Study Report

P2-C1-012

Monitoring prescription of medicines for public health emergencies at risk of shortages

11/07/2024

Version 3.1

TABLE OF CONTENTS

DOCUMENT HISTORY

Version	Date	Description
V1.0	02/04/2024	**Final Version for EMA review**
V2.0	02/05/2024	**Version incorporating EMA comments after first assessment**
V3.0	30/05/2024	**Final Version after acceptance**
V3.1	17/07/2024	**Final report uploaded in the HMA-EMA Catalogue**

Study Title	DARWIN EU® – Monitoring prescription of medicines for public health emergencies at risk of shortages
Study Report Version identifier	V3.1
Dates Study Report updates	11/07/2024
EU PAS register number	EUPAS107932
Active substance	

Antibiotic	Class	ATC code
Azithromycin	Macrolides	J01FA10
Clarithromycin	Macrolides	J01FA09
Phenoxymethylpenicillin	Beta-lactamase sensitive penicillins	J01CE02
Benzylpenicillin	Beta-lactamase sensitive penicillins	J01CE01
Amoxicillin	Penicillins with extended spectrum	J01CA04
Amoxicillin + clavulanic acid	Combinations of penicillins	J01CR02
Ceftriaxone	Third-generation cephalosporins	J01DD04
Cefotaxime	Third-generation cephalosporins	J01DD01
Meropenem	Carbapenems	J01DH02
Cefuroxime	Second-generation cephalosporins	J01DC02
Piperacillin + Tazobactam	Combinations of penicillins	J01CR05

Medicinal product	N/A
Research question and objectives	This study aims to characterise the incidence of use (prescription/dispensation) of 11 antibiotics to understand time trends, cycles and seasonality in the use of those medicines; and to forecast the use rates of such medicines under assumed scenarios, which could help anticipate and prevent potential shortages, or manage them. To estimate monthly incidence rates of use of the 11 selected medicines during a 10-year period counting backwards from the most recent data available (e.g., 2014-2023) stratified by age and sex, in each of the databases. To conduct time series modelling by fitting an ARIMA model to data

	generated in objective 1 for short-term (6-month) forecasting.
Country(-ies) of study	Belgium, Germany, Spain and the UK
Author	Marta Pineda Moncusí Yuchen Guo Daniel Prieto-Alhambra

1. DESCRIPTION OF STUDY TEAM

Study team Role	Names	Organisation
Study Project Manager/Principal Investigator	Marta Pineda Moncusí	University of Oxford
Data Scientist	Yuchen Guo Edward Burn	University of Oxford University of Oxford
Epidemiologist	Daniel Prieto Alhambra	University of Oxford and Erasmus MC University
Data Analyst	Yuchen Guo	University of Oxford
Data Partners*	**Names**	**Organisation**
Local Study Coordinator/ Data Analyst	Talita Duarte Salles	IDIAP Jordi Gol, Erasmus MC
	Laura Pérez-Crespo Irene López Sánchez	IDIAP Jordi Gol
	Jasmine Gratton Dina Vojinovic James Brash	IQVIA
	Angela Leis Machin Miguel Angel Mayer Pujadas Juan Manuel Ramirez Anguita	IMASIS

*Data partners' role is only to execute code at their data source, and they do not have an investigator role.

2. DATA SOURCES

Country	Name of Database	Health Care setting	Type of Data	Number of active subjects	Calendar period covered by each data source
UK	CPRD GOLD	Primary Care	EHR	3 million	09/09/1987 to 23/06/2023
Spain	IMASIS	Hospital	EHR	0.6 million	01/01/1990 to 09/09/2023
Germany	IQVIA DA Germany	Primary & Secondary Care	EHR	41.9 million	01/01/1992 to 30/06/2023
Belgium	IQVIA LPD Belgium	Primary Care	EHR	8.5 million	01/06/2013 to 30/06/2023
Spain	SIDIAP	Primary Care	EHR	5.8 million	01/01/2006 to 30/06/2023

CPRD = Clinical Practice Research Datalink GOLD, IMASIS= Institut Municipal Assistència Sanitària Information System, DA = Disease Analyzer, LPD = Longitudinal Patient Data, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària, EHR = Electronic Heath records.

3. ABSTRACT (STAND-ALONE SUMMARY OF THE STUDY REPORT)

Title

DARWIN EU® - Monitoring prescription of medicines for public health emergencies at risk of shortages

Rationale and Background

Scientific and commercial data on monthly prescriptions of medicines that may be critical in public health emergencies (PHE) can help understanding time trends and seasonal variations. In conjunction with time series and forecasting models, as well as data on medicines supply, such data can contribute to the on-going efforts of the European Medicines Agency to better monitor and coordinate its response to shortages of critical medicines.

Research question and objectives

1) To estimate monthly incidence rates of use (prescription or dispensation) of 11 selected medicines during a 10-year period counting backwards from the most recent data available, stratified by age and sex, in each of the databases.
2) To conduct time series modelling by fitting an ARIMA model to data generated in objective 1 for short-term (6-month) forecasting.

Research Methods

Study design

- Population-level cohort (Objective 1, Population-level drug utilisation study)
- Time series modelling (Objective 2, Time series modelling based on Objective 1 findings)

Population

All individuals present in the database in the last 10 years of available data were included in the analysis. For this population, incidence of use of antibiotics was estimated.

Variables

Drugs of interest: list of 11 antibiotics that may be critical in PHE.

Calendar month, age, and sex were used for stratification.

Data sources

IQVIA LPD Belgium [IQVIA Belgium] (Primary Care Database)

1. CPRD GOLD [CPRD] (UK, Primary Care Database)
2. SIDIAP (Spain, Primary Care Database)
3. IMASIS (Spain, Secondary Care Database)
4. IQVIA DA Germany [IQVIA Germany] (Germany, combination of primary and secondary care (outpatient visits) database).

Sample size

No sample size was calculated for this study. The expected number of prescriptions/dispensations across data sources was anticipated to be roughly between <1000 and 25 million.

Data analyses

Population-level drug utilisation study on antibiotics: estimation of monthly incidence rates of antibiotic use (prescription or dispensation) per 100,000 person-years, as described in section 9.9.2.1.

Time series modelling: forecast of the 6-month incidence rates of antibiotic use after the end of available data using seasonal Auto-Regressive Integrated Moving Average (sARIMA) models, and sARIMA models incorporating the effect of the COVID-19 pandemic as an exogenous variable (sARIMAX), for time series analysis, as described in section 9.9.3.3.

Results

Population-level cohort study

The results showed consistent patterns across all 5 databases, with seasonal patterns of use (prescription or dispensation) of most of the antibiotics in winter, and a substantial drop after the start of the COVID-19 pandemic (March 2020). The age-stratified analysis showed the use of antibiotics was generally higher in individuals aged 65 or older. When stratified by sex, women had a generally higher incidence in primary care (CPRD, IQVIA Belgium and SIDIAP); whilst men presented generally higher use in hospital (IMASIS) and outpatient specialist data (IQVIA Germany).

Cefuroxime use decreased over time in CPRD, benzylpenicillin decreased in SIDIAP, and phenoxymethylpenicillin in IQVIA Belgium was discontinued after September 2019. A few peaks at isolated time points were observed in the use rates, for example in the use of amoxicillin, azithromycin, clarithromycin, amoxicillin/clavulanate and phenoxymethylpenicillin during December 2022 in CPRD.

Crude number of antibiotic users in IMASIS, IQVIA Belgium and IQVIA Germany were increased during winter 2022 when compared to previous months. However, incidence rates estimated in the last 6 months of data available in these three databases (which contained the winter months from 2022), were inaccurate due to artefactual increases in rates of antibiotic use associated with the computation of the observation period in IMASIS, IQVIA Belgium and IQVIA Germany.

Time series modelling

Based on the model diagnostics Mean Absolute Error and Mean Absolute Percentage Error, sARIMAX models fitted the trend changes due to the impact of the COVID-19 pandemic better than the models without a term for the pandemic (i.e., sARIMA). Moreover, sARIMAX models demonstrate adequate levels of goodness of fit, which validates the reliability of their respective coefficients.

A number of observations were out of the 95% predictive intervals of the forecast based on sARIMAX, including the use of amoxicillin, azithromycin, clarithromycin, amoxicillin/clavulanate, phenoxymethylpenicillin in December 2022 in CPRD when compared with the predicted values.

Despite the exclusion of the latest 6 months of data available in IMASIS, IQVIA Belgium and IQVIA Germany, observed values after mid-2022 for IMASIS, and end-2022 for both IQVIA databases may still contain artefacts related to artefactual decrease in the denominator. The observed incidence rates from these periods (from the latest 6 to 12 months of available data) were used for tuning the sARIMA(X) models. The values forecasted for the tune periods follow a reasonable trend despite the differences with the observed values. IQVIA Germany seems the least affected of these three databases. To compensate for the months excluded in IMASIS and IQVIA databases, we provided an extension of the forecasted values (from 6 months to 12 months) as a sensitivity analysis, which provides an insight in the trends of the antibiotics use over a more relevant and recent period of time. However, considering that uncertainty of predicted values increases with the distance between the time of the last observation and the timepoint of the predicted value, accuracy of our six-months prediction is more reliable than the 12-month prediction included in the sensitivity analysis section.

Conclusion

sARIMA(X) models fitted to time series data on prescription rates enabled forecasting future incidence rates. The models produced more accurate predictions when adding an indicator term to account for the impact of the COVID-19 pandemic between January 2020 and December 2021 (i.e., sARIMAX models). Since the selected databases are representative for the study population in the respective countries, the fitted models can be used to anticipate the incidence rates of antibiotic use in the four European countries of the study. The observation of drug usage beyond the sARIMAX-based forecast could be useful for the management of potential shortages. The forecasted values for IMASIS, IQVIA Belgium and IQVIA Germany databases need to be interpreted with caution, as the forecasted usage may be slightly overestimated.

4. LIST OF ABBREVIATIONS

Acronyms/terms	Description
ATC	Anatomical Therapeutic Chemical code
CDM	Common Data Model
CPRD	Clinical Practice Research Datalink GOLD
DA	Disease Analyzer
DARWIN EU®	Data Analysis and Real World Interrogation Network
DUS	Drug Utilization Study
EEA	European Economic Area
EHR	Electronic Health Records
EMA	European Medicines Agency
EU	European Union
GP	General Practitioner
IMASIS	Institut Municipal Assistència Sanitària Information System
IR	Incidence Rates
LPD	Longitudinal Patient Data
OMOP	Observational Medical Outcomes Partnership
PHE	Public Health Emergencies
sARIMA	seasonal Autoregressive Integrated Moving Average predictive models
sARIMAX	sARIMA models incorporating the effect of the COVID-19 pandemic as an exogenous variable
SIDIAP	Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària

5. AMENDMENTS AND UPDATES

Number	Date	Section of study protocol	Amendment or update	Reason
1	30/05/2024	8.3 Study period and follow up	Exclusion of the latest 6-month of available data in IMASIS, IQVIA Belgium and IQVIA Germany	To reduce the artifacts in the IR estimates from the latest available data driven by the methodology to obtain the observation period in these databases.
2	30/05/2024	8.7 Analysis	Inclusion of sensitivity analysis extending the forecasted period up to 12 months.	To observe a further period of forecasted use after restricting the follow up in IMASIS and IQVIA databases.

