diagnosis data only; 100% sensitivity when using discharge data plus medical record review
			ICD-9 Codes: Pertussis/whooping cough 033.0–033.9	
Bozzola et al.	Infectious Diseases and Vaccination Strategies: How to Protect the ''Unprotectable''?	Pertussis	hospital admission data and medical record review; validated using PCR	
Breakwell et al.	Pertussis Vaccine Effectiveness in the Setting of Pertactin-Deficient Pertussis	Pertussis	Department of Health Registry clinical case reports of cough illness lasting ≥2 weeks with paroxysms of coughing, inspiratory "whoop, " or posttussive vomiting	
Bellettini et al.	Clinical, laboratory and radiographic predictors of Bordetella pertussis infection	Pertussis	Medical record review. predictors were identified cyanosis (OR 8.0; 95% CI 1.8 to 36.3; p = 0.007) and lymphocyte counts> 10 4 / uL (OR 10.0, 95% CI 1.8 -54.5;p = 0.008) in children under 6 months of age.	Clinical symptom presentation in pertussis: cough (100%), cyanosis (59.6%), post-cough vomiting (37.9%), fever (34.2%), respiratory distress (36%)
De Serres et al.	Effectiveness of a whole cell pertussis vaccine in child-care centers and schools	Pertussis	questionnaire collected information on all cough illnesses lasting for at least 2 weeks and on other pertussis-associated symptoms (paroxysmal cough, posttussive vomiting, apnea, whoop). Medical records of children who reported a cough illness present for at least 2 weeks were evaluated independently by physicians	
Hurtado-Mingo et al	Clinical and epidemiological features of pertussis Among infants hospitalized During 2007-2011 in Seville	Pertussis	medical records with a diagnosis of pertussis using ICD-9 codes 033, 033.0 and 033.9 and/or positive isolates by polymerase chain reaction	Cough was the most common symptom (87%), followed by cyanosis (44%), respiratory distress (33%) and apneas (26%). P-values for symptoms: Whooping cough, 0.012; Cyanosis, 0.288; Breathlessness, 0.078; Apneas, 0.134; Pneumonia, 0.562; Fever, 0.187
Chen et al.	Estimated incidence of pertussis in people aged <50 years in the United States	Pertussis	database of medical insurance claims; database of laboratory test results; and surveillance database records for diagnosed pertussis, defined as a claim for pertussis (ICD-9 033.0, 033.9, 484.3); and medical records of cough illness (ICD-9 033.0, 033.9, 484.3, 786.2, 466.0, 466.1, 487.1) attributed to laboratory confirmed pertussis	Comparison of medical claims data and laboratory data indicated that the three pertussis ICD-9 diagnosis codes had low sensitivity (30.4%) and high specificity (94.0%), with a positive predictive value of 68.2%.

Ferronatoa et al.	Respiratory viral infections in infants with clinically suspected pertussis	Pertussis	hospital admissions records for dry cough for at least two weeks accompanied by inspiratory stridor, paroxysmal cough, or vomiting after coughing	Cough followed by inspiratory stridor and cough accompanied by cyanosis were significant predictors of pertussis (positive predictive values of 100% and 84%, respectively). Leukocyte count > 20,000 cells/mm3 and lymphocyte count > 10,000 cells/mm3 showed predictive values of 92% and 85%, respectively. However, these variables showed low negative predictive values for the diagnosis of pertussis (40%, 60%, 52% and 64%, respectively). Laboratory predictors had 90% sensitivity.
Guinto-Ocampo et al.	Predicting pertussis in infants.	Pertussis	medical record review for clinical and laboratory predictors of pertussis. Exclusion: Infants who received macrolide antimicrobials for pertussis	p-values of predictors: Cough 0.47, Coughing contact 0.77, ALTE 0.37, Posttussive vomiting 0.33, Fever 0.99 Respiratory rate (breaths/min) 0.02, Hypoxia (%O2 saturation e94%) 0.09, WBC count, 1000/2L 0.02, Lymphocytes 0.00, ALC/2L 0.00. Sensitivity of molecular methods: WBC count 89%; % Lymphocytes 89%, ALC 89%. Specificity of molecular methods: WBC count 45%; % Lymphocytes 66%, ALC 75%. PPV of molecular methods: WBC count 26%; % Lymphocytes 37%, ALC 44%. NPV of molecular methods: WBC count 95%; % Lymphocytes 96%, ALC 97%.
Danica E. Kuncio	Health Care Worker Exposures to Pertussis: Missed Opportunities for Prevention	Pertussis. Health record (EHR) data identifying laboratory-confirmed pertussis cases. Not specified	Not specified	Potential exposures were defined as HCWs involved in the care of the index case before investigation. A confirmed exposure as defined as any HCW who had direct face-to-face contact within 3 feet of an index case regardless of the length of contact time and the vaccination status of the exposed person. Fulfillment of these criteria was confirmed by the OHD through interviews with each potentially exposed person

