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

Version 1.2
April 30, 2021

<table>
<thead>
<tr>
<th>Title</th>
<th>Background rates of Adverse Events of Special Interest for monitoring of COVID-19 vaccines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version identifier of the final study report</td>
<td>1.2 (draft final)</td>
</tr>
<tr>
<td>Date of last version of the final study report</td>
<td>April 30, 2021</td>
</tr>
<tr>
<td>EU PAS register number</td>
<td>EUPAS37273</td>
</tr>
<tr>
<td>Active substance</td>
<td>NA</td>
</tr>
<tr>
<td>Medicinal product</td>
<td>NA</td>
</tr>
<tr>
<td>Product reference</td>
<td>NA</td>
</tr>
<tr>
<td>Procedure number</td>
<td>NA</td>
</tr>
<tr>
<td>Marketing authorisation holder(s)</td>
<td>NA</td>
</tr>
<tr>
<td>Research question and objectives</td>
<td>To generate incidence rates of adverse event of special interest for COVID-19 vaccine safety monitoring in 7 EU countries.</td>
</tr>
<tr>
<td>Country(-ies) of study</td>
<td>Seven participating electronic health care databases in 5 countries: Germany, Italy, Spain, UK and Denmark. Data from Netherlands and France will arrive in the final version.</td>
</tr>
</tbody>
</table>
| Authors | Corinne Willame, University Medical Center Utrecht, The Netherlands  
Miriam Sturkenboom, University Medical Center Utrecht  
Daniel Weibel, University Medical Center Utrecht, The Netherlands |
| Marketing authorisation holder(s) | none |
| MAH contact person | Not relevant |
Table of contents

1 ABSTRACT .................................................................................................................. 7
2 LIST OF ABBREVIATIONS .......................................................................................... 16
3 INVESTIGATORS .......................................................................................................... 17
4 OTHER RESPONSIBLE PARTIES ............................................................................... 18
5 MILESTONES ............................................................................................................... 18
6 RATIONALE AND BACKGROUND ........................................................................... 20
  6.1 BACKGROUND ........................................................................................................ 20
  6.2 RATIONALE FOR THE STUDY ............................................................................. 20
7 RESEARCH QUESTION AND OBJECTIVES ............................................................... 21
  7.1 CO-PRIMARY OBJECTIVES .................................................................................. 21
  7.2 SECONDARY OBJECTIVES .................................................................................... 21
8 AMENDMENTS AND UPDATES .................................................................................. 21
9 RESEARCH METHODS ................................................................................................. 22
  9.1 STUDY DESIGN ........................................................................................................ 22
  9.2 SETTING .................................................................................................................. 22
  9.3 SUBJECTS ............................................................................................................... 23
  9.4 VARIABLES ............................................................................................................ 24
     9.4.1 Person-time & Follow-up .................................................................................. 24
     9.4.2 AESI, At-risk medical conditions & Operationalization .................................. 24
     9.4.3 Other variables ............................................................................................... 28
  9.5 DATA SOURCES AND MEASUREMENT ................................................................. 28
     9.5.1 Germany: GePaRD .......................................................................................... 28
     9.5.2 Netherlands: PHARMO Database Network ...................................................... 29
     9.5.3 Denmark: Danish Registries (DCE-AU) ............................................................ 29
     9.5.4 Spain: BIFAP .................................................................................................... 30
     9.5.5 Spain: SIDIAP .................................................................................................. 30
     9.5.6 Spain: FISABIO ............................................................................................... 31
     9.5.7 Italy: PEDIANET database .............................................................................. 31
     9.5.8 Italy: ARS database ......................................................................................... 32
     9.5.9 United Kingdom: CPRD & HES ...................................................................... 32
     9.5.10 France: Système National des Données de Santé (SNDS) ................................ 33
  9.6 BIAS ........................................................................................................................ 34
  9.7 STATISTICAL METHODS ........................................................................................ 37
     9.7.1 Analysis of Demographics and Baseline Characteristics .................................. 37
     9.7.2 Hypotheses ....................................................................................................... 37
     9.7.3 Statistical Methods ........................................................................................... 37
     9.7.4 Statistical Analysis ............................................................................................ 38
     9.7.5 Missing data ..................................................................................................... 39
List of Figures
Figure 1 Study design .............................................................................................................. 23
Figure 2 Data management plan ............................................................................................. 36
Figure 3: Person-time (years) over calendar and by data source ........................................... 44
Figure 4: Forest plot for incidence rates of GBS per 100,000 PY by data source and calendar year using a narrow definition ................................................................. 47
Figure 5: Forest plot for incidence rates of GBS per 100,000 PY by data source and age for narrow definition ................................................................. 48
Figure 6: Forest plot for incidence rates of ADEM per 100,000 PY by year ................................ 49
Figure 7: Forest plot of incidence rate of ADEM by age category ........................................... 50
Figure 8: Incidence rates of narcolepsy per 100,000 PY and calendar year ............................ 52
Figure 9: Incidence rates of narcolepsy per 100,000 PY and age ........................................... 53
Figure 10: Incidence rates of acute aseptic arthritis (broad definition) per 100,000 PY and calendar year ................................................................. 54
Figure 11: Incidence rates of diabetes mellitus type 1 per 100,000 PY and calendar year .......... 55
Figure 12: Incidence of Diabetes (type 1) by age ................................................................. 55
Figure 13: Incidence rates of thrombocytopenia (primary & secondary) ................................. 56
Figure 14: Incidence rates of thrombocytopenia (primary and secondary) by age and data source ........................................................................................................... 57
Figure 15: TP incidence rates (primary and secondary) from IMI-ADVANCE separated by source of data (Willame et al.) ........................................................................................................... 58
Figure 16: Incidence rates of microangiopathy by calendar year and data source ..................... 58
Figure 17: Incidence rate of microangiopathy by data source and age category ....................... 59
Figure 18: Incidence of heart failure by data source and calendar year ................................... 60
Figure 19: Incidence of heart failure by data source and age ................................................... 61
Figure 20: Incidence of stress cardiomyopathy by data source and calendar year ................... 62
Figure 21: Incidence of stress cardiomyopathy by data source and age ................................... 63
Figure 22: Incidence of CAD (narrow) by data source and calendar year ................................... 64
Figure 23: Incidence of CAD (narrow) by data source and age ................................................ 65
Figure 24: Incidence of CAD (broad) by data source and age .................................................. 66
Figure 25: Incidence of cardiovascular disease according to Global burden of disease study ........ 67
Figure 26: Incidence of arrhythmia by data source and calendar year ..................................... 68
Figure 27: Incidence of arrhythmia by data source and age ..................................................... 69
Figure 28: Incidence of arrhythmia in the UK Biobank ............................................................ 70
Figure 29: Incidence of myocarditis and pericarditis by data source and calendar year ............. 71
Figure 30: Incidence of myocarditis and pericarditis by data source and age ............................ 71
Figure 31: Incidence of DIC by data source and calendar year ............................................... 73
Figure 32: Incidence of DIC by data source and age ............................................................... 73
Figure 33: Incidence of TTP by data source and calendar year ............................................... 74
Figure 34: Incidence of TTP by data source and age ............................................................... 75
Figure 35: Incidence of VTE by data source and calendar year ................................................ 76
Figure 36: Incidence of VTE by data source and age ............................................................... 77
Figure 37: Incidence of Cerebral venous thrombosis by data source and calendar year .............. 78
Figure 38: Incidence of Cerebral venous thrombosis by data source and age ............................ 79
Figure 39: Incidence of Ischemic stroke by data source and calendar year ............................... 80
Figure 40: Incidence of Ischemic stroke by data source and age ............................................. 81
Figure 41: Incidence of hemorrhagic stroke by data source and calendar year ....................... 82
Figure 42: Incidence of hemorrhagic stroke by data source and age ....................................... 83
Figure 43: Incidence of single organ cutaneous vasculitis by data source and calendar year ........ 84
Figure 44: Incidence of single organ cutaneous vasculitis by data source and age .................... 85
Figure 45: Incidence of Acute Liver Injury by data source and calendar year ............................ 86
Figure 46: Incidence of acute liver injury by data source and age ............................................ 87
Figure 47: Incidence of Acute kidney Injury by data source and calendar year ....................... 88
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-CoV2 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

<table>
<thead>
<tr>
<th>Body system / Classification</th>
<th>AESI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto-immune diseases</td>
<td>Guillain-Barré Syndrome (GBS)</td>
</tr>
<tr>
<td></td>
<td>Acute disseminated encephalomyelitis (ADEM)</td>
</tr>
<tr>
<td></td>
<td>Narcolepsy</td>
</tr>
<tr>
<td></td>
<td>Acute aseptic arthritis</td>
</tr>
<tr>
<td></td>
<td>(Type I) Diabetes</td>
</tr>
<tr>
<td></td>
<td>(Idiopathic)Thrombocytopenia</td>
</tr>
<tr>
<td>Cardiovascular system</td>
<td>Acute cardiovascular injury including: Microangiopathy, Heart failure, Stress cardiomyopathy, Coronary artery disease, Arrhythmia, Myocarditis</td>
</tr>
<tr>
<td>Circulatory system</td>
<td>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</td>
</tr>
<tr>
<td>Hepato-gastrointestinal and renal system</td>
<td>Acute liver injury</td>
</tr>
<tr>
<td></td>
<td>Acute kidney injury</td>
</tr>
<tr>
<td>Nerves and central nervous system</td>
<td>Generalized convulsion</td>
</tr>
<tr>
<td></td>
<td>Meningoencephalitis</td>
</tr>
<tr>
<td></td>
<td>Transverse myelitis</td>
</tr>
<tr>
<td>Respiratory system</td>
<td>Acute respiratory distress syndrome</td>
</tr>
<tr>
<td>Skin and mucous membrane, bone and joints system</td>
<td>Erythema multiforme</td>
</tr>
<tr>
<td></td>
<td>Chilblain – like lesions</td>
</tr>
<tr>
<td>Other system</td>
<td>Anosmia, ageusia</td>
</tr>
<tr>
<td></td>
<td>Anaphylaxis</td>
</tr>
<tr>
<td></td>
<td>Multisystem inflammatory syndrome in children (MIS-C)</td>
</tr>
<tr>
<td></td>
<td>Death (any causes)</td>
</tr>
<tr>
<td></td>
<td>Coronavirus disease 2019 (COVID-19)</td>
</tr>
<tr>
<td></td>
<td>Sudden Death</td>
</tr>
</tbody>
</table>

Pregnancy outcomes (to be included in final version of report):

| 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 will finally include 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 report data from 8 data sources are included.