6. MILESTONES

STUDY SPECIFIC DELIVERABLE	TIMELINE (planned)	TIMELINES (actual)
Draft Study Protocol	27/10/2023	20/10/2023
Final Study Protocol	04/12/2023	04/12/2023
Creation of Analytical code	31/01/2024	31/01/2024
Execution of Analytical Code on the data	After 31/02/2024 (TBC)	31/02/2024
Interim Study Report (if applicable)	n/a	n/a
Draft Study Report	After 31/01/2024 (TBC)	20/03/2024
Final Study Report	After 31/01/2024 (TBC)	02/04/2024
Draft Manuscript (if agreed on)	-	-
Final Manuscript (if agreed on)	-	-

7. RATIONALE AND BACKGROUND

The extended mandate of EMA reinforcing the role of the Agency in crisis preparedness and management of medicinal products and medical devices became applicable on 1st March 2022 (Regulation on EMA's extended mandate becomes applicable | European Medicines Agency (europa.eu)).

EMA is responsible for monitoring medicine shortages that might lead to a crisis situation, as well as reporting shortages of critical medicines during public health emergencies (PHE). Such shortages would make it difficult or impossible to meet the treatment needs of individual patients or populations.(1) The Agency has also the mandate to coordinate responses of EU / EEA countries to shortages of critical medical devices and in-vitro diagnostics in crisis situations.

Scientific and commercial data on monthly prescriptions of medicines that may be critical in PHE can help understanding trends and seasonal variations. In conjunction with time series and forecasting models, as well as data on medicines supply, such data can contribute to the on-going efforts of the Agency to better monitor and coordinate its response to shortages of critical medicines.

This study focused on generating monthly use (prescription or dispensation) rates of selected medicines over the last 10 years of available data and then conducting a time series modelling by fitting Autoregressive Integrated Moving Average (ARIMA) models to such data to then forecast short-term (6 months) use rates.

8. RESEARCH QUESTION AND OBJECTIVES

Table 1: Primary and secondary research questions and objective

A. Primary research question and objective

Objective:	To estimate monthly incidence rates of use (prescription or dispensation) of the 11 selected medicines during the last 10 years of available data, stratified by age and sex, in each of the databases.
Hypothesis:	Not applicable
Population (mention key inclusion-exclusion criteria):	The study cohort comprises all individuals present in the database in a 10-year period counting backwards from the most recent data available. Additional sensitivity analysis was applied for the calculation of incidence rates where observation time of the respective use of the antibiotic of interest was excluded during use and 30 days afterwards.
Exposure:	Eleven antibiotics identified as potentially critical in public health emergencies (PHE)
Comparator:	None
Outcome:	None
Time (when follow up begins and ends):	Follow-up started on a pre-specified calendar time point i.e., 1st of January for each month between the 10-year study period for the calculation of monthly incidence rates. End of follow-up was defined as the earliest of loss to follow-up, end of data availability, death, or end of the 10-year study period.
Setting:	Inpatient and outpatient setting using data from the following data sources: 1) CPRD GOLD (UK, Primary Care Database) 2) IMASIS (Spain, Secondary Care Database) 3) IQVIA DA Germany (primary and outpatient care) 4) IQVIA LPD Belgium (Primary Care Database) 5) SIDIAP (Spain, Primary Care Database)
Main measure of effect:	Incidence rates of antibiotic use

CPRD = Clinical Practice Research Datalink GOLD, IMASIS = Institut Municipal Assistència Sanitària Information System, DA = Disease Analyzer, LPD = Longitudinal Patient Data, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

B. Secondary research question and objective

Objective:	To conduct a time series modelling by fitting an ARIMA model to data generated in objective 1 for short-term (6-month) forecasting
Hypothesis:	Not applicable
Population (mention key inclusion-exclusion criteria):	The study cohort comprises the monthly incidence rates of use (prescription or dispensation) of the 11 selected medicines during the last 10 years of available data, from the primary research question.
Exposure:	Eleven antibiotics identified as potentially critical in PHE
Comparator:	None
Outcome:	None
Time (when follow up begins and ends):	We used the overall (i.e., not stratified nor sensitivity analysis) monthly incidence rates from the last 10 years of available data calculated in objective 1 to forecast the following 6-months
Setting:	Inpatient and outpatient setting using data from the following data sources: 1) CPRD GOLD (UK, Primary Care Database) 2) IMASIS (Spain, Secondary Care Database) 3) IQVIA DA Germany (primary and outpatient care) 4) IQVIA LPD Belgium (Primary Care Database) 5) SIDIAP (Spain, Primary Care Database)
Main measure of effect:	Forecast of the 6-month incidence rates of antibiotic use after the end of study period

PHE = public health emergencies, CPRD = Clinical Practice Research Datalink GOLD, IMASIS = Institut Municipal Assistència Sanitària Information System, DA = Disease Analyzer, LPD = Longitudinal Patient Data, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

9. RESEARCH METHODS

9.1 Study Type and Study Design

Retrospective cohort studies were conducted using routinely-collected health data from 5 databases. The study comprised two consecutive parts:

1. A population-based cohort study was conducted to address objective 1, estimating the incidence rates of use (prescription or dispensation) of the 11 selected antibiotics.
2. A time series modelling study based on the Population-level drug utilisation study on antibiotics to address objective 2, fitting the time series data generated in objective 1 into an ARIMA model

to then forecast use rates for the subsequent 6 months.

9.2 Study Setting and Data Sources

9.2.1 Data sources selected

Inpatient and outpatient setting using data from the following data sources:

1. Clinical Practice Research Datalink GOLD [CPRD] (UK, Primary Care Database)
2. Institut Municipal Assistència Sanitària Information System [IMASIS, Parc Salut Mar Barcelona - Hospital del Mar] (Spain, Secondary Care Database)
3. IQVIA Disease Analyzer Germany [IQVIA Germany]
4. IQVIA Longitudinal Patient Data Belgium [IQVIA Belgium] (Primary Care Database)
5. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària [SIDIAP] (Spain, Primary Care Database)

Information on data source(s) used is described in a **Table 2**.

9.2.2 Rationale for the data sources selected

This study has been conducted using 5 of the 10 databases onboarded for DARWIN EU® in 2022. The selection of databases for this study was performed based on data quality, particularly on the reliability of the data and the relevance for the proposed research question. All selected databases fulfil the criteria required for a population-level drug utilisation study, as they systematically and adequately capture all drug use.

Three databases include records from primary care (CPRD, IQVIA Belgium, SIDIAP), one from hospital (IMASIS), and one includes records from primary and secondary care (IQVIA Germany). Since certain medications observed in primary care might not be observed at secondary care, an vice versa, given the nature of their own indication, we included data sources from both health care settings, and the administration path (primary and/or secondary care) has been considered when interpreting the results. We selected these four outpatient databases (CPRD GOLD, IQVIA Belgium, SIDIAP and IQVIA Germany, covering primary care and outpatient specialist care) as many antibiotics, particularly beta-lactam antibiotics/combinations and macrolides, were expected to be prescribed in that ambulatory health care setting. In addition, IMASIS includes in- and outpatient records from the hospital, where antibiotics used for the treatment of more severe conditions, including the third-generation cephalosporins and carbapenems, which was expected to be administered during hospitalisation. This data source also captures data on antibiotics that were expected to be initiated and prescribed for outpatient use following hospital discharge. With that, the diversity of healthcare settings enables to capture all the antibiotics of interest with a relevant number of prescriptions/dispensations that ranged from <1000 to 25 million across data sources. While two of the databases included in the study are from the same country (SIDIAP and IMASIS), they cover different healthcare settings. This enabled to capture differences in medicines usage in hospitals compared to primary care within the same health care region.

Previous checks evaluating the reliability of the data sources performed at the time of their onboarding showed that drug records represented between the 25% and 70% of the total number of records in the

selected databases. Data on use of antibiotics (prescription or dispensation) in 4 of the 5 databases (CPRD, SIDIAP, IMASIS and IQVIA Germany) were used in a previous DARWIN EU study which focussed on the use of antibiotics of the WHO Watch list (EUPAS103381).

The included data sources capture the relevant data necessary to answer the research questions of this study, they have recent data covering the 10-year study period, and they cover different healthcare settings where the drugs of interest may be used (i.e., primary and secondary care). Moreover, previous studies confirm its capacity to capture the use of the antibiotics of interest.

Information on data source(s) used is described below:

1. Clinical Practice Research Datalink GOLD [CPRD] (UK, Primary Care Database)

The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service, jointly funded by the National Institute for Health and Care Research and the Medicines and Healthcare products Regulatory Agency, a part of the Department of Health, United Kingdom (UK) (https://cprd.com). CPRD (2) comprises computerized records of all clinical and referral events in primary care in addition to comprehensive demographic information and medication prescription data in a sample of UK patients (predominantly from Scotland (52% of practices) and Wales (28% of practices). The prescription records include information on the type of product, date of prescription, strength, dosage, quantity, and route of administration. Data from contributing practices are collected and processed into research databases. Quality checks on patient and practice level are applied during the initial processing. Data are available for 20 million patients, including 3.2 million currently registered patients. Access to CPRD data requires approval via the Research Data Governance Process.

2. Institut Municipal Assistència Sanitària Information System [IMASIS, Parc Salut Mar Barcelona - Hospital del Mar] (Spain, Secondary Care Database)

The Institut Municipal Assistència Sanitària Information System (IMASIS) is the Electronic Health Record (EHR) system of Parc de Salut Mar Barcelona (PSMar) which is a complete healthcare services organisation. Currently, this information system includes and shares the clinical information of two general hospitals (Hospital del Mar and Hospital de l'Esperança), one mental health care centre (Centre Dr. Emili Mira) and one social-healthcare centre (Centre Fòrum) including emergency room settings, which are offering specific and different services in the Barcelona city area (Spain). At present, IMASIS includes clinical information more than 1 million patients with at least one diagnosis and who have used the services of this healthcare system since 1990 and from different settings such as admissions, outpatients, emergency room and major ambulatory surgery. The diagnoses are coded using The International Classification of Diseases ICD-9-CM and ICD-10-CM. The average follow-up period per patient in years is 6.37 (SD±6.82). IMASIS-2 is the anonymized relational database of IMASIS which is used for mapping to OMOP including additional sources of information such as the Tumours Registry.(3)

3. IQVIA Disease Analyzer Germany [IQVIA Germany] (Primary and Secondary Care Database)

IQVIA Germany is collected from extracts of patient management software used by general practitioners (GPs) and specialists practicing in ambulatory care settings.(4) Data coverage includes more than 34M distinct person records out of at total population of 80M (42.5%) in the country and collected from 2,734 providers. Patient visiting more than one provider are not cross identified for data protection reasons and therefore recorded as separate in the system. Dates of service include from 1992 through present. Observation time is defined by the first and last consultation dates. Germany has no mandatory GP system

and patient have free choice of specialist. As a result, data are collected from visits to 28.8% General, 13.4% Orthopaedic Surgery, 11.8% Otolaryngology, 11.2% Dermatology, 7.7% Obstetrics/Gynaecology, 6.2% various Neurology and Psychiatry 7.0% Paediatric, 4.6% Urology, 3.7% Cardiology, 3.5% Gastroenterology, 1.5% Pulmonary and 0.7% Rheumatology practices.