Yury V. Lobzin	Retrospective Study of the Clinical Epidemiological Characteristics of Pertussis in Infants Prior to Their First Vaccination in the Russian Federation	Pertussis. Archived medical records and a questionnaire were	Not specified	A diagnosis of pertussis was made based on standard clinical epidemiological data, in accordance with the World Health Organization International Classification of Diseases (ICD-10) and the literature. cases verified by laboratory analysis (if facilities were available) or diagnosed only clinically (if laboratory facilities were not available); if laboratory data were negative but characteristic clinical symptoms were present the diagnosis was maintained. If performed, laboratory analysis was done preferably by polymerase chain reaction (PCR), or by serology or culture.

L. K. MCCALLUM	Estimating the burden of pertussis in young children on hospitals and emergency departments: a study using linked routinely emergency departments: a study using linked routinely	pertussis notifications , hospitalizati ons and emergency, department (ED) presentatio ns.	Information available on the records included data about the notification (who and when), laboratory confirmation (specimens, type and dates) and patient outcome (hospitalization, death). A calculated onset date was available which is defined as the earliest of notification, patient reported onset or specimen collection dates. The Admitted Patient Data Collection contains demographic, administrative, diagnostic and procedural information from all NSW public and private hospitals and day procedure centres. Diagnoses and procedures for each admission are coded according to ICD-10-AM [20]. Records were available from 1 July 2000. The Emergency Department Data Collection captures an estimated 83% of presentations to EDs in NSW public hospitals [18]. Information on patient demographics and provisional diagnoses is available; diagnoses are coded using ICD-9 [21] ICD-10-AM [20] or SNOMED-CT [22] classification schemes. Records were available from 1 January 2005. We used death records from two sources: NSW Registry of Births, Deaths and Marriages (RBDM) 696 L. K. McCallum and others death registrations and Australian Bureau of Statistics (ABS) mortality records. NSW RBDM records contain only fact of death information and include data from 1 January 1994. ABS records contain information on the cause of death coded according to ICD-10 [23] and include data from 1 January 1994. We used records from all datasets up until 31 December 2008, with the exception of the ABS mortality records which were available only up until 31 December 2007 due to a delay in the release of data by the ABS.	Pertussis notifications included cases with definitive laboratory evidence; or laboratory suggestive evidence together with clinical evidence; or clinical evidence together with an established epidemiological link to a confirmed case with laboratory evidence. Laboratory definitive evidence required isolation of B. pertussis from a clinical specimen or detection of B. pertussis by nucleic acid testing. Laboratory suggestive evidence required seroconversion or significant increase in antibody (IgA or IgG) level or a5 fourfold rise in titre to B . pertussis whole cell (IgA only) or B. pertussis specific antigen (in absence of recent vaccination) or a single high IgA titre to whole cells or detection of B. pertussis antigen by immunofluorescence assay. Clinical evidence required a coughing illness lasting 5 2 weeks or paroxysms of coughing or inspiratory whoop or post-tussive vomiting [24] The Admitted Patient Data Collection records were classifi ed as having a coded diagnosis of ‘ pertussis’ if any diagnosis fi eld had an ICD-10-AM code of A37. Emergency Department Data Collection records were classified as having a coded diagnosis of ‘ pertussis’ if the provisional diagnosis had an ICD-10-AM code of A37, ICD-9 code of 033 or if the SNOMEDCT code description contained ‘ whooping cough’ or ‘ pertussis’ .

11.4. Appendix P7 (BoD)

Appendix Tables A1-A4 present the results of the literature search for relative risk estimates, for defined set of vaccine-event pairs.

Appendix Table B presents details on the studies selected to provide background incidence rates for ITP, febrile and generalised convulsions/seizures.

Appendix Table C details the disability weights and disability durations compiled for the selected adverse events.

Appendix Table D provides the selection and classification of the provisional set of candidate adverse events for investigation (initial phase of the project).

Computation of vaccination-attributable event incidence using published PAR(%) or relative risks

Depending on availability, published vaccine-attributable risks AR (%) (i.e., from safety or related studies) for the selected events can be applied, to estimate the proportion of events among exposed (i.e., vaccinated persons) that are associated with vaccination. As a published attributable risk (AR%) may be restricted to specific age-group(s), decisions have to made about whether or not the AR (%) should be extrapolated to other age-groups for which published values are unavailable. Further decisions need to be taken regarding application of vaccine-attributable risks derived from a particular national population, to the event data for other countries.

