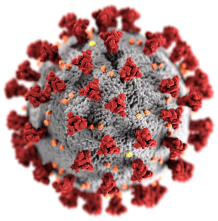
ACCESS

vACCine covid-19  monitoring readinESS

D3- Final Report

Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines

Version 2.0
June 30, 2021

Willame C, Dodd C, Gini R, Durán CE, Thomsen RM, Lei Wang, Anne Gedebjerg, Kahlert J, Ehrenstein V, Bartolini C, Paoletti O, Droz C, Moore N, Haug U, Schink T, Diez-Domingo J, Mira-Iglesias A, Vergara-Hernández C, Carreras JJ, Villalobos F, Pallejà M, Aragón M, Perez-Gutthann S, Arana A, Giaquinto C, Barbieri E, Stona L, Huerta C, Martín-Pérez M, García Poza P, de Burgos A, Martínez-González M, Bryant V, Trifiro G, Souverein P, Gardarsdottir H, Siiskonen SJ, Mahy P, Weibel D, Klungel O, Sturkenboom MCJM. DOI: 10.5281/zenodo.5255870

DISCLAIMER

This protocol has been accepted by EMA as a deliverable of the framework contract No EMA/2018/28/PE. The protocol expresses the expertise of the authors and the ACCESS consortium as well as feedback received from EMA. It may not be understood or quoted as being made on behalf, or reflecting the position of the European Medicines Agency or one of its Committees or Working Parties

Marketing authorisation holder(s)

Title	Background rates of Adverse Events of Special Interest for monitoring of COVID-19 vaccines
Version identifier of the final study report	2.0 (Final)
Date of last version of the final study report	June 30, 2021
EU PAS register number	*EUPAS37273*
Active substance	*NA*
Medicinal product	*NA*
Product reference	*NA*
Procedure number	*NA*
Marketing authorisation holder(s)	*NA*
Research question and objectives	To generate incidence rates of adverse events of special interest for COVID-19 vaccine safety monitoring in 7 EU countries.
Country(-ies) of study	Seven participating electronic health care databases in 6 countries: Germany, Italy, Spain, UK, The Netherlands and Denmark.
Authors	Corinne Willame, University Medical Center Utrecht, The Netherlands Miriam Sturkenboom, University Medical Center Utrecht, The Netherlands

Marketing authorisation holder(s)	none
MAH contact person	Not relevant

Table of contents

1 ABSTRACT 8

2 LIST OF ABBREVIATIONS 20

3 INVESTIGATORS 21

4 OTHER RESPONSIBLE PARTIES 22

5 MILESTONES 22

6 RATIONALE AND BACKGROUND 24

6.1 BACKGROUND 24
6.2 RATIONALE FOR THE STUDY 24

7 RESEARCH QUESTION AND OBJECTIVES 25

7.1 CO-PRIMARY OBJECTIVES 25
7.2 SECONDARY OBJECTIVES 25

8 AMENDMENTS AND UPDATES 25

9 RESEARCH METHODS 26

9.1 STUDY DESIGN 26
9.2 SETTING 26
9.3 SUBJECTS 27
9.4 VARIABLES 28
9.4.1 Person-time & Follow-up 28
9.4.2 AESI, At-risk medical conditions & Operationalization 29
9.4.3 Other variables 33
9.5 DATA SOURCES AND MEASUREMENT 33
9.5.1 Germany: GePaRD 33
9.5.2 Netherlands: PHARMO Database Network 33
9.5.3 Denmark: Danish Registries (DCE-AU) 34
9.5.4 Spain: BIFAP 34
9.5.5 Spain: SIDIAP 35
9.5.6 Spain: FISABIO 36
9.5.7 Italy: PEDIANET database 36
9.5.8 Italy: ARS database 37
9.5.9 United Kingdom: CPRD & HES 37
9.5.10 France: Système National des Données de Santé (SNDS) 38
9.5 BIAS 39
9.6 STUDY SIZE 40
9.7 DATA TRANSFORMATION 40
9.7.1 Data extraction 41
9.7.2 Data engineering and analysis 41
9.7.3 Software and Hardware 41
9.7.4 Storage 41
9.7.5 Access 42
9.7.6 Archiving and record retention 42
9.8 STATISTICAL METHODS 42
9.8.1 Analysis of Demographics and Baseline Characteristics 42
9.8.2 Hypotheses 42
9.8.3 Statistical Methods 42
9.8.4 Statistical Analysis 43
9.8.5 Missing data 45

10 RESULTS 45

10.1 UPDATES FROM APRIL REPORT: 46

10.2 DESCRIPTIVE DATA 46
10.3 OUTCOME DATA 50
10.3.1 AESIs 52
10.3.2 Control events 139
10.3.3 Pregnancy outcomes 142
11 OTHER ANALYSES 148
12 DISCUSSION 148
12.1 KEY RESULTS 148
12.2 LIMITATIONS 149
12.3 STRENGTHS 149
12.4 INTERPRETATION 149
12.5 GENERALISABILITY 150
13 CONCLUSION 150
14 ANNEXES 151
APPENDIX 1. LIST OF STAND-ALONE DOCUMENTS 151

List of Tables

Table 1 AESI 9
Table 2 Attrition diagram 11
Table 3 Crude total AESI incidence rates per 100,000 PY in the year 2017 based on narrow codes 12
Table 4 Overview of data sources used for the study 26
Table 5. List of AESI 29
Table 6 Comorbid conditions with evidence of increased COVID-19 severity 31
Table 7 Incidence rates and prevalence rates calculations for the main analyses 44
Table 8 Study flowchart 48
Table 9 Demographics of study population 49
Table 10 Number of incident events over the study period according to the narrow and broad definition for each AESI in each data source 51

List of Figures

Figure 1 Study design 27
Figure 2 Data management plan 41
Figure 3: Person-time (years) over calendar and by data source 50
Figure 4: Forest plot for incidence rates of GBS per 100,000 PY by data source and calendar year using a narrow definition 53
Figure 5: Forest plot for incidence rates of GBS per 100,000 PY by data source and age for narrow definition 54
Figure 6: Forest plot for incidence rates of ADEM per 100,000 PY by year 55
Figure 7: Forest plot of incidence rate of ADEM by age category 56
Figure 8: Incidence rates of narcolepsy per 100,000 PY and calendar year 58
Figure 9: Incidence rates of narcolepsy per 100,000 PY and age 59
Figure 10: Incidence rates of acute aseptic arthritis (broad definition) per 100,000 PY and calendar year 60
Figure 11: Incidence rates of diabetes mellitus type 1 per 100,000 PY and calendar year and age up to 40 61
Figure 12: Incidence rates of thrombocytopenia (primary & secondary). 62
Figure 13: Incidence rates of thrombocytopenia (primary and secondary) by age and data source 63
Figure 14: ITP incidence rates (primary and secondary) from IMI-ADVANCE separated by source of data (Willame et al.) 63
Figure 15: Incidence rates of microangiopathy by calendar year and data source 64
Figure 16: Incidence rate of microangiopathy by data source and age category 65
Figure 17: Incidence of heart failure by data source and calendar year 66
Figure 18: Incidence of heart failure by data source and age 67
Figure 19: Incidence of stress cardiomyopathy by data source and calendar year 68
Figure 20: Incidence of stress cardiomyopathy by data source and age 69
Figure 21 : Incidence of CAD (narrow) by data source and calendar year 70
Figure 22: Incidence of CAD (narrow) by data source and age 71
Figure 23: Incidence of CAD (broad) by data source and age 72
Figure 24: Incidence of cardiovascular disease according to Global burden of disease study 73
Figure 25: Incidence of arrhythmia by data source and calendar year 74
Figure 26: Incidence of arrhythmia by data source and age 75
Figure 27: Incidence of arrhythmia in the UK Biobank 76
Figure 28 Incidence of myocarditis alone by data source and calendar year 77
Figure 29 Incidence of myocarditis alone by data source and age 78
Figure 30 Incidence of myocarditis and pericarditis by data source and calendar year 79

Figure 31 Incidence of myocarditis and pericarditis by data source and age 80
Figure 32 Incidence of DIC by data source and calendar year 81
Figure 33 Incidence of DIC by data source and age 82
Figure 34 Incidence of Thrombotic microangiopathy by data source and calendar year 83
Figure 35 Incidence of TTP by data source and age 84
Figure 36 Incidence of VTE by data source and calendar year 85
Figure 37 Incidence of VTE by data source and age 86
Figure 38 Incidence of Cerebral venous thrombosis by data source and calendar year 87
Figure 39 Incidence of Cerebral venous thrombosis by data source and age 88
Figure 40 Incidence of Ischemic stroke by data source and calendar year 89
Figure 41 Incidence of Ischemic stroke by data source and age 90
Figure 42 Incidence of hemorrhagic stroke by data source and calendar year 91
Figure 43 Incidence of hemorrhagic stroke by data source and age 92
Figure 44 Incidence of single organ cutaneous vasculitis by data source and calendar year 93
Figure 45 Incidence of single organ cutaneous vasculitis by data source and age 94
Figure 46 Incidence of Acute Liver Injury by data source and calendar year 95
Figure 47 Incidence of acute liver injury by data source and age 96
Figure 48 Incidence of Acute kidney Injury by data source and calendar year 97
Figure 49 Incidence of acute kidney injury by data source and age 98
Figure 50 Incidence of generalized convulsions by data source and calendar year 99
Figure 51 Incidence of generalized convulsions by data source and age 100
Figure 52 Incidence of meningoencephalitis by data source and calendar year 101
Figure 53 Incidence of meningoencephalitis by data source and age 102
Figure 54 Incidence of transverse myelitis by data source and calendar year 103
Figure 55 Incidence of transverse myelitis by data source and age 104
Figure 56 Incidence of ARDS by data source and calendar year 105
Figure 57 Incidence of ARDS by data source and age 106
Figure 58 Incidence of erythema multiforme by data source and calendar year 107
Figure 59 Incidence of erythema multiforme by data source and age 108
Figure 60 Incidence of chilblain like lesions by data source and calendar year 109
Figure 61 Incidence of chilblain like lesions by data source and age 110
Figure 62 Incidence of anosmia/ageusia by data source and calendar year 111
Figure 63 Incidence of anosmia/ageusia by data source and age (using broad definition) 112
Figure 64 Incidence of anaphylaxis by data source and calendar year 113
Figure 65 Incidence of anaphylaxis by data source and age 114
Figure 66 Incidence of multi-system inflammatory condition by data source and calendar year 115
Figure 67 Monthly incidence of multi-system inflammatory syndrome in 2020 in children by data source 116
Figure 68 Incidence of multi-system inflammatory syndrome in 2020 in children by age and data source 116
Figure 69 Incidence of multi-system inflammatory condition by data source and age 117
Figure 70 Incidence of death by data source and calendar year 118
Figure 71 Incidence of death by data source and age 119
Figure 72 Incidence of sudden death by data source and calendar year 120
Figure 73 Incidence of sudden death by data source and age 121
Figure 74 Incidence of COVID in 2020 by data source and age 122
Figure 75 Weekly incidence of COVID-19 for each data source 123
Figure 76 Incidence of VTE with TP by data source and calendar year 124
Figure 77 Incidence of VTE with TP by data source and age 125

Figure 78 Incidence of VTE without TP by data source and calendar year 126
Figure 79 Incidence of VTE without TP by data source and age..... 127
Figure 80 Incidence of CVST with TP by data source and calendar year..... 128
Figure 81 Incidence of CVST with TP by data source and age 129
Figure 82 Incidence of CVST without TP by data source and calendar year..... 130
Figure 83 Incidence of CVST without TP by data source and age 131
Figure 84 Incidence of Arterial thrombotic events with TP by data source and calendar year 132
Figure 85 Incidence of Arterial thrombotic events with TP by data source and age 133
Figure 86 Incidence of Arterial thrombotic events without TP by data source and calendar year 134
Figure 87 Incidence of Arterial thrombotic events without TP by data source and age..... 135
Figure 88 Incidence of Arterial or VTE with TP by data source and calendar year 136
Figure 89 Incidence of Arterial or VTE with TP by data source and age 137
Figure 90 Incidence of Arterial or VTE without TP by data source and calendar year 138
Figure 91 Incidence of Arterial or VTE without TP by data source and age 139
Figure 92 Incidence of colonic diverticulitis by data source and year 140
Figure 93 Incidence of hypertension by data source and calendar year..... 141
Figure 94 Incidence of hypertension by data source and age 142
Figure 95 Prevalence rates of pregnancy outcome (/1,000 pregnant women)..... 146

1 Abstract

Title: Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines

Main authors:
Drs. C. Willame, University Medical Center Utrecht, Utrecht, The Netherlands
Prof. dr. M.C.J.M. Sturkenboom, University Medical Center Utrecht, The Netherlands.

Rationale and background:
The global rapid spread of COVID-19 caused by the SARS-CoV-2 triggered the need for developing vaccines to control for this pandemic. This study aimed to generate background incidence rates of adverse events of special interest (AESI) that may be used to monitor benefit-risk profile of COVID-19 vaccines.

Research question and objectives:

Co-primary:

- To estimate the incidence rates of adverse events of special interest (AESI) in the general population by calendar year and data source over the period 2017 to 2020.
- To estimate the incidence of pregnancy outcomes among pregnant women aged between 12 to 55 years old by calendar year and data source over the period 2017 to 2020.
- To estimate the weekly and monthly incidence rates of COVID-19 (overall and by severity level) in 2020 by data source.
- To estimate the monthly incidence rates of multisystem inflammatory syndrome in children (MIS-C) aged between 0 to 19 years old in 2020 by data source.

Secondary:

- To estimate the incidence rates of AESI in the general population by calendar year, sex, age group, and data source over the period 2017 to 2020.
- To estimate the incidence rates of AESI in the general population by month, sex, age group, and data source over the period 2017 to 2020.
- To estimate the incidence rates of multisystem inflammatory syndrome (MIS-C) in children in 2020 by month, sex, age group, and data source.
- To estimate the prevalence of high-risk medical conditions for developing severe COVID-19 by year and data source over the period 2017 to 2020.
- To estimate the incidence rates of AESI in the at-risk population for developing severe COVID-19 by calendar year, sex, age group, and data source over the period 2017 to 2020.

Study design:
A retrospective multi-database dynamic cohort study, conducted in 2021 and covering data from 2017 to 2020 (2010-2013 for Denmark and 2014-2017 for Germany), until the date of last data availability for each data source.

Population:
The study population included all individuals observed in one of the participating data sources for at least one day during the study period and who had at least 1 year of data availability before cohort entry, except for individuals with data available since birth.

Variables of interest are

- Person-time: birth and death dates as well as periods of observation.
- Events: dates of medical and/or procedure and/or prescription/dispensing codes to identify AESI, pregnancy outcomes and at-risk medical conditions.

Table 1 ***AESI***

Body system / Classification	**AESI**
Auto-immune diseases	Guillain-Barré Syndrome (GBS)
	Acute disseminated encephalomyelitis (ADEM)
	Narcolepsy
	Acute aseptic arthritis
	(Type I) Diabetes
	(Idiopathic)Thrombocytopenia
Cardiovascular system	Acute cardiovascular injury including: Microangiopathy, Heart failure, Stress cardiomyopathy, Coronary artery disease, Arrhythmia, Myocarditis alone and Myocarditis/pericarditis
Circulatory system	Coagulation disorders including: Disseminated intravascular coagulation, Venous thromboembolism (including Pulmonary embolism and Deep vein thrombosis), Thrombotic microangiopathy, Hemorrhagic stroke, Ischemic stroke, Cerebral venous thrombosis, thrombotic thrombocytopenia syndrome (TTS)
	Single Organ Cutaneous Vasculitis
Hepato-gastrointestinal and renal system	Acute liver injury
	Acute kidney injury
Nerves and central nervous system	Generalized convulsion
	Meningoencephalitis
	Transverse myelitis
Respiratory system	Acute respiratory distress syndrome
Skin and mucous membrane, bone and joints system	Erythema multiforme
	Chilblain – like lesions
Other system	Anosmia, ageusia
	Anaphylaxis
	Multisystem inflammatory syndrome in children (MIS-C)
	Death (any causes)
	Coronavirus disease 2019 (COVID-19)
	Sudden Death

Pregnancy outcomes:

Pregnancy outcome - Maternal	Gestational Diabetes
	Pre-eclampsia
	Maternal death
Pregnancy outcome - Neonates	Fetal growth restriction
	Spontaneous abortions

	Stillbirth
	Preterm birth
	Major congenital anomalies
	Microcephaly
	Neonatal death
	Termination Of Pregnancy for Fetal Anomaly (TOPFA)

Control events: colonic diverticulitis, hypertension

Data sources:
This study included 10 data sources from 7 European countries (Denmark, Germany, France, Italy, Netherlands, Spain, United Kingdom). Data sources contain health insurance data (GePaRD, SNDS), hospitalisation record linkage data (PHARMO, Danish registries (DCE-AU), SIDIAP, ARS) or data from general practitioners (CPRD, PEDIANET, BIFAP, FISABIO). For this final report data from 9 data sources were included.

Study size:
The study population for the total study comprised approximately 141.6 million individuals. In this final report, a total number of 45 million individuals were included.

Data analysis:
Incidence rates (and 95%CI) of AESI by calendar year were calculated by dividing the number of incident (new) cases by the total person-time (for AESIs) at risk.

Prevalence of pregnancy outcomes by calendar year were calculated by dividing the number of maternal or neonates' events by the total number of pregnant women.

Prevalence rates of at-risk medical conditions for developing severe COVID-19 by calendar year were calculated by dividing the number of existing cases in a year by the average of the total number of persons recorded monthly. Incidence rates (and 95%CI) of AESI among at-risk populations were also computed.

Results

This report comprises background rate data on AESI from 6 countries (UK, ES, IT, DK, NL, DE) and 9 data sources (BIFAP, Pedianet (children only), CPRD, ARS, Danish registries, FISABIO, SIDIAP, PHARMO, GeParD). Data from France (SNDS) could not be generated in a timely manner due to administrative constraints in data release. Data sources included different subpopulations based on the availability of numerator data of the observed persontime (Hosp= hospital based, PC= primary care, HOSP-PC= overlap between hospitalization and primary care).

The attrition diagram is shown in Table 2. The incidence rates (per 100,00 PY) for each AESIs and per databases are presented in Table 3. Databases were classified according to the type of datasource (GP only, inpatient and GP, inpatient only, inpatient and emergency room, in-and-outpatient and claims insurance database).

Table 2 ***Attrition diagram***

	ARS		PEDIANET		GEPARD		DCE-AU		CPRD		FISABIO		BIFAP		BIFAP_PC_HOSP		SIDIAP_PC		SIDIAP_PC_HOSP		PHARMO_HOSP		PHARMO_PCHOSP	
Subjects disposition	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Subjects present during 2016-2020	3903982	100	199174	100	764269	100	6143446	100	5309729	100	6646465	100	14007619	100	14007619	100	6550377	100	6550377	100	9356863	100	9356863	100
Subjects with sex or birth date missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	<0.01	3	<0.01
Subjects without valid birth date	4245	0.11	15	0.01	-	-	-	-	1394	0.03	11936	0.18	46	<0.01	46	<0.01	-	-	-	-	2063	0.02	2063	0.02
Subjects with death date before study entry	2082	0.05	-	-	12	<0.01	-	-	-	-	2018	0.03	-	-	-	-	-	-	-	-	-	-	-	-
Subjects without overlap in observed time	127559	3.27	17869	8.97	214018	28	188086	3.06	537702	10.13	744596	11.2	3741105	26.71	9583730	68.42	263079	4.02	4769969	72.82	163702	1.75	8858600	94.67
Subjects without sufficient look-back period	702494	17.99	-	-	-	-	-	-	81923	1.54	1355	0.02	-	-	-	-	81725	1.25	22169	0.34	6263	0.07	-	-
Total number of subjects included in the study	3067602	78.58	181290	91.02	550239	72	5955360	96.94	4688710	88.3	5886560	88.57	10266468	73.29	4423843	31.58	6205573	94.74	1758239	26.84	9184832	98.16	496197	5.3

*Aarhus University (DCE-AU) did not extract 2017-2020 but 2010-2013. GeParD covers data from 2014 to 2017 and data from only 800.000 out of 25 million were included.

Table 3 Crude total AESI incidence rates per 100,000 PY in the year 2017 based on narrow codes

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
1. GBS	BIFAP_PC:1.34 (1.12-1.60) SIDIAP_PC: 2.17 (1.80-2.59) CPRD: 1.54 (1.16-2.0) Pedianet (2018): 0.68 (0.02-3.79)	BIFAP_PC_HOSP: 2.05 (1.65-2.53) PHARMO_PC_HOSP: 3.53 (2.02-5.73) SIDIAP_PC_HOSP: 3.83 (2.91-4.95)	PHARMO_HOSP: 1.39 (1.15-1.65)	ARS: 3.98 (3.16-4.94)	DCE-AU (2010): 3.06 (2.63-3.55) FISABIO: 4.42 (3.87-5.02)	GePaRD: 2.32 (1.16-4.14)	No clear pattern	Increase with age consistently and lowering after 80
2. ADEM	BIFAP_PC: 0.15 (0.08-0.25) SIDIAP_PC: 0.02 (0.00-0.10) CPRD: NA Pedianet: no event	BIFAP_PC_HOSP: 0.35 (0.20-0.58) SIDIAP_PC_HOSP (2018): 0.07 (0.0-0.37)	PHARMO_HOSP: no event	ARS: 0.05 (0.00-0.27)	DCE-AU:no event FISABIO: 0.45 (0.29-0.67)	GePaRD: no event	BIFAP: Decreased in rates in 2020	Lower rates in 80+
3. Narcolepsy	BIFAP_PC: 2.35 (2.05-2.68) SIDIAP_PC: 0.97 (0.73-1.26) CPRD: 1.13 (0.81-1.53) Pedianet: no event	BIFAP_PC_HOSP: 1.75 (1.37-2.19) PHARMO_PC_HOSP: 0.22 (0.01-1.23) SIDIAP_PC_HOSP: 1.39 (0.86-2.12)	PHARMO_HOSP: 0.30 (0.20-0.44)	ARS: 0.49 (0.24-0.90)	DCE-AU (2010): 3.25 (2.81-3.76) FISABIO: 2.35 (1.96-2.80)	GePaRD: NA	Decreased in rates in 2020	Higher rates in 20-29
4. Acute Aseptic Arthritis	No narrow codes	No narrow codes	No narrow codes	No narrow codes	No narrow codes	-	-	-
5. Type 1 Diabetes mellitus (only up to 40 years old)	BIFAP_PC: 11.63 (10.59-12.74) SIDIAP_PC: 23.80 (21.94-25.78) CPRD: 37.65 (34.80-40.66) Pedianet (2018): no event	BIFAP_PC_HOSP: 21.55 (19.38-23.90) SIDIAP_PC_HOSP: 34.15 (29.82-38.93) PHARMO_PC_HOSP: 42.92 (34.24-53.14)	PHARMO_HOSP: 16.67 (15.40-18.02)	ARS: 20.28 (17.05-23.94)	DCE-AU: Not included FISABIO: 38.25 (35.77-40.85)	GePaRD: NA	No clear pattern	No clear pattern
6. Thrombocytopenia	BIFAP_PC: 45.96 (44.61-47.35) SIDIAP_PC: 99.93 (97.34-102.56) CPRD: 21.63 (20.15-23.20)	BIFAP_PC_HOSP: 65.97 (63.55-68.47) PHARMO_PC_HOSP: 37.04 (31.65-43.09) SIDIAP_PC_HOSP: 142.42	PHARMO_HOSP: 19.70 (18.79-20.65)	ARS: 40.56 (37.85-43.43)	DCE-AU (2010): 18.40 (17.31-19.56) FISABIO: 150.95 (147.66-154.29)	GePaRD: NA	Decreased in rates in 2020	Increase with age

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
	Pedianet (2018): 6.81 (3.26-12.52)	(136.47-148.56)						
7. Microangiopathy	BIFAP_PC: 0.53 (0.39-0.70) SIDIAP_PC: 0.35 (0.22-0.54) CPRD: 0.63 (0.40-0.95) Pedianet: no event	BIFAP_PC_HOSP: 0.64 (0.42-0.93) PHARMO_PC_HOSP: 1.32 (0.49-2.88) SIDIAP_PC_HOSP: 5.67 (4.54-7.01)	PHARMO_HOSP: 1.15 (0.94-1.40)	ARS: 0.79 (0.45-1.28)	DCE-AU (2010):3.63 (3.16-4.17) FISABIO: 7.33 (6.62-8.09)	GePaRD: NA	Slight decrease in rates in 2020	Increase with age
8. Heart failure	BIFAP_PC: 231.53 (228.47-234.62) SIDIAP_PC: 241.06 (237.02-245.14) CPRD: 155.80 (151.76-159.92) Pedianet: no event	BIFAP_PC_HOSP: 359.59 (353.89-365.37) PHARMO_PC_HOSP: 426.13 (407.28-445.62) SIDIAP_PC_HOSP: 491.57 (480.43-502.90)	PHARMO_HOSP: 143.97 (141.49-146.48)	ARS: 725.90 (714.20-737.74)	DCE-AU (2010): 276.39 (272.06-280.79) FISABIO: 510.84 (504.75-516.98)	GePaRD: 100.67 (91.77-110.19)	Decreased in rates in 2020	Increase with age
9. Stress cardiomyopathy	BIFAP_PC: 0.24 (0.15-0.36) SIDIAP_PC: 0.05 (0.01-0.15) CPRD: no event Pedianet: no event	BIFAP_PC_HOSP: 1.58 (1.23-2.01) SIDIAP_PC_HOSP: 3.63 (2.73-4.72)	PHARMO_HOSP: no event	ARS: 7.12 (6.01-8.38)	DCE-AU: no event FISABIO: 3.48 (2.99-4.01)	GePaRD: NA	No clear pattern	Increase with age
10. Coronary artery disease	BIFAP_PC: 84.99 (83.14-86.87) SIDIAP_PC: 83.14 (80.79-85.55) CPRD: 165.23 (161.07-169.47) Pedianet: no event	BIFAP_PC_HOSP: 119.21 (115.94-122.55) PHARMO_PC_HOSP: 253.21 (238.76-268.32) SIDIAP_PC_HOSP: 124.61 (119.05-130.36)	PHARMO_HOSP: 109.75 (107.59-111.94)	ARS: 322.04 (314.27-329.94)	DCE-AU (2010): 196.65 (193.01-200.36) FISABIO: 195.01 (191.27-198.82)	GePaRD: 169.30 (155.58-183.90)	Decreased in rates in 2020	Increase with age
11. Arrhythmia	BIFAP_PC: 719.40 (713.97-724.85) SIDIAP_PC: 885.49 (877.71-893.32) CPRD: 495.71 (488.47-503.03)	BIFAP_PC_HOSP: 872.38 (863.45-881.38) PHARMO_PC_HOSP: 1497.25 (1461.42-1533.74) SIDIAP_PC_HOSP: 1345.61	PHARMO_HOSP: 353.91 (350.01-357.84)	ARS: 1207.50 (1192.33-1222.80)	DCE-AU (2010): 650.83 (644.15-657.58) FISABIO: 1161.45 (1152.16-1170.79)	GePaRD: NA	Reduction in rates in 2020	Increase with age

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
	Pedianet (2018): 130.88 (113.02-150.76)	(1327.02-1364.40)						
12.1 Myocarditis/ pericarditis	BIFAP_PC: 16.54 (15.73-17.38) SIDIAP_PC: 18.63 (17.53-19.79) CPRD: 12.38 (11.27-13.58) Pedianet (2018): 3.40 (1.11-7.94)	BIFAP_PC_HOSP: 18.73 (17.45-20.08) PHARMO_PC_HOSP: 23.59 (19.33-28.51) SIDIAP_PC_HOSP: 31.61 (28.84-34.57)	PHARMO_HOSP: 4.18 (3.77-4.62)	ARS: 33.49 (31.03-36.11)	DCE-AU (2010): 16.05 (15.03-17.13) FISABIO: 11.01 (10.14-11.94)	GePaRD: NA	Reduction in rates in 2020	Higher in 20-39
12.2 Myocarditis alone	BIFAP_PC: 2.00 (1.72-2.30) SIDIAP_PC: 0.85 (0.62-1.12) CPRD: 2.86 (2.34-3.47) Pedianet (2018): 0.68 (0.02-3.79)	BIFAP_PC_HOSP: 2.50 (2.05-3.03) PHARMO_PC_HOSP: 15.87 (12.42-19.99) SIDIAP_PC_HOSP: 3.43 (2.56-4.50)	PHARMO_HOSP: 1.31 (1.08-1.57)	ARS: 6.28 (5.24-7.47)	DCE-AU (2010): 3.66 (3.19-4.20) FISABIO: 2.35 (1.96-2.80)	GePaRD: NA	No clear pattern	Higher in 20-39
13.1 Disseminated Intravascular Coagulation	BIFAP_PC: 0.18 (0.10-0.29) SIDIAP_PC: 0.02 (0.00-0.10) CPRD: 0.14 (0.04-0.32) Pedianet: no event	BIFAP_PC_HOSP: 0.35 (0.20-0.58) PHARMO_PC_HOSP: 0.44 (0.05-1.59) SIDIAP_PC_HOSP: 4.16 (3.19-5.32)	PHARMO_HOSP: 0.67 (0.51-0.86)	ARS: 2.16 (1.57-2.90)	DCE-AU (2010): 2.79 (2.38-3.26) FISABIO: 5.90 (5.27-6.59)	GePaRD: NA	No clear pattern	Increase with age
13.2 Thrombotic Thrombocytopenic Purpura/Thrombotic microangiopathy	BIFAP_PC: 0.88 (0.70-1.09) SIDIAP_PC: 0.34 (0.20-0.52) CPRD: 0.30 (0.15-0.54) Pedianet: no event	BIFAP_PC_HOSP: 0.87 (0.62-1.20) PHARMO_PC_HOSP: 0.88 (0.24-2.26) SIDIAP_PC_HOSP: 1.52 (0.96-2.28)	PHARMO_HOSP: 0.50 (0.37-0.67)	ARS: 0.79 (0.45-1.28)	DCE-AU (2010): 0.97 (0.74-1.27) FISABIO: 1.52 (1.21-1.89)	GePaRD: NA	Decreased in rates in 2020	Increase with age
13.3 Venous Thromboembolism	BIFAP_PC: 216.28 (213.32-219.27) SIDIAP_PC: 202.99 (199.30-206.74) CPRD: 174.70 (170.42-179.06)	BIFAP_PC_HOSP: 267.34 (262.43-272.32) PHARMO_PC_HOSP: 237.93 (223.93-252.59) SIDIAP_PC_HOSP: 254.94 (246.95-263.12)	PHARMO_HOSP: 46.83 (45.45-48.27)	ARS: 226.66 (220.16-233.30)	DCE-AU (2010): 181.90 (178.40-185.47) FISABIO: 201.01 (197.22-204.87)	GePaRD: NA	No clear pattern	Increase with age

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
	Pedianet (2018): 0.68 (0.02-3.79)							
13.4 Ischemic stroke	BIFAP_PC: 123.09 (120.86-125.35) SIDIAP_PC: 161.09 (157.80-164.43) CPRD: 132.19 (128.47-135.98) Pedianet: no event	BIFAP_PC_HOSP: 195.30 (191.11-199.56) PHARMO_PC_HOSP: 269.56 (254.64-285.12) SIDIAP_PC_HOSP: 201.30 (194.22-208.58)	PHARMO_HOSP: 76.74 (74.94-78.58)	ARS: 263.60 (256.59-270.76)	DCE-AU: 168.39 (165.02-171.83) FISABIO: 308.15 (303.43-312.93)	GeParRD: NA	No clear pattern	Increase with age
13.5 Hemorrhagic stroke	BIFAP_PC: 18.53 (17.68-19.42) SIDIAP_PC: 13.59 (12.64-14.58) CPRD: 7.95 (7.06-8.92) Pedianet: (2018): 2.04 (0.42-5.97)	BIFAP_PC_HOSP: 33.41 (31.69-35.19) PHARMO_PC_HOSP: 36.82 (31.45-42.84) SIDIAP_PC_HOSP: 38.54 (35.48-41.80)	PHARMO_HOSP: 18.75 (17.87-19.67)	ARS: 95.54 (91.34-99.89)	DCE-AU: 41.54 (39.88-43.26) FISABIO: 62.17 (60.07-64.32)	GePaRD: NA	No clear pattern	Increase with age
13.8 Cerebral venous thrombosis	BIFAP_PC: 0.32 (0.21-0.45) SIDIAP_PC: 0.14 (0.06-0.28) CPRD: 0.14 (0.04-0.32) Pedianet (2020): 1.26 (0.03-7.0)	BIFAP_PC_HOSP: 0.47 (0.29-0.73) PHARMO_PC_HOSP: 0.88 (0.24-2.26) SIDIAP_PC_HOSP: 0.59 (0.27-1.13)	PHARMO_HOSP: 0.79 (0.62-1.00)	ARS: 1.33 (0.87-1.93)	DCE-AU: 0.97 (0.74-1.27) FISABIO: 1.15 (0.88-1.47)	GePaRD: NA	No clear pattern	No clear pattern
14. SOCV	BIFAP_PC: 6.16 (5.67-6.68) SIDIAP_PC: 4.73 (4.18-5.33) CPRD: 15.93 (14.66-17.29) Pedianet (2018): 26.55 (18.88-36.30)	BIFAP_PC_HOSP: 6.78 (6.02-7.61) PHARMO_PC_HOSP: 10.36 (7.61-13.78) SIDIAP_PC_HOSP: 7.65 (6.32-9.18)	PHARMO_HOSP: 1.49 (1.24-1.76)	ARS: 7.07 (5.96-8.32)	DCE-AU (2010): 14.31 (13.36-15.35) FISABIO: 31.81 (30.31-33.36)	GePaRD: NA	Reduction in rates in 2020	Increase with age
15. Acute liver injury	BIFAP_PC: 20.69 (19.78-21.63) SIDIAP_PC: 25.66 (24.36-27.01)	BIFAP_PC_HOSP: 22.73 (21.31-24.31) PHARMO_PC_HOSP: 34.62 (29.42-40.48)	PHARMO_HOSP: 8.15 (7.57-8.76)	ARS: 36.34 (33.77-39.05)	DCE-AU 2010: 20.65 (19.49-21.88) FISABIO: 48.00 (46.15-49.89)	GePaRD: 7.39 (5.15-10.27)	Reduction in rates in 2020	Increase with age

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
	CPRD: 6.16 (5.38-7.03) Pedianet (2019): 3.86 (1.25-9.00)	SIDIAP_PC_HOSP: 41.45 (38.27-44.82)						
16. Acute kidney injury	BIFAP_PC: 70.75 (69.07-72.47) SIDIAP_PC: 576.87 (570.61-583.18) CPRD: 118.22 (114.71-121.81) Pedianet (2018): 1.36 (0.16-4.92)	BIFAP_PC_HOSP: 151.37 (147.68-155.13) PHARMO_PC_HOSP: 474.05 (454.15-494.60) SIDIAP_PC_HOSP: 992.56 (976.67-1008.67)	PHARMO_HOSP: 223.61 (220.51-226.73)	ARS: 348.59 (340.52-356.80)	DCE-AU (2010): 185.64 (182.10-189.25) FISABIO: 612.57 (605.92-619.29)	GeParRD: 46.63 (40.62-53.28)	Reduction in rates in 2020 (except FISABIO)	consistent and strong increase with age
17. Generalized convulsion	BIFAP_PC: 65.39 (63.77-67.04) SIDIAP_PC: 89.28 (86.83-91.77) CPRD: 112.69 (109.27-116.20) Pedianet (2018): 150.85 (132.61-172.10)	BIFAP_PC_HOSP: 90.65 (87.80-93.56) PHARMO_PC_HOSP: 154.04 (142.81-165.91) SIDIAP_PC_HOSP: 152.36 (146.20-158.71)	PHARMO_HOSP: 47.12 (45.71-48.56)	ARS: 165.34 (159.80-171.03)	DCE-AU 2010: 219.70 (215.84-223.63) FISABIO: 230.92 (226.84-235.05)	GePaRD: 73.38 (65.83-81.56)	Reduction in rates in 2020	Highest in children and elderly
18. Meningoencephalitis	BIFAP_PC: 4.36 (3.95-4.81) SIDIAP_PC: 1.52 (1.21-1.87) CPRD: 2.67 (2.16-3.26) Pedianet (2019): 2.31 (0.48-6.76)	BIFAP_PC_HOSP: 8.62 (7.76-9.55) PHARMO_PC_HOSP: 4.85 (3.04-7.34) SIDIAP_PC_HOSP: 3.63 (2.73-4.72)	PHARMO_HOSP: 1.53 (1.29-1.81)	ARS: 5.99 (4.97-7.15)	DCE-AU 2010: 3.84 (3.36-4.39) FISABIO: 10.32 (9.47-11.22)	GePaRD: NA	No clear pattern	Increase with age
19. Transverse myelitis	BIFAP_PC: 0.16 (0.09-0.26) SIDIAP_PC: 0.67 (0.47-0.92) CPRD: 1.05 (0.74-1.44) Pedianet: no event	BIFAP_PC_HOSP: 0.26 (0.13-0.36) PHARMO_PC_HOSP: 0.22 (0.01-1.23) SIDIAP_PC_HOSP: 0.79 (0.41-1.38)	PHARMO_HOSP: 0.23 (0.14-0.36)	ARS: 1.52 (1.03-2.16)	DCE-AU (2010): 1.32 (1.05-1.66) FISABIO: 0.60 (0.41-0.85)	GePaRD: NA	No clear pattern	same magnitude across ages
20. ARDS	BIFAP_PC: 48.24 (46.85-49.66) SIDIAP_PC: 7.74 (7.04-8.50)	BIFAP_PC_HOSP: 74.17 (71.60-76.81) PHARMO_PC_HOSP: 38.36 (32.87-44.50)	PHARMO_HOSP: 23.55 (22.55-24.58)	ARS: 30.69 (28.33-33.19)	DCE-AU (2010): 96.38 (93.84-98.99) FISABIO: 141.15	GeParRD: 5.48 (3.58-8.03)	No clear pattern	Increase with age

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
	CPRD: 22.87 (21.34-24.48) Pedianet (2018): 2.72 (0.74-6.97)	SIDIAP_PC_HOSP: 13.13 (11.37-15.08)			(137.97-144.38)			
21.Erythema multiforme	BIFAP_PC: 7.02 (6.49-7.57) SIDIAP_PC: 8.64 (7.90-9.45) CPRD: 8.28 (7.37-9.27) Pedianet (2018): 8.85 (4.71-15.13)	BIFAP_PC_HOSP: 7.68 (6.86-8.56) PHARMO_PC_HOSP: 1.54 (0.62-3.18) SIDIAP_PC_HOSP: 11.02 (9.41-12.82)	PHARMO_HOSP: 0.37 (0.25-0.52)	ARS: 9.77 (8.46-11.23)	DCE-AU (2010): 4.63 (4.10--5.23) FISABIO: 15.09 (14.07-16.17)	GePaRD: NA	Decreased in rates in 2020	Higher in children and elderly
22. Chilblain–like lesions	BIFAP_PC: 31.36 (30.24-32.51) SIDIAP_PC: 15.86 (14.84-16.93) CPRD: 12.08 (10.98-13.27) Pedianet (2018): 2.72 (0.74-6.97)	BIFAP_PC_HOSP: 25.02 (23.53-26.57) PHARMO_PC_HOSP: 2.20 (1.06-4.05) SIDIAP_PC_HOSP: 15.97 (14.02-18.11)	PHARMO_HOSP: 0.01 (0.00-0.06)	ARS: 0.25 (0.08-0.57)	DCE-AU (2010): 0.46 (0.31-0.68) FISABIO: 64.05 (62.92-66.24)	GePaRD: NA	No clear pattern	No clear age patterns
23. Anosmia, Ageusia	BIFAP_PC: 10.19 (9.55-10.85) SIDIAP_PC: 18.69 (17.58-19.85) CPRD: 22.38 (20.86-23.97) Pedianet: no event	BIFAP_PC_HOSP: 14.29 (13.18-15.48) PHARMO_PC_HOSP: 23.15 (18.93-28.02) SIDIAP_PC_HOSP: 19.13 (17.00-21.47)	PHARMO_HOSP: 0.12 (0.06-0.22)	ARS: 0.05 (0.00-0.27)	DCE-AU (2010): 1.14 (0.89-1.46) FISABIO: 28.82 (27.39-30.30)	GePaRD: NA	Increased in 2020	No clear age pattern
24. Anaphylaxis	BIFAP_PC: 5.64 (5.17-6.14) SIDIAP_PC: 11.33 (10.47-12.24) CPRD: 19.71 (18.29-21.20) Pedianet (2018): 3.40 (1.11-7.94)	BIFAP_PC_HOSP: 8.81 (7.94-9.75) PHARMO_PC_HOSP: 10.58 (7.80-14.03) SIDIAP_PC_HOSP: 14.96 (13.09-17.06)	PHARMO_HOSP: 3.21 (2.85-3.60)	ARS: 7.90 (6.73-9.22)	DCE-AU (2010): 10.06 (9.26-10.93) FISABIO: 22.07 (20.84-23.37)	GePaRD: 6.85 (4.34-10.28)	No clear pattern	Highest in children
26. MIS	BIFAP_PC: 0.59 (0.45-0.77) SIDIAP_PC: 0.23 (0.12-0.39) CPRD: 0.50 (0.29-1.78) Pedianet (2018): 3.40 (1.11-7.94)	BIFAP_PC_HOSP: 0.66 (0.44-0.96) PHARMO_PC_HOSP: 1.54 (0.62-3.18) SIDIAP_PC_HOSP: 0.53 (0.23-1.04)	PHARMO_HOSP: 0.42 (0.30-0.58)	ARS: 0.83 (0.49-1.34)	DCE-AU (2010): 0.66 (0.48-0.91) FISABIO: 2.61 (2.20-3.08)	GePaRD: 0.42 (0.05-1.52)	No clear pattern	Most of events in [0-19]

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
27. Death	BIFAP_PC: 875.73 (869.78-881.71) CPRD: 860.68 (851.17-870.27) Pedianet: not reported	BIFAP_PC_HOSP: 955.43 (946.14-964.78) PHARMO_PC_HOSP: not reported SIDIAP_PC_HOSP: 985.21 (969.47-1001.14)	PHARMO_HOSP: not reported	ARS: 1859.40 (1840.72-1878.22)	DCE-AU (2010): 895.76 (977.57-994.03) FISABIO: 796.69 (789.12-804.31)	GePaRD: NA	Increase in 2020	Increase with age
28. Sudden death	BIFAP_PC: 2.22 (1.93-2.54) SIDIAP_PC: 205.17 (201.46-208.94) CPRD: 1.07 (0.76-1.47) Pedianet: not reported	BIFAP_PC_HOSP: 1.87 (1.48-2.33) PHARMO_PC_HOSP: not reported SIDIAP_PC_HOSP: 218.85 (211.45-226.43)	PHARMO_HOSP: not reported	ARS: 3.19 (2.46-4.07)	DCE-AU (2010): 60.15 (58.15-62.22) FISABIO: 26.46 (25.09-27.87)	GePaRD: NA	Increase in rates in 2020	Increase with age
29. VTE with TP	BIFAP_PC: 0.11 (0.05-0.19) SIDAP_PC: 0.18 (0.08-0.32) CPRD: 0.14 (0.04-0.32) Pedianet: no event	BIFAP_PC_HOSP: 0.19 (0.06-0.37) PHARMO_PC_HOSP: 0.88 (0.24-2.26) SIDIAP_PC_HOSP: 2.44 (1.72-3.36)	PHARMO_HOSP: 0.70 (0.54-0.90)	ARS: 1.52 (1.03-2.16)	DCE-AU (2010): 0.29 (0.18-0.47) FISABIO: 3.18 (2.71-3.69)	GePaRD: NA	Slight increase in rates in 2020	Increase with age
30.VTE without TP	BIFAP_PC: 216.21 (213.25-219.20) SIDIAP_PC: 202.82 (199.12-206.56) CPRD: 174.56 (170.29-178.92) Pedianet: no event	BIFAP_PC_HOSP: 267.20 (262.29-272.18) PHARMO_PC_HOSP: 237.05 (223.07-251.67) SIDIAP_PC_HOSP: 253.28 (245.32-261.44)	PHARMO_HOSP: 46.22 (44.82-47.65)	ARS: 225.38 (218.90-232.00)	DCE-AU (2010): 181.70 (178.20-185.27) FISABIO: 199.20 (195.42-203.04)	GePaRD: NA	No clear pattern	Increase with age
31.CVST with TP	no event	SIDIAP_PC_HOSP (2018): 0.07 (0.00-0.37)	No event	ARS: 0.05 (0.00-0.27)	DCE-AU (2013): 0.02 (0.00-0.14) FISABIO: 0.09 (0.03-0.22)	GePaRD: NA	-	-
32. CVST without TP	BIFAP_PC: 0.32 (0.21-0.45) SIDIAP_PC: 0.14 (0.06-0.28) CPRD: 0.14 (0.04-0.32)	BIFAP_PC_HOSP: 0.47 (0.29-0.73) PHARMO_PC_HOSP: 0.88 (0.24-2.26) SIDIAP_PC_HOSP: 0.59 (0.27-1.13)	PHARMO_HOSP: 0.79 (0.62-1.00)	ARS: 1.28 (0.83-1.87)	DCE-AU (2010):0.97 (0.74-1.27) FISABIO: 1.05 (0.79-1.37)	GePaRD: NA	No clear pattern	Slight increase with age

AESI	Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)	Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)	Incident inpatient only (PHARMO_HOSP), 2017 (95%CI)	Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)	Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)	Incidence by on claims data (GePaRD – only a subset of AESIs), 2017, (95% CI)	Calendar year change in 2020 compared to previous years	Age pattern
	Pedianet: no event							
33. Arterial with TP	BIFAP_PC: 0.12 (0.06-0.21) SIDIAP_PC: 0.11 (0.04-0.23) CPRD: 0.06 (0.01-0.20) Pedianet: no event	BIFAP_PC_HOSP: 0.07 (0.01-0.21) PHARMO_PC_HOSP: 1.10 (0.36-2.57) SIDIAP_PC_HOSP: 3.63 (2.73-4.72)	PHARMO_HOSP: 0.80 (0.63-1.01)	ARS: 1.47 (0.99-2.10)	DCE-AU (2010): 0.38 (0.25-0.58) FISABIO: 5.92 (5.28-6.61)	GePaRD: NA	No clear pattern	Increase with age
34. Arterial without TP	BIFAP_PC: 206.41 (203.53-209.33) SIDIAP_PC: 241.74 (237.70-245.82) CPRD: 293.33 (287.78-298.97) Pedianet: no event	BIFAP_PC_HOSP: 311.12 (305.82-316.49) PHARMO_PC_HOSP: 511.68 (491.01-533.00) SIDIAP_PC_HOSP: 317.86 (308.93-326.99)	PHARMO_HOSP: 182.59 (179.79-185.41)	ARS: 567.09 (556.76-577.56)	DCE-AU (2010): 357.91 (352.98-362.91) FISABIO: 478.57 (472.66-484.53)	GePaRD: NA	Decrease in rates in 2020	Increase with age
35. Arterial or VTE with TP	BIFAP_PC: 0.22 (0.14-0.34) SIDIAP_PC: 0.26 (0.15-0.44) CPRD: 0.19 (0.08-0.40) Pedianet: no event	BIFAP_PC_HOSP: 0.26 (0.13-0.46) PHARMO_PC_HOSP: 1.98 (0.91-3.77) SIDIAP_PC_HOSP: 6.00 (4.83-7.37)	PHARMO_HOSP: 1.46 (1.22-1.74)	ARS: 2.90 (2.20-3.74)	DCE-AU (2010): 0.63 0.45-0.88) FISABIO: 8.76 (7.96-9.59)	GePaRD: NA	No clear pattern	Increase with age
36.Arterial or VTE without TP	BIFAP_PC: 418.55 (414.42-422.71) SIDIAP_PC: 440.34 (434.88-445.86) CPRD: 463.74 (456.74-470.81) Pedianet: no event	BIFAP_PC_HOSP: 571.26 (564.05-578.53) PHARMO_PC_HOSP: 741.80 (716.83-767.43) SIDIAP_PC_HOSP: 562.67 (550.74-574.78)	PHARMO_HOSP: 225.77 (222.67-228.91)	ARS: 775.11 (763.02-787.36)	DCE-AU (2010): 530.96 (524.94-537.05) FISABIO: 656.10 (649.16-663.09)	GePaRD: NA	No clear pattern	Increase with age

NA: not applicable, GeParD only provided rates for a subset of the AESIs; No event= no event identified in the data source; Not reported: event not available in the datasource

2 List of abbreviations

ACCESS	vACCine covid-19 monitoring readinESS
ADEM	Acute disseminated encephalomyelitis
ADVANCE	Accelerated Development of VAccine beNefit-risk Collaboration in Europe
AESI	Adverse Event of Special Interest
AKI	Acute Kidney Injury
ALI	Acute Liver Injury
ARDS	Acute respiratory distress requiring ventilation
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CDM	Common Data Model
CEPI	Coalition for Epidemic Preparedness Innovations
CI	Confidence interval
COVID-19	Coronavirus disease 2019
DAP	Data Access Provider
DIC	Disseminated Intravascular coagulation
DNA	Desoxyribonucleic acid
DRE	Digital Research Environment
DVT	Deep Vein Thrombosis
ECDC	European Centre for Disease Prevention and Control
EMA	European Medicines Agency
EMR	Electronic Medical Records
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance.
ETL	Extract, Transform, and Load
EU PAS	The European Union electronic Register of Post-Authorisation Studies
GBS	Guillain-Barré Syndrome
GDPR	General Data Protection Regulation
GP	General Practitioner
GPP	Good Participatory Practice
HIV	Human Immunodeficiency Virus
HF	Heart Failure
ICD	International Classification of Diseases
ICMJE	International Committee of Medical Journal Editors
ICU	Intensive Care Unit
IMI	Innovative Medicines Initiative
ITP	(Idiopathic)Thrombocytopenia
MIS	Multisystem Inflammatory Syndrome
MIS-C	Multisystem Inflammatory Syndrome in children
mRNA	messenger Ribonucleic acid
NHS	National Health Service
PE	Pulmonary Embolism

QC	Quality Control
RNA	Ribonucleic acid
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SPEAC	Safety Platform for Emergency vACcines
SOCV	Single Organ Cutaneous Vasculitis
TOPFA	Termination of Pregnancy for Fetal Anomaly
VAC4EU	Vaccine monitoring Collaboration for Europe

3 Investigators

Coordinating Center University Medical Center Utrecht, The Netherlands
Corinne Willame
Miriam Sturkenboom, ACCESS coordinator
Roel Elbers
Daniel Weibel
Caitlin Dodd (until January 15, 2021)

Collaborating Institutions (by alphabetical order)	Study Sites	Key persons
Aarhus University (DCE-AU)	Denmark	Vera Ehrenstein, Reimar W. Thomsen, Johnny Kahlert
Agenzia Regionale di Sanita Toscana (ARS)	Italy	Rosa Gini, Claudia Bartolini, Olga Paoletti
Bordeaux PharmacoEpi (BPE), University of Bordeaux	France	Cecile Droz, Nicholas Moore
Leibniz Institute for Prevention Research and Epidemiology - BIPS	Germany	Ulrike Haug, Tania Schink
FISABIO	Spain	Javier Diez-Domingo, Ainara Mira-Iglesias, Carlos Vergara-Hernández
IDIAP-Jordi Gol (SIDIAP)	Spain	Bonaventura Bolibar, Villalobos Felipe
PHARMO/STIZON	The Netherlands	Josine Kuipers & Michiel Meulendijk
RIVM	The Netherlands	Hester de Melker
RTI-HS	Spain & United States of America	Susana-Perez-Gutthann Alejandro Arana
SoSeTe-Pedianet	Italy	Carlo Giaquinto, Elisa Barbieri, Luca Stona

Spanish Agency of Medicines and Medical Devices (AEMPS - BIFAP)	Spain	Consuelo Huerta, Mar Martín-Pérez, Patricia García-Poza; Airam de Burgos, María Martínez-González, Verónica Bryant
Utrecht University	The Netherlands	Patrick Souverein, Olaf Klungel, Helga Gardarsdottir, Satu Siiskonen
VAC4EU secretariat	Belgium	Patrick Mahy, Juul Klaassen, Nathalie Vigot

4 Other responsible parties

None

5 Milestones

Milestone	Planned date	Actual date	Comments
Start of data collection	October 1st, 2020	November 1st, 2020	Protocol approvals
End of data collection	December 1	To be updated	Partial data are delivered in this report.
Registration in the EU PAS register	-	September 13, 2020	-
Study progress report 1	December 15, 2020	December 15, 2020	Due to governance approvals delays and work performed to validate the data workflow process, only partial data has been provided.
Final draft report of study results	December 15, 2020	February 15, 2021	Due to governance approvals delays and work performed to validate the data workflow process, only partial data has been provided.
Update of final draft report and annexes (version 1.1)		March 3, 2021	○ Update of German data (prior years), and exclusion of meanings of events indicating suspicion of diagnosis

			○ inclusion of comments of EMA on initial draft final report ○ correction of output error in excel sheet for at risk population ○ updating of graphics to improve readability
Update of draft final report (version 1.2)	April 30, 2021		○ Refinement of algorithms (medical codes) for all AESIs ○ Inclusion of SIDIAP and PHARMO ○ Inclusion of all data from BIFAP (all regions) ○ Inclusion of 6 subtypes of coagulation disorders ○ Incidence rate of COVID-19 by severity level ○ Monthly rates ○ Updated benchmarking data
Final report of study results	June 30, 2021		○ Refinement of code lists ○ Inclusion of pregnancy outcomes ○ Inclusion of rates for myocarditis alone beyond the old myocarditis/pericarditis ○ Updated benchmarking data ○ Updated rates for all datasources ○ Revise of algorithms for GeParD

6 Rationale and background

6.1 Background

COVID-19 vaccine development has been triggered on a global level following the release of the genetic sequence of SARS-CoV-2 on 11 January 2020[1]

The landscape for COVID-19 vaccines is characterized by a wide range of technology platforms including nucleic acid (DNA and RNA), virus-like particle, peptide, viral vector (replicating and non-replicating), recombinant protein, live attenuated virus and inactivated virus approaches.

6.2 Rationale for the study

When new vaccines are launched on a market and used at a large scale, monitoring of adverse events post-immunisation are necessary to ensure a proper evaluation of the benefit-risk profile of vaccines. Different methods for signal evaluation such as observed versus expected analysis and signal detection exist to identify safety signal and to assess the relationship between vaccine exposure and the occurrence of an event. These methods rely on accurate background rates of the event under evaluation. In the absence of these background rates, occurrence of rare events or an apparent increase in more common events can be interpreted as a signal of an unsafe vaccine. This stresses the importance of generating background rates of potential adverse events of special interest (AESI) in regions or countries where upcoming COVID-19 vaccines may be used[2].

To support safety signal evaluation, this study generated background rates of AESI; those may be used to contextualize data from prospective monitoring studies and spontaneous reporting databases, and thereby, to help identify potential safety signals.

[1] Le, T. Thanh, et al. "The COVID-19 vaccine development landscape." *Nat Rev Drug Discov* 19.5 (2020): 305-6

[2] Black, Steven, et al. "Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines." *The Lancet* 374.9707 (2009): 2115-2122.

7 Research question and objectives

7.1 Co-Primary objectives

- To estimate the incidence rates of adverse events of special interest (AESI) in the general population by calendar year and data source over the period 2017 to 2020.
- To estimate the incidence of pregnancy outcomes among pregnant women aged between 12 to 55 years old by calendar year and data source over the period 2017 to 2020.
- To estimate the weekly and monthly incidence rates of COVID-19 (overall and by severity level) in 2020 by data source.
- To estimate the monthly incidence rates of multisystem inflammatory syndrome in children (MIS-C) aged between 0 to 19 years old in 2020 by data source.

7.2 Secondary objectives

- To estimate the incidence rates of AESI in the general population by calendar year, sex, age group, and data source over the period 2017 to 2020.
- To estimate the incidence rates of AESI in the general population by month, sex, age group, and data source over the period 2017 to 2020.
- To estimate the incidence rates of multisystem inflammatory syndrome (MIS-C) in children in 2020 by month, sex, age group, and data source.
- To estimate the prevalence of high-risk medical conditions for developing severe COVID-19 by year and data source over the period 2017 to 2020.
- To estimate the incidence rates of AESI in the at-risk population for developing severe COVID-19 by calendar year, sex, age group, and data source over the period 2017 to 2020.

8 Amendments and updates

Date	Amendment	Justification	Protocol Section
September 21,2020	Adding transverse myelitis	Request EMA	Events of special interest
April 14 ,2021	Adding thrombotic thrombocytopenia	Request EMA	Events of special interest
June 3, 2021	Isolating myocarditis from pericarditis	Request EMA	Events of special interest

9 Research methods

9.1 Study design

The study was a retrospective multi-database dynamic cohort study. The study was conducted during the years 2017 to 2020, including the period of SARS-CoV-2 circulation in Europe until the date of last data availability for each data source, where possible. Since Denmark and Germany could not get access to recent data so quickly it used available data from 2010-2013 and from 2014-2017, respectively, to generate background incidence rates.

9.2 Setting

The study results included data from 10 data sources in 7 European countries (Table 4). Data sources are described in section 9.4.

Table 4 Overview of data sources used for the study

Country	Data Access Provider	Name Data source	Coding system	Active population	Type of data source	Provenance for diagnoses
Germany	BIPS	GePaRD	ICD10GM	25 million*	Health insurance	Claims by hospitals and outpatient specialists and GPs**
Netherlands	PHARMO	PHARMO	ICD10CM (HOSP) & ICPC (GP)	6 million	Record linkage	GP medical records, Hospital discharge diagnoses
Denmark	Aarhus University (DCE-AU)	Danish Registries	ICD10CM	5.8 million	Record linkage	Hospital discharge, outpatient specialist diagnoses
Spain	AEMPS	BIFAP	SNOMED, ICD9CM (GP) & ICD10CM (HOSP)	10 million	GP medical records	GP medical records, communication from specialists &hospitalization discharge diagnoses for a subpopulation
Spain-Valencia	FISABIO	FISABIO	ICD9CM ICD10CM	5.2 million	Record linkage	GP medical records, outpatient specialist, Hospitalization discharge and Emergency visits
Spain-Catalunya	IDIAPJGol	SIDIAP	ICD10CM	5.7 million	Record linkage	GP medical records & communication from specialists, hospitalization discharge diagnoses primary and secondary
Italy	SoSeTe	PEDIANET	ICD9CM	0.5 million	Pediatric medical record	Family Pediatricians & communication from specialists
Italy	ARS	ARS	ICD9CM	3.6 million	Record linkage	Emergency room visits Hospitalization discharge
United Kingdom	Utrecht University	CPRD-Gold	RCD2, SNOMED	13 million	GP medical records	GP medical records & communication from specialists
France	BPE	SNDS		67 million	Health insurance	Claims by Hospital, and outpatient specialists, GP

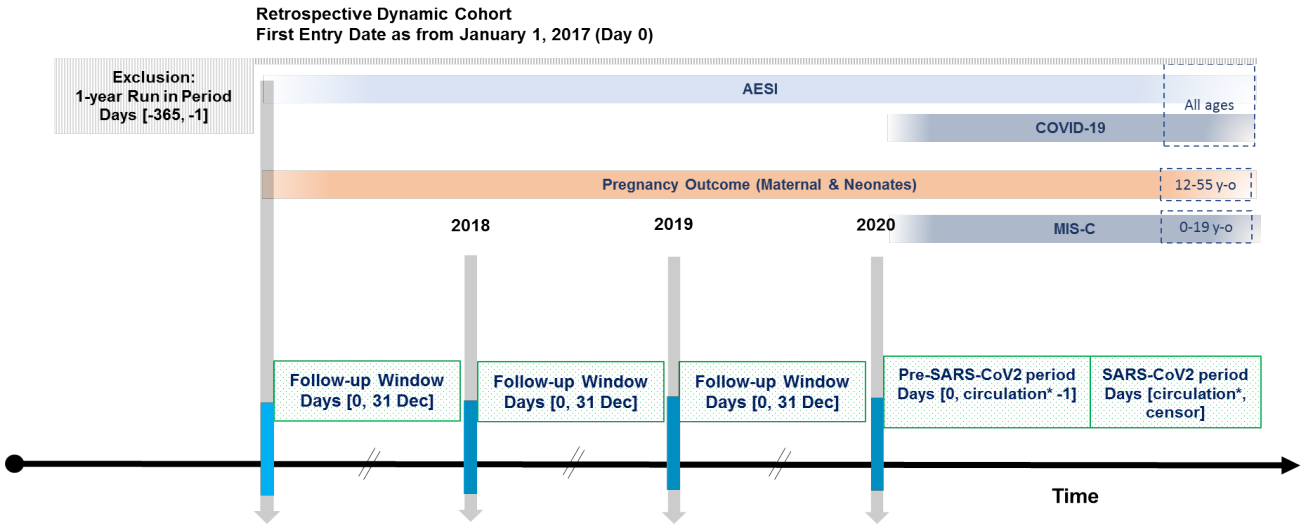
GP: General practitioner

*BIPS: extracted only data from 1 statutory health insurance which represents 800.000 individuals of out 25 million, data extraction process for the largest health insurer is ongoing.
**Only events with hospital diagnoses were included in the study.

9.3 Subjects

The base population included all individuals observed in one of the participating data sources for at least one day during the study period (01 January 2017 - last data availability) and who had at least 1 year of data availability before cohort entry, except for individuals with data available since birth. For Danish and German data the study period differed due to data availability.
Per event, for calculation of incidence, individuals were followed until the earliest of date of the event, death, exiting the data source, or last data draw-down. Because person-time was censored at the occurrence of the event, person-time varies between events.
For calculation of prevalence, individuals were followed until death, exiting the data source, or last data draw-down.

Sub-populations such as pregnant women or children were created according to the outcome under assessment (Figure 1).



y-o: years old
Study period from January 2017 until last data collected (e.g. October 2020)
*Start of SARS-CoV2 circulation is country-specific and based on ECDC data
Censoring occurred at event date, last data collected, last data draw-down, or death, whichever occurred first

Figure 1 Study design

For incidence rates of non-pregnancy AESI, start of follow-up time was defined as the latest of having one year of valid data in the data source, or 01 January 2017 (2010 for Denmark; 2014 for Germany), for those who were not in the data source at birth; or as the latest between birth and 01 January 2017 otherwise. End of follow-up was defined per event as the earliest of date of event, death, last data draw-down, or exiting the data source. Individual person-time varied according to the event under evaluation based on censoring conditions.

For incidence rates of pregnancy outcomes, start of follow-up time was defined at the start date of the pregnancy. For subjects pregnant on 01 January 2017 with one year of valid data prior to 01 January 2017, 01 January 2017 was used as the start of follow-up. For subjects reaching one year of valid data

in the data source during a pregnancy, the date of one year of valid data is used as the start of follow-up. End of follow-up was defined per pregnancy as the date of the event, end date of pregnancy (this may be equal to the date of the event), death, last data draw-down, or exiting the data source. Subjects could contribute more than one pregnancy during the study period.

For prevalence of at-risk medical conditions, start of follow-up time for identification of at-risk conditions was defined as the latest of 01 January 2016, for those who were not in the data source at birth; or as the latest between birth and 01 January 2016 otherwise. Start of follow-up for inclusion in at risk group was defined as the latest of having one year of valid data in the data source, or 01 January 2017, for those who were not in the data source at birth; or as the latest between birth and 01 January 2017 otherwise. End of follow-up was defined as the earliest of date of death, last data draw-down, or exiting the data source.

9.4 Variables

Variables of interest for the calculation of background incidence rates and prevalence rates were those relevant for creation of:

- Person-time: birth and death dates as well as periods of observation based on the person-time calculation.
- Events: dates of medical and/or procedure and/or prescription/dispensing codes to identify AESI and at-risk medical conditions.

9.4.1 Person-time & Follow-up

For incidence rates of non-pregnancy AESI, start of follow-up time was defined as the latest of having one year of valid data in the data source, or 01 January 2017 (except for DCE-AU and GePaRD), for those who are not in the data source at birth; or as the latest between birth and 01 January 2017 otherwise. End of follow-up was defined per event as the earliest of date of event, death, last data availability, or exiting the data source. Individual person-time varied according to the event under evaluation.

For incidence rates of pregnancy outcomes, start of follow-up time was defined at the start date of the pregnancy. For subjects pregnant on 01 January 2017 with one year of valid data prior to 01 January 2017, 01 January 2017 was used as the start of follow-up. For subjects reaching one year of valid data in the data source during a pregnancy, the date of one year of valid data was used as the start of follow-up. End of follow-up was defined per pregnancy as the date of the event, end date of pregnancy (this may be equal to the date of the event), death, last data draw-down, or exiting the data source. Subjects could contribute more than one pregnancy during the study period.

Variables of interest for the calculation of background incidence rates and prevalence rates were those relevant for creation of:

- Person-time: birth and death dates as well as periods of observation.
- Events: dates of medical and/or procedure and/or prescription/dispensing codes to identify AESI and at-risk medical conditions.

9.4.2 AESI, At-risk medical conditions & Operationalization

9.4.2.1 AESI

The list of AESI has been defined based on events that are or are potentially related to marketed vaccines, events related to vaccine platforms or adjuvants, and events that may be associated with COVID-19, because they would fit in the pathogenesis or events occurring in target groups (e.g. pregnancy outcomes). As part of the harmonization of COVID-19 vaccine safety monitoring during clinical development phase, the Coalition for Epidemic Preparedness Innovations (CEPI) has created a preliminary list of AESI for COVID-19 vaccine safety monitoring together with the Brighton Collaboration. This preliminary list did not yet include AESI related to adjuvants nor maternal/neonatal outcomes. Since AS03 adjuvant will be made available for COVID-19 vaccine development and a potential association between vaccine containing AS03 (i.e. Pandemrix, GSK Vaccines, Belgium) and narcolepsy has been identified during the H1N1 pandemic, the list of AESI also included narcolepsy[3]. The list of AESI has been discussed and agreed with the European Medicine Agency (EMA) advisory group monitoring committee on 9th July 2020, after which sudden death, diabetes, transverse myelitis and death were added. In March 2021, further to safety issues with the use of some COVID-19 vaccines, the coagulation disorders have been reclassified and included 6 subtypes of events. Myocarditis became of relevance end of May 2021, and the definition was narrowed.