4. IQVIA Longitudinal Patient Data Belgium [IQVIA Belgium] (Primary Care Database)

IQVIA Belgium is a computerised network of GPs who contribute to a centralised database of anonymised data of patients with ambulatory visits. Currently, around 300 GPs from 234 practices are contributing to the database covering 1.1M patients from a total of 11.5M Belgians (10.0%). The database covers time from 2005 through the present. Observation time is defined by the first and last consultation dates. Drug information is derived from GP prescriptions. Drugs obtained over the counter by the patient outside the prescription system are not reported. No explicit registration or approval is necessary for drug utilisation studies.

5. Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària [SIDIAP] (Spain, Primary Care Database)

SIDIAP is collected from EHR records of patients receiving primary care delivered through Primary Care Teams (PCT), consisting of GPs, nurses and non-clinical staff.(5, 6) The Catalan Health Institute manages 286 out of 370 such PCT with a coverage of 5.6M patients, out of 7.8M people in the Catalan population (74%). The database started to collect data in 2006. The mean follow-up is 10 years. The observation period for a patient can be the start of the database (2006), or when a person is assigned to a Catalan Health Institute primary care centre. Date of exit can be when a person is transferred-out to a primary care centre that does not pertain to the Catalan Health Institute, or date of death, or date of end of follow-up in the database. Drug information is available from prescriptions and from dispensing records in pharmacies. Drugs not prescribed in the GP setting might be underreported; and disease diagnoses made at specialist care settings are not included. Studies using SIDIAP data require previous approval by both a Scientific and an Ethics Committee.

9.3 Study Period

The study comprised the last 10 years of available data counting backwards from the respective latest date of data lock of the respective databases (**Table 2**).

Table 2. Study period used in each of the selected Data Sources.

Country	Name of Database	Health Care setting	Type of Data	Number of active subjects	Data lock for the last update
UK	CPRD	Primary Care	EHR	3 million	01-07-2023
Spain	IMASIS	Hospital	EHR	0.6 million	09-09-2023
Germany	IQVIA Germany	Primary & Secondary Care	EHR	41.9 million	30-06-2023
Belgium	IQVIA Belgium	Primary Care	EHR	8.5 million	30-09-2023
Spain	SIDIAP	Primary Care	EHR	5.8 million	30-09-2023

UK = United Kingdom, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària, IMASIS= Institut Municipal Assistència Sanitària Information System, CPRD = Clinical Practice Research Datalink GOLD, EHR = Electronic Heath records.

9.4 Follow-up

To estimate the monthly incidence rates of each medicine of interest, we required the eligible population and their contributed observation time to first be identified separately for each medicine. Thus, the following methodology was conducted for each of the antibiotics of the study in parallel:

Individuals included in the study started contributing person-time on the respective date of the latest of the following: 1) study start date/start of each month for each calendar year, 2) date at which the observation period starts. Individuals stop contributing person-time at the earliest date of the following: 1) end of each month for each calendar year (or end of available data in each of the data sources when earlier than end of the month), 2) date at which the observation period of the specific person ends.

An example of entry and exit into the denominator population is shown in **Figure 1**. In this example, person ID 1 and 3 were included as denominators at the study start date, or at the start of each month for each calendar year, as both were being observed in the data base from a prior date. Person ID 2 entered the study at the date of starting their period of observation, which was later than the study start date/start of the month. Person ID 1 and 2 were followed until the study end date (end of available data in each of the data sources/end of each month for each calendar year) whilst Person ID 3 left when exiting the database (the end of observation period). Lastly, person ID 4 had two observation periods in the database. The first period contributed time from study start/start of the month until end of observation period, the second started contributing time again on the date of their second observation period start and exit at study end date/end of each month for each calendar year.

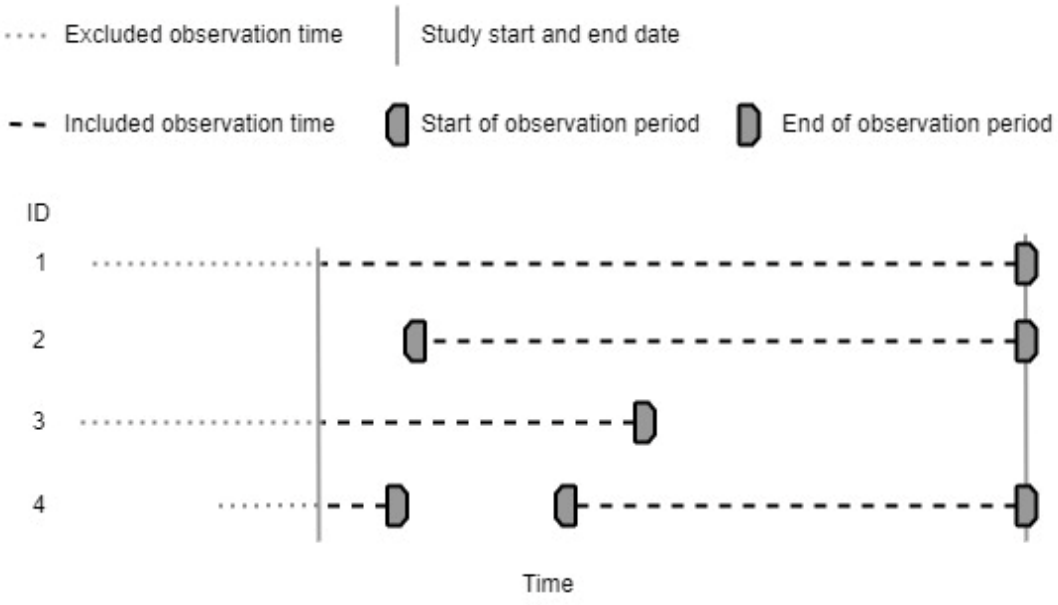


Figure 1: Included observation time for the denominator population

CPRD and SIDIAP end of observation period are based on individuals' registration and/or enrolment in their health care system. This information is administrative and independent of the use of the health care system. Individuals are followed until they deregister, either by moving out of the catchment area or die. Thus, patients who do not interact with the health system frequently (e.g., healthier patients) are followed up until the end of data availability unless they deregister.

IMASIS, IQVIA Belgium and IQVIA Germany databases do not have registration or enrolment timelines. Thus, these three databases calculate the end of observation period based on the last interaction of the individual with their health care system: last visit or last observation record date. However, this methodology has the limitation that the follow up in patients who do not interact with the health system frequently stop earlier, decreasing the number of individuals active at the end of data availability.

9.5 Study Population with inclusion and exclusion criteria

The study cohort comprised all individuals present in the databases during the study period (the last 10 years of available data counting backwards from the respective latest date of data lock of the respective databases).

Additional eligibility criteria for a sensitivity analysis of Incidence rates were applied for the calculation of incidence rates where the observation time of users of the antibiotic of interest was excluded during use and 30 days afterwards. The rates estimated in sensitivity analysis were not analysed with sARIMA(X) models.

9.6 Variables

9.6.1 Exposure/s (where relevant)

For this study, the exposure of interest was use (during study period) of antibiotics identified as potentially critical in PHE. This list of antibiotics (with respective ATC code) is described in **Table 3**.

Table 3. Exposure of interest

Antibiotic	Class	ATC code
Azithromycin	Macrolides	J01FA10
Clarithromycin	Macrolides	J01FA09
Phenoxymethylpenicillin	Beta-lactamase sensitive penicillins	J01CE02
Benzylpenicillin	Beta-lactamase sensitive penicillins	J01CE01
Amoxicillin	Penicillins with extended spectrum	J01CA04
Amoxicillin + Clavulanic acid	Combinations of penicillins	J01CR02
Ceftriaxone	Third-generation cephalosporins	J01DD04
Cefotaxime	Third-generation cephalosporins	J01DD01

Antibiotic	Class	ATC code
Meropenem	Carbapenems	J01DH02
Cefuroxime	Second-generation cephalosporins	J01DC02
Piperacillin + Tazobactam	Combinations of penicillins	J01CR05

ATC = Anatomical Therapeutic Chemical code

9.6.2 Outcome/s (where relevant)

NA

9.6.3 Other covariates, including confounders, effect modifiers and other variables (where relevant)

9.6.3.1 Covariates for stratification in population-level drug utilisation study:

- Age: age was stratified in three groups (<18, 18-64, 65+ years).
- Time period: Monthly
- Sex (Male, Female)

9.7 Study size

No sample size was calculated for this study. The expected number of prescriptions/dispensations across data sources was expected to be roughly between <1000 and 25 million.

9.8 Data transformation

NA

9.9 Statistical Methods

This section describes the details of the analysis approach and rationale for the choice of analysis, with reference to the DARWIN EU® Catalogue of Data Analyses, available at https://www.darwin-eu.org/methods/standardised-analytics, which describes the type of analysis in function of the study type.

The analysis includes calculation of population-based incidence rates as described in section *9.9.3.2 – Population-level drug utilisation study* on antibiotics; and the fitting of the sARIMA models into the data generated in objective 1 is described in section *9.9.3.3 – Time series modelling*.

9.9.1 Federated Network Analyses

Analyses have been conducted separately for each database. Before study initiation, test runs of the analytics are performed on a subset of the data sources or on a simulated set of patients and quality control checks are performed. Once all the tests were passed, the final package was released in the version-controlled Study Repository for execution against all the participating data sources.

The data partners locally executed the analytics against the OMOP-CDM in R Studio and reviewed and approved the by default aggregated results before returning them to the DARWIN EU Coordination Centre.

Where necessary, multiple execution iterations were performed, with additional fine tuning of the code base. A service desk was available during the study execution for support.

The study results of all data sources were checked after which they were made available to the DARWIN EU Coordination Centre team and the write-up of the study report started. All results were locked and timestamped for reproducibility and transparency.

9.9.2 Main Summary Measures

Monthly incidence rates of prescription of selected antibiotics. To prevent confidentiality issues, cell counts lower than 5 were reported as "<5".

9.9.3 Main Statistical Methods

9.9.3.1 Methods to derive parameters of interest in Population-level drug utilisation study

Calendar time

Calendar time were based on monthly periods (e.g., 01/01/2013 – 31/01/2013) of the index prescription/dispensation in the population level analysis, based on the calendar periods covered by each data source as presented in the section *2. DATA SOURCES*.

Age

Age at index date were calculated using January 1st of the year of birth as proxy for the actual birthday. The following age groups were used for stratification: <18, 18-64, 65+ years.

Sex

Results are presented stratified by sex: Male, Female.

9.9.3.2 Population-level drug utilisation study

Incidence calculations were conducted separately for each antibiotic of interest.

Incidence calculations

Users (for incidence rates estimation) were selected based on their use of the respective drug of interest after the start of the study and/or in each month for each calendar year..

Monthly incidence rates (IR) of the antibiotics of interest were estimated as the of number of users per 100,000 person-years of the population at risk of getting exposed during each month for each calendar year. Individuals included in the study contributed time at risk up to their first use (prescription or dispensation) during the study month. Or if they do not have a drug exposure, they have contributed time at risk until the end of the month (or data availability if earlier), as described above in Section 9.4 (Follow-up). Incidence rates are reported with 95% Poisson exact confidence intervals.

An illustration of the calculation of incidence of antibiotic use is shown below in **Figure 2**.