If the AR (%) for a particular event is not available from the literature, it can be estimated based on relative risks (RRs) from relevant published pharmaco-epidemiological or other studies:

$$AR(\%) = \frac{RR - 1}{RR} * 100$$

To obtain the *population-attributable risk,* PAR (%) – or the proportion of events within the [partially] exposed population that may be associated with exposure – the relative risk needs to be combined with national vaccination coverage data. The PAR (%) can be estimated from the RR and vaccination coverage (*vc*) as:

$$PAR(\%) = \frac{vc(RR - 1)}{1 + vc(RR - 1)} * 100$$

If AR (%) and vaccination coverage are available, we can calculate PAR (%), or the proportion of disease within the partially-vaccinated population due to vaccination. So, given the overall ('background') event incidence rate (for a particular age-group and time period, in *events per person*), inc_{backgr}, and PAR (%), we can estimate the incidence rate of the event associated with vaccination:

$$inc_{vacc} = inc_{backgr} * \text{PAR (\%)}$$

Then, the expected number of events within the population that are associated with vaccination is simply:

$$n_{vacc} = inc_{vacc} * \text{pop}$$

with population size *pop*. However, most often RRs are calculated with respect to a 'window' risk-period, for instance 1-14 days following immunization. In this case, the above methods to calculate AR% and PAR% cannot be applied as is, as they implicitly assume a 'constant' exposure (thus application is clearer when the exposure variable is for example smoker vs. non-smoker).

Possible scenarios for published data on incidence/AR%/relative risks

1. Study already computes and applies AR (%) to national incidence rates, and reports inc_{vacc} (sometimes called 'excess incidence rate') or possibly the number of vaccination-attributable events (pop * inc_{vacc}). Thus, we can already use the published inc_{vacc} in the DALY calculation without doing anything further (but should verify that study correctly calculated inc_{vacc}).

2. Study reports a relative risk (RR), which is the ratio of the event incidence in a 'window period' following vaccination and the event incidence in a 'control period'. To apply a RR to a given background event incidence rate, we need to estimate the incidence rates for the control period and in the window period.

   In a 100% vaccinated population, the observed background event incidence rate, inc_{backgr}, is (n_{vacc} + $n_{control}$)/ total person-time. n_{vacc} and $n_{control}$ refer to number of events in the risk 'window' post-vaccination of size *t* (in days) and the number of events in the 'control period', respectively. Both are unknown.

   So, inc_{backgr} is actually a weighted average; if the unit is one vaccinated person-year:

   $$inc_{backgr} = inc_{vacc} * (t/365.25) \quad + \quad inc_{control} * ((365.25-t)/365.25) \qquad (1)$$

   Note that multiple doses within the period over which inc_{backgr} is defined, can be easily handled by setting *t* to the size of the summed risk windows; for instance for *n* vaccine doses with associated risk window sizes t_d:

   $$t = \sum_d^n t_d$$

   If we let $w = (t/365.25)$, and rearrange to define inc_{vacc} in terms of constants and one unknown:

   $$inc_{vacc} = \frac{inc_{backgr} - (inc_{control} * (1-w))}{w} \qquad (2)$$

   Given that RR = $(inc_{vacc} / inc_{control})$, *we can re-arrange as:*

   $$inc_{control} = (inc_{vacc} / RR) \qquad (3)$$

   and substitute in (2) to define inc_{vacc} in terms of known values only:

$$inc_{vacc} = \frac{inc_{backgr} - (inc_{vacc}/RR * (1-w))}{w} \quad (4)$$

after re-arrangement:

$$inc_{vacc} = \frac{inc_{backgr}}{(\frac{1-w}{RR}+w)} \quad (5)$$

which is the incidence rate per vaccinated person-year, attributable to immunization.

Then, the expected number of adverse events associated with immunization in a population of *pop* vaccinated persons within one year of follow-up, where the period at risk is *t* days (and recall *w* = (*t*/365.25)) is:

$$n_{vacc} = inc_{vacc} * w * pop \quad (6)$$

As the incidence rate inc_{vacc} refers to a *vaccinated* population (more precisely, events per vaccinated person-year), it should be further adjusted for vaccination coverage *vc* (ie. if *vc* was 0%, then no adverse events could be attributed to immunization!).

$$n_{vacc} = inc_{vacc} * vc * w * pop \quad (7)$$

This result is sufficient for input to the DALY calculation; we don't need to do anything more.

Uncertainty in the RR can be incorporated by computing within a simulation (e.g., @Risk for Excel) or in a sampling/MCMC framework. This method can easily be generalised to use multiple RRs, for instance when reported separately for distinct risk windows (e.g., same day as immunization, 1-3 days following, 4-7 days following) and for multiple risk windows within a year (i.e., four DTP immunizations given within the first year of life).

For example, Barlow et al. (2001) report excess rate of febrile seizures (i.e., inc_{vacc}) of 5.6 per 100,000 children receiving the DTP immunization in the USA in 1991-93. The relative risk for the 1-day window period 'same day as immunization' was 5.70 (95% CI: 1.98-16.42). Thus we could already use the published inc_{vacc} in our DALY calculation, but as a 95% CI was not reported, additional work/assumptions would be required. Instead, better to use the RR estimate provided, and apply to background incidence rates obtained from the ADVANCE databases.