Table 5. List of AESI

Body system / Classification	**AESI**
Auto-immune diseases	Guillain-Barré Syndrome
	Acute disseminated encephalomyelitis
	Narcolepsy
	Acute aseptic arthritis
	Type 1 Diabetes
	(Idiopathic) Thrombocytopenia
Cardiovascular system	Acute cardiovascular injury including: Microangiopathy, Heart failure, Stress cardiomyopathy, Coronary artery disease, Arrhythmia, Myocarditis/Pericarditis, Myocarditis alone
Circulatory system	Coagulation disorders including: Disseminated intravascular coagulation, Venous thromboembolism (including Pulmonary embolism and Deep vein thrombosis), Thrombotic microangiopathy, Hemorrhagic stroke, Ischemic stroke, Cerebral venous thrombosis, thrombotic thrombocytopenia syndrome (TTS)
	Single Organ Cutaneous Vasculitis
Hepato-gastrointestinal and renal system	Acute liver injury
	Acute kidney injury
Nerves and central nervous system	Generalized convulsion
	Meningoencephalitis

[3] *Miller, E, et al. Risk of narcolepsy in children and young people receiving AS03 adjuvanted pandemic A/H1N1 2009 influenza vaccine: retrospective analysis. BMJ 346 (2013).*

	Transverse myelitis
Respiratory system	Acute respiratory distress syndrome
Skin and mucous membrane, bone and joints system	Erythema multiforme
	Chilblain – like lesions
Other system	Anosmia, ageusia
	Anaphylaxis
	Multisystem inflammatory syndrome in children
	Death (any causes)
	COVID-19 disease (by levels of severity)
	Sudden death
Pregnancy outcome - Maternal	Gestational Diabetes
	Preeclampsia
	Maternal death
Pregnancy outcome - Neonates	Fetal growth restriction
	Spontaneous abortions
	Stillbirth
	Preterm birth
	Major congenital anomalies
	Microcephaly
	Neonatal death
	Termination Of Pregnancy for Fetal Anomaly

- Two additional events: colonic diverticulitis and hypertension, were included as control events. These events serve as indicators to investigate potential changes in health care behaviours during the pandemic and associated lockdown periods. Colonic diverticulitis was chosen as a serious event necessitating urgent healthcare contact while hypertension was chosen as a less serious event for which healthcare contact may be delayed.
- AESI are defined using event definition forms (see annex 1) and identification in the data sources makes use of medical and/or procedure and/or prescription/dispensing codes. Using information contained in event definition forms together with data access provider experience, various algorithms for definition of each AESI may be explored, algorithm development is part of the study.

9.4.2.2 At-Risk Medical Conditions to develop severe COVID-19

At risk medical conditions for developing severe COVID-19 have been defined based on scientific evidence available on Center of Disease Control and National Health Services websites when we wrote the protocol[4].

The selected at-risk medical conditions are considered as at higher risk to develop severe COVID-19 as per protocol (Table 6).

[4] Dodd, CN, Willame C, Sturkenboom M et al. Protocol for Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines.
http://www.encepp.eu/encepp/openAttachment/fullProtocol/37296;jsessionid=8dHqQmMa7kW7URDzEbQkAIR57zM6WItos9bXKY6uP0kZnnBf1hpi!-1960461856

The following variables will be created:

- At-Risk groups: medical codes and associated dates for at-risk medical conditions characterizing at-risk groups for developing severe COVID-19 as well as prescription and/or dispensing records for drug exposures which may be used as proxies for their identification. At-risk groups will be created for each of the at-risk medical conditions listed in Table 6. Multimorbidity will be considered (subjects may belong to more than one at-risk group).

Table 6 Comorbid conditions with evidence of increased COVID-19 severity [5]

At-risk medical conditions	**Medicinal product proxy(ies) (ATC code)**
Cancer (with chemo/immuno/radio-therapy, cancer treatment, immunosuppressant; targeted cancer treatment (such as protein kinase inhibitors or PARP inhibitors); blood or bone marrow cancer (such as leukemia, lymphoma, myeloma))	Alkylating agents (L01A) Antimetabolites (L01B) Plant alkaloids and other natural products (L01C) Cytotoxic antibiotics and related substances (L01D) Other antineoplastic agents (L01X) Hormones and related agents (L02A) Hormone antagonists and related agents (L02B) Immunostimulants (L03) Immunosuppressants (L04)
Type 1 & 2 Diabetes	Blood glucose lowering drugs (A10B/A10A)
Obesity (BMI > 30)	Peripherally acting antiobesity products (A08AB) Centrally acting antiobesity products (A08AA)
Cardiovascular disease/ Serious heart conditions including heart failure, coronary artery disease, cardiomyopathies	Antiarrhythmics, class I and III (C01B) Cardiac stimulants excl. Cardiac glycosides (C01C) Vasodilators used in cardiac diseases (C01D) Other cardiac preparations (C01E) Antithrombotic agents (B01A)
Chronic lung disease including COPD, cystic fibrosis, severe asthma	Drugs for obstructive airway diseases (R03) Lung surfactants (R07AA) Respiratory stimulants (R07AB)
Chronic kidney disease	Erythropoietin (B03XA01)
HIV	Protease inhibitors (J05AE) Combinations to treat HIV (J05AR) NRTI (J05AF) NNRTI (J05AG)
Immunosuppression	Immunosuppressants (L04A) Corticosteroids (H02)
Sickle Cell Disease	Hydroxyurea (L01XX05)

[5] CDC: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html
https://www.nhs.uk/conditions/coronavirus-covid-19/people-at-higher-risk/who-is-at-high-risk-from-coronavirus-clinically-extremely-vulnerable/

	Other hematological agents (B06AX)
Negative Control Conditions	**Medicinal product proxy(ies) (ATC code)**
Colonic Diverticulitis	-
Hypertension	First anti-hypertensive drugs (C02, C03, C07, C08, C09)
Pregnancy	
Start date of pregnancy	
End date of pregnancy	

9.4.2.3 Operationalization

For each of the events of interest living event definition forms have been created comprising the following chapters:

- Event definition: using the Brighton Collaboration definitions if available and otherwise definitions from European learned societies
- Synonyms / lay terms used for the event: these show how an event may be described/called in free text
- Laboratory tests done specific for event (may be used as confirmation)
- Diagnostic tests done specific for event (may be used as confirmation in building algorithms)
- Drugs used to treat event (may be used as confirmation in building algorithms)
- Procedures used specific for event treatment (may be used as confirmation in building algorithms)
- Setting (outpatient specialist, in-hospital, GP, emergency room) where condition will be most frequently diagnosed
- Diagnosis codes or algorithms used in different papers to identify the events in Europe/USA
- Experience of participating data sources to identify or validate the events (to be completed by each data source)
- Proposed codes by Codemapper[6]
- Algorithm proposal for event identification; several algorithms will be built during the execution of the protocol using diagnosis codes, provenance, and confirmatory tests/drugs/procedures[7]
- Published background rates
- Extracted codes (upon characterization)
- Study design related information
    - Estimated lag time from onset to diagnosis
    - Is condition a contraindication to any vaccination?
    - Is this a chronic or potentially recurrent condition?
    - Does this condition cause increased fatality?
    - Time to onset (from vaccination and/or infection)
- References

The event definition form was used throughout the project to transparently track how an event is defined and identified in each of the data sources. It was the basis for the creation of study variables and algorithms and are evolving documents capturing which codes and algorithms were used. All definitions are retrievable from the Zenodo community.

[6] *https://vac4eu.org/codemapper/*

[7] *Gini, Rosa, et al. "Data extraction and management in networks of observational health care databases for scientific research: a comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE strategies." Egems 4.1 (2016).*

9.4.3 Other variables

- Demographic characteristics: dates of birth and death, sex, country and /or region, data source.

In those data sources in which full date of birth is not available, date of birth was derived as follows:

- Date of birth will be defined as the 15th of the birth month and birth year. If the birth month is missing, the birth date will be defined as the 30th June of the birth year.

9.5 Data sources and measurement

9.5.1 Germany: GePaRD

GePaRD is based on claims data from four statutory health insurance providers in Germany and currently includes information on approximately 25 million persons who have been insured with one of the participating providers since 2004 or later. Per data year, there is information on approximately 20% of the general population and all geographical regions of Germany are represented. In addition to demographic data, GePaRD contains information on dispensations of reimbursable prescription drugs as well as outpatient (i.e., from general practitioners and specialists) and inpatient services and diagnoses. GePaRD also contains information on influenza vaccinations and routine childhood immunizations and there is experience with studies on utilization and risk of vaccination and on background incidence of adverse events following vaccinations. GePaRD data have been used for vaccine safety studies. GePaRD is listed under the ENCePP resources database. http://www.encepp.eu/encepp/viewResource.htm?id=26534. For the purpose of this study, only one statutory health insurance was included which represents around 800,000 people.

9.5.2 Netherlands: PHARMO Database Network

The PHARMO Database Network is a population-based network of electronic healthcare databases and combines anonymous data from different primary and secondary healthcare settings in the Netherlands. These different data sources, including data from general practices, in- and out-patient pharmacies, clinical laboratories, hospitals, the cancer register, pathology register and perinatal register, are linked on a patient level through validated algorithms. To ensure the privacy of the data in the PHARMO Database Network, the collection, processing, linkage and anonymization of the data is performed by STIZON. STIZON is an independent, ISO/IEC 27001 certified foundation, which acts as a Trusted Third Party between the data sources and the PHARMO Institute. The longitudinal nature of the PHARMO Database Network system enables to follow-up more than 9 million persons of a well-defined population in the Netherlands for an average of twelve years. Currently, the PHARMO Database Network covers over 6 million active persons out of 17 million inhabitants of the Netherlands. Data collection period, catchment area and overlap between data sources differ. Therefore, the final cohort size for any study will depend on the data sources included. As data sources are linked on an annual basis, the average lag time of the data is one year. All electronic patient records in the PHARMO Database Network include information on age, sex, socioeconomic status and mortality. Other available information depends on the data source. A detailed description of the different data sources is given below. PHARMO is always seeking new opportunities to link with healthcare databases. Furthermore,

it is possible to link additional data collections, such as data from chart reviews, patient-reported outcomes or data from general practice trials.
The General Practitioner database comprises data from electronic patient records registered by GPs. The records include information on diagnoses and symptoms, laboratory test results, referrals to specialists and healthcare product/drug prescriptions. The prescription records include information on type of product, prescription date, strength, dosage regimen, quantity and route of administration. Drug prescriptions are coded according to the WHO ATC Classification System [www.whocc.no]. Diagnoses and symptoms are coded according to the International Classification of Primary Care - ICPC [www.nhg.org], which can be mapped to the International Classification of Diseases - ICD codes, but can also be entered as free text. GP data cover a catchment area representing 3.2 million residents (~20% of the Dutch population).
The Out-patient Pharmacy Database comprises GP or specialist prescribed healthcare products dispensed by the out-patient pharmacy. The dispensing records include information on type of product, date, strength, dosage regimen, quantity, route of administration, prescriber specialty and costs. Drug dispensings are coded according to the WHO ATC Classification System. Out-patient pharmacy data cover a catchment area representing 4.2 million residents (~25% of the Dutch population). PHARMO is listed under the ENCePP resources database. PHARMO data capture influenza vaccine and may be linked to the PRAEVENTIS database that is held by RIVM, based on specific permissions. http://www.encepp.eu/encepp/viewResource.htm?id=22271

9.5.3 Denmark: Danish Registries (DCE-AU)

Denmark has a tax-funded health care system ensuring easy and equal access to health care for all its citizens, and with this system all contacts are recorded in administrative and medical registers. The records carry a unique personal identification number, called the CPR-number, assigned to every Danish citizen. Linkage between registers at an individual level is possible because this CPR-number is used in all Danish registers and assigned by the Danish Civil Registration System. All registers have a nationwide coverage and an almost 100% capture of contacts covering information on currently 5.8 million inhabitants plus historical information. For the purpose of the study we will obtain information from the following registries. The Danish National Prescription Registry (DNPR) includes data on all outpatient dispensing of medications and vaccines at Danish pharmacies from 1995 and onwards, including dispensing date, ATC code, product code and amount. The Danish National Health Service Register includes data on primary care services, including general practitioner contacts, examinations, procedures, and vaccinations; psychologist or psychiatrist and other primary care provider visits; etc. From the Danish Civil Registration System, data on demographics (sex, date of birth) and censoring (migration, vital status). The Danish National Patient Registry contains diagnoses and procedures from all hospitalizations since 1977 and contacts to hospital outpatient clinics since 1995. The Danish National Health Service Register contains information on referral for vaccine administration from GPs. The Danish databases were characterized in the ADVANCE project and considered fit for purpose for vaccine coverage, benefits and risk assessment and could participate in near real-time monitoring. http://www.encepp.eu/encepp/viewResource.htm?id=36221

9.5.4 Spain: BIFAP

BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atencion Primaria), a computerized database of medical records of primary care (www.bifap.aemps.es) is a non-profit

research project funded by the Spanish Agency for Medicines and Medical Devices (AEMPS). The project started in 2001 and the current complete version of the database with information until December 2019 includes clinical information of 10.153 primary care practices (PCPs) and pediatricians. Nine participant autonomous regions send their data to BIFAP every year. BIFAP database currently includes anonymized clinical and prescription/dispensing data from around 14 million (9,4 active population) patients representing 85% of all patients of those regions participating in the database, and 29% of the Spanish population. Mean duration of follow-up in the database is 8.7 years. Information collected by PCPs includes administrative, socio-demographic, lifestyle, and other general data, clinical diagnosis and health problems, results of diagnostic procedures, interventions, and prescriptions/dispensations. Diagnoses are classified according to the International Classification of Primary Care (ICPC)-2, ICD-9 and SNOMEDCT system, and a variable proportion of clinical information is registered in "medical notes" in free text fields in the EMR. Additionally, information on hospital discharge diagnoses coded in ICD-10 terminology is linked to patients included in BIFAP for a subset of periods and regions participating in the database. All information on prescriptions of medicines by the PCP is incorporated and linked by the PCP to a health problem (episode of care), and information on the dispensation of medicines at pharmacies is extracted from the e-prescription system that is widely implemented in Spain.

Information on 2020 and COVID-19 is also available for a number of regions from registries linked to the database. The BIFAP database was characterized in the ADVANCE project and considered fit for purpose for vaccine coverage, benefits and risk assessment. http://www.encepp.eu/encepp/viewResource.htm?id=21501

9.5.5 Spain: SIDIAP

The Information System for Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària' - SIDIAP; www.sidiap.org) was created in 2010 by the Catalan Health Institute (CHI) and the IDIAPJGol Institute. It includes information collected since 01 January 2006 during routine visits at 278 primary care centers pertaining to the CHI in Catalonia (North-East Spain) with 3,414 participating GPs. SIDIAP has pseudo-anonymized records for 5.7 million people (80% of the Catalan population) being highly representative of the Catalan population.

The SIDIAP data comprises the clinical and referral events registered by primary care health professionals (GPs, paediatricians and nurses) and administrative staff in electronic medical records, comprehensive demographic information, community pharmacy invoicing data, specialist referrals and primary care laboratory test results. It can also be linked to other data sources, such as the hospital discharge database, on a project-by-project basis. Health professionals gather this information using ICD-10 codes, ATC codes and structured forms designed for the collection of variables relevant for primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, blood and urine test results. In relation to vaccines, SIDIAP includes all routine childhood and adult immunizations, including the antigen and the number of administered doses. Encoding personal and clinic identifiers ensures the confidentiality of the information in the SIDIAP database. The SIDIAP database is updated annually at each start of the year.

Nowadays, with the COVID-19 pandemic, there is the possibility to have shorter term updates in order to monitor the evolution of the pandemic. Recent reports have shown the SIDIAP data to be useful for epidemiological research. SIDIAP is listed under the ENCePP resources database www.encepp.eu/encepp/resourcesDatabase.jsp). The SIDIAP database was characterized in the

ADVANCE project and considered fit for purpose for vaccine coverage, benefits and risk assessment. http://www.encepp.eu/encepp/viewResource.htm?id=4646

9.5.6 Spain: FISABIO

The region of Valencia, with 5 million inhabitants, is part of the Spanish National Health System, a universal public healthcare system. Information will be obtained from the population-based electronic information systems of the Valencia health system integrated database: (1) The Population Information System (SIP) provides an identification number for each person under Valencian Health Service (VHS) coverage, and registers some demographic characteristics, and dates and causes of VHA discharge, including death. (2) The minimum basic dataset at hospital discharge is a synopsis of clinical and administrative information on all hospital discharges, including diagnoses and procedures (all electronic health systems in the VHS use the ICD- 9-CM). (3) The Emergency Department module (ED) including ED dates of visit and discharge and reason for discharge. (4) The electronic medical record (EMR) for ambulatory care, available in all primary healthcare centers and other ambulatory settings. It has all the information on patients regarding diagnoses, their personal and family medical history, laboratory results, lifestyle, etc. (5) The pharmaceutical module (prescription information system), part of EMR, includes information about both physician prescriptions and dispensations from pharmacy claims. (6) The Corporate Resource Catalogue provides information about the geographical and functional organization of VHS, its health centers, health services provided and professionals in healthcare. Specific public health registries are available and linkable at an individual level (such as the perinatal register and the congenital anomalies register, from which pregnancy outcomes can be obtained) All the information in these systems can be linked at an individual level through the SIP number. The FISABIO database was used for research into Narcolepsy in the SOMNIA study[8].[9]

9.5.7 Italy: PEDIANET database

PEDIANET, a pediatric general practice research database, contains reason for accessing healthcare, health status (according to the Guidelines of Health Supervision of the American Academy of Paediatrics), demographic data, diagnosis and clinical details (free text or coded using the ICD-9 CM), prescriptions (pharmaceutical prescriptions identified by the ATC code), specialist appointments, diagnostic procedures, hospital admissions, growth parameters and outcome data of the children habitually seen by about 140 family paediatricians distributed throughout Italy.
PEDIANET can link to other databases using unique patient identifiers. In the first database, information on routine childhood vaccination are captured including vaccine brand and dose. In the second database, information on patient hospitalization date, reason for hospitalization, days of hospitalizations and discharge diagnosis (up to six diagnosis) are captured. The family paediatrician's participation in the database is voluntary and patients and their parents provide consent for use of their data for research purposes. In Italy each child is assigned to a family paediatrician, who is the referral for any health visit or any drug prescription, thus the database contains a very detailed personal medical history.
The data, generated during routine practice care using common software (JuniorBit®), are anonymized and sent monthly to a centralized database in Padua for validation. The PEDIANET database can be

[8] *Dodd, Caitlin, et al. Incidence rates of narcolepsy diagnoses in Taiwan, Canada, and Europe: The use of statistical simulation to evaluate methods for the rapid assessment of potential safety issues on a population level in the SOMNIA study. Plos One (2018).*
[9] *Weibel, Daniel, et al. Narcolepsy and adjuvanted pandemic influenza A (H1N1) 2009 vaccines – Multi-country assessment. Vaccine 36(41): 6202–6211 (2019).*

linked to regional vaccination data which was successfully tested in the ADVANCE project where it was characterized and deemed fit for purpose for paediatric routine vaccines. In Italy, a national register for COVID-19 cases has been implemented and a linkage with the PEDIANET database is available.

9.5.8 Italy: ARS database

The Italian National Healthcare System is organized at regional level: the national government sets standards of assistance and a tax-based funding for each region, and regional governments are responsible to provide to all their inhabitants. Tuscany is an Italian region, with around 3.6 million inhabitants. The Agenzia Regionale di Sanita' della Toscana (ARS) is a research institute of the Tuscany Region. The ARS database comprises all information that are collected by the Tuscany Region to account for the healthcare delivered to its inhabitants. Moreover, ARS collects data from regional initiatives. All the data in the ARS data source can be linked with each other at the individual level, through a pseudo-anonymous identifier. The ARS database routinely collects primary care and secondary care prescriptions of drugs for outpatient use, and is able to link them at the individual level with hospital admissions, admissions to emergency care, records of exemptions from copayment, diagnostic tests and procedures, causes of death, mental health services register, birth register, spontaneous abortion register, induced terminations register. A pathology register is available, mostly recorded in free text, but with morphology and topographic Snomed codes. Mother-child linkage is possible through the birth register. Vaccine data is available since 2016 for children and since 2019 for adults. However, to date, 2019 vaccination data for adults may still be incomplete. In Italy, a national register for COVID-19 cases has been implemented and a linkage with the ARS database is available. The ARS database was characterized in the ADVANCE project and considered fit for purpose for vaccine coverage, benefits and risk assessment when using the new vaccine register (from 2019)[10].

9.5.9 United Kingdom: CPRD & HES

The Clinical Practice Research Datalink (CPRD) from the UK collates the computerized medical records of general practitioners (GPs) in the UK who act as the gatekeepers of healthcare and maintain patients' life-long electronic health records. As such they are responsible for primary healthcare and specialist referrals, and they also store information stemming from specialist referrals, and hospitalizations. GPs act as the first point of contact for any non-emergency health-related issues, which may then be managed within primary care and/or referred to secondary care as necessary. Secondary care teams also feedback information to GPs about their patients, including key diagnoses. The data recorded in the CPRD include demographic information, prescription details, clinical events, preventive care, specialist referrals, hospital admissions, and major outcomes, including death. The majority of the data are coded in Read Codes. Validation of data with original records (specialist letters) is also available.
The dataset is generalizable to the UK population based upon age, sex, socioeconomic class and national geographic coverage when GOLD & Aurum versions are used.
There are currently approximately 42 million patients (acceptable for research purposes) – of which 13 million are active (still alive and registered with the GP practice) – in approximately 1,700 practices (https://cprd.com/Data). Data include demographics, all GP/healthcare professional consultations (phone, letter, email, in surgery, at home), diagnoses and symptoms, laboratory test results, treatments, including all prescriptions, all data referrals to other care, hospital discharge summary (date and Read codes), hospital clinic summary, preventive treatment and immunizations, death (date and cause). For a

[10] *http://www.encepp.eu/encepp/viewResource.htm?id=24417*

proportion of the CPRD panel practices (>80%), the GPs have agreed to permit CPRD to link at patient level to the Hospital Episode Statistics (HES) data. CPRD is listed under the ENCePP resources database, access will be provided by the Utrecht University. The CPRD was not yet characterized in the ADVANCE project, where the UK THIN (The Health Improvement Network) and RCGP RSC (Royal College of General Practitioners Research and Surveillance Centre) databases were used, but has been largely used in vaccine studies.

The HES database contains details of all admissions to National Health System (NHS) hospitals in England; approximately 60% of GP practices in the CPRD are linked to the HES database. Not all patients in the CPRD have linked data (e.g. if they live outside England or if their GP has not agreed that their data should be used in this way). As with standard CPRD patients, HES data are limited to research-standard patients. CPRD records are linked to the HES using a combination of the patient's NHS number, gender and date of birth.

For the purpose of this study, only CPRD GOLD was used.

9.5.10 France: Système National des Données de Santé (SNDS)

The SNDS (Système National des Données de Santé)[11] is the French nationwide healthcare database. It currently covers the overall French population (about 67 million persons) from birth (or immigration) to death (or emigration), even if a subject changes occupation or retires. Using a unique pseudonymized identifier, the SNDS merges all reimbursed outpatient claims from all French health care insurance schemes (SNIIRAM database), hospital-discharge summaries from French public and private hospitals (PMSI database), and the national death register. SNDS data are available since 2006 and contains information on:

- General characteristics: gender, year of birth, area of residence, etc.
- Death: month, year and cause
- Long-Term Disease registration associated with an ICD-10 diagnostic codes
- Outpatient reimbursed healthcare expenditures with dates and codes (but not the medical indication nor result): visits, medical procedures, nursing acts, physiotherapy, lab tests, dispensed drugs and medical devices, etc. For each expenditure, associated costs, prescriber and caregiver information (specialty, private/public practice) and the corresponding dates are provided.
- Inpatients details: primary, related and associated ICD-10 diagnostic codes resulting from hospital discharge summaries with the date and duration of the hospital stay, the performed medical procedures, and the related costs. Drugs included in the diagnosis related group cost are not captured.

Outpatient data (SNIIRAM) are uploaded to the SNDS throughout the year. It is admitted that a lag of around 6 months is required to catch 90% of the dispensings. Inpatient data (PMSI) are uploaded in one time, at the end of the following year. Hence, we consider that complete SNDS data of year Y are available in January of the year Y+2. SNDS access is regulated.

Each study and data extraction need approval from the CESREES (Comité Ethique et Scientifique pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé) in charge of assessing scientific quality of the project, and authorization from the CNIL (French data protection commission), and then contracts with the SNDS data holder (CNAM) for data extraction. Bordeaux PharmacoEpi

[11] *https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS*

(BPE), a research platform of the University of Bordeaux specialized in real world studies, will be in charge of requesting access to SNDS data. The SNIIRAM data were not yet characterized in the ADVANCE project but have been used for vaccine studies. http://www.encepp.eu/encepp/viewResource.htm?id=38744

9.5 Bias

This final study report includes data from 9 data sources in 7 countries to compute incidence rates of AESI, based on available and permissions in October-November 2020. These data sources were chosen based on availability, ability to run multisite studies and experience in using common data models plus ability to join the consortium quickly in May 2020 during a very short tender period. These data sources contain various type of data which are either representative of the national population (eg. CPRD, Danish registry), or have a regional/multiregional scope (eg. BIFAP, SIDIAP, PEDIANET). Some data are collected at hospital level including or not emergency department or at GPs level only, others are collected at both hospital and GPs level. Given the heterogeneity in the type of encounters recorded, our analyses are computed per data source and no pooled estimates are generated.
Some of the participating data sources in this protocol have long lag times, which means that they cannot contribute to all calendar years for the estimation of the background rates in the first analysis. Six data sources may contribute 2020 data: three will contribute hospitalization data only (ARS and SIDIAP) in adults and children while PEDIANET contributed hospitalization and GP data in paediatrics only, and CPRD and BIFAP contributed GP data, FISABIO contributed GP and hospitalization data.
Some of the data sources do not encompass a birth register, many do not encompass information on induced terminations and/or spontaneous abortions. Quality of information on the pregnancy start and end dates and pregnancy outcome is conditional on this availability and delivery of this data is ongoing. Most of the data sources were characterized in the ADVANCE project and considered fit for purpose for vaccines benefits and risk assessment[12].
A broad set of AESI that are known for being related to vaccination or associated with COVID-19 have been included in this study. Some of them have a well-established clinical definition but for events such as MIS-C, ARDS, Coagulation disorders the Brighton Collaboration definition was under development by the CEPI funded SPEAC project at the time of this protocol development. Case definition for MIS-C and ARDS were made available during the course of the study and were used appropriately and were not available at the time of data extraction and analysis.
For each of the events we used broad (Sensitive: including narrow and possible codes) and narrow (specific_ definitions where possible, to assess and quantify the range of potential misclassification. Due to limited resources, further case ascertainment could not be conducted to confirm disease diagnoses as part of this study, therefore misclassification of outcomes cannot be excluded. Recorded disease diagnosis will be used as date to classify a case as incident. For long latency diseases (e.g., autoimmune diseases), the disease onset may have started months prior to the recorded diagnosis, however this cannot be estimated without review of records, which is not resourced in this study.

Enhanced COVID-19 diseases following vaccination is a theoretical concern at the moment, and not yet shown in any of the studies. Since this event is conditioned on vaccination we cannot assess background rates during the pre-licensure vaccination period. To have some standard to measure against, we

[12] *Sturkenboom M, Braeye T, van der Aa L, Danieli G, Dodd C, Duarte-Salles T, Emborg HD, Gheorghe M, Kahlert J, Gini R, Huerta-Alvarez C, Martín-Merino E, McGee C, de Lusignan S, Picelli G, Roberto G, Tramontan L, Villa M, Weibel D, Titievsky L. ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations. Vaccine. 2020 Dec 22;38 Suppl 2:B8-B21. doi: 10.1016/j.vaccine.2020.01.100. Epub 2020 Feb 12. PMID: 32061385.*

assessed COVID-19 according to five levels that are defined as: Level 1/ any recorded COVID-19 diagnosis and no hospitalisation; Level 2/ hospitalisation for COVID-19 disease with moderate symptoms; Level 3/ hospitalisation for COVID-19 disease with severe symptoms but without mechanical respiratory support; 4/ hospitalisation for COVID-19 disease with severe symptoms and with mechanical respiratory support; Level 5/ death due to COVID-19. The analyses described here are not intended to ascertain the incidence of COVID-19 which is not feasible as not all subjects are tested or diagnosed, but to assess time trends where possible and at least estimate the incidence of severe COVID-19 (Levels 2-4) in preparation for monitoring of enhanced disease following vaccination.

9.6 Study size

The study population included all individuals registered with at least one year of data prior to the start of the study period or follow-up from birth. Overall, the study population aimed to comprise approximately 141.6 million individuals (see Table 4), although Germany decided to run analyses first on a smaller population. The full analysis will allow detailing of actual size.

9.7 Data transformation

This study was conducted in a distributed manner using a common protocol, common data model (CDM), and common analytics programs. This process was used successfully in several other European multi-database projects. The data pipeline has been further improved in the IMI-ConcePTION project (https://www.imi-conception.eu/). This process maximized the involvement of the data providers in the study by utilizing their knowledge on the characteristics and the process underlying the data collection which made analysis more efficient.

1. First, to harmonize the structure of the data sets held by each partner, a shared syntactic foundation was utilized. Syntactic foundation is described in Annex 1 and refers to the syntactically harmonized CDM. In this common data model, data were represented in a common structure but the content of the data remained in their original format.
2. Second, to reconcile differences across terminologies a shared semantic foundation was built for the definition of events under study by collecting relevant concepts in a structured fashion using a standardized event definition template (see annex 1). The Codemapper tool was used to create diagnosis code lists based upon completed event definition templates for each AESI and comorbid risk condition (Becker et al., 2017). Based on the relevant diagnostic medical codes and keywords one or more algorithms were constructed (typically one sensitive, or broad, algorithm and one specific, or narrow algorithm) to operationalize the identification and measurement of each event by medically trained persons. These algorithms can differ per database. No validation was planned done for this study, as there were no resources for this within the budget of the EMA tender. Wherever possible the event definition sheet specified prior validation of algorithms and codes. Scripts for semantic harmonization were coded in R, distributed to data access providers for local deployment, and shared on the catalogue. The impact of choices of different algorithms were assessed quantitatively. This resulted in a set of study variables which were both semantically and syntactically harmonized.
3. Third, following conversion to harmonized study variable sets, R programs for calculation of incidence and prevalence were distributed to data access providers for local deployment. The aggregated results produced by these scripts were then uploaded to the Digital Research Environment (DRE) for pooled analysis and visualization (see Figure 2). The DRE was made available through UMCU/VAC4EU (https://www.andrea-consortium.org/). The DRE is a cloud

based, globally available research environment where data is stored and organized securely and where researchers can collaborate (https://www.andrea-consortium.org/azure-dre/).

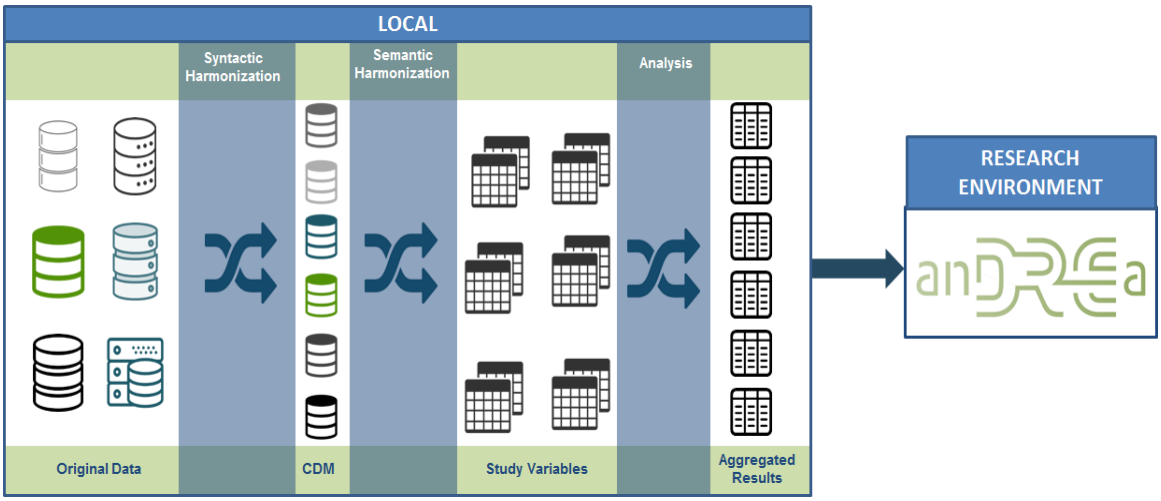


Figure 2 Data management plan

9.7.1 Data extraction

Each database access provider (DAP) created ETL specifications using the standard ConcePTION ETL design template. Following completion of this template and review with study statisticians, each DAP extracted the relevant study data locally using their software (eg Stata, SAS, R, Oracle). This data was loaded into the CDM structure in csv format. These data remained local (Figure 2). Data that were loaded to the CDM were verified using quality checks (level 1 and 2). Specifics of quality checks can be found at the IMI-ConcePTION website.

9.7.2 Data engineering and analysis

Centrally written R scripts were sent to the DAPs and this script transformed the data in the syntactically harmonized CDM to semantically harmonized study variables (see Figure 2). The R scripts were structured in modular form with validated functions. Functions were either standard R packages or packages designed, developed and tested on purpose for multi-database studies. The DAPs ran the R code locally and sent aggregated analysis results to the anDREa digital research environment using a secure file transfer protocol. In the anDREa platform, results were aggregated using SAS and rates were calculated and plotted using R. DAPs that were not able to share low cell counts ran SAS code locally and submitted the incidence rates. All submitted data was inspected (for quality assessment) and pooled (if needed) for final reporting. All steps were detailed in the statistical plan.

9.7.3 Software and Hardware

All final statistical computations were performed on the DRE using SAS. Data access providers had access to the workspace for verification of the scripts.

9.7.4 Storage

Aggregated results, ETL specifications, and a repository of study scripts were stored in the DRE.

9.7.5 Access

Within the DRE, each project-specific area consisted of a separate, secure folder, called a 'workspace'. Each workspace was completely secure, so researchers were in full control of their data. Each workspace had its own list of users, which was managed by its administrators. Access to this workspace was only possible with double authentication using an ID and password together with the user's mobile phone for authentication.

Upload of files was possible for all researchers with access to the workspace within the DRE. Download of files was only possible after requesting and receiving permission from a workspace member with an 'owner' role.

9.7.6 Archiving and record retention

The final study aggregated results sets and statistical programs will be archived and stored on the VAC4EU Sharepoint. The final study protocol and possible amendments, the final statistical report, statistical programs and output files will be archived on the VAC4EU Sharepoint.

Documents that individually and collectively permit evaluation of the conduct of a study and the quality of the data produced will be retained for a period of 5 years in accordance with GPP guidelines. These documents could be retained for a longer period, however, if required by the applicable regulatory requirements or by an agreement between study partners. It is the responsibility of the principal investigator to inform the other investigators/institutions as to when these documents no longer need to be retained. Study records or documents may also include the analyses files, syntaxes (usually stored at the site of the database), ETL specifications, and output of data quality checks.

9.8 Statistical methods

All analyses have been detailed in a Statistical Analysis Plan which was delivered earlier.

9.8.1 Analysis of Demographics and Baseline Characteristics

Demographic characteristics (age at study entry and sex) and baseline characteristics such as at-risk medical conditions and pregnancy were summarized for each data source using descriptive statistics.
Frequency tables including numbers and percentages were generated for categorical variables (age at study entry in categories, sex and at-risk medical conditions).
Mean, standard error, median and range were provided for continuous variables (age at study entry).

9.8.2 Hypotheses

Not applicable. This study is not hypothesis testing.

9.8.3 Statistical Methods

Incidence rates for each AESIs by calendar year were calculated by dividing the number of incident cases (not in run-in year) (numerator) by the total person-time at risk (denominator). A 95%CI was computed using an exact method (Ulm, 1990).
Prevalence rates for each pregnancy outcomes by calendar year were calculated by dividing the number of events (numerator) by the total number of pregnant women (denominator).
Prevalence rates for at-risk conditions were calculated by dividing the number of existing cases in a year (numerator) by the average of the total number of persons recorded monthly (denominator).
Incidence rates were reported by calendar including the year 2020 which corresponds to the SARS-CoV-2 circulation period to investigate potential changes in health care behaviors during the pandemic and associated lockdown periods on the incidence rates, as well as in at-risk populations.

9.8.4 Statistical Analysis

9.8.4.1 Analysis of co-primary objectives

- Incidence rate (and 95% CI) of AESI were calculated for all individuals by calendar years and data sources: the numerator was the number of incident cases (not in the run-in year) in each calendar year (2017, 2018, 2019, 2020) and each data source. The denominator was the total person-years at risk, i.e. from 1st January or birth until date of event, death, last data draw-down, or leaving the database, whichever occurs first, in each calendar year and each data source.

- Prevalence rate of pregnancy outcomes are calculated in women aged 12 to 55 years by calendar year and data sources: the numerator is the number of pregnancy outcomes among women aged 12 to 55 years in each calendar year (2017, 2018, 2019) and each data source. The denominator is the total number of pregnant women in each calendar year and each data source.

- Incidence rate (and 95% CI) of recorded COVID-19 disease (overall and by severity level) was calculated by calendar months and calendar weeks for the year 2020 and data sources: the numerator was the number of incident COVID-19 cases and the denominator were the total person-months or person-weeks at risk, i.e. from 1st January 2020 or birth until date of event, death, last data draw-down or leaving the database whichever occurs first, in each calendar month or week and each data source.

- Incidence rates (and 95% CI) of MIS-C are calculated in children aged 0 to 19 years by calendar month for the year 2020 and data sources: the numerator is the number of incident cases among children aged 0 to 19 years in each data source. The denominator is the total person-years at risk in those up to 19 years old, i.e. from 1st January 2020 or birth until date of event, death, last data draw-down, leaving the database, end of the month or 19th birthday, whichever occurs first, in each calendar month and each data source.

9.8.4.2 Analysis of secondary objectives

- Incidence rates (and 95% CI) of AESI using further stratifications were estimated using the same approach as described for AESI. Rates were stratified by calendar year, sex, age group (Year of age in subjects <20, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80 and older) and data source.

- Monthly incidence rates (and 95% CI) of AESI are estimated for all individuals by month, sex, age group and data source: the numerator was the number of incident cases (not in the run-in

year) by months from 01 January 2017 until last data available (e.g. October 2020), sex, age group (Year of age in subjects <20, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80 and older) and each data source. The denominator was the total person-months at risk, i.e. from 1st January or birth until date of event, death, last data draw-down, leaving the database or end of the month whichever occurs first, in each month and each data source. Monthly incidence rates of AESI were presented graphically to help interpretation of potential seasonality patterns among selected AESIs.

- Monthly incidence rates (and 95% CI) of MIS-C using further stratifications were estimated using the same approach as above. Rates were stratified by calendar month, sex, year of age and data source.

- Prevalence rates (and 95% CI) of at-risk medical conditions for developing severe COVID-19 and prevalence of the use of immunosuppressants were calculated by dividing the number of individuals identified with an at-risk medical condition by the average of the total number of individuals recorded in a month. Prevalence rates were estimated for each calendar year (2017, 2018, 2019, 2020), by sex, age groups (Year of age in subjects <20, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80 and older) and data source. Subjects identified as having an at-risk condition in the run-in period were considered prevalent cases and at-risk at study start (01 January 2017).

- Incidence rates (and 95% CI) of AESI in each at-risk population were estimated by calendar year, sex, age group (Year of age in subjects <20, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80 and older) and data source using the same approach as described above.

Table 7 Incidence rates and prevalence rates calculations for the main analyses

Analysis	Numerator	Denominator	Stratification factors
Incidence rate of AESI	# of new cases of any AESI	Total person-years (person-months) at risk of all subjects	Data sources Calendar time (in years and months) Sex Age group #1
Incidence rate of pregnancy outcomes	# of new events of any pregnancy outcomes	Total pregnancies in women aged 12 to 55 years	Data sources Calendar time (in years) Age group #2
Incidence rate of recorded COVID-19	# of new cases of recorded COVID-19 split by severity	Total person-months or person-weeks at risk of all subjects	Data sources Calendar time (in months and weeks in 2020) Sex Age group #1 Disease severity
Incidence rate of MIS-C	# of new cases of MIS-C	Total person-months at risk of subjects aged 0 to 19 years	Data sources Calendar time (in month in 2020) Sex Age group #3
Proportion of subjects with each at-risk medical condition	# of existing individuals with at-risk medical conditions	Average of total # of individuals registered monthly	Data sources Calendar time (in years) Sex Age group #1
Incidence rate of AESI in each at-risk population	# of new cases of any AESI	Total person-years of existing individuals with at-risk medical conditions	Data sources Calendar time (in years) Sex Age group #1

AESI: Adverse Event of Special Interest; MIS-C: multisystem inflammatory syndrome
Age group #1: Year of age for subjects <20, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80 and older
Age group #2: 12-19, 20-29, 30-39, 40-55
Age group #3: 0-4, 5-9, 10-14, 15-19

9.8.5 Missing data

Since the underlying data represent attended medical care, we cannot assume that absence of information of clinical events means absence of that condition. For this reason, broad algorithms have been included in the report. No imputation was planned or done for missing data.

10 Results

This study was performed as part of a feasibility analysis of a European infrastructure for COVID-19 vaccine monitoring. The study was conducted using a distributed data network across Europe. Several risks and challenges have been encountered during the study that have affected the completion of the full study within pre-planned timelines. The challenges were related to:

- Governance approval process and data access: For most of the databases, governance approvals from scientific ethics committees were obtained within few weeks after submission of the protocol. However, the access to the SNDS database (France) requires a 3-steps process approval which could not allow to access the data in a timely manner for this project. At the time of submission of this final report, the process to access the French data is still ongoing, all scientific and ethic approvals have been received but data transfer is still ongoing.

- Data management workflow: this study used a data management workflow developed in the IMI-Conception project. Some steps needed adjustments specific for this study, which could only be implemented in January 2021. In addition, a close follow-up was provided to the DAPs for the ETL development and the running of the R script.

- Pregnancy outcomes: the ACCESS project is running in parallel to the IMI-Conception project in which identification of start and end of pregnancy and identification of pregnancy outcomes will be validated through complex medical algorithms. The development of the algorithms is still under development for most of the data access providers (eg. GePaRD, BIFAP, SIDIAP), therefore prevalence rates for pregnancy outcomes could only be generated for data sources with an existing pregnancy registry (ie. ARS and CPRD). DCE-AU, FISABIO and PHARMO did not commit to provide pregnancy outcomes in ACCESS.

In this report, data from ARS, PEDIANET, FISABIO, BIFAP, SIDIAP, CPRD, PHARMO DCE-AU and GePaRD (only for AESIs requiring hospitalization) are included. Given the delay in governance process, we anticipate data from SNDS (BPE) to be available in Q3 2021. The French data are therefore not included in this report but will be made available on the VAC4EU dashboard once available (https://vac4eu.org/covid-19-tool/). Updated data from GePaRD including data from the largest health insurer in Germany will be available in Q3 2021 and will also made available on the VAC4EU dashboard once available (https://vac4eu.org/covid-19-tool/).

10.1 Updates from April report:

- In the previous version of the report, we highlighted that incidence rates from the GePaRD database were of higher magnitude compared to the other data sources included in the study. Further investigations have been conducted to understand possible issue related to meaning of events. The outcomes of the investigation concluded that identification of comorbidities requires more sophisticated algorithms with a mixed use of medical diagnosis and prescription, for this reason, GePaRD will not contribute to at-risk population assessment. In addition, only rates for AESIs requiring hospitalisation have been considered for inclusion in the final report.
- As per EMA's request, medical codes for myocarditis/pericarditis have been revised. In addition, rates for the composite endpoint myocarditis/pericarditis and myocarditis alone have been generated.
- A refinement of the concepts and classification as broad/narrow has been conducted for all events and descending SNOMED codes were added. This revision has little impact on the previous data.
- Updated data including 2020, when possible and a revision of the codelist resulting in a new data extraction process for all DAPs.
- Subpopulations based on only primary care or primary care & hospital are described for PHARMO, BIFAP and SIDIAP.

The incidence rates of AESIs in the at-risk population are available for all data sources included in this report, except for the GePaRD data source which could not contribute to the identification of comorbidities.

Changes according to plan are mentioned here below:

- The prevalence rates for pregnancy outcomes were generated for two data sources (ARS and CPRD) for which a pregnancy registry was available.
- To ensure the timely inclusion of Danish register data in this study, the Danish team prioritized the use of a set of data for which ethics approval was previously approved. Therefore, data from 2010 to 2013 were included for DCE-AU database.
- To ensure the timely inclusion of GePaRD data in this study, the BIPS team used of a set of data from the smallest SHI which are only available from 2004 to 2017. Therefore, data from 2014 to 2017 were included for the GePaRD database.

On request of EMA, detailed data is made available in excel sheets annexes 2-8.

10.2 Descriptive data

Subjects were included in the study according to pre-defined inclusion and exclusion criteria. Reasons for exclusion included invalid birth date, death before study entry (01 January 2017; 01 January 2010 for DCE-AU), observation periods not overlapping with the study period (01 January 2017/2010 (for DCE-AU) – date of last data availability), and unavailability of one year of look-back time prior to study entry. For all databases, a small proportion of subjects met the exclusion criteria. The study flow chart is presented in Table 5. Data on a total of 45 million subjects were included in this study report.

Subjects were described in terms of age at study entry, person time contributed during the study period, and presence of chronic at-risk conditions at study entry. Descriptive data for each data source are presented in Table 6. A total number of 45,986,634 subjects were included in the study, contributing to 148.3 million person-years. The largest contribution in person-time was from BIFAP (22.3%) followed by PHARMO (20.0%), SIDIAP (13.5%), DCE-AU (13.0%), FISABIO (12.8%), CPRD (10.2%), ARS (6.7%) and PEDIANET (0.4%). Figure 3 presents the person-time contribution by database over the study period.

Three databases, BIFAP, SIDIAP and PHARMO, provided subpopulation data. The subpopulation is a subset of the full population and corresponds to subjects recorded both at primary care level and/or hospital level. The subpopulation accounted for 43.1%, 28.3% and 5.4% of the full included population for BIFAP, SIDIAP and PHARMO, respectively (Table 8).

Table 8 Study flowchart

	ARS		PEDIANET		GEPARD		DCE-AU		CPRD		FISABIO		BIFAP		BIFAP_PC_HOSP		SIDIAP_PC		SIDIAP_PC_HOSP		PHARMO_HOSP		PHARMO_PCHOSP	
Subjects disposition	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Subjects present during 2016-2020	3903982	100	199174	100	764269	100	6143446	100	5309729	100	6646465	100	14007619	100	14007619	100	6550377	100	6550377	100	9356863	100	9356863	100
Subjects with sex or birth date missing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	<0.01	3	<0.01
Subjects without valid birth date	4245	0.11	15	0.01	-	-	-	-	1394	0.03	11936	0.18	46	<0.01	46	<0.01	-	-	-	-	2063	0.02	2063	0.02
Subjects with death date before study entry	2082	0.05	-	-	12	<0.01	-	-	-	-	2018	0.03	-	-	-	-	-	-	-	-	-	-	-	-
Subjects without overlap in observed time	127559	3.27	17869	8.97	214018	28	188086	3.06	537702	10.13	744596	11.2	3741105	26.71	9583730	68.42	263079	4.02	4769969	72.82	163702	1.75	8858600	94.67
Subjects without sufficient look-back period	702494	17.99	-	-	-	-	-	-	81923	1.54	1355	0.02	-	-	-	-	81725	1.25	22169	0.34	6263	0.07	-	-
Total number of subjects included in the study	3067602	78.58	181290	91.02	550239	72	5955360	96.94	4688710	88.3	5886560	88.57	10266468	73.29	4423843	31.58	6205573	94.74	1758239	26.84	9184832	98.16	496197	5.3

Table 9 Demographics of study population

		ARS		PEDIANET		GePaRD		DCE-AU		CPRD		FISABIO		BIFAP_PC		BIFAP_PC_HOSP		SIDIAP_PC		SIDIAP_PC_HOSP		PHARMO_PC		PHARMO_PC_HOSP	
Characteristics	**Parameters**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**	**Value**	**%**
Total number of subjects		3067602	.	181290	.	550239	.	5955360	.	4688710	.	5886560	.	10266468	.	4423843	.	6205573	.	1758239	.	9184832	.	496197	.
Total Person-Years across follow-up period		9065271	.	356184	.	1542723	.	22490217	-	12762349	.	20911202	.	34525598	.	8494349	.	19845706	.	5355124	.	26568111	.	1164810	.
Age (in years)	Min	0	.	0	.	0	.	-	-	0	.	0	.	0	.	0	.	0	.	0	.	0	.	0	.
	P25	27	.	1	.	22	.	-	-	20	.	22	.	24	.	27	.	23	.	23	.	25	.	23	.
	P50	47	.	5	.	38	.	-	-	38	.	42	.	43	.	45	.	41	.	41	.	44	.	43	.
	Mean	46	.	5	.	37	.	-	-	39	.	41	.	43	.	45	.	41	.	41	.	44	.	42	.
	P75	65	.	9	.	53	.	-	-	57	.	59	.	61	.	63	.	57	.	58	.	61	.	60	.
	Max	117	.	16	.	106	.	-	-	113	.	117	.	113	.	112	.	116	.	111	.	118	.	117	.
Age category at start of follow-up (in years)	0-19	560532	18.27	181290	100	123229	22.4	1631807	27.4	1157919	24.7	1311335	22.28	2127401	20.72	823019	18.6	1353935	21.82	3861	4.55	1673052	18.22	106551	21.47
	20-29	265845	8.67	0	0	77879	14.15	698402	11.73	612393	13.06	601889	10.22	1032307	10.06	416957	9.43	678016	10.93	5221	6.16	1127744	12.28	60266	12.15
	30-39	335846	10.95	0	0	85204	15.48	775766	13.03	658828	14.05	827324	14.05	1409363	13.73	578177	13.07	956644	15.42	10077	11.89	1208378	13.16	59004	11.89
	40-49	472615	15.41	0	0	95781	17.41	834099	14.01	631115	13.46	916646	15.57	1572316	15.32	675357	15.27	1018150	16.41	7461	8.8	1358055	14.79	68768	13.86
	50-59	457716	14.92	0	0	83286	15.14	726340	12.2	616935	13.16	774435	13.16	1416951	13.8	633111	14.31	795481	12.82	10587	12.49	1312117	14.29	73728	14.86
	60-69	376819	12.28	0	0	46446	8.44	679841	11.42	477933	10.19	641876	10.9	1093273	10.65	496620	11.23	611296	9.85	15309	18.06	1064322	11.59	62772	12.65
	70-79	360187	11.74	0	0	30288	5.5	406935	6.83	352337	7.51	532903	9.05	901369	8.78	421301	9.52	473386	7.63	20172	23.8	798490	8.69	45388	9.15
	80+	238042	7.76	0	0	8126	1.48	202170	3.39	181250	3.87	280152	4.76	713488	6.95	379301	8.57	318665	5.14	12078	14.25	642674	7	19720	3.97
Person-Years across follow-up period (per age category)	0-19	1396241	15.4	356184	100	314386	20.38	5417245	24.09	2863126	22.43	4103324	19.62	6368798	18.45	1475624	17.37	3878177	19.54	1002231	18.72	4285628	16.13	221438	19.01
	20-29	796689	8.79	-	-	187022	12.12	2639524	11.74	1437077	11.26	2073398	9.92	3294295	9.54	780125	9.18	2014182	10.15	536922	10.03	3201274	12.05	134601	11.56
	30-39	908683	10.02	-	-	216633	14.04	2890365	12.85	1709558	13.4	2746498	13.13	4412677	12.78	1070247	12.6	2801475	14.12	747012	13.95	3507399	13.2	133829	11.49
	40-49	1391120	15.35	-	-	256492	16.63	3294106	14.65	1754888	13.75	3460336	16.55	5490532	15.9	1319480	15.53	3447041	17.37	933846	17.44	3862330	14.54	160034	13.74
	50-59	1478656	16.31	-	-	256223	16.61	2923975	13	1854253	14.53	3012592	14.41	5095982	14.76	1255375	14.78	2782154	14.02	762148	14.23	3934798	14.81	182981	15.71
	60-69	1180125	13.02	-	-	157667	10.22	2778684	12.36	1446057	11.33	2399175	11.47	4003342	11.6	995612	11.72	2129361	10.73	589617	11.01	3153788	11.87	157884	13.55
	70-79	1068981	11.79	-	-	111715	7.24	1618270	7.2	1074750	8.42	1887325	9.03	3034922	8.79	765961	9.02	1577614	7.95	447477	8.36	2368347	8.91	115293	9.9
	80+	844776	9.32	-	-	42586	2.76	928047	4.13	622638	4.88	1228553	5.88	2825050	8.18	831925	9.79	1215702	6.13	335871	6.27	2254547	8.49	58751	5.04
Person-Years across follow-up period (per sex)	Female	4706024	51.91	170278	47.81	803484	52.08	11319983	50.33	6408155	50.21	10641698	50.89	17701930	51.27	4347542	51.18	10067763	50.73	2713540	50.67	13967221	52.57	589204	50.58
	Male	4359247	48.09	185906	52.19	739239	47.92	11170234	49.67	6354194	49.79	10269503	49.11	16823667	48.73	4146807	48.82	9777943	49.27	2641584	49.33	12600891	47.43	575606	49.42
At-risk population in the year before study start*	Cardiovascular disease	448606	17.13	<5	<0.01	-	-	438670	7.95	349525	8.06	155803	2.72	722656	7.57	308597	7.5	441907	7.67	132624	8.16	468206	5.37	44764	9.92
	Cancer	69092	2.3	<5	<0.01	-	-	45090	0.76	28094	0.6	86696	1.49	101228	1.00	42687	0.97	78529	1.28	26587	1.54	75034	0.82	13243	2.74
	Chronic lung disease	181345	6.28	2285	1.28	-	-	309020	5.47	403630	9.42	181890	3.19	458595	4.68	154488	3.62	332547	5.66	107038	6.48	294824	3.32	28056	5.99
	HIV	7133	0.23	14	0.01	-	-	487	0.01	331	0.01	7362	0.13	2063	0.02	543	0.01	383	0.01	320	0.02	3485	0.04	317	0.06
	Chronic kidney disease	15670	0.51	5	0	-	-	7391	0.12	19042	0.41	63101	1.08	22411	0.22	11238	0.25	31226	0.51	14991	0.86	11994	0.13	3279	0.67
	Type 2 Diabetes	166839	5.75	69	0.04	-	-	154150	2.66	170149	3.77	376405	6.83	508261	5.21	188534	4.45	353886	6.05	109886	6.67	226023	2.52	22471	4.74
	Severe obesity (BMI > 30)	4550	0.15	1138	0.63	-	-	35377	0.6	16122	0.35	84494	1.46	74578	0.73	27598	0.63	140851	2.32	51838	3.04	9524	0.1	2463	0.5
	Sickle cell disease	2933	0.1	53	0.03	-	-	421	0.01	1621	0.03	1111	0.02	5714	0.06	1943	0.04	4836	0.08	1800	0.1	917	0.01	119	0.02
	Use of immunosuppressants	181348	6.28	<5	<0.01	-	-	86139	1.47	16564	0.35	35715	0.61	114371	1.13	42584	0.97	93600	1.53	32927	1.91	114507	1.26	10455	2.15

** at risk population identified based on medical diagnosis or drug proxies*

BMI: Body Mass Index. Note: PEDIANET only captures children aged 0-14 years

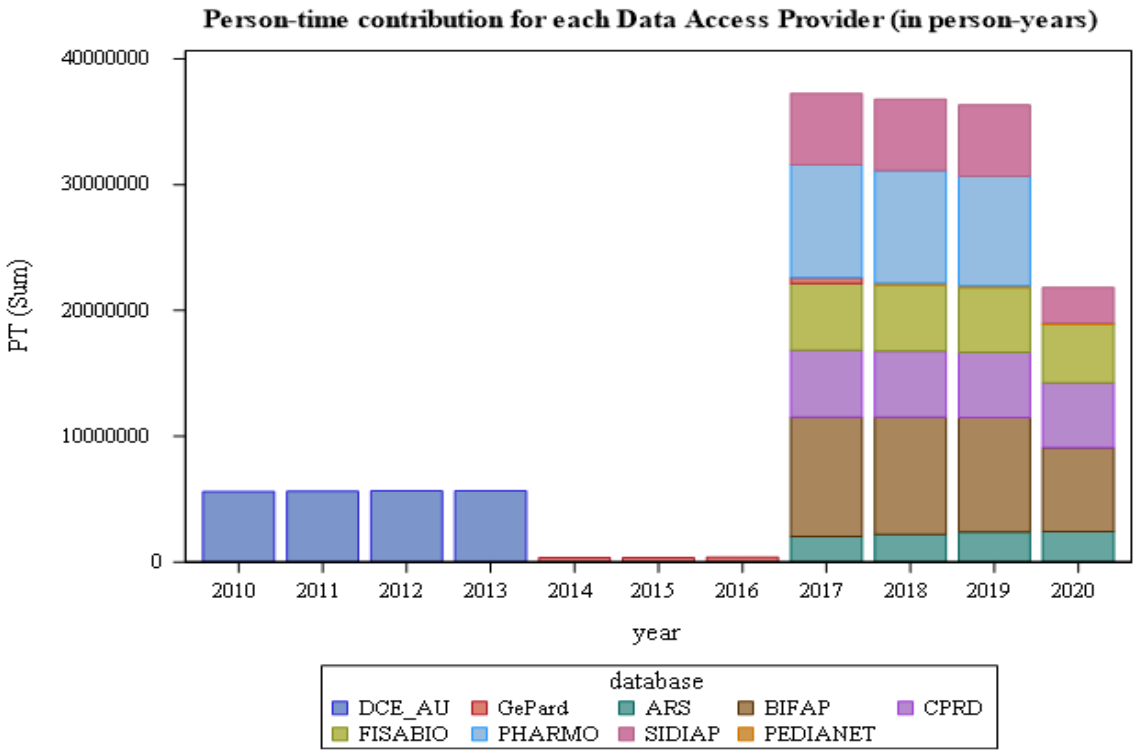


Figure 3: Person-time (years) over calendar and by data source

Figure 3 shows the overview of the amount of person time that could be contributed in each calendar year by DAPs. Full 2020 data was available on the entire population for ARS, whereas for BIFAP, the amount of person time was about 75% with respect to 2019, due to lag time of data, similarly for CPRD. 2020 data was complete for FISABIO and Pedianet, whereas SIDIAP had only information for half of the period. PHARMO could not yet deliver updated data on 2020.

10.3 Outcome data

The number of subjects with an incident occurrence of each AESI according to narrow and broad clinical definitions are presented in Table 10. For most of the AESIs, the broad clinical definition increased drastically the number of events identified in the data sources. For GBS, arrhythmia and erythema multiforme, stress cardiomyopathy, sudden death and TTS, only narrow definition was available. Acute Aseptic Arthritis did not contain medical codes for narrow definition.

Table 10 Number of incident events over the study period according to the narrow and broad definition for each AESI in each data source

AESI	DE_GEPARD		ES_BIFAP_PC		ES_BIFAP_PCHOSP		ES_FISABIO		ES_SIDIAP_PC		ES_SIDIAP_PCHOSP		IT_ARS		IT_PEDIANET		NL_PHARMO_HOSP		NL_PHARMO_PCHOSP		UK_CPRD		DK_DCE_AU	
	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow	n_Broad	n_narrow
ADEM	391	-	4076	44	1975	25	12904	111	2498	<5	2299	5	1894	7	15	-	3929	-	396	-	10922	<5	-	24591
Acute Aseptic Arthritis	94	-	103657	1218	26709	266	139965	-	53931	-	19299	-	7115	-	<5	-	4897	-	5817	-	42409	161	-	20086
Guillain Barre Syndrome	47	47	411	411	178	178	791	791	453	453	211	211	326	326	<5	<5	332	332	5529	34	209	209	625	625
Narcolepsy	1202	15	87969	645	16995	122	299300	448	50021	218	23236	63	529	37	485	-	23571	88	4939	7	26376	129	694	42889
Thrombocytopenia	85	78	14709	14662	5504	5480	27692	27578	16339	16317	7099	7086	2891	2868	27	27	5153	5113	444	440	2679	2634	4195	4307
Type 1 Diabetes Mellitus	461	206	180957	3180	44864	1037	147298	13205	154115	4629	70336	2132	55495	3884	86	70	68832	4285	11617	500	60317	3946	38182	110730
Arrhythmia	2737	2737	230935	230935	71238	71238	203807	203807	160146	160146	64451	64451	84143	84143	496	496	83894	83894	22681	22681	58114	58114	135732	135732
Coronary Artery Disease	3786	2352	177958	28108	43734	10288	90177	38868	50964	16807	29212	6669	56552	24072	7	-	82365	27222	19124	4156	35838	20315	39512	139681
Heart Failure	1674	1630	87593	79694	33010	30035	97527	92546	45909	45664	24279	23698	64196	51894	<5	<5	36521	36106	5211	5179	33620	19184	55714	60017
Microangiopathy	12	7	403	147	101	45	1919	1511	207	62	307	235	79	52	<5	0	417	286	2161	18	2972	73	754	4985
Stress cardiomyopathy	-	-	101	101	118	118	772	772	42	42	212	212	664	664	-	-	-	-	-	-	-	-	-	-
Arterial or VTE with TP	-	<5	-	60	-	23	-	1929	-	47	-	354	-	197	-	-	-	412	-	30	-	31	132	-
Arterial or VTE without TP	-	6081	-	137255	-	46673	-	131462	-	85803	-	29777	-	57892	-	<5	-	56421	-	12364	-	54901	113465	-
Arterial with TP	-	<5	-	31	-	11	-	1176	-	17	-	193	-	101	-	-	-	225	-	16	-	15	69	-
Arterial without TP	-	5008	-	70694	-	26470	-	85019	-	47740	-	17027	-	42352	-	-	-	46113	-	8437	-	35997	77112	-
CVST with TP	-	-	-	-	-	-	-	11	-		-	<5	-	<5	-	-	-	<5	-	-	-	<5	-	-
CVST without TP	-	23	-	98	-	54	-	270	-	25	-	38	-	113	-	<5	-	205	-	13	-	24	232	-
Cerebral venous thrombosis	21	21	98	98	54	54	280	280	25	25	40	40	115	115	<5	<5	205	205	13	13	24	23	233	233
Disseminated Intravascular Coagulation	<5	<5	69	69	38	38	1242	1242	14	14	178	178	153	153	<5	<5	161	161	8	8	22	22	567	567
Hemorrhagic stroke	434	434	6718	6718	3246	3246	12892	12892	2820	2820	2366	2366	7131	7131	7	7	4681	4681	4067	441	1157	1157	8912	8912
Ischemic stroke	2556	2409	66284	43475	22982	16596	65286	51975	41598	31646	13595	10934	30339	20065	<5	0	27434	20127	7331	4577	16495	16495	39772	56643
Myocarditis	106	106	663	663	184	184	618	618	153	153	185	185	517	517	<5	<5	350	350	335	335	374	374	3517	3517
Myocarditis or Pericarditis	177	177	6641	5436	1892	1586	2288	2288	3433	3433	1529	1529	3375	2848	10	10	1072	1072	406	406	2361	1580	770	770
SOCV	94	74	3697	2021	1037	614	10616	5484	1594	864	717	405	592	592	92	92	1721	394	709	119	3149	1927	3345	3585
Thrombotic microangiopathy	6	6	262	262	79	79	289	289	48	48	57	57	52	52	0	0	133	133	10	10	35	35	214	214
VTE with TP	-	-	-	30	-	13	-	809	-	32	-	172	-	100	-	-	-	194	-	14	-	16	68	-
VTE without TP	-	1147	-	69176	-	21062	-	53256	-	39716	-	13616	-	17383	-	<5	-	11390	-	4273	-	19795	39387	-
Venous Thromboembolism	941	941	69196	69196	21072	21072	53643	53643	39743	39743	13726	13726	17460	17460	<5	<5	11549	11549	4282	4282	19808	19808	39428	39428
Acute Kidney Injury	1496	718	58768	24840	27178	13007	172582	154323	101530	94916	50991	45256	44649	26457	57	<5	63421	53895	11156	4871	23226	16026	42579	97076
Acute Liver Injury	188	122	6622	6547	1897	1862	10636	9663	5274	4931	2591	2353	2348	2348	7	7	2208	1979	1664	386	821	793	4350	5578
Generalized Convulsion	1200	1200	30326	21152	9019	7728	51095	51095	17056	17056	7734	7734	12730	12730	523	523	11408	11408	2576	1802	16981	13332	45839	45839
Meningoencephalitis	158	129	2511	1562	1216	736	7859	2112	1057	364	1386	261	1720	555	15	8	955	332	358	48	565	336	919	1411
Transverse Myelitis	17	13	64	43	57	23	331	190	95	83	50	38	127	99	-	-	113	63	8	3	130	106	275	320
Anaphylaxis	854	114	418023	1999	99401	761	361020	4556	149196	2166	48418	736	14022	819	470	11	19064	857	11774	140	120200	2355	2396	30304
Anosmia	<5	<5	5618	5604	1384	1384	8601	8592	4556	4552	1303	1303	12	9	<5	<5	32	30	668	668	2752	2748	246	258
Death	-	9306	-	306753	-	85218	-	174415	-	184318	-	51971	-	145052	-	-	-	<5	-	<5	-	111286	215353	-
MIS	20	7	285	170	92	54	1058	411	114	69	135	21	1223	65	19	19	643	97	51	12	69	69	141	512
Sudden death	<5	<5	556	556	94	94	20592	20592	46700	46700	12605	12605	237	237	<5	<5	-	-	-	-	136	136	8170	8170
ARDS	99	85	18495	16414	6971	6257	37446	32478	1477	1154	2660	2123	8704	4021	7	7	6612	6364	507	472	4554	2953	22156	24995
Chilblain like lesions	<5	<5	9471	9471	2051	2051	9759	9759	2845	2845	702	702	13	13	15	15	<5	<5	19	19	1405	1405	102	102
Erythema multiforme	17	17	2096	2096	647	647	2514	2514	1494	1494	492	492	718	718	21	21	92	92	9713	20	946	946	1021	1021
Colonic diverticulitis	-	18618	-	48274	-	15925	-	54254	-	42170	-	17924	-	11954	-	-	-	23228	-	2304	-	27525	42388	-
Hypertension	-	38486	-	1726424	-	396073	-	336831	-	1000972	-	274637	-	667603	-	36	-	263768	-	23168	-	311454	367437	-

'-' indicates no event detected in the database

Main results

Incidences rates using narrow codes are presented in the core text of this report for each AESI. Incidence rates for broad codes definition, age- and sex stratified incidence rates, monthly incidence rates and incidence rates in the at-risk population are presented in annexes to this report. These should be considered as sensitivity analyses.

Yearly incidence rates for all participating data sources are presented using forest plots, as well as age specific incidence rates. Where possible incidence rates from the literature are also described to allow benchmarking with external references.

10.3.1 AESIs

10.3.1.1 Guillain Barre syndrome

For definition codes see
https://doi.org/10.5281/zenodo.5109436

Guillain Barré syndrome is an immune-mediated disorder which can lead to autoimmune antibodies and/or inflammatory cells that cross react with components of peripheral nerves and roots, leading to demyelination or axonal damage or both. This results into various degrees of weakness, sensory abnormalities and autonomic dysfunction. The clinical findings patients with GBS present with are acute or subacute onset of varying degrees of weakness in limbs or cranial nerve-innervated muscles, associated with hypo –or areflexia and a characteristic profile in the cerebrospinal fluid (CSF).[13]

GBS was detected as event in most data sources. The incidence of GBS was consistent between 1 and 5 per 100,000 person-years (figure 4). The rate increased with age as expected and lowers again in the highest age category (figure 5). Data based on GP data alone have consistently lower rates than analyses including both GP and hospitalizations, which is understandable since GBS is a condition that requires hospitalization.