Patient ID 1, 2 and 3 contributed time at risk between the month start and end dates except when they became incident users of antibiotics. For Patient ID 2 and 5, use of antibiotics before month start were not count as incident use. Patient ID 4 and 5 were not seen to use antibiotics between the month start and end date, and so they contributed time at risk but had no incident outcomes.

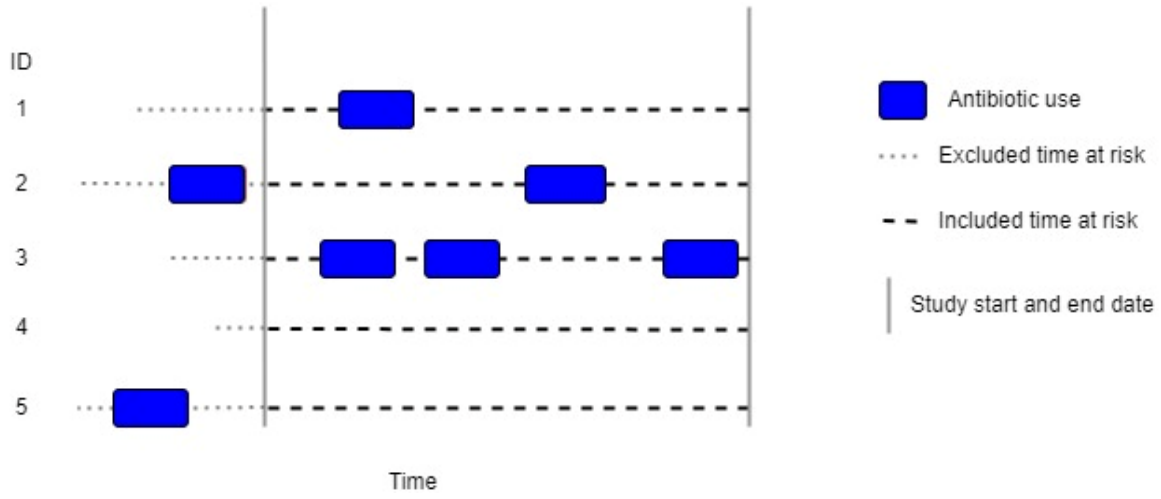


Figure 2: Incidence example for antibiotic use

9.9.3.3 Time series modelling

We conducted a time series modelling where AutoRegressive Integrated Moving Average (ARIMA) models were used to model trends and seasonality over time for the monthly incidence rate (IR) estimates (overall IR from main analysis, i.e. without age and sex stratification). We forecast the use of antibiotics in each database of the study when their respective cohorts had a minimum of 1,000 users to ensure numerical stability of the estimated IRs.

Period forecasted in each database of the study

As detailed in *9.4 Follow-up* section, IMASIS, IQVIA Belgium and IQVIA Germany the end of the observation period of an individual is based on the date of the last visit/last observation with the health care system, reducing the follow-up of infrequent users. This translates into an artefactual decrease in the denominator of these three databases: users of antibiotics are recorded as they interacted with their health care system, but person-time in the denominator appears reduced due to the shorter follow-up of those individuals who did not interact with the system towards the end of data availability. This artifact became more relevant when reaching the last months before date lock of the database (i.e., end of data availability), and therefore the corresponding incidence rates were not reliable. Thus, the IR estimates corresponding to the latest 6 months available of these three databases were excluded from the sARIMA(X) modelling.

Table 4 reports the last date from the Population-level drug utilisation data that was used to fit the sARIMA(X) models in each database of the study.

Table 4. Date of the last month of observed use of antibiotics (i.e., last reported incidence rate estimate) used to fit the sARIMA(X) models, and 6-month period where use of antibiotics was forecasted, for each database of the study.

Data base	Last month of observed data used in ARIMA(X)	6-month period forecasted: From	Until
IQVIA Germany	01-12-2022	01-01-2023	01-06-2023
IMASIS	01-02-2023	01-03-2023	01-08-2023
IQVIA Belgium	01-03-2023	01-04-2023	01-09-2023
CPRD	01-05-2023	01-06-2023	01-11-2023
SIDIAP	01-06-2023	01-07-2023	01-12-2023

CPRD = Clinical Practice Research Datalink GOLD, IMASIS = Institut Municipal Assistència Sanitària Information System, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

ARIMA models

The two types of ARIMA models used were:

1. Seasonal AutoRegressive Integrated Moving Average (sARIMA):
   These models were characterized by varying combinations of seasonal autoregressive (AR) and moving average (MA) parameters, possibly on the integrated time series.
2. Seasonal AutoRegressive Integrated Moving Average with an eXogenous variable (sARIMAX):
   These models were characterized by varying combinations of seasonal autoregressive (AR) and moving average (MA) parameters, possibly on the integrated time series. Additionally, a model term was incorporated to account for the effect of the COVID-19 pandemic from January 2020 to December 2021, transforming the models into sARIMAX models to better capture this external influence.

The steps to select and evaluate the sARIMA and the sARIMAX models (abbreviated as sARIMA(X) when referring to both) were the same. The model selection and evaluation process for the sARIMA(X) are outlined in the following sub-sections (**Figure 3**):

1. Train: First, sARIMA(X) models were fitted using the auto.arima() function from the forecast R package, which considers stationary, seasonality and trend testing. When computing the optimal values of the p, q, d, P, Q parameters determining the order of differences and of the (seasonal) AR and MA components of each sARIMA(X) model using the auto.arima() function, we established the maximum value of the p+q+P+Q sum of 5 to diminish the risk of overfitting the data and to ensure the statistical parsimony of the model. A range of different combinations of values of AR and MA parameters was assessed, and the best model was chosen based on the lowest Akaike's Information Criterion (AIC). The models were estimated using the period from the start of the study follow-up until the start of the "tune period" (see below) in each database.
2. Tune: The model's predictive performance was evaluated using out-of-sample validation on the tune period, which comprised the latest 6 months of available and reliable incidence rate (IR per 100,000 person-years, with 95%CI) data. This enabled comparison of model predictions with the actually observed values for the last 6 months. In this phase, we explored different max.order settings in the

training process to tune the model parameters, with the goals of enhancing forecasting capability and preventing overfitting. Each parameter configuration was assessed based on its Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) on the validation set, enabling us to decide the max.order that minimized these metrics and ensured accurate forecasts. Diagnostic assessments, including plots of residuals and Q-Q plots for normality were conducted to ensure the model's fitness.

3. Forecast: The selected sARIMA(X) model(s) were employed to forecast the incidence rates of use for the next six months after end of data availability. This step aims to provide a reliable prediction of future incidence rates based on historical data, including potential external impact of the COVID-19 pandemic.

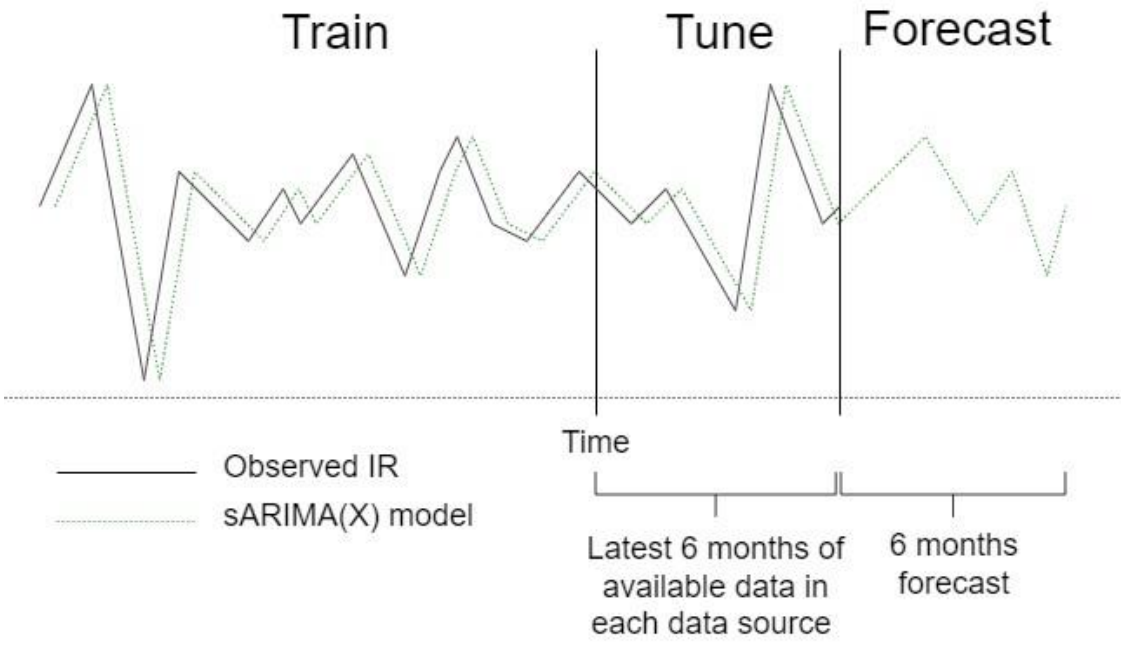


Figure 3: Graphical representation of the fitting of sARIMA(X) models

9.9.4 Missing Values

Individuals who stop their follow-up or observation period before the end of the study period had missing follow-up data. For these individuals, the time at risk of experiencing the event of interest (use of one of the medicines listed in section 9.6.1) was censored at the time of loss-to-follow-up/end of observation period as defined in section 9.4. Estimated incidence rates assume censoring occurred at random.

9.9.5 Sensitivity Analysis

9.9.5.1 Population-level drug utilisation

A sensitivity analysis for new users has been run, where users are required to not have been exposed to the drug of interest for at least 30 days prior the current use. If the start date of a prescription/dispensation

does not fulfil the exposure washout criteria of 30 days of no use, the whole exposure was eliminated. The estimates obtained in the sensitivity analysis have not been analysed with sARIMA(X) models.

Sensitivity analysis for Incidence calculations:

The denominator population for the sensitivity analysis restricting to new users of the antibiotic of interest excluded any study subjects with use of the medication of interest 30 days prior to the date at which they would have otherwise satisfied the criteria to enter the denominator population (as described above). An illustration of how the denominator population was calculated is shown in **Figure 4.**

Patient ID 1 and 3 contribute time at risk up to the point at which they become incident users of antibiotics. Patient ID 4 and 5 are not seen to use antibiotics and so contribute time at risk but no incident outcomes. Meanwhile, patient ID 2 first contributes time at risk starting at the day when the washout period of a previous exposure, before study start, has ended, and ending when the next exposure of antibiotic was starting. A second period of time at risk again starts after the washout period. For person ID 3, only the first and third exposures of antibiotics count as incident use, while the second exposure starts within the washout period of the first exposure. The time between start of the first exposure until the washout period after the second exposure was not considered as time at risk.