3. Study reports an incidence rate ratio (IRR), which is the ratio of the event incidence in a 'window period' following vaccination and the event incidence in a control period; the IRR can be estimated directly from incidence rates, or by fitting a Poisson model (so allowing adjustment for demographic and other variables). One can follow the same procedure as for (2) above, to 'adjust' the published IRR so it can be used to derive AR% and eventually $inc_{vacc.}$

   For example, Bakken et al (2015) report, for children up to 45 months of age in Norway in 2009, an IRR of 2.00 (95% CI:1.15-3.51) for the risk of febrile seizures occurring 1-3 days after pandemic influenza vaccination, compared with a 166-day control period. This study does not report inc_{vacc} , but it can be calculated from the provided incidence rates and person-time, by applying the supplied incidence rate in the control period to the 'window' person-time, which gives the expected number of events. Then subtract this from the observed

number of events to calculate the excess incidence rate, inc_{vacc}. Thus, DALYs could be calculated directly from the values provided in this study.

4. For certain AEFI (eg. anaphylaxis), a RR is typically not reported because the event occurs immediately after immunization and can be considered to be caused by immunization. Risks are often reported as events per 1 million doses. In such a case, we can easily convert a risk provided as cases per million doses directly to the number of vaccination-attributable events, $n_{vacc,}$ as:

$$n_{vacc} = cases\ per\ 1M/1000000 * vc * pop \quad (8)$$

For instance, if in a given country the first dose of MMR is given at 12-13 months, the relevant population size (*pop*) can be approximated as size of birth cohort in the year previous (thus, for the DALY calculation only a single year-wide age-group needs to be specified). Note that this method does *not* use background incidence rate data.

For example, Cheng et al. (2015) report, for children <18 years old in Victoria, Australia (period 2007-2013), 12.5 cases per million MMR vaccination doses (95% CI: 11.8-13.2). DALYs can be calculated by first converting the provided risk estimate (with 95% CIs) to the number of vaccination-attributable events, n_{vacc} (with 95% CIs), according to the above expression.

Tools for computing the burden of disease

Data collection spreadsheet

An Excel spreadsheet has been developed to enable centralised collection of all relevant data and parameters required for the computation of adverse event burden (as YLD). Tables organised on separate spreadsheet tabs store population size data, vaccination scheme information, vaccination coverage, background incident rates, and PAR(%) and/or RR estimates. Point estimates of vaccination-attributable incident rates and number of vaccination-attributable cases, and various YLD measures, both stratified by age-group, country/data source, and event can be calculated with this tool. The set of events, age-groups, study period, and countries/databases are all modifiable by the user, requiring only basic knowledge of Excel.

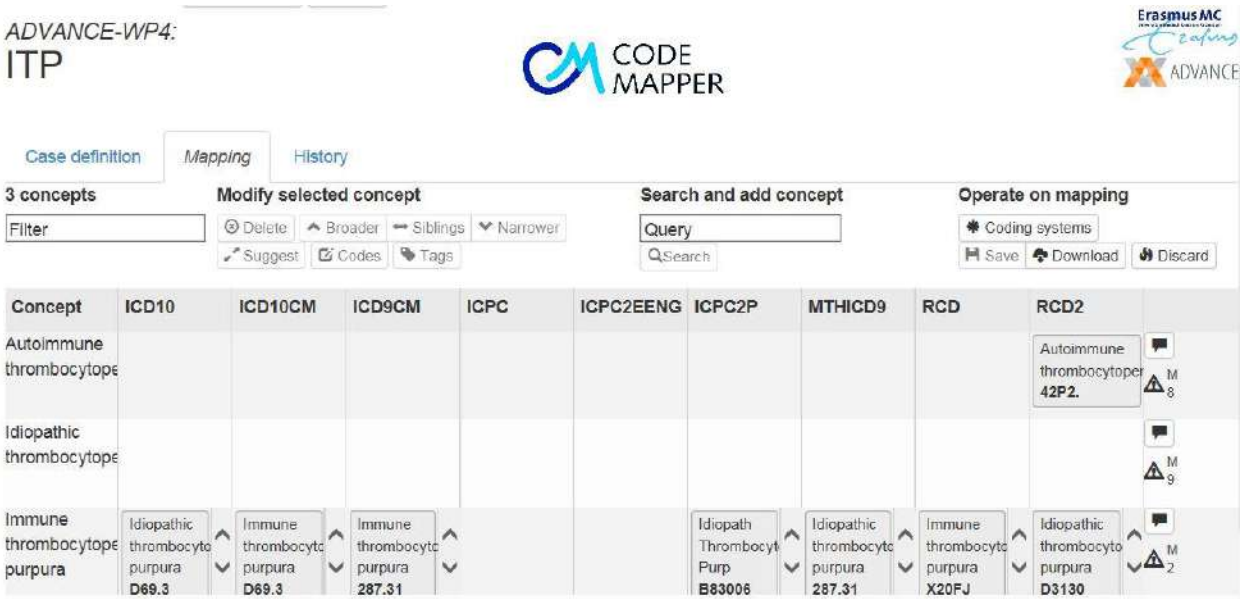
R functions for computation of YLD and YLL, with uncertainty intervals

To enable the computation of point estimates (as well as correct 95% uncertainty intervals) of a number of outcome measures (e.g., vaccination-attributable event incidence rates, YLD, YLD per 100,000 population, YLL), a calculation tool in the form of a suite of R functions has been developed. This tool is suitable for users having a basic- to intermediate-level knowledge of R.