**Study size:**
The study population for the total study comprises approximately 130.6 million individuals. In this report, a total number of 35 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 rates (and 95%CI) 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 5 countries (UK, ES, IT, DK, NL) and 8 data sources (BIFAP, Pedianet (children only), CPRD, ARS, Danish registries, FISABIO, SIDIAP, PHARMO). Data from France (SNDS) will be added in the last version, as well as the data on the pregnancy outcomes. Data sources included different subpopulations based on the availability of numerator data (Hosp: hospital based, PC= primary care, HOSP-PC=overlap between hospitalization and primary care

Below the below diagram is shown

---

**Table 2 Attriution diagram**

<table>
<thead>
<tr>
<th>Subjects disposition</th>
<th>ARS</th>
<th>FISABIO</th>
<th>SIDIAP</th>
<th>CPRD</th>
<th>PHARMO</th>
<th>PHARMO_PC_HOSP</th>
<th>SIDIAP_PC</th>
<th>SIDIAP_PC_HOSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects without sex or missing birth date</td>
<td>90036</td>
<td>71711</td>
<td>764981</td>
<td>762617</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
</tr>
<tr>
<td>Subjects without sex or missing birth date and data</td>
<td>8715</td>
<td>71711</td>
<td>764981</td>
<td>762617</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
</tr>
<tr>
<td>Subjects without sex or missing birth date</td>
<td>90041</td>
<td>71711</td>
<td>764981</td>
<td>762617</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
</tr>
<tr>
<td>Subjects without sex or missing birth date and data</td>
<td>8715</td>
<td>71711</td>
<td>764981</td>
<td>762617</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
<td>2307785</td>
</tr>
</tbody>
</table>

*Aarhus University (DCE-AU) did not extract 2016-2020 but 2010-2013.

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

<table>
<thead>
<tr>
<th>AESI</th>
<th>Incidence GP based (BIFAP_PC, SIDIAP_PC, CPRD, Pedianet) 2017, (95% CI)</th>
<th>Incidence GP and inpatient (BIFAP_PC_HOSP, PHARMO_PC_HOSP)</th>
<th>Incidence inpatient only and Emergency room (ARS), 2017, (95% CI)</th>
<th>Incidence in and outpatient (DCE-AU, FISABIO), 2017, (95% CI)</th>
<th>Claims data (GePard), 2017, (95% CI)</th>
<th>Calendar year change in 2020</th>
<th>Age pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. GBS</td>
<td>BIFAP_PC: 1.34 (1.13-1.59) SIDIAP_PC: 2.28 (1.91-2.72) CPRD: 1.54 (1.19-2.0)</td>
<td>BIFAP_PC_HOSP: 2.05 (1.66-2.53) PHARMO_PC_HOSP: 2.89 (1.74-4.79)</td>
<td>ARS: 3.09 (4.48)</td>
<td>DCE-AU (2010): 3.06 (2.63-3.55) FISABIO: 2.00 (1.65-2.42)</td>
<td>-</td>
<td>Increase with age consistently and lowering after 80</td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>BIFAP_PC (2019): 0.03 (0.01-0.09)</td>
<td>SIDIAP_PC: 0.36 (0.23-0.56)</td>
<td>CPRD: NA, Pedianet: NA</td>
<td>BIFAP_PC_HOSP: 0.76 (0.54-1.07)</td>
<td>ARS: 0.07 (0.02-0.28)</td>
<td>DCE-AU (2010): 0.79 (0.59-1.06)</td>
<td>FISABIO: 0.32 (0.20-0.51)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>2. ADEM</td>
<td>BIFAP_PC: 2.35 (2.06-2.68)</td>
<td>SIDIAP_PC: 0.97 (0.74-1.27)</td>
<td>CPRD: 1.13 (0.83-1.53)</td>
<td>BIFAP_PC_HOSP: 1.75 (1.39-2.20)</td>
<td>PHARMO_PC_HOSP: 0.19 (0.03-0.31)</td>
<td>ARS: 0.50 (0.30-0.83)</td>
<td>DCE-AU (2010): 3.25 (2.81-3.76)</td>
</tr>
<tr>
<td>3. Narcolepsy</td>
<td>BIFAP_PC: 2.35 (2.06-2.68)</td>
<td>SIDIAP_PC: 0.97 (0.74-1.27)</td>
<td>CPRD: 1.13 (0.83-1.53)</td>
<td>BIFAP_PC_HOSP: 1.75 (1.39-2.20)</td>
<td>PHARMO_PC_HOSP: 0.19 (0.03-0.31)</td>
<td>ARS: 0.50 (0.30-0.83)</td>
<td>DCE-AU (2010): 3.25 (2.81-3.76)</td>
</tr>
<tr>
<td>4. Acute Aseptic Arthritis</td>
<td>No narrow codes</td>
<td>No narrow codes</td>
<td>No narrow codes</td>
<td>No narrow codes</td>
<td>No narrow codes</td>
<td>No narrow codes</td>
<td>No narrow codes</td>
</tr>
<tr>
<td>7. Microangiopathy</td>
<td>BIFAP_PC: 0.53 (0.40-0.70)</td>
<td>SIDIAP_PC: 1.26 (1.00-1.60)</td>
<td>CPRD: 0.63 (0.42-0.95)</td>
<td>BIFAP_PC_HOSP: 0.64 (0.44-0.93)</td>
<td>PHARMO_PC_HOSP: 0.77 (0.29-2.05)</td>
<td>0.67 (0.43-1.04)</td>
<td>DCE-AU (2010): 3.63 (3.16-4.17)</td>
</tr>
<tr>
<td>9. Stress cardiomyopathy</td>
<td>BIFAP_PC: 0.24 (0.16-0.36)</td>
<td>SIDIAP_PC: 0.57 (0.40-0.81)</td>
<td>CPRD: no event</td>
<td>BIFAP_PC_HOSP: 1.58 (1.24-2.01)</td>
<td>6.72 (5.85-7.71)</td>
<td>DCE-AU: no event</td>
<td>FISABIO: 2.92 (2.49-3.42)</td>
</tr>
<tr>
<td>10. Coronary artery disease</td>
<td>BIFAP_PC: 89.45 (87.56-91.38)</td>
<td>SIDIAP_PC: 91.30 (88.80-93.87)</td>
<td>CPRD: 149.67 (145.74-153.70)</td>
<td>Pedianet: no event</td>
<td>BIFAP_PC_HOSP: 139.03 (135.52-142.63)</td>
<td>PHARMO_PC_HOSP: 80.57 (74.42-87.23)</td>
<td>292.81 (286.75-299.00)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>12. Myocarditis/ pericarditis</td>
<td>BIFAP_PC: 11.76 (11.09-12.47)</td>
<td>SIDIAP_PC: 0.33 (0.21-0.52)</td>
<td>CPRD: 2.56 (2.09-3.14)</td>
<td>Pedianet (2018): 0.68 (0.10-4.83)</td>
<td>BIFAP_PC_HOSP: 11.76 (10.77-12.84)</td>
<td>PHARMO_PC_HOSP: 1.54 (0.77-3.08)</td>
<td>9.11 (8.09-10.26)</td>
</tr>
<tr>
<td>13.1 Disseminated Intravascular Coagulation</td>
<td>BIFAP_PC: 7.53 (7.00-8.10)</td>
<td>SIDIAP_PC: 0.51 (0.35-0.74)</td>
<td>CPRD: 0.14 (0.06-0.34)</td>
<td>Pedianet (2019): 0.77 (0.11-5.47)</td>
<td>BIFAP_PC_HOSP: 6.97 (6.22-7.81)</td>
<td>1.63 (1.23-2.16)</td>
<td>DCE-AAU (2010): 2.79 (2.38-3.26)</td>
</tr>
<tr>
<td>13.2 Thrombotic Thrombocytopenic Purpura</td>
<td>BIFAP_PC: 0.88 (0.71-1.09)</td>
<td>SIDIAP_PC: 0.75 (0.55-1.02)</td>
<td>CPRD: 0.30 (0.17-0.54)</td>
<td>Pedianet: no event</td>
<td>BIFAP_PC_HOSP: 0.87 (0.63-1.20)</td>
<td>0.67 (0.43-1.04)</td>
<td>DCE-AU: 0.97 (0.74-1.27)</td>
</tr>
<tr>
<td>13.3 Venous Thromboembolism</td>
<td>BIFAP_PC: 190.85 (188.08-193.66)</td>
<td>SIDIAP_PC: 192.35 (188.22-196.57)</td>
<td>CPRD: 174.70 (170.45-179.08)</td>
<td>Pedianet (2018): 0.68 (0.10-4.83)</td>
<td>BIFAP_PC_HOSP: 226.73 (222.23-231.31)</td>
<td>182.96 (178.18-187.86)</td>
<td>DCE-AU: 181.36 (177.86-184.93)</td>
</tr>
<tr>
<td>13.3 Ischemic stroke</td>
<td>BIFAP_PC: 170.54 (167.93-173.19)</td>
<td>SIDIAP_PC: 169.44 (166.03-172.93)</td>
<td>CPRD: 224.83 (220.00-229.77)</td>
<td>Pedianet (2018): 0.68 (0.10-4.83)</td>
<td>BIFAP_PC_HOSP: 269.76 (264.85-274.76)</td>
<td>378.11 (371.21-385.14)</td>
<td>DCE-AU: 248.14 (244.04-252.31)</td>
</tr>
</tbody>
</table>