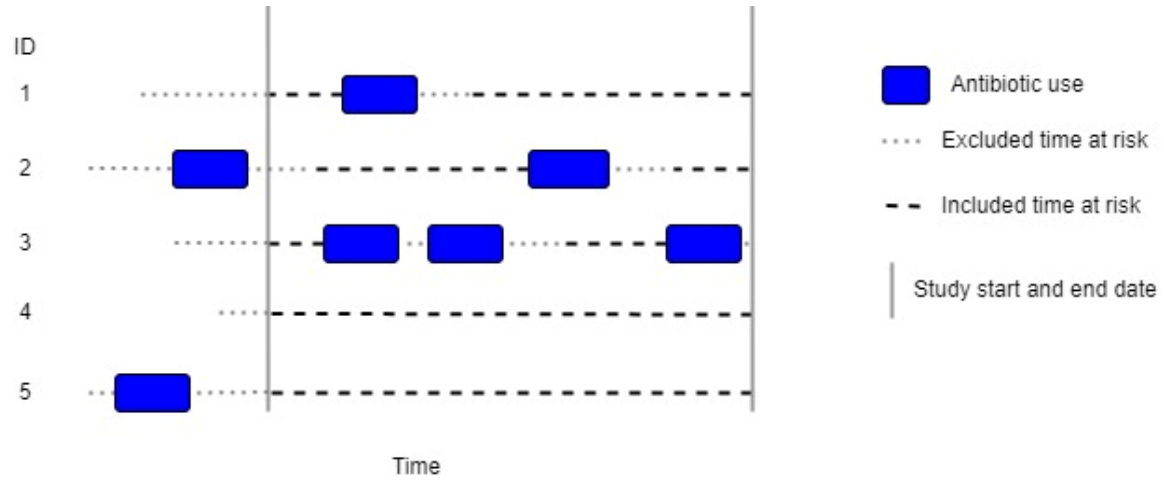


Figure 4: Incidence example for antibiotic use in the sensitivity analysis

9.9.5.2 Time series modelling

We conducted a 12-month forecast of the antibiotic use, where the initial forecast period (**Figure 5**) of 6-month was extended by an additional six months, totalling twelve months.

Table 5 reports the last date from the Population-level drug utilisation data that was used to fit the sARIMA(X) models in each database of the study, and the 12-month period forecasted.

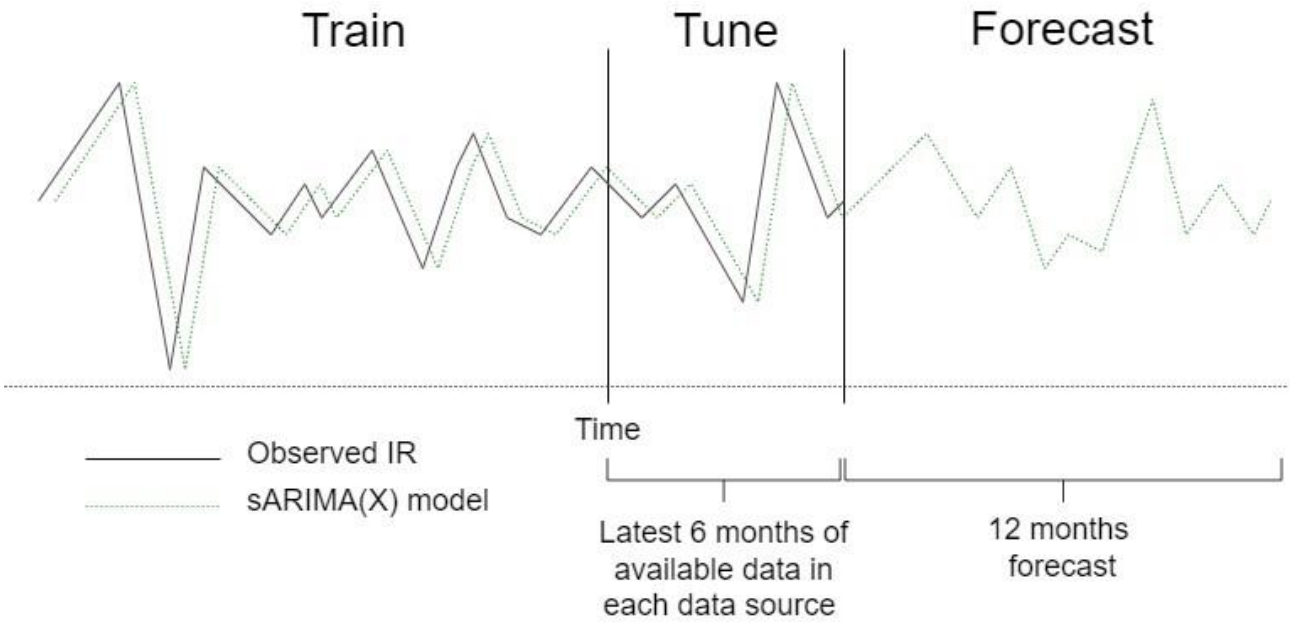


Figure 5: Graphical representation of the fitting of sARIMA(X) models forecasting 12-month of the antibiotic use

Table 5. Date of the last month of observed use of antibiotics (i.e., last reported incidence rate estimate) used to fit the sARIMA(X) models, and 12-month period where use of antibiotics was forecasted, for each database of the study.

Data base	Last month of observed data used in ARIMA(X)	12-month period forecasted: From	Until
IQVIA Germany	01-12-2022	01-01-2023	01-12-2023
IMASIS	01-02-2023	01-03-2023	01-02-2024
IQVIA Belgium	01-03-2023	01-04-2023	01-03-2024
CPRD	01-05-2023	01-06-2023	01-05-2024
SIDIAP	01-06-2023	01-07-2023	01-06-2024

CPRD = Clinical Practice Research Datalink GOLD, IMASIS= Institut Municipal Assistència Sanitària Information System, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

9.9.6 R-packages

We used the R package "IncidencePrevalence" (https://github.com/darwin-eu/IncidencePrevalence). for the population-level estimation of drug utilisation. Additionally, we used the R package "forecast" (https://cran.r-project.org/package=forecast) for the sARIMA(X) modelling.

10 DATA MANAGEMENT

All databases have been mapped to the OMOP common data model. This enabled the use of standardised analytics and tools across the network since the structure of the data and the terminology system has been harmonised. The OMOP CDM is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org. This analytic code for this study was written in R. Each data partner executed the study code against their database containing patient-level data and then returned the results set which only contained aggregated data. The results from each of the contributing data sites have been combined in tables and figures for the study report.

11 QUALITY CONTROL

General database quality control

A number of open-source quality control mechanisms for the OMOP CDM were developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, it was expected that data partners have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality was solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

When defining drug cohorts, non-systemic products were excluded from the list of included codes summarised on the ingredient level. A pharmacist reviewed the codes of the antibiotics of interest.
When defining cohorts for indications, a systematic search of possible codes for inclusion was conducted using CodelistGenerator R package (https://github.com/darwin-eu/CodelistGenerator). This software allows the user to define a search strategy and using this could then query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the DrugExposureDiagnostics R package (https://github.com/darwin-eu/DrugExposureDiagnostics) was run to assess the use of different codes across the databases contributing to the study and identify any codes potentially omitted in error.
The study code was based on two R packages currently being developed to (1) estimate Incidence and Prevalence and (2) characterise drug utilisation using the OMOP common data model. These packages include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing.

The R package containing the study code is available via GitHub (https://github.com/darwin-eu-studies/P2-C1-012-DrugShortages).

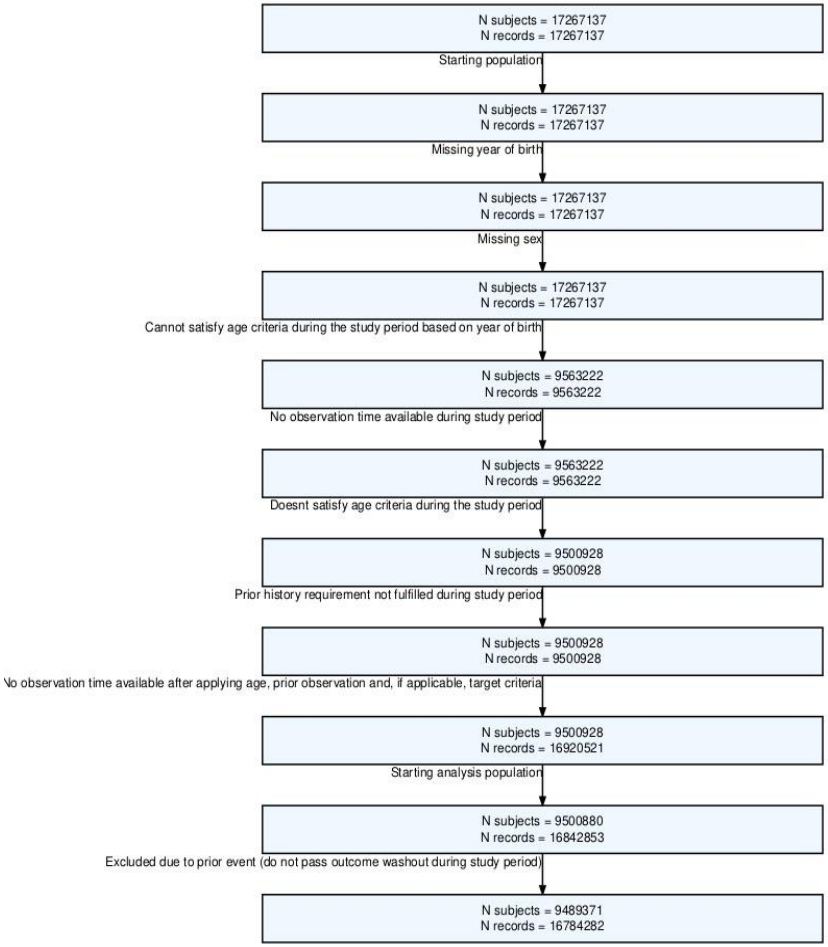
12 RESULTS

All results are available in a web-application ("shiny app") at https://data-dev.darwin-eu.org/connect/#/apps/71c3f554-ef71-4ccd-b135-4a5d34809a63/access.

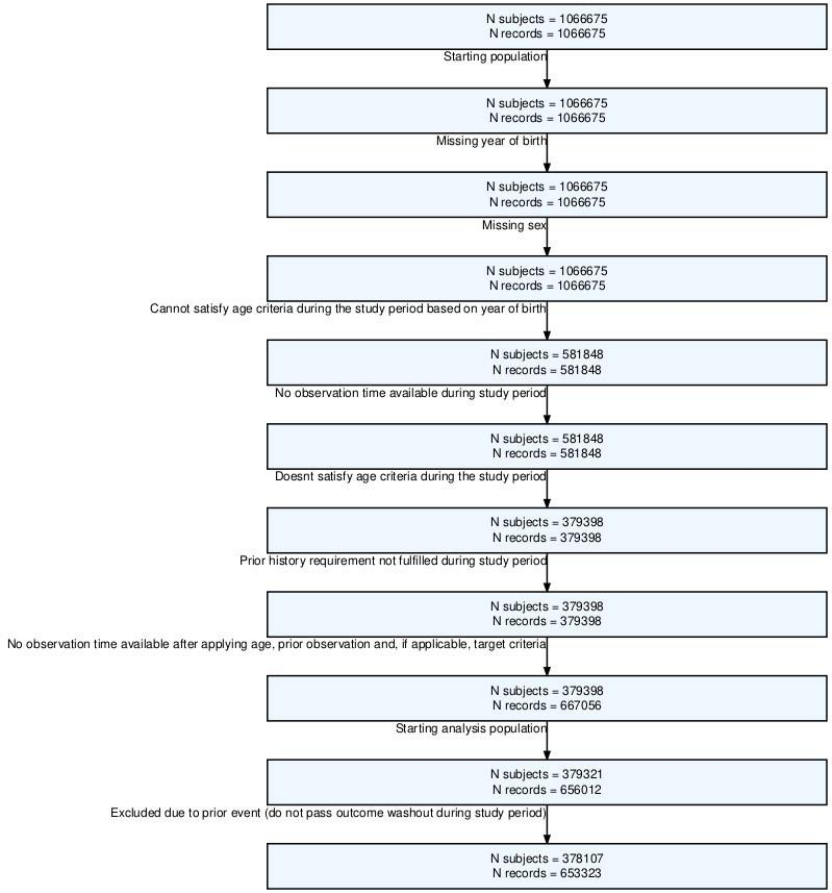
12.1 Individuals

A representation of the attrition process to obtain the study cohorts for studying the usage of antibiotics of interest, using as example the overall population at risk of using amoxicillin is provided in **Figure 6**. The attrition of all the study cohorts is provided in the shiny app.