CodeMapper results

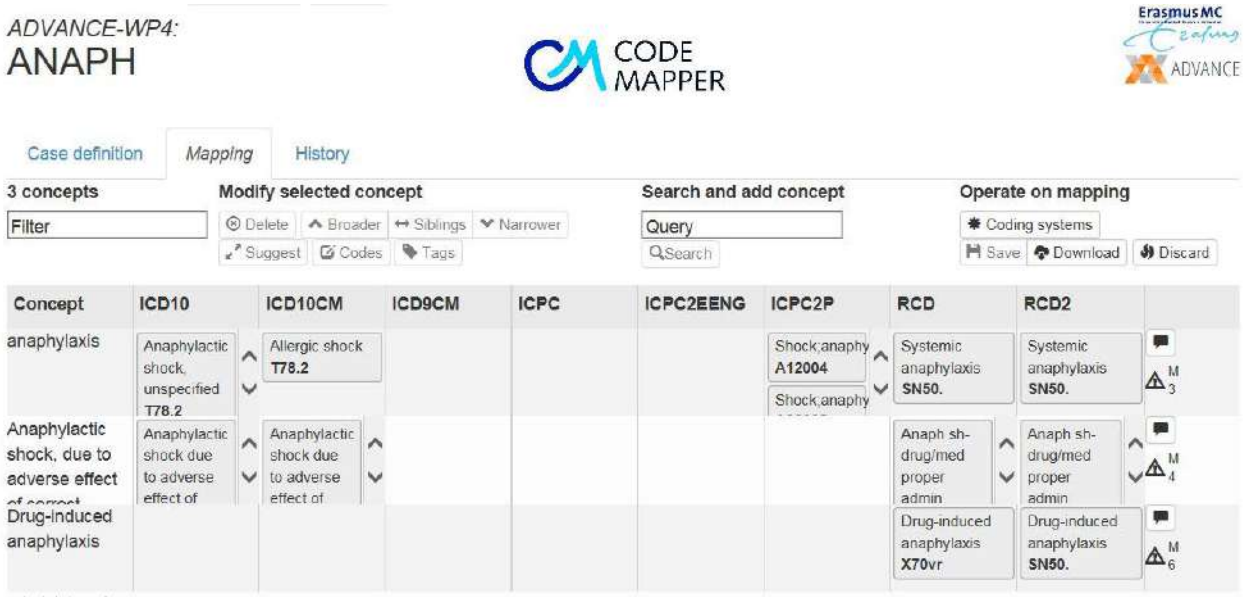
ITP

Results of the code-mapping session (participants: Benus Becker, Caitlin Dodd, Kartini Gadroen, Nicoline van der Maas, Scott McDonald, Danielle Nijsten) are shown below, as a screengrab from the CodeMapper tool.



Anaphylaxis

Results of the code-mapping session (participants: Benus Becker, Nicoline van der Maas, Scott McDonald, Danielle Nijsten) are shown below, as a screengrab from the CodeMapper tool.



Febrile convulsions/seizures and generalised convulsions/seizures

Code-mapping was carried out as part of ADVANCE WP5 activities.

APPENDIX A

A1.1 MMR – ITP

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Black et al, 2003	1988-1999	UK		13-24 mo	Case-control				6.3 (1.3-30.1)	6 weeks	
Miller et al, 2001	1991-1994	UK		12-23 mo					3.27 (1.49-7.16)	6 weeks	
O'Leary et al, 2012	2000-2009	USA		12-19 mo	Retrospective cohort study				IRR 5.48 (1.61-18-64)	6 weeks	
Bertola et al, 2010	1999-2007	Italy		1 mo – 18 years	Case-control				OR 2.4 (1.2-4.7)	6 weeks	
France et al, 2008	1991-2000	USA		12-23 mo 12-15 mo	Retrospective cohort				IRR 3.94 (2.01-7.69) 7.10 (2.03-25.03)	6 weeks	

A1.2 DTaP – ITP

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
O'Leary et al, 2012	2000-2009	USA	1.8 million	12-19 mo 4-6 y	Retrospective cohort	DTaP			IRR 1 (0.21-4.81) IRR 2.57 (0.53-12.37)	6 weeks	

Note. (DN) It was very difficult to find suitable studies; most articles are either case reports, or patients had received a combination of vaccines.