D3 Draft final report with partial results 30-04-2021 11
<table>
<thead>
<tr>
<th>13.5 Hemorrhagic stroke</th>
<th>BIFAP_PC: 10.80 (10.16-11.48)</th>
<th>SIDIAP_PC: 11.16 (10.31-12.08)</th>
<th>CPRD: 13.21 (12.08-14.45) (Pedianet: 2018: 2.04 (0.66-6.33))</th>
<th>BIFAP_PC_HOSP: 21.73 (20.37-23.18)</th>
<th>80.16 (77.02-83.43)</th>
<th>DCE-AU: 37.48 (35.91-39.12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.8 Cerebral venous thrombosis</td>
<td>BIFAP_PC: 0.30 (0.21-0.43)</td>
<td>SIDIAP_PC: 0.24 (0.14-0.41)</td>
<td>CPRD: 0.14 (0.06-0.34) (Pedianet: 2018: 1.26 (0.66-6.33))</td>
<td>BIFAP_PC_HOSP: 0.47 (0.30-0.73)</td>
<td>1.20 (0.87-1.66)</td>
<td>DCE-AU: 0.97 (0.74-1.27)</td>
</tr>
<tr>
<td>16. Acute kidney injury</td>
<td>BIFAP_PC: 70.74 (69.06-72.46)</td>
<td>SIDIAP_PC: 78.87 (78.46-95.93)</td>
<td>CPRD: 118.22 (114.74-121.81) (Pedianet: 2018: 1.36 (0.34-5.44))</td>
<td>BIFAP_PC_HOSP: 151.37 (147.71-155.12)</td>
<td>245.49 (239.51-251.62)</td>
<td>DCE-AU 2010: 185.64 (182.10-189.25)</td>
</tr>
<tr>
<td>17. Generalized convulsion</td>
<td>BIFAP_PC: 51.18 (49.76-52.64)</td>
<td>SIDIAP_PC: 97.11 (94.53-99.76)</td>
<td>CPRD: 99.80 (96.60-103.10) (Pedianet 2018: 1.50.85 (1.32-172.11))</td>
<td>BIFAP_PC_HOSP: 79.87 (77.22-82.61)</td>
<td>160.34 (155.87-164.94)</td>
<td>DCE-AU 2010: 219.70 (215.84-223.63)</td>
</tr>
<tr>
<td>18. Meningoencephalitis</td>
<td>BIFAP_PC: 4.24 (3.84-4.68)</td>
<td>SIDIAP_PC: 1.81 (1.49-2.20)</td>
<td>PHARMO_PC: 6.28 (6.06-6.50)</td>
<td>CPRD: 2.67 (2.19-3.26) (Pedianet 2018: 2.31 (0.75-7.16))</td>
<td>BIFAP_PC_HOSP: 8.50 (7.67-9.43)</td>
<td>PHARMO_PC_HOSP: 14.48 (11.55-18.16)</td>
</tr>
<tr>
<td>19. Transverse myelitis</td>
<td>BIFAP_PC: 0.16 (0.10-0.27)</td>
<td>SIDIAP_PC: 0.71 (0.52-0.97)</td>
<td>CPRD: 0.15 (0.76-1.44)</td>
<td>Pedianet: NA</td>
<td>BIFAP_PC_HOSP: 0.02 (0.00-0.14)</td>
<td>PHARMO_PC_HOSP: 0.19 (0.03-1.35)</td>
</tr>
<tr>
<td>22. Chilblain-like lesions</td>
<td>BIFAP_PC: 31.36 (30.25-32.51)</td>
<td>SIDIAP_PC: 25.02 (23.56-26.57)</td>
<td>CPRD: 12.08 (11.00-13.26)</td>
<td>Pedianet 2018: 2.72 (1.02-7.25)</td>
<td>BIFAP_PC_HOSP: 25.02 (23.56-26.57)</td>
<td>PHARMO_PC_HOSP: 1.54 (0.77-3.08)</td>
</tr>
<tr>
<td>23. Anosmia, Ageusia</td>
<td>BIFAP_PC: 14.17 (13.43-14.95)</td>
<td>SIDIAP_PC: 19.41 (18.28-20.61)</td>
<td>CPRD: 22.38 (20.89-23.97)</td>
<td>Pedianet: NA</td>
<td>BIFAP_PC_HOSP: 9.19 (8.32-10.15)</td>
<td>PHARMO_PC_HOSP: 1.16 (0.52-2.58)</td>
</tr>
<tr>
<td>26. MIS</td>
<td>BIFAP_PC: 0.59 (0.45-0.77)</td>
<td>SIDIAP_PC: 0.31 (0.19-0.50)</td>
<td>CPRD: 0.50 (0.32-0.86)</td>
<td>Pedianet (2018): 3.40 (1.42-8.17)</td>
<td>BIFAP_PC_HOSP: 0.66 (0.46-0.96)</td>
<td>PHARMO_PC_HOSP: 1.54 (0.77-3.08)</td>
</tr>
<tr>
<td>27. Death</td>
<td>BIFAP_PC: 87.53 (869.79-881.71)</td>
<td>CPRD: 860.68 (851.19-870.27)</td>
<td>Pedianet: not reported</td>
<td>BIFAP_PC_HOSP: 955.43 (946.16-964.79)</td>
<td>1375.94 (1362.75-1389.26)</td>
<td>DCE-AU (2010): 985.76 (977.57-994.02)</td>
</tr>
<tr>
<td>28. Sudden death</td>
<td>BIFAP_PC: 2.22 (1.94-2.54)</td>
<td>BIFAP_PC_HOSP: 1.87 (1.50-2.33)</td>
<td>2.13 (1.67-2.72)</td>
<td>DCE-AU (2010): 0.82 (0.61-1.09)</td>
<td>No difference</td>
<td>Increase with age</td>
</tr>
<tr>
<td>29. VTE with TP</td>
<td>BIFAP_PC: 0.08 (0.04-0.16)</td>
<td>BIFAP_PC_HOSP: 0.19 (0.10-0.38)</td>
<td>1.00 (0.70-1.43)</td>
<td>DCE-AU: 0.29 (0.18-0.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. VTE without TP</td>
<td>BIFAP_PC: 190.77 (188.00-193.58)</td>
<td>BIFAP_PC_HOSP: 226.58 (222.08-231.17)</td>
<td>182.06 (177.30-186.95)</td>
<td>DCE-AU: 181.17 (177.67-184.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. CVST with TP</td>
<td>SIDIAP_PC: 0.04 (0.01-0.16)</td>
<td>-</td>
<td>0.03 (0.00-0.21)</td>
<td>DCE-AU: 0.02 (0.00-0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. CVST without TP</td>
<td>BIFAP_PC: 0.32 (0.22-0.46)</td>
<td>BIFAP_PC_HOSP: 0.47 (0.30-0.73)</td>
<td>1.16 (0.83-1.62)</td>
<td>DCE-AU: 0.97 (0.74-1.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Arterial with TP</td>
<td>BIFAP_PC: 0.14 (0.06-0.24)</td>
<td>BIFAP_PC_HOSP: 0.14 (0.06-0.31)</td>
<td>1.46 (1.09-1.96)</td>
<td>DCE-AU: 0.38 (0.25-0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. Arterial without TP</td>
<td>BIFAP_PC: 257.63 (254.41-260.59)</td>
<td>BIFAP_PC_HOSP: 403.76 (399.40-411.55)</td>
<td>654.49 (645.37-663.74)</td>
<td>DCE-AU: 468.67 (462.91-474.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. Arterial or VTE with TP</td>
<td>BIFAP_PC: 0.22 (0.14-0.34)</td>
<td>BIFAP_PC_HOSP: 0.33 (0.20-0.56)</td>
<td>2.43 (1.93-3.06)</td>
<td>DCE-AU: 0.63 (0.45-0.88)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. Arterial or VTE without TP</td>
<td>BIFAP_PC: 443.58 (439.34-447.80)</td>
<td>BIFAP_PC_HOSP: 621.96 (614.46-629.55)</td>
<td>822.37 (812.13-832.74)</td>
<td>DCE-AU: 639.42 (632.80-646.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedianet: 1.36 (0.34-5.44)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2 List of abbreviations

ACCESS  vACCine covid-19 monitoring readinESS
ADDEM  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)

<table>
<thead>
<tr>
<th>Collaborating Institutions (by alphabetical order)</th>
<th>Study Sites</th>
<th>Key persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarhus University (DCE-AU)</td>
<td>Denmark</td>
<td>Vera Ehrenstein, Reimar W. Thomsen, Johnny Kahlert</td>
</tr>
<tr>
<td>Agenzia Regionale di Sanita Toscana (ARS)</td>
<td>Italy</td>
<td>Rosa Gini, Claudia Bartolini, Olga Paoletti</td>
</tr>
<tr>
<td>Bordeaux PharmacoEpi (BPE), University of Bordeaux</td>
<td>France</td>
<td>Cecile Droz, Nicholas Moore</td>
</tr>
<tr>
<td>Leibniz Institute for Prevention Research and Epidemiology - BIPS</td>
<td>Germany</td>
<td>Ulrike Haug, Tania Schink</td>
</tr>
<tr>
<td>FISABIO</td>
<td>Spain</td>
<td>Javier Diez-Domingo, Ainara Miraglesias, Carlos Vergara-Hernández</td>
</tr>
<tr>
<td>IDIAP-Jordi Gol</td>
<td>Spain</td>
<td>Bonaventura Bolibar, Villalobos Felipe</td>
</tr>
<tr>
<td>PHARMO/STIZON</td>
<td>The Netherlands</td>
<td>Josine Kuipers &amp; Michiel Meulendijk</td>
</tr>
<tr>
<td>RIVM</td>
<td>The Netherlands</td>
<td>Hester de Melker</td>
</tr>
<tr>
<td>RTI-HS</td>
<td>Spain &amp; United States of America</td>
<td>Susana-Perez-Gutthann Alejandro Arana</td>
</tr>
<tr>
<td>SoSeTe-Pedianet</td>
<td>Italy</td>
<td>Carlo Giaquinto, Elisa Barbieri, Luca Stona</td>
</tr>
<tr>
<td>Spanish Agency of Medicines and Medical Devices (AEMPS)</td>
<td>Spain</td>
<td>Consuelo Huerta, Mar Martín-Pérez, Patricia García-Poza; Airam</td>
</tr>
</tbody>
</table>
4 Other responsible parties