These rates are consistent with prior incidence studies where incidence rates ranged between 1 and 2 per 100,000 person-years and increased with age[14]. Data generated by Li et al. showed higher point estimates ranging between 1 and 12/100,000 [15]. Gubernot et al. provided rates within our observed ranges[16].

[13] Sejvar JJ, Kohl KS, Gidudu J, Amato A, Bakshi N, Baxter R, e.a. Guillain–Barré syndrome and Fisher syndrome: Case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 10 January 2011;29(3):599–612. https://doi.org/10.1016/j.vaccine.2010.06.003.

[14] van der Maas NA, Kramer MA, Jacobs BC, van Soest EM, Dieleman JP, Kemmeren JM, de Melker HE, Sturkenboom MC. Guillain-Barré syndrome: background incidence rates in The Netherlands. J Peripher Nerv Syst. 2011 Sep;16(3):243-9. doi: 10.1111/j.1529-8027.2011.00356.x. PMID: 22003939.

[15] Li X, Ostropolets A, Makadia R, Shaoibi A, Rao G, Sena AG, Martinez-Hernandez E, Delmestri A, Verhamme K, Rijnbeek PR, Duarte-Salles T, Suchard M, Ryan P, Hripcsak G, Prieto-Alhambra D. Characterizing the incidence of adverse events of special interest for COVID-19 vaccines across eight countries: a multinational network cohort study. medRxiv [Preprint]. 2021 Mar 28:2021.03.25.21254315. doi: 10.1101/2021.03.25.21254315. PMID: 33791732; PMCID: PMC8010764.

[16] Gubernot D, Jazwa A, Niu M, Baumblatt J, Gee J, Moro P, Duffy J, Harrington T, McNeil MM, Broder K, Su J, Kamidani S, Olson CK, Panagiotakopoulos L, Shimabukuro T, Forshee R, Anderson S, Bennett S. U.S. Population-Based background incidence rates of medical conditions for use in safety assessment of COVID-19 vaccines. Vaccine. 2021 Jun 23;39(28):3666-3677. doi: 10.1016/j.vaccine.2021.05.016. Epub 2021 May 14. PMID: 34088506; PMCID: PMC8118666.

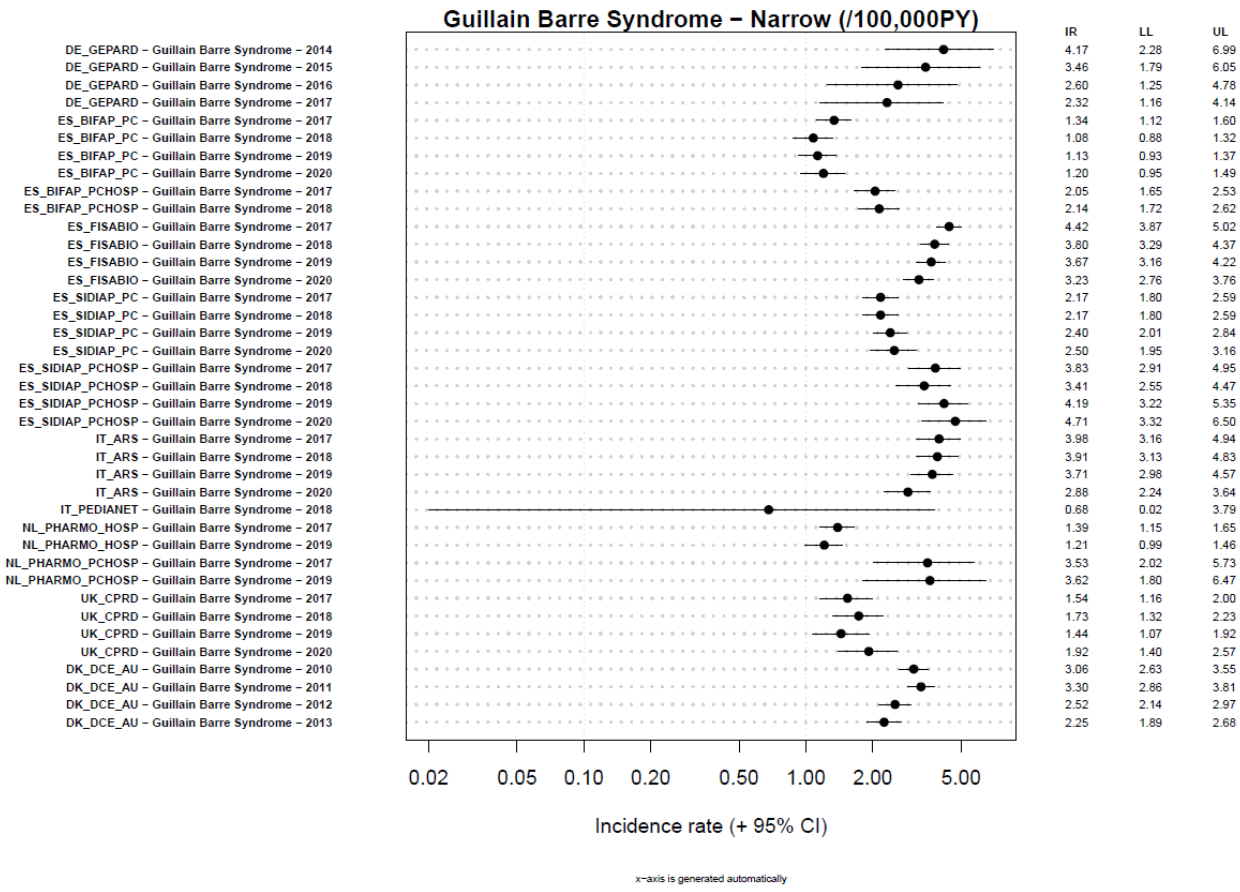


Figure 4: Forest plot for incidence rates of GBS per 100,000 PY by data source and calendar year using a narrow definition

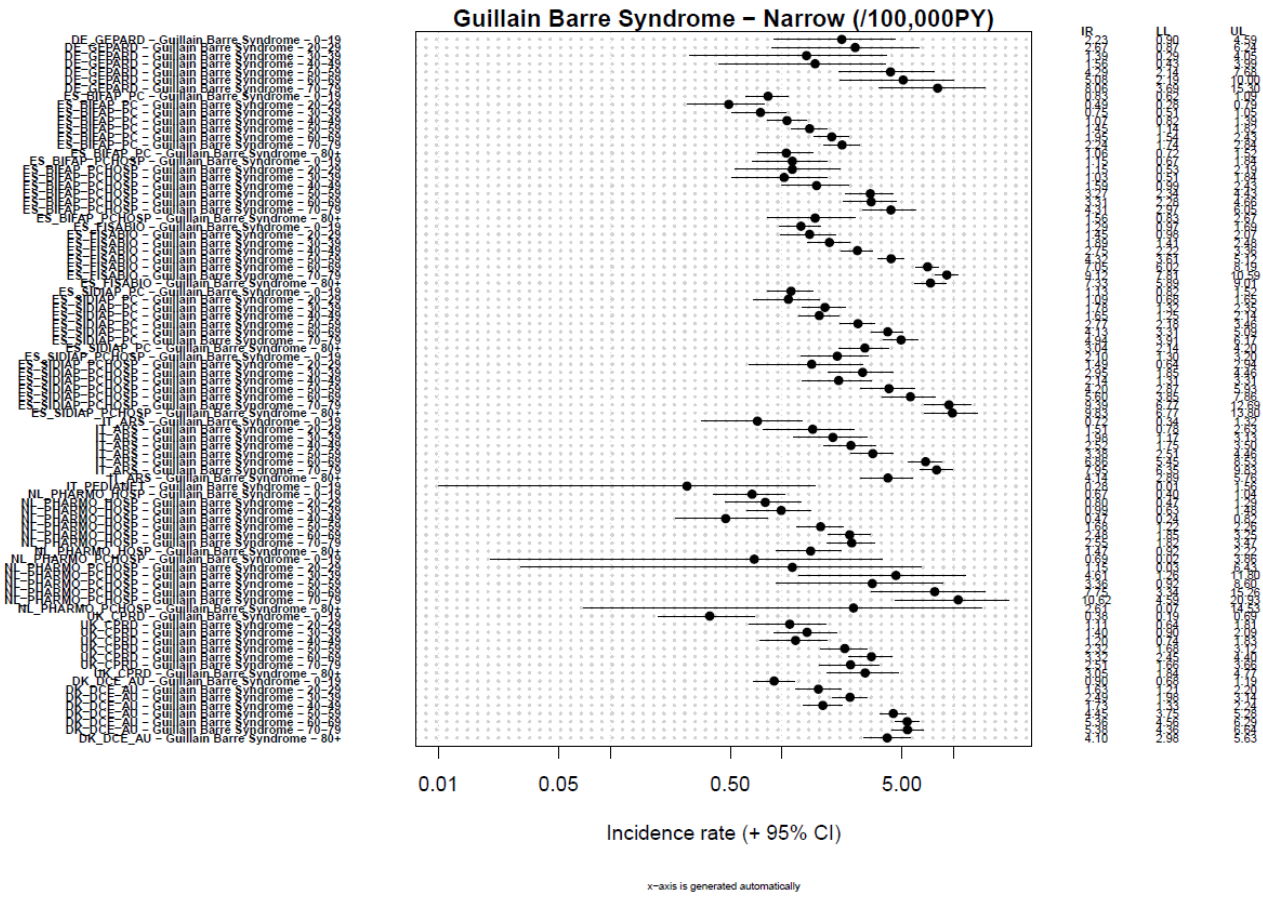


Figure 5: Forest plot for incidence rates of GBS per 100,000 PY by data source and age for narrow definition

10.3.1.2 Acute Disseminated Encephalomyelitis (ADEM)

For definition codes see
https://doi.org/10.5281/zenodo.5109555

Acute disseminated encephalomyelitis, also known as ADEM, is a uni-phasic syndrome where autoantibodies lead to brain inflammation and demyelination, an immune-mediated demyelinating central nervous system disorder. It most likely occurs after an infection or an immunization. ADEM is distinguished from acute encephalitis by a predominance of demyelinating, rather than cytotoxic injury, and a temporal association with a specific inciting immunogenic challenge. It can occur at any age group, but especially in children[14].

ADEM is rare and specific events that can only be seen with SNOMED and ICD-9 codes were only observed in ARS, BIFAP, SIDIAP and FISABIO with rates < 1/100,000 person-years (figure 6). No cases of ADEM using the narrow definition have been identified in PHARMO and ARS. Specific ICD10 codes require two decimals which often are not coded as such. Rates were stable over calendar time. In FISABIO, higher rates were observed in the 0-19 age category (figure 7). There is only a low

[14] Sejvar JJ, Kohl KS, Bilynsky R, Blumberg D, Cvetkovich T, Galama J, e.a. Encephalitis, myelitis, and acute disseminated encephalomyelitis (ADEM): Case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 1 August 2007;25(31):5771–92.

impact of the provenance of data. Our rates are comparable to a literature review published by Gubernot et al. with a range of 0.1-0.5/100,000 for ADEM. Li et al did not report rates for ADEM.

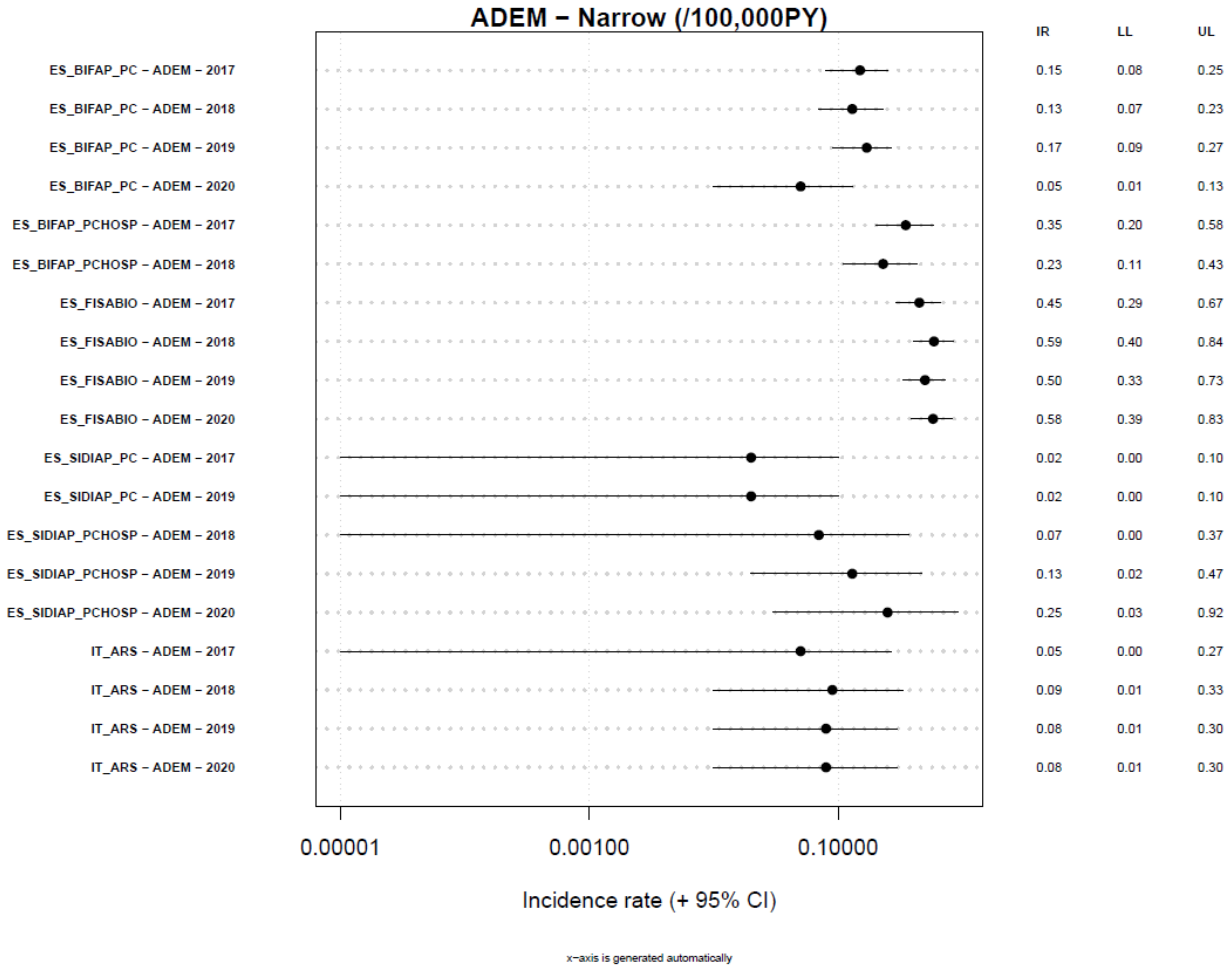


Figure 6: Forest plot for incidence rates of ADEM per 100,000 PY by year

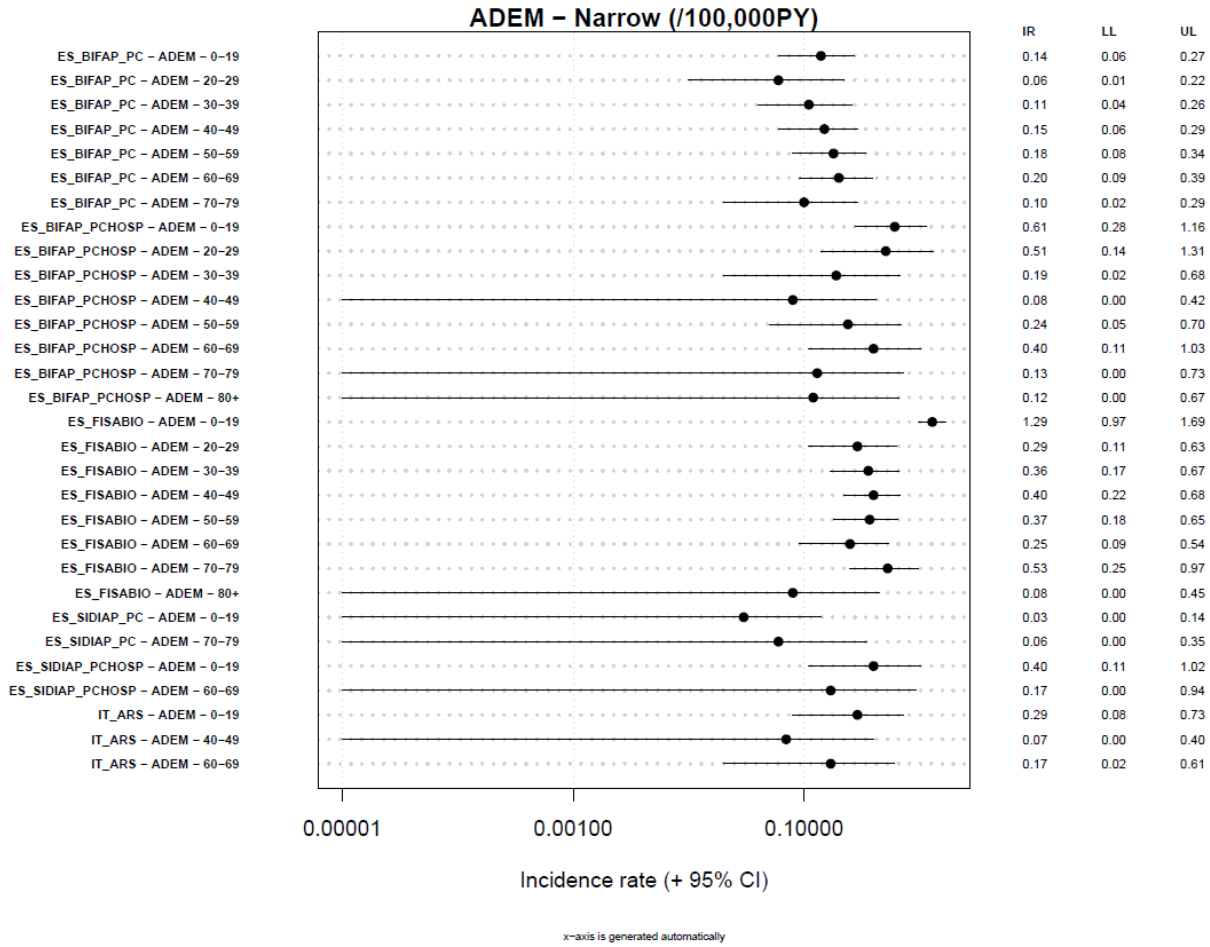


Figure 7: Forest plot of incidence rate of ADEM by age category

Incidence rates from the literature in the USA showed an incidence rate of 0.4 per 100,000 PY [15] for persons below age 20, very similar to what we observe. ADEM overlaps with encephalitis and may be misclassified especially in elderly.

10.3.1.3 Narcolepsy

For definition codes see
https://doi.org/10.5281/zenodo.5110083

Narcolepsy is a sleep disorder primarily characterized by excessive daytime sleepiness and cataplexy-episodes of muscle weakness brought on by emotions. Additional symptoms may comprise hypnagogic hallucinations, sleep paralysis, fragmented nocturnal sleep, as well as impaired ability for sustained attention and non-sleep symptoms such as obesity, anxiety, cognitive and emotional disturbances, and behavioral problems and precocious puberty in children.

[15] *Leake JAD , Albani S , Kao AS et al . Acute disseminated encephalomyelitis in childhood: epidemiologic, clinical and laboratory features . Pediatr Infect Dis J 2004 ; 23 : 756 – 764*

Events of narcolepsy were observed in all databases, except Pedianet. Incidence rates were slightly higher in Denmark, ranging from 2.39/100,000 person-years in 2013 to 4.03/100,000 person-years in 2011. In databases for which the year 2020 was included in the study (all except DCE-AU), smaller rates were observed, which may be due to the lock down (less access to sleeping test) or lack of full year data. The observed rates compare well to the crude rates observed in the VAESCO[16], SOMNIA[17] and ADVANCE[18].
ARS underestimates the rate of narcolepsy as this is data source captures only hospitalization data and emergency room visits and narcolepsy does not require a hospitalization, it is generally picked up in data sources that capture outpatient or primary care diagnoses. (figure 8).

In a prior study by Oberle et al. a total of 233 sleep centers participated in estimation of incidence using ICD-9 code G47.4). A total of 1,198 patients with an initial diagnosis of narcolepsy within the observed period were included, of whom 106 (8.8%) were children and adolescents under the age of 18 years and 1,092 (91.2%) were adults. In children and adolescents, the age-standardized adjusted incidence rate significantly increased from 0.14/100,000 person-years in the pre-pandemic period to 0.50/100,000 person-years in the post-pandemic period (incidence density ratio, IDR 3.57; 95% CI 1.94–7.00). In adults, no significant change was detectable. The increase started in spring 2009[19]. Our data are consistent with the data from the literature reported by Gubernot et al. (2021)[16] . Recent data by Li et al. showed rates up to 10-fold higher in their data (15/100,000) in the age categories between 18-34[20].

[16] Wijnans L, Lecomte C, de Vries C, Weibel D, Sammon C, Hviid A, et al. The incidence of narcolepsy in Europe: before, during, and after the influenza A(H1N1)pdm09 pandemic and vaccination campaigns. Vaccine 2013;31(8):1246–54.
[17] Dodd CN, de Ridder M, Huang W-T, Weibel D, Giner-Soriano M, Perez-Vilar S, et al. Incidence rates of narcolepsy diagnoses in Taiwan, Canada, and Europe: The use of statistical simulation to evaluate methods for the rapid assessment of potential safety issues on a population level in the SOMNIA study. PLoS ONE 2018;13(10):e0204799.
[18] Willame, C., Dodd, C., van der Aa, L. et al. Incidence Rates of Autoimmune Diseases in European Healthcare Databases: A Contribution of the ADVANCE Project. Drug Saf (2021). https://doi.org/10.1007/s40264-020-01031-1
[19] Oberle D, Drechsel-Bäuerle U, Schmidtmann I, Mayer G, Keller-Stanislawski B. Incidence of Narcolepsy in Germany. Sleep 2015;38(10):1619–28.
[20] Li X, Ostropolets A, Makadia R, Shaoibi A, Rao G, Sena AG, Martinez-Hernandez E, Delmestri A, Verhamme K, Rijnbeek PR, Duarte-Salles T, Suchard M, Ryan P, Hripcsak G, Prieto-Alhambra D. Characterizing the incidence of adverse events of special interest for COVID-19 vaccines across eight countries: a multinational network cohort study. medRxiv [Preprint]. 2021 Mar 28:2021.03.25.21254315. doi: 10.1101/2021.03.25.21254315. PMID: 33791732; PMCID: PMC8010764.

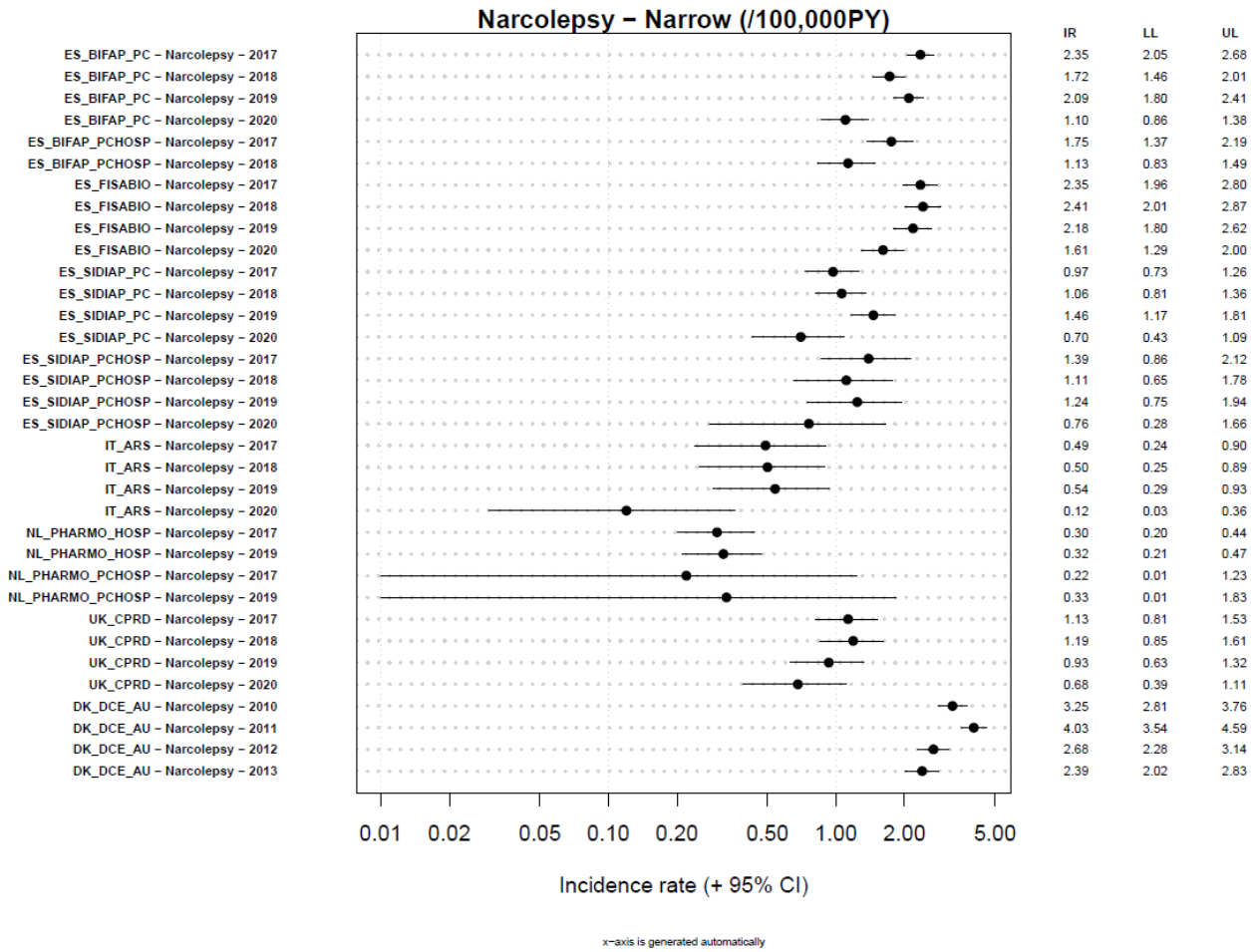


Figure 8: Incidence rates of narcolepsy per 100,000 PY and calendar year

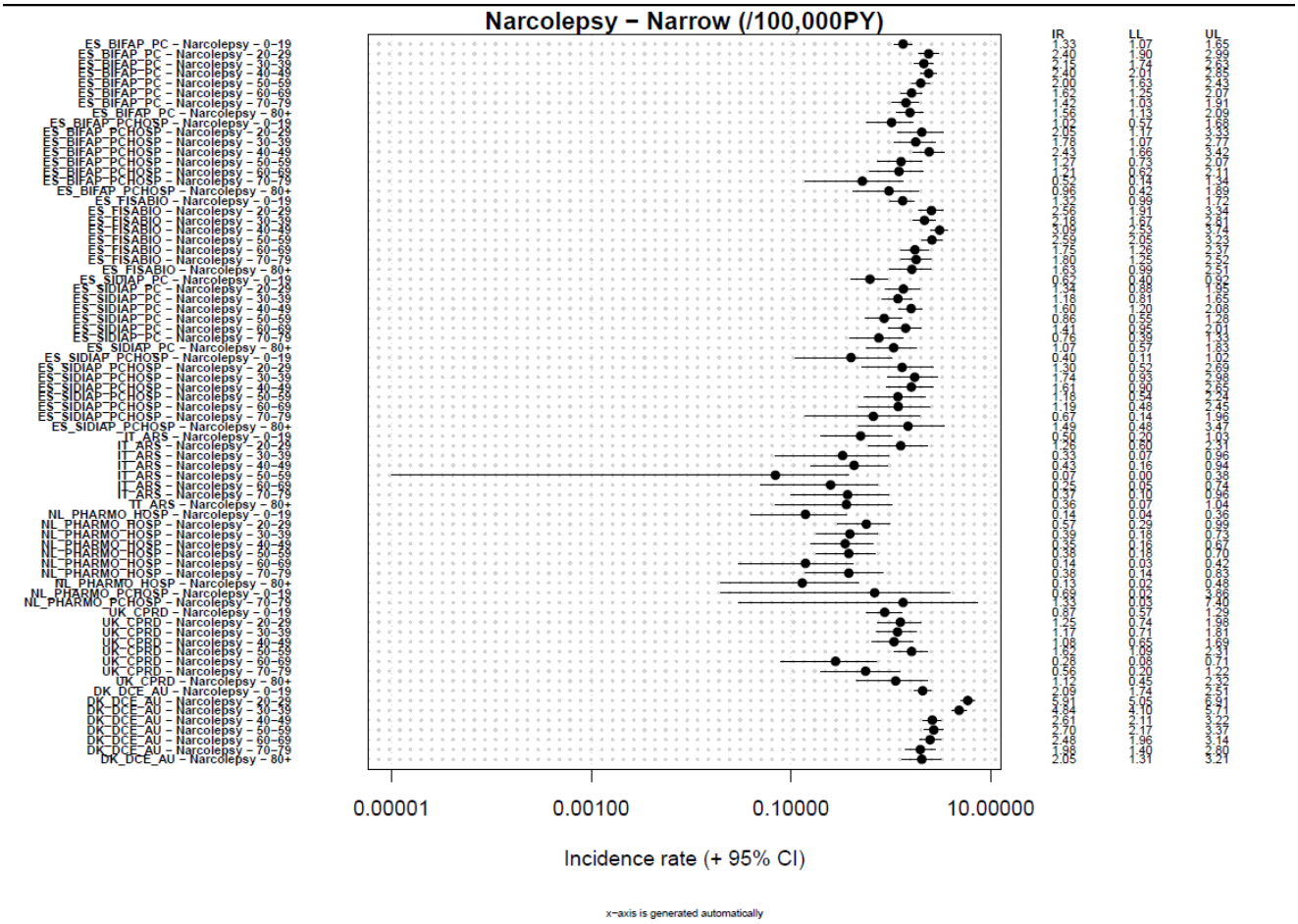


Figure 9: Incidence rates of narcolepsy per 100,000 PY and age

10.3.1.4 Acute Aseptic Arthritis
Definition and codes in: https://doi.org/10.5281/zenodo.5110155

Acute aseptic arthritis (AAA) is a clinical syndrome characterized by acute onset of signs and symptoms of joint inflammation for a period of no longer than 6 weeks, synovial increased leucocyte count and the absence of microorganisms on Gram stain, routine culture and/or PCR. AAA doesn't include chronic inflammatory conditions such as rheumatoid arthritis (RA), connective tissue diseases, osteoarthritis vasculitis or spondylarthropathies. These conditions are chronic and are diagnosed later than within 6 weeks[21]

We could not identify narrow codes specific for AAA in any of the vocabularies. Therefore, only a broad definition that includes many other arthritic diseases could be used to generate incidence rates (see broad excel sheets).

[21] Woerner A, Pourmalek F, Panozzo C, Pileggi G, Hudson M, Caric A, et al. Acute aseptic arthritis: Case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. Vaccine. 2019 Jan 7;37(2):384–91.

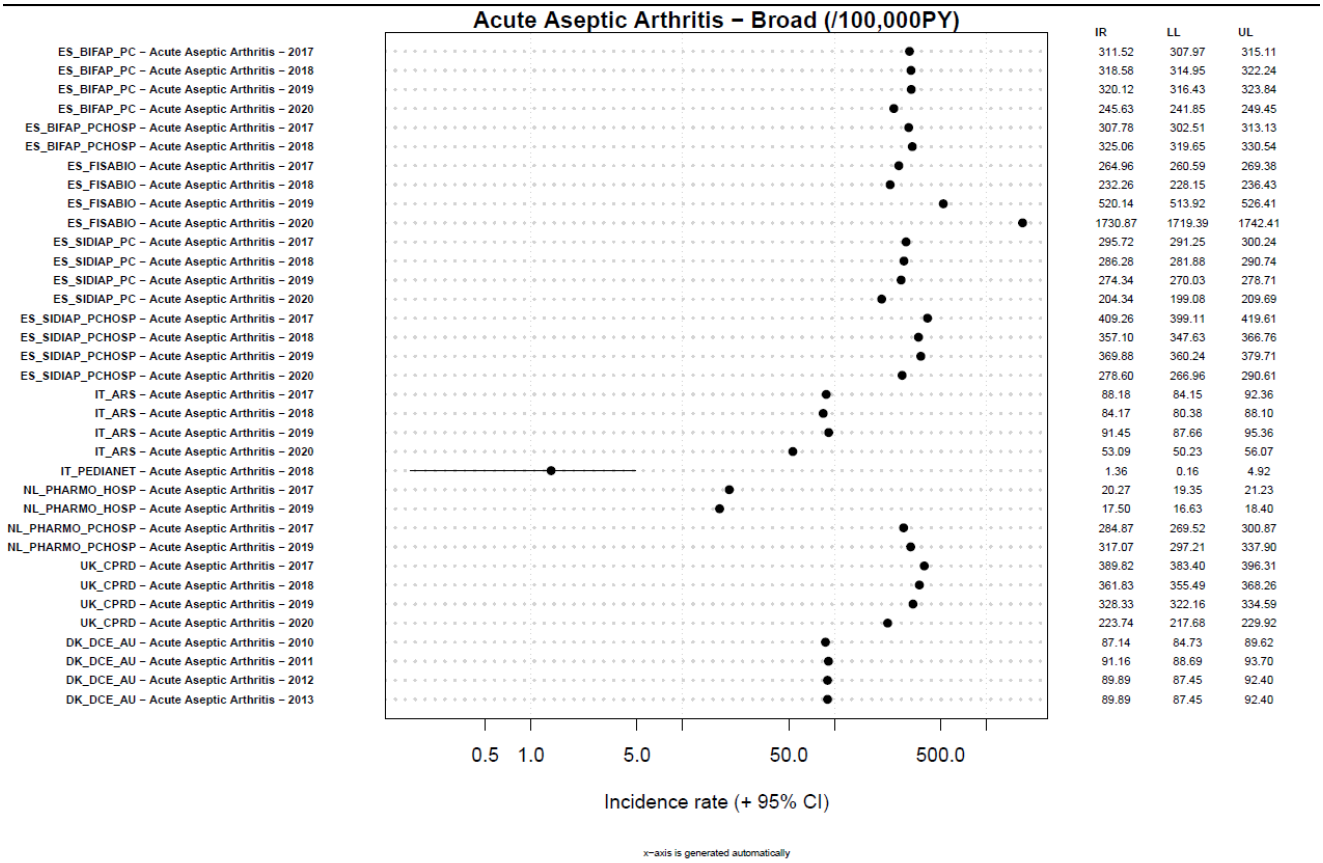


Figure 10: Incidence rates of acute aseptic arthritis (broad definition) per 100,000 PY and calendar year

10.3.1.5 Diabetes Mellitus

Definition and codes in Zenodo
Sturkenboom, MCJM, Willame, C, Belbachir, L, & Duran, C. (2021). ACCESS-Background rate of adverse events-definition -Diabetes mellitus type1 (1.0). Zenodo. https://doi.org/10.5281/zenodo.5110781

Events of diabetes mellitus using the narrow definition diagnosis codes were observed in all participating databases. The rates for Type 1 diabetes mellitus showed a sharp increased with age suggesting a lack of sensitivity of the algorithm. It is likely that Type 2 diabetes mellitus were also reported. For this reason, only rates up to the age of 40 have been included in the report. Neither Li et al. nor Gubernot et al, reported on Type 1 diabetes mellitus rates. A recent meta-analysis by Mobasseri[22] reported an incidence of Type 1 diabetes mellitus of 15/100,000 PY in Europe, we have slightly higher rates as we limit to age 40. In the US, Klein et al.[23] estimated incidence of Type 1 diabetes mellitus at 45.9/100,000 PY.

[22] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7146037/

[23] Klein NP, Ray P, Carpenter D, Hansen J, Lewis E, Fireman B, Black S, Galindo C, Schmidt J, Baxter R. Rates of autoimmune diseases in Kaiser Permanente for use in vaccine adverse event safety studies. Vaccine. 2010 Jan 22;28(4):1062-8. doi: 10.1016/j.vaccine.2009.10.115. Epub 2009 Nov 5. PMID: 19896453.

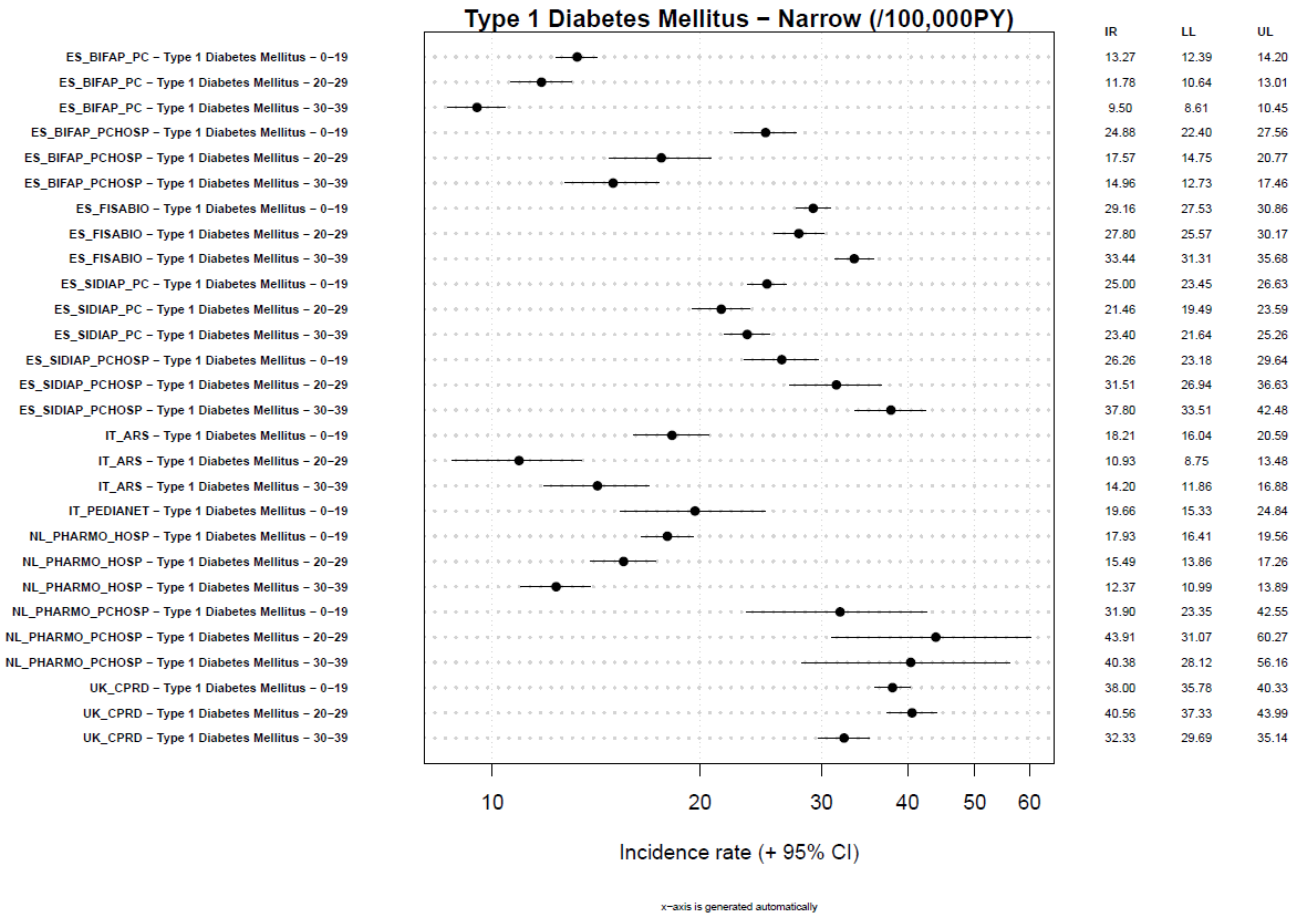


Figure 11: Incidence rates of diabetes mellitus type 1 per 100,000 PY and calendar year and age up to 40

10.3.1.6 Thrombocytopenia

Definition and codes in Zenodo:
Sturkenboom, M, Willame, C, Duran, C, & Belchabir, L. (2021). ACCESS-Background rate of adverse events-definition -thrombocytopenia. Zenodo. https://doi.org/10.5281/zenodo.5169150

Incidence rates or primary and secondary TP are as we expect, and go down in 2020, maybe because of lock down effects since this requires laboratory assessment. Rates are highly age dependent (figure 13), which we would expect especially for secondary TP.

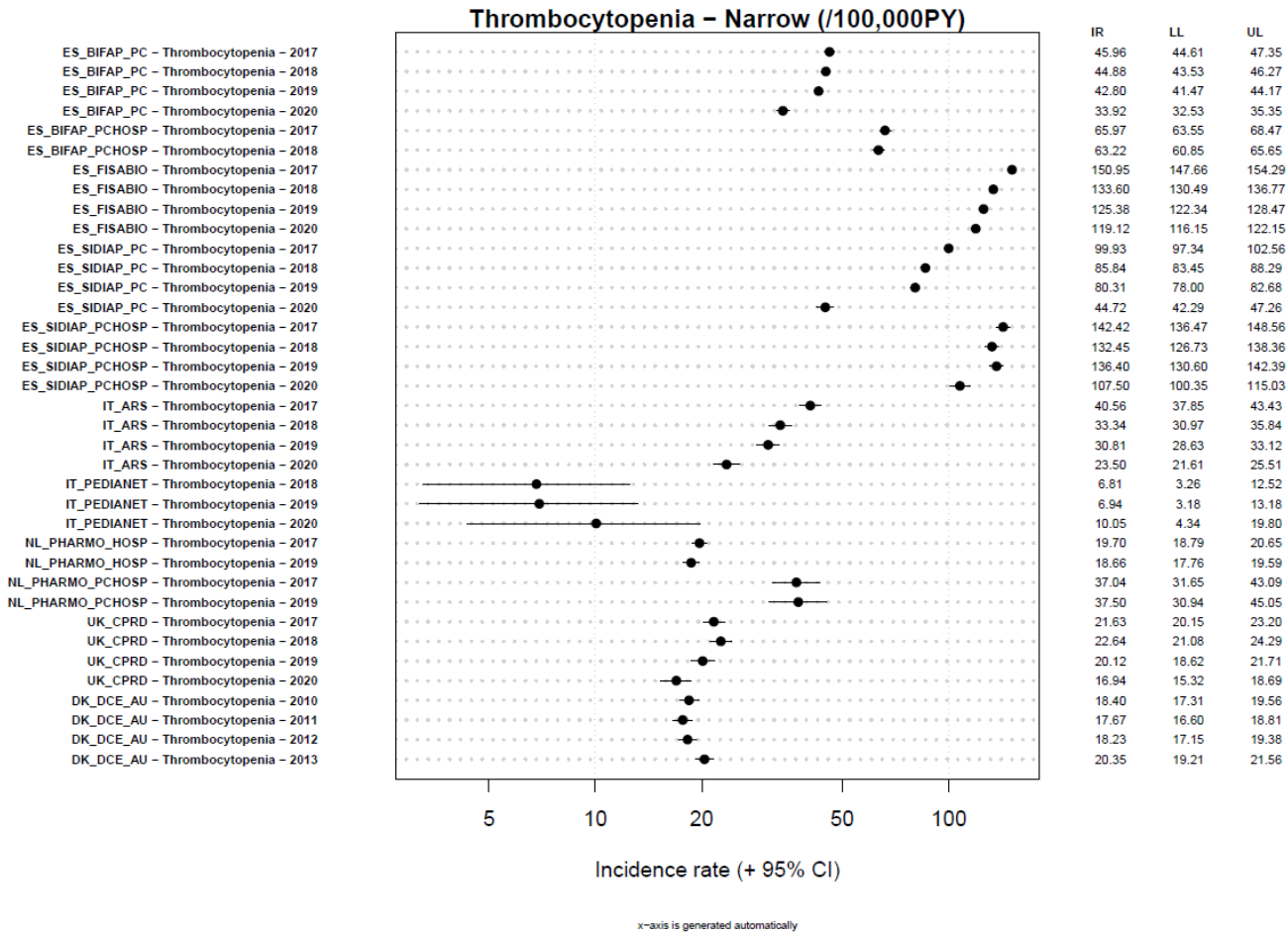


Figure 12: Incidence rates of thrombocytopenia (primary & secondary).

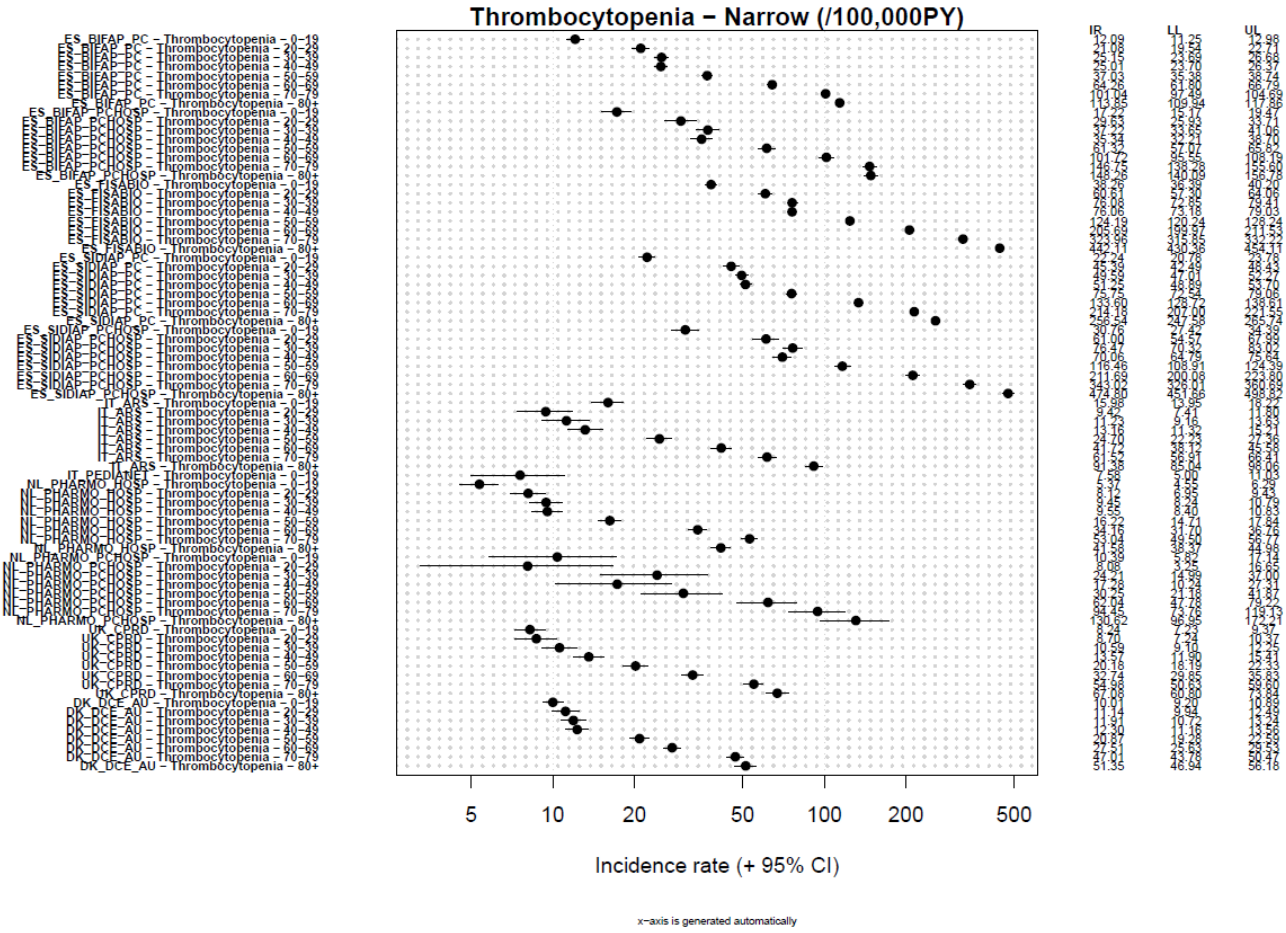


Figure 13: Incidence rates of thrombocytopenia (primary and secondary) by age and data source

Reference data on ITP were available from the ADVANCE study which were separated by type of source data (GP and Hospital based) (figure 14). Since we included both ITP and secondary TP, our rates were higher. Li et al. reported on ITP with rates up to 53/100,000 PY.

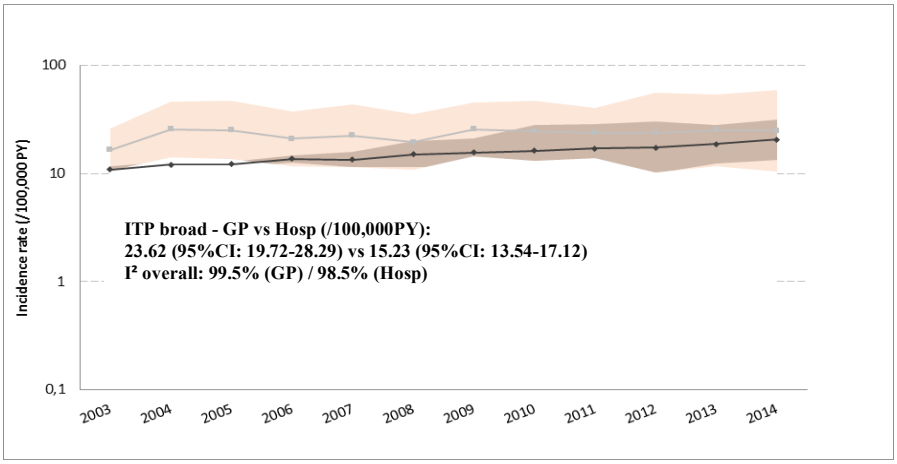


Figure 14: ITP incidence rates (primary and secondary) from IMI-ADVANCE separated by source of data (Willame et al.)[23]

[23] Willame, C., Dodd, C., van der Aa, L. et al. Incidence Rates of Autoimmune Diseases in European Healthcare Databases: A Contribution of the ADVANCE Project. Drug Saf (2021). https://doi.org/10.1007/s40264-020-01031-1

10.3.1.7 Microangiopathy

Definition and codes in Zenodo
Kelters, L, Sturkenboom, MCJM, Willame, C, Belchabir, L, & Durán, L. (2021). ACCESS-Background rate of adverse events-definition –Microangiopathy. Zenodo. https://doi.org/10.5281/zenodo.5169451

Events of microangiopathy using narrow definition were observed in all participating databases, except PEDIANET, which has children only. Incidence rates were stable overtime (figure 15). The rates in FISABIO were higher maybe because of age, FISABIO includes GP, in and outpatient specialist diagnoses. Rates are lowest in GP only datasources. A pattern of increased rates with age was observed in all databases, whereas rates lowered in 2020, potentially because of the lock down.

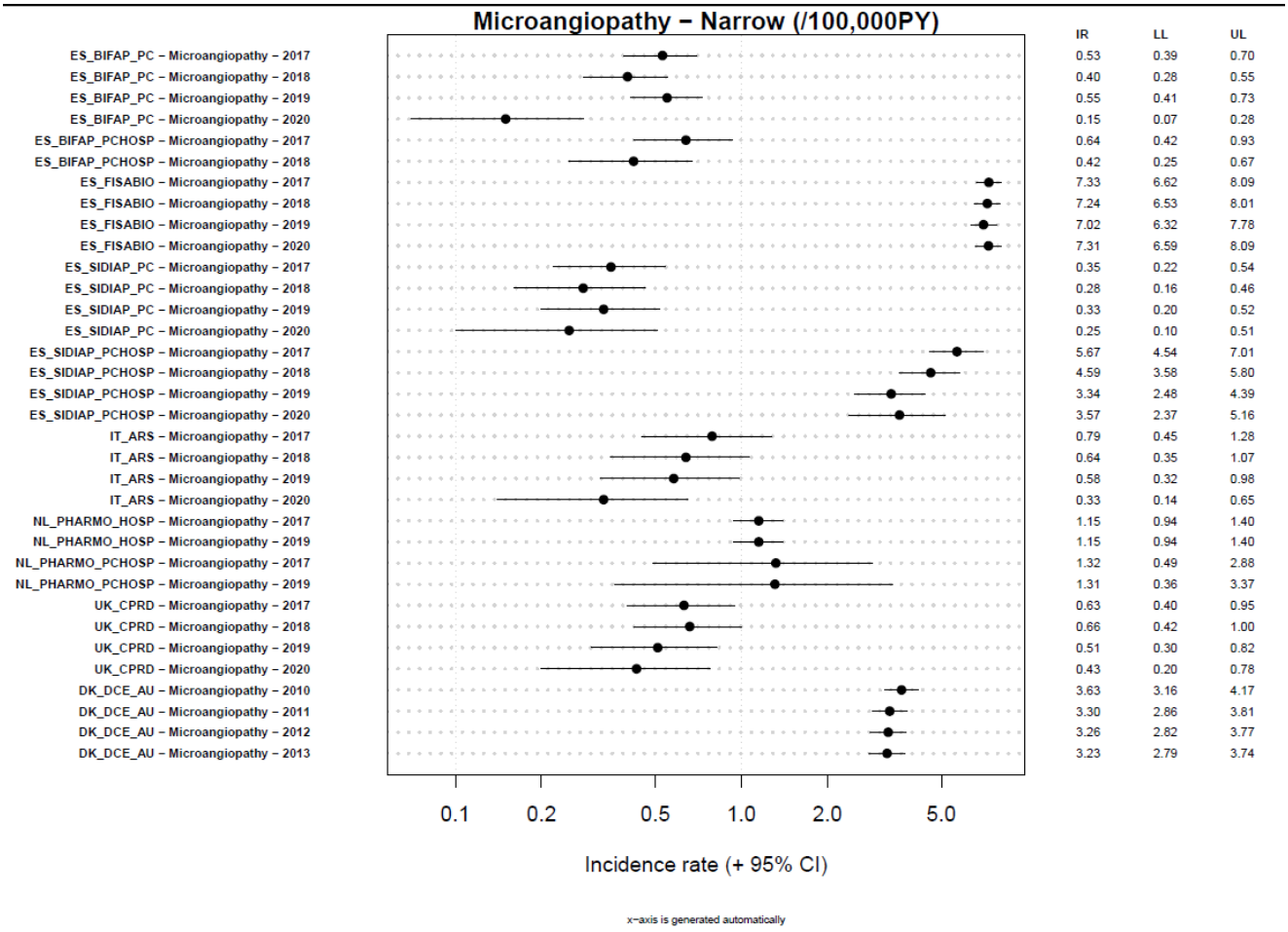


Figure 15: Incidence rates of microangiopathy by calendar year and data source

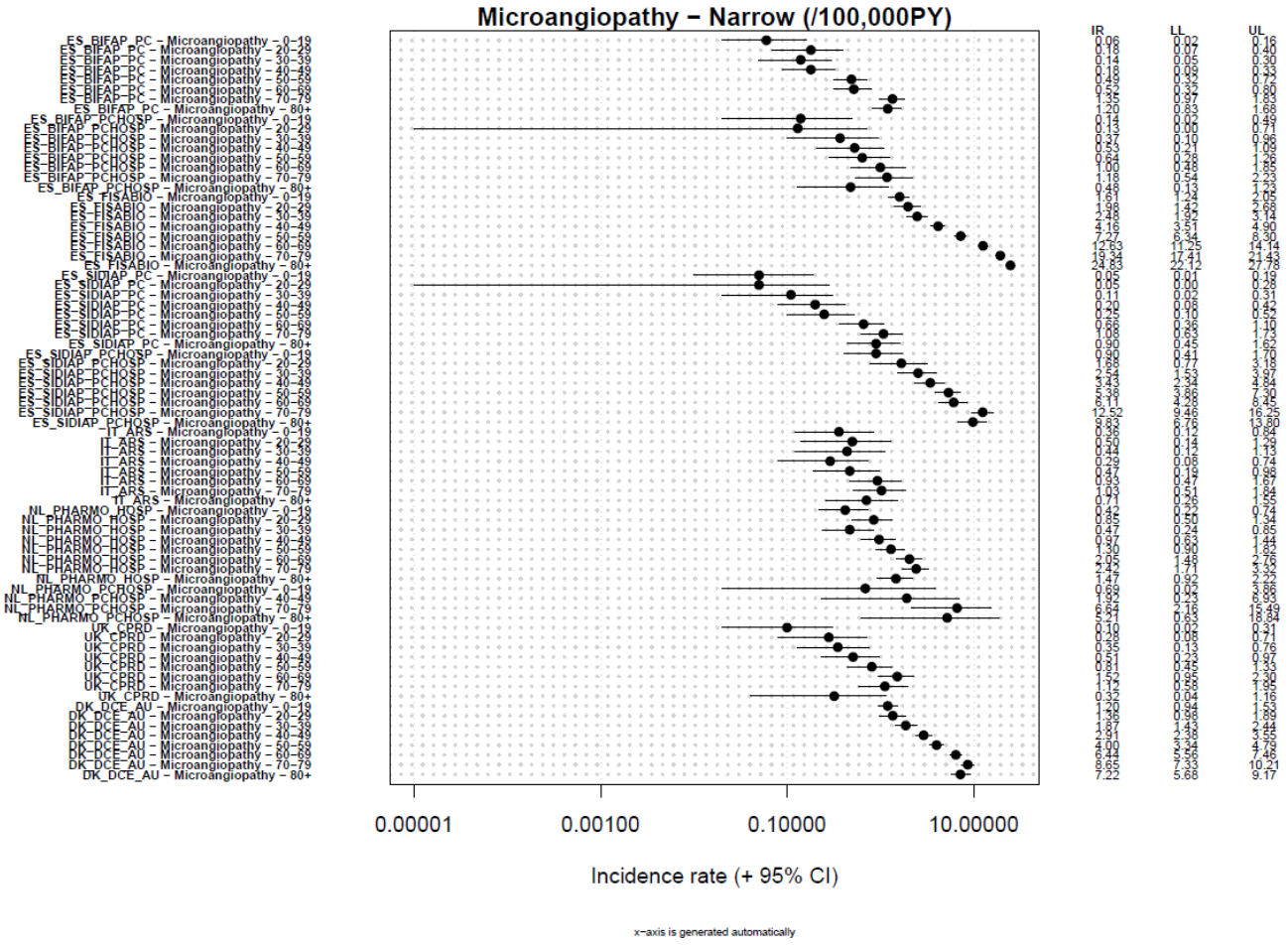


Figure 16: Incidence rate of microangiopathy by data source and age category

There were no background rates as benchmark in the published literature.

10.3.1.8 Heart Failure

Definition and codes in Zenodo
Kelters, I, Souverein, P, Huerta, C, Martín-Pérez, M, García-Poza, P, Belbachir, L, Willame, C, Durán, C, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition – Heart Failure (1.1). Zenodo. https://doi.org/10.5281/zenodo.5226393

Heart failure was observed in all participating databases (figure 17). A clear pattern of increased rates with age was observed in all databases (figure 18). In PEDIANET the rate was very low, as this data source captures children only.

According to Groenewegen et al. the incidence of heart failure in European countries and the USA ranges widely from 1 to 9 cases per 1000 person-years and strongly depends on the population studied and the diagnostic criteria used. In developed countries, incidence rates have stabilized between 1970 and 1990 and are now thought to be decreasing. Our observed incidence rates are in line with that range.

From published articles, incidence rates were estimated at 295/100,000 person-years [24] (Corrao, 2014) from Italy and 306/100,000 person-years from Canada [25]. A recent study conducted in claims database in Germany found an incidence rate of heart failure of 655/100,000 person-years [26]. A study from the US showed increased incidence in older population [27]. Our rates are in line with published rates and show the strong increase with age. Neither Li nor Gubernot published data on heart failure incidence.

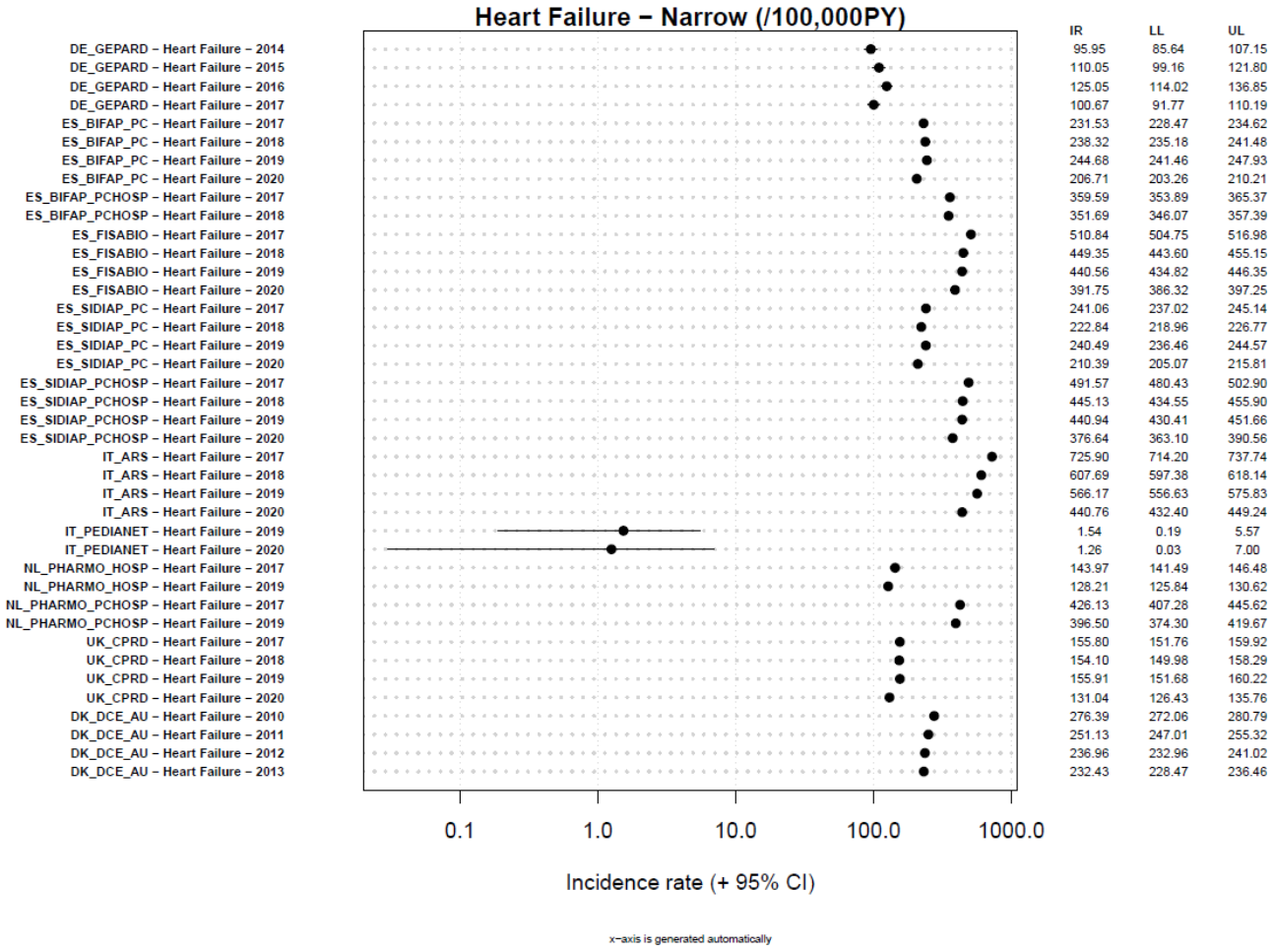


Figure 17: Incidence of heart failure by data source and calendar year

[24] Corrao G, Ghirardi A, Ibrahim B, Merlino L, Maggioni AP. Burden of new hospitalization for heart failure: a population-based investigation from Italy. Eur J Heart Fail. 2014 Jul;16(7):729-36. doi: 10.1002/ejhf.105. Epub 2014 May 7. PMID: 24806352.

[25] Yeung DF, Boom NK, Guo H, Lee DS, Schultz SE, Tu JV. Trends in the incidence and outcomes of heart failure in Ontario, Canada: 1997 to 2007. CMAJ. 2012 Oct 2;184(14):E765-73. doi: 10.1503/cmaj.111958. Epub 2012 Aug 20. PMID: 22908143; PMCID: PMC3470643.

[26] Störk S, Handrock R, Jacob J, et al. Treatment of chronic heart failure in Germany: a retrospective database study. *Clin Res Cardiol*. 2017;106(11):923-932. doi:10.1007/s00392-017-1138-6

[27] Huffman MD, Berry JD, Ning H, Dyer AR, Garside DB, Cai X, Daviglus ML, Lloyd-Jones DM. Lifetime risk for heart failure among white and black Americans: cardiovascular lifetime risk pooling project. J Am Coll Cardiol. 2013 Apr 9;61(14):1510-7. doi: 10.1016/j.jacc.2013.01.022. PMID: 23500287; PMCID: PMC3618527.