A1.3 HBV – ITP

(DN) No suitable literature could be located.

A2.1 MMR – Anaphylaxis

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Bohlke et al, 2003	1991-1997	US	848,945	<18 years	Retrospective analysis				3.5 cases/million doses (0.7-10.3)		
Patja 2001	1982-1996	Finland	2,990,000 doses, 1.8 million individuals	14 months – 23 years	Prospective follow-up			M-M-R II	1.0 cases/100,000 doses	All developd reaction within 15 minutes	
D'Souza et al, 2000	1998	Australia	1.7 million individuals	4-5 years	Retrospective analysis and follow-up	live attenuated measles virus (Edmonston strain), mumps virus (Jeryl Lynn strain), and rubella virus (Wistar RA 27/3 strain), and 25mcg neomycin per 0.5ml dose.	Merck, Sharp and Dohme	M-M-R II	0.06 cases/100,000 doses	30 days	1 case
Pool et al, 2002	1991-1997	US	94,000,000 doses		Retrospective analysis			M-M-R II	1.8 cases/1,000,000 doses		
Cheng et al, 2015	2011-2013	Australia	481,297 doses	<18 years	Retrospective analysis		GSK	Priorix	1.25 cases/100,000 (1.18-1.32)		6 cases in total from passive surveillance

A2.3 DTaP – Anaphylaxis

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Bohlke et al, 2003	1991-1997	US	448,456	<18 years	Retrospective analysis				0 cases/million doses (0-8.2)		
Nakayama and Onoda 2007	1994-2004	Japan				Thimerosal was removed from vaccine after 2003	Kitasato		*1994-2003:* 0.93 cases/million doses *2004:* 1.14 cases/million	0-48 hours	
McNeil et al, 2016	2009-2011	USA	584,103	0+ years	Retrospective				5.14/million (1.06-15.01)		3 cases only. Not broken down by age-group

A2.4 DTwP – Anaphylaxis

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Bohlke et al, 2003	1991-1997	US	778,807	<18 years	Retrospective analysis				1.3 cases/million doses (0.03-7.1)		1 case

A2.5 HBV – Anaphylaxis

	Setting		Study population		Study design	Pharmaceutical details			Results		Remarks
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	
Bohlke et al, 2003	1991-1997	US	1,852,147 doses	<18 years	Retrospective analysis				1.1 cases/million doses (0.1-3.9)	Cases develop anaphylaxis within 2 days	This is the study WHO uses for its HBV-factsheet
Dobson et al, 1995	1992	Canada	127,922 doses	11-12 years	Cohort study			Engerix-B, 20 µg	7.8 cases/million doses		Can't access full text – 1 case only
CDC, 1996		US			Retrospective analysis				1.67 cases/million dosse		Data based on *Vaccine Adverse Events Reporting System*

A2.6 MenC – Anaphylaxis

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups	Study design	Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Yergeau et al, 1996	December 1992-march 1993	Canada	1,198,751	6 months-20 years	Retrospective descriptive study		Pasteur Merieux, Lyons SmithKline Beecham Pharma Inc., Connaught Laboratories Limited	96% received: Polysaccharide Meningococcal A and C Vaccine 4% received either: Mencevax AC Menomune	0.1 cases/100,000 doses		Just 1 case
McNeil et al, 2016	2009-2011	USA	649,199	0+ years	Retrospective	MCV4			6.16/million (1.68-15.78)		4 cases only. Not broken down by age-group

A2.7 VZV – Anaphylaxis

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups		Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Bohlke et al, 2003)	1991-1997	USA	254,186	<18 years	Retrospective analysis				0 cases/million doses (0-14.5)		
Chaves et al, 2008	1995-2005	USA	47,733,950 doses		Retrospective analysis				0.1 cases/ 100,00 doses		In 63% of the cases developing anaphylaxis the varicella vaccine was given in combination with another vaccine (mostly MMR).
Ozaki et al, 2005	1994-1999 1999-2000	Japan	1,410,000 doses 1,300,000 doses			Live varicella vaccine/Oka strain gelatin-containing Gelatin free	Kannonji Institute, The Research Foundation for Microbial Diseases of Osaka University		2.27 cases/100,000 doses 0,08 cases/100,000 doses		
Wise et al, 2000	1995-1998	USA	~9,730,000 doses		Retrospective analysis	Live varicella vaccine	Merck & co.	Varivax	0.31 cases/100,000		
McNeil et al, 2016	2009-2011	USA	304,001	0+ years	Retrospective				6.58/million (0.80-23.77)		2 cases only. Not broken down by age-group