None

5 Milestones

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Planned date</th>
<th>Actual date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start of data collection</td>
<td>October 1st, 2020</td>
<td>November 1st, 2020</td>
<td>Protocol approvals</td>
</tr>
<tr>
<td>End of data collection</td>
<td>December 1</td>
<td>To be updated</td>
<td>Partial data are delivered in this report.</td>
</tr>
<tr>
<td>Registration in the EU PAS register</td>
<td>-</td>
<td>September 13, 2020</td>
<td>-</td>
</tr>
<tr>
<td>Study progress report 1</td>
<td>December 15, 2020</td>
<td>December 15, 2020</td>
<td>Due to governance approvals delays and work performed to validate the data workflow process, only partial data has been provided.</td>
</tr>
<tr>
<td>Final draft report of study results</td>
<td>December 15, 2020</td>
<td>February 15, 2021</td>
<td>Due to governance approvals delays and work performed to validate the data workflow process, only partial data has been provided.</td>
</tr>
</tbody>
</table>
| Update of final draft report and annexes (version 1.1) | | March 3, 2021 | o Update of German data (prior years), and exclusion of meanings of events indicating suspicion of diagnosis  
  o inclusion of comments of EMA on initial draft final report  
  o correction of output error in excel sheet for at risk population  
  o 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 | • Inclusion of pregnancy outcomes  
• Inclusion of SNDS data source and updates other datasources |
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-CoV2 on 11 January 2020\textsuperscript{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\textsuperscript{2}.

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


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. *(Will be delivered in Q2 2021)*
- 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

<table>
<thead>
<tr>
<th>Date</th>
<th>Amendment</th>
<th>Justification</th>
<th>Protocol Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 21, 2020</td>
<td>Adding transverse myelitis</td>
<td>Request EMA</td>
<td>Events of special interest</td>
</tr>
<tr>
<td>April 14, 2021</td>
<td>Adding thrombotic thrombocytopenia</td>
<td>Request EMA</td>
<td>Events of special interest</td>
</tr>
</tbody>
</table>
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-CoV2 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

<table>
<thead>
<tr>
<th>Country</th>
<th>Data Access Provider</th>
<th>Name Data source</th>
<th>Active population</th>
<th>Type of data source</th>
<th>Provenance for diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>BIPS</td>
<td>GePaRD*</td>
<td>16 million</td>
<td>Health insurance</td>
<td>Claims by hospitals and outpatient specialists and GPs</td>
</tr>
<tr>
<td>Netherlands</td>
<td>PHARMO</td>
<td>PHARMO</td>
<td>6 million</td>
<td>Record linkage</td>
<td>GP medical records, Hospital discharge diagnoses</td>
</tr>
<tr>
<td>Denmark</td>
<td>Aarhus University (DCE-AU)</td>
<td>Danish Registries</td>
<td>5.8 million</td>
<td>Record linkage</td>
<td>Hospital discharge, outpatient specialist diagnoses</td>
</tr>
<tr>
<td>Spain</td>
<td>AEMPS</td>
<td>BIFAP</td>
<td>10 million</td>
<td>GP medical records</td>
<td>GP medical records, communication from specialists &amp; hospitalization discharge for a subpopulation</td>
</tr>
<tr>
<td>Spain-Valencia</td>
<td>FISABIO</td>
<td>FISABIO</td>
<td>5 million</td>
<td>Record linkage</td>
<td>GP medical records, outpatient specialist &amp; Hospitalization discharge</td>
</tr>
<tr>
<td>Spain-Catalunya</td>
<td>IDIAPJGol</td>
<td>SIDIAP</td>
<td>5.7 million</td>
<td>Record linkage</td>
<td>GP medical records, communication from specialists, hospitalization discharge diagnoses(primary and secondary)</td>
</tr>
<tr>
<td>Italy</td>
<td>SoSeTe</td>
<td>PEDIANET</td>
<td>0.5 million</td>
<td>Pediatric medical record</td>
<td>Family Pediatricians &amp; communication from specialists</td>
</tr>
<tr>
<td>Italy</td>
<td>ARS</td>
<td>ARS</td>
<td>3.6 million</td>
<td>Record linkage</td>
<td>Emergency room visits Hospitalization discharge</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Utrecht University</td>
<td>CPRD-Gold</td>
<td>13 million</td>
<td>GP medical records</td>
<td>GP medical records &amp; communication from specialists</td>
</tr>
<tr>
<td>France</td>
<td>BPE</td>
<td>SNDS</td>
<td>67 million</td>
<td>Health insurance</td>
<td>Claims by Hospital, and outpatient specialists, GP</td>
</tr>
</tbody>
</table>

GP: General practitioner
*BIPS: extracted only data from 1 statutory health insurance which represents 0.5 million of people, data extraction process for the largest health insurer is ongoing.
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).

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 will be 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 will be 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 narcolepsy1. 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.

Table 5. List of AESI

<table>
<thead>
<tr>
<th>Body system / Classification</th>
<th>AESI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto-immune diseases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Guillain-Barré Syndrome</td>
</tr>
<tr>
<td></td>
<td>Acute disseminated encephalomyelitis</td>
</tr>
<tr>
<td></td>
<td>Narcolepsy</td>
</tr>
<tr>
<td></td>
<td>Acute aseptic arthritis</td>
</tr>
<tr>
<td></td>
<td>Type 1 Diabetes</td>
</tr>
<tr>
<td></td>
<td>(Idiopathic) Thrombocytopenia</td>
</tr>
<tr>
<td>Cardiovascular system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acute cardiovascular injury including: Microangiopathy, Heart failure, Stress cardiomyopathy, Coronary artery disease, Arrhythmia, Myocarditis</td>
</tr>
<tr>
<td>Circulatory system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coagulation disorders including: Disseminated intravascular coagulation, Venous thromboembolism (including Pulmonary embolism and Deep vein thrombosis), Thrombotic microangiopathy, Hemorrhagic stroke, Ischemic stroke, Cerebral sinus thrombosis, thrombotic thromcytopenia syndrome (TTS)</td>
</tr>
<tr>
<td></td>
<td>Single Organ Cutaneous Vasculitis</td>
</tr>
<tr>
<td>Hepato-gastrointestinal and renal system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acute liver injury</td>
</tr>
<tr>
<td></td>
<td>Acute kidney injury</td>
</tr>
<tr>
<td>Nerves and central nervous system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Generalized convulsion</td>
</tr>
<tr>
<td></td>
<td>Meningoencephalitis</td>
</tr>
<tr>
<td></td>
<td>Transverse myelitis</td>
</tr>
<tr>
<td>Respiratory system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acute respiratory distress syndrome</td>
</tr>
<tr>
<td>Skin and mucous membrane, bone and joints system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erythema multiforme</td>
</tr>
<tr>
<td></td>
<td>Chilblain – like lesions</td>
</tr>
<tr>
<td>Other system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anosmia, ageusia</td>
</tr>
<tr>
<td></td>
<td>Anaphylaxis</td>
</tr>
<tr>
<td></td>
<td>Multisystem inflammatory syndrome in children</td>
</tr>
<tr>
<td></td>
<td>Death (any causes)</td>
</tr>
<tr>
<td></td>
<td>COVID-19 disease (by levels of severity)</td>
</tr>
<tr>
<td></td>
<td>Sudden death</td>
</tr>
<tr>
<td>Pregnancy outcome - Maternal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gestational Diabetes</td>
</tr>
<tr>
<td></td>
<td>Preeclampsia</td>
</tr>
<tr>
<td></td>
<td>Maternal death</td>
</tr>
</tbody>
</table>

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 Disease Control and National Health Services websites when we wrote the protocol⁴.

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

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 severity⁵

<table>
<thead>
<tr>
<th>At-risk medical conditions</th>
<th>Medicinal product proxy(ies) (ATC code)</th>
</tr>
</thead>
</table>

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=8dHqQmMa7kW7URDzEbQkAIR57zM6WItos9bXY6uP0kZnnBf1hpil-1960461856

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 2 Diabetes

- Blood glucose lowering drugs, excluding insulins (A10B)

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)
- Other hematological agents (B06AX)

Negative Control Conditions

- Medicinal product proxy(ies) (ATC code)

Colonic Diverticulitis

- 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
• 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
• Published background rates
• Extracted codes (upon characterization)
• Study design related information
  o Estimated lag time from onset to diagnosis
  o Is condition a contraindication to any vaccination?
  o Is this a chronic or potentially recurrent condition?
  o Does this condition cause increased fatality?
  o 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 (see Annex 1). The codes, classified as narrow (specific) and possible (codes that may comprise the event) that were utilized to extract events and covariates are attached as excel sheet in Annex 1.

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

---

6 https://vac4eu.org/codemapper/
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 500,000 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 Atención 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 version of the database with information until December 2019 includes clinical information of 10.153 GPs 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 8.6 years. Diagnoses are classified according to the International Classification of Primary Care (ICPC)-2,ICD-9 code and SNOMEDCT system. Information on hospital discharge diagnoses is being progressively included. 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 Agency (VHA) and the regional Government of Valencia: (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 dispensions 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.³ 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 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 Sanità’ 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, 10 http://www.encepp.eu/encepp/viewResource.htm?id=24417

03 Draft final report with partial results 30-04-2021
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 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) 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 (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.

https://www.snds.gouv.fr/SNDS/Qu-est-ce-que-le-SNDS
9.5 Bias

This draft final study report includes data from 9 data sources in 6 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 contributed GP data, BIFAP and 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. 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 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 130.6 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.