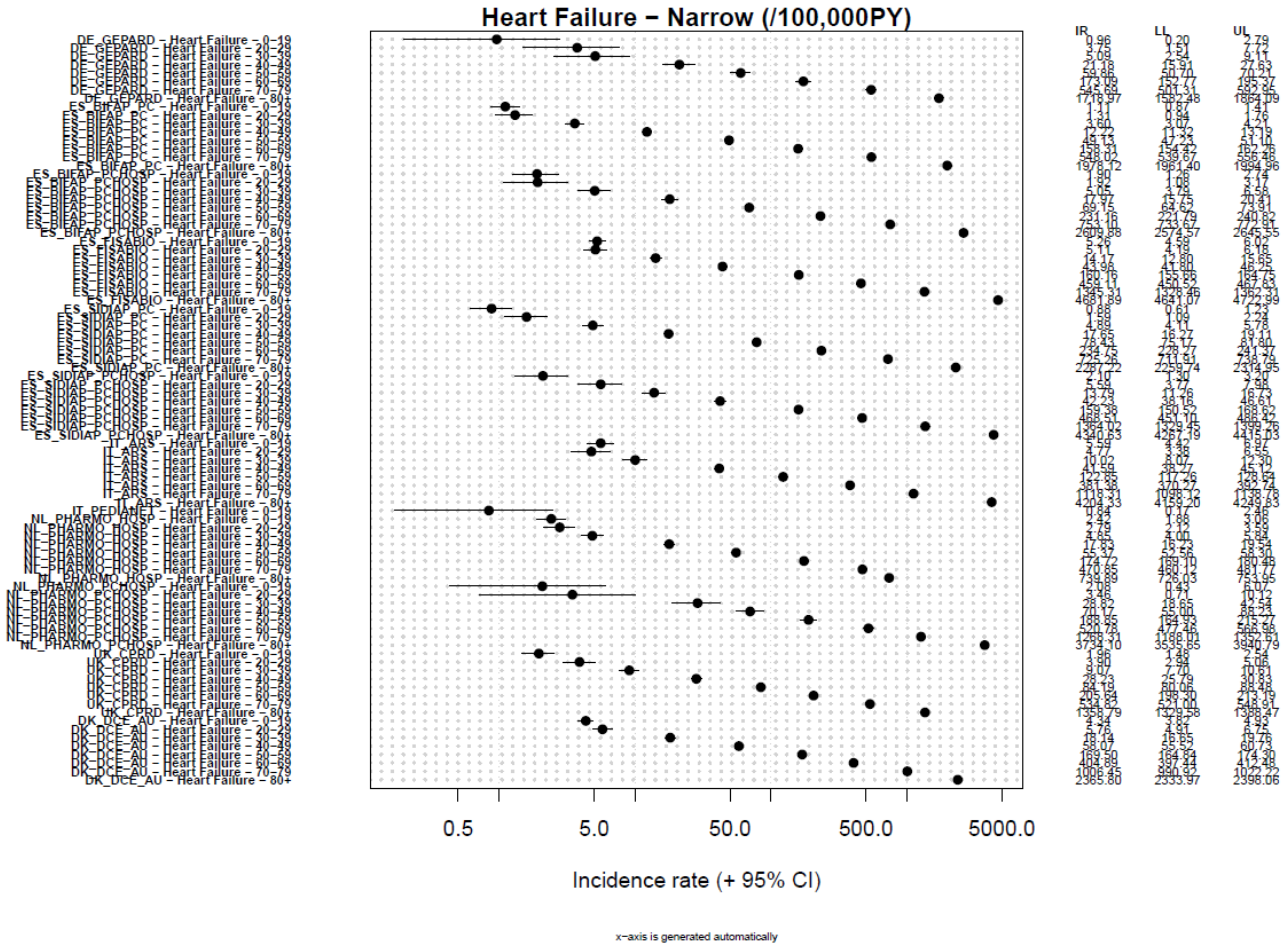


Figure 18: Incidence of heart failure by data source and age

10.3.1.9 Stress cardiomyopathy

Definition and codes in Zenodo
Kelters, I, Willame, C, Belbachir, L, Durán, C, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Stress Cardiomyopathy (1.0). Zenodo. https://doi.org/10.5281/zenodo.5226504

Takotsubo syndrome is a stress cardiomyopathy. The diagnosis of stress cardiomyopathy is difficult because of its clinical phenotype may closely resemble AMI regarding ECG abnormalities and biomarkers. Two additional features that are helpful in distinguishing TTS from acute MI are QTc prolongation > 500 ms during the acute phase and the recovery of LV function over 2 – 4 weeks.

Events of stress cardiomyopathy using narrow definition were observed in ARS, BIFAP, SIDIAP and FISABIO (figure 19). Higher rates were observed in both databases including inpatients. IRs were stable overtime while a small decrease was observed for the year 2020, potentially because of the lockdown and decreased health care access. A clear pattern of increased rates with age was observed in the 4 databases. There is a READ code but this is used very infrequently which explains why there was no event in CPRD. For Spain we recommend the use of rates from BIFAP_PCHOSP, FISABIO and SIDIAP which capture both primary and inpatient diagnoses.

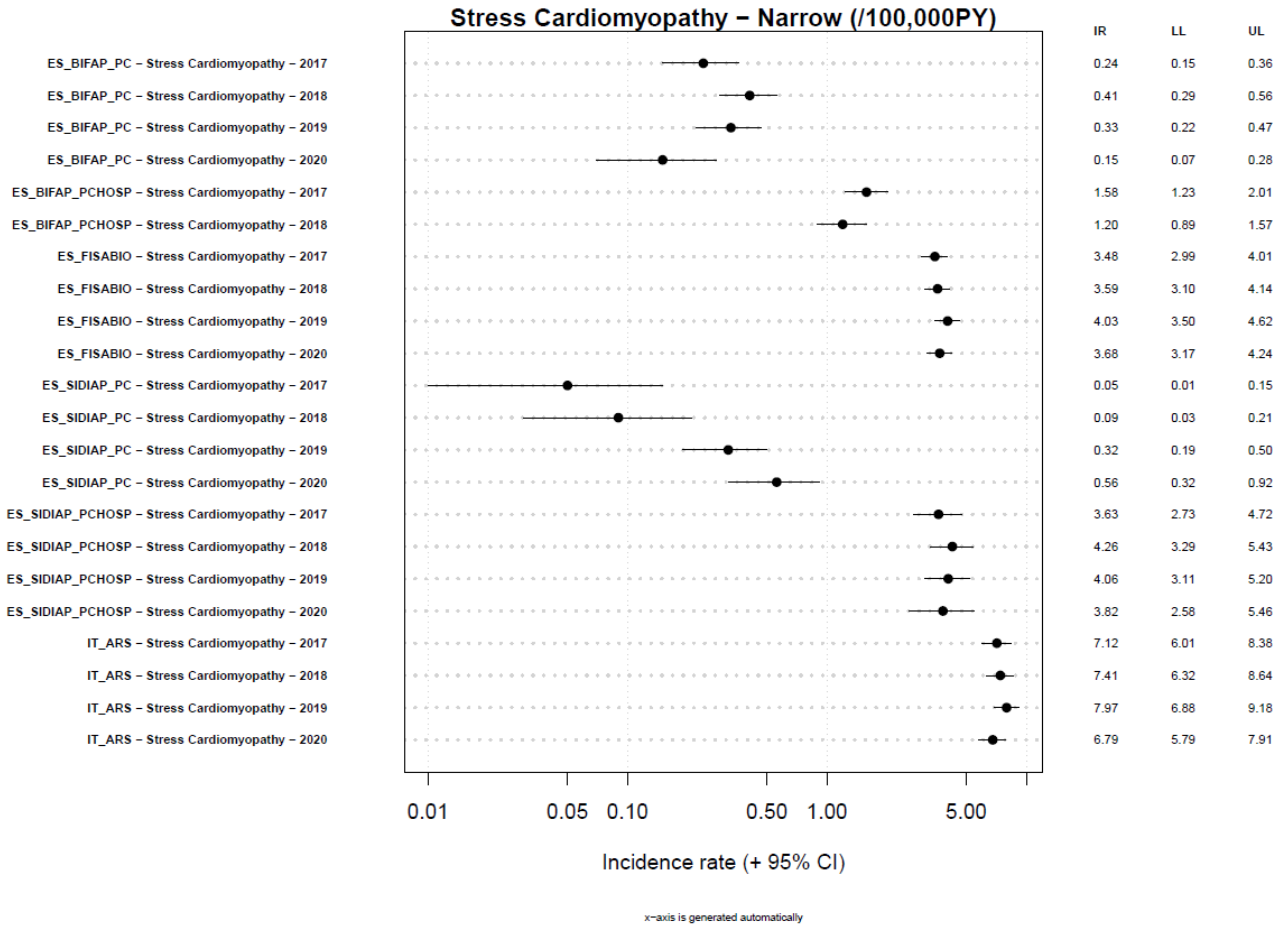


Figure 19: Incidence of stress cardiomyopathy by data source and calendar year

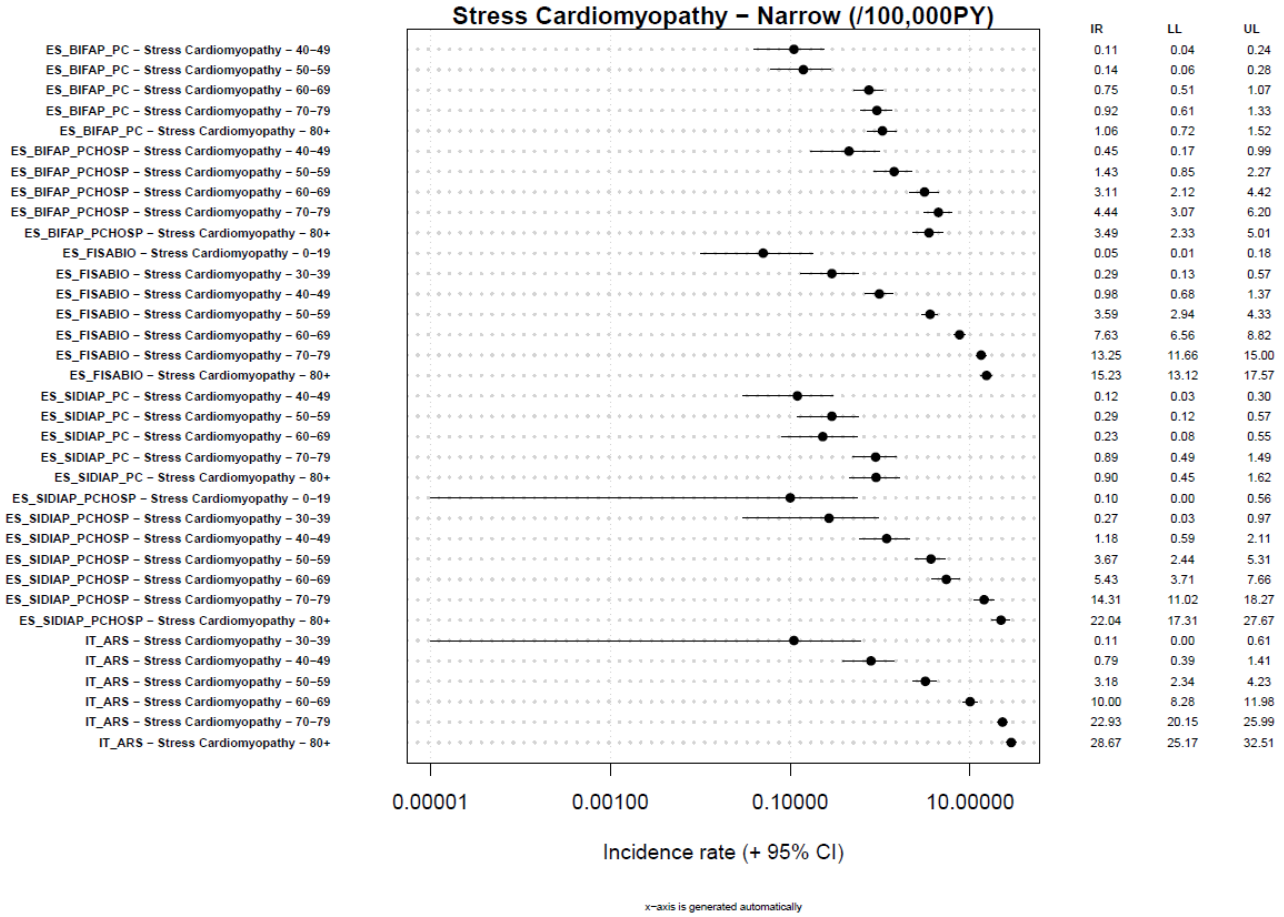


Figure 20: Incidence of stress cardiomyopathy by data source and age

There are not so many published incidence data, as the disease is only recently recognized. Minhas et al. reported on a significant increase in the incidence of takotsubo cardiomyopathy from 2006 to 2012[28]. In that study, the incidence of increased almost 20-fold over the time-period. Similarly, a study by Murugiah et al. showed that hospitalization rates for stress cardiomyopathy are increasing. In that study, the incidence of primary TS increased from 2.3 hospitalizations per 100,000 person-years in 2007 to 7.1 in 2012[29]. Jabri et al reported an increase of stress cardiomyopathy during the COVID-19 pandemic[30]. The rates we observe are within the range reported by Murugiah et al.

10.3.1.10 Coronary artery disease (CAD)

Definition and codes in Zenodo

[28] Minhas AS, Hughey AB, Kolias TJ. Nationwide trends in reported incidence of takotsubo cardiomyopathy from 2006 to 2012. Am J Cardiol. 2015;116:1128–1131. doi: 10.1016/j.amjcard.2015.06.042

[29] Murugiah K, Wang Y, Desai NR, Spatz ES, Nuti SV, Dreyer RP, et al. Trends in short- and long-term outcomes for takotsubo cardiomyopathy among medicare fee-for-service beneficiaries, 2007 to 2012. JACC Heart Fail. 2016;4:197–205. doi: 10.1016/j.jchf.2015.09.013.

[30] Jabri A, Kalra A, Kumar A, et al. Incidence of Stress Cardiomyopathy During the Coronavirus Disease 2019 Pandemic. JAMA Netw Open. 2020;3(7):e2014780. doi:10.1001/jamanetworkopen.2020.14780

Kelters, I, Willame, C, Durán, C, Belbachir, L, Martín-Pérez, M, García-Poza, P, Souverein, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Coronary Artery Disease (1.0). Zenodo. https://doi.org/10.5281/zenodo.5226602

Coronary artery disease or ischemic heart disease describes a set of clinical symptoms caused by an inadequate blood supply to the myocardium. This pathological process is characterized by atherosclerotic plaque accumulation in the epicardial arteries, whether obstructive or non-obstructive. Our narrow code set focus on proof of obstruction, the broader code set include also cardiovascular disease. We present both rates in the graphics.
Events of coronary artery diseases using narrow definition were observed in all databases, except PEDIANET. IRs were stable overtime while a small decrease was observed for the year 2020, most likely due to the lockdown (figure 21). IRs differed based on the provenance of the diagnosis it was lowest 95.33/100,000 person-years in PHARMO (primary care records) to 322.04/100,000 person-years in ARS (discharge diagnoses), also in PHARMO-PC-HOSP the rate was much higher than the PHARMO-PC. A clear pattern of increasing rates with age was observed in all databases.

The recently published article from the European Society of Cardiology concluded on an IRs of coronary artery diseases of 176.3/100,000 person-years (95%CI: 150-238) (Atlas Writing Group, 2020).

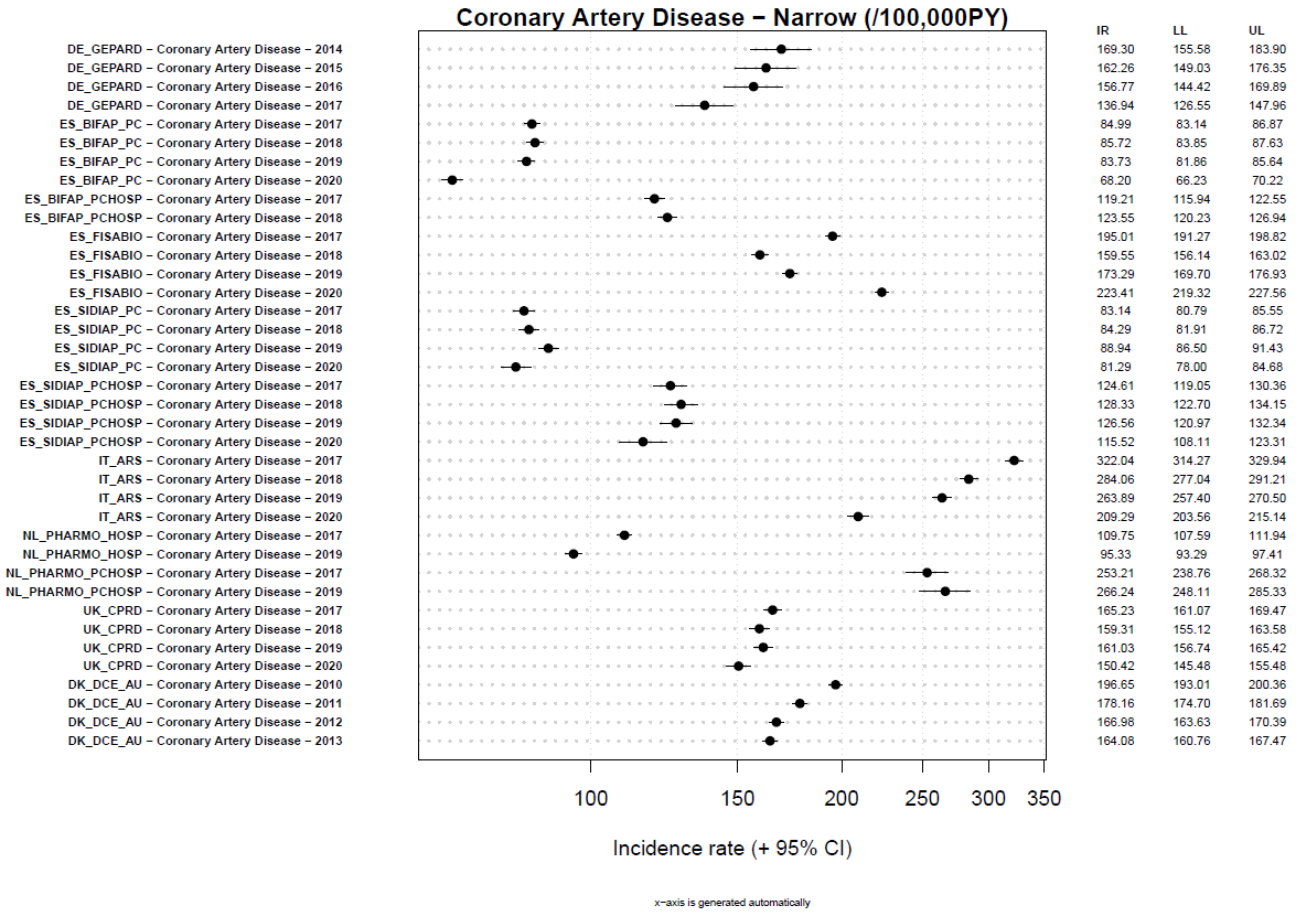


Figure 21 : Incidence of CAD (narrow) by data source and calendar year

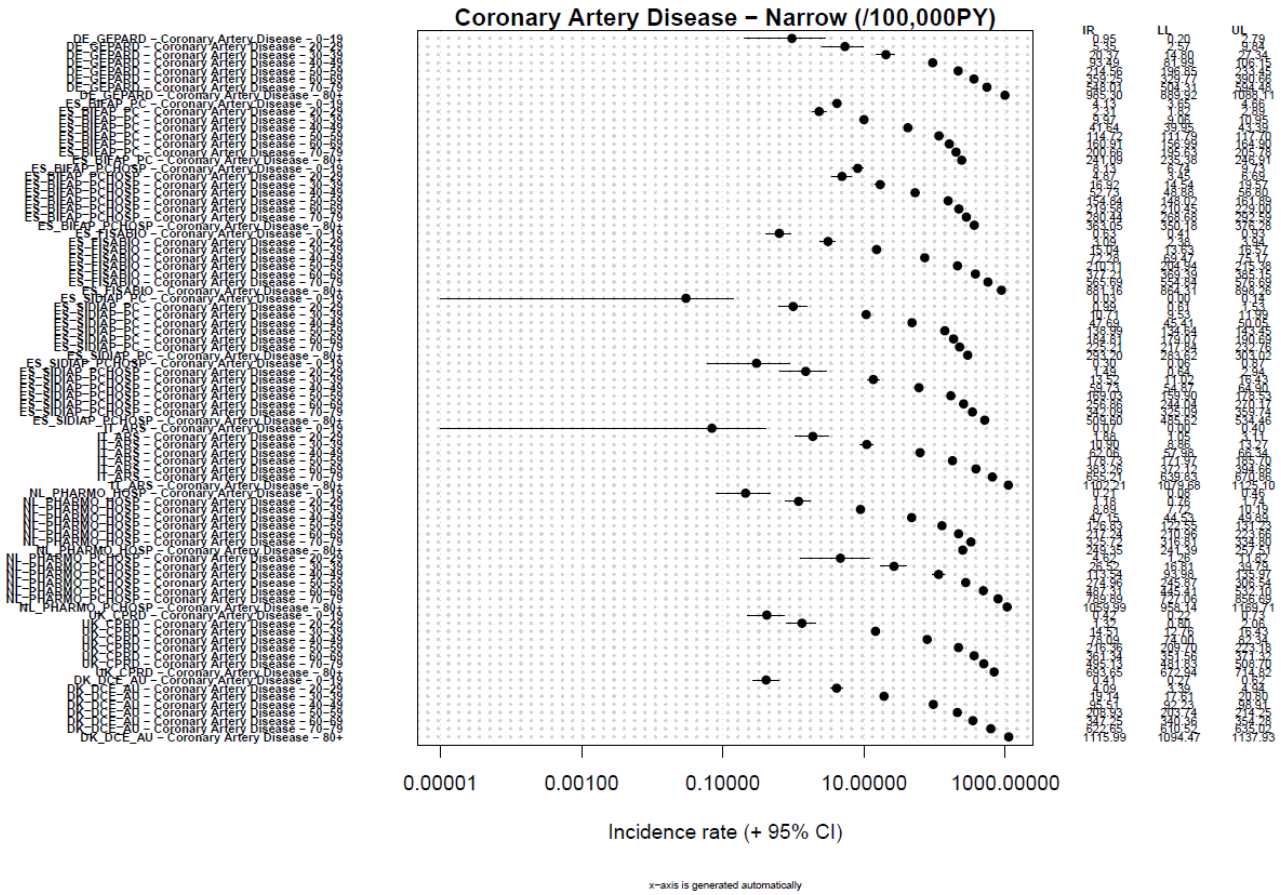


Figure 22: Incidence of CAD (narrow) by data source and age

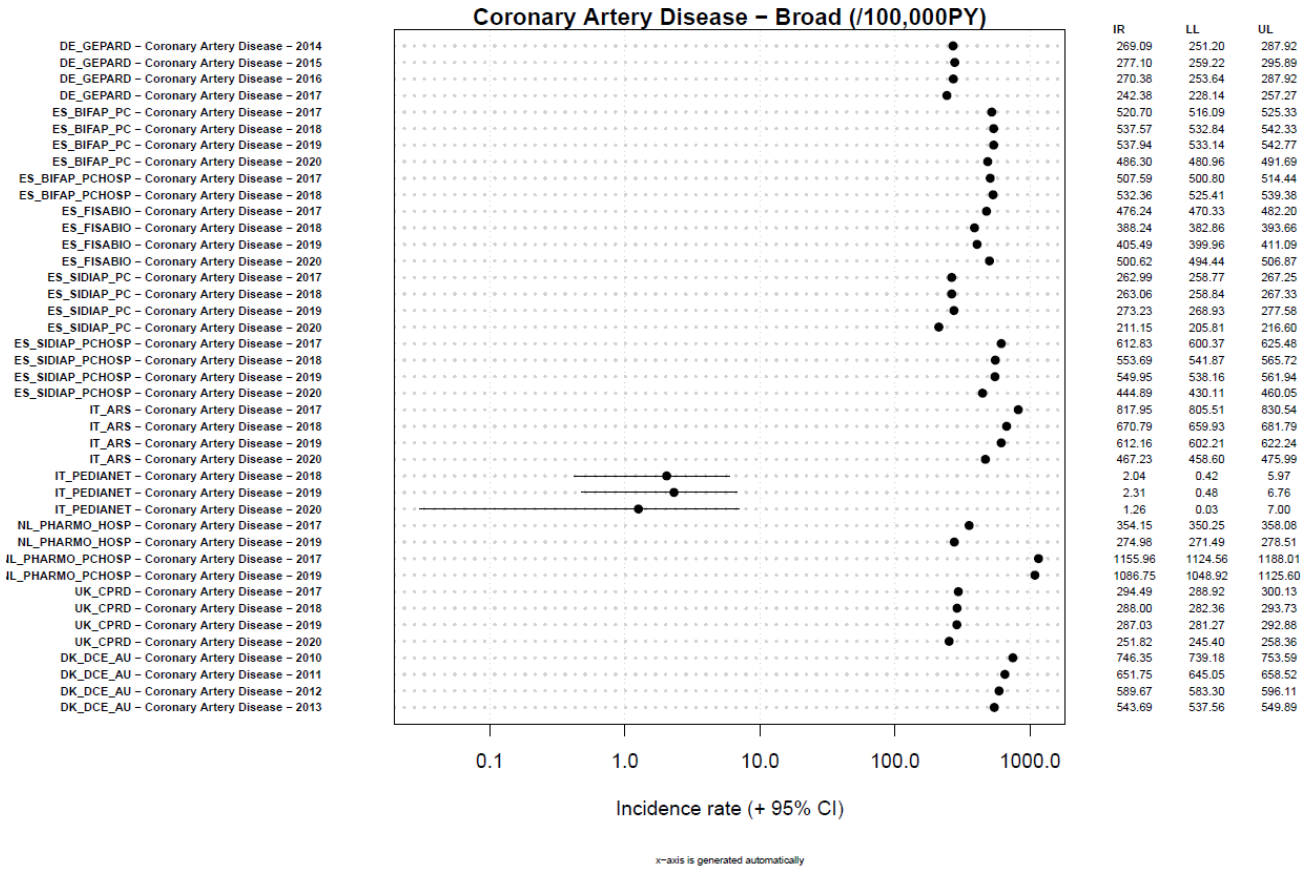


Figure 23: Incidence of CAD (broad) by data source and age

According to the global burden study the incidence of cardiovascular disease in Europe ranges between 600 and 1600 per 100,000 person-years, this is consistent with the broader definition of CAD[31].

[31] *Data source*: Global Burden of Disease Study 2017, Institute for Health Metrics and Evaluation, http://ghdx.healthdata.org/gbd-results-tool

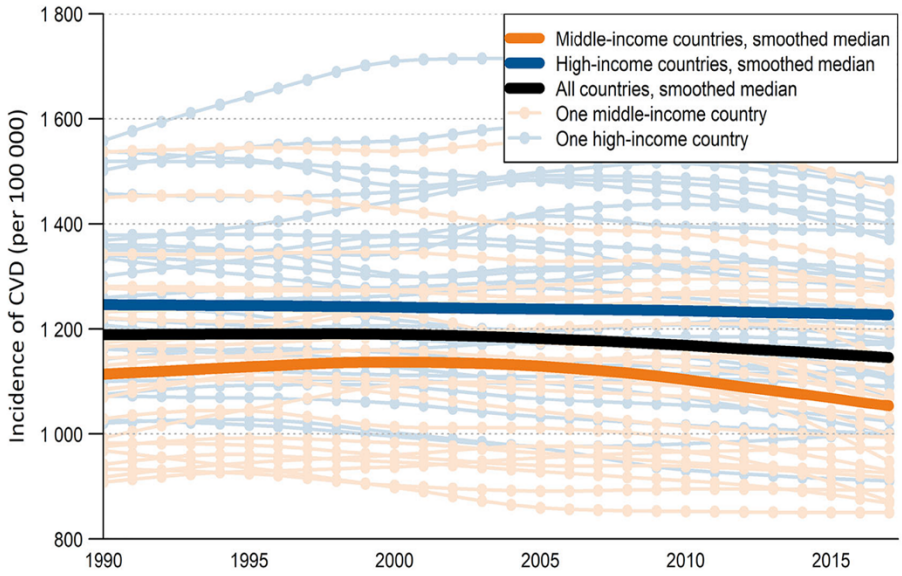


Figure 24: Incidence of cardiovascular disease according to Global burden of disease study

10.3.1.11 Arrhythmia

Definition and codes in Zenodo
Engelen, R, Willame, C, Durán, C, Belbachir, L, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Arrhythmia (1.0). Zenodo. https://doi.org/10.5281/zenodo.5226644

A cardiac arrhythmia is an abnormality or perturbation in the normal activation or beating of the heart myocardium. There are different types of cardiac arrhythmias and they can be classified by the origin in the heart of the arrhythmia: ventricular or supraventricular or whether there is an increase or decrease in the heart rate: tachycardia or bradycardia. In this study we consider all together, all codes enter in the narrow definition, there is no broad definition. In the next run we will classify tachycardia as possible, as it is symptomatic and may fit better with a broad definition.

IRs were quite stable over time while a significant decrease was observed for the year 2020 (figure 25). Datasources with hospital and outpatient/GP data had higher rates than datasources with only GP-based data. Variable rates were observed across years in PHARMO, with a high peak in 2018, which requires investigation as it is also seen in other conditions.

From published articles, IRs of arrhythmia ranges between 208/100,000 in Denmark up to 780/100,000 person-years in UK, which is consistent with our rates

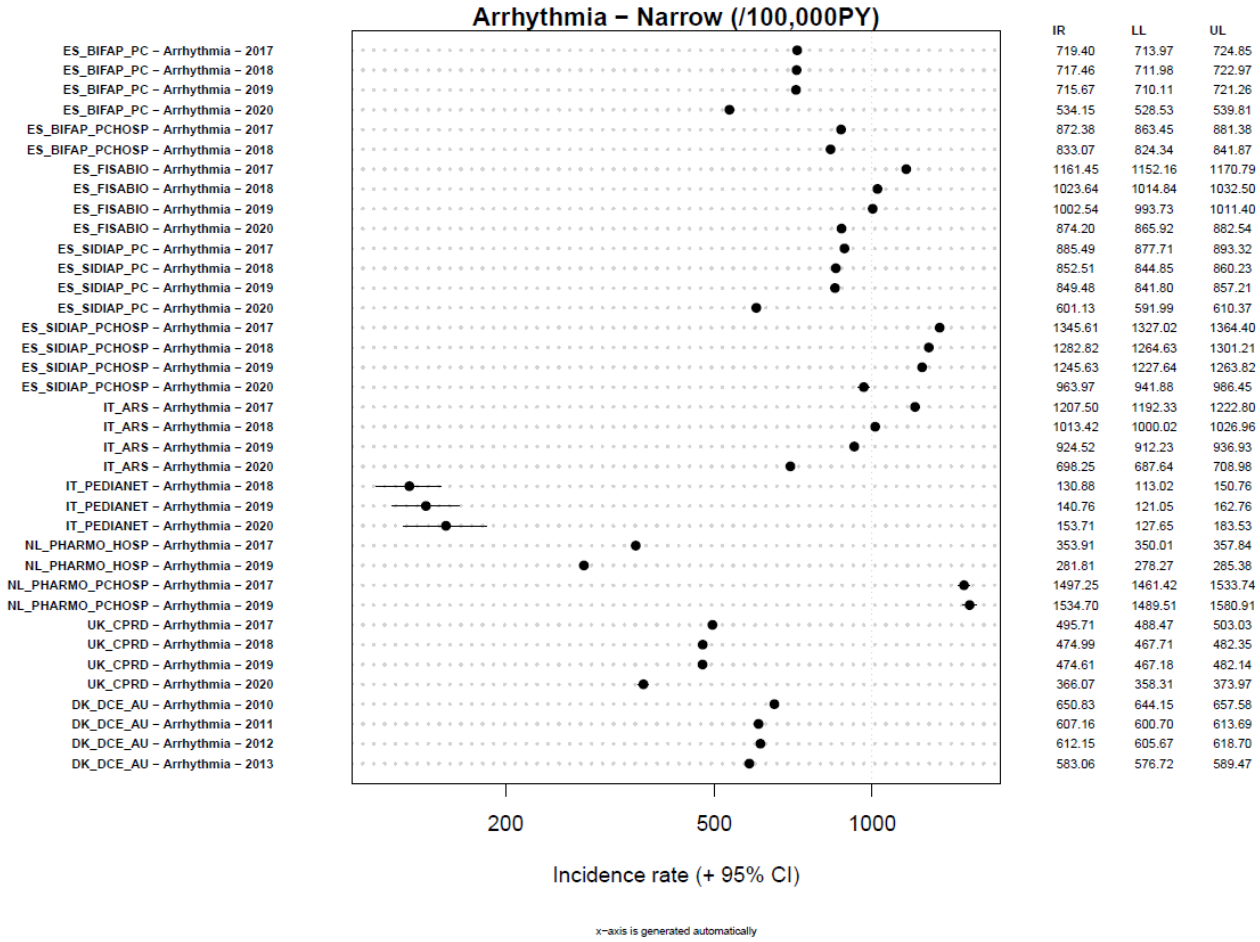


Figure 25: Incidence of arrhythmia by data source and calendar year

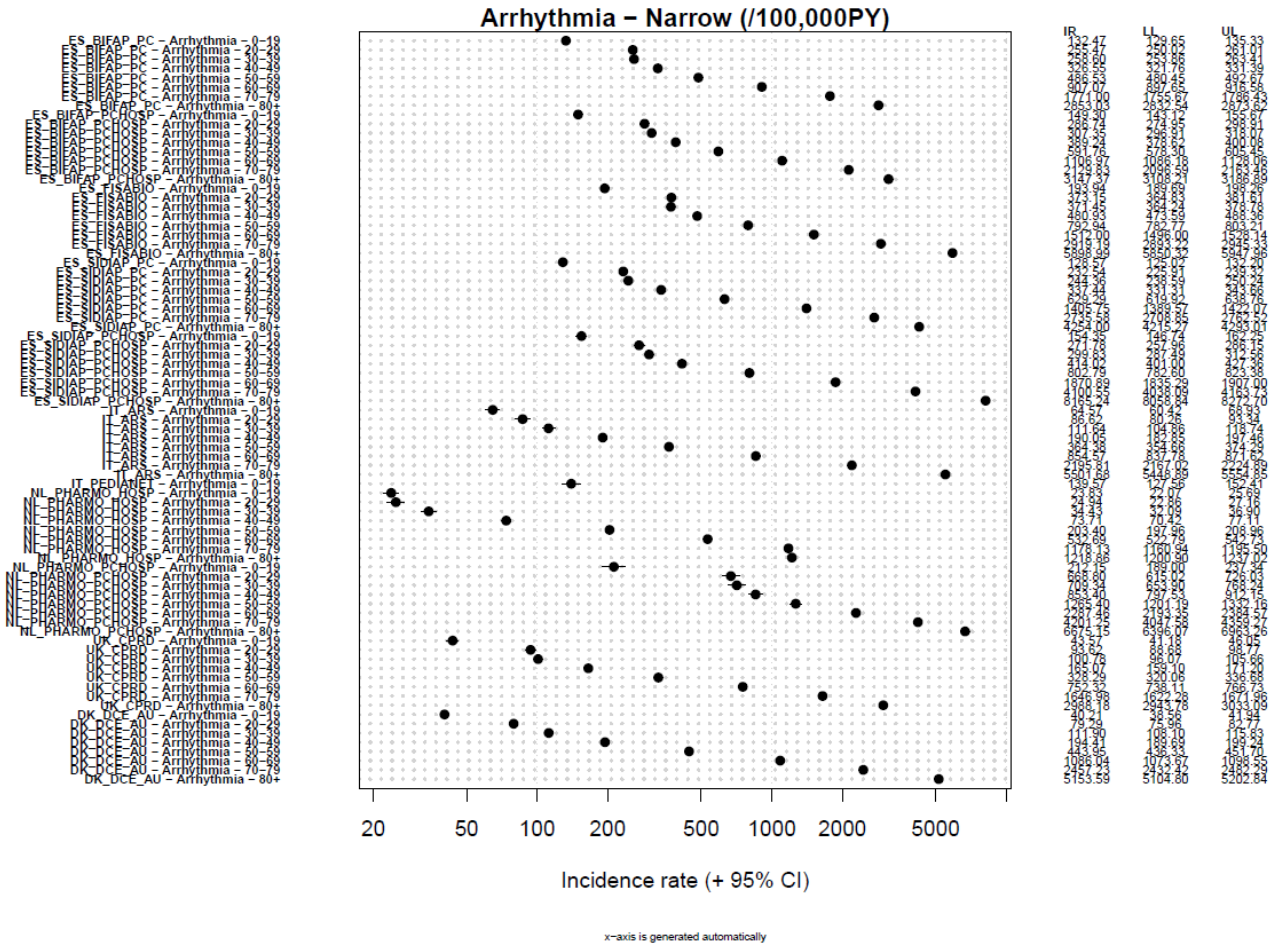


Figure 26: Incidence of arrhythmia by data source and age

Reference data can be obtained from the UK Biobank publication which reported rates of overall arrhythmia (figure 28):

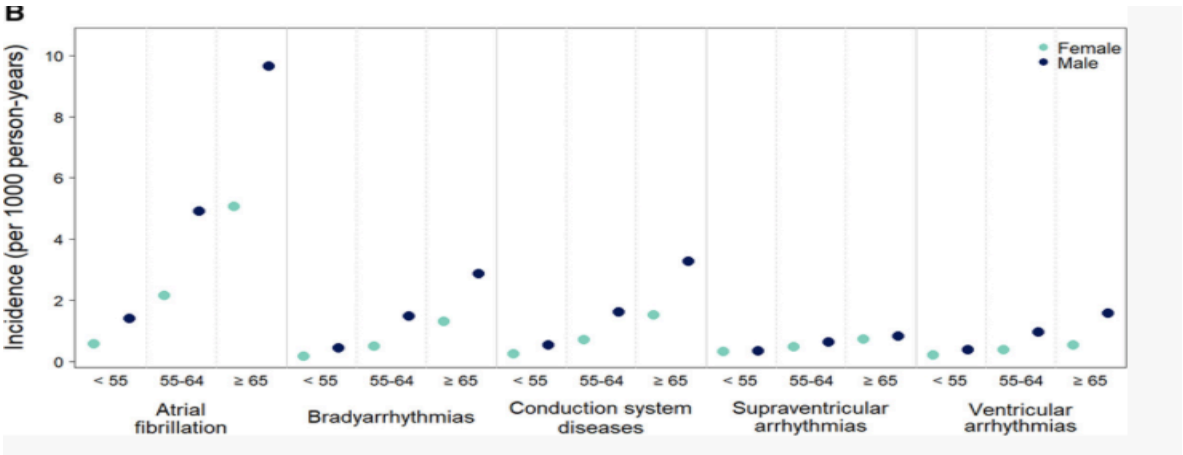


Figure 27: Incidence of arrhythmia in the UK Biobank [32]

The overall rates in males were 242/100,000 in males <55 years of age, 739/100,000 for 55-64 years of age and 1370/100,000 for > 65 years. Rates for women in these age categories were 117, 342 and 729 respectively. This is aligned with our observations.

10.3.1.12 Myocarditis/pericarditis

Definition and codes in Zenodo
Sturkenboom, MCJM, Willame, C, Belbachir, L, & Duran, C. (2021). ACCESS-Background rate of adverse events-definition –Myocarditis and/or pericarditis. Zenodo. https://doi.org/10.5281/zenodo.5172798

Myocarditis is an inflammatory disease of the myocardium caused by different infectious (viral and non-viral) and non-infectious triggers (autoimmune diseases, hypersensitivity reactions to drugs, toxic reactions to drugs, toxics, etc.) Pericarditis is the inflammation of the pericardium from various origins, such as infection, neoplasm, autoimmune process, injuries, or drug-induced. Pericarditis usually leads to pericardial effusion, or constrictive pericarditis.

10.3.1.12.1 Myocarditis alone

The clinical definition of myocarditis contained only medical codes specific for this condition. Rates of myocarditis alone have been generated and were much lower when compared to myocarditis/pericarditis suggesting that the composite endpoint myocarditis/pericarditis was much driven by pericarditis clinical events. Myocarditis is more frequent in younger age groups, then decreases but increases again in higher age groups. Gubernot reported incidence rates of myocarditis alone between 1-10/100,000 PY which is consistent with our data. PHARMO 2018 rates seem out of range and should be discarded.

[32] Khurshid S, Choi SH, Weng LC, Wang EY, Trinquart L, Benjamin EJ, Ellinor PT, Lubitz SA. Frequency of Cardiac Rhythm Abnormalities in a Half Million Adults. Circ Arrhythm Electrophysiol. 2018 Jul;11(7):e006273. doi: 10.1161/CIRCEP.118.006273. PMID: 29954742; PMCID: PMC6051725.

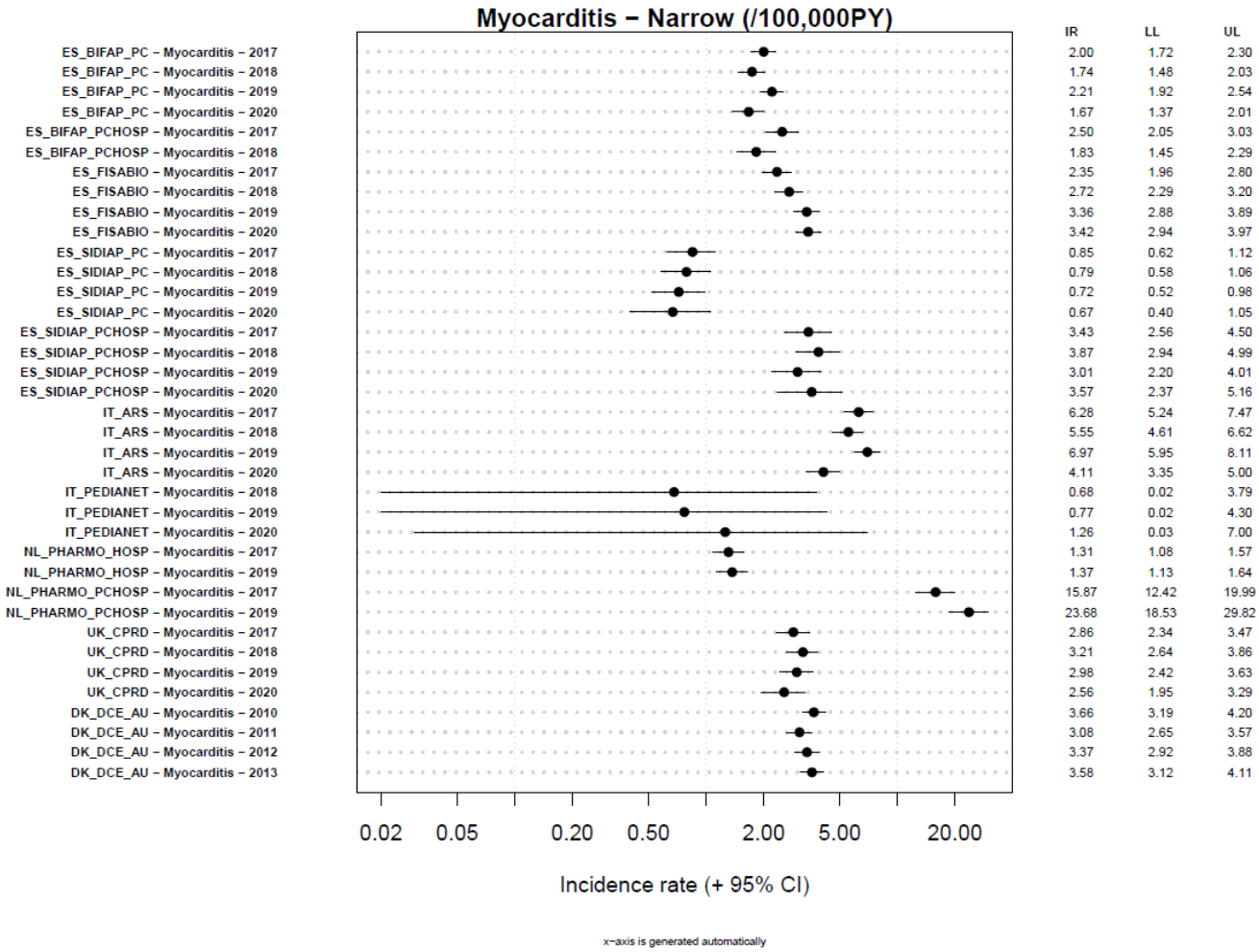


Figure 28 Incidence of myocarditis alone by data source and calendar year

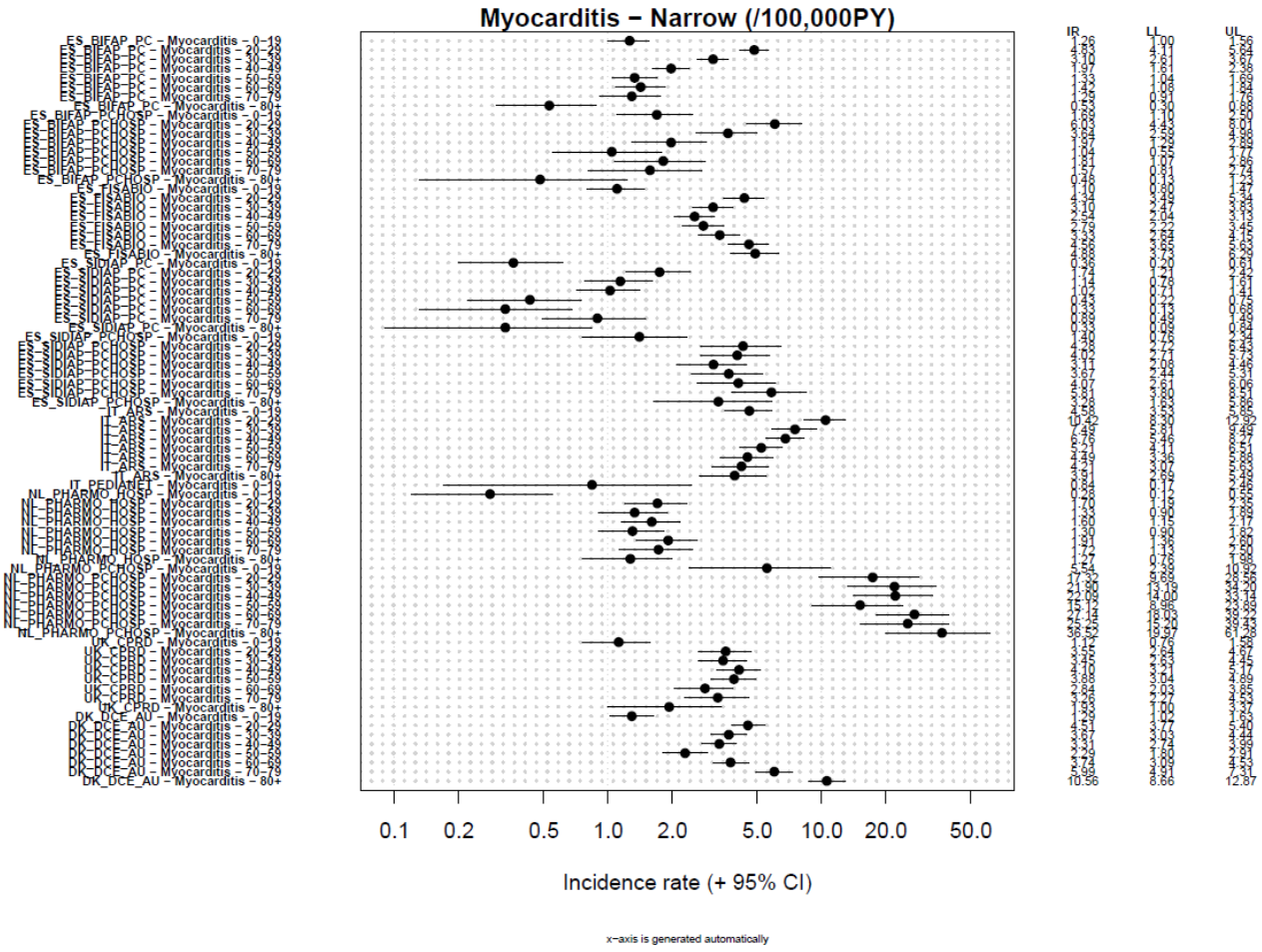


Figure 29 Incidence of myocarditis alone by data source and age

10.3.1.12.2 Myocarditis or pericarditis

Rates for myocarditis/pericarditis have been generated using a narrow and a broad medical definition. IRs for myocarditis/pericarditis were quite stable overtime in each of the data sources (figure 30) with a slight decrease in 2020. Rates increased with age. Li et al. Reported rates of myocarditis/pericarditis ranging from 6-57 across age/gender, consistent with our rates.

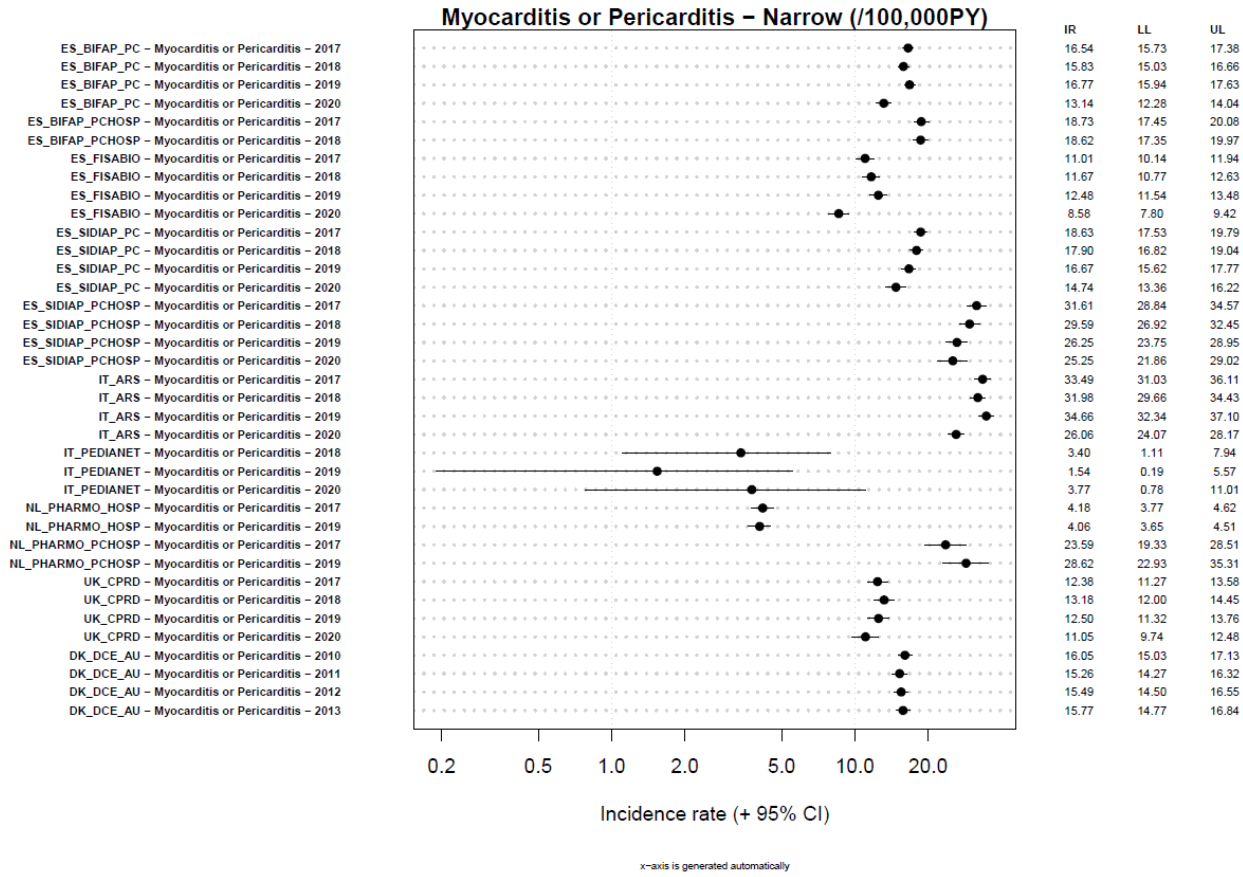


Figure 30 Incidence of myocarditis and pericarditis by data source and calendar year

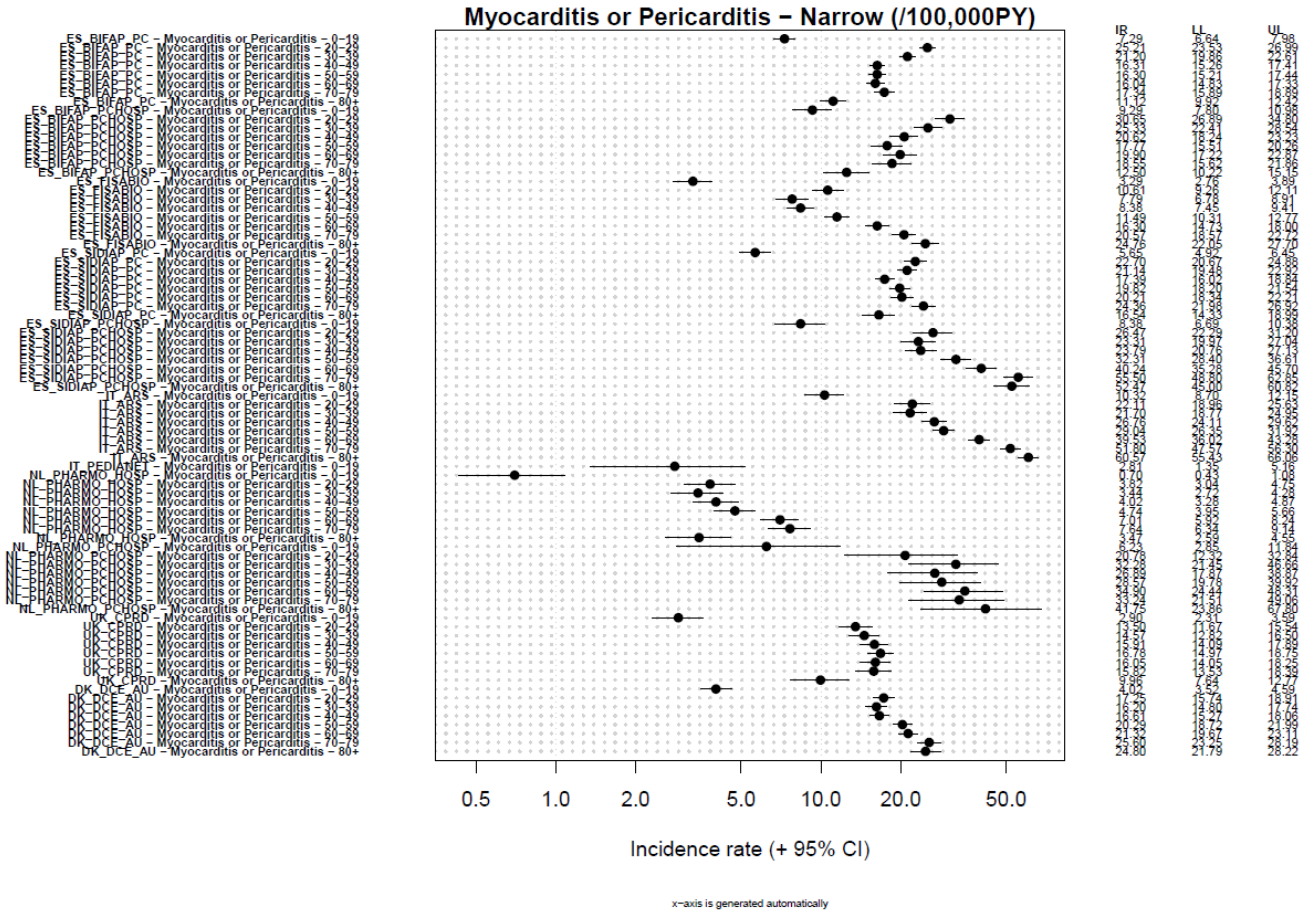


Figure 31 Incidence of myocarditis and pericarditis by data source and age

10.3.1.13 Coagulation disorders

Definition and codes in Zenodo
Egbers, T, Belbachir, L, Durán, C, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Coagulation disorders (1.2). Zenodo. https://doi.org/10.5281/zenodo.5228687

A coagulation disorder is a problem with blood clotting. Blood clotting usually occurs when there is damage to a blood vessel. Platelets immediately adhere to the cut edges of the vessel and release chemicals to attract even more platelets. A platelet plug is formed and the external bleeding stops.

Coagulation disorder can either be too much clotting leading to thrombosis, emboli or ischemic stroke, or too little clotting leading to bleeding and hemorrhagic stroke. Coagulation disorders were classified in 6 subtypes:

- Disseminated intravascular coagulation,
- Thrombotic Thrombocytopenic Purpura or Thrombotic microangiopathy,
- Venous Thromboembolism including Pulmonary embolism and Deep Vein Thrombosis,
- Cerebral Venous Thrombosis,
- Ischemic stroke,
- Hemorrhagic stroke

10.3.1.13.1 Disseminated intravascular coagulation (DIC)

Disseminated intravascular coagulation is a syndrome that may develop in the course of various clinical conditions. DIC is a result of generalized activation of coagulation with a concomitant activation or inhibition of fibrinolysis. It can be acute such as unexplained thrombocytopenia or chronic due to a mild to moderate platelet count reduction. Limited evidence is available on background incidence rates for this condition. The paper from Singh et al. (2010) reports overall estimate of 18.6/100,000 person-years in 2010[33]. The incidence rate of DIC was shown to increase with age in both men and women and was consistently higher in men. Li et al. reported incidence rates between 2 and 24 with increasing age which is a similar magnitude but higher than we observed.

IRs were stable overtime while a small decrease was observed for the year 2020, potentially because of the lockdown and decreased health care access. A clear pattern of increased rates with age was observed in the databases, except for CPRD and SIDIAP.

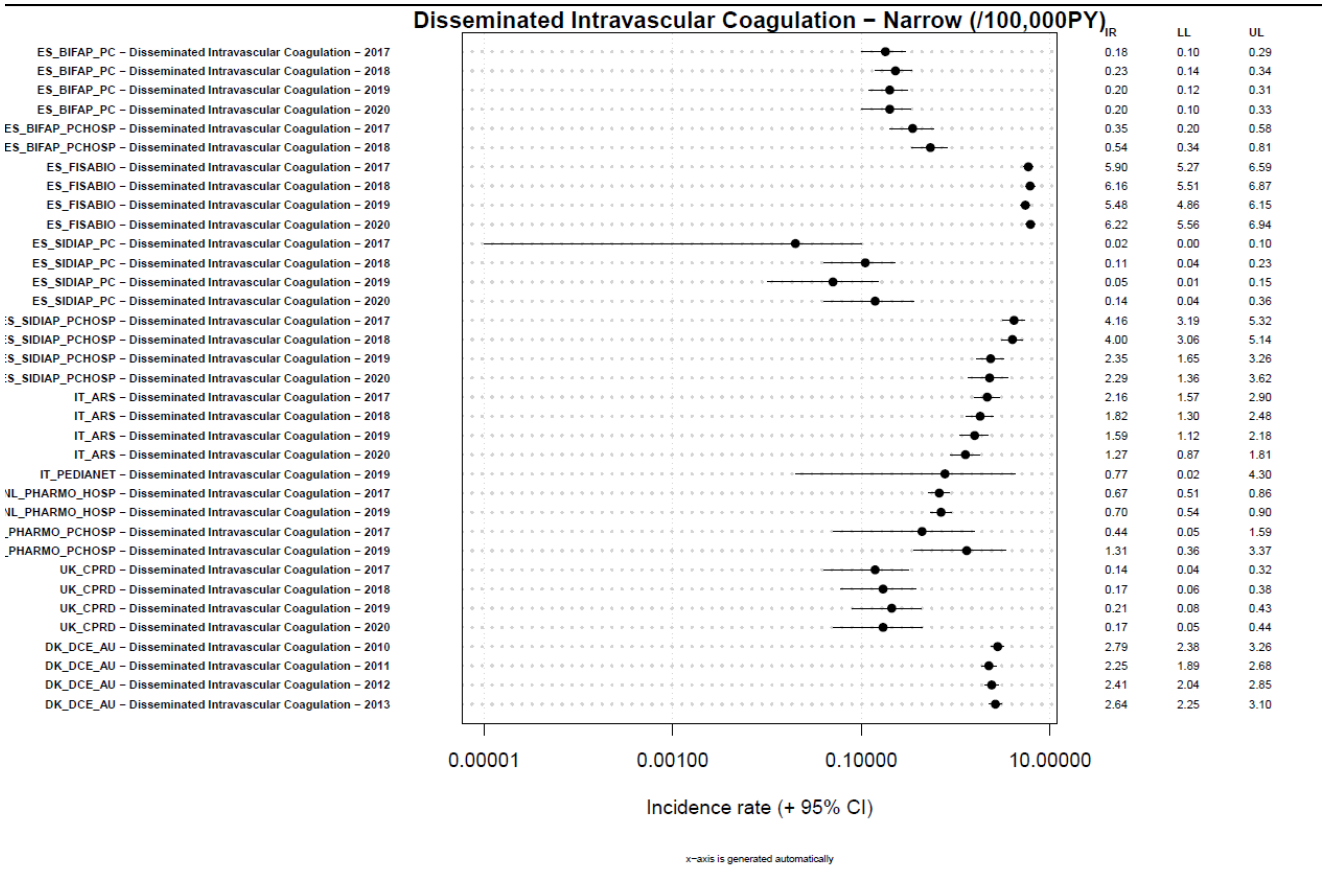


Figure 32 Incidence of DIC by data source and calendar year

[33] Singh B, Hanson AC, Alhurani R, Wang S, Herasevich V, Cartin-Ceba R, Kor DJ, Gangat N, Li G. Trends in the incidence and outcomes of disseminated intravascular coagulation in critically ill patients (2004-2010): a population-based study. Chest. 2013 May;143(5):1235-1242. doi: 10.1378/chest.12-2112. PMID: 23139140.

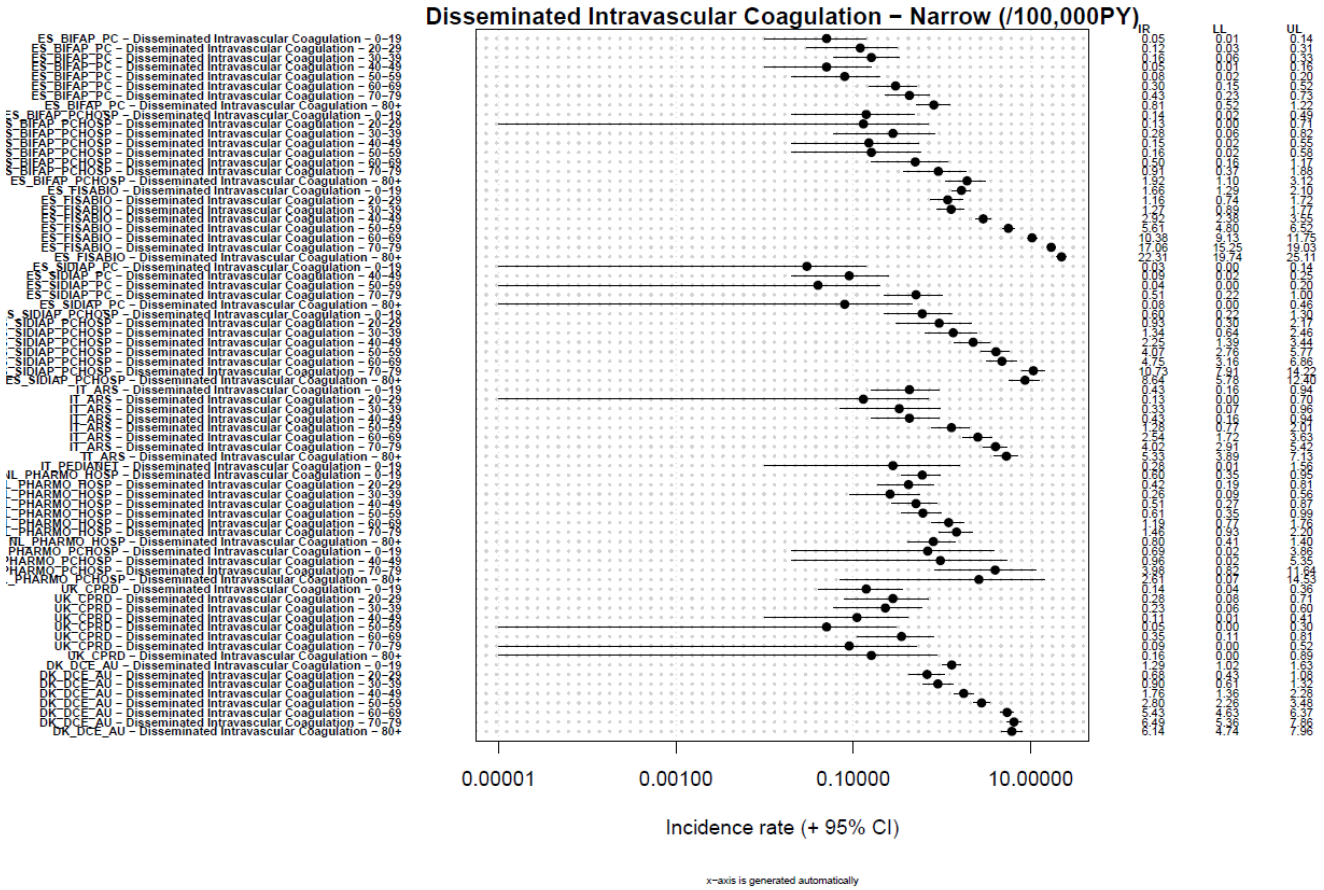


Figure 33 Incidence of DIC by data source and age

10.3.1.13.2 Thrombotic microangiopathy or thrombotic Thrombocytopenic Purpura (TTP)

Thrombotic thrombocytopenic purpura (TTP) is a thrombotic microangiopathy characterized by thrombocytopenia, schistocytic anemia, neurologic impairment, renal impairment, and fever. Thrombocytopenia is caused by the formation of intravascular platelet aggregates, which develop due to endothelial injury and the presence of ultra large von Willebrand factor (vWF) molecules in plasma. TTP is more prevalent in women, usually between the ages of 30 to 40 years. Acquired TTP is an ultra-orphan disease with an annual incidence between 1.5 and 6.0 cases per million and mainly affecting young and healthy adults aged 40 years on average[34].

IRs were stable overtime with a small decrease for the year 2020, potentially because of the lockdown and decreased health care access. A clear pattern of increased rates with age was observed in the databases. There no estimates of incidence of TTP by OHDSI nor VSD. From the literature we know that Thrombotic thrombocytopenic purpura (TTP) is a rare and often fatal disorder with an estimated incidence of 0.37 cases per 100,000 PY[35] which is similar to our rates.

[34] Miesbach W, Menne J, Bommer M. Incidence of acquired thrombotic thrombocytopenic purpura in Germany: a hospital level study. Orphanet Journal of Rare Diseases. 2019; 14:260. https://doi.org/10.1186/s13023-019-1240-0

[35] Torok TJ, Holman RC, Chorba TL. Increasing mortality from thrombotic thrombocytopenic purpura in the United States: analysis of national mortality data, 1968-1991. *Am J Hematol.*1995;50:84-90.