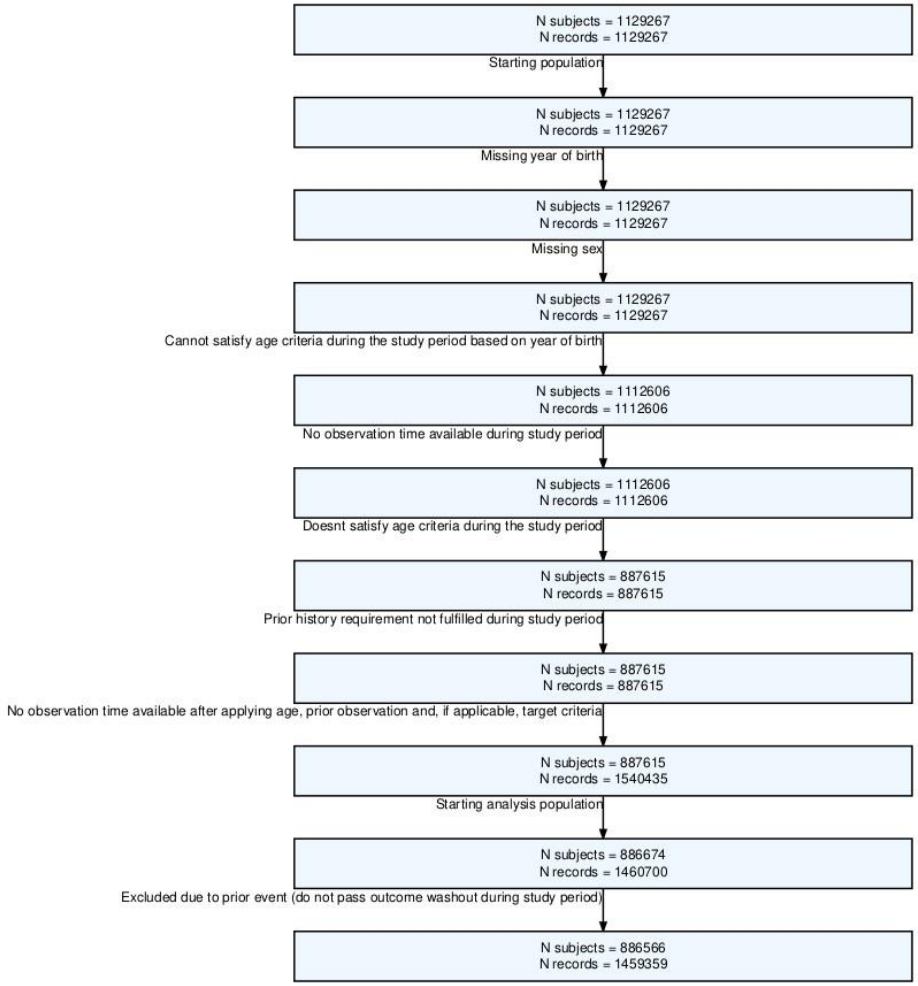
A) CPRD amoxicillin attrition



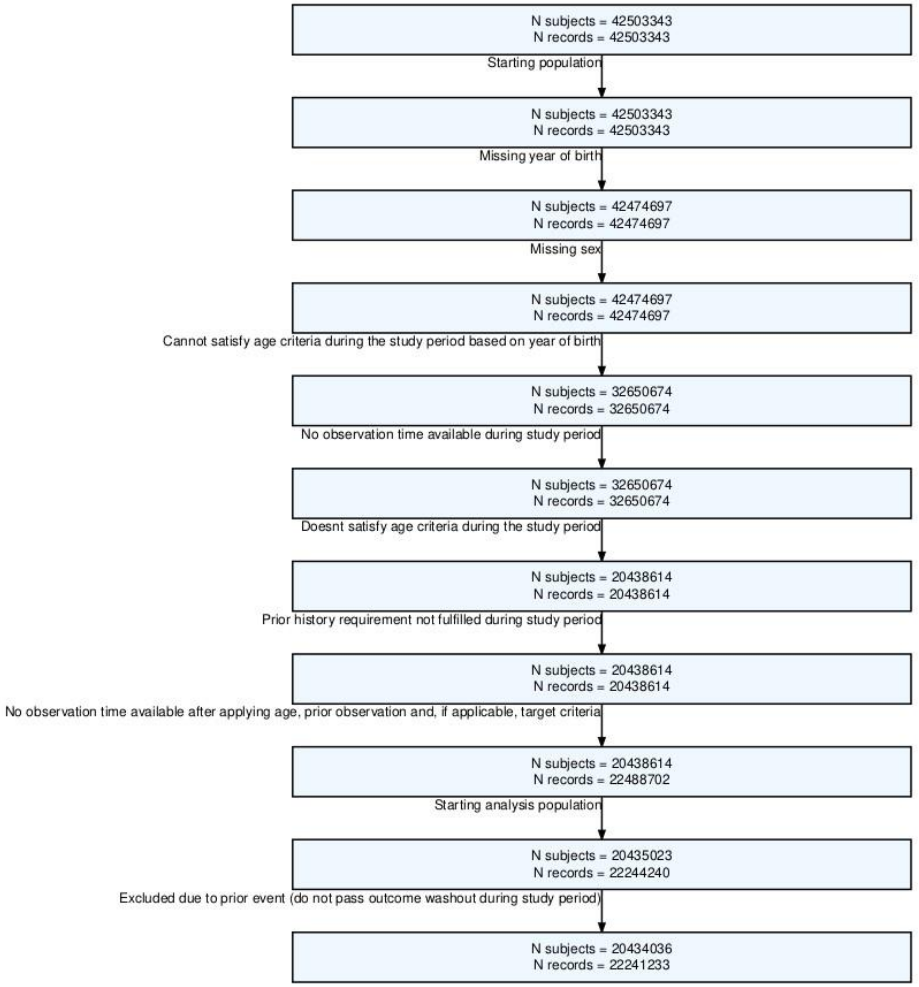
B) IMASIS amoxicillin attrition:



C) IQVIA Belgium amoxicillin attrition:



D) IQVIA Germany amoxicillin attrition:



E) SIDIAP amoxicillin attrition:

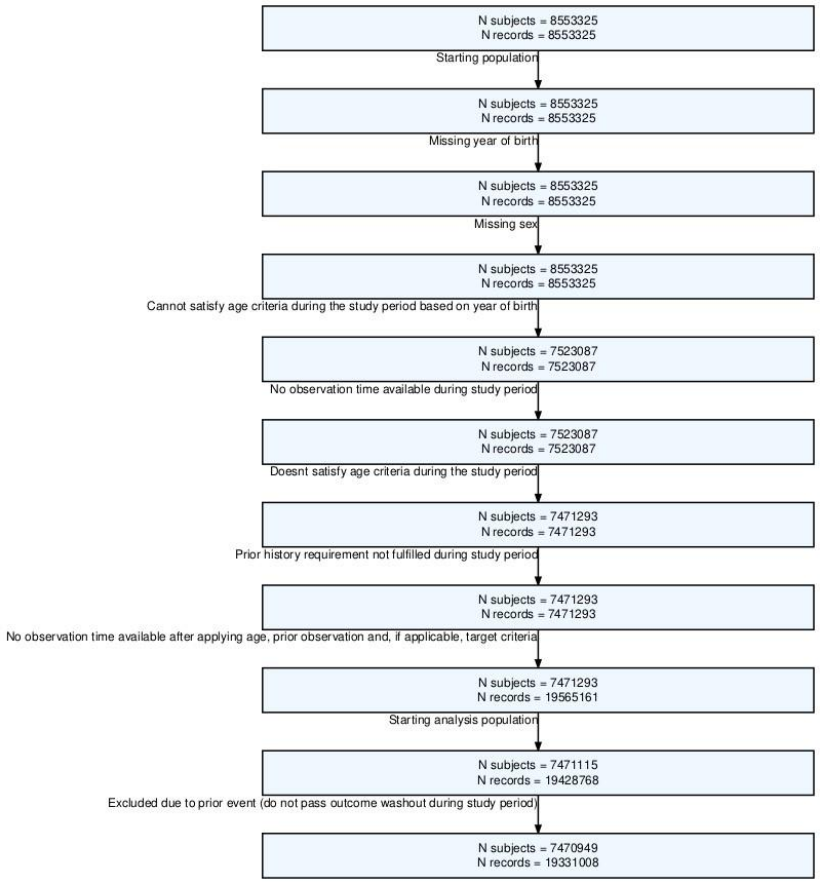


Figure 6: Attrition for the amoxicillin study cohorts in the overall population in A) CPRD, B) IMASIS, C) IQVIA Belgium, D) IQVIA Germany and E) SIDIAP.

Table 6 describes the number of users and the number of records of the antibiotics of interest included for each database during the overall study period.

Amoxicillin cohorts were the largest in all databases: CPRD n=276,925, IMASIS n=67,723, IQVIA Belgium n=309,585, IQVIA Germany n=1,405,702, SIDIAP n=4,036,463. Tazobactam/Piperacillin, cefotaxime and meropenem cohorts, which are antibiotics typically administered in hospitals, presented low user counts (n ≤ 750 users) except in IMASIS (n=11214, n=7600 and n=5713, respectively).

Table 6. Number of individuals and records per antibiotic of interest in each source population during the overall study period.