9.6 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 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/).
9.6.1 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 (e.g., 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.

9.6.2 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.

9.6.3 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.

9.6.4 Storage

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

9.6.5 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.

9.6.6 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.

9.7 Statistical methods

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

9.7.1 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).

9.7.2 Hypotheses

Not applicable. This study is not hypothesis testing.

9.7.3 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 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-CoV2 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.
9.7.4 Statistical Analysis

9.7.4.1 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.

- *Incidence rate (and 95% CI) 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, 2020 pre-SARS-CoV2 period and 2020 SARS-CoV2 period) and each data source. The denominator is the total pregnancies at risk among pregnant women in each calendar year and each data source. (Pending for next update)*

- 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-months 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.

9.7.4.2 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

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

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

9.7.5 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. Data access was delayed for the PHARMO database because of the need to remotely access the data. IDIAP-Jordi Gol, had lack of capacity to start early because of COVID-19 related research overload.

- Data management workflow: this study tested 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.

- In this report, data from ARS, PEDIANET, FISABIO, BIFAP, SIDIAP, CPRD, PHARMO (only coagulation disorders) and DCE-AU are included. We anticipate data from SNDS (BPE) to be available by end of Q2 2021. Updated data from GePard including data from the largest health insurer in Germany will be available by Q2 2021. Timely linkage to medical birth registers is challenging, which requires more work on defining start of pregnancy. Pregnancy data should be available by end of Q2 2021.

10.1 Updates from March 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 are currently ongoing to understand possible issue related to meaning of events. Until this is finalized data from GePard have been removed from this report. Updated data including data from the largest health insurer in Germany will be available by Q2 2021.

- On request of EMA and in collaboration with the Early Covid Vaccine Monitor project, background rate data for the year 2020 were generated.

- As per EMA’s request, coagulations disorders have been classified in 6 subtypes. The subtypes include Disseminated intravascular coagulation, Venous thromboembolism (including Pulmonary embolism and Deep vein thrombosis), Thrombotic microangiopathy, Hemorrhagic stroke, Ischemic stroke, Cerebral sinus thrombosis.

- A refinement of the concepts and classification as broad/narrow has been conducted for all events. This revision has little impact on the previous data. Main changes were made for the following events: 1/ Type 1 diabetes mellitus medical codes now exclude medical codes possibly related to Type 2 diabetes; 2/ Thrombocytopenia narrow now includes secondary thrombocytopenia; 3/ reclassification of coagulations disorders into 6 subtypes. 4) broadening of VTE codes

- Updated data including 2020, when possible and a revision of the codelist resulting in a new data extraction process for all DAPs.

- 2020 Data from FISABIO and PHARMO could not be updated in time with newest data for this new report. Therefore, data extracted for the February 2021 reported were included. Updated data will be available for the final report by Q2 2021.

- Data from BIFAP includes all 8 regions as per protocol.

- Graphical presentations of monthly data are included in Annex 9.

- Subpopulations based on only primary care or primary care & hospital are described for PHARMO and
BIFAP. Similar subpopulations for SIDIAP will be delivered for the final report by Q2 2021.

- Listing of the background rates on the dashboard [https://vac4eu.org/covid-19-tool/](https://vac4eu.org/covid-19-tool/)

In this report, the incidence rates of COVID-19 according to severity level (Level 1 to 5 as per WHO classification) are included.

The incidence rates of AESIs in the at-risk population are available for all data sources included in this report. Changes according to plan are mentioned here below:

- The incidence rates for pregnancy outcomes could not be delivered in this report. Algorithms for pregnancy start and end dates, that do not require birth registers are currently under construction, the incidence rates should be available by end of Q2 2021.
- Due to late submission of data, data from SNDS (BPE) could not be delivered on time. Those data will be delivered in Q2 2021.
- 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.

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 – 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 39 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 35,573,297 subjects were included in the study, contributing to 117 million person-years. The largest contribution in person-time was from BIFAP (23.9%) followed by DCE-AU (19.2%), FISABIO (17.4%), SIDIAP (16.2%), CPRD (10.9%), ARS (10.4%), PHARMO (1.9%) and PEDIANET (0.30%). Figure 3 presents the person-time contribution by database over the study period.

Two databases, BIFAP 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% and 67.6% of the full population for BIFAP and PHARMO, respectively (Table 9).
### Table 8  Study flowchart

<table>
<thead>
<tr>
<th>Subjects disposition</th>
<th>ARS</th>
<th>PEDANET</th>
<th>AUH</th>
<th>CPRD</th>
<th>FISABIO</th>
<th>BIFAP</th>
<th>BIFAP_PC_HOSP</th>
<th>SIDIAP_PC</th>
<th>SIDIAP_PC_HOSP</th>
<th>PHARMO_PC</th>
<th>PHARMO_PC_HOSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects present during 2016-2020</td>
<td>390398</td>
<td>100</td>
<td>199174</td>
<td>100</td>
<td>614344</td>
<td>100</td>
<td>14007619</td>
<td>100</td>
<td>5903957</td>
<td>100</td>
<td>9853916</td>
</tr>
<tr>
<td>Subjects without sex or missing birth date</td>
<td>424</td>
<td>0.11</td>
<td>15</td>
<td>0.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Subjects without valid birth date</td>
<td>424</td>
<td>0.11</td>
<td>15</td>
<td>0.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Subjects with death date before study entry</td>
<td>1849</td>
<td>0.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Subjects without overlap in observed time</td>
<td>80245</td>
<td>2.06</td>
<td>17786</td>
<td>8.97</td>
<td>18886</td>
<td>3.06</td>
<td>537702</td>
<td>10.13</td>
<td>695892</td>
<td>10.62</td>
<td>3741105</td>
</tr>
<tr>
<td>Subjects without sufficient look-back period</td>
<td>285622</td>
<td>7.32</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total number of subjects included in the study</td>
<td>353202</td>
<td>90.47</td>
<td>181290</td>
<td>91.02</td>
<td>5855390</td>
<td>96.94</td>
<td>4688710</td>
<td>88.3</td>
<td>5855296</td>
<td>88.35</td>
<td>10266668</td>
</tr>
</tbody>
</table>

| Subjects present during 2016-2020                         | 390398 | 100     | 199174| 100   | 614344  | 100    | 14007619      | 100       | 5903957        | 100       | 9853916        |
| Subjects without sex or missing birth date                 | 424   | 0.11    | 15    | 0.07  | -       | -      | -             | -         | -              | -         | -              |
| Subjects without valid birth date                          | 424   | 0.11    | 15    | 0.07  | -       | -      | -             | -         | -              | -         | -              |
| Subjects with death date before study entry                | 1849  | 0.05    | -     | -     | -       | -      | -             | -         | -              | -         | -              |
| Subjects without overlap in observed time                  | 80245 | 2.06    | 17786 | 8.97  | 18886   | 3.06   | 537702        | 10.13     | 695892         | 10.62     | 3741105        |
| Subjects without sufficient look-back period               | 285622| 7.32    | -     | -     | -       | -      | -             | -         | -              | -         | -              |
| Total number of subjects included in the study             | 353202| 90.47   | 181290| 91.02 | 5855390 | 96.94  | 4688710       | 88.3      | 5855296        | 88.35     | 10266668       |

**D3 Draft final report with partial results 30-04-2021**
### Table 9: Demographics of study population

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Parameters</th>
<th>Value</th>
<th>%</th>
<th>Value</th>
<th>%</th>
<th>Value</th>
<th>%</th>
<th>Value</th>
<th>%</th>
<th>Value</th>
<th>%</th>
<th>Value</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (in years)</strong></td>
<td>Min</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30-39</td>
<td>80+</td>
<td>70-79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>BD</td>
<td>51</td>
<td>24</td>
<td>47</td>
<td>16.52</td>
<td>47.74</td>
<td>15.86</td>
<td>5.01</td>
<td>29</td>
<td>7.44</td>
<td>3.18</td>
<td>3.18</td>
</tr>
<tr>
<td><strong>People of color</strong></td>
<td>Min</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30-39</td>
<td>80+</td>
<td>70-79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>BD</td>
<td>51</td>
<td>24</td>
<td>47</td>
<td>16.52</td>
<td>47.74</td>
<td>15.86</td>
<td>5.01</td>
<td>29</td>
<td>7.44</td>
<td>3.18</td>
<td>3.18</td>
</tr>
</tbody>
</table>

**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 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, whereas for some AESIs (acute liver injury, meningoencephalitis) it stays the same. For GBS, arrhythmia and erythema multiforme, only narrow definition was available. Acute Aseptic Arthritis did not contain medical codes for narrow definition. All coagulations disorders only contained narrow medical codes.
Table 10 Number of incident events over the study period according to the narrow and broad definition for each AESI in each data source