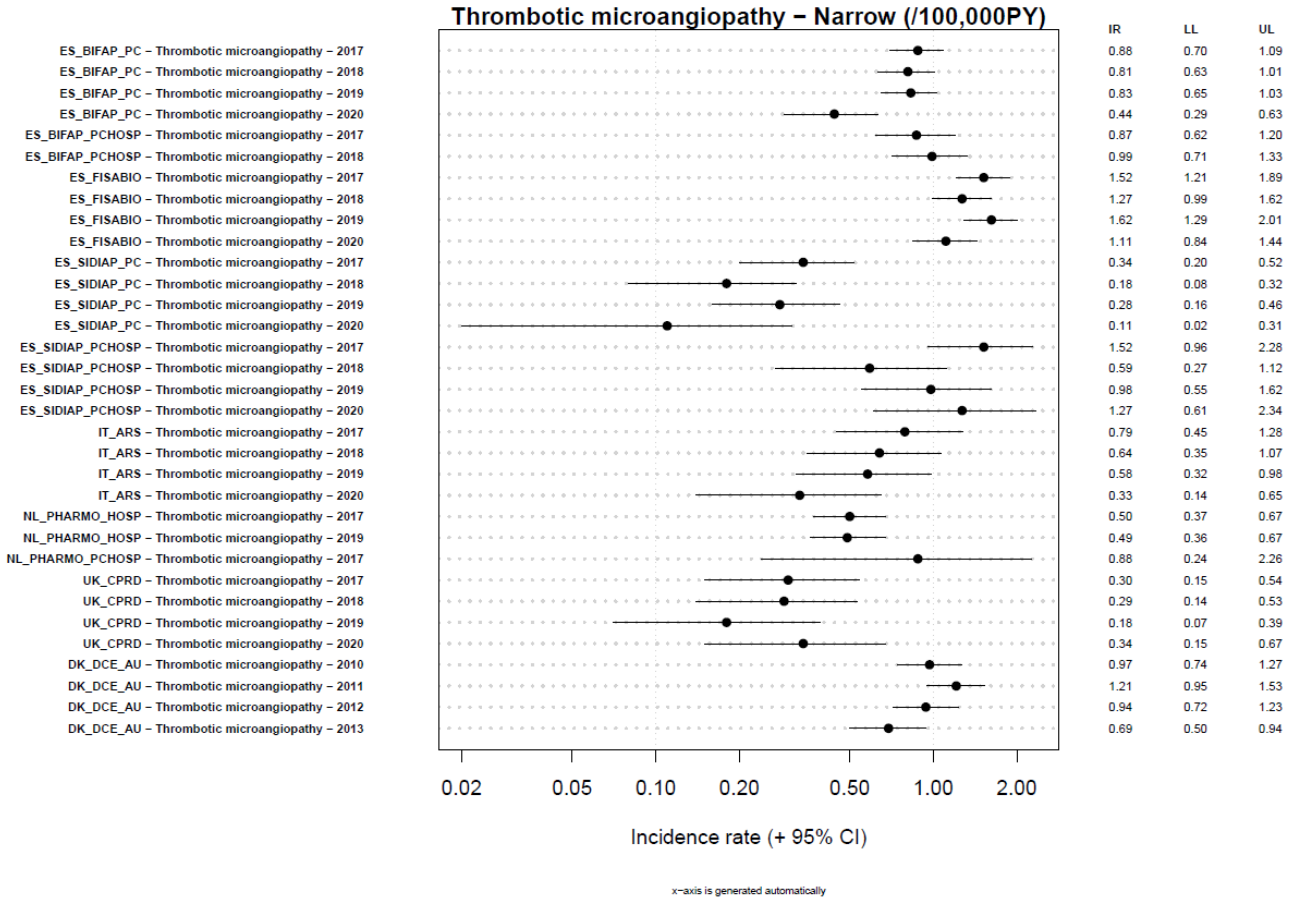


Figure 34 Incidence of Thrombotic microangiopathy by data source and calendar year

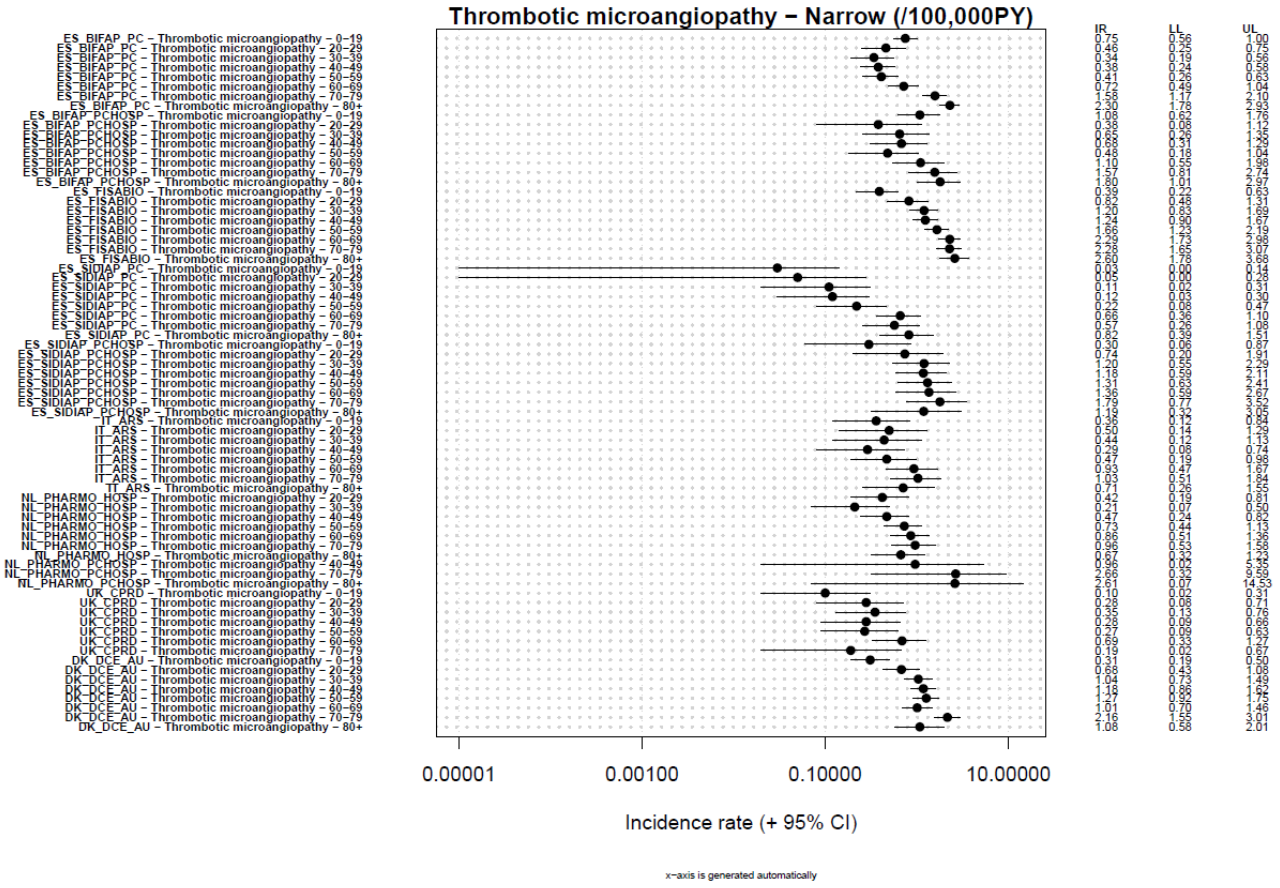


Figure 35 Incidence of TTP by data source and age

10.3.1.13.3 Venous Thromboembolism (VTE)

Venous thromboembolism (VTE), defined as pulmonary embolism (PE) and deep-vein thrombosis (DVT) of the lower limbs, is the third most common cardiovascular illness after acute coronary syndromes[36]. DVT refers to the development of a thrombus in the deep venous system of the lower extremities or, less commonly, the upper extremities. PE refers to the occlusion of the pulmonary artery or some of its branches by an embolus. The embolus may be formed by thrombi which usually originate from deep veins of the lower extremities or the pelvis. Published incidence rates show a strong age-dependent pattern with incidence increasing with age. Rates of DVT were found to vary from 117/100,000 person-years for all types of VTE. Small differences are reported according to DVT only (48/100,000 person-years) or PE with and without DVT (69/100,000 person-years)[37]. Recent data suggests an increase in incidence over time with estimates increasing from 95 to 133/100,000 person-years from 1999 to 2009 [38] (Huang, 2014). In US general population, VTE rates vary from 108 to 167/100,000 person-years (Gubernot et al., 2021).

[36] Gillum RF. Pulmonary embolism and thrombophlebitis in the United States, 1970–1985. Am Heart J 1987; 114: 1262–1264.

[37] Silverstein MD, Heit JA, Mohr DN, et al. Trends in the Incidence of Deep Vein Thrombosis and Pulmonary Embolism A 25-Year Population-Based Study. Arch Intern Med. 1998;158(6):585-593. doi:10.1001/archinte.158.6.585

[38] Huang W, Goldberg RJ, Anderson FA, Kiefe CI, Spencer FA. Secular trends in occurrence of acute venous thromboembolism: the Worcester VTE study (1985-2009). Am J Med. 2014 Sep;127(9):829-39.e5. doi: 10.1016/j.amjmed.2014.03.041. Epub 2014 May 6. PMID: 24813864; PMCID: PMC4161646.

IRs were stable overtime with a small decrease for the year 2020, potentially because of the lockdown and decreased health care access. A clear pattern of increased rates with age was observed in the databases. Pottegard recently reported rates of 158/100,000 PY in Denmark and 126/100,000 PY in Norway. (Pottegard et al., 2021) which is comparable to the Danish rates, but other datasources had higher rates. Data from 2018 in PHARMO PCHOSP seem outlier.

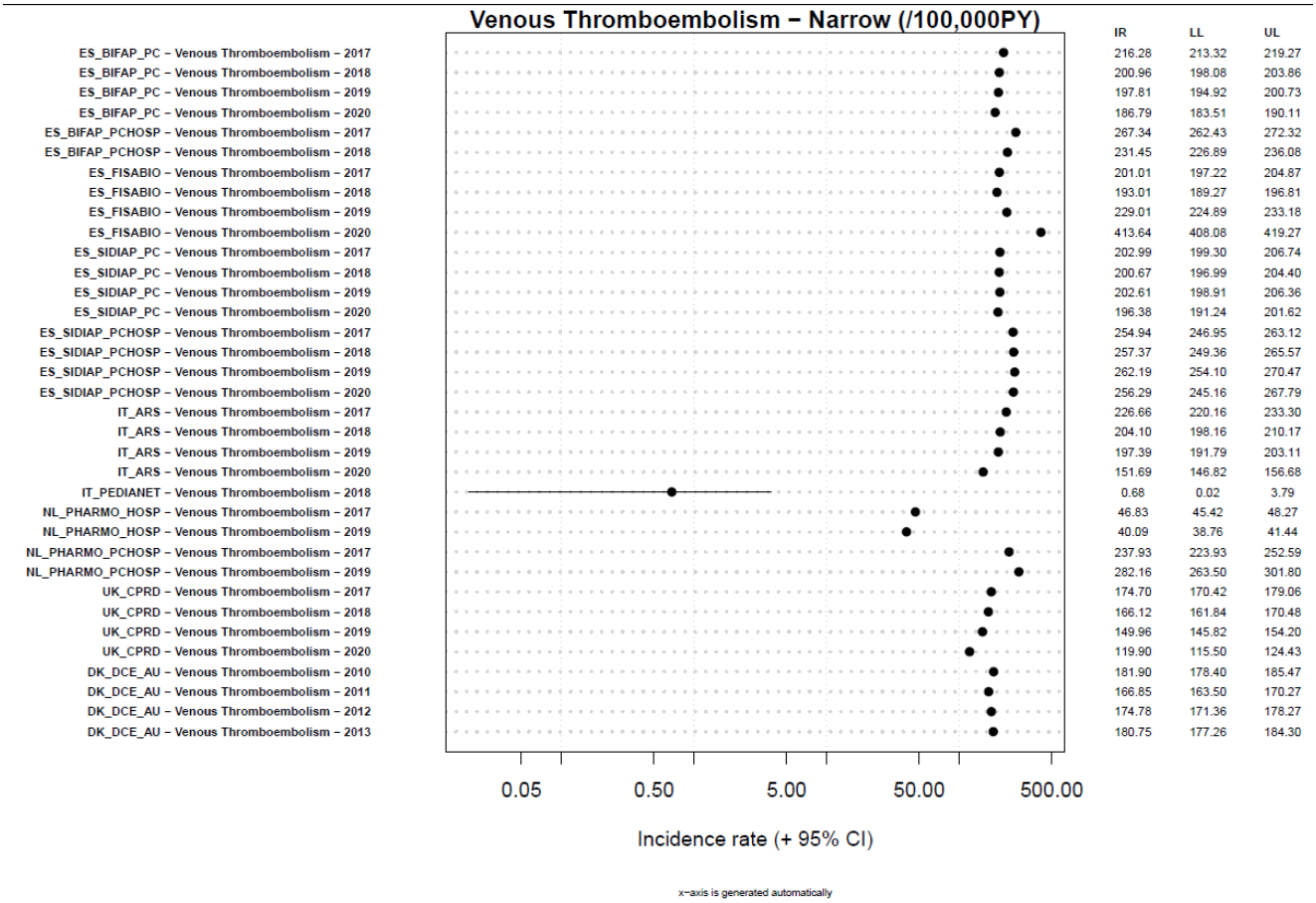


Figure 36 Incidence of VTE by data source and calendar year

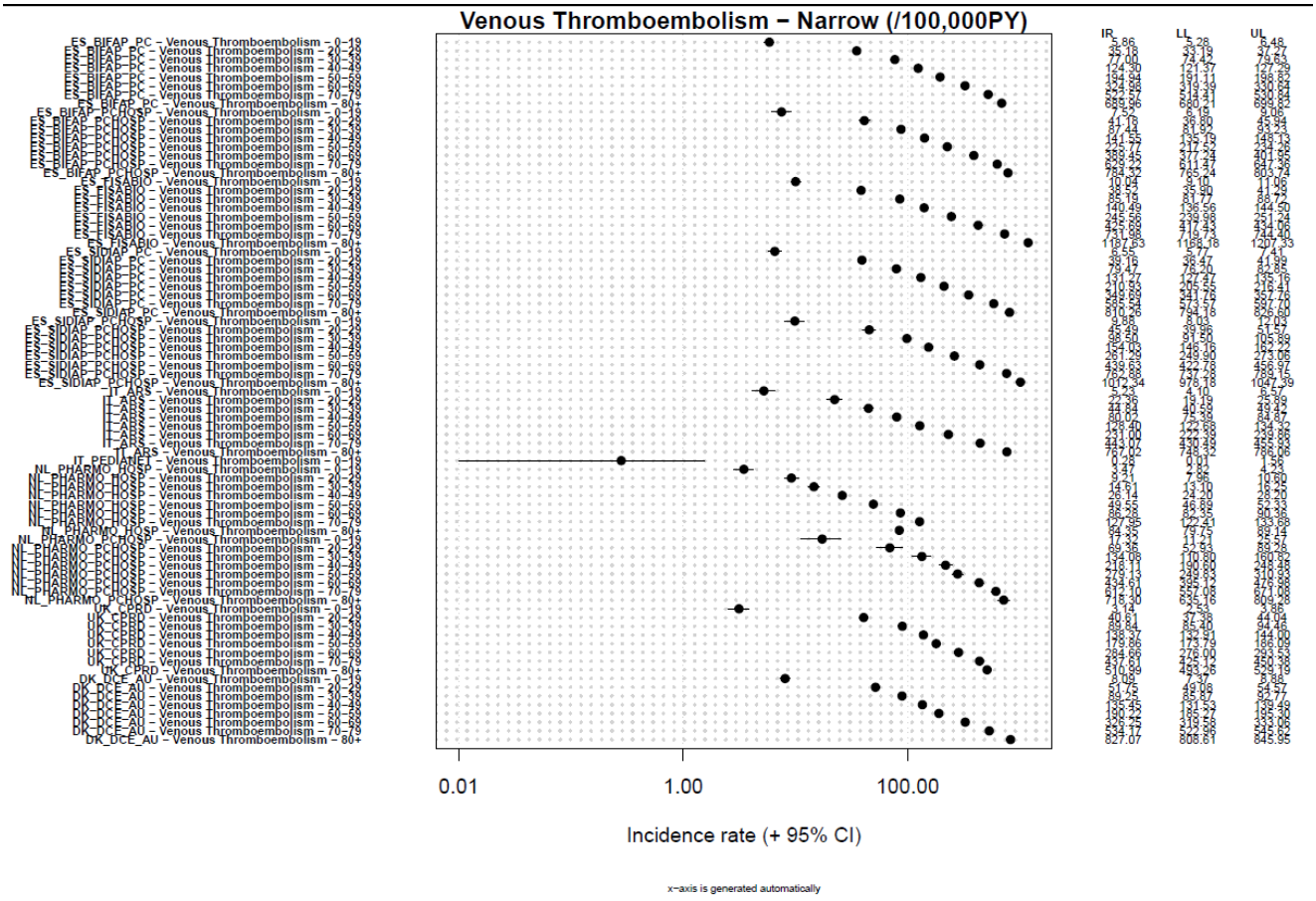


Figure 37 Incidence of VTE by data source and age

10.3.1.13.4 Cerebral Venous Thrombosis

Cerebral vein and cerebral venous sinus thrombosis (CVST) are blood clots that form in the veins that drain the blood from the brain called the sinuses and cerebral veins. CVST is a multifactorial condition with gender-related specific causes, with a wide clinical presentation. CVST has an annual incidence estimated to be two to five cases per million [39]. Two other studies found higher incidence rates than previously reported with annual rates ranging between 13.2 to 15.7 cases per million [40] [41]

Rates of CVST were observed in all databases. Cases were identified in PEDIANET in 2020 only and higher in CPRD for the same year. Higher rates were observed in ARS, database which encompassed inpatient and emergency room. Overall, IRs were stable overtime and no specific age-pattern was observed in the databases. Pottegard reported incidence rates of 2/100,000 PY and 1/100,000 PY in Denmark and Norway, which is aligned with our data. Data based on primary care are consistently lower than data based on hospital diagnoses, which may explain the lower rates reported by Prieto et al. for the DE-disease analyser and French LPD data.

[39] Capecchi M, Abbattista M, Martinelli I. Cerebral venous sinus thrombosis. J Thromb Haemost 2018; 16: 1918–31. https://doi.org/10.1111/jth.14210
[40] Coutinho JM, Zuurbier SM, Aramideh M, Stam J. The incidence of cerebral venous thrombosis: a cross-sectional study. Stroke. 2012 Dec;43(12):3375-7. doi: 10.1161/STROKEAHA.112.671453. Epub 2012 Sep 20. PMID: 22996960.
[41] Devasagayam S, Wyatt B, Leyden J, Kleinig T. Cerebral Venous Sinus Thrombosis Incidence Is Higher Than Previously Thought: A Retrospective Population-Based Study. Stroke. 2016 Sep;47(9):2180-2. doi: 10.1161/STROKEAHA.116.013617. Epub 2016 Jul 19. PMID: 27435401.

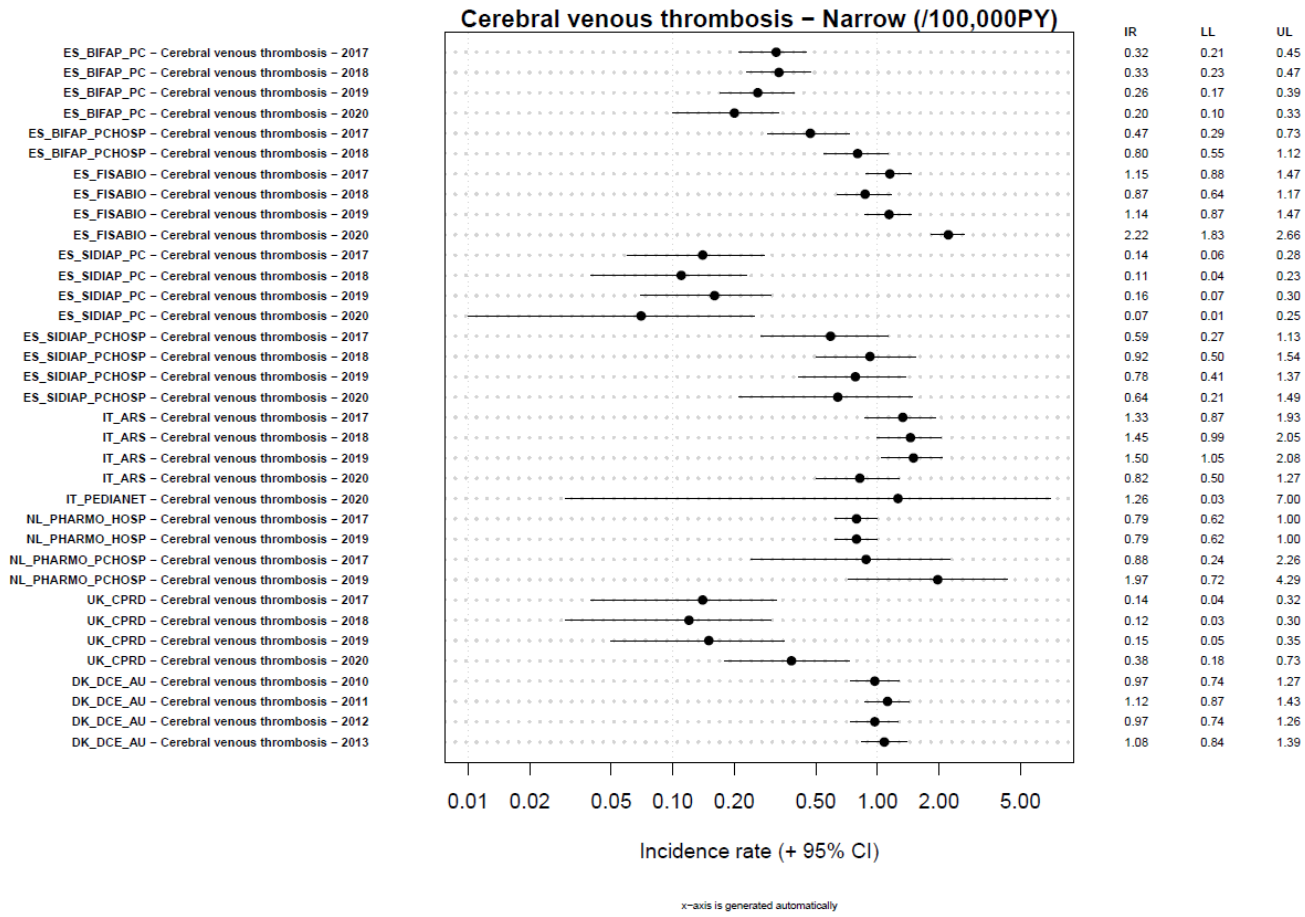


Figure 38 Incidence of Cerebral venous thrombosis by data source and calendar year

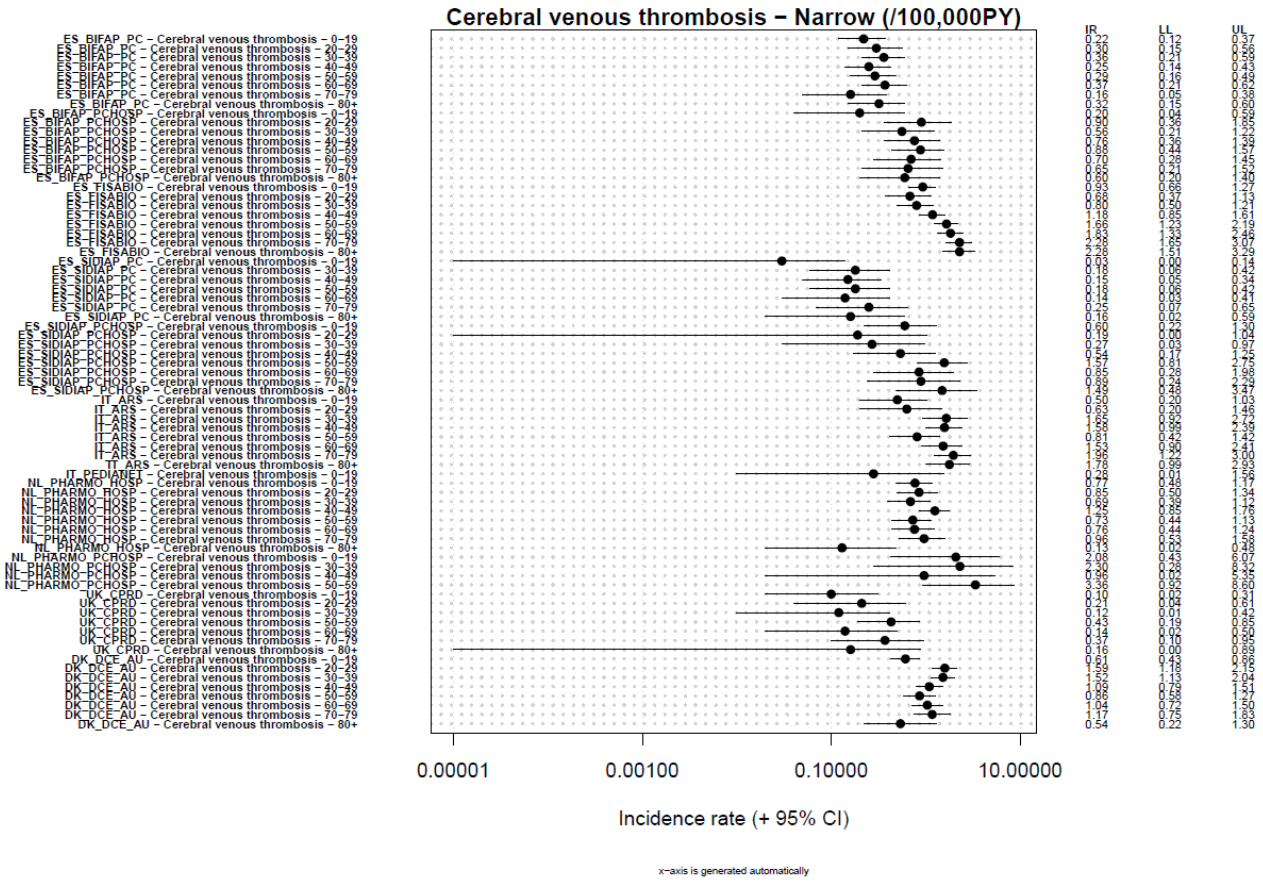


Figure 39 Incidence of Cerebral venous thrombosis by data source and age

10.3.1.13.5 Ischemic stroke

Stroke is a family of diseases often stratified into ischemic and hemorrhagic stroke. It is defined as an abrupt onset of focal brain, spinal cord, or retinal injury due to abnormalities of cerebral blood flow. On the basis of pathomechanism and etiology, stroke can be classified as ischemic stroke (around 80% of all strokes), hemorrhagic stroke (around 15-20% of all strokes) and cerebral venous thrombosis (< 1% of all strokes).

Ischemic stroke is a thrombotic condition similar to ischemic heart disease, but manifested as an occlusion of an artery and resulting in a reduction of focal cerebral perfusion. It may be caused by atherosclerotic plaques, degenerative lesions, cardiac embolism or less common causes such as coagulopathies. The incidence of ischemic stroke was estimated at 134/100,000 person-years in the general population from a study conducted in the Danish registries between 1997 and 2017 [42]. A clear increased age pattern was observed.
The rates of ischemic are lower in GP only data sources as compared to data sources including hospital data sources. Rates decrease over time in FISABIO and ARS. The rate in 2018 in PHARMO PCHOSP

[42] Charlotte Andreasen, Gunnar H. Gislason, MD, PhD, et al. Incidence of Ischemic Stroke in Individuals With and Without Aortic Valve Stenosis A Danish Retrospective Cohort Study. Stroke Volume 51, Issue 5, May 2020, Pages 1364-1371.https://doi.org/10.1161/STROKEAHA.119.028389

is an outlier and should be not be included in calculations, it is currently being explored what is the cause of this outlier. Pottegard reported rates of 103 and 75 per 100,000 in Denmark and Norway which is consistent with our data.

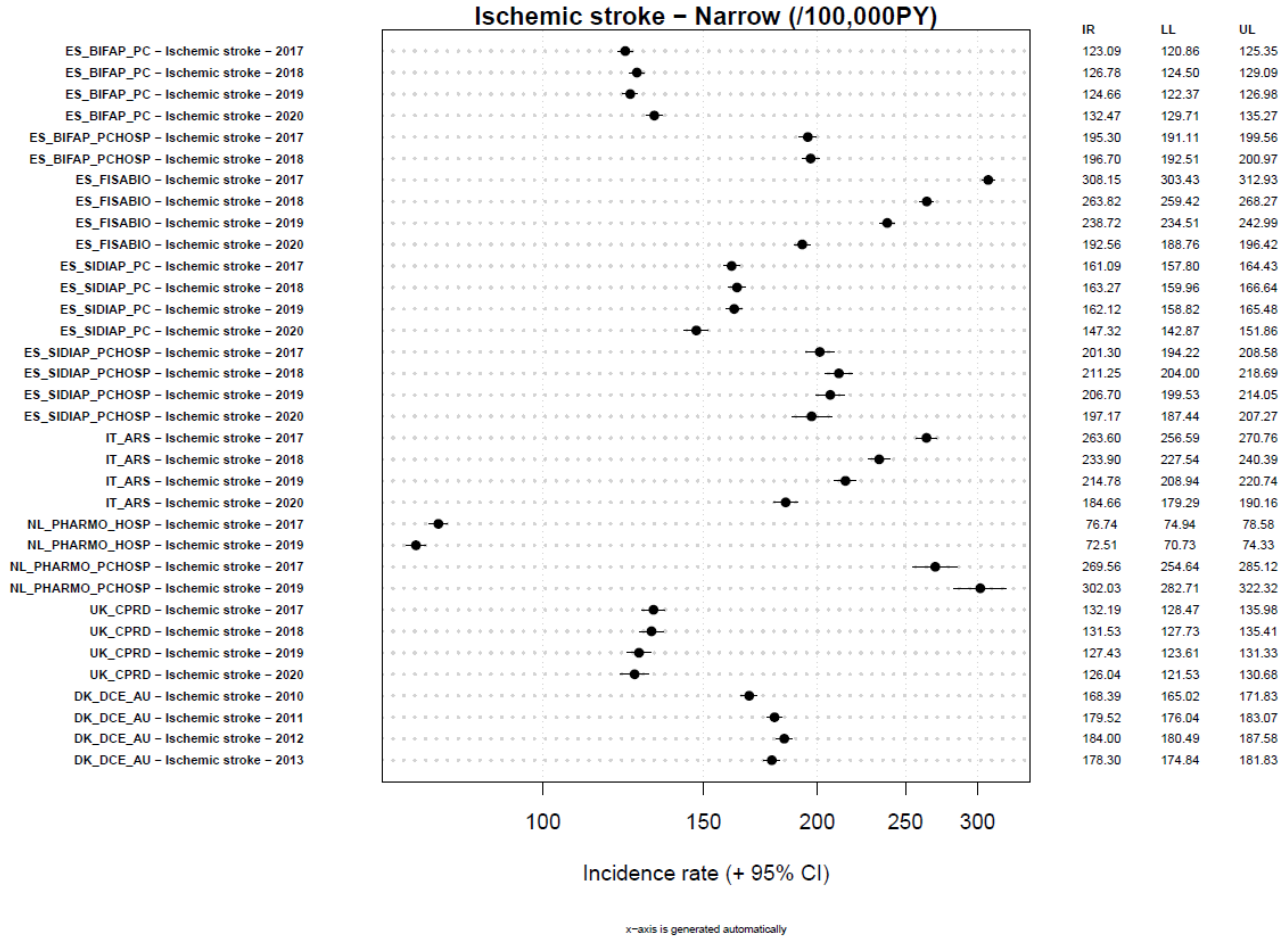


Figure 40 Incidence of Ischemic stroke by data source and calendar year

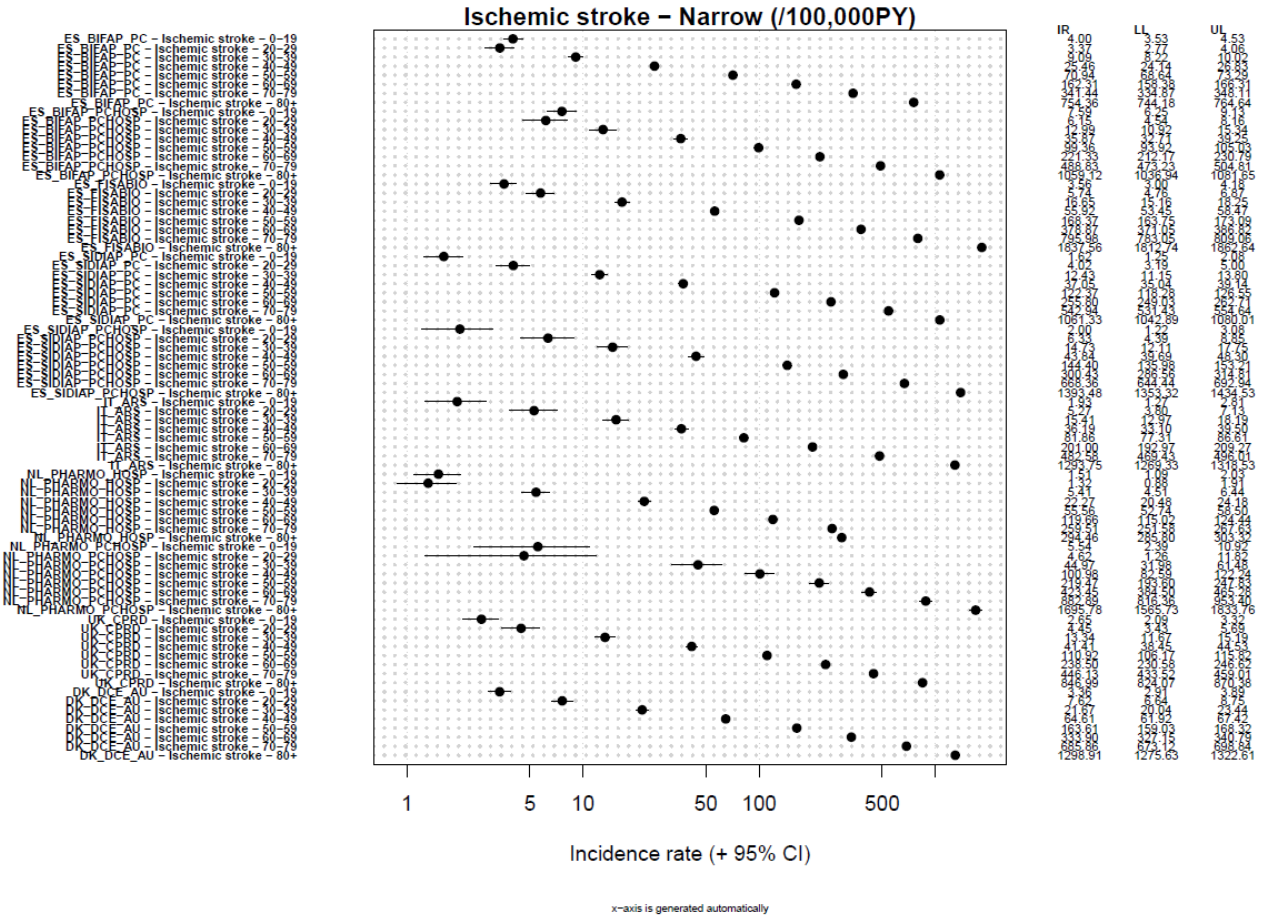


Figure 41 Incidence of Ischemic stroke by data source and age

10.3.1.13.6 Hemorrhagic stroke

A hemorrhagic stroke is bleeding (hemorrhage) that suddenly interferes with the brain's function. This bleeding can occur either within the brain or between the brain and the skull. Hemorrhagic strokes account for about 20% of all strokes, for this incidence we did not include subarachnoid hemorrhage because of the different etiology. Intracerebral hemorrhage has an overall incidence of 24.6/100,000 person-years. A study conducted in The Netherlands showed stable incidence rates over time with a strong age pattern. In the year 2020, rates were of 5.9/100,000 person-years, 37.2/100,000 person-years and 176.3/100,000 person-years among the age groups 35-54, 55-74 and 75-94, respectively [43]. Rates for intracerebral hemorrhage are 20 and 14 per 100,000 PY in Denmark and Norway as reported by Pottegard *et al*. IRs were stable overtime with higher rates observed in ARS (figure 42). Li et al. reported pooled rates of hemorrhagic stroke ranging between 7 and 500 /100,000 with increasing age, higher than what we observed.

[43] Jolink WM, Klijn CJ, Brouwers PJ, Kappelle LJ, Vaartjes I. Time trends in incidence, case fatality, and mortality of intracerebral hemorrhage. Neurology. 2015 Oct 13;85(15):1318-24. doi: 10.1212/WNL.0000000000002015. Epub 2015 Sep 16. PMID: 26377254.

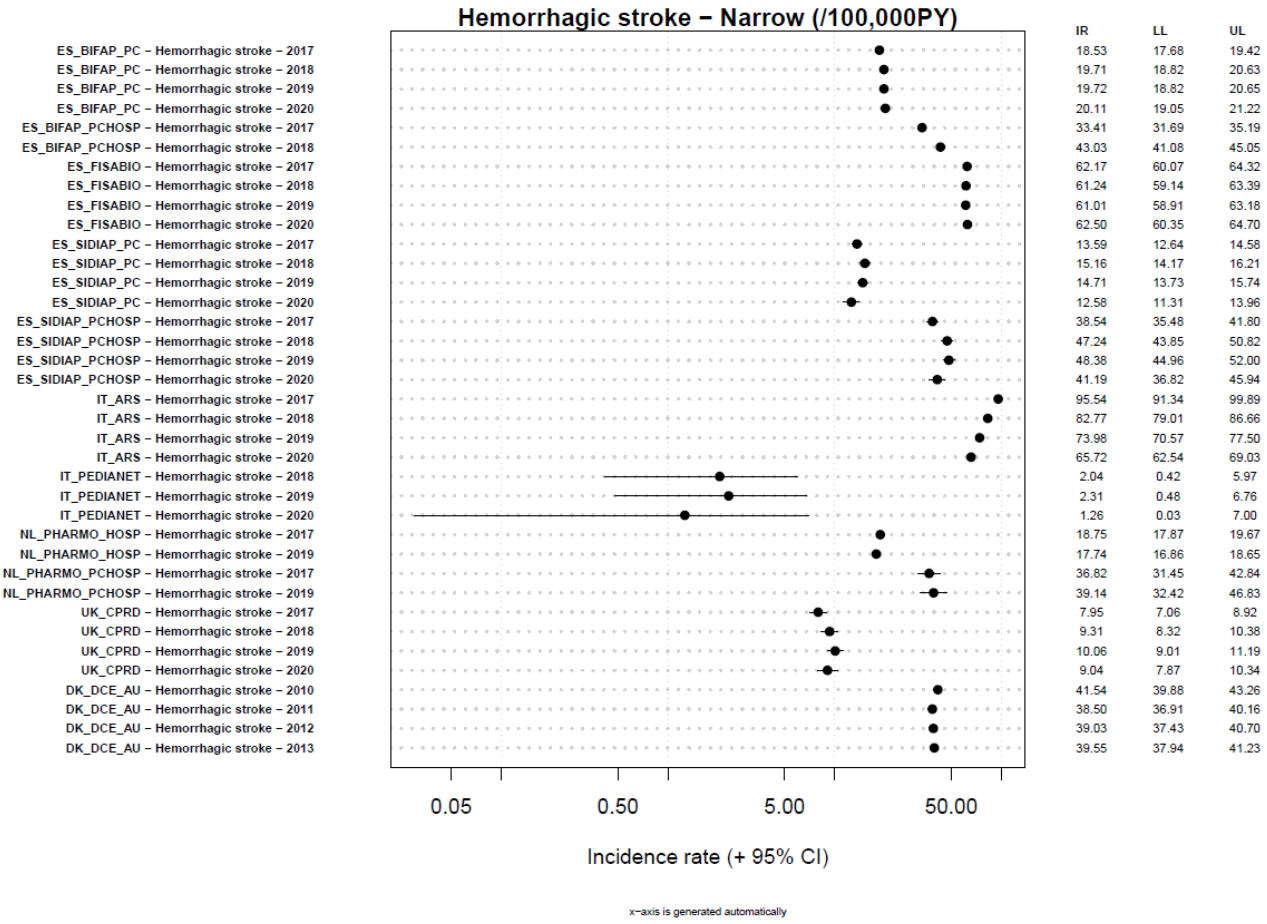


Figure 42 Incidence of hemorrhagic stroke by data source and calendar year

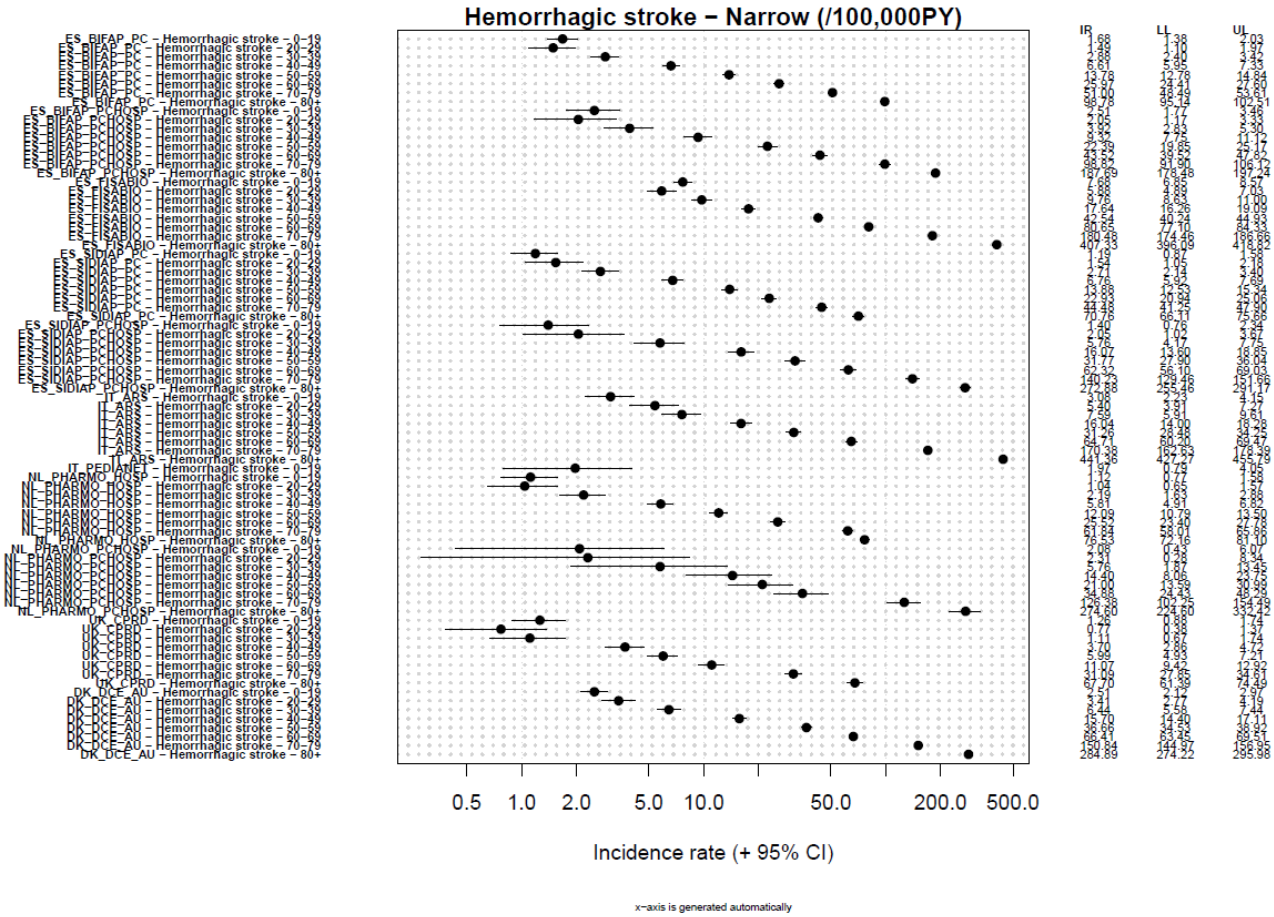


Figure 43 Incidence of hemorrhagic stroke by data source and age

10.3.1.14 Single organ cutaneous vasculitis

Definition and codes in Zenodo
Engelen, R, Willame, C, Martín-Pérez, M, García-Poza, P, Souverein, P, Belbachir, L, Durán, C, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Single Organ Cutaneous Vasculitis (1.0). Zenodo. https://doi.org/10.5281/zenodo.5234977

Single Organ Cutaneous Vasculitis is a syndrome characterized by clinical and histological features of small vessel vasculitis of the skin without involvement of other organ systems. It can be the first sign of systemic vasculitis. It is a disease that is diagnosed by outpatient visits.

Events of single organ cutaneous vasculitis using the narrow definition were observed in all databases. IRs were stable over calendar time while a significant decrease was observed for the year 2020 in all databases potentially due to the lockdown. The rates vary based on provenance of diagnosis data in particular in PHARMO. FISABIO has the highest rate but the ICD10CM codes were considered more generic than other dictionaries, so there may be some misclassification.

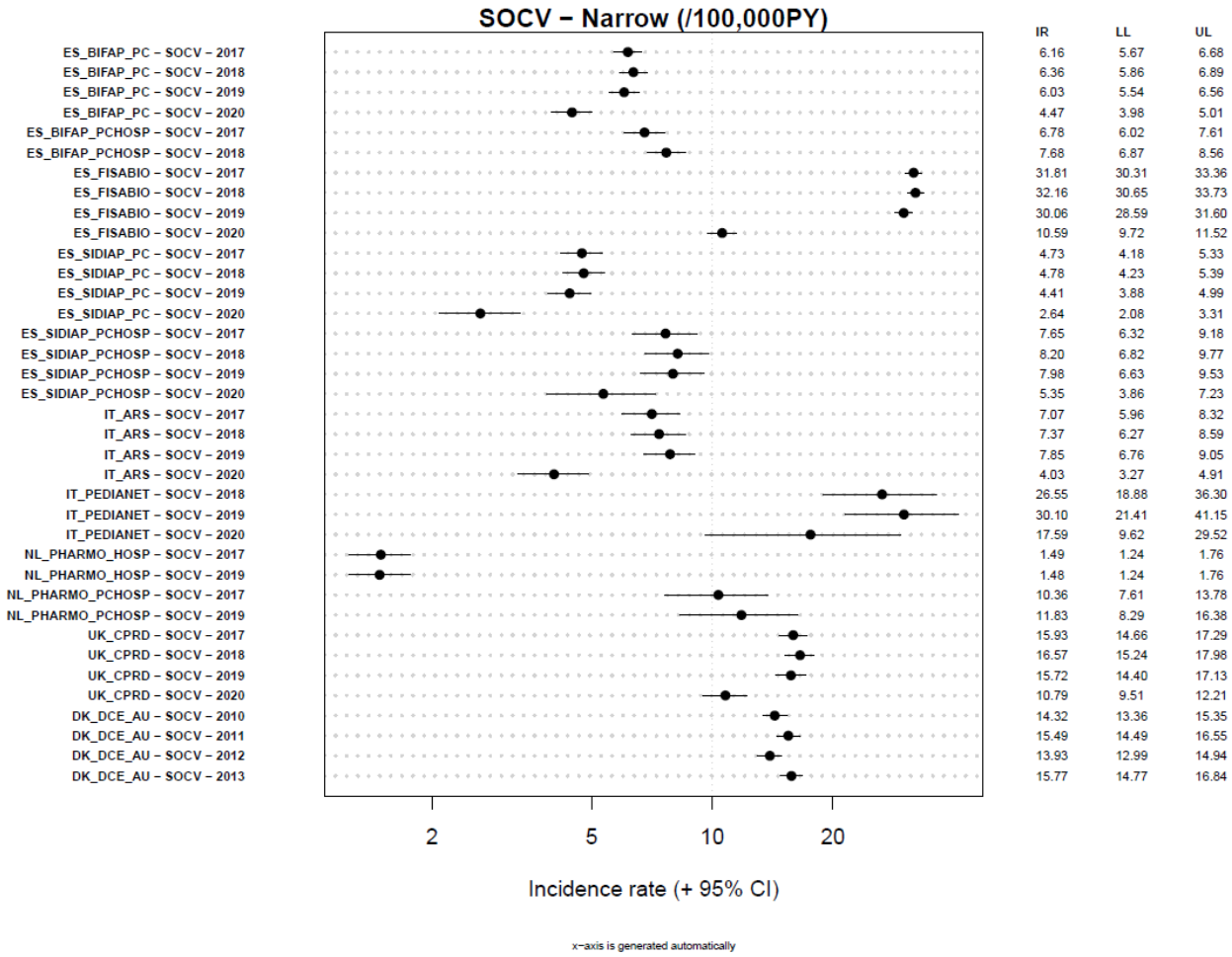


Figure 44 Incidence of single organ cutaneous vasculitis by data source and calendar year

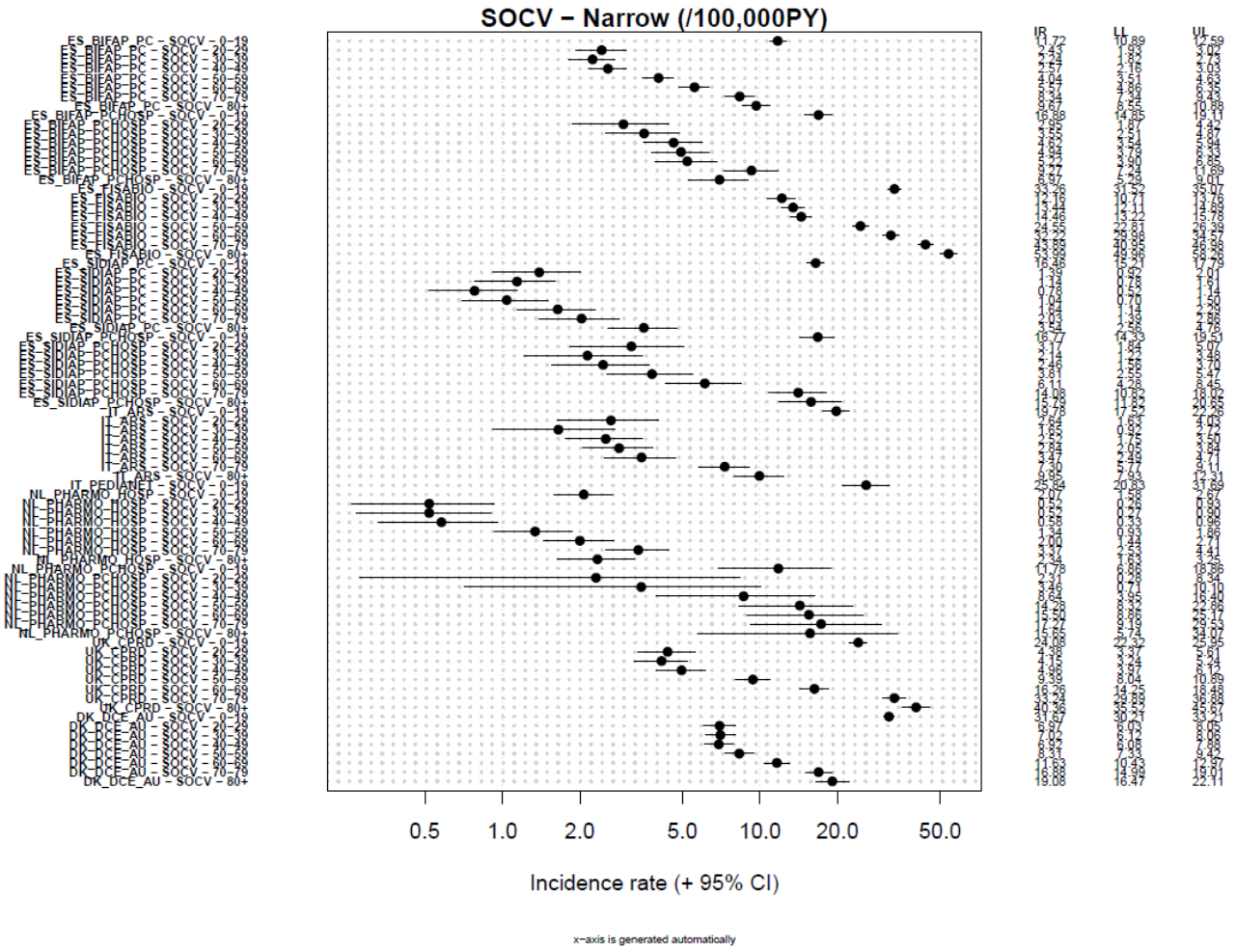


Figure 45 Incidence of single organ cutaneous vasculitis by data source and age

Since this is a disease that has been specified since 2012, there are no good incidence rate studies in the general population as a benchmark.

10.3.1.15 Acute liver injury

Definition and codes in Zenodo
Rojo Villaescusa, M, Belbachir, L, Willame, C, Martín-Pérez, M, García-Poza, P, Durán, C, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Acute Liver Injury (1.0). Zenodo. https://doi.org/10.5281/zenodo.5235027

The European Association for the study of the Liver defines acute liver failure (ALF) as highly specific and rare syndrome, characterised by an acute abnormality of liver blood tests in an individual without underlying chronic liver disease. The disease process is associated with development of a coagulopathy of liver aetiology, and clinically apparent altered level of consciousness due to hepatic encephalopathy (HE). The condition of patients who develop coagulopathy, but do not have any alteration to their level of consciousness is defined as acute liver injury (ALI). ALI is a diagnosis that needs to be made upon testing of liver enzymes.

The incidence rates were low in PEDIANET. A clear pattern with age was observed (Figure 46 and 47). Rates based on both hospital and GP data were consistently higher.

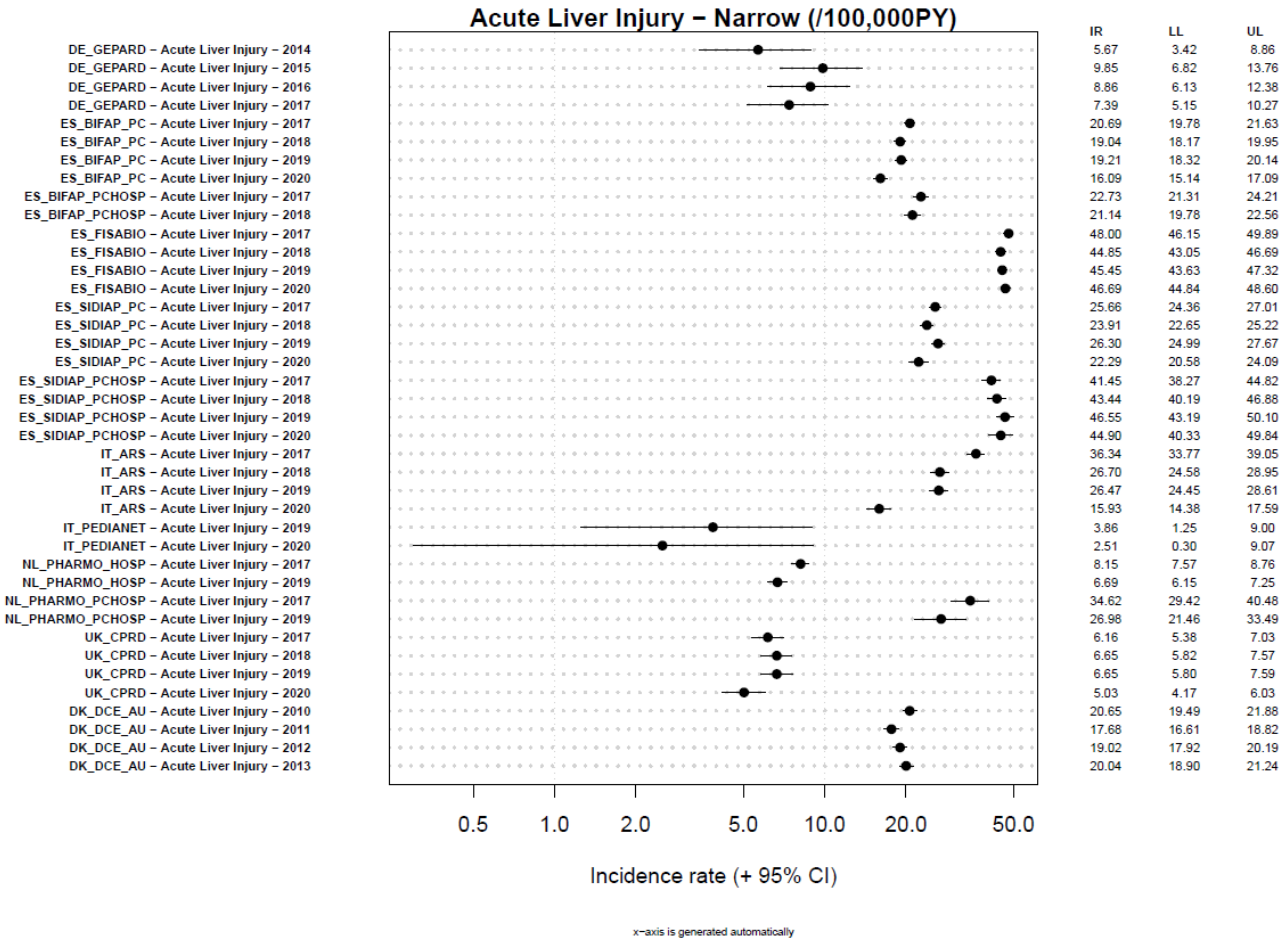


Figure 46 Incidence of Acute Liver Injury by data source and calendar year

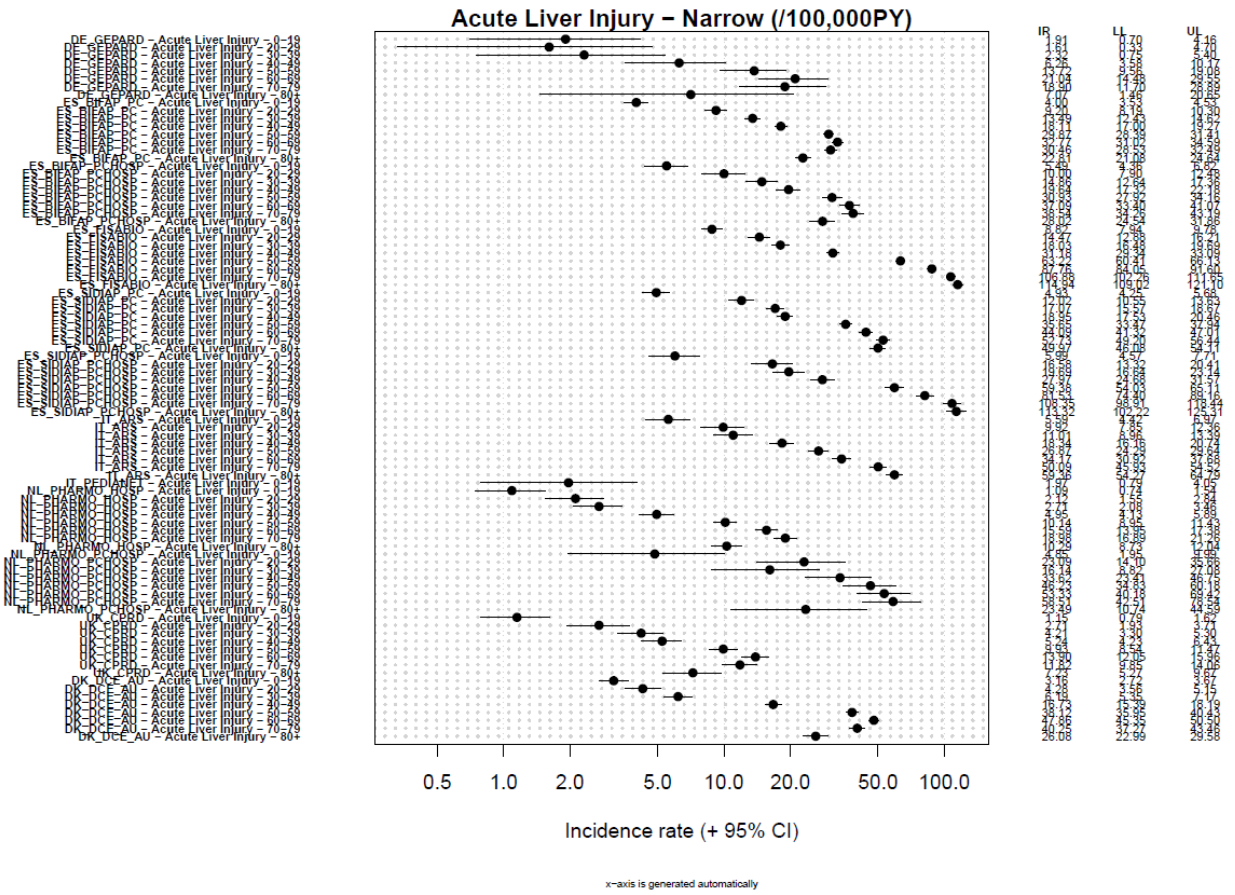


Figure 47 Incidence of acute liver injury by data source and age

10.3.1.16 Acute kidney injury

Definition and codes in Zenodo
Kelters, I, Willame, C, Souverein, P, Martín-Pérez, M, García-Poza, P, Belbachir, L, Durán, C, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Acute Kidney Injury (1.0). Zenodo. https://doi.org/10.5281/zenodo.5235557

AKI is defined as an abrupt (within hours) decrease in kidney function, which encompasses both injury (structural damage) and impairment (loss of function). It is a syndrome that rarely has a sole and distinct pathophysiology. AKI is not a single disease entity. It's a heterogeneous group of conditions characterized by sudden decrease in glomerular filtration rate (GFR) followed by an increase in serum creatinine concentration or oliguria. It occurs in the setting of acute or chronic illness.

IRs for AKI are highest when both hospital and GP data are used, they increased in 2020 for FISABIO but decreased in other data sources. IRs for FISABIO and SIDIAP were slightly higher compared to the other databases, one explanation could be the identification of a large proportion of individuals with an unspecified code (i.e. the ICD-10: N17.9 – Acute renal failure, unspecified).

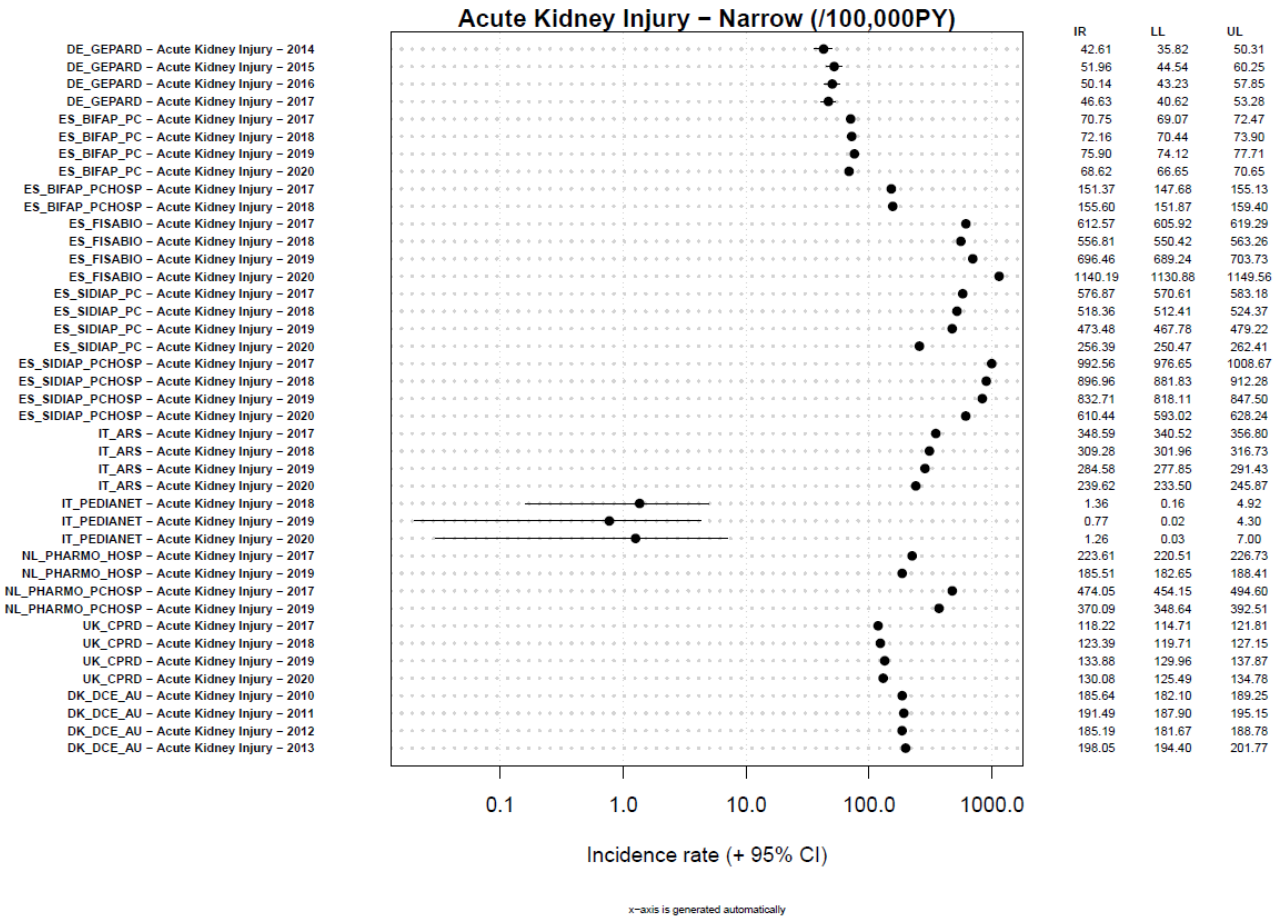


Figure 48 Incidence of Acute kidney Injury by data source and calendar year

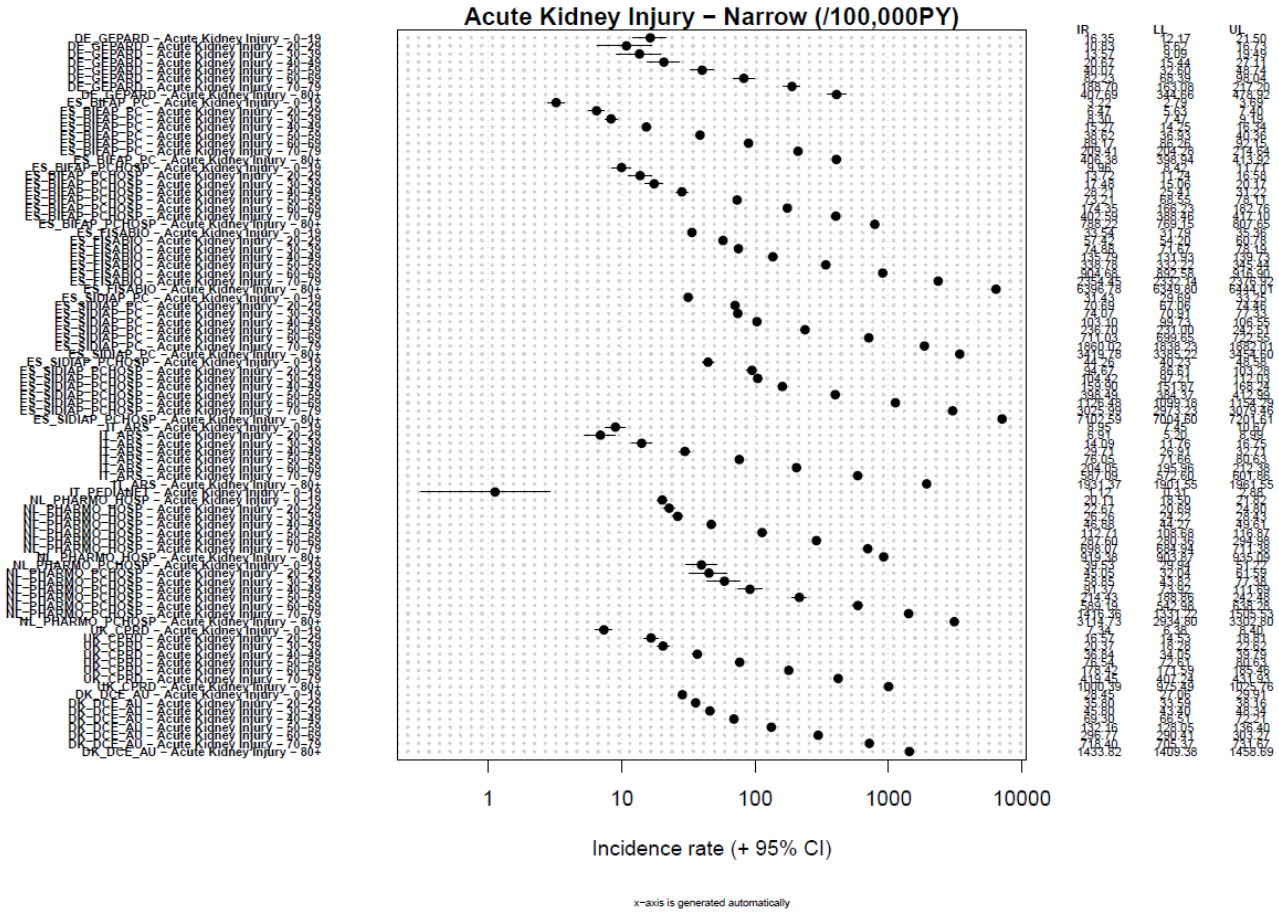
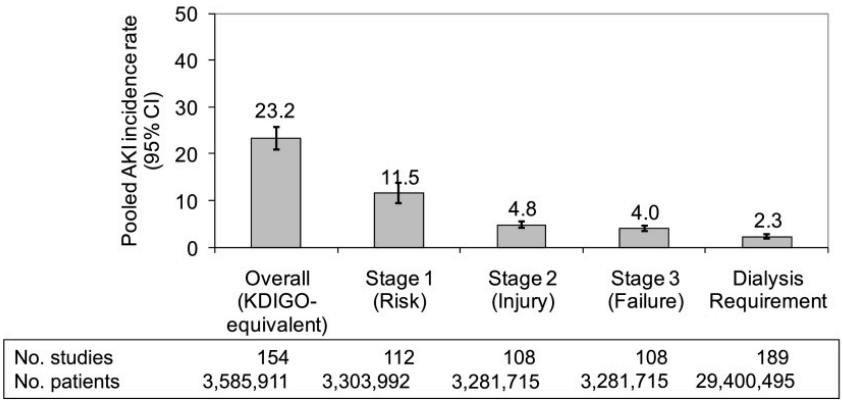


Figure 49 Incidence of acute kidney injury by data source and age

Reference rates can be obtained from the worldwide meta-analysis of AKI[44] showing a rate of 4.8% of AKI in hospitalized patients, our rates are lower as they are in the general population.