Data source name	Cohort Name	Number Records	Number Subjects
CPRD_GOLD	amoxicillin	7040444	2769255
CPRD_GOLD	amoxicillin_clavulanate	966186	582603
CPRD_GOLD	azithromycin	374330	78147
CPRD_GOLD	cefotaxime	143	133
CPRD_GOLD	ceftriaxone	1813	1375
CPRD_GOLD	cefuroxime	5170	2840
CPRD_GOLD	clarithromycin	1509607	800012
CPRD_GOLD	meropenem	301	95
CPRD_GOLD	benzylpenicillin	668	549
CPRD_GOLD	phenoxymethylpenicillin	1586526	856830
CPRD_GOLD	tazobactam_piperacillin	256	173
IMASIS	amoxicillin	295553	67723
IMASIS	amoxicillin_clavulanate	274864	61852
IMASIS	azithromycin	36646	16281
IMASIS	cefotaxime	42854	7600
IMASIS	ceftriaxone	84846	19049
IMASIS	cefuroxime	4403	3360
IMASIS	clarithromycin	3160	1848
IMASIS	meropenem	63053	5713
IMASIS	benzylpenicillin	4921	1753
IMASIS	phenoxymethylpenicillin	NA	NA
IMASIS	tazobactam_piperacillin	104706	11214
IQVIA_Belgium	amoxicillin	639469	309585

Data source name	Cohort Name	Number Records	Number Subjects
IQVIA_Belgium	amoxicillin_clavulanate	281466	158081
IQVIA_Belgium	azithromycin	174598	105172
IQVIA_Belgium	cefotaxime	<5	<5
IQVIA_Belgium	ceftriaxone	994	839
IQVIA_Belgium	cefuroxime	58156	39169
IQVIA_Belgium	clarithromycin	67526	49560
IQVIA_Belgium	meropenem	NA	NA
IQVIA_Belgium	benzylpenicillin	1333	380
IQVIA_Belgium	phenoxymethylpenicillin	2160	1922
IQVIA_Germany	amoxicillin	2115414	1405702
IQVIA_Germany	amoxicillin_clavulanate	554559	419966
IQVIA_Germany	azithromycin	901241	612845
IQVIA_Germany	cefotaxime	201	152
IQVIA_Germany	ceftriaxone	5231	4138
IQVIA_Germany	cefuroxime	1287060	894745
IQVIA_Germany	clarithromycin	412263	310087
IQVIA_Germany	meropenem	306	165
IQVIA_Germany	benzylpenicillin	6099	3299
IQVIA_Germany	phenoxymethylpenicillin	524409	402478
IQVIA_Germany	tazobactam_piperacillin	180	155
SIDIAP	amoxicillin	11269004	4036463
SIDIAP	amoxicillin_clavulanate	5373659	2711651
SIDIAP	azithromycin	1838095	1189116
SIDIAP	cefotaxime	983	750
SIDIAP	ceftriaxone	22292	18000
SIDIAP	cefuroxime	717642	489773
SIDIAP	clarithromycin	510332	414479
SIDIAP	meropenem	NA	NA
SIDIAP	benzylpenicillin	5532	4688
SIDIAP	phenoxymethylpenicillin	NA	NA

NA = Not available (no users or counts were found for that specific drug of interest), CPRD = Clinical Practice Research Datalink GOLD, IMASIS = Institut Municipal Assistència Sanitària Information System, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

12.2 Descriptive Data

This population-level drug utilisation study does not contain descriptive information on antibiotic users. Information on dimension of study cohorts is reported in the "Individuals" section above (12.1).

12.3 Outcome Data

Outcome data is provided in the "Main Results" section below (12.4).

12.4 Main Results

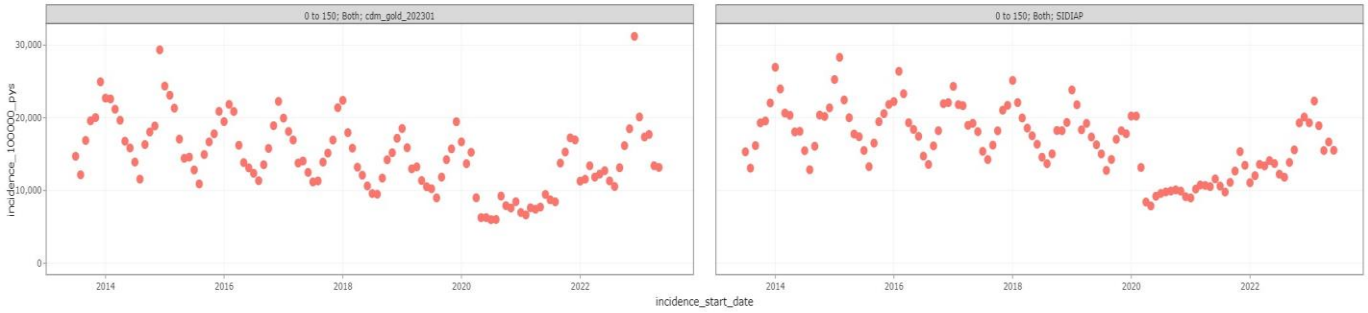
All results for each individual drug and database are available in the shiny app at https://data-dev.darwin-eu.org/connect/#/apps/71c3f554-ef71-4ccd-b135-4a5d34809a63/accesshttps://data-dev.darwin-eu.org/connect/.

12.4.1 Population-level drug utilisation

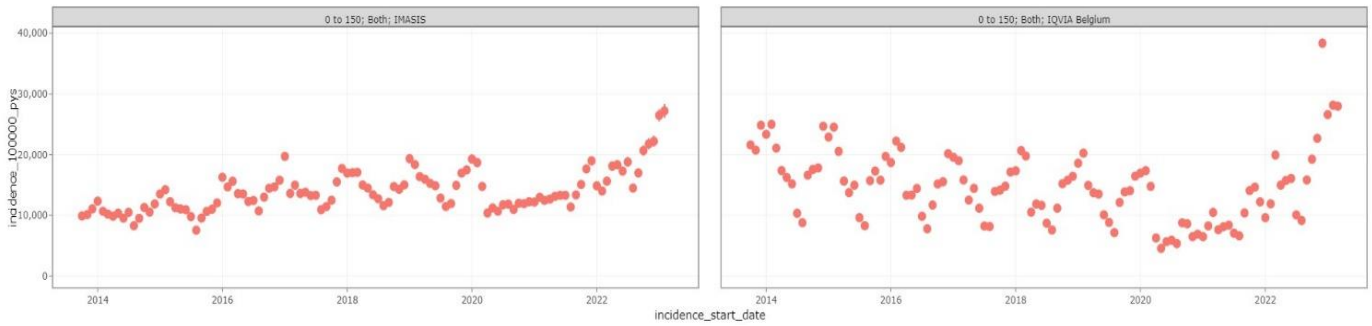
Antibiotics use with an indication for infections frequent in winter showed a seasonality pattern with annual peaks around November-March months, such as observed for amoxicillin (**Figure 7**). The results also showed a substantial drop in the use of most of the antibiotics of interest after the start of the COVID-19 pandemic (March 2020) across the 5 databases, with the exception of azithromycin and ceftriaxone in IMASIS that presented a peak during March and April 2020 (IR [95%CI] per 100,000 person-years: 4,389 [13,799-14,998] and 13,217 [12,642-13,812] for azithromycin, and 15,535 [14,924-16,164] and 18,741 [18,058-19,444] for ceftriaxone, respectively).

Artefactual increases in rates of antibiotic use were observed in the last 6 months of data available from IMASIS, IQVIA Belgium and IQVIA Germany, due to a known drop in the denominator as described in *9.9.3.3 Time series modelling* section. Observing the number of users of amoxicillin, azithromycin and ceftriaxone in IMASIS, IQVIA Belgium and IQVIA Germany (i.e., the numerator from the IR, which is provided in the column *n_events* within the shiny app), we can see an increment of users during the winter 2022. (**Table 7**)

A) IR amoxicillin in CPRD **B) IR amoxicillin in SIDIAP**



C) IR amoxicillin in IMASIS **D) IR amoxicillin in IQVIA Belgium**



E) IR amoxicillin in IQVIA Germany

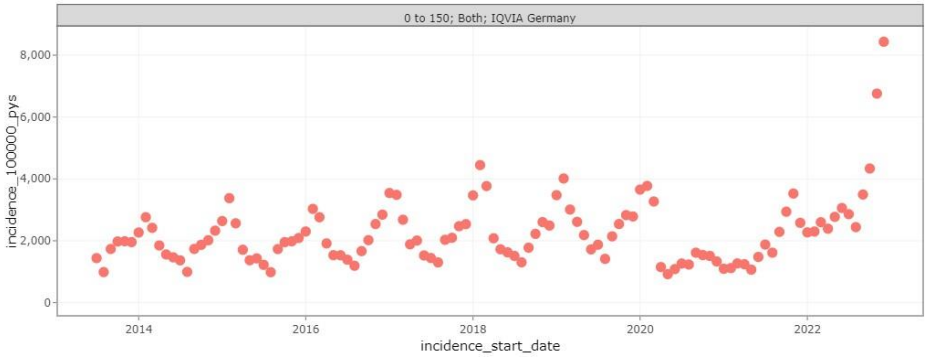


Figure 7: Incidence rates of amoxicillin use per 100,000 person-years in the overall population (i.e., any age and sex) for each of the databases of the study: A) CPRD, B) SIDIAP, C) IMASIS, D) IQVIA Belgium and E) IQVIA Germany.

Table 7. Number of users of amoxicillin, azithromycin and ceftriaxone in IMASIS, IQVIA Belgium and IQVIA Germany from September to December 2022.

Cohort Name	Date (month observed)	IMASIS (Number of users during that month)	IQVIA Belgium (Number of users during that month)	IQVIA Germany (Number of users during that month)
Amoxicillin	01/12/2022	2,278	9,126	35,865
	01/11/2022	2,256	5,514	29,499
	01/10/2022	2,301	5,115	20,484
	01/09/2022	1,892	4,355	16,656
Azithromycin	01/12/2022	311	3,022	15,717
	01/11/2022	281	1,859	11,355
	01/10/2022	261	1,854	8,608
	01/09/2022	221	1,473	6,868
Ceftriaxone	01/12/2022	927	11	48
	01/11/2022	752	14	43
	01/10/2022	792	8	48
	01/09/2022	584	5	52

Additionally, we observed that use of phenoxymethylpenicillin in IQVIA Belgium was discontinued after September 2019 (**Figure 8**).

The age-stratified analysis showed the use of antibiotics was generally higher in individuals aged 65 or older. When stratified by sex, women had a generally higher incidence of antibiotic use in primary care (CPRD, IQVIA Belgium and SIDIAP); whilst men presented generally higher use in hospital (IMASIS) and IQVIA Germany data. Full IR results including stratifications for age groups and sex, or sensitivity analyses where new users had a prior 30 days wash out of the drug of interest, are presented in the shiny app.

IR showing the trends of antibiotic use which were used for the forecasting are presented below in the *12.4.2 Time series modelling* subsection.

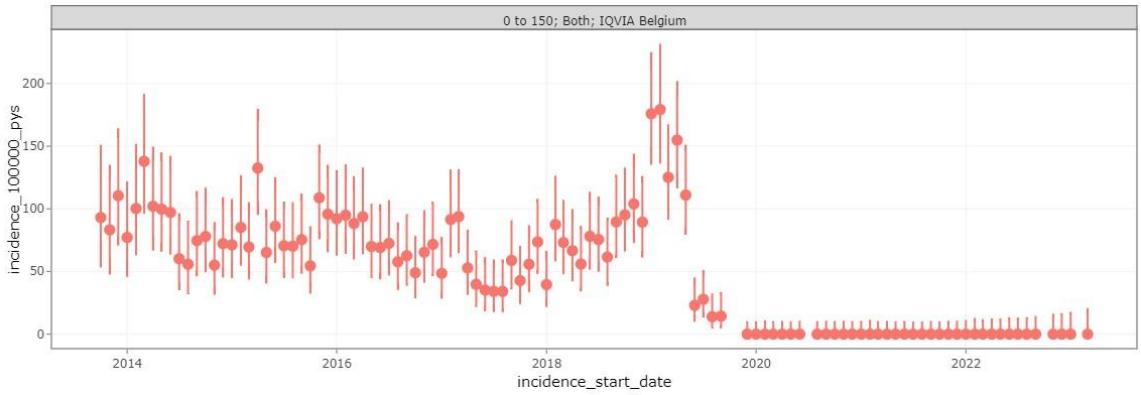


Figure 8: Incidence rates with 95% confidence interval of phenoxymethylpenicillin use per 100,000 person-years in the overall population of IQVIA Belgium

12.4.2 Time series modelling and diagnostics

The attrition numbers specific to each of the antibiotics of interest used for forecasting in each of the databases are summarised in **Table 8**.

All results of time series modelling study based on the Population-level drug utilisation, including the model diagnostics and parameter estimates, are available in the shiny app at https://data-dev.darwin-eu.org/connect/#/apps/71c3f554-ef71-4ccd-b135-4a5d34809a63/accesshttps://data-dev.darwin-eu.org/connect/.

Table 8. Number of drug users per cohort and attrition of antibiotics modelled in each database. Usage of the antibiotics of interest in each database was forecasted when study cohort had a minimum of 1,000 users.

Step	CPRD		IMASIS		IQVIA Germany		IQVIA Belgium		SIDIAP	
	Counts	Forecast based on sARIMA(X)	Counts	Forecast based on sARIMA(X)	Counts	Forecast based on sARIMA(X)	Counts	Forecast based on sARIMA(X)	Counts	Forecast based on sARIMA(X)
Amoxicillin	2769255	YES	67723	YES	309585	YES	1405702	YES	4036463	YES
Phenoxymethylpenicillin	856830	YES	NA	NO	1922	YES	402478	YES	NA	NO
Clarithromycin	800012	YES	1848	YES	49560	YES	310087	YES	414479	YES
Amoxicillin/Clavulanate	582603	YES	61852	YES	158081	YES	419966	YES	2711651	YES
Azithromycin	78147	YES	16281	YES	105172	YES	612845	YES	1189116	YES
Cefuroxime	2840	YES	3360	YES	39169	YES	894745	YES	489773	YES
Ceftriaxone	1375	YES	19049	YES	839	NO	4138	YES	18000	YES
Benzylpenicillin	549	NO	1753	YES	380	NO	3299	YES	4688	YES
Tazobactam/Piperacillin	173	NO	11214	YES	NA	NO	155	NO	NA	NO
Cefotaxime	133	NO	7600	YES	<5	NO	152	NO	750	NO
Meropenem	95	NO	5713	YES	NA	NO	165	NO	NA	NO

NA = Not available, CPRD = Clinical Practice Research Datalink GOLD, IMASIS= Institut Municipal Assistència Sanitària Information System, SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària.

12.4.2.1 Model diagnostics

sARIMAX models (those which added a model term for the effect of COVID-19 pandemic) showed marginally better fit evaluation metrics than sARIMA models (i.e., models not including this term). Using CPRD as an example, the average ratio between sARIMA and sARIMAX in the tune periods were 1.02 (for MAE), and 1.04 (for MAPE), favouring the sARIMAX models. Thus, the models reported below focus on sARIMAX. Equivalent results from sARIMA models are available in the shiny app.

The ACF and PACF plots of the residuals indicated that at least 95% of the values across the plotted lags fell within the 95% confidence intervals. Specifically, for amoxicillin, the ACF and PACF plots showed no more than one value outside these intervals, except for the initial value at lag 0 in the ACF plot. It is an autocorrelation of the series with itself, which by definition equals one. (**Figure 9**)

QQ-plots of the residuals demonstrate that while some models exhibit outliers resulting in deviations from normality, others maintain distributions that align with normality. For instance, the residuals of amoxicillin in the CPRD dataset show such deviation, whilst residuals in IMASIS conform closely to a normal distribution. (**Figure 10**).

Scatterplots of standardised residuals over time for homoscedasticity test revealed that all models exhibited constant variance and no discernible patterns, indicating an appropriate fit of the models. (

Figure 11. Scatter of sARIMAX residuals for amoxicillin in each of the databases of the study: A) CPRD, B) IMASIS, C) IQVIA Belgium, D) IQVIA Germany and E). SIDIAP

12.4.2.2 sARIMAX models forecasting 6-month of antibiotics use

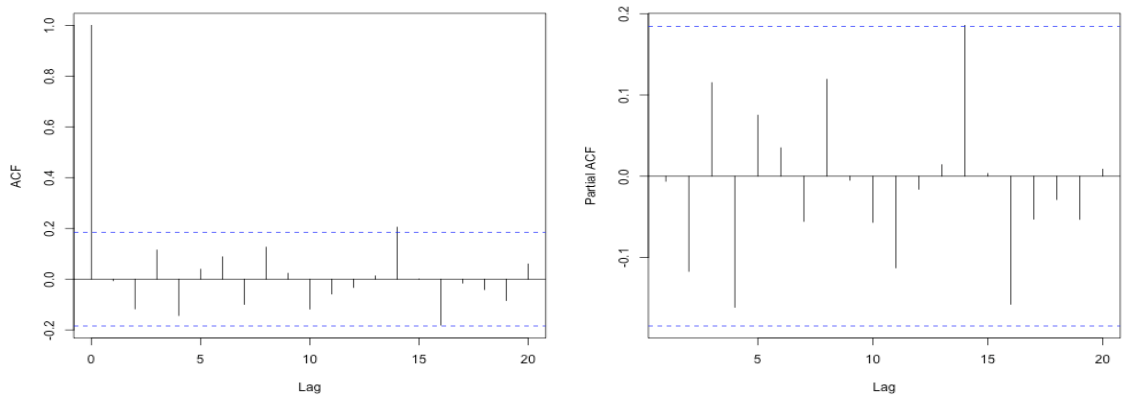
This section contains the sARIMAX models fitted to the times series data produced in section 12.4.1. Data generated in CPRD and SIDIAP was used until the end of data availability, whilst last 6 months of data generated in IMASIS, IQVIA Belgium and IQVIA Germany was removed due to the artefactual decrease in the denominator as described in 9.9.3.3 Time series modelling section.

The plots shown in this section display the times series data generated up to the date displayed in the column "Last month of observed data used in ARIMA(X)" from **Table 4** using a dashed green line (i.e., the observed values); the train and tune values using an orange line (i.e. expected values); and the values forecasted by the model for the next 6 months using a purple line. Vertical back dashed lines indicate the start of the tune and forecast periods.

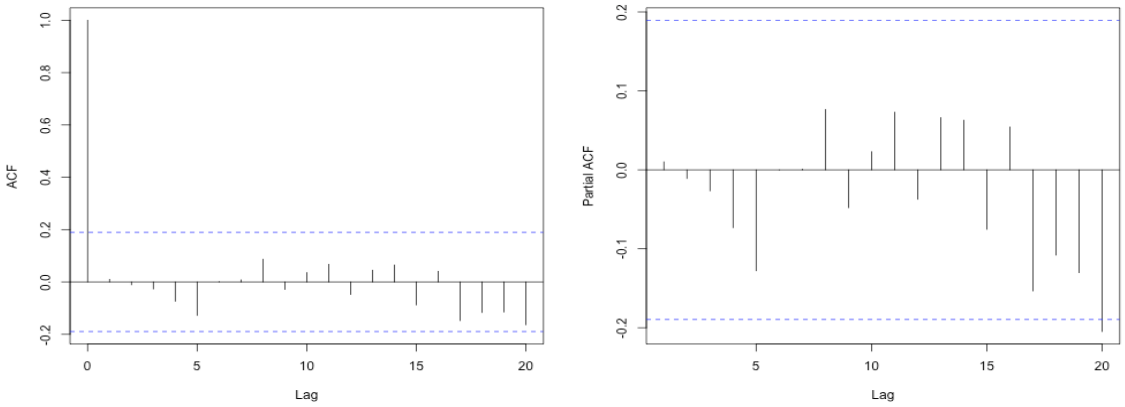
)

Consequently, we conclude that the models demonstrate acceptable levels of goodness of fit, which validates the reliability of their respective coefficients.

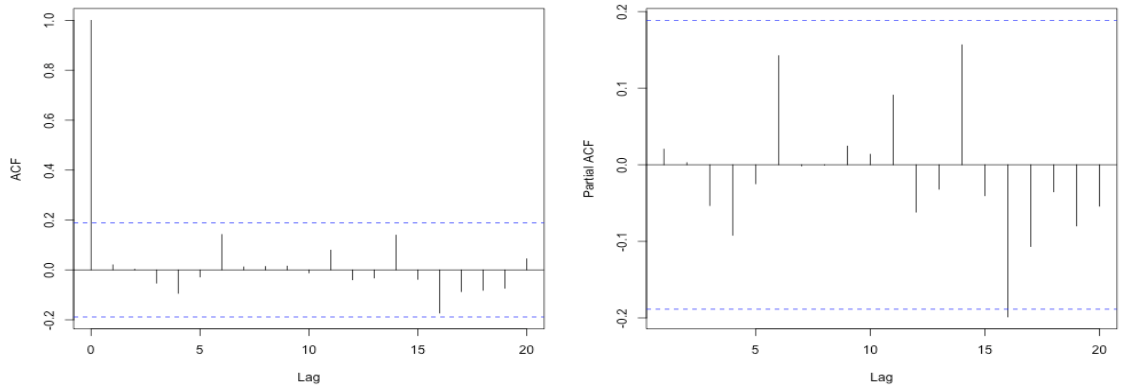
A) ARIMAX residual diagnostics for amoxicillin in CPRD model



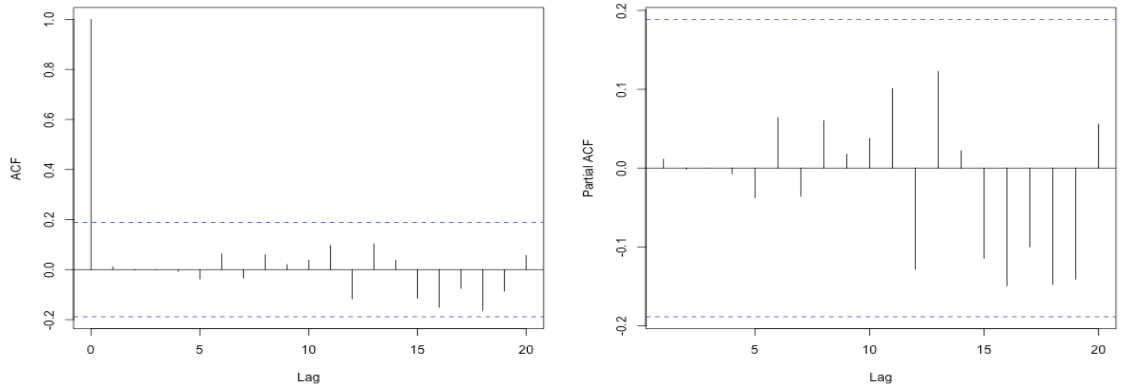
B) ARIMAX residual diagnostics for amoxicillin in IMASIS model



C) ARIMAX residual diagnostics for amoxicillin in IQVIA Belgium model



D) ARIMAX residual diagnostics for amoxicillin in IQVIA Germany model



E) ARIMAX residual diagnostics for amoxicillin in SIDIAP model