<table>
<thead>
<tr>
<th>AESI</th>
<th>ARS</th>
<th>PEDIANET</th>
<th>DECE-AD</th>
<th>CPRD</th>
<th>FISABIO</th>
<th>SODAP</th>
<th>BIFAP_PC</th>
<th>BIFAP_PC_HOSP</th>
<th>PHARMO_PC</th>
<th>PHARMO_PC_HOSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2594</td>
<td>11</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n_broad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n_narrow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Aseptic Arthritis</td>
<td>9996</td>
<td>441</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>6102</td>
<td>9391</td>
<td>323</td>
<td>140</td>
<td>1851</td>
<td>541</td>
</tr>
<tr>
<td>Guillain Barre Syndrome</td>
<td>441</td>
<td>441</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>625</td>
<td>831</td>
<td>323</td>
<td>140</td>
<td>1851</td>
<td>541</td>
</tr>
<tr>
<td>Necrotizing fasciitis</td>
<td>754</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>64</td>
<td>831</td>
<td>323</td>
<td>140</td>
<td>1851</td>
<td>541</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>3968</td>
<td>3667</td>
<td>27</td>
<td>27</td>
<td>4307</td>
<td>2679</td>
<td>2634</td>
<td>2272</td>
<td>17305</td>
<td>140</td>
</tr>
<tr>
<td>Arterial or VTE</td>
<td>2898</td>
<td>0365</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>903</td>
<td>6007</td>
<td>9391</td>
<td>323</td>
<td>140</td>
<td>541</td>
</tr>
<tr>
<td>Acute Liver Injury</td>
<td>87569</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S0CV</td>
<td>87569</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Failure</td>
<td>9203</td>
<td>205</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>993</td>
<td>567</td>
<td>567</td>
<td>22</td>
<td>22</td>
<td>500</td>
</tr>
<tr>
<td>Venous Thromboembolism</td>
<td>19911</td>
<td>19911</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>3907</td>
<td>3907</td>
<td>19808</td>
<td>19808</td>
<td>3579</td>
<td>757</td>
</tr>
<tr>
<td>Acute Kidney Injury</td>
<td>56888</td>
<td>31565</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>97076</td>
<td>42579</td>
<td>23226</td>
<td>16026</td>
<td>84695</td>
<td>64137</td>
</tr>
<tr>
<td>Acute Liver Injury</td>
<td>3057</td>
<td>1057</td>
<td>7</td>
<td>7</td>
<td>5578</td>
<td>5207</td>
<td>821</td>
<td>793</td>
<td>7397</td>
<td>7397</td>
</tr>
<tr>
<td>Generalized Convulion</td>
<td>16964</td>
<td>523</td>
<td>523</td>
<td>523</td>
<td>45839</td>
<td>45839</td>
<td>16981</td>
<td>12040</td>
<td>46145</td>
<td>25820</td>
</tr>
<tr>
<td>Meningoencephalitis</td>
<td>2339</td>
<td>755</td>
<td>15</td>
<td>8</td>
<td>1411</td>
<td>919</td>
<td>565</td>
<td>336</td>
<td>4888</td>
<td>4888</td>
</tr>
<tr>
<td>Anaphylaxis</td>
<td>19017</td>
<td>1151</td>
<td>470</td>
<td>11</td>
<td>29600</td>
<td>1374</td>
<td>12020</td>
<td>2355</td>
<td>31492</td>
<td>45240</td>
</tr>
<tr>
<td>Eosinophilia</td>
<td>1800</td>
<td>849</td>
<td>849</td>
<td>849</td>
<td>849</td>
<td>849</td>
<td>849</td>
<td>849</td>
<td>849</td>
<td>849</td>
</tr>
<tr>
<td>Death</td>
<td>16196</td>
<td>215353</td>
<td>11286</td>
<td>11286</td>
<td>16752</td>
<td>&lt;5</td>
<td>306753</td>
<td>306753</td>
<td>306753</td>
<td>306753</td>
</tr>
<tr>
<td>Sudden death</td>
<td>247</td>
<td>247</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>87</td>
<td>87</td>
<td>69</td>
<td>69</td>
<td>402</td>
<td>402</td>
</tr>
<tr>
<td>ARDS</td>
<td>10662</td>
<td>4820</td>
<td>7</td>
<td>7</td>
<td>23054</td>
<td>20136</td>
<td>4331</td>
<td>2728</td>
<td>26100</td>
<td>24612</td>
</tr>
<tr>
<td>Chilblain like lesions</td>
<td>29</td>
<td>29</td>
<td>15</td>
<td>15</td>
<td>102</td>
<td>102</td>
<td>1405</td>
<td>1405</td>
<td>9272</td>
<td>9272</td>
</tr>
<tr>
<td>Erythema multiforme</td>
<td>1043</td>
<td>1043</td>
<td>21</td>
<td>21</td>
<td>1021</td>
<td>1021</td>
<td>946</td>
<td>946</td>
<td>2420</td>
<td>2420</td>
</tr>
<tr>
<td>Transverse Myelitis</td>
<td>170</td>
<td>131</td>
<td>320</td>
<td>275</td>
<td>130</td>
<td>106</td>
<td>55</td>
<td>55</td>
<td>107</td>
<td>89</td>
</tr>
<tr>
<td>MIES</td>
<td>1431</td>
<td>86</td>
<td>19</td>
<td>19</td>
<td>512</td>
<td>141</td>
<td>69</td>
<td>69</td>
<td>573</td>
<td>366</td>
</tr>
<tr>
<td>Arterial or VTE with TP</td>
<td>234</td>
<td>146</td>
<td>38</td>
<td>38</td>
<td>234</td>
<td>53</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arterial or VTE without TP</td>
<td>85330</td>
<td>2</td>
<td>125102</td>
<td>63348</td>
<td>34797</td>
<td>34797</td>
<td>10844</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Articular with TP</td>
<td>135</td>
<td>83</td>
<td>22</td>
<td>22</td>
<td>135</td>
<td>83</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arterial without TP</td>
<td>67845</td>
<td>69608</td>
<td>44677</td>
<td>45150</td>
<td>89182</td>
<td>34123</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVST with TP</td>
<td>3</td>
<td>2</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVST without TP</td>
<td>150</td>
<td>1</td>
<td>232</td>
<td>24</td>
<td>150</td>
<td>54</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTE with TP</td>
<td>19826</td>
<td>1</td>
<td>39625</td>
<td>17975</td>
<td>35726</td>
<td>61392</td>
<td>17708</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colonic diverticulitis</td>
<td>16142</td>
<td>&lt;5</td>
<td>42388</td>
<td>27525</td>
<td>40367</td>
<td>44720</td>
<td>48585</td>
<td>15925</td>
<td>2372</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>83212</td>
<td>36</td>
<td>367437</td>
<td>31012</td>
<td>319179</td>
<td>948098</td>
<td>204441</td>
<td>196084</td>
<td>25831</td>
<td>18797</td>
</tr>
</tbody>
</table>

Note: Data not yet reclassified based on narrow and broad for PHARMO and FISABIO

D3 Draft final report with partial results 30-04-2021
**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, this will be further improved towards the final report.

10.3.1 AESIs

10.3.1.1 Guillain Barre syndrome

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).

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 (figure 5).

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. Data generated by the OHDSI collaboration and published in a pre-print server showed significantly higher estimates.

---


Figure 4: Forest plot for incidence rates of GBS per 100,000 PY by data source and calendar year using a narrow definition.
10.3.1.2 Acute Disseminated Encephalomyelitis (ADEM)

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\textsuperscript{14}.

ADEM is rare and specific events that can only be seen with ICD-9/10CM codes were only observed in DCE-AU, ARS, BIFAP, SIDIAP and FISABIO with rates < 1/100,000 person-years (figure 6). Rates were stable over calendar time. In FISABIO, higher rates were observed in the 0-19 age category (figure 7). BIFAP subpopulations rates show that the inclusion of hospital based events really increases the event rates. For observed/expected analyses we would recommend to also use broad rates (which include encephalitis).

Figure 6: Forest plot for incidence rates of ADEM per 100,000 PY by year
Incidence rates from the literature in the USA showed an incidence rate of 0.4 per 100,000 PY\textsuperscript{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

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.

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. 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 in those DAPs compare well to the crude rates observed in the VAESCO\textsuperscript{16}, SOMNIA\textsuperscript{17} and ADVANCE\textsuperscript{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 (Danish registries, BIFAP, CPRD, SIDIAP, FISABIO) (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\textsuperscript{19}. In all data sources that capture 2020, the rates decrease in 2020. Our data differ with the recent data from the OHDSI consortium which showed rates up to 10-fold higher (15/100,000) in the age categories between 18-34.

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

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.\(^{20}\)

We could not identify narrow codes 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).

---

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

10.3.1.5 Diabetes Mellitus

Events of diabetes mellitus using the narrow definition diagnosis codes were observed in all participating databases, except ARS. The codes have been refined and excluded the use of oral glucose lowering drugs. Rates from FISABIO and PHARMO were not yet generated with the updated algorithm and this will be available in the final report.
Figure 11: Incidence rates of diabetes mellitus type 1 per 100,000 PY and calendar year

Figure 12: Incidence of Diabetes (type 1) by age
10.3.1.6 Thrombocytopenia

**Primary ITP** is an autoimmune disorder characterized by isolated thrombocytopenia (peripheral blood platelet count <100x 109/L in the absence of other causes or disorders that may be associated with thrombocytopenia. The diagnosis of primary ITP remains one of exclusion; no robust clinical or laboratory parameters are currently available to establish its diagnosis with accuracy. The main clinical problem of primary ITP is an increased risk of bleeding, although bleeding symptoms may not always be present\(^{21}\).

**Secondary TP**

All forms of thrombocytopenia except primary ITP.

Incidence rates from other data sources 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 14), which we would expect especially for secondary TP.

---

Figure 13: Incidence rates of thrombocytopenia (primary & secondary).

---

**Figure 14**: Incidence rates of thrombocytopenia (primary and secondary) by age and data source

Reference data were available from the ADVANCE study which were separated by type of source data (GP and Hospital based) (figure 15).
10.3.1.7 Microangiopathy

Events of microangiopathy using narrow definition were observed in all participating databases, except PEDIANET, which has children only. Incidence rates were in the same ranges and stable overtime (figure 15). A slight pattern of increased rates with age was observed in all databases, whereas rates lowered in 2020, potentially because of the lock down.

Figure 16: Incidence rates of microangiopathy by calendar year and data source

---

There were no background rates as benchmark in the published literature. Microangiopathy frequently occurs in diabetic patients.