Pooled incidence rate of AKI in studies that used KDIGO-equivalent serum creatinine-based AKI definition and staging system, or dialysis requirement. Some studies and patients are included in more than one category. KDIGO, Kidney Disease Improving Global Outcomes; 95% CI, 95% confidence interval.

[44] Susantitaphong P, Cruz DN, Cerda J, et al. World incidence of AKI: a meta-analysis [published correction appears in Clin J Am Soc Nephrol. 2014 Jun 6;9(6):1148]. *Clin J Am Soc Nephrol*. 2013;8(9):1482-1493. doi:10.2215/CJN.00710113

10.3.1.17 Generalized convulsion

Definition and codes in Zenodo
van Wijngaarden, P, Willame, C, Durán, C, Belbachir, L, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition – Generalized Convulsions (Version 1). Zenodo. https://doi.org/10.5281/zenodo.5236092

Seizures are episodes of neuronal hyperactivity most commonly resulting in sudden, involuntary muscular contractions. They may also manifest as sensory disturbances, autonomic dysfunction and behavioral abnormalities, and impairment or loss of consciousness. Descriptions and classifications of seizures are complex and subject to change, because the etiology and pathogenesis of most seizures remain to be elucidated[45].

IRs from PHARMO-HOSP were significantly lower than the IRs from the other databases. The incidence rate was higher in children than in other subsequent age groups, and increased again with older age, this pattern was consistent across all data sources (Figure 50).

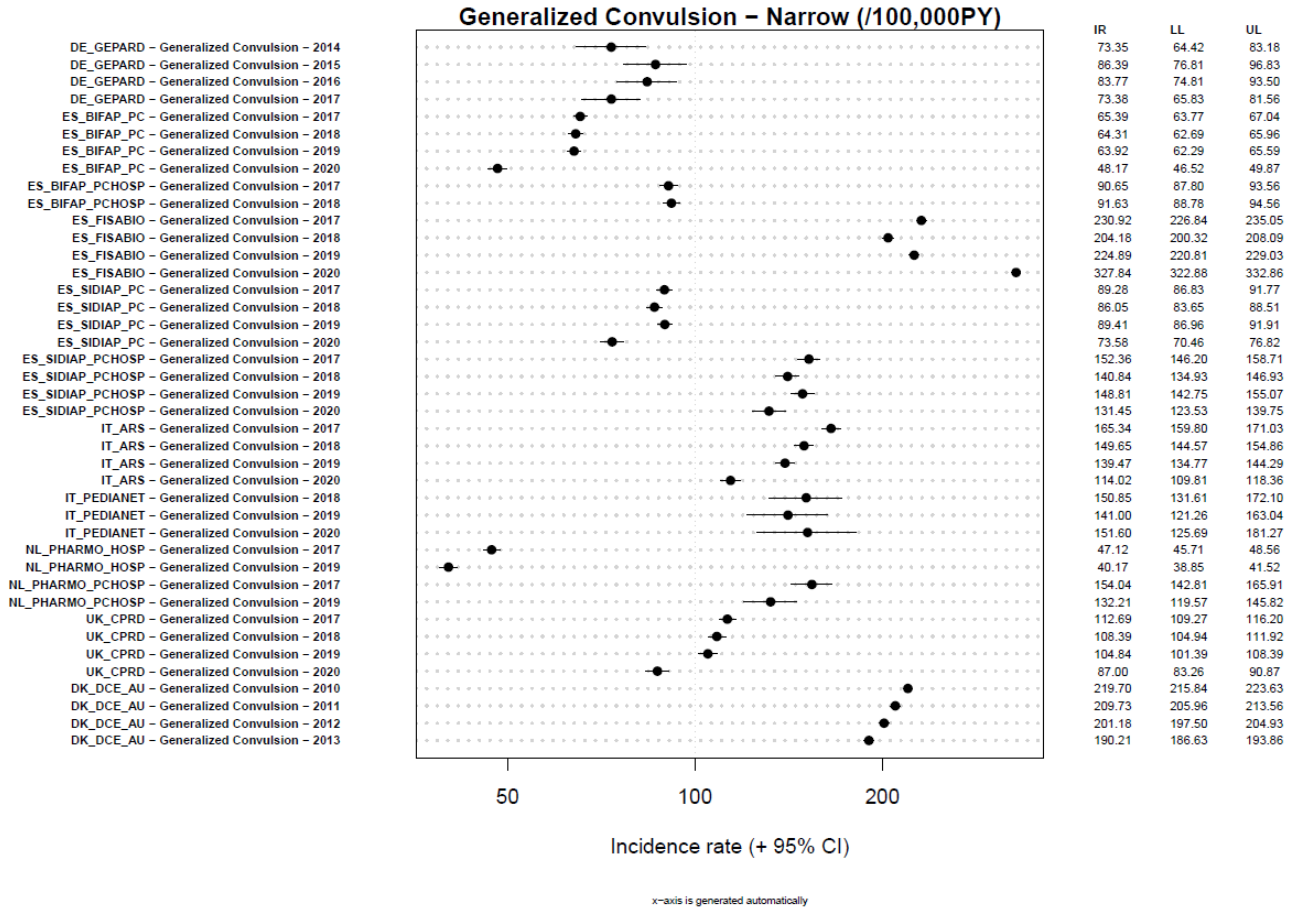


Figure 50 Incidence of generalized convulsions by data source and calendar year

[45] Bonhoeffer J, Menkes J, Gold MS, et al. Generalized convulsive seizure as an adverse event following immunization: case definition and guidelines for data collection, analysis, and presentation. Vaccine. 2004;22(5-6):557-562. doi:10.1016/j.vaccine.2003.09.008

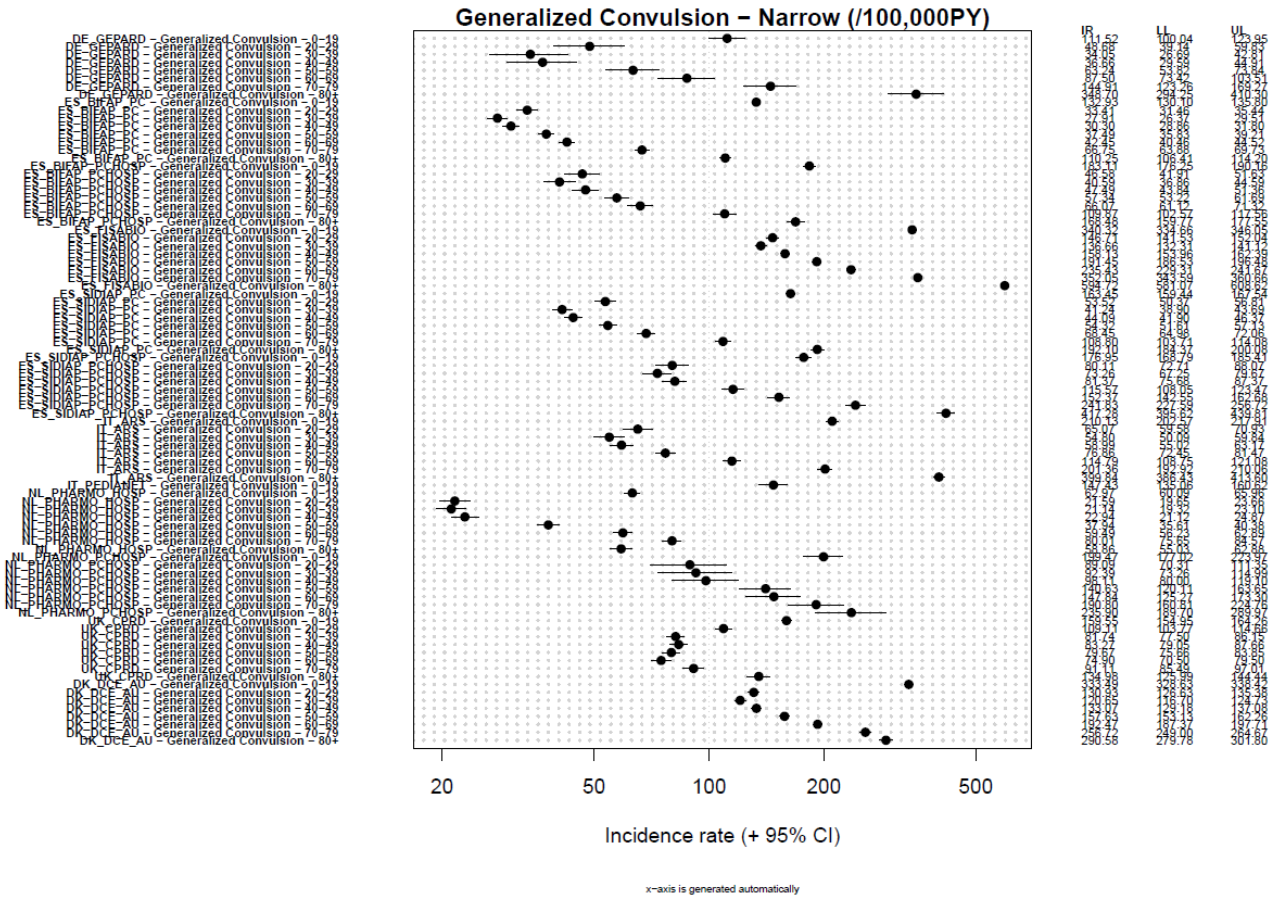


Figure 51 Incidence of generalized convulsions by data source and age

10.3.1.18 (Meningo)encephalitis

Definition and codes in Zenodo

van Wijngaarden, P, Belbachir, L, Durán, C, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition – (Meningo)encephalitis (Version 1). Zenodo. https://doi.org/10.5281/zenodo.5236137

Encephalitis is defined as inflammation of the parenchyma of the brain. Strictly speaking, it is a pathologic diagnosis, in which the presence of inflammation, edema, and neuronophagia (neuronal cell death) is demonstrated by histopathology[46].

The incidence of meningo-encephalitis increased with age (Figure 52). From the literature, IRs of encephalomyelitis in children aged between 0 to 17 years were 0.79/100,000 person-years in the UK

[46] Sejvar JJ, Kohl KS, Bilynsky R, et al. Encephalitis, myelitis, and acute disseminated encephalomyelitis (ADEM): case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 2007;25(31):5771-5792. doi:10.1016/j.vaccine.2007.04.060

[47]. Other studies conducted on the total population suggest rates of 4.3/100,000 person-years in Canada [48] and 4.32/100,000 person-years in the UK [49]. Reference rates of acute encephalitis usually are below 10/100,000 PY which is consistent with our findings.

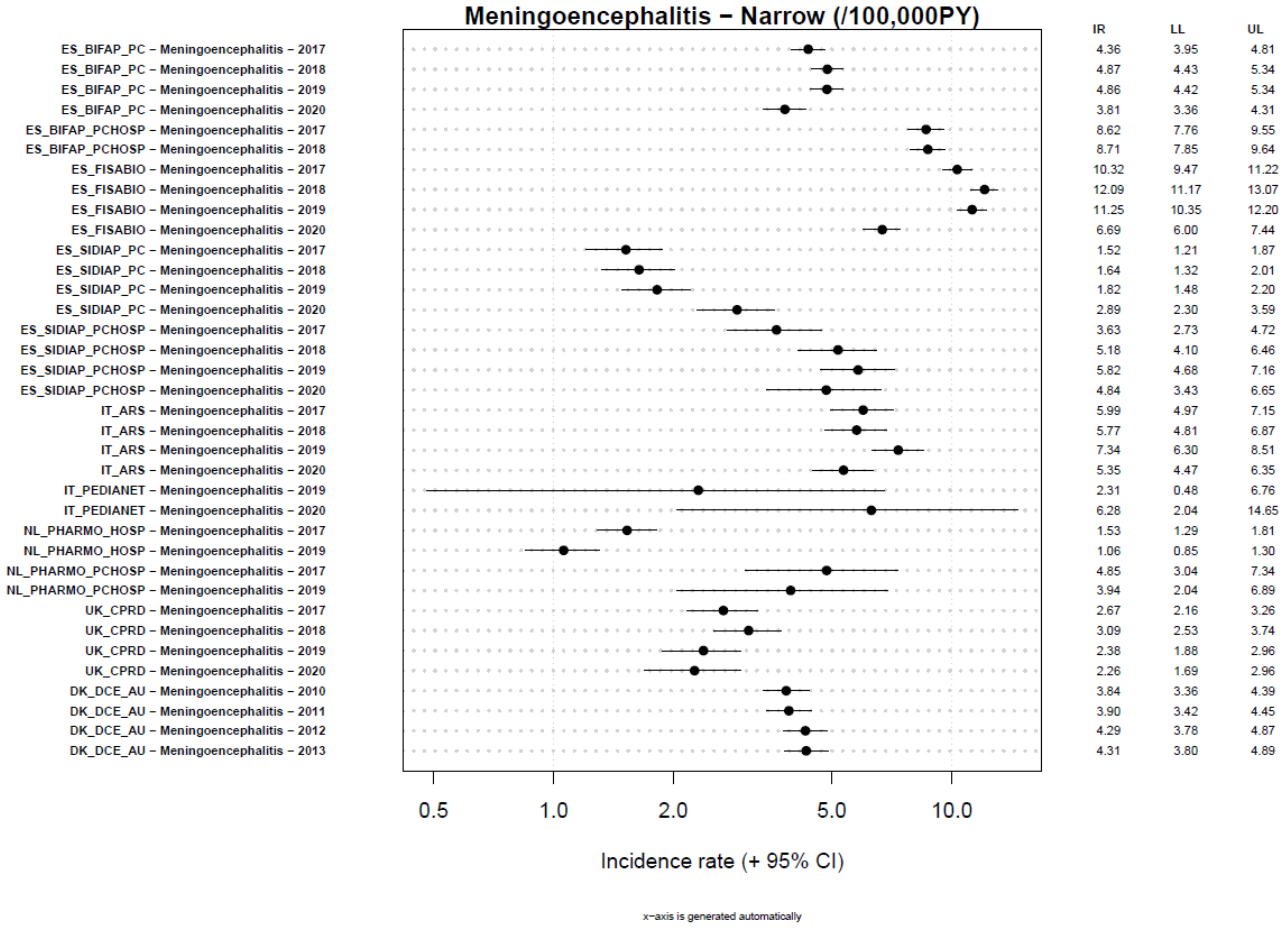


Figure 52 Incidence of meningoencephalitis by data source and calendar year

[47] Iro MA, Sadarangani M, Goldacre R, Nickless A, Pollard AJ, Goldacre MJ. 30-year trends in admission rates for encephalitis in children in England and effect of improved diagnostics and measles-mumps-rubella vaccination: a population-based observational study. Lancet Infect Dis. 2017 Apr;17(4):422-430. doi: 10.1016/S1473-3099(17)30114-7. Epub 2017 Mar 2. PMID: 28259562.

[48] Parpia AS, Li Y, Chen C, Dhar B, Crowcroft NS. Encephalitis, Ontario, Canada, 2002-2013. *Emerg Infect Dis*. 2016;22(3):426-432. doi:10.3201/eid2203.151545

[49] Granerod J, Cousens S, Davies NW, Crowcroft NS, Thomas SL. New estimates of incidence of encephalitis in England. *Emerg Infect Dis*. 2013;19(9):1455-1462. doi:10.3201/eid1909.130064

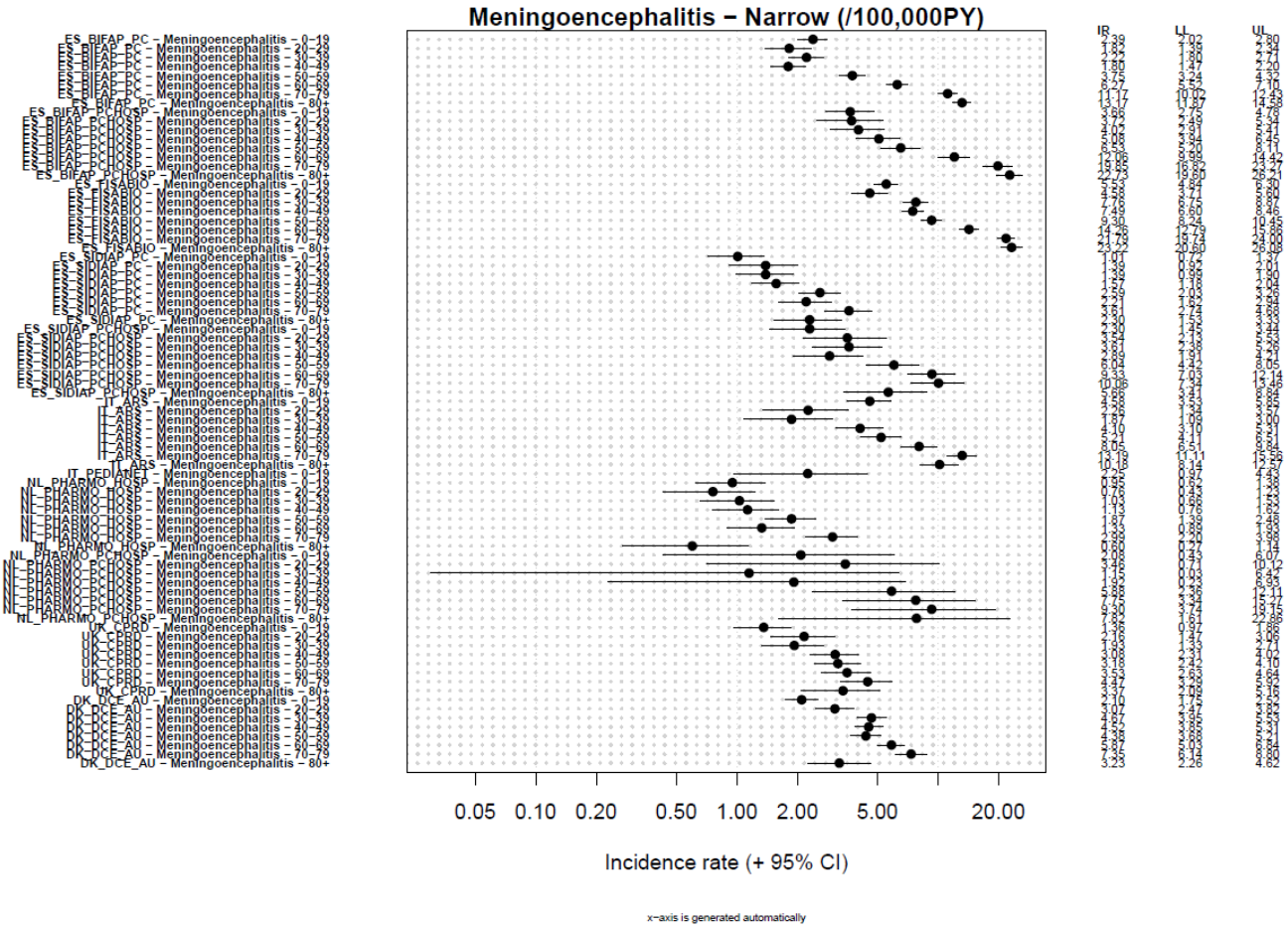


Figure 53 Incidence of meningoencephalitis by data source and age

10.3.1.19 Transverse myelitis

Definition and codes in Zenodo
Sturkenboom, MCJM, Belbachir, L, Souverein, P, Martín-Pérez, M, García-Poza, P, & Durán, C. (2021). ACCESS-Background rate of adverse events-definition –transverse myelitis (1.0). Zenodo. https://doi.org/10.5281/zenodo.5237332

Transverse myelitis is a neurological disorder causing acute spinal cord injury as a result of acute inflammation, often associated with para infectious processes and autoimmune disease[50].

Rates of transverse myelitis were available from most of the databases, except Pedianet. (Figure 54). The incidences did not show a specific age pattern. TM rates were estimated between 1 and 8 new cases per million per year, which is consistent with our data.

[50] *Bhat A, Naguwa S, Cheema G, Gershwin ME. The epidemiology of transverse myelitis. Autoimmun Rev. 2010 Mar;9(5):A395-9. doi: 10.1016/j.autrev.2009.12.007. Epub 2009 Dec 24. PMID: 20035902.*

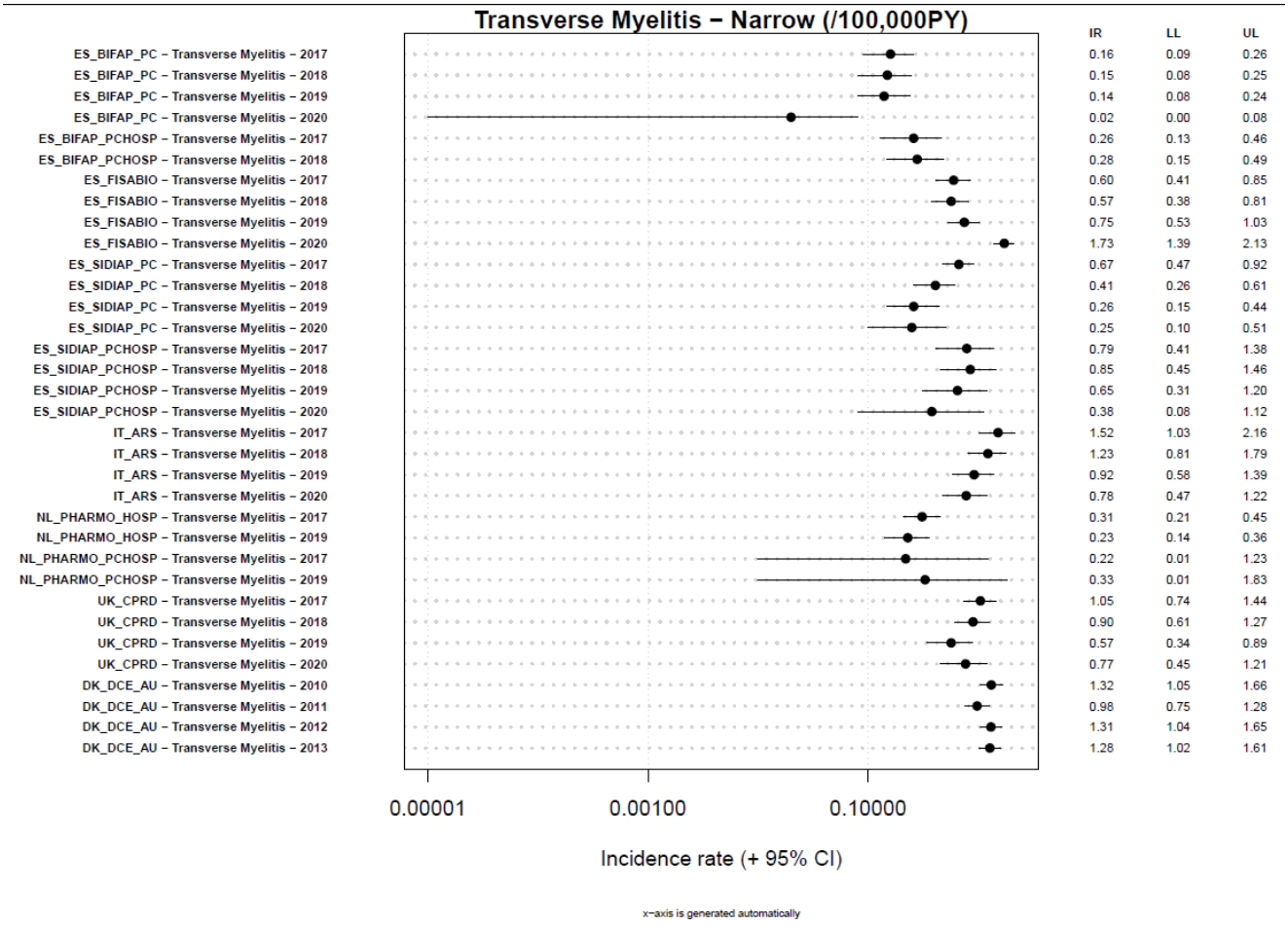


Figure 54 Incidence of transverse myelitis by data source and calendar year

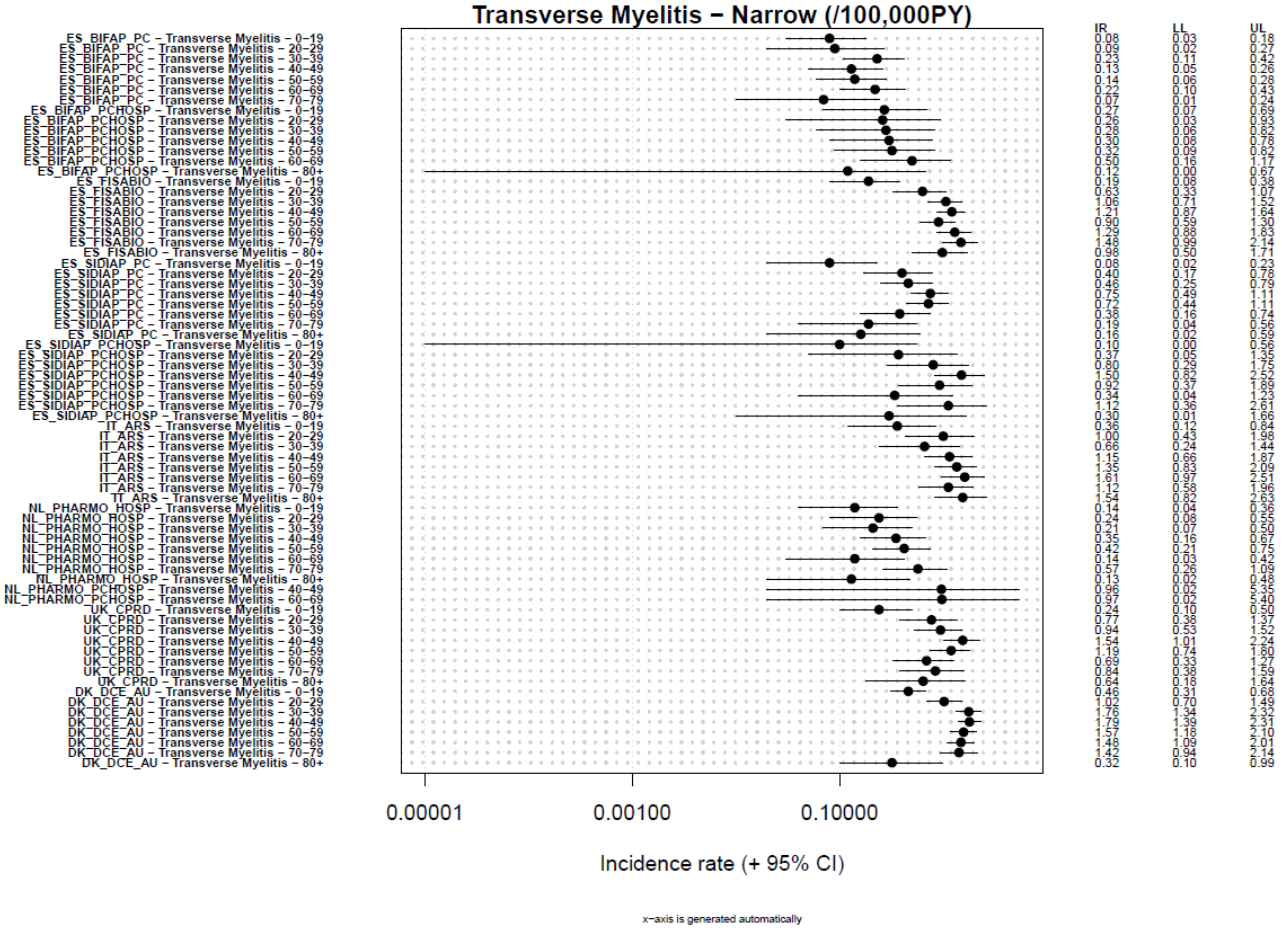


Figure 55 Incidence of transverse myelitis by data source and age

10.3.1.20 Respiratory system – Acute respiratory distress syndrome

Definition and codes in Zenodo

Rojo Villaescusa, M, Dodd, C, Belbachir, L, Martín-Pérez, M, García-Poza, P, Souverein, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Acute Respiratory Distress Syndrome (1.0). Zenodo. https://doi.org/10.5281/zenodo.5236188

Acute respiratory distress syndrome (ARDS) is an acute inflammatory lung process, which leads to protein-rich non-hydrostatic pulmonary edema, causes refractory hypoxemia, increases lung "stiffness" and impairs the ability of the lung to eliminate carbon dioxide[51]

Hospitalisation databases (SIDIAP, ARS and FISABIO) showed significant higher rates for 2020 compared to the other years in their database (Figure 56). The incidence increased consistently with age (Figure 57).

[51] Rezoagli E, Fumagalli R, Bellani G. Definition and epidemiology of acute respiratory distress syndrome. *Ann Transl Med.* 2017;5(14):282. doi:10.21037/atm.2017.06.62

Data from Iceland show that the age-standardised incidence of ARDS was 7.2 cases per 100,000 person-years and was increased by 0.2 cases per year (P < 0.001). The most common causes of ARDS were pneumonia (29%) and sepsis (29%). An overview paper reports rates between 10 and 79 per 100,000 PY [52]

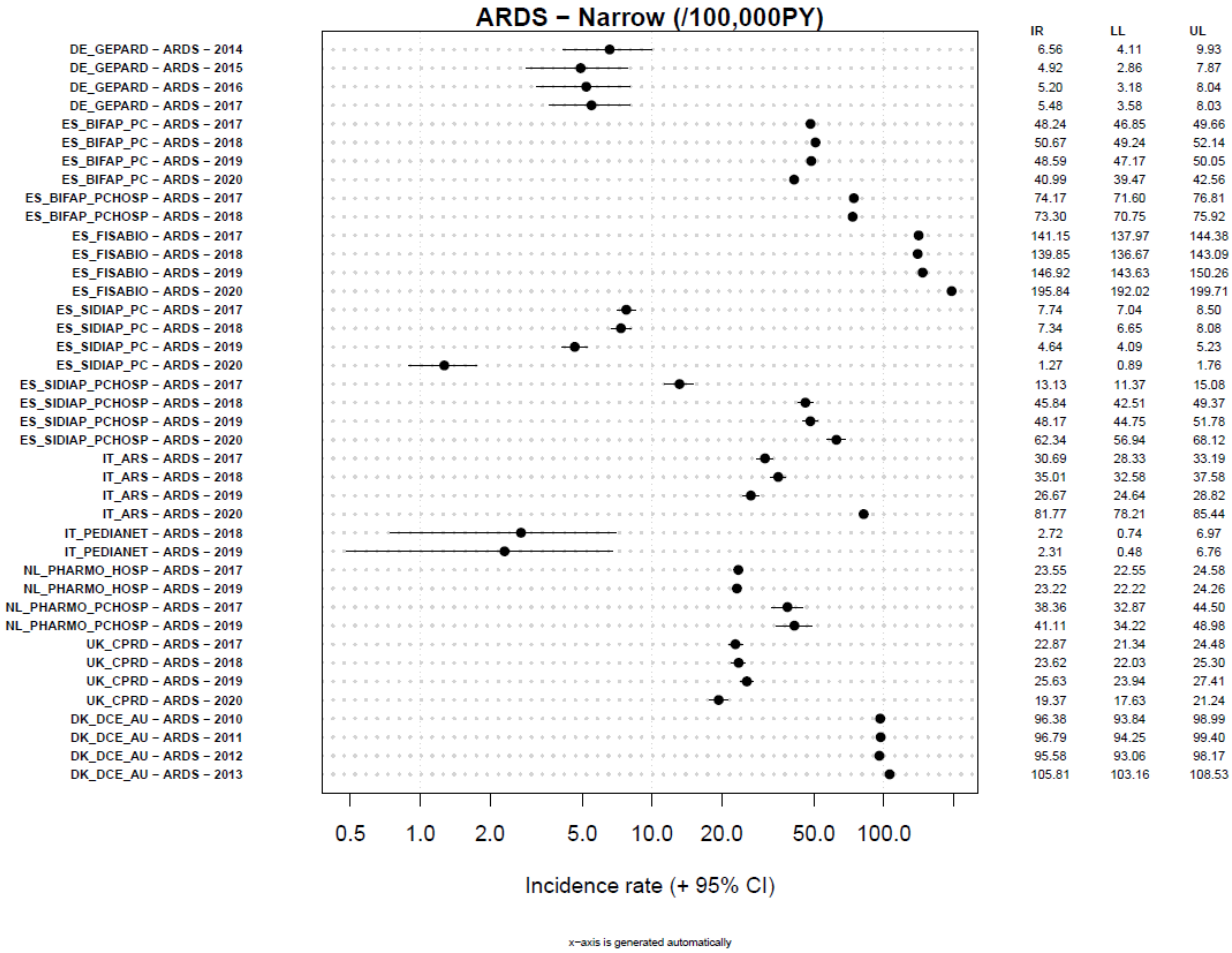


Figure 56 Incidence of ARDS by data source and calendar year

[52] Rezoagli E, Fumagalli R, Bellani G. Definition and epidemiology of acute respiratory distress syndrome. *Ann Transl Med.* 2017;5(14):282. doi:10.21037/atm.2017.06.62

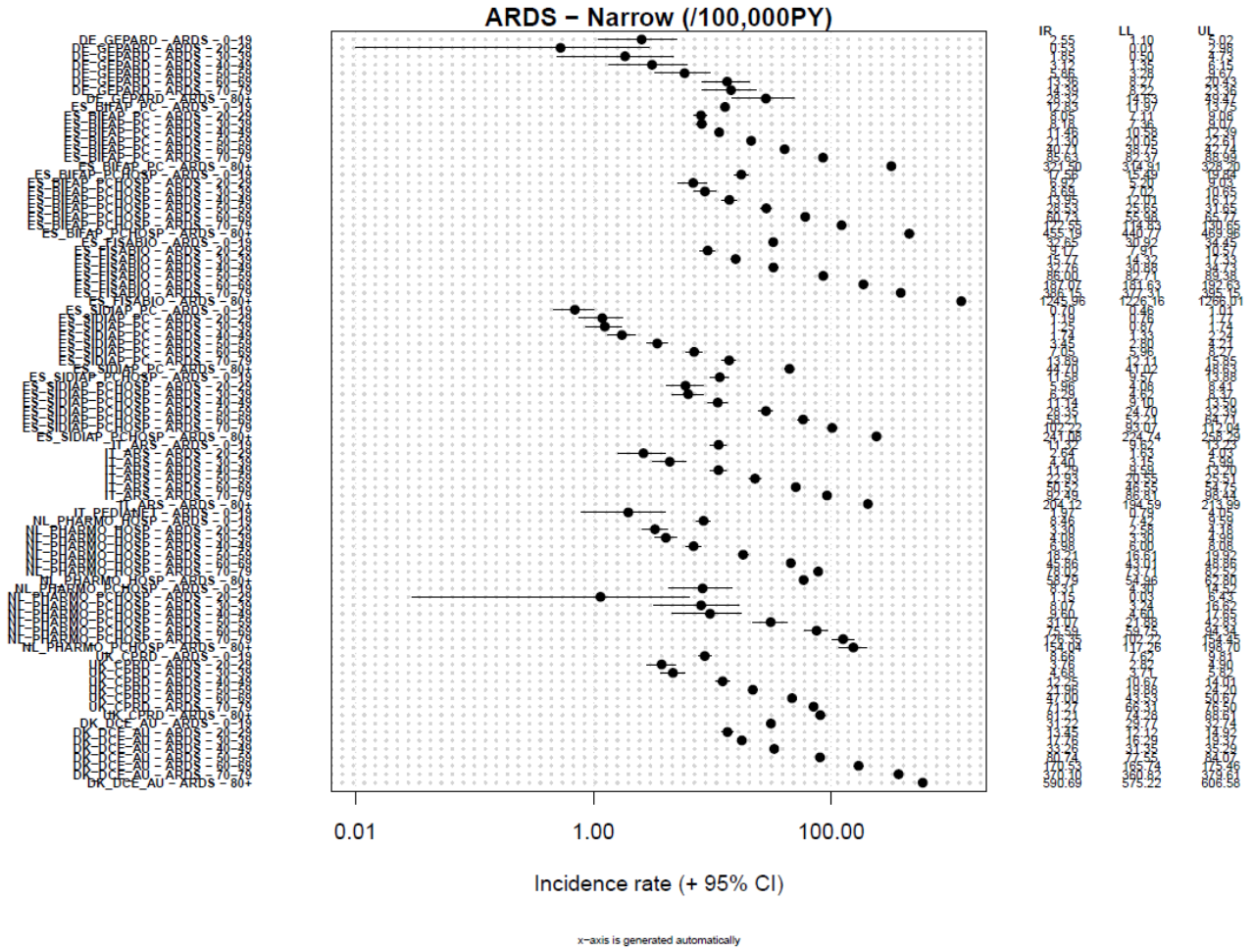


Figure 57 Incidence of ARDS by data source and age

10.3.1.21 Erythema multiforme

Definition and codes in Zenodo

Rojo Villaescusa, M, Willame, C, Durán, C, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Erythema Multiforma (1.0). Zenodo. https://doi.org/10.5281/zenodo.5236231

Erythema multiforme (EM) is an acute, self-limited disease that is typically associated with hypersensitivity reactions to viruses, as well as drugs. It is characterized by targetoid erythematous lesions with predominant acral localization and can be subdivided into isolated cutaneous and combined mucocutaneous forms[53].

IRs were distributed between 0.25/100,000 person-years (CI 95% 0.16 – 0.38) from PHARMO and 15.09/100,000 person-years (IC95%: 14.08-16.17) from FISABIO (Figure 58). For 2020 we observe significant lower IRs of 3.99/100,000 person-years versus 8.85/100,000 person-years in 2019 in ARS,

[53] Lerch M, Mainetti C, Terziroli Beretta-Piccoli B, Harr T. Current Perspectives on Erythema Multiforme. Clinic Rev Allerg Immunol 2018;54(1):177–84.

3.87/100,000 person-years versus 6.25 in 2019 in BIFAP, and 6.38/100,000 person-years versus 12.58 in 2019 in FISABIO. Rates were highest in children.

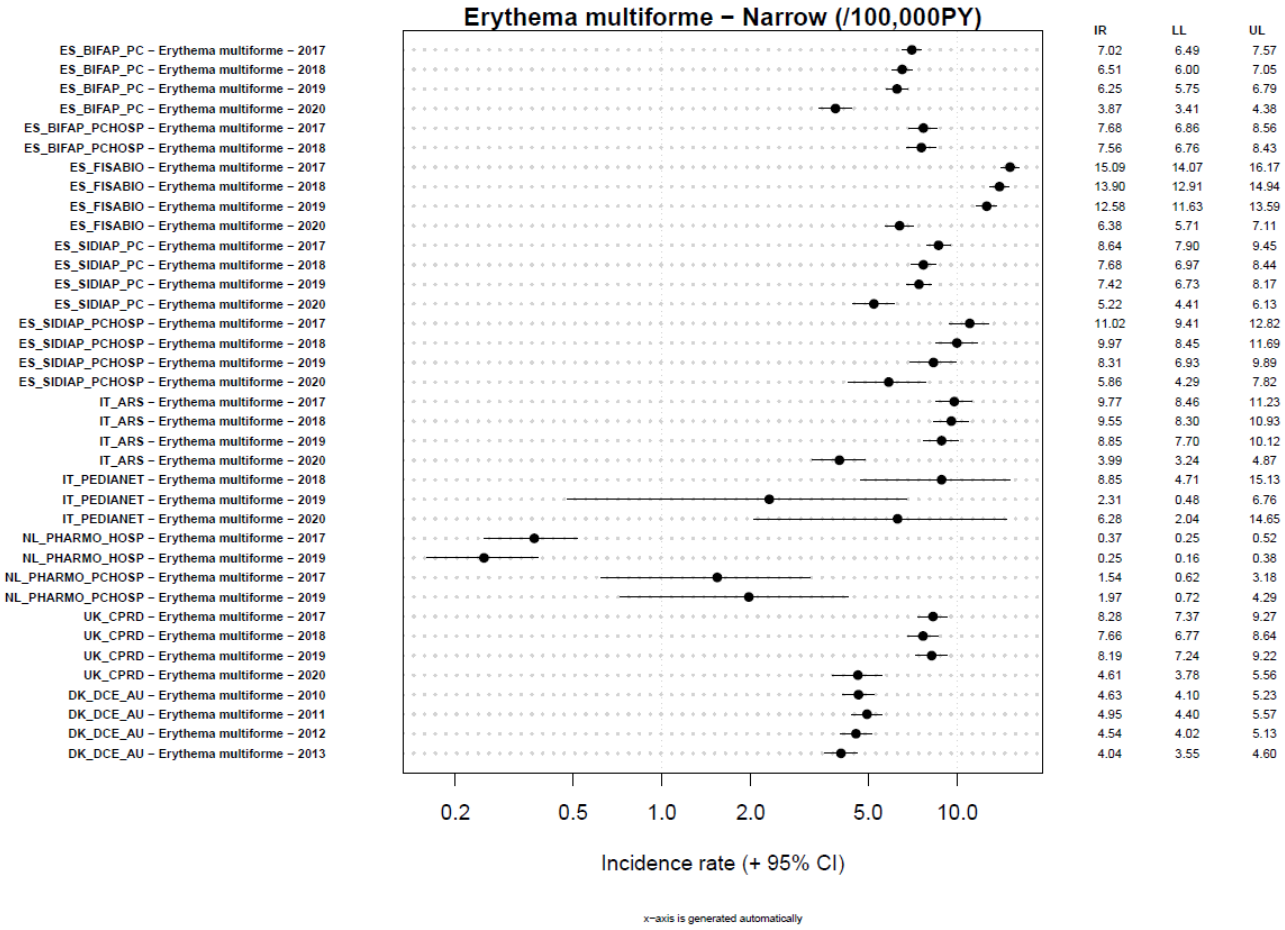


Figure 58 Incidence of erythema multiforme by data source and calendar year

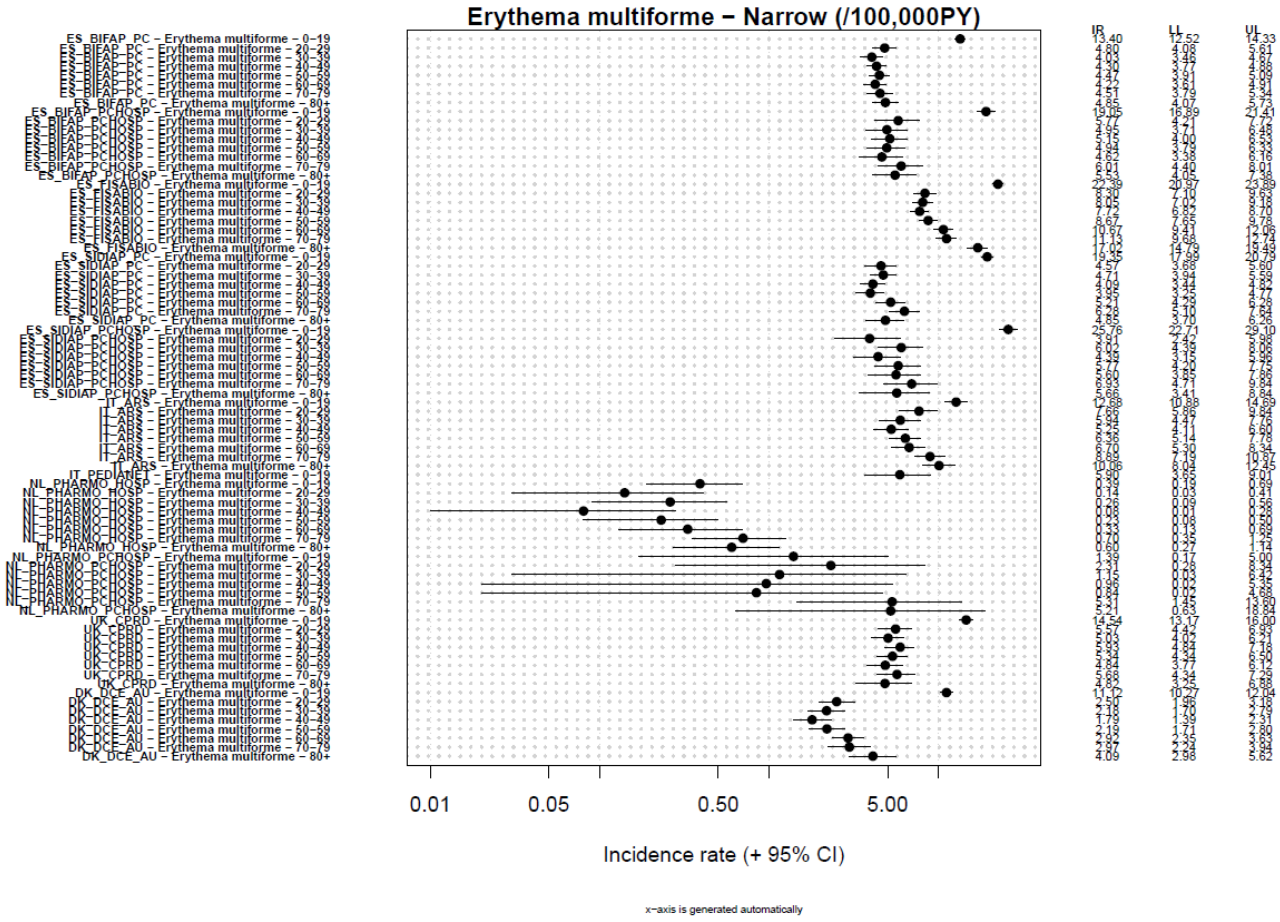


Figure 59 Incidence of erythema multiforme by data source and age

10.3.1.22 Chilblain–like lesions

Definition and codes in Zenodo
van Wijngaarden, P, Willame, C, Belbachir, L, Durán, C, Martín-Pérez, M, García-Poza, P, Souverein, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Chilblain Like lesions (1.0). Zenodo. https://doi.org/10.5281/zenodo.5236280

During the recent COVID-19 pandemic patients with little or no symptoms presented themselves with chilblain-like lesions located on the toes and fingers. These patients had no underlying autoimmune disease (such as lupus erythematosus), Raynaud's phenomenon or previous episodes of idiopathic chilblains. It mostly affected children and young adults and the lesions took place later in the course of the (suspected) COVID-19 disease. The chilblain-like lesions manifest as multiple red-violaceous edematous lesions with papules and macules located on acral regions such as toes, the feet (heel, sole) and/or the fingers, asymptomatic or associated with pruritis of mild pain. Because of the presentation similar with chilblain, it is referred to as pseudo-chilblain of chilblain-like lesions (See event definition form).

We observed a range of IRs for chilblain-like lesions (figure 60). The yearly rates presented significant differences in each year for FISABIO ranging from 37.05/100,000 person-years (CI 95%: 35.42 –

38.75) in 2020 to 64.05/100,000 person-years (CI 95%: 62.93 – 66.24) in 2017. Rates were very low in hospital based datasources only.

We do not observe a clear age pattern in incidence (Figure 61).

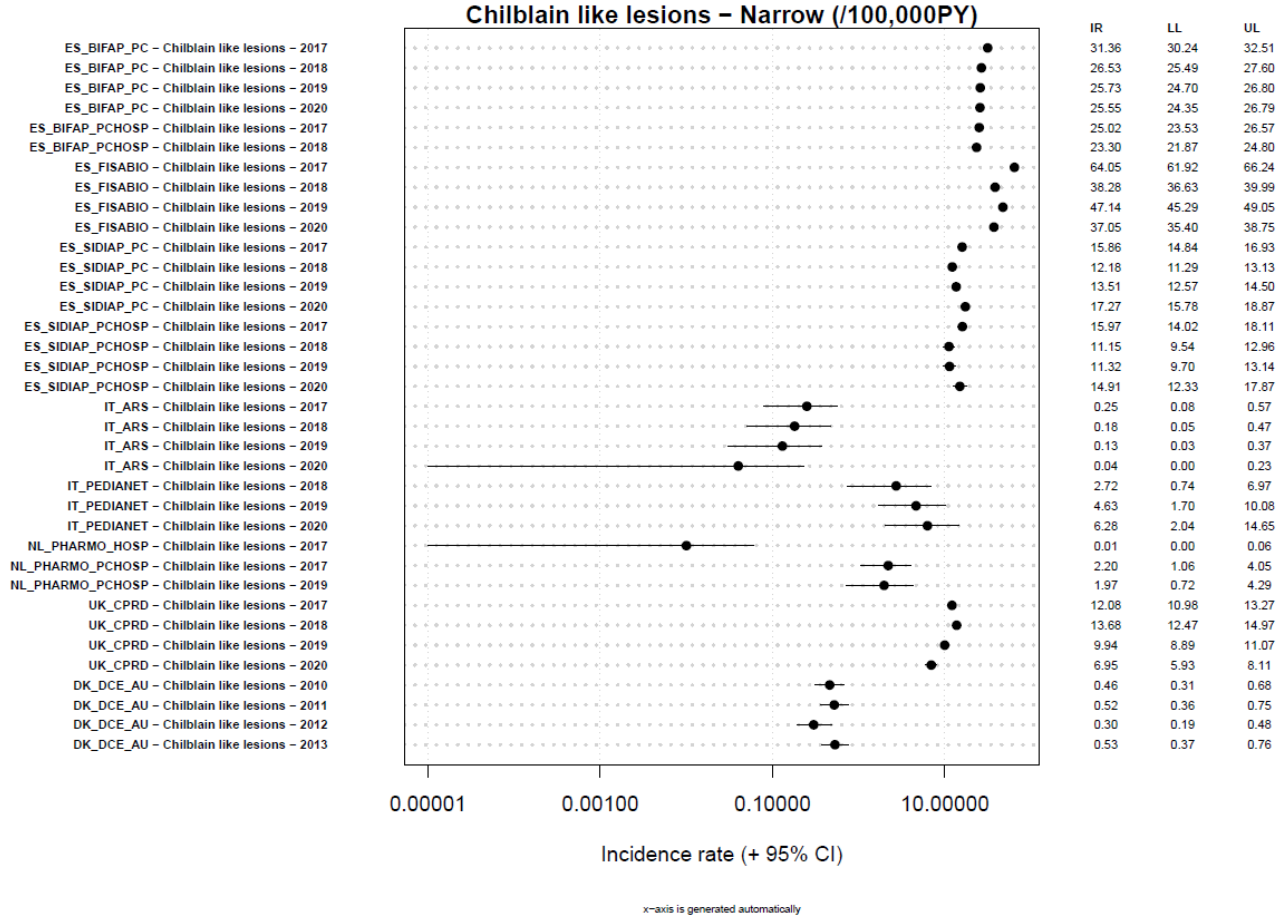


Figure 60 Incidence of chilblain like lesions by data source and calendar year

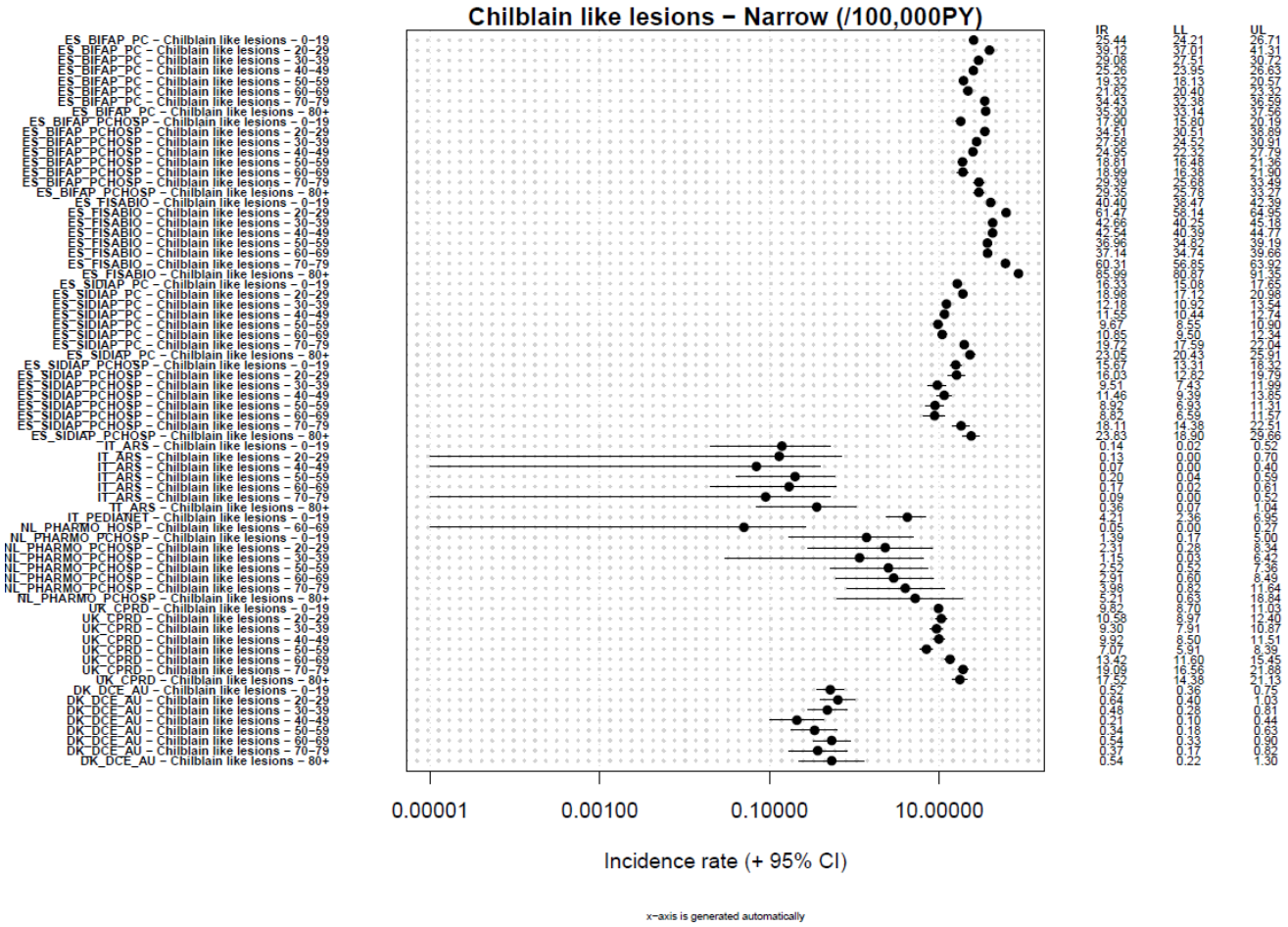


Figure 61 Incidence of chilblain like lesions by data source and age

We did not find reference rates on this condition.

10.3.1.23 Anosmia, Ageusia

Definition and codes in Zenodo
Egbers, T, Willame, C, Belbachir, L, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Anosmia & Ageusia (1.0). Zenodo. https://doi.org/10.5281/zenodo.5236687

Anosmia is lack of smell and ageusia is lack of taste.

The IRs were lowest from ARS with 0.05/100,000 person-years (95% CI 0.01 – 0.35) and highest from FISABIO with 28.82/100,000 person-years (95%: 27.41-30.30) in 2017. Significantly higher rates were observed in the year 2020 in all databases, except CPRD (Figure 60). Rates increase with age but diminish in oldest age group (figure 63)

Hospital /specialist-based data sources (ARS, DCE-AU) may underestimate the incidence and recommend to use from primary care based data sources

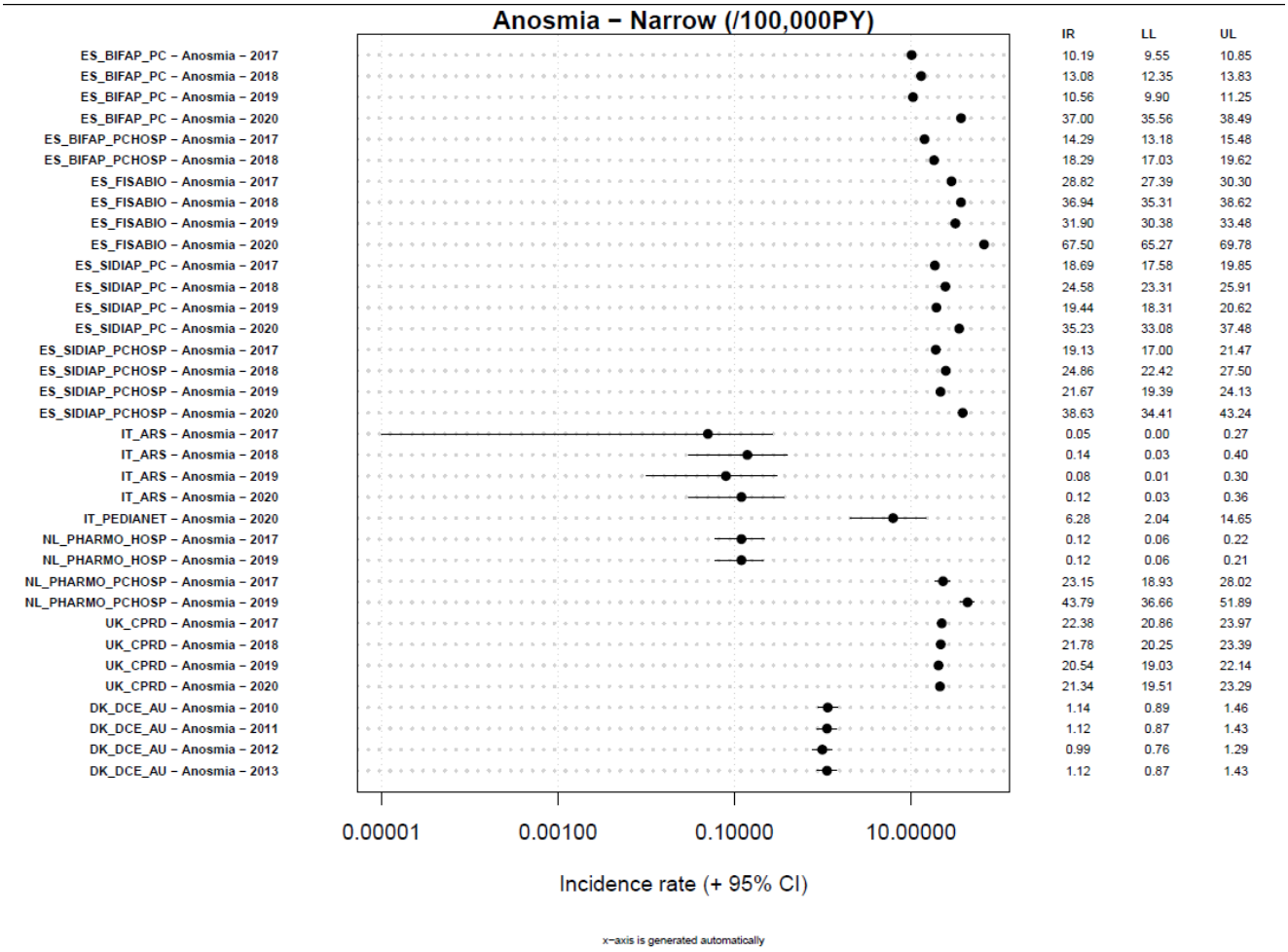


Figure 62 Incidence of anosmia/ageusia by data source and calendar year

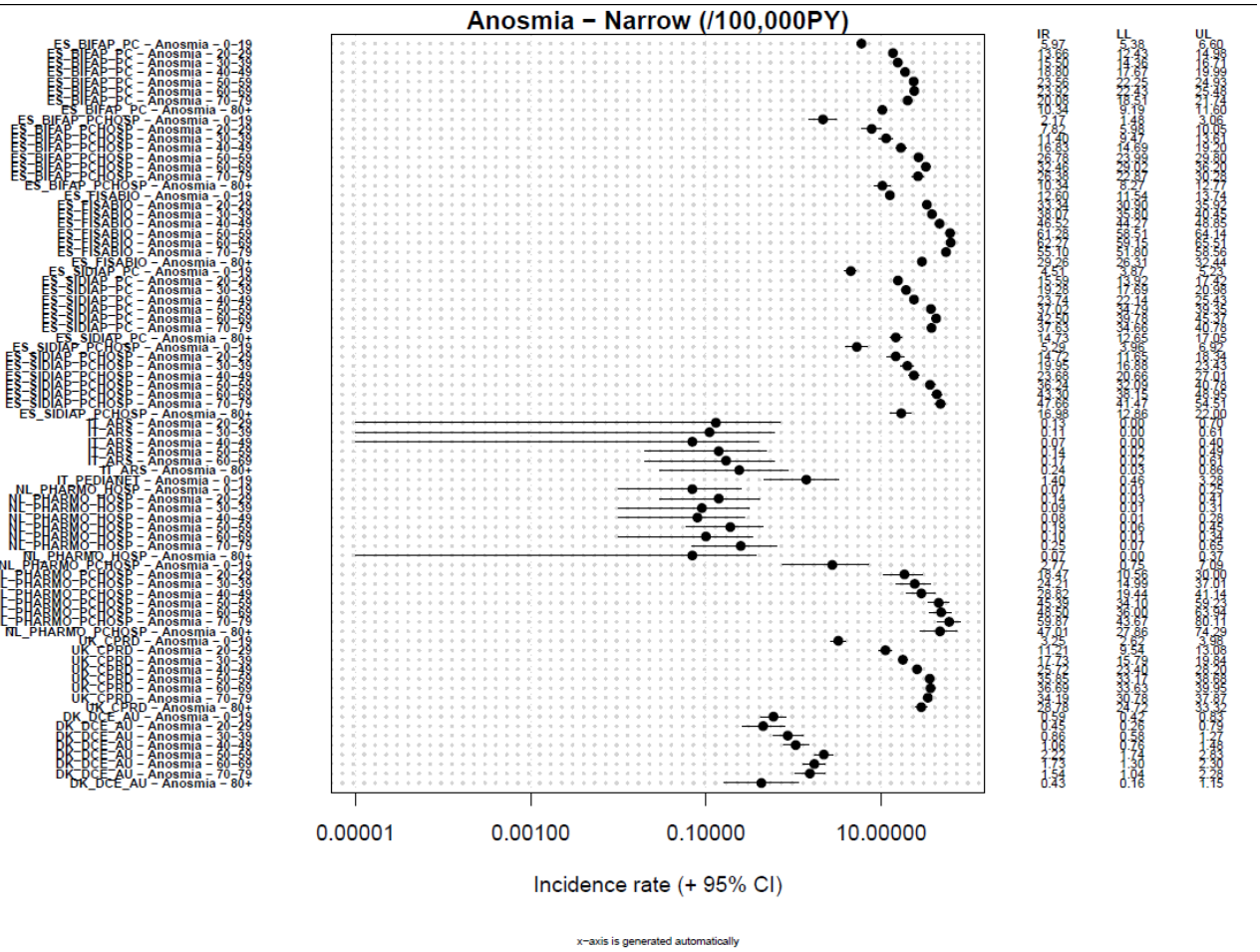


Figure 63 Incidence of anosmia/ageusia by data source and age (using broad definition)

10.3.1.24 Anaphylaxis

Definition and codes in Zenodo
Kelters, I, Willame, C, Belbachir, L, Souverein, P, Martín-Pérez, M, García-Poza, P, Durán, C, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Anaphylaxis (1.0). Zenodo. https://doi.org/10.5281/zenodo.5236723

Anaphylaxis is a serious systemic hypersensitivity reaction that is usually rapid in onset and may cause death. Severe anaphylaxis is characterized by potentially life-threatening compromise in breathing and/or the circulation and may occur without typical skin features or circulatory shock being present[54].

Anaphylaxis IRs vary between 1.54/100,000 person-years (95% CI 0.39 – 6.16) in PEDIANET to 24.63/100,000 person-years (95% CI 23.31 – 26.03) in FISABIO (figure 64). The rates from CPRD, SIDIAP and BIFAP were significantly lower for 2020 with an IR of 13.14, 7.37 and 4.07/100,000 person-years, respectively, compared to the other years. The rate of anaphylaxis is lower in the elderly

[54] Rüggeberg, J. U., Gold, M. S., Bayas, J. M., Blum, M. D., Bonhoeffer, J., Friedlander, S., ... & Erlewyn-Lajeunesse, M. (2007). Anaphylaxis: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine, 25(31), 5675-5684. https://doi.org/10.1016/j.vaccine.2007.02.064

(figure 65). Li et al. reported rates much higher than ours starting around 75 in youngest ages and decreasing to 10/100,000 in eldest.