A3.1 MMR – Febrile convulsions/seizures

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups		Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Ma et al, 2015					Trials, restrospective cohort, and SCCS	MMRV MMR					Systematic review and meta-analysis. Included studies compare MMR+V vs. MMR-only.
Vesterg aard et al, 2004	1991-1999	Denmark	537,171	3 mo – <10 years	Restrospective cohort (register-based)	MMR			2.75 (2.55-2.97)	14 days	

A4.1 MMR – Generalised convulsions/seizures

	Setting		Study population		Study design	Pharmaceutical details			Results		
Article	Study period	Country	Total size (n)	Age groups		Vaccine type	Manufacturer	Product name	RR (95% CI)	Window sizes	Remarks
Bino et al, 2003	2000	Albania	876,000 doses	1-14 years		Combined MR vaccine	Serum Institute India		1.15 per 1,000,000 doses		1 seizure case
D'Souza et al, 2000	1998	Australia	1.7 million children	4-13 years		Live attenuated measles virus (Edmonston strain), mumps virus (Jeryl Lynn strain), and rubella virus (Wistar RA 27/3 strain), and 25mcg neomycin per 0.5ml dose.	Merck, Sharp and Dohme lyophilised product	M-M-R II-	0.24 per 100,000 doses	30 days	
Barlow et al, 2001	1991-1993	USA	137,457 doses	<7 years	Cohort study		.		1.11 (0.11-11.28) 0.48 (0.05-4.64)	8-14 days 15-30 days	Both window sizes: only 1 case of nonfebrile seizure.

APPENDIX B

Table B1 – Selected background incidence rates

		Setting		Study population		Study design	Results	Remarks
Adverse event	Article	Study period	Country	Total size (n)	Age groups		Inc rate (95% CI)	
ITP	Yong et al, 2010	1990-2005	UK		<18 years <2 years 2-5 years		4.2 per 100,000 p-yrs (3.7–5.8) 6.8 per 100,000 p-yrs (4.9–9.2) 7.2/100,000 p-yrs (5.9–8.8)	
Febrile convulsions/ seizures	Sammon et al, 2015	1999-2011	UK	1,532,992	2-12 mos 13-24 mos 25-60 mos 61-120 mos 121-180 mos	GP registry-based	5.56/1000 p-yrs (5.37–5.75) 13.77/1000 p-yrs (13.48–14.07) 4.32/1000 p-yrs (4.13–4.33) 0.58/1000 p-yrs (0.54–0.61) 0.23/1000 p-yrs (0.18–0.28)	Used General Practice Research Database
Generalised convulsions/ seizures	Sammon et al, 2015	1999-2011	UK	1,532,992	2-12 mos 13-24 mos 25-60 mos 61-120 mos 121-180 mos	GP registry-based	2.79/1000 p-yrs (2.66–2.93) 2.88/1000 p-yrs (2.75–3.02) 1.91/1000 p-yrs (1.85–1.98) 1.38/1000 p-yrs (1.32–1.43) 1.19/1000 p-yrs (1.09–1.29)	Used General Practice Research Database. (Data are for nonfebrile only)

APPENDIX C

Table C1: Disability weights (DW) and disability durations for the four selected adverse events, plus diarrhoea.

Adverse event	DW (95% uncertainty interval)	Disability duration
Idiopathic thrombocytopenia purpura (ITP)	Thrombocytopenia purpura: 0.159 (0.106–0.226) [source: Salomon et al, 2015] ***Rare sequela; possible proxy:*** Gastric bleeding: 0.325 (0.209–0.462) [source: Salomon et al, 2015]	2-8 weeks (max. 6 months) [sources: http://patient.info/health/immune-thrombocytopenia-leaflet; http://www.netdoctor.co.uk/conditions/heart-and-blood/a1176/idiopathic-thrombocytopenic-purpura-itp/] 93% resolved by 6 months [source: Mantadakis et al, 2010]
Anaphylaxis	***Possible proxy:*** Epilepsy: severe: 0.552 (0.375–0.710) [source: Salomon et al, 2015]	1 day [source: Ginsberg et al, 2015]
Febrile convulsions/seizures	***Possible proxies:*** Epilepsy: less severe: 0.263 (0.173–0.367) Epilepsy: severe: 0.552 (0.375–0.710) [source: Salomon et al, 2015]	1 day [source: Oluwabusi & Sood, 2012]
Generalised convulsions/seizures	***Possible proxies:*** Epilepsy: less severe: 0.263 (0.173–0.367) Epilepsy: severe: 0.552 (0.375–0.710) [source: Salomon et al, 2015]	Average duration of disability in children between 0 and 4 years old is 1.5 years [source: Murray & Lopez, 1996] *or* 1 day (assume same as for febrile convulsions)
Diarrhoea	Diarrhoea: mild: 0.074 (0·049–0.104) Diarrhoea: moderate: 0·188 (0.125–0.264) Diarrhoea: severe: 0·247 (0.164–0.348) [source: Salomon et al, 2015]	1 to 10 days [source: Kirk et al, 2015]