10.3.1.8 Heart Failure

Heart failure was observed in all participating databases (figure 18). A clear pattern of increased rates with age was observed in all databases (figure 19). 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 \(^{23}\) (Corrao, 2014) from

---

Italy and 306/100,000 person-years from Canada 24. A recent study conducted in claims database in Germany found an incidence rate of heart failure of 655/100,000 person-years 25. A study from the US suggested increased incidence in older population 26.

Figure 18: Incidence of heart failure by data source and calendar year

---


10.3.1.9 Stress cardiomyopathy

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 20). Higher rates were observed in both inpatient databases (ARS, FISABIO). 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. We could only identify the event using ICD-9, 10 and SNOMED, no code existed in ICPC which may explain why the rate is low in BIFAP. 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 FISABIO which capture both primary and inpatient diagnoses.
Figure 20: Incidence of stress cardiomyopathy by data source and calendar year
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\textsuperscript{27}. 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\textsuperscript{28}. Jabri et al reported an increase of stress cardiomyopathy during the COVID-19 pandemic\textsuperscript{29}. The rates we observe are within the range reported by Murugiah et al.

10.3.1.10 Coronary artery disease (CAD)

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

\textsuperscript{27} 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
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. IRs differed based on the provenance of the diagnosis it was lowest 80.57/100,000 person-years in PHARMO (primary care records) to 292.81/100,000 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 22: Incidence of CAD (narrow) by data source and calendar year
Figure 23: Incidence of CAD (narrow) by data source and age
Figure 24: 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\textsuperscript{30}.

\textsuperscript{30} Data source: Global Burden of Disease Study 2017, Institute for Health Metrics and Evaluation, \url{http://ghdx.healthdata.org/gbd-results-tool}
10.3.1.11 Arrhythmia

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 26). In the year 2017, IRs ranged from 495.71/100,000 person-years in CPRD to 1118/100,000 person-years in ARS. A clear pattern of increased rates with age was observed in all databases. Variable rates were observed across years in PHARMO, and this would need to be further investigated.

From published articles, IRs of arrhythmia ranges between 208/100,000 in Denmark up to 780/100,000 person-years in UK.
Figure 26: Incidence of arrhythmia by data source and calendar year
Reference data can be obtained from the UK Biobank publication which reported rates of overall arrhythmia (figure 28):
Figure 28: Incidence of arrhythmia in the UK Biobank

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

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.

IRs were quite stable overtime in each of the data sources (figure 28) with a slight decrease in 2020. Rates ranged from 0.33/100,000 person-years in SIDIAP_PC to 11.76/100,000 person-years in BIFAP-PC, both general practitioner databases, but using different coding (ICPC versus ICD-10). Rates increased with age.

The pre-printed publication from OHDSI, rates for myocarditis/pericarditis ranged from 37/100,000 person-years (95%CI: 16-88) in males aged 18-34 to 41/100,000 person-years (95%CI: 9-193) in males aged 85+. Those rates are much higher than our observations.

10.3.1.13 Coagulation disorders

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 201032. The incidence rate of DIC was shown to increase with age in both men and women and was consistently higher in men.

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. In 2017, rates ranged from to 0.13/100,000 person-years in CPRD to 7.53/100,000 person-years in BIFAP. Rates in BIFAP are higher compared to the other databases, this is explained by the presence of an extra medical code in the SNOMED dictionary that is not available in the other coding system.

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.

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. In

2017, rates ranged from to 0.26/100,000 person-years in PHARMO to 1.24/100,000 person-years in FISABIO.

Figure 33 Incidence of TTP by data source and calendar year
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. 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). Recent data suggests an increase in incidence over time with estimates increasing from 95 to 133/100,000 person-years from 1999 to 2009 (Huang, 2014).
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. In 2017, rates ranged from to 174.70/100,000 person-years in CPRD to 190.85/100,000 person-years in BIFAP, except in PEDIANET where cases were identified only in 2018 with a rate estimated at 0.88/100,000 person-years in the pediatric population.

![Figure 35 Incidence of VTE by data source and calendar year](image)

*Figure 35 Incidence of VTE by data source and calendar year*
Figure 36 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. Two other studies found higher incidence rates than previously reported with annual rates ranging between 13.2 to 15.7 cases per million.

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. In 2017, rates ranged from to 0.14/100,000 person-years in CPRD to 1.20/100,000 person-years in ARS. These rates are aligned with data previously published.

---

Figure 37 Incidence of Cerebral venous thrombosis by data source and calendar year
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 1340/100,000 person-years in the general population from a study conducted in the Danish registries between 1997 and 2017. A clear increased age pattern was observed.

IRs were stable over time with a clear pattern of increased rates with age in all databases. In 2017, rates ranged from to 169.44/100,000 person-years in SIDIAP to 378.11/100,000 person-years in ARS.

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.

IRs were stable overtime with higher rates observed in ARS. In the year 2017, IRs ranged from 10.8/100,000 person-years in BIFAP to 80.16/100,000 person-years in ARS (figure 30). The rates were low in the pediatric PEDIANET data source. A clear age pattern was observed with rates in hospital setting reaching 405/100,000 person-years in the 80+. The magnitude of rates in ARS are close to the rates recently published by OHDSI. We

---

would recommend to use data from ARS for this event.

Figure 41 Incidence of hemorrhagic stroke by data source and calendar year
10.3.1.14 Single organ cutaneous vasculitis

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.

In the year 2017, IRs ranged from 5.56/100,000 person-years in SIDIAP to 32.88/100,000 person-years (95%CI: 31.37-34.46) in FISABIO (Figure 43). IRs were higher in children lowered and then increased again with age (Figure 44).
Figure 43 Incidence of single organ cutaneous vasculitis by data source and calendar year
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

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 (3.66/100,000 person-years in 2019). For the other databases, rates ranged from 6.16/100,000 person-years in CPRD to 42.70/100,000 person-years in FISABIO (in 2017). A clear pattern with age was observed (Figure 45 and 46).
Figure 45 Incidence of Acute Liver Injury by data source and calendar year
10.3.1.16 Acute kidney injury

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 per 100,000 PY with a narrow definition for Acute kidney injury are shown for Denmark (DCE-AU), Italy (ARS, PEDIANET), Spain (BIFAP, FISABIO, SIDIAP), and the UK (CPRD). Kidney injury IR are lowest in PEDIANET (pediatrics only) and significantly higher in the other databases. In 2017, the IRs ranged between 70.74/100,000 person-years in BIFAP and 588.87/100,000 person-years in SIDIAP with some significant differences between them. The rates of 2020 are significantly lower in FISABIO compared to the other years within their databases. There is a consistent and strong increase in incidence with increasing age across all data sources.
Figure 47 Incidence of Acute kidney Injury by data source and calendar year
Reference rates can be obtained from the worldwide meta-analysis of AKI showing a rate of 4.8% of AKI in hospitalized patients, our rates are lower as they are in the general population.

10.3.1.17 Generalized convulsion

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.\(^{43}\)

IRs from BIFAP were significantly lower than the IR from the other databases and between 35.74/100,000 person-years in 2020 and 51.18/100,000 person-years in 2017. The IRs of all the other databases ranged between 97.11/100,000 person-years in SIDIAP and 219.70/100,000 person-years in DCE-AU (Figure 49). FISABIO showed a significantly higher rate of 158.62/100,000 person-years in 2020 compared to the other years while BIFAP, SIDIAP, ARS and CPRD showed significantly lower rates in 2020. From PHARMO, only rates from the subpopulation were available, these rates were significantly lower compared to the other databases, this will be further investigated once newest data will be available.

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).

10.3.1.18 Meningoencephalitis

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.

Meningoencephalitis IRs per 100,000 PY ranging from 1.81/100,000 person-years (95% CI 1.49 – 2.20) in SIDIAP to 32.08/100,000 person-years (95% CI 30.59 – 33.64) in FISABIO (figure 51) in 2017. FISABIO showed significantly higher rates compared to the other databases, this can be explained by the capture of primary care and hospital data. Rates from the BIFAP and PHARMO subpopulations were also higher. The incidence of meningo-encephalitis increased with age (figure 52).

From the literature, IRs of encephalomyelitis in children aged between 0 to 17 years was of 0.79/100,000 person-years in the UK. Other studies conducted on the total population suggest rates of 4.3/100,000

---


person-years in Canada and 4.32/100,000 person-years in the UK. Reference rates of acute encephalitis usually are below 10/100,000 PY which is consistent with our findings. Reference rates of acute encephalitis usually are below 10/100,000 PY which is consistent with our findings.

Figure 51 Incidence of meningoencephalitis by data source and calendar year

Epub 2017 Mar 2. PMID: 28259562.


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. We did not make an event definition form because of the late entry of the event. Codes are listed in the table below and all are narrow.

Table 11 Codes for transverse myelitis

<table>
<thead>
<tr>
<th>Coding system</th>
<th>Code</th>
<th>Code name</th>
<th>Concept</th>
<th>Concept name</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD10/CM</td>
<td>G37.3</td>
<td>Acute transverse myelitis NOS</td>
<td>C0270627</td>
<td>Myelitis, Acute Transverse</td>
<td>narrow</td>
</tr>
<tr>
<td>ICD9</td>
<td>323.82</td>
<td>Transverse myelitis NOS</td>
<td>C0026976</td>
<td>Myelitis, Transverse</td>
<td>narrow</td>
</tr>
<tr>
<td>ICD9</td>
<td>341.2</td>
<td>Acute (transverse) myelitis</td>
<td>C0270627</td>
<td>Myelitis, Acute Transverse</td>
<td>narrow</td>
</tr>
<tr>
<td>ICD9</td>
<td>341.20</td>
<td>Acute (transverse) myelitis NOS</td>
<td>C0270627</td>
<td>Myelitis, Acute Transverse</td>
<td>narrow</td>
</tr>
</tbody>
</table>

Rates of transverse myelitis were available from most of the databases, except Pedianet. Rates were low and ranged between 0.16/100,000 person-years in BIFAP to 1.32/100,000 person-years in DCE-AU (Figure 53). 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.