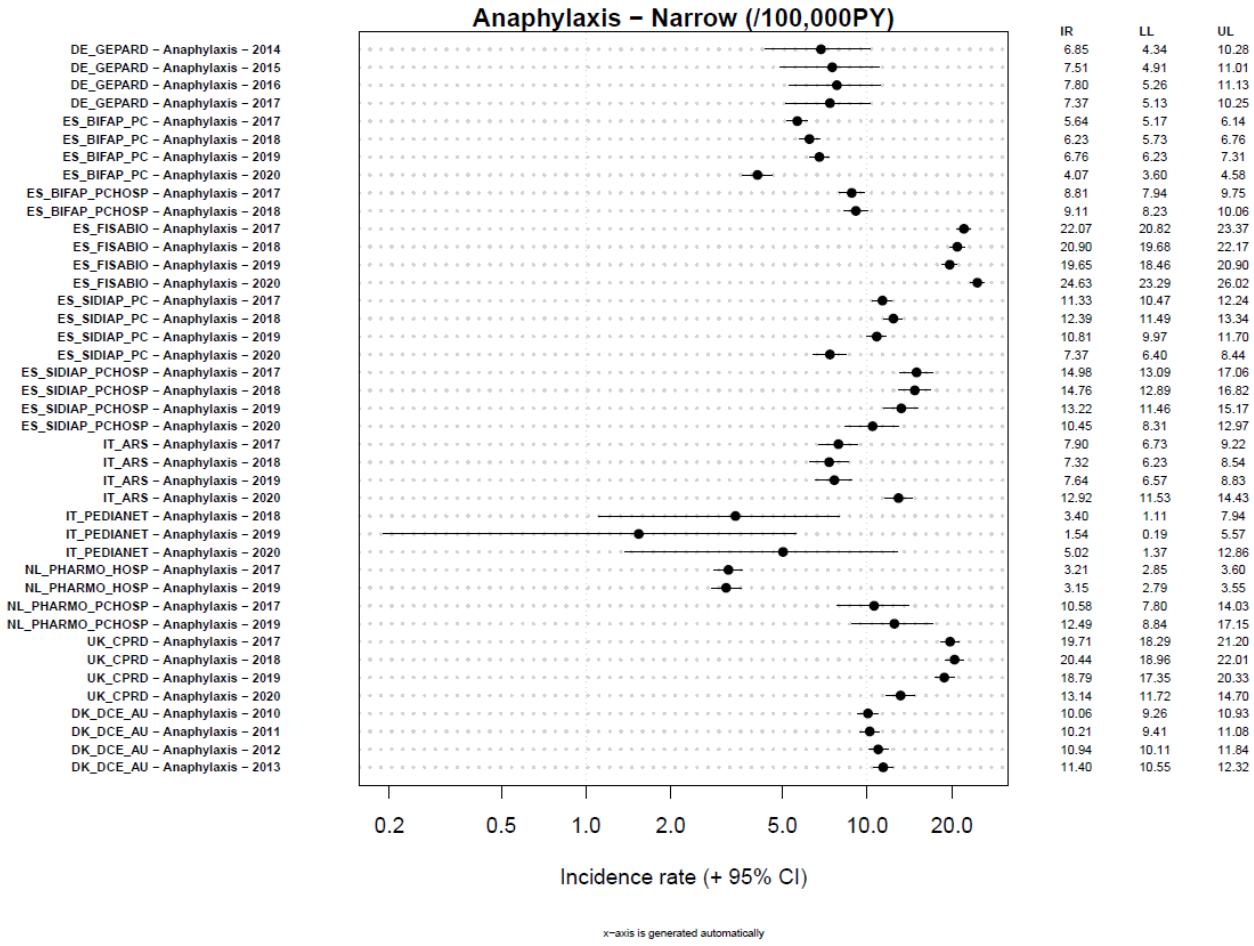


Figure 64 Incidence of anaphylaxis by data source and calendar year

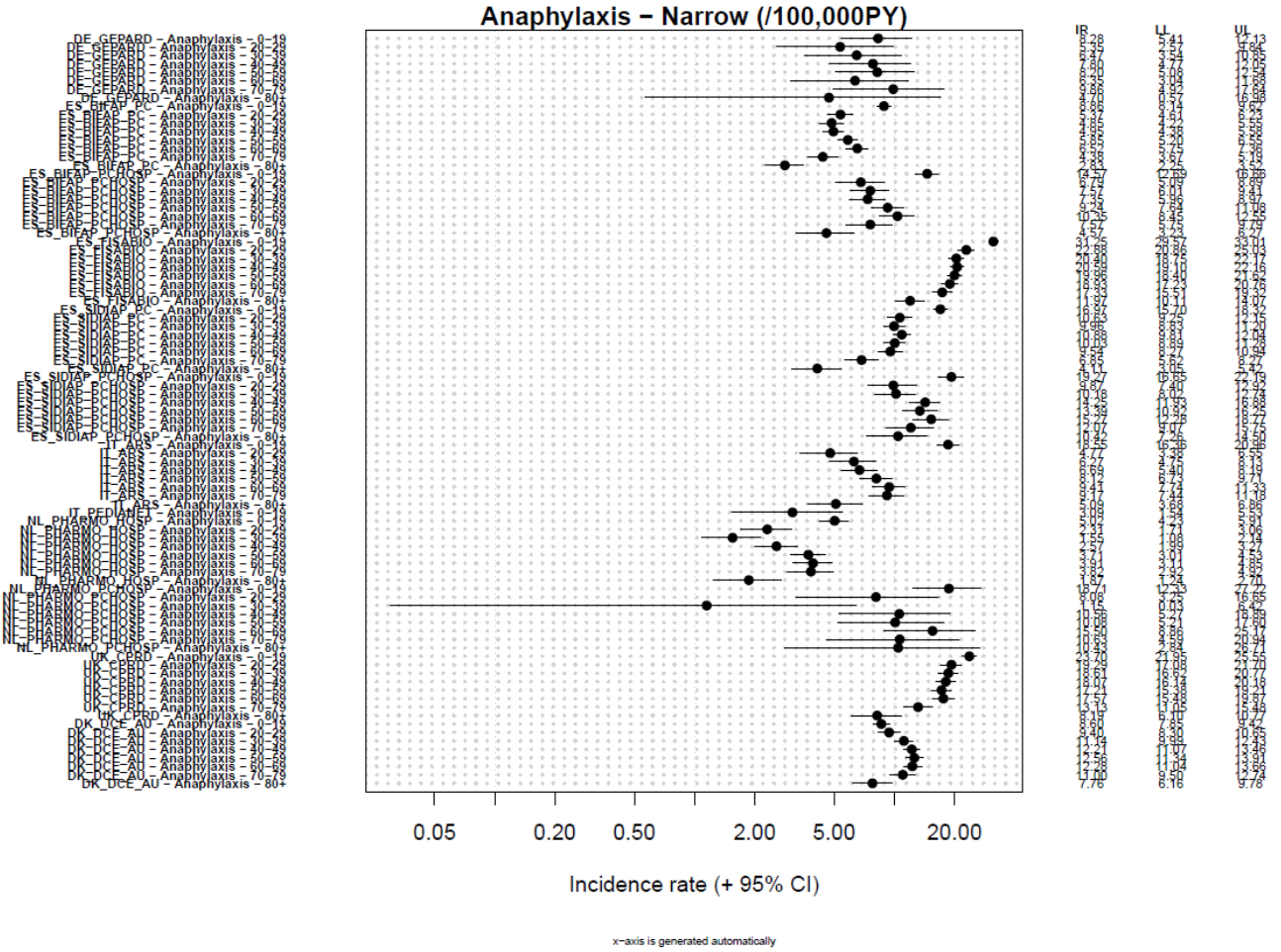


Figure 65 Incidence of anaphylaxis by data source and age

There are many reference rates based on all types of anaphylaxis, in most countries rates vary between 1-10/100,000 PY (see anaphylaxis companion guide Brighton Collaboration). These rates were consistent with what we observed.

10.3.1.25 Multisystem inflammatory syndrome

Definition and codes in Zenodo

Engelen, R, Belbachir, L, Dodd, C, Durán, C, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Multi-Inflammatory Syndrome (in Children) (1.0). Zenodo. https://doi.org/10.5281/zenodo.5236781

Multisystem inflammatory syndrome in children, also known as MIS-C, is a syndrome that appears to be a rare complication of COVID-19 in children. The syndrome is similar to incomplete Kawasaki disease (KD), a febrile illness of young childhood involving inflammation of the blood vessels that can result in coronary artery aneurysms. Symptoms often occur 1-6 weeks following infection with COVID-19 and may overlap with an acute respiratory COVID-19 presentation. Recently the syndrome is also found in adults. This is why we also assessed over the entire age range.

Events of multisystem inflammatory syndrome using narrow definition were observed in all databases and for all study years. Overall, IRs were low ranging between 0.30/100,000 person-years in SIDIAP in 2018 to 6.94/100,000 person-years in PEDIANET, which is pediatric population only (Figure 66).

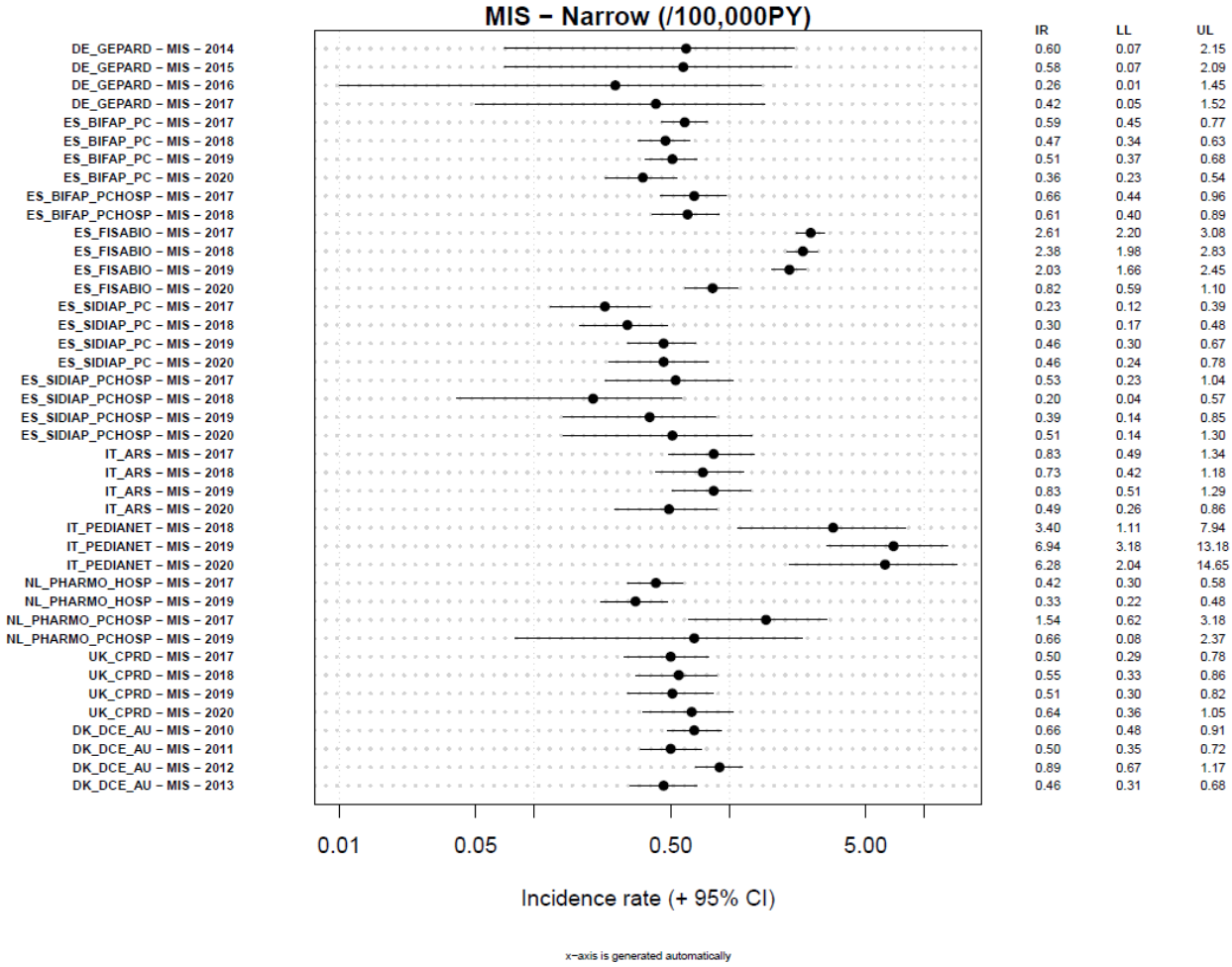


Figure 66 Incidence of multi-system inflammatory condition by data source and calendar year

Monthly incidence rates of MISC are available in Annex 2 and 5 (for narrow and broad definition respectively). Monthly rates in children are graphically depicted in Figure 67. Six databases provided data in 2020: ARS, PEDIANET, CPRD, BIFAP, SIDIAP (PC and PC_HOSP) and FISABIO. Incidence rate for PEDIANET showed potential reporting bias with no cases reported in March, May, July and September and higher rates in the following months, March, June, August and October. As shown in Figure 68, majority of the cases were observed in the [0-4] age group.

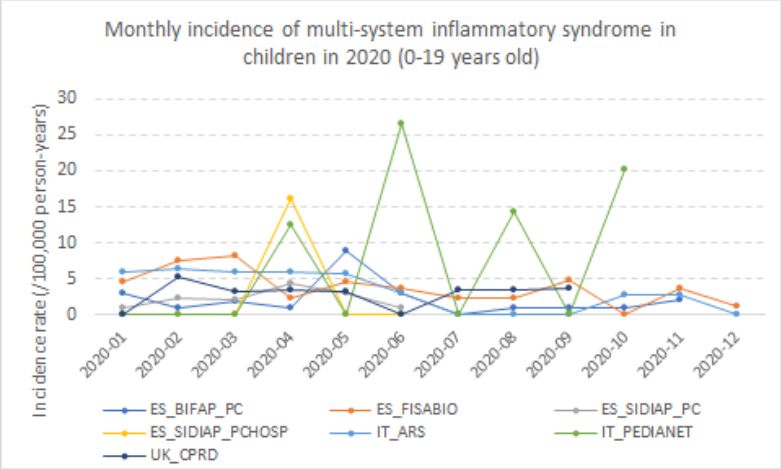


Figure 67 Monthly incidence of multi-system inflammatory syndrome in 2020 in children by data source

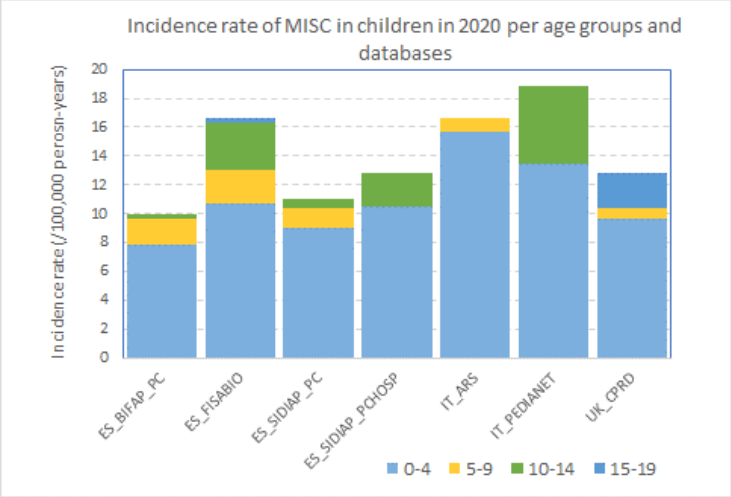


Figure 68 Incidence of multi-system inflammatory syndrome in 2020 in children by age and data source

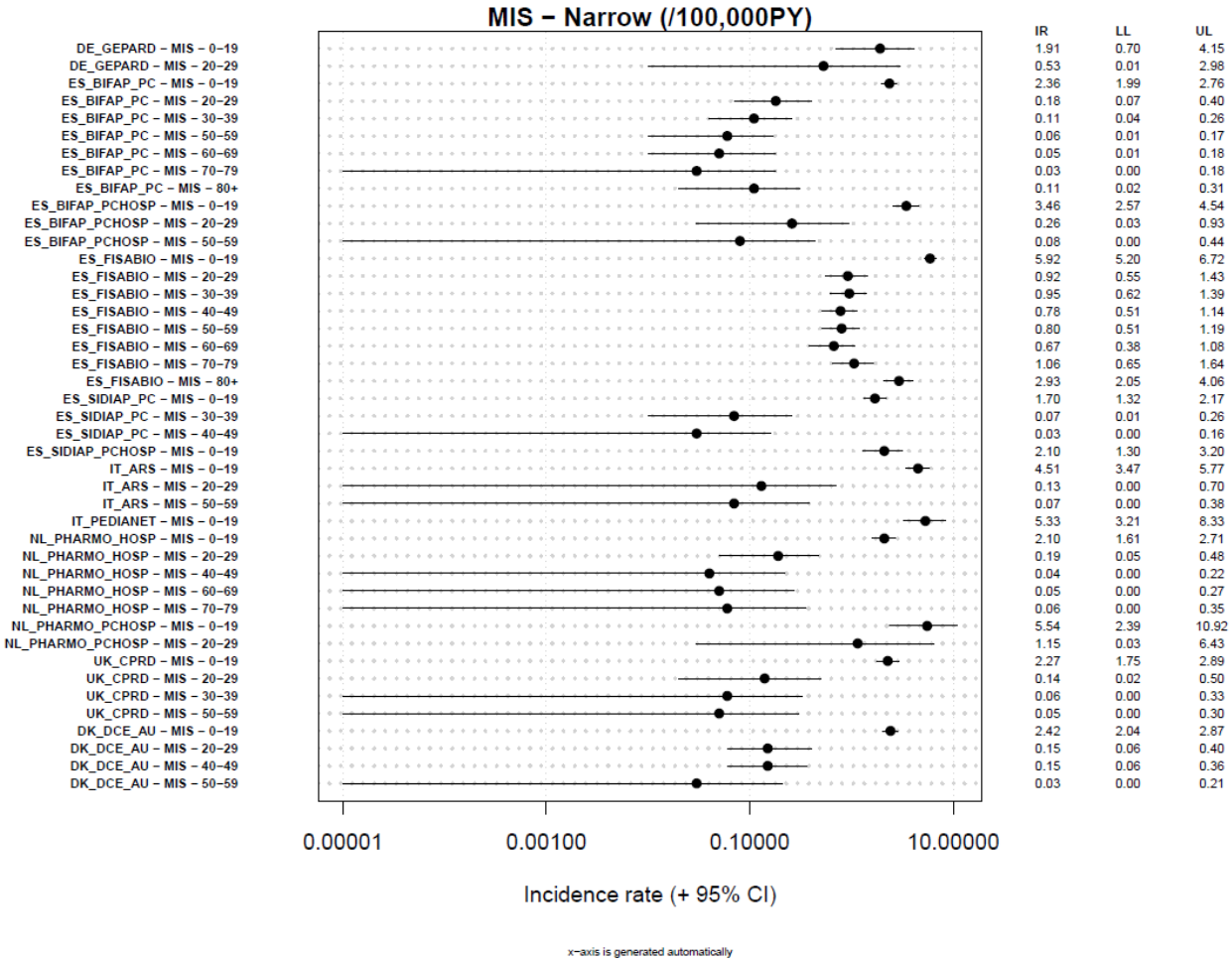


Figure 69 Incidence of multi-system inflammatory condition by data source and age

10.3.1.26 Death (any causes)

Deaths were identified from all participating databases, except in PEDIANET (pediatric) and PHARMO. The incidence rates of death were significantly higher in ARS. A clear pattern of increased incidence rates with age was observed across databases.

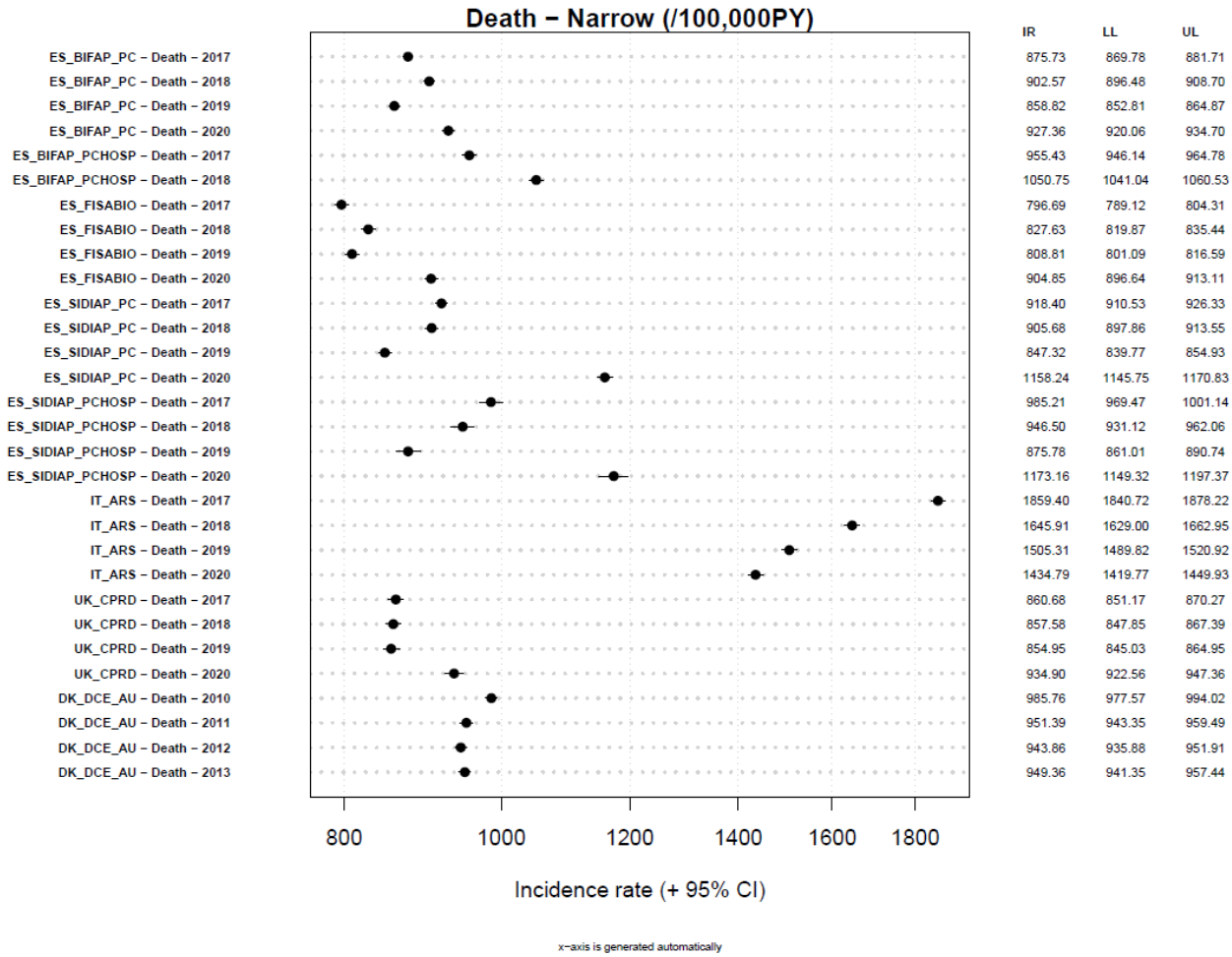


Figure 70 Incidence of death by data source and calendar year

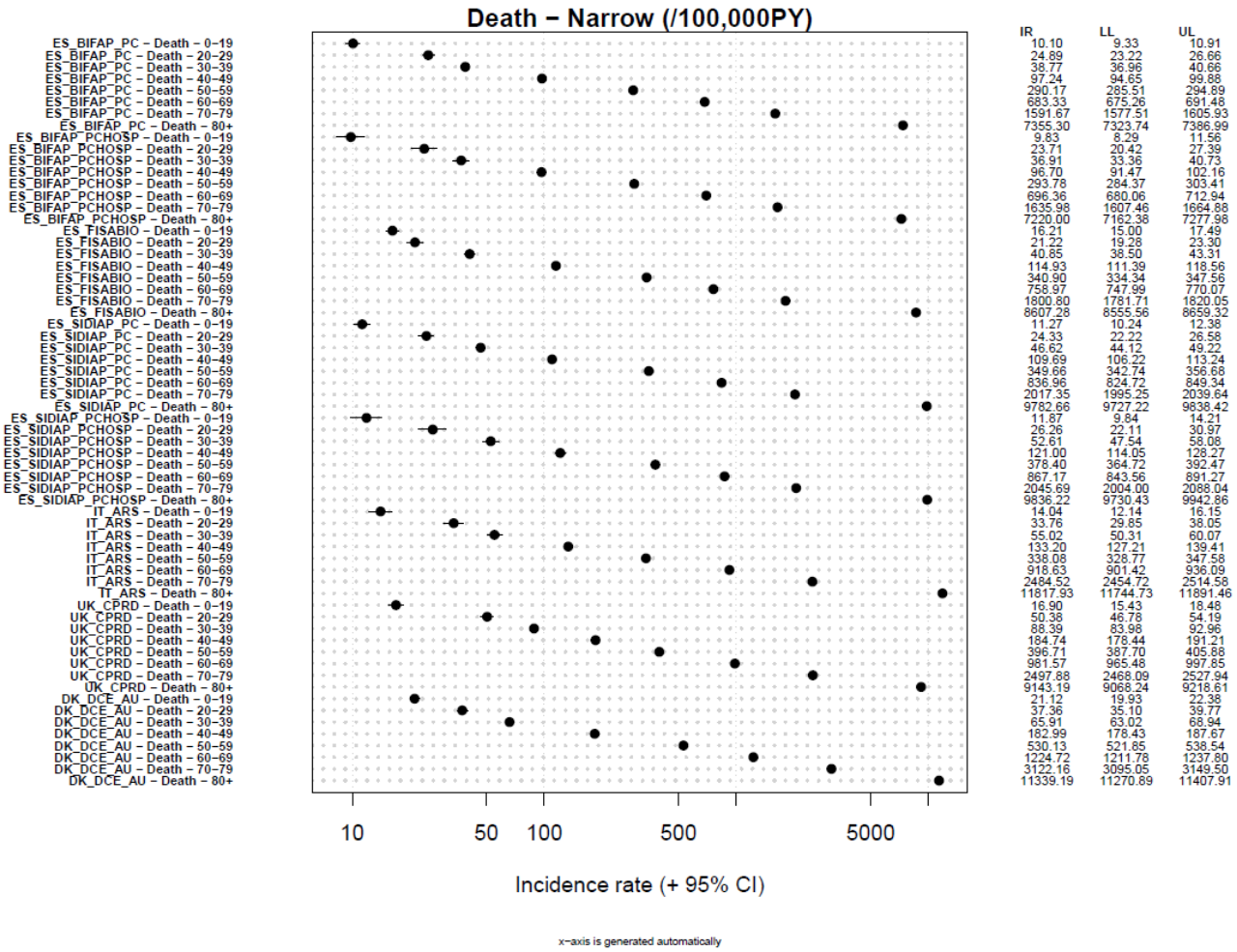


Figure 71 Incidence of death by data source and age

10.3.1.27 Sudden death

Zenodo definition
Sturkenboom, MCJM, Belbachir, L, & Durán, C. (2021). ACCESS-Background rate of adverse events-definition –Sudden death (1.0). Zenodo. https://doi.org/10.5281/zenodo.5237497

The diagnosis and definition of sudden death are variable, but the generally recognized definition is based on the length of time between the onset of symptoms and death. The World Health Organization (WHO) definition of sudden death according to the International classification of diseases, version 10 (ICD-10) is death, non-violent and not otherwise explained, occurring less than 24 hours from the onset of symptoms[55].

For sudden death, we expect that this diagnosis may be underestimated in the data sources, if the causes of death are not well recorded. Higher rates were observed for SIDIAP and FISABIO, this could be

[55] International classification of diseases (ICD-10). Geneva, World Health Organization, 2005

explained by the use of an unspecified ICD-10 code (R99: Other ill-defined and unspecified causes of mortality) which is the most frequently identified code. As for death, no sudden death could be extracted from PHARMO.

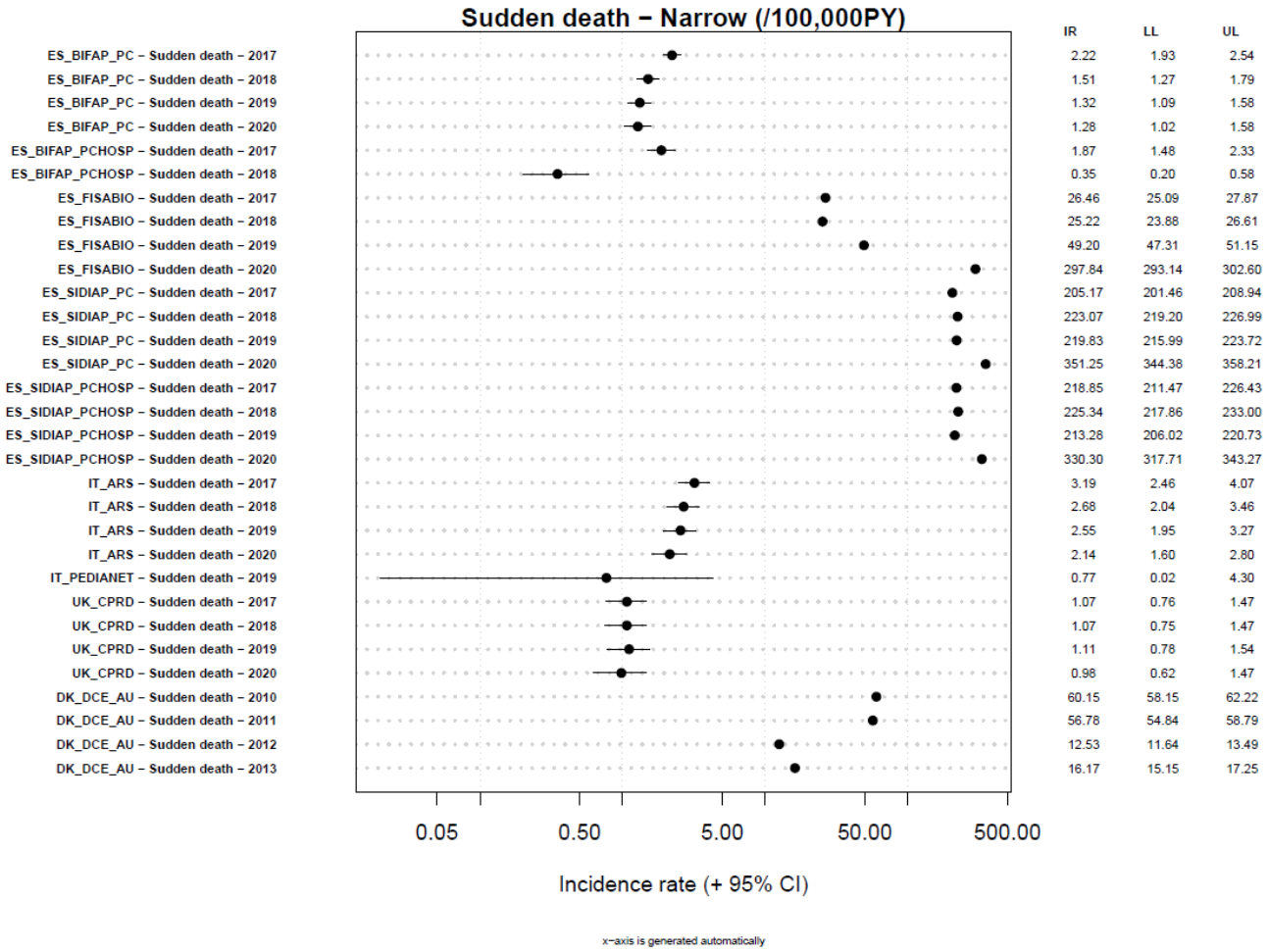


Figure 72 Incidence of sudden death by data source and calendar year

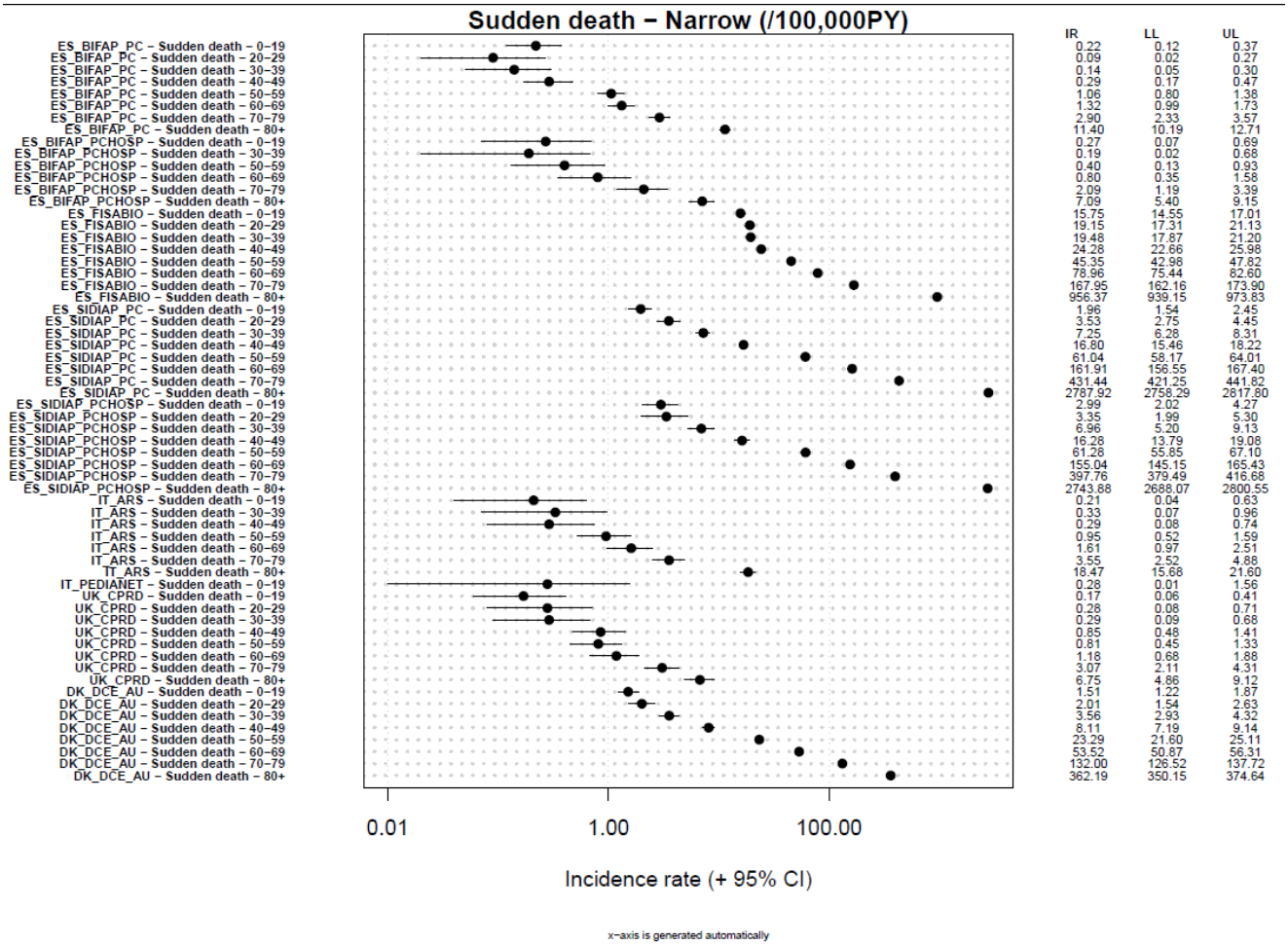


Figure 73 Incidence of sudden death by data source and age

10.3.1.28 COVID-19

Definition and codes in Zenodo
van Wijngaarden, P, Belbachir, L, Dodd, C, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition –Vaccine Associated Enhanced covid disease (1.0). Zenodo. https://doi.org/10.5281/zenodo.5237304

ARS, PEDIANET, CPRD and SIDIAP could provide data on COVID-19 in 2020. In Italy, a COVID registry is available which allowed to identify confirmed diagnosis. A similar registry was also available for the regions covered by the BIFAP database. Data from BIFAP COVID registry were available. For FISABIO, a database with laboratory confirmed SARS-CoV2 cases, and the primary care and hospital databases were used to identify COVID-19 cases. However, for both databases (BIFAP and FISABIO) the definition of COVID-19 according to severity level could not be generated for this report and, therefore they are not presented in Figure 73. For the other databases, COVID-19 cases were identified directly from the database using specific medical codes for coronavirus (see Annex 2_COVID_narrow for list of codes). An algorithm was built to classify the COVID-19 cases according to severity level. The algorithm used a combination of COVID-19 medical codes and

COVID-19 symptoms medical codes (symptoms are available in clinical form definition). Five severity levels were defined based on the WHO case definition and included the following (mutually exclusive):

- Level 1: recorded diagnosis (narrow definition) and no recording of hospitalisation;
- Level 2: hospitalization for COVID-19 with moderate symptoms;
- Level 3: hospitalization for COVID-19 with severe symptoms but without mechanical respiratory support;
- Level 4: hospitalization for COVID-19 with severe symptoms and with mechanical respiratory support;
- Level 5: death due to COVID-19.

Figure 74 Incidence of COVID in 2020 by data source and age

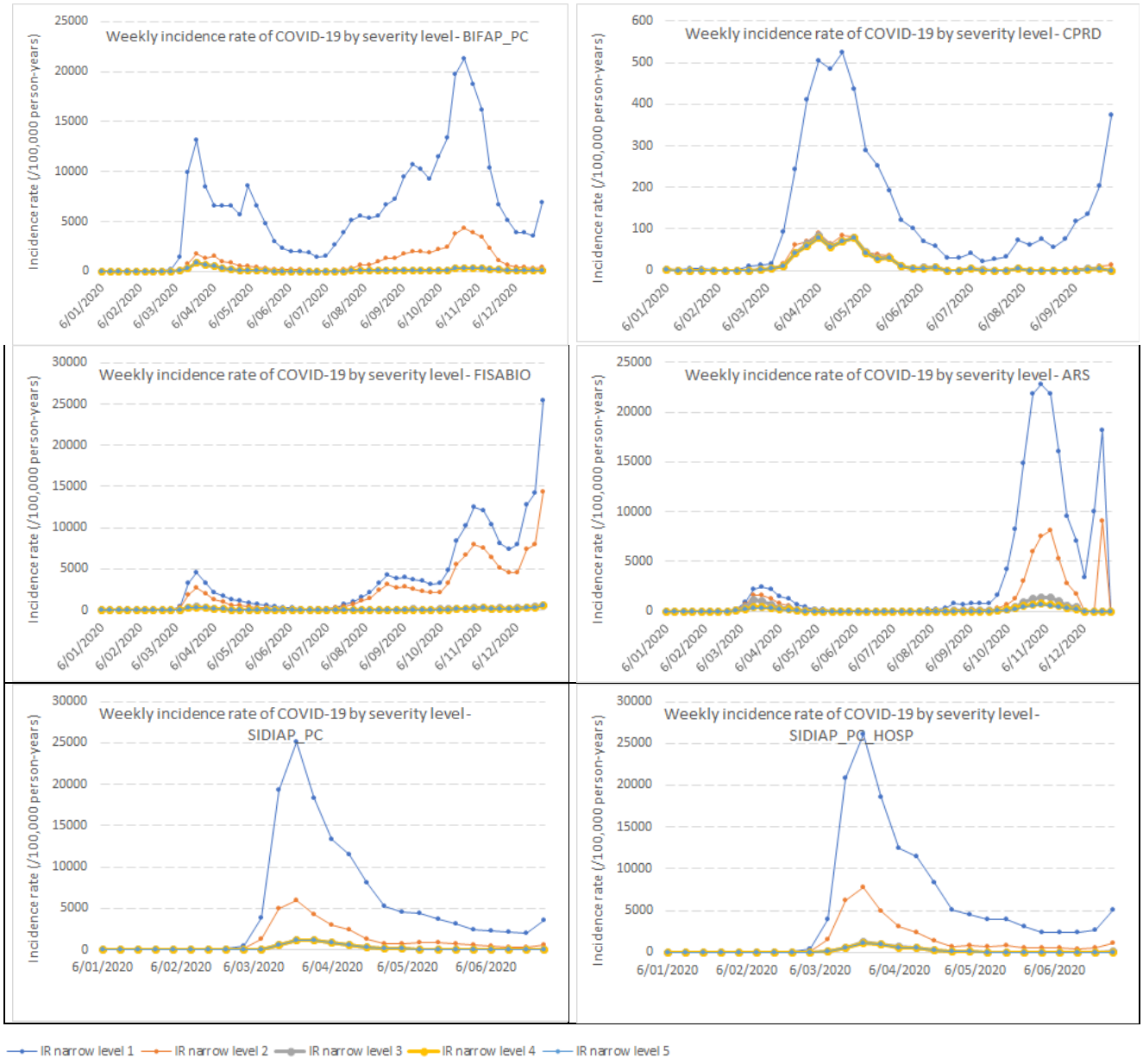


Figure 75 Weekly incidence of COVID-19 for each data source

10.3.1.29 Thrombosis with Thrombocytopenia (TTS)

Part of the coagulation disorder definition
van Wijngaarden, P, Willame, C, Durán, C, Belbachir, L, Souverein, P, Martín-Pérez, M, García-Poza, P, & Sturkenboom, MCJM. (2021). ACCESS-Background rate of adverse events-definition – Generalized Convulsions (Version 1). Zenodo. https://doi.org/10.5281/zenodo.5236092

The incidence rate for the co-occurrence of thrombosis and thrombocytopenia (within 10 days before or after distance) were also computed. Events of thrombosis were defined as VTE (DVT & PE), Arterial (CAD narrow & Ischemic Stroke), VTE or Arterial, CVST (broad). Rates for all four types of

thrombosis were computed with and without the co-occurrence of thrombocytopenia. Rates of thrombosis with thrombocytopenia were extremely low, very few cases of CVST with TP could be identified from databases with hospitalisation data.

10.3.1.29.1 VTE with and without TP

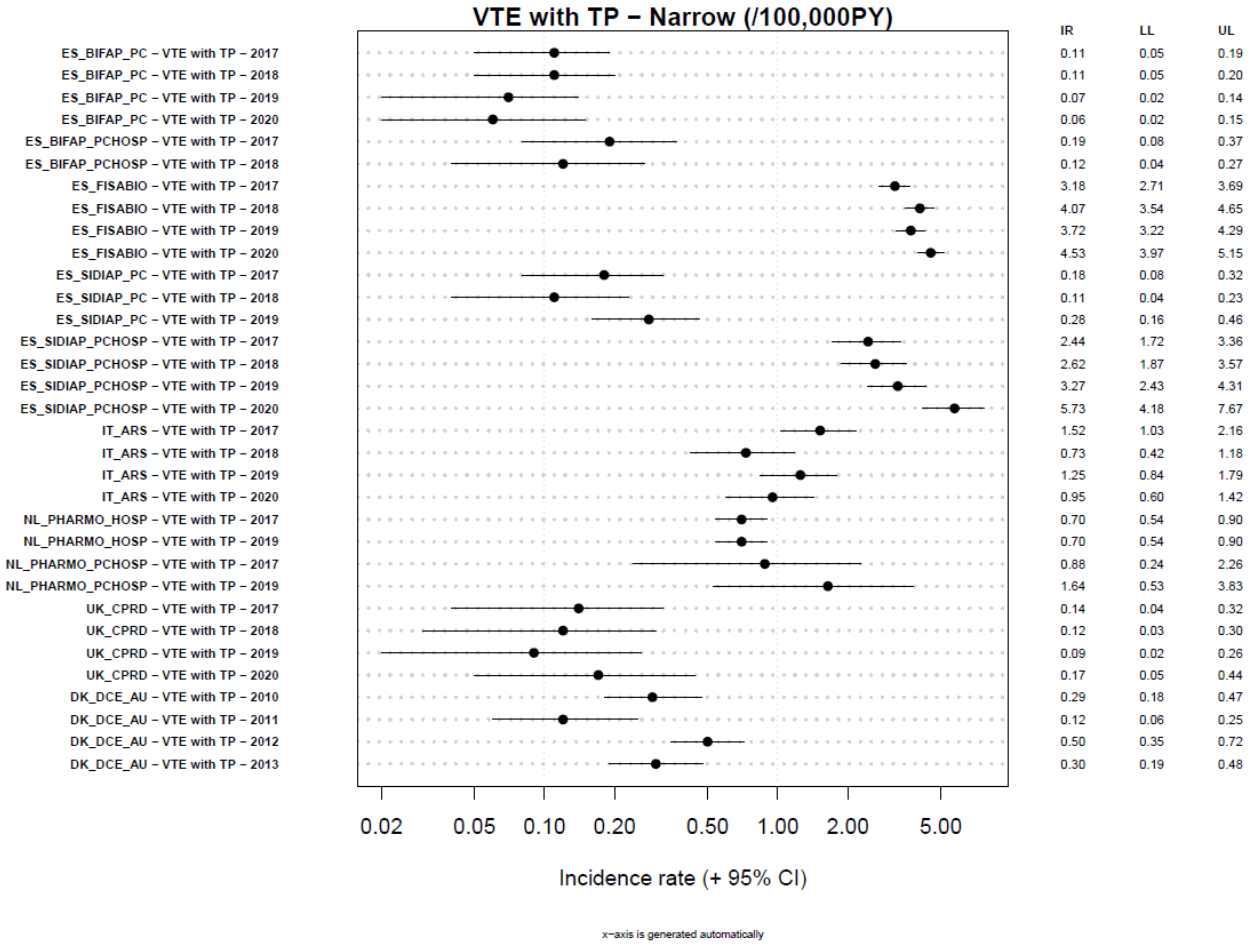


Figure 76 Incidence of VTE with TP by data source and calendar year

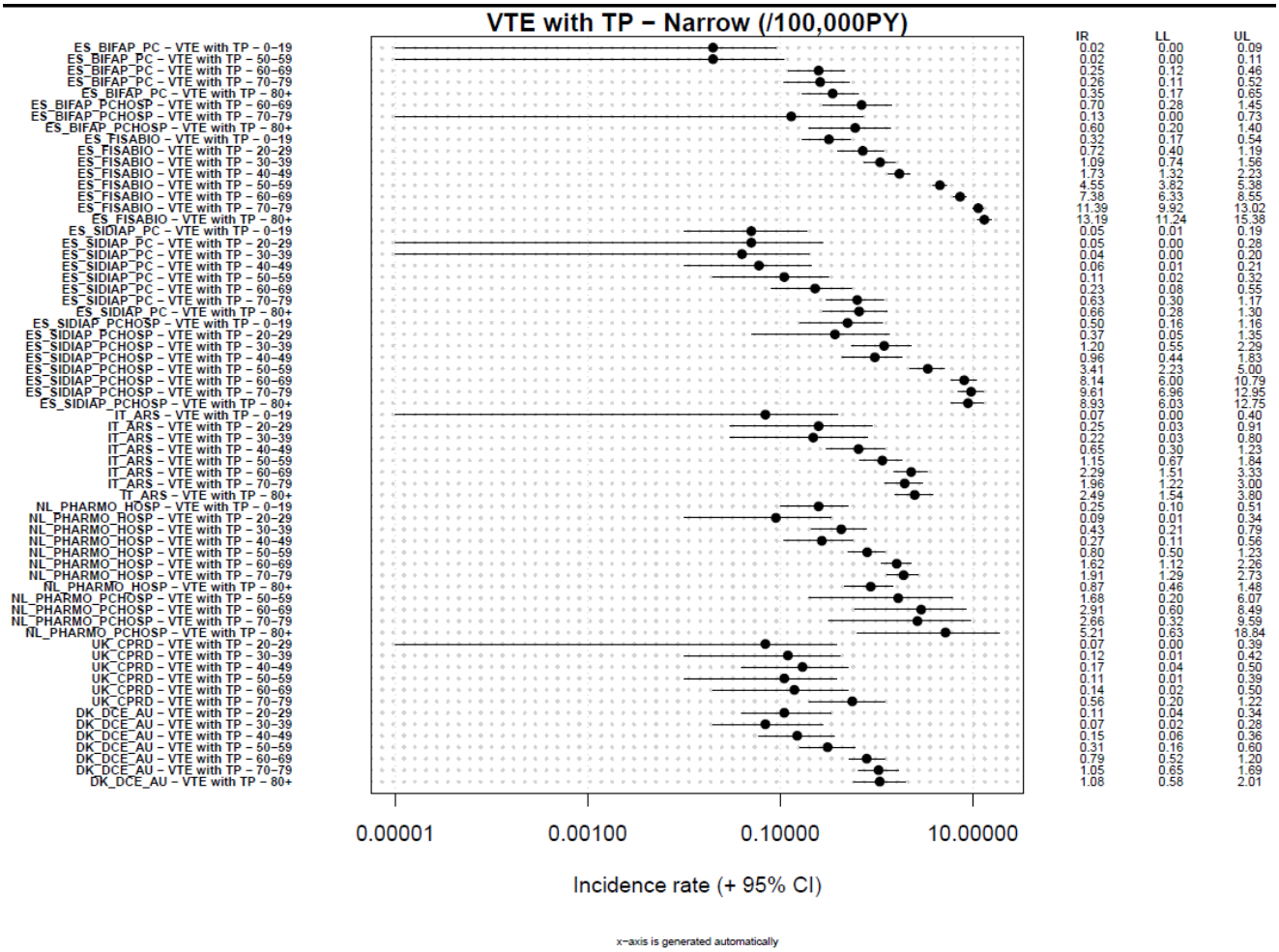


Figure 77 Incidence of VTE with TP by data source and age

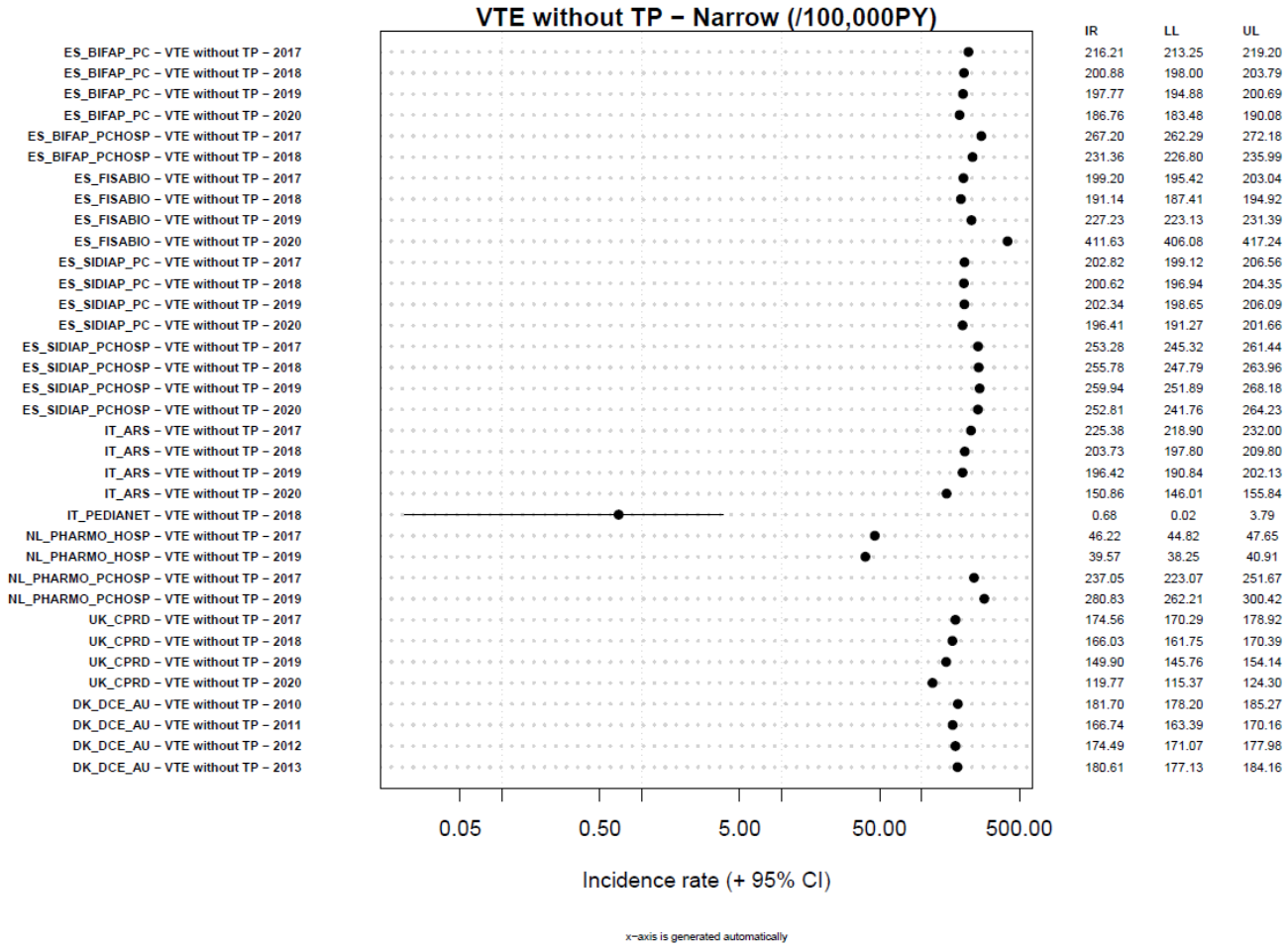


Figure 78 Incidence of VTE without TP by data source and calendar year

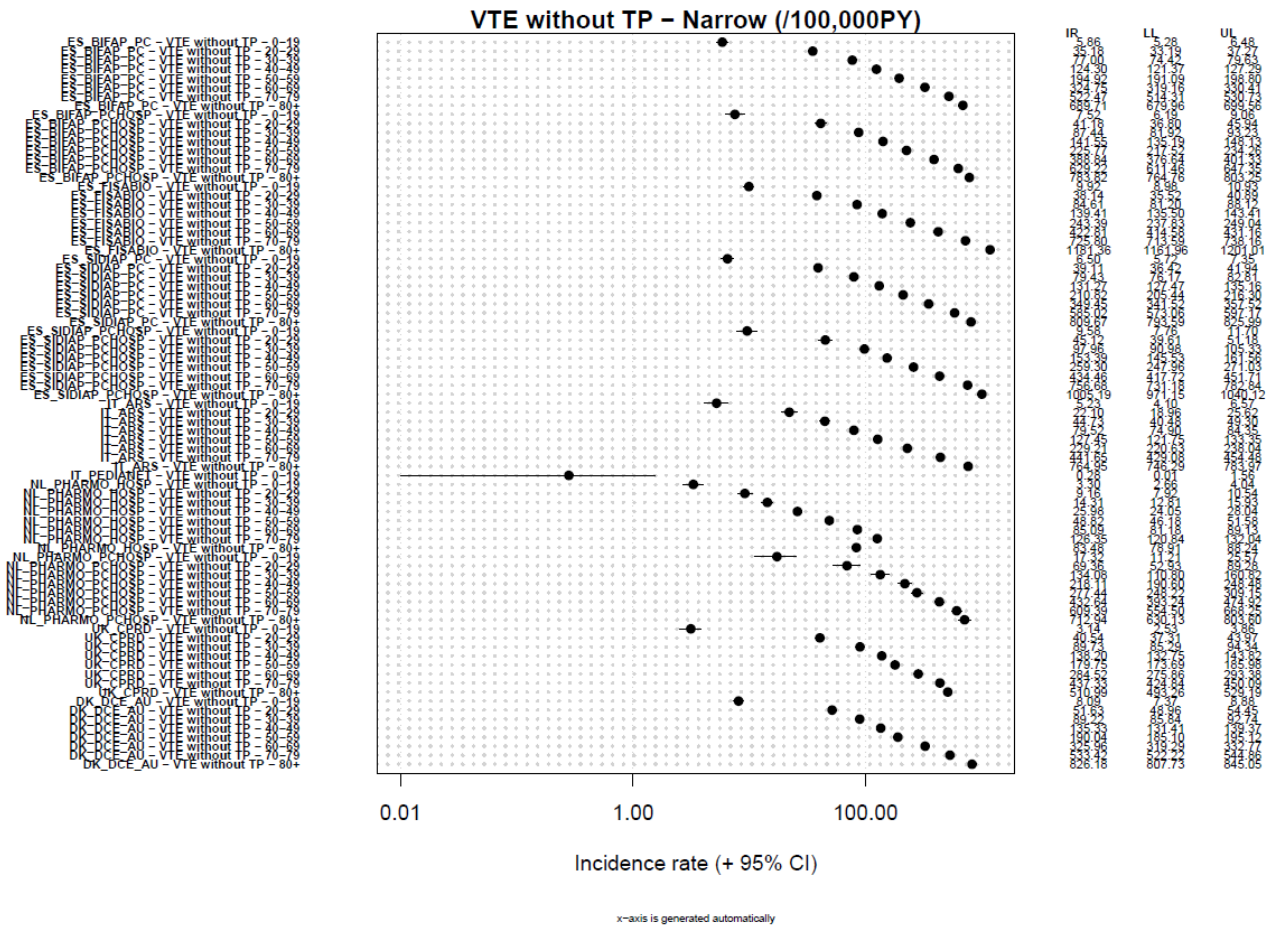


Figure 79 Incidence of VTE without TP by data source and age

10.3.1.29.2 CVST with and without TP

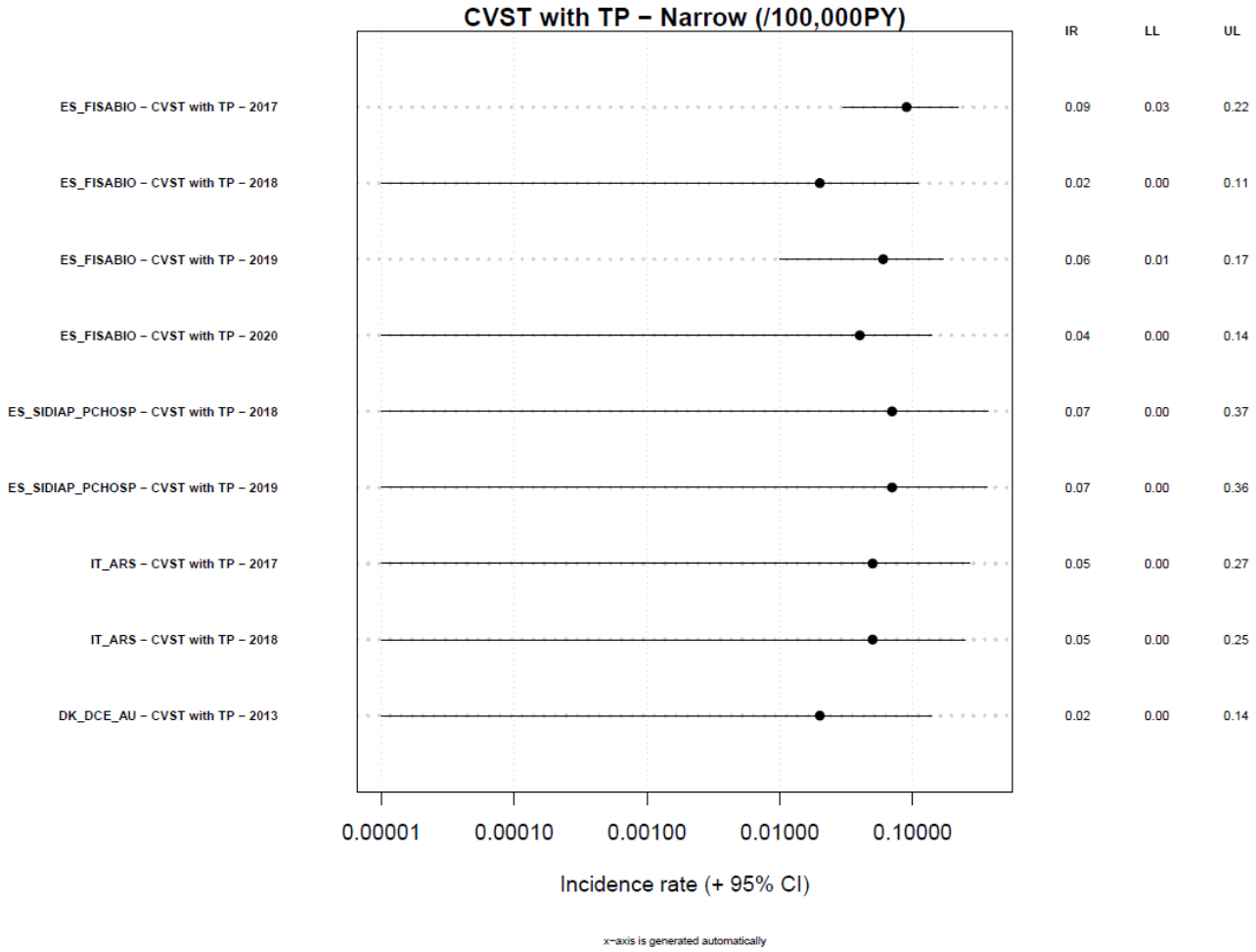


Figure 80 Incidence of CVST with TP by data source and calendar year

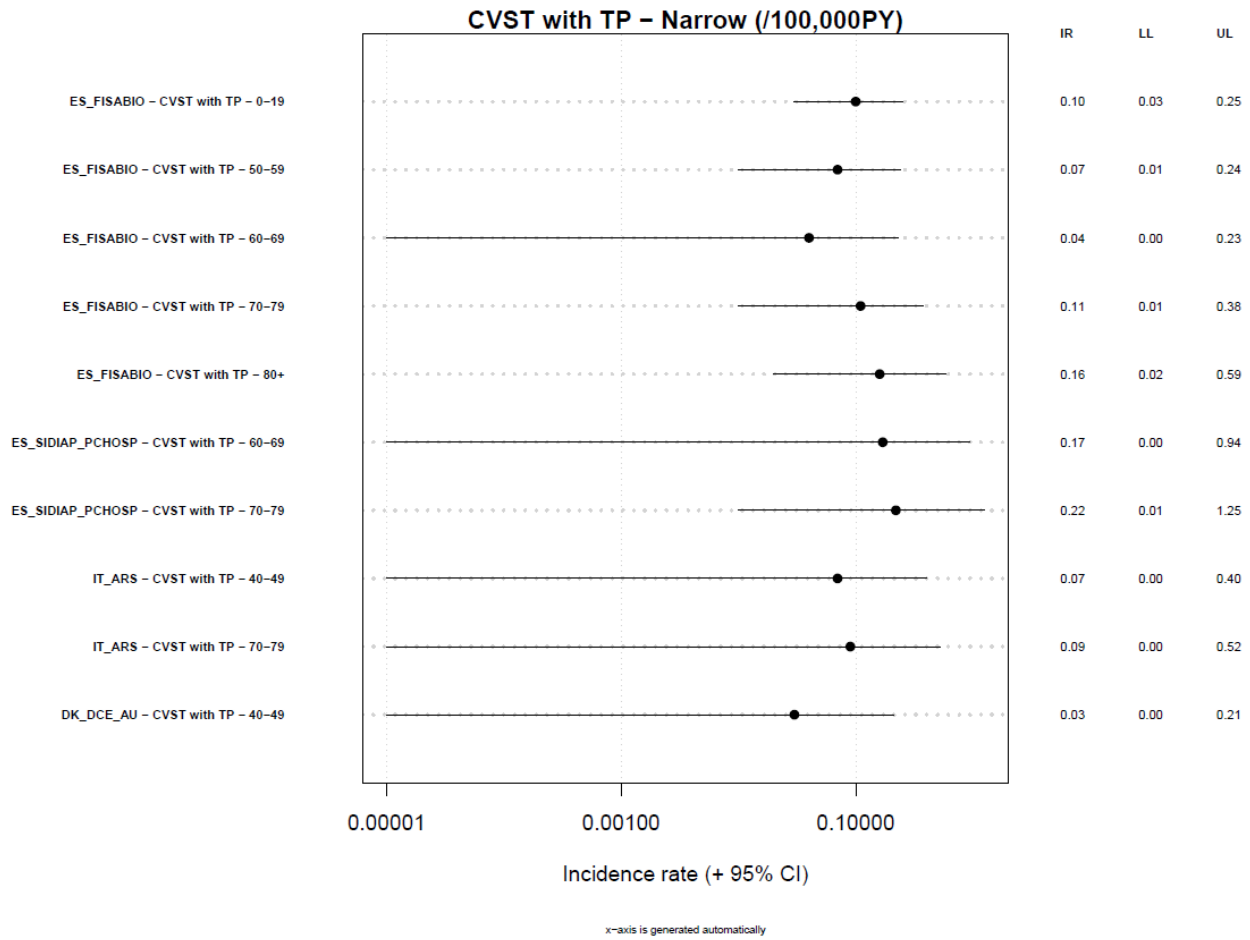


Figure 81 Incidence of CVST with TP by data source and age

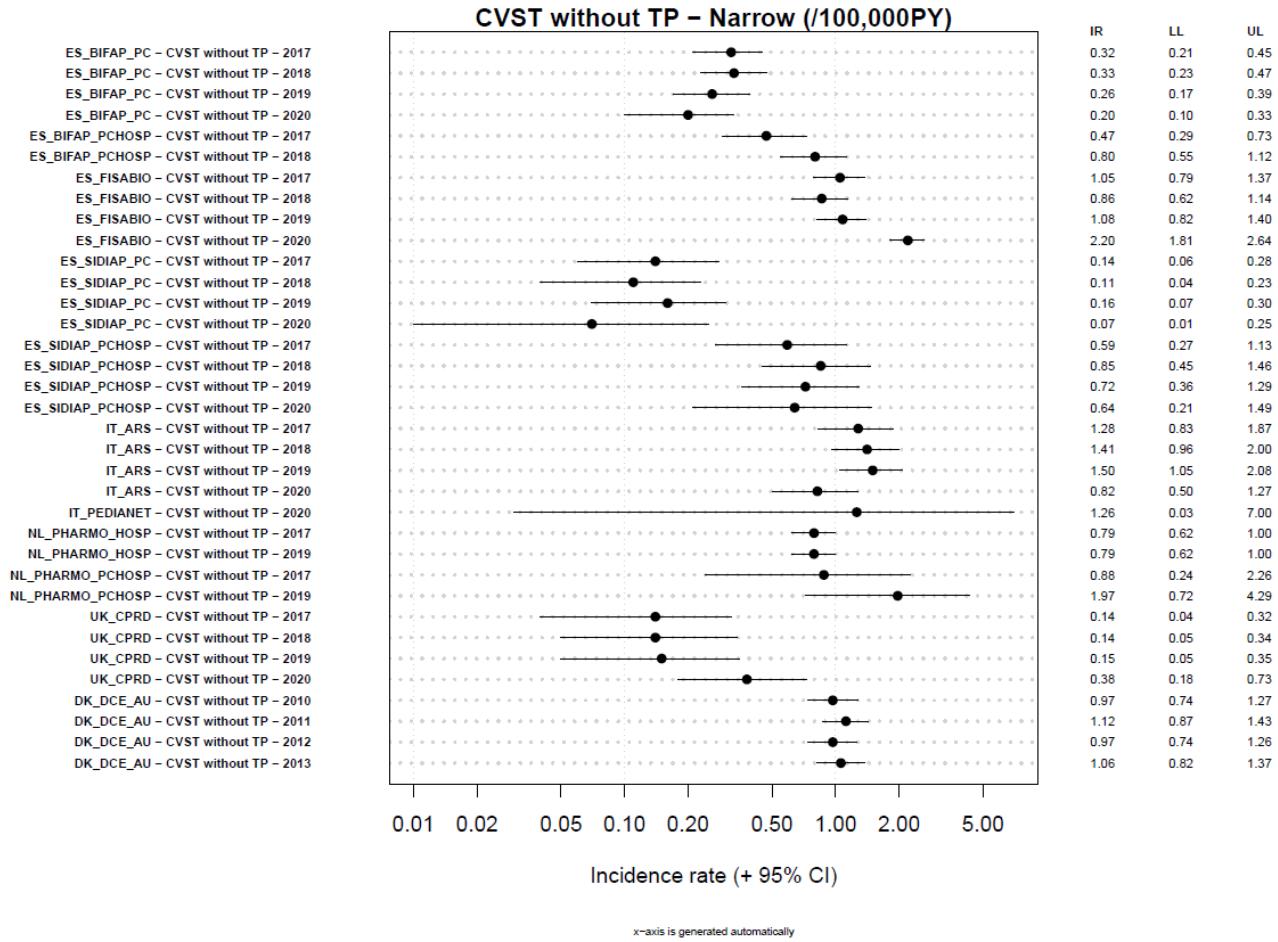


Figure 82 Incidence of CVST without TP by data source and calendar year

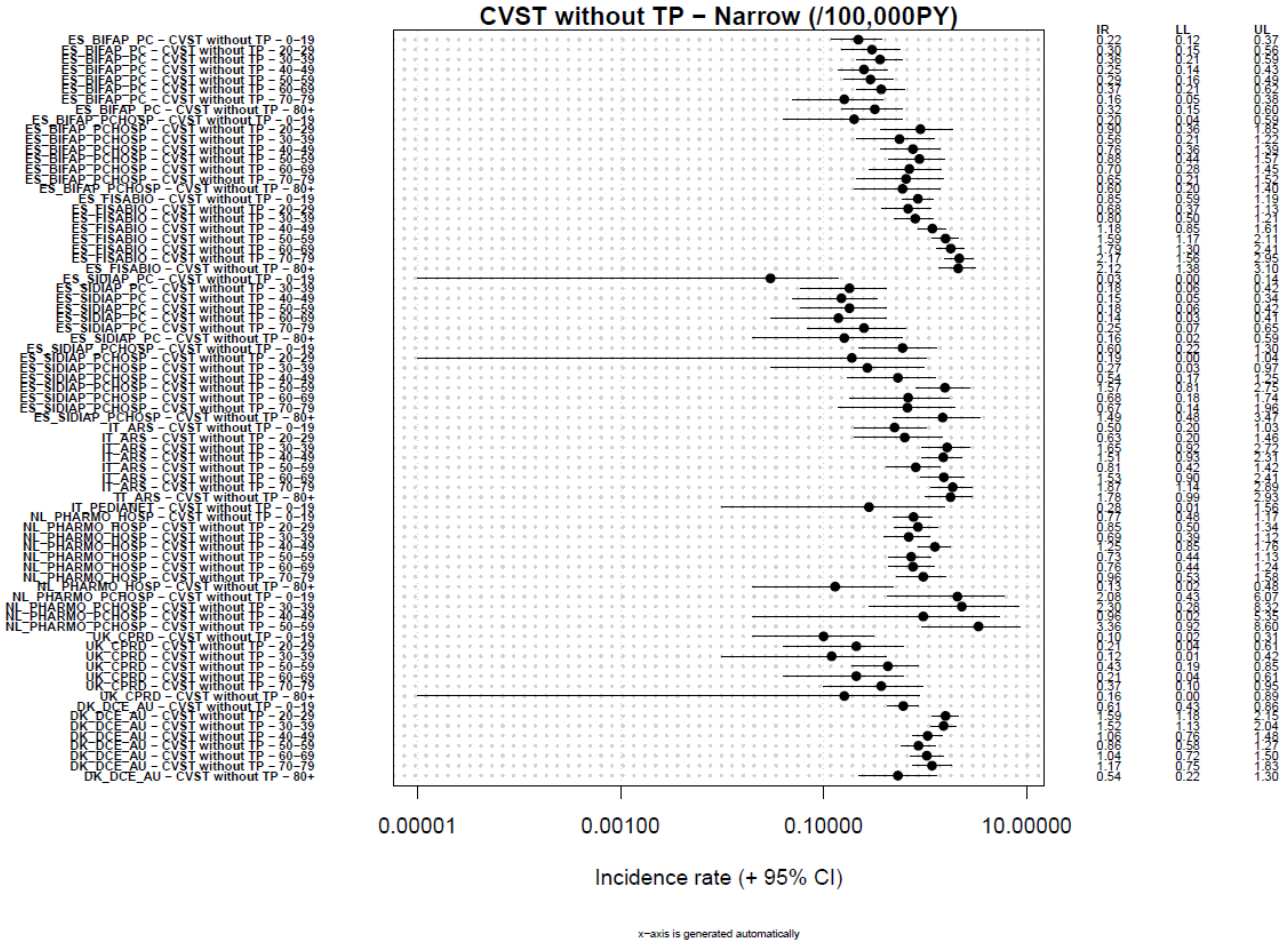


Figure 83 Incidence of CVST without TP by data source and age

10.3.1.29.3 Arterial thrombotic events with and without TP

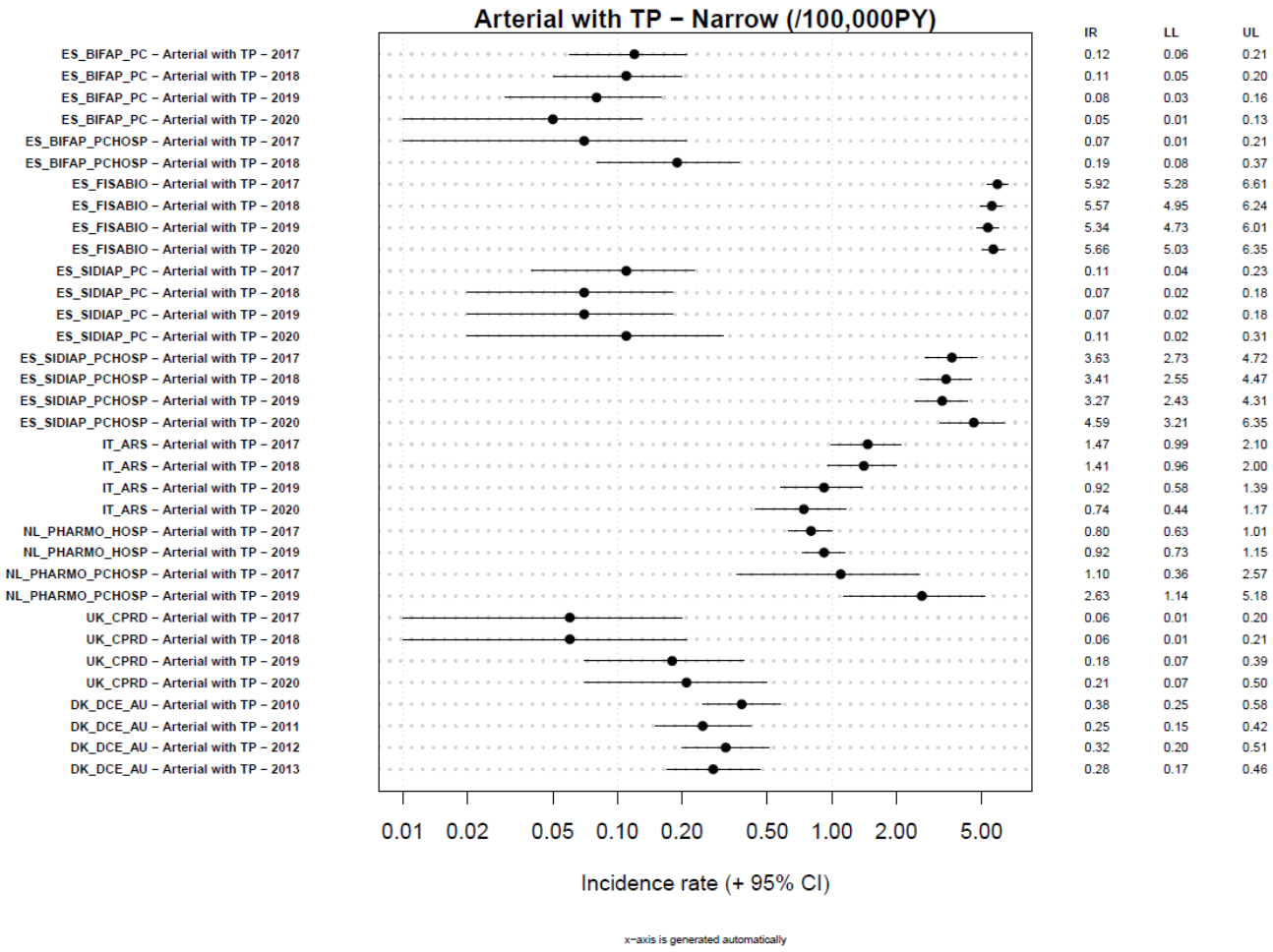


Figure 84 Incidence of Arterial thrombotic events with TP by data source and calendar year

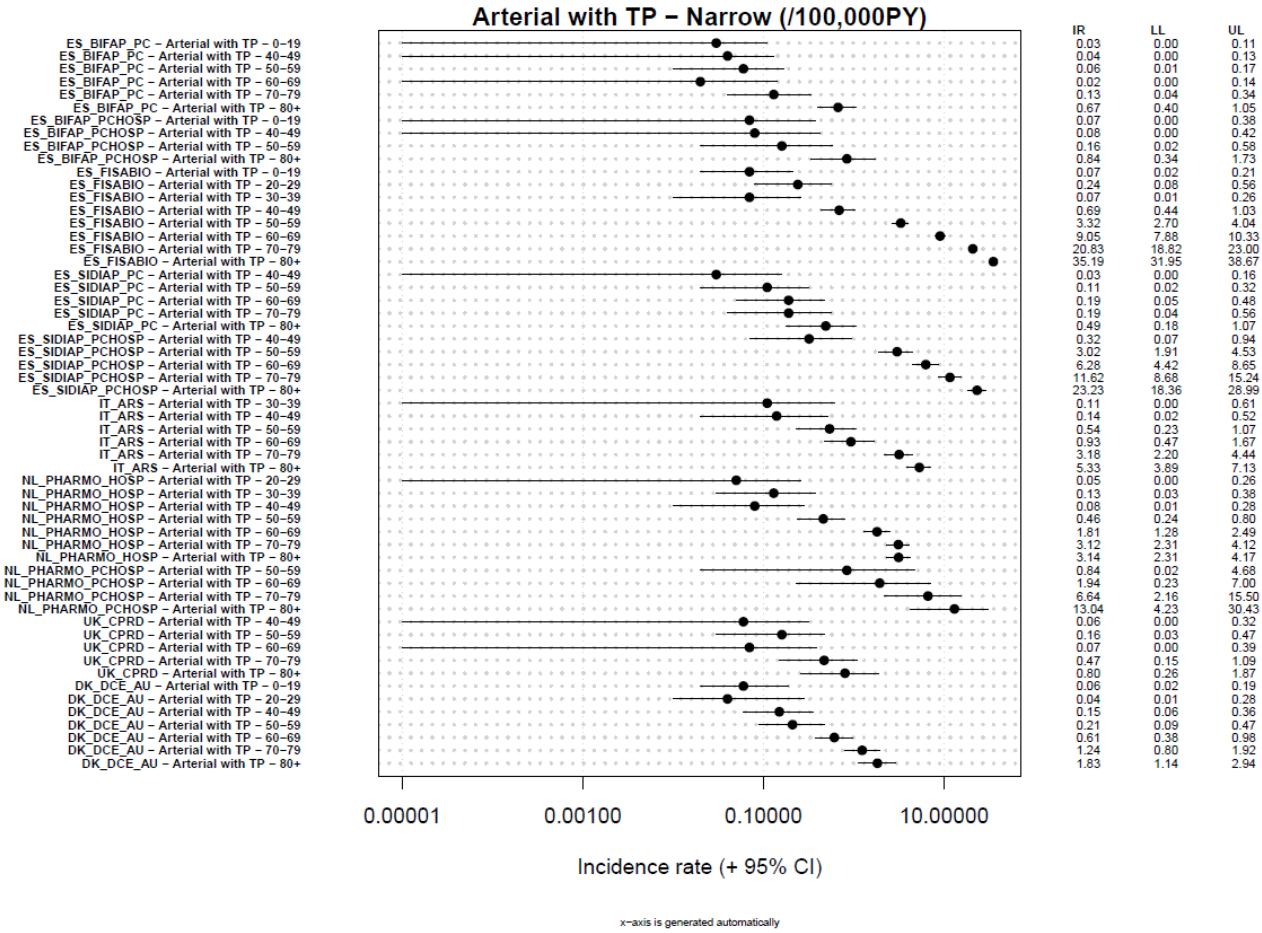


Figure 85 Incidence of Arterial thrombotic events with TP by data source and age

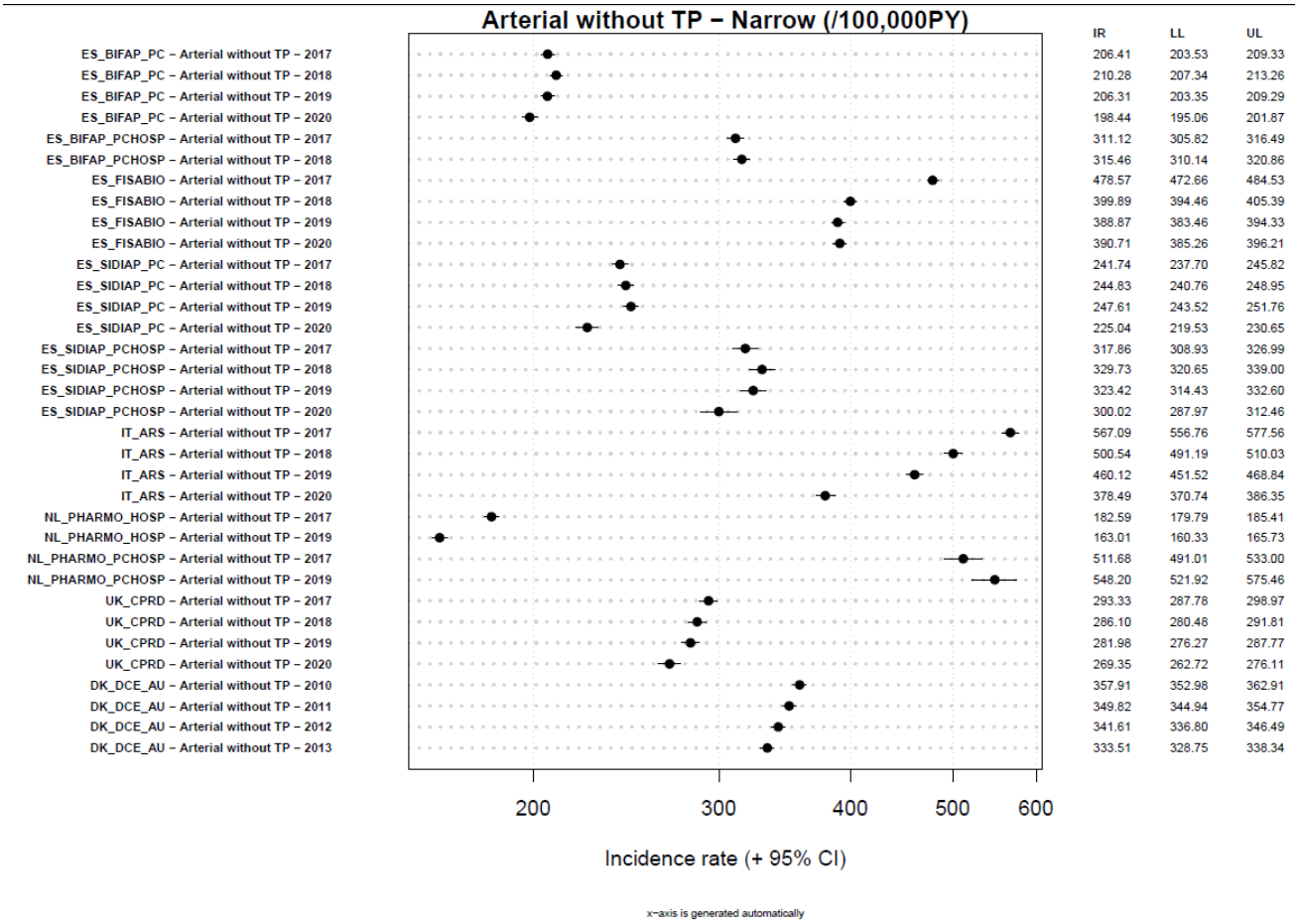


Figure 86 Incidence of Arterial thrombotic events without TP by data source and calendar year

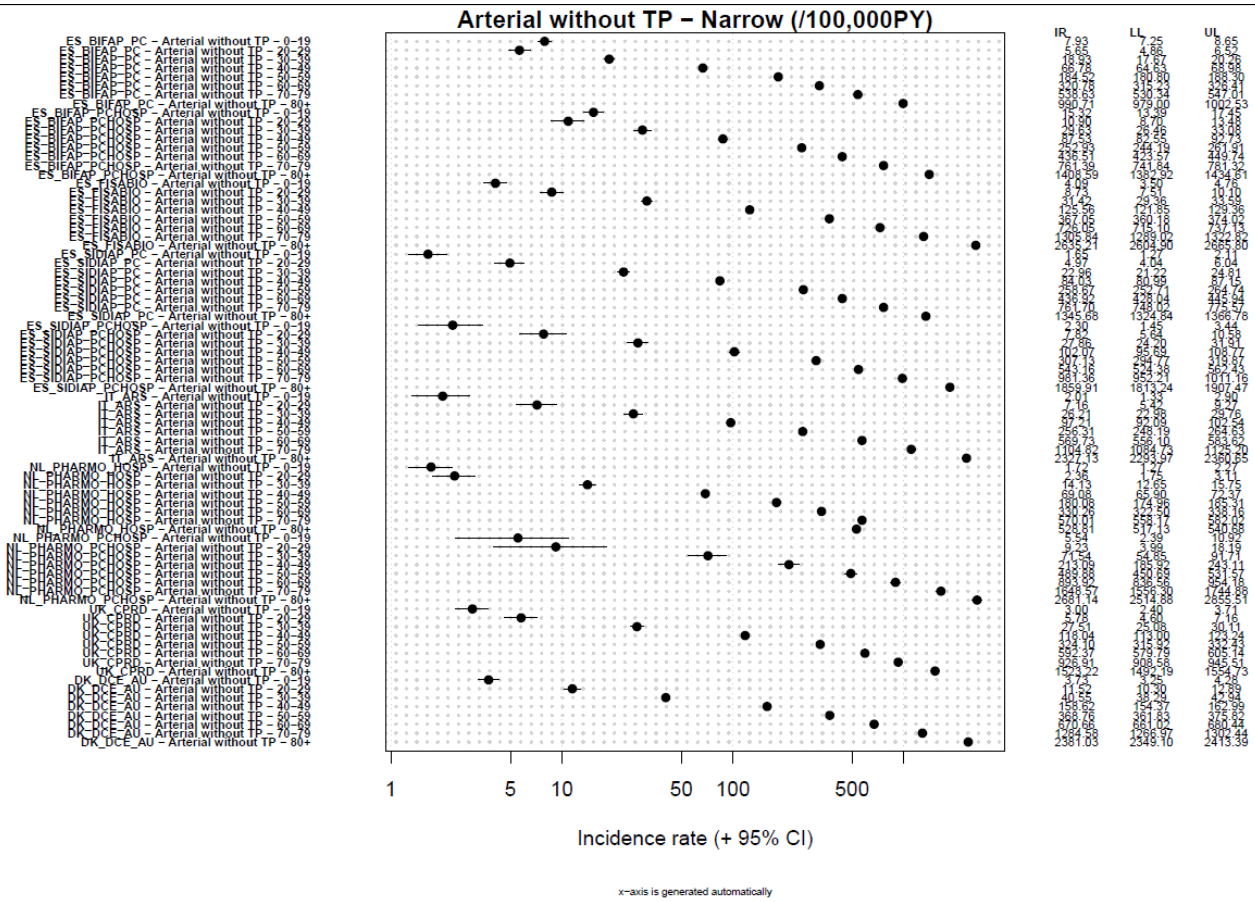


Figure 87 Incidence of Arterial thrombotic events without TP by data source and age

10.3.1.29.4 Arterial thrombotic events or VTE with and without TP

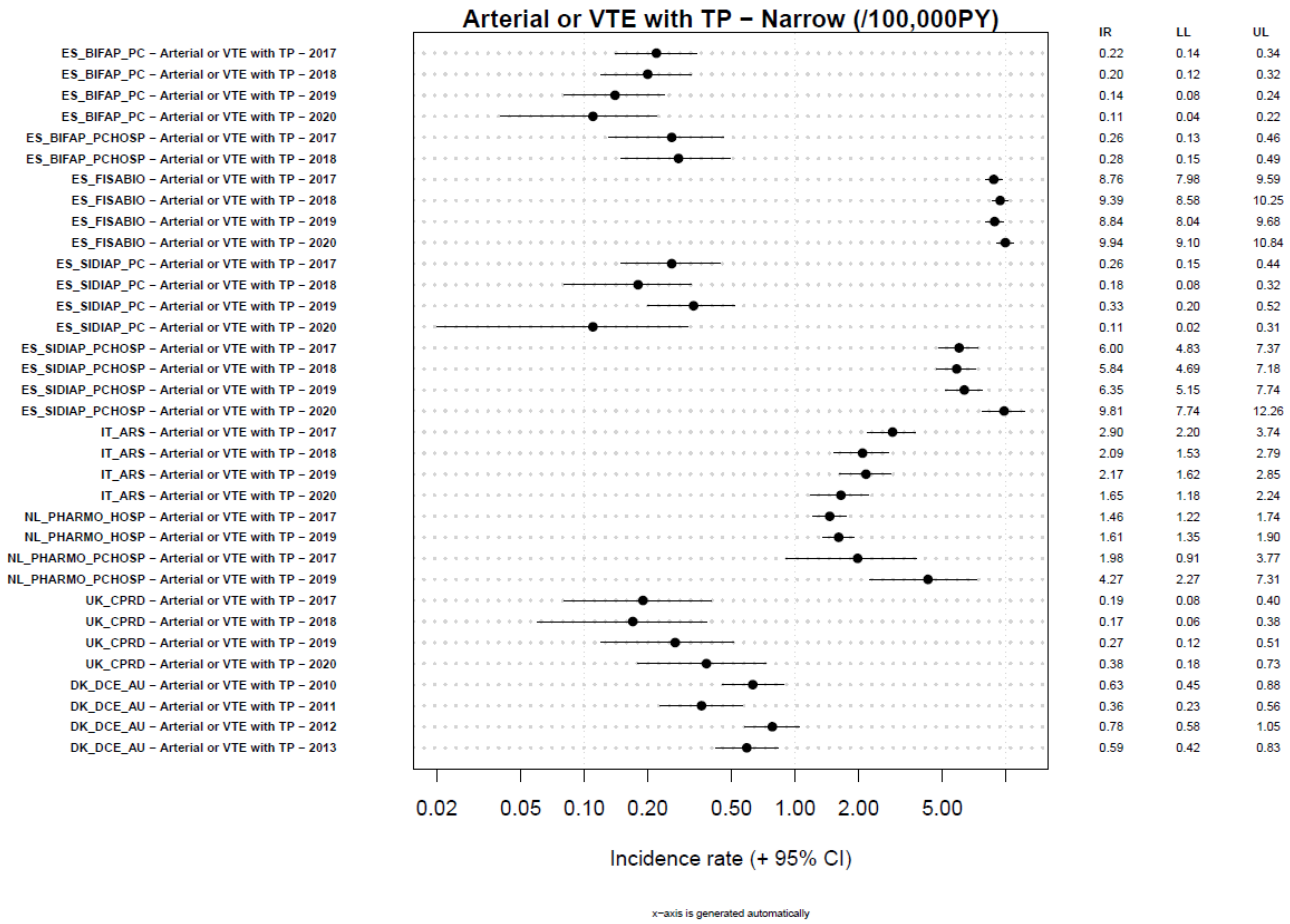


Figure 88 Incidence of Arterial or VTE with TP by data source and calendar year

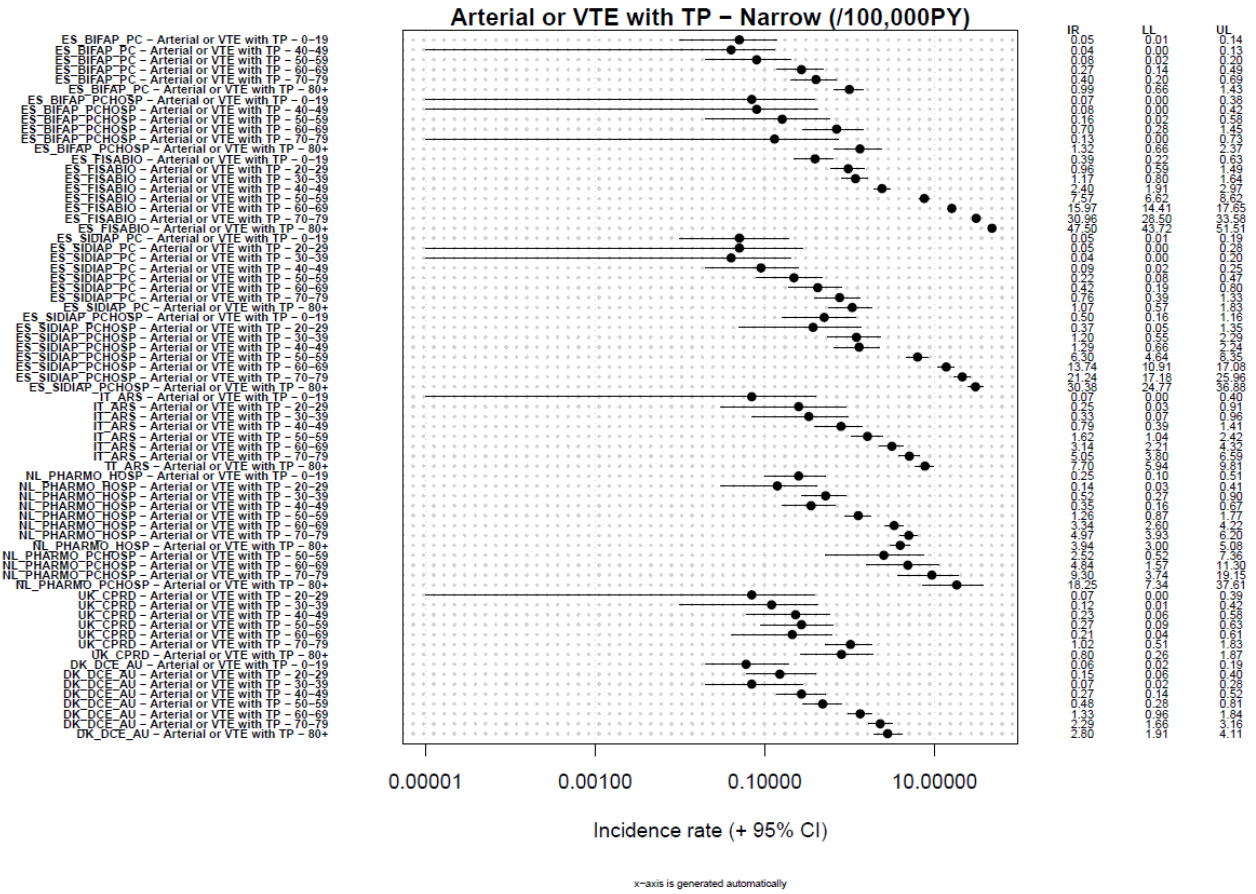


Figure 89 Incidence of Arterial or VTE with TP by data source and age

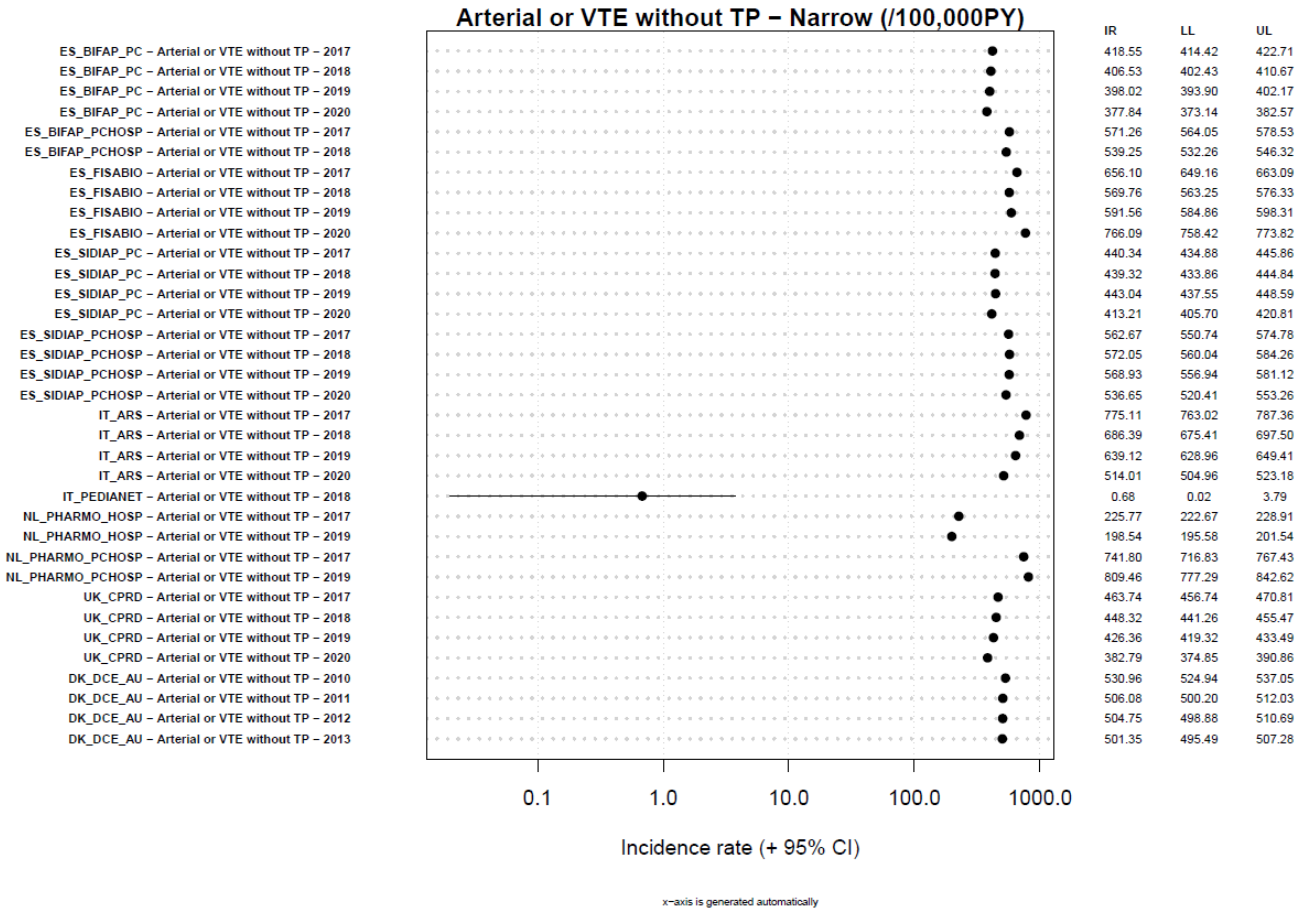


Figure 90 Incidence of Arterial or VTE without TP by data source and calendar year

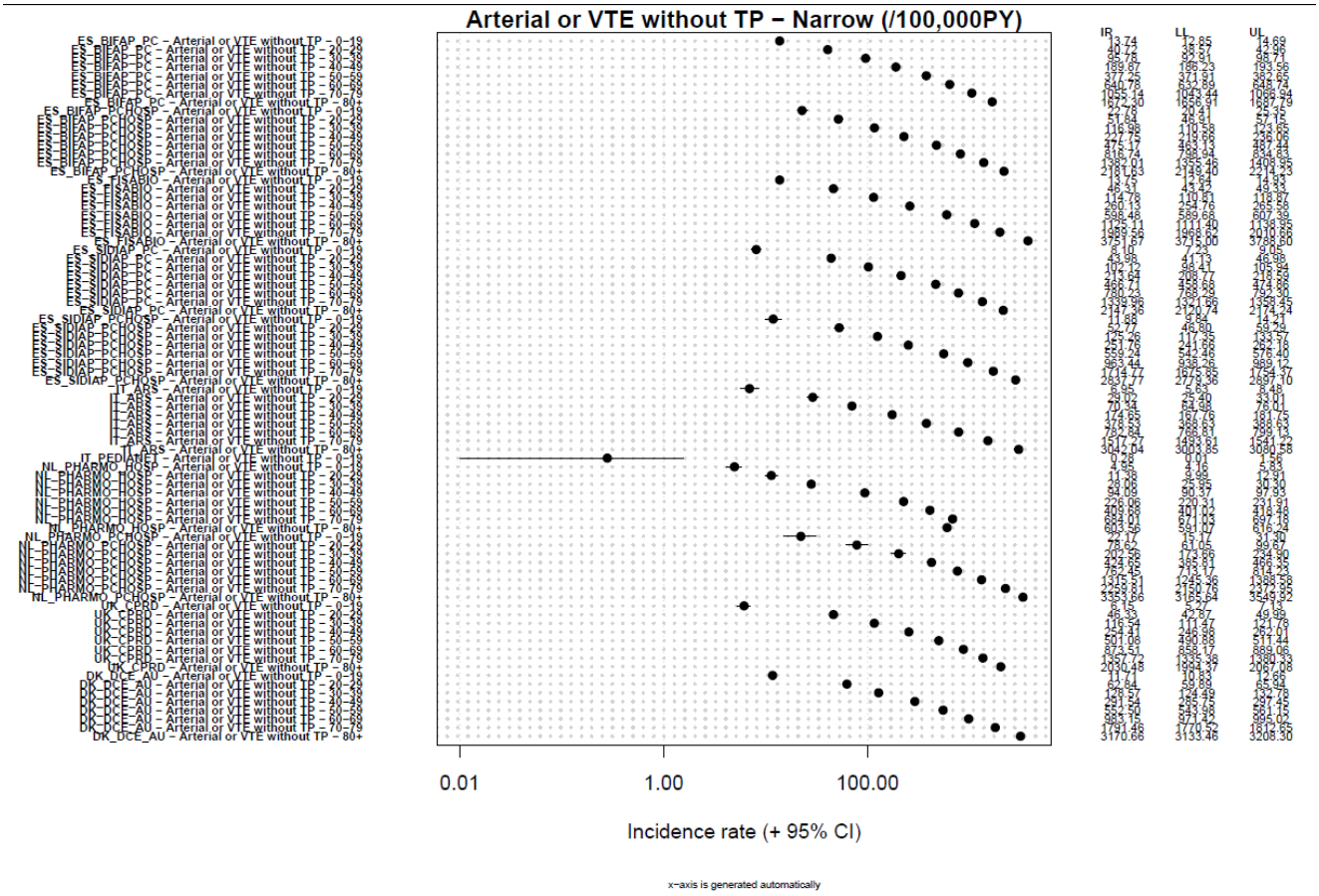


Figure 91 Incidence of Arterial or VTE without TP by data source and age

10.3.2 Control events

10.3.2.1 Colonic diverticulitis

Events of colonic diverticulitis using narrow definition were observed in all databases, except Pedianet (Figure 82). IRs were stable overtime. Significantly lower rates were observed in 2020 in ARS, CPRD, BIFAP and SIDIAP while significantly higher rates were observed in FISABIO for the same year.

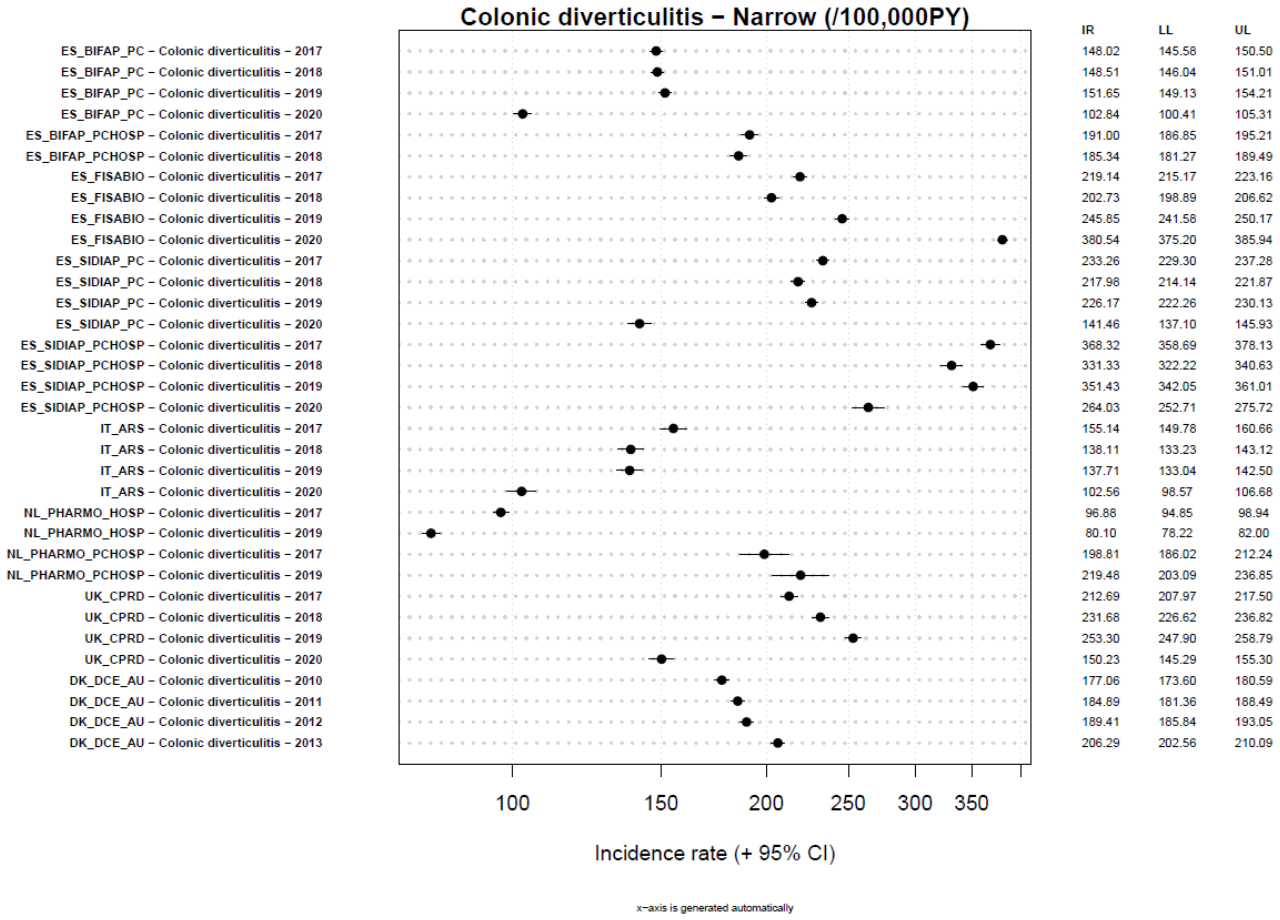


Figure 92 Incidence of colonic diverticulitis by data source and year

10.3.2.2 Hypertension

Hypertension using narrow definition & medication use were observed in all databases. IRs were stable overtime with a significant increase observed in 2020 in BIFAP and ARS. Increasing rates of hypertension diagnoses & medication use with age were observed in all databases.

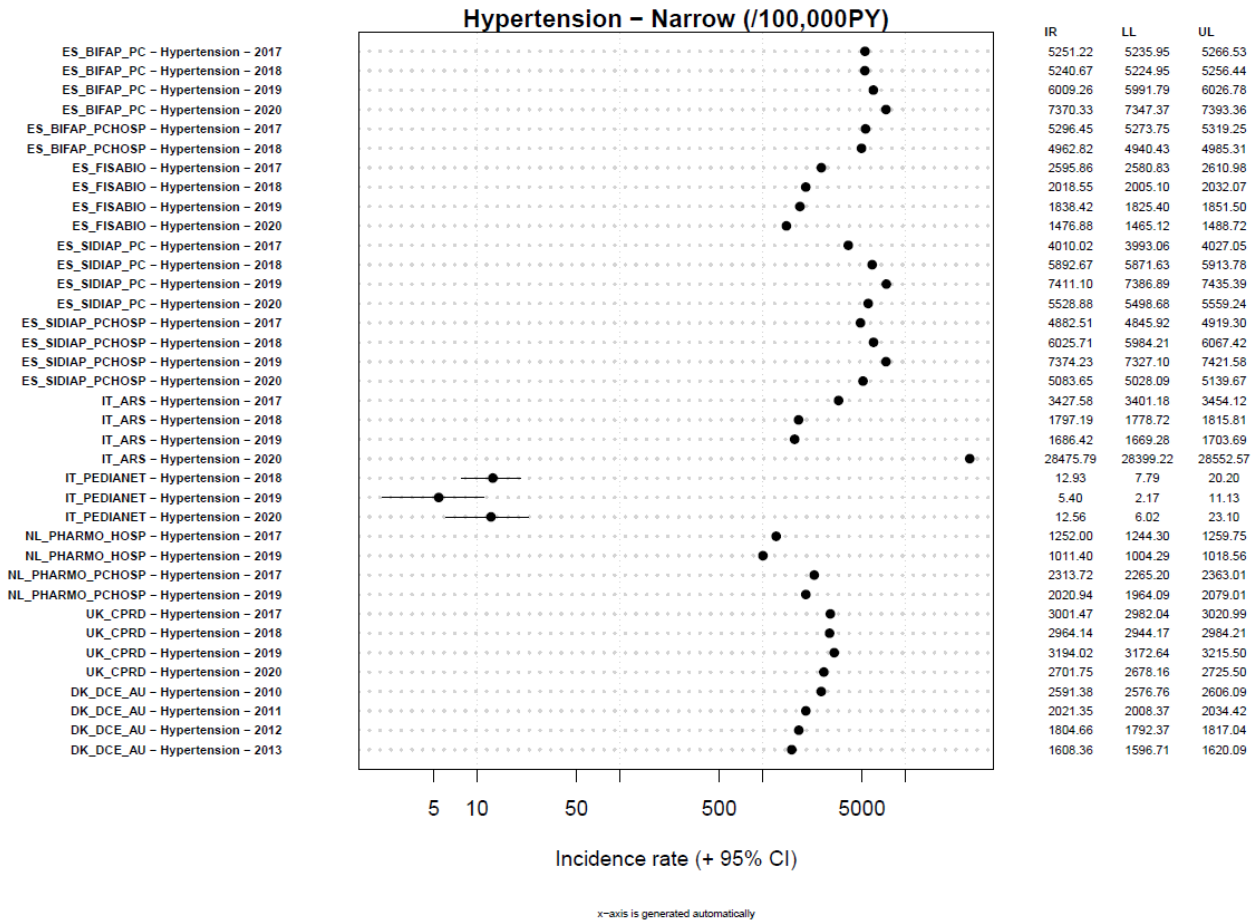


Figure 93 Incidence of hypertension by data source and calendar year

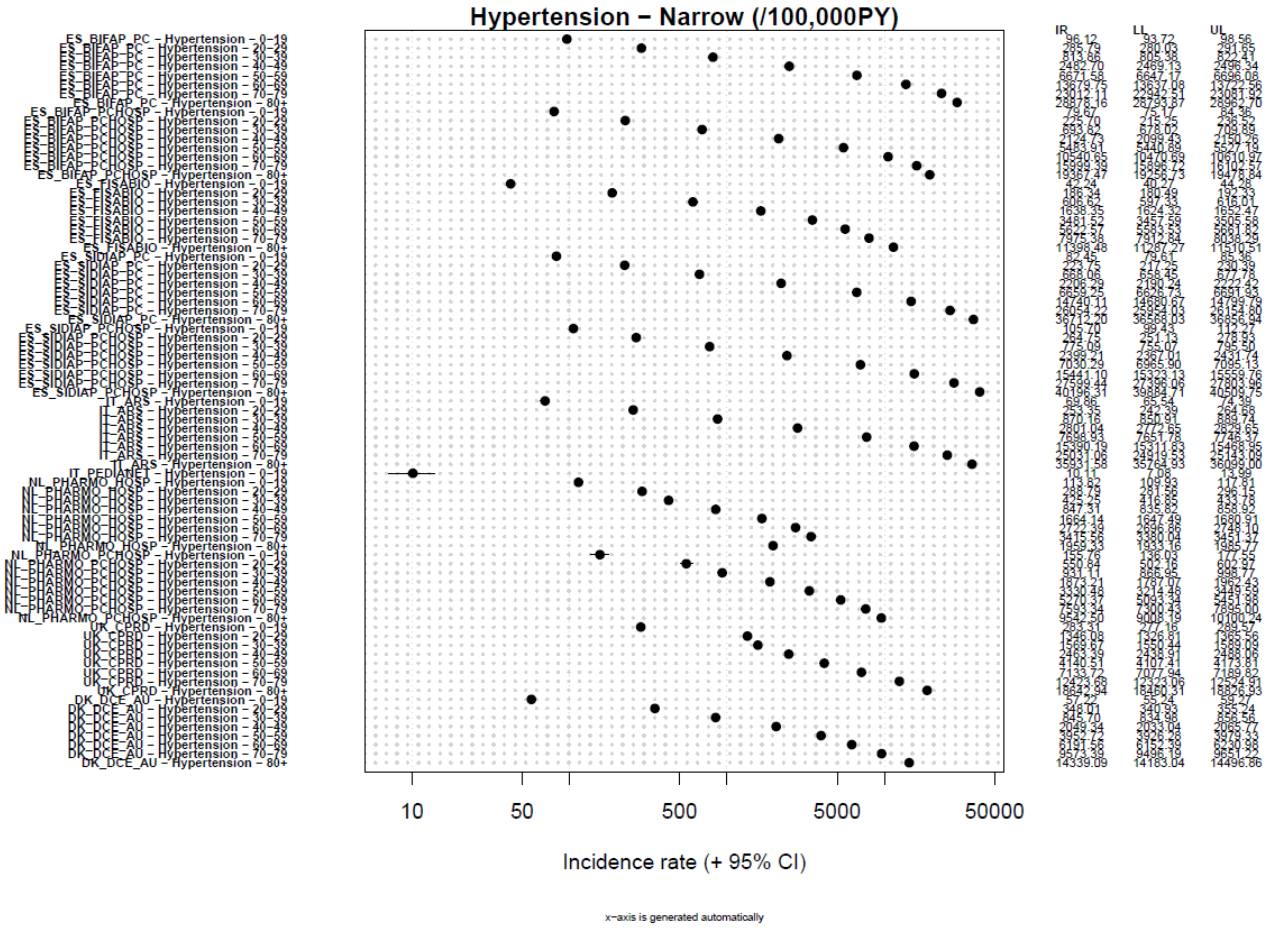


Figure 94 Incidence of hypertension by data source and age

10.3.3 Pregnancy outcomes

Prevalence rates for pregnancy outcomes have been generated from two data sources for which pregnancy registries were available (ie. ARS and CPRS). For all maternal and neonatal outcomes, the unit of analysis was pregnant women. Prevalence rates, expressed in per 1,000 pregnant women, are shown in the Figure 85 below.

In the ARS data source, a sensitivity analysis including all pregnant women captured in pregnancy registry and at in-and-outpatient hospital database were included (it concerns the following outcomes: spontaneous abortion, induced abortion, maternal death). In addition, narrow and broad clinical definitions for some AESIs (TOFPA, pre-eclampsia, microencephaly, major congenital anomalies, gestational diabetes, fetal growth retardation) are also reported as sensitivity analysis in Figure 85. In the CPRD data source, spontaneous abortion and induced abortion could not be captured and are therefore not presented. Both the pregnancy registry and medical records at GPs level were used to identify the outcomes.

From ARS database, the prevalence of maternal outcomes (gestational diabetes, pre-eclampsia and maternal death) increased with age. Rates of gestational diabetes increased with age and reached up to 55/1,000 pregnant women in 2019. The rates are aligned with published references (5.4% [95%: 3.8-7.8]), Eades, 2017). Prevalence of pre-eclampsia has been estimated at 2.8% of love birth (Osungbade,

2011), our analysis in ARS (an inpatient database) showed slightly elevated rates. Maternal death has been identified within 2 months post-delivery, the sensitivity analysis (dotted line in Figure 87) showed an increase of this event with age.
From the CPRD database, gestational diabetes showed an increase with age. Based on the data, maternal death could not be reliably captured and pre-eclampsia in 2019 indicated no event in the older age groups. This analysis indicated the need to further validate algorithms for the identification of pregnancy outcomes.

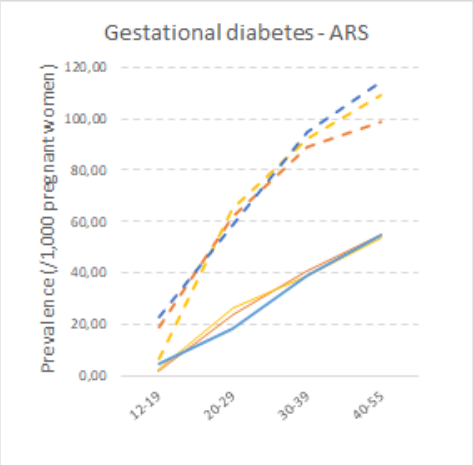
Spontaneous abortion increased with age and reached up to 231 events/1,000 pregnant women in the age 40-55 in 2019. The rates are higher compared to published reference (Strumpf, 2021). Preterm birth also increased with age with rates up to 55 events/1,000 pregnant women in 2019. Prevalence rates for preterm birth are aligned with WHO published data (6.2 [95%CI: 5.8-6.7], WHO, 2010). Prevalence rate of induced abortion is higher in younger women (41.3/1,000 pregnant women aged between 12-19).

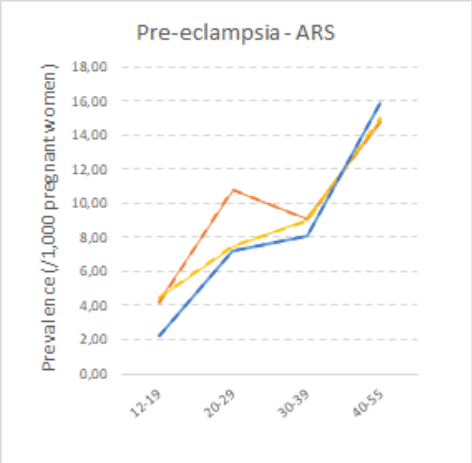
From the CPRD database, prevalence rates for preterm birth are higher compared to ARS. Congenital anomalies and TOFPA could be identified in the database. However, the algorithms to identify the other neonatal outcomes need further validation.

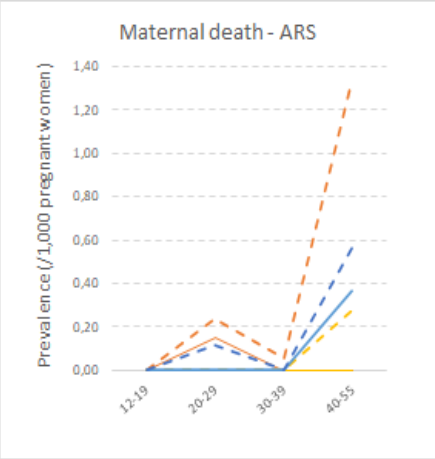
Eades CE, Cameron DM, Evans JMM. Prevalence of gestational diabetes mellitus in Europe: A meta-analysis. Diabetes Res Clin Pract. 2017 Jul;129:173-181. doi: 10.1016/j.diabres.2017.03.030. Epub 2017 May 9. PMID: 28531829.

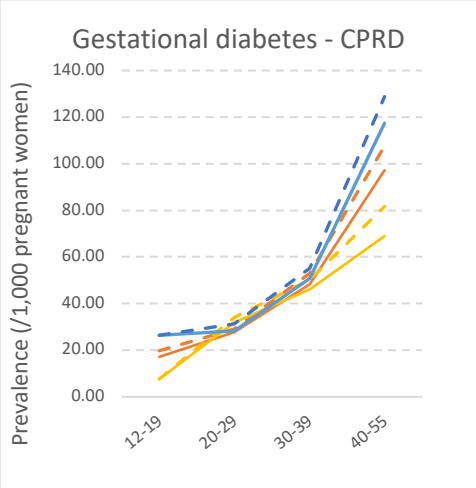
Osungbade KO, Ige OK. Public health perspectives of preeclampsia in developing countries: implication for health system strengthening. J Pregnancy. 2011;2011:481095. doi: 10.1155/2011/481095. Epub 2011 Apr 4. PMID: 21547090; PMCID: PMC3087154.
Strumpf, E., Lang, A., Austin, N. *et al.* Prevalence and clinical, social, and health care predictors of miscarriage. *BMC Pregnancy Childbirth* **21,** 185 (2021). https://doi.org/10.1186/s12884-021-03682-z
Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, Rubens C, Menon R, Van Look PF. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bull World Health Organ. 2010 Jan;88(1):31-8. doi: 10.2471/BLT.08.062554. Epub 2009 Sep 25. PMID: 20428351; PMCID: PMC2802437.

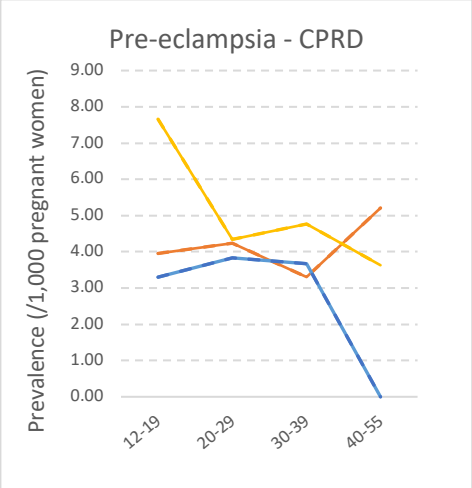
A- Maternal outcomes

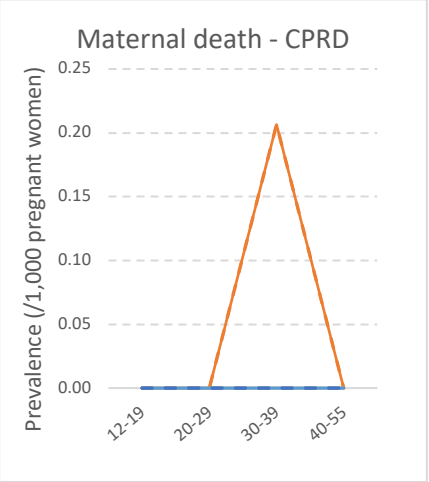




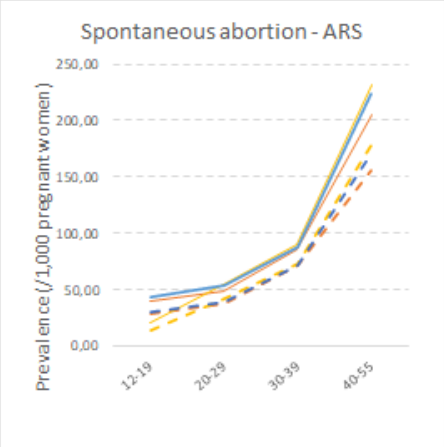


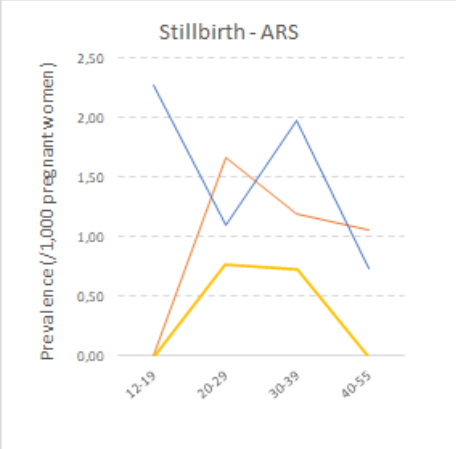


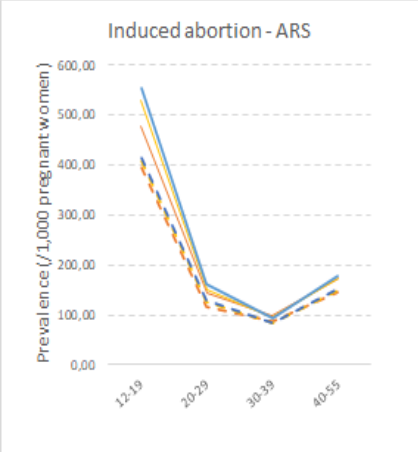


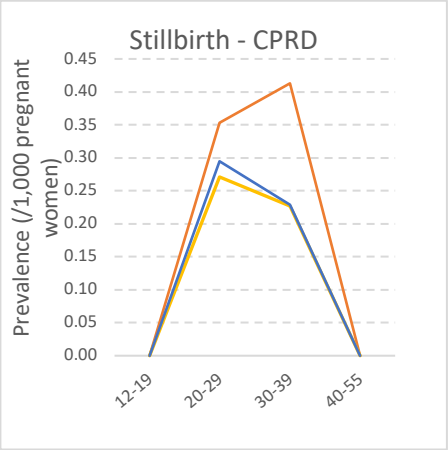


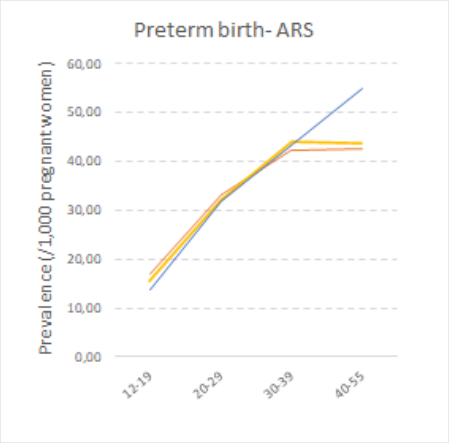
B- Neonates outcomes

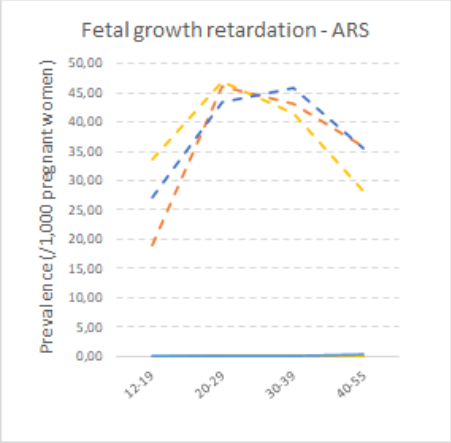

Spontaneous abortion - ARS
Prevalence (/1,000 pregnant women)
250,00
200,00
150,00
100,00
50,00
0,00
12-19
20-29
30-39
40-55


Stillbirth - ARS
Prevalence (/1,000 pregnant women)
2,50
2,00
1,50
1,00
0,50
0,00
12-19
20-29
30-39
40-55


Induced abortion - ARS
Prevalence (/1,000 pregnant women)
600,00
500,00
400,00
300,00
200,00
100,00
0,00
12-19
20-29
30-39
40-55


Stillbirth - CPRD
Prevalence (/1,000 pregnant women)
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
12-19
20-29
30-39
40-55


Preterm birth- ARS
Prevalence (/1,000 pregnant women)
60,00
50,00
40,00
30,00
20,00
10,00
0,00
12-19
20-29
30-39
40-55


Fetal growth retardation - ARS
Prevalence (/1,000 pregnant women)
50,00
45,00
40,00
35,00
30,00
25,00
20,00
15,00
10,00
5,00
0,00
12-19
20-29
30-39
40-55

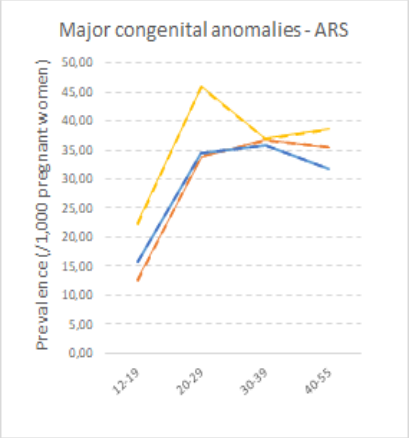

Major congenital anomalies - ARS
Prevalence (/1,000 pregnant women)
50,00
45,00
40,00
35,00
30,00
25,00
20,00
15,00
10,00
5,00
0,00
12-19
20-29
30-39
40-55

Figure 95 *Prevalence rates of pregnancy outcome (/1,000 pregnant women)*

11 Other analyses

Counts of codes identified for each AESI, at-risk condition, and at-risk condition drug proxy can be found in Annex 2.

Monthly incidence rates are available in Annex 9. Line graphs are displayed for each AESI and each database.

12 Discussion

12.1 Key results

This study generated incidence rates for 41 AESIs, including maternal and neonatal pregnancy outcomes.

This final report highlights the challenges of conducting studies using distributed data network. While the process to access data was facilitated for some data sources, the governance approval process for some other data sources did not allow to provide up-to-date data (eg. DCE-AU) or delayed considerably the data extraction process (eg. SNDS).

This report comprises data from 6 countries (UK, ES, IT, DK, NL, DE) and 9 data sources (BIFAP, Pedianet, CPRD, ARS, Danish registries, FISABIO, SIDIAP, PHARMO, GeParD). Data from France (SNDS) could not be included but will be made available on the VAC4EU dashboard. Due to high rates that were observed from the German database, data from Germany (GeParD) were further investigated with the conclusion that only events which require hospitalisation can be accurately estimated. For this reason, only a limited number of events were included in the study for the GePaRD database (ADEM, GBS, Acute kidney injury, acute liver injury, heart failure, coronary artery disease, acute respiratory distress syndrome, anaphylaxis and multi-system inflammatory syndrome). In addition, the data from GePaRD represents 800,000 individuals out of 25 million (one of the smallest SHI in Germany), the team will extract data from the largest healthcare insurer that will allow for a highest representativeness of the population and the data will be made available on the VAC4EU dashboard. A total number of 45 million individuals were included in this study contributing to 148 million person-years.

In this report, rates for myocarditis alone beyond to myocarditis/pericarditis were generated. In addition, a refinement of medical codes has been conducted which has little impact on the rates that were produced for the previous version of the report.

Each AESI was defined according to narrow and broad clinical definitions. The narrow definition included medical codes that are specific for the identification of the event of interest. Broad definition included a larger set of medical codes that were considered (possible) thus sensitive for the identification of events. The incidence rates increased drastically when broad definitions were used (e.g. ADEM, ITP, narcolepsy, cardiovascular diseases in general, generalized convulsion, anaphylaxis). It also shows the dependency of results on the type of algorithm that is chosen. This indicates the range by which we should interpret results.

Overall, the incidence rates were shown to be quite consistent from one year to another and between databases. The results showed an increased in mortality in 2020 and a decreased in some diseases such as cardiovascular diseases in 2020. Age patterns were clearly observed for most of the events.

The negative control events (not causative of COVID-19 disease) showed patterns we expected, for countries with 2020 data, the rates decreased, especially for hypertension, where medical visits are required. AESI that are not really symptomatic but require medical attention might be affected in a similar fashion.

12.2 Limitations

Due to very limited resources in ACCESS funding, a validation of the identified events could not be performed. For this reason, the risk of misclassifications cannot be excluded, we tried to show the impact of potential misclassification by using narrow and broad definitions. Further investigations on the impact of misclassification could be performed by adding in an analysis focusing on the provenance (meaning) of the diagnosis code (e.g. primary care, hospital discharge, specialist, laboratory etc), to be prepared for association safety studies.

Most of the AESIs included in this study require visit to specialists or hospitalization. Databases using exclusively general practitioner's data may underestimate the incidence of these events. On the other side, data sources with just hospitalizations (ARS) might underestimate events that are diagnosed mostly in an outpatient setting. Incidence rates from claims databases should be considered with caution as it is likely that incidences were slightly overestimated. For most the AESIs, an age pattern was observed increasing with age. This age-specific pattern should be taken into consideration for future use of these background incidences.

12.3 Strengths

The study generated background incidence rates of 42 AESIs with a high precision. Given the size of the study population covered by the databases, the estimates are likely representative of the European population. In addition, the concept of subpopulation could be introduced for 3 databases (PHARMO, SIDIAP and BIFAP). The subpopulation includes patients who has a health care follow-up at hospital and primary care levels, it allows to provide more accurate estimates for diseases that required emergency room visit and/or hospitalisation. In addition, the pipeline could generate background incidence for newly identified syndrome like TTS, it shows the strength of the infrastructure in rapid response to specific research questions.

12.4 Interpretation

We provided rates from various data sources in different countries as EMA requested. European data sources are quite heterogeneous because of different coding systems, health care practices and availability of data. We used a two-step approach to harmonize, first a syntactic harmonization, putting all data in the same structure, and secondly a semantic harmonization, based on code mapping. Sematic harmonization is complex, and infinite. It comprises of harmonization of different coding systems with different granularity, coding practices in different settings and an impact analysis of this. In this initial analysis we harmonized on the basis of coding, and developed algorithms based on codes and meanings of codes. We will continue this work, to investigate the impact of the use of different data provenances. For signal detection interpretation we now recommend that both narrow and broad definitions are used in

the triaging process. The component analysis will impact most on definitions we may use for signal evaluation.

The generated incidence rates were within the range of background incidences reported in the literature obtained through a rapid assessment of literature. Due to restricted resources, we could not do a systematic literature research, but will rely on the systematic Brighton Collaboration/SPEAC literature background rate assessment, for the final report and paper.

12.5 Generalisability

This study used a wide population range, without restrictions beyond study period. This ensures the generalizability of the incidence rates.

13 Conclusion

The study generated background incidence rates with high precision for a pre-specified list of AESIs which will be used for further assessment of the safety of COVID-19 vaccines. It was building on the IMI-Conception CDM and pipeline.

14 Annexes

Appendix 1. List of stand-alone documents

[Documents listed in Annex 1 can be maintained separately from the study final study report. They should be clearly identifiable and provided on request. Write "None" if there is no document or list documents in a table as indicated below.]

Number	Document reference number	Title
1	*Annex 1*	*Event definition forms:* https://zenodo.org/communities/vac4eu/
2	*Annex 2*	*Excel sheet – IR narrow*
3	*Annex 3*	*Forest plot – by year - narrow*
4	*Annex 4*	*Forest plot – by age - narrow*
5	*Annex 5*	*Excel sheet – IR broad*
6	*Annex 6*	*Forest plot – by year - broad*
7	*Annex 7*	*Forest plot – by age - broad*
8	*Annex 8*	*Forest plot – IR at risk population*
9	*Annex 9*	*Line plot – Monthly IR*