APPENDIX D

Table D1 – Selected candidate adverse events for investigation, ranked by Eudravigilance frequency

Event	EudraVigilance (%)	PRISM	GRiP	Frequency/ seriousness
Convulsions/seizures	4.9	yes	yes	Frequent/low
Febrile convulsions/seizures	2.2	yes	yes	Frequent/low
Diarrhea	3.4	no	no	Frequent/low
Urticaria	3.1	no	no	Frequent/low
Anaphylaxis	0.6	yes	yes	Infrequent/high
Guillain-Barre Syndrome	0.5	yes	yes	Infrequent/high
Idiopathic thrombocytopenia purpura	0.2	yes	yes	Infrequent/high
Bell's Palsy	0.0	yes	yes	Infrequent/high
Acute disseminated encephalomyelitis/encephalitis	0.0	yes	yes	Infrequent/high

Note. PRISM (Baker et al, 2013); GRiP (Brauchli Pernus et al, 2015); EudraVigilance fromhttp://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/q_and_a/q_and_a_detail_000166.jsp.

REFERENCES

Baker MA, Nguyen M, ColeDV, LeeGM, Lieu TA. Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. Vaccine 2013; **31**: K98-K112

Bakken IJ, AabergK M, GhaderiS, Gunnes N, Trogstad L, Magnus P, Håberg SE. Febrile seizures after 2009 influenza A (H1N1) vaccination and infection: a nationwide registry-based study. BMC Infect Dis 2015; **15**(1): 1.

Barlow WE, et al. The risk of seizures after receipt of whole-cell pertussis or measles, mumps, and rubella vaccine. N Eng J Med 2001; **345**(9): 656-661.

Bino S, et al. Measles-rubella mass immunization campaign in Albania, November 2000. J Infect Dis 2003; **187** Suppl 1: S223-229.

Bohlke K, et al. Risk of anaphylaxis after vaccination of children and adolescents. Pediatrics 2003; **112**(4): 815-820.

Brauchli Pernus Y, et al. Reference set for performance testing of pediatric vaccine safety signal detection methods and systems. Vaccine 2016; **34**(51): 6626-6633.

CDC. Update: vaccine side effects, adverse reactions, contraindications, and precautions: recommendations of the Advisory Committee on Immunization Practices (ACIP), MMWRRecomm Rep 1996; **45**(RR-12): 1-35.

Cheng DR, et al. Pediatric anaphylactic adverse events following immunization in Victoria, Australia from 2007 to 2013. Vaccine 2015; **33**(13): 1602-1607.

Dobson S, et al. Assessment of a universal, school-based hepatitis B vaccination program. JAMA 1995; **274**(15): 1209-1213.

D'Souza RM, et al. Adverse events following immunisation associated with the 1998 Australian Measles Control Campaign. Commun Dis Intell 2000; **24**(2): 27-33.

Ginsberg, GM, et al. Should Israel screen all mothers-to-be to prevent early-onset of neonatal group B streptococcal disease? A cost-utility analysis. Israel JHealth Pol Res 2013;**2**(1): 1.

Kirk MD, et al. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. PLoS Med 2015;**12**(12): e1001921.

Mantadakis E, Farmaki E, Buchanan GR. Thrombocytopenic purpura after measles-mumps-rubella vaccination: a systematic review of the literature and guidance for management. J Pediatrics 2010; **156**(4): 623-628.

McNeil MM, et al. Risk of anaphylaxis after vaccination in children and adults. J Allergy Clin Immunol 2016;**137**(3): 868-878.

Murray CJL, Lopez AD.Global Health Statistics. Cambridge, Harvard University Press (Global Burden of disease and Injury Series, Vol. 2), 1996.

Nakayama T, Onoda K. Vaccine adverse events reported in post-marketing study of the Kitasato Institute from 1994 to 2004. Vaccine 2007; **25**(3): 570-576.

O'Leary ST, et al. The risk of immune thrombocytopenic purpura after vaccination in children and adolescents. Pediatrics 2012; **129**(2): 248-255.

Oluwabusi T, Sood SK. Update on the management of simple febrile seizures: emphasis on minimal intervention. Curr Opin Ped 2012; **24**(2): 259-265.

Patja, A. Allergic reactions to Measles-Mumps-Rubella vaccination. Pediatrics 2001;**107**(2): e27-e27.

Pool V, et al. Prevalence of anti-gelatin IgE antibodies in people with anaphylaxis after Measles-Mumps-Rubella vaccine in the United States. Pediatrics 2002; **110**(6): e71-e71.

Salomon JA, et al. Disability weights for the Global Burden of Disease 2013 study. Lancet Glob Heal2015; **3**(11): e712-e723.

Sammon CJ, et al The incidence of childhood and adolescent seizures in the UK from 1999 to 2011: A retrospective cohort study using the Clinical Practice Research Datalink. Vaccine2015; **33**(51): 7364-7369.