**Figure 53** Incidence of transverse myelitis by data source and calendar year

### 10.3.1.20 Respiratory system – Acute respiratory distress syndrome

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\(^9\)

IRs from Denmark, Italy, Spain and the UK range from 11.57/100,000 person-years (95% CI 15.43 – 19.00) in SIDIAP to 143.25/100,000 person-years (95% CI 140.06 – 146.51) in FISABIO. SIDIAP and FISABIO are showing significant lower rates for 2020 compared to the other years in their database (Figure 54). Significantly higher rates were observed in ARS for the year 2020. The incidence increased consistently with age (Figure 55).

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\(^{50}\)

\[\text{Figure 54 Incidence of ARDS by data source and calendar year}\]

---


Erythema multiforme

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\textsuperscript{51}.

IRs were distributed between 2.31/100,000 person-years (CI 95% 0.75 – 7.16) from PEDIANET and 14.93/100,000 person-years from FISABIO (Figure 56). For 2020 we observe significant lower IRs of 4.17/100,000 person-years (CI 95% 3.49 – 4.96) in ARS, 3.90/100,000 person-years (CI 95% 3.45 – 4.40) in BIFAP, and 5.87/100,000 person-years (CI 95% 5.21 – 6.61) in FISABIO as compared to the other years. Rates were highest in children.

Figure 56 Incidence of erythema multiforme by data source and calendar year
Chilblain–like lesions

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 from 0.13/100,000 person-years (CI 95%: 0.05 – 0.35) from ARS to 64.25/100,000 person-years (CI 95%: 62.13 – 66.44) from FISABIO. The rates from data sources without primary care diagnoses ARS and DCE-AU are significantly lower and the rates from FISABIO are significantly higher compared to BIFAP, SIDIAP, CPRD, and PEDIANET (Figure 58). The yearly rates presented significant differences in each year for FISABIO ranging from 30.33/100,000 person-years (CI 95%: 28.79 – 31.95) in 2020 to 64.25/100,000 person-years (CI 95%: 62.13 – 66.44) in 2017.

We do not observe a clear age pattern in incidence (Figure 59).
Figure 58 Incidence of chilblain like lesions by data source and calendar year
We did not find reference rates on this condition.

10.3.1.23 Anosmia, Ageusia

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

The IRs were lowest from FISABIO with 0.30/100,000 person-years (95% CI 0.18 – 0.49) and highest from SIDIAP with 19.41/100,000 person-years (95%: 18.28-20.61) in 2017. Significantly higher rates were observed in the year 2020 in all databases, except CPRD. No events were detected in ARS which comprise hospitalisation records only (Figure 60).

Using a broad definition, the rates of BIFAP increased (Figure 61), due to the wide code N16 (ICPC), and were comparable with other data sources that capture primary care. Hospital /specialist-based data sources (ARS, DCE-AU) may underestimate the incidence and recommend to use from primary care based data sources.

---

Figure 59 Incidence of chilblain like lesions by data source and age

Figure 60 Incidence rate (+ 95% CI) for chilblain like lesions by data source and age.
Figure 60 Incidence of anosmia/ageusia by data source and calendar year
Figure 61 Incidence of anosmia/ageusia (broad definition) by data source and calendar year
10.3.1.24 Anaphylaxis

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\footnote{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}.

Anaphylaxis IRs vary between 1.54/100,000 person-years (95% CI 0.39 – 6.16) in PEDIANET to 24.43/100,000 person-years (95% CI 23.14 – 25.80) in FISABIO (figure 63). The rates from CPRD and BIFAP were significantly lower for 2020 with an IR of 13.14 and 4.35/100,000 person-years, respectively, compared to the other years. The rate of anaphylaxis is lower in the elderly (figure 64).
Figure 63 Incidence of anaphylaxis by data source and calendar year
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. Rates of similar magnitude and similar pattern with lower rates in elderly were observed from the OHDSI collaboration.

10.3.1.25 Multisystem inflammatory syndrome

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 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 year. Overall, IRs were low ranging between 0.28/100,000 person-years in SIDIAP in 2018 to 6.94/100,000 person-years in PEDIANET, which is pediatric only (Figure 65).
Figure 65 Incidence of multi-system inflammatory condition by data source and calendar year

Monthly incidence rates of MISC are available for children in Annex 2 and 5 (for narrow and broad definition respectively). Monthly rates in children are graphically depicted in Figure 66. Six databases provided data in 2020: ARS, PEDIANET, CPRD, BIFAP, SIDIAP and FISABIO.
Figure 66 Incidence of multi-system inflammatory condition by data source and month in children
Figure 67 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 SIDIAP. The incidence rates of death were significantly higher in ARS, which contains exclusively hospital data. Higher rates were observed in the year 2020 in CPRD, BIFAP and in FISABIO in 2020 compared to the other years. A clear pattern of increased incidence rates with age was observed across databases.
Figure 68 Incidence of death by data source and calendar year
10.3.1.27 Sudden death

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.\textsuperscript{53}

For sudden death, we expect that this diagnosis may be underestimated, if the causes of death are not well recorded. Very few cases were observed in 0-19, most data sources did not have rates in the younger age categories (figure 71). As for death, no sudden death could be identified from SIDIAP, this will be further investigated.

\textsuperscript{53} International classification of diseases (ICD-10). Geneva, World Health Organization, 2005
Figure 70 Incidence of sudden death by data source and calendar year
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. However, data from BIFAP needs further validation and, therefore they are not presented in Figure 72. For the other database, 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-10 medical codes and COVID 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:

- Level 1: any 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.
10.3.1.29  Thrombosis with Thrombocytopenia (TTS)

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 with thrombocytopenia were extremely low with a higher rate of 1.46/100,000 person-years for arterial thrombosis with thrombocytopenia.

10.3.1.29.1  VTE with and without TP

Figure 72  Incidence of COVID in 2020 by data source and age
Figure 73 Incidence of VTE with TP by data source and calendar year
Figure 74 Incidence of VTE without TP by data source and calendar year

10.3.1.29.2 CVST with and without TP
Figure 75: Incidence of CVST with TP by data source and calendar year
Figure 76 Incidence of CVST without TP by data source and calendar year

10.3.1.29.3 Arterial thrombotic events with and without TP
Figure 77 Incidence of Arterial thrombotic events with TP by data source and calendar year
Figure 78 Incidence of Arterial thrombotic events without TP by data source and calendar year

10.3.1.29.4 Arterial thrombotic events or VTE with and without TP
Figure 79 Incidence of Arterial or VTE with TP by data source and calendar year
10.3.1.30  Pregnancy outcomes (maternal and neonates)

Incidence rates for pregnancy outcomes could not yet be delivered in this report and will be delivered in the final report.

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 73). IRs were stable overtime. Significantly lower rates were observed in 2020 in ARS and CPRD while significantly higher rates were observed in FISABIO for the same year.
Hypertension using narrow definition 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 with age were observed in all databases.
Figure 82: Incidence of hypertension by data source and calendar year
Figure 83 Incidence of hypertension by data source and age
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 42 AESIs, except pregnancy outcomes for which algorithms are currently under validation as part of the IMI-funded ConcePTION project. Incidence rates for pregnancy outcomes will be available by Q2 2021.

In this report, a refinement of medical codes according to narrow and broad definitions has been conducted. As a consequence, the coagulations disorders have been classified according to 6 subtypes. In addition, rates for co-occurrence of thrombosis with thrombocytopenia have been generated. Type 1 diabetes is more specific and no longer includes medical codes that could be attributed to Type 2 diabetes. Overall, the refinement of the medical codes has little impact on the rates that were produced for the previous version of the report.

This report comprises data from 5 countries (UK, ES, IT, DK, NL) and 8 data sources (BIFAP, Pedianet, CPRD, ARS, Danish registries, FISABIO, SIDIAP, PHARMO). Data from France (SNDS) and SIDIAP (ES), will be added in the next version. Because of high rates that were observed from the German database, data from Germany (GePard) are currently under investigation. In addition, the team will extract data from the largest healthcare insurer in Germany that will allow for a highest representativeness of the population. The refined medical code lists could not be implemented for PHARMO and FISABIO, for this reason data previously generated from FISABIO and PHARMO are presented. A total number of 39 million individuals were included contributing to 117 million person-years.

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. For the final updates of the report we will investigate the impact of the meaning/provenance of a diagnosis.

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. We will extend this further in the next update 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 2 databases (PHARMO and BIFAP), a third database (SIDIAP) will also provide information of subpopulation in the final report.

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. Updates will be made with additional databases (BPE), pregnancy events and a component analyses assessing the impact of the provenance of data.
## Annexes

### 15 Annex 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.]

<table>
<thead>
<tr>
<th>Number</th>
<th>Document reference number</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Annex 1</td>
<td>Event definition forms</td>
</tr>
<tr>
<td>2</td>
<td>Annex 2</td>
<td>Excel sheet – IR narrow</td>
</tr>
<tr>
<td>3</td>
<td>Annex 3</td>
<td>Forest plot – by year - narrow</td>
</tr>
<tr>
<td>4</td>
<td>Annex 4</td>
<td>Forest plot – by age - narrow</td>
</tr>
<tr>
<td>5</td>
<td>Annex 5</td>
<td>Excel sheet – IR broad</td>
</tr>
<tr>
<td>6</td>
<td>Annex 6</td>
<td>Forest plot – by year - broad</td>
</tr>
<tr>
<td>7</td>
<td>Annex 7</td>
<td>Forest plot – by age - broad</td>
</tr>
<tr>
<td>8</td>
<td>Annex 8</td>
<td>Forest plot – IR at risk population</td>
</tr>
<tr>
<td>9</td>
<td>Annex 9</td>
<td>Forest plot – Monthly IR</td>
</tr>
</tbody>
</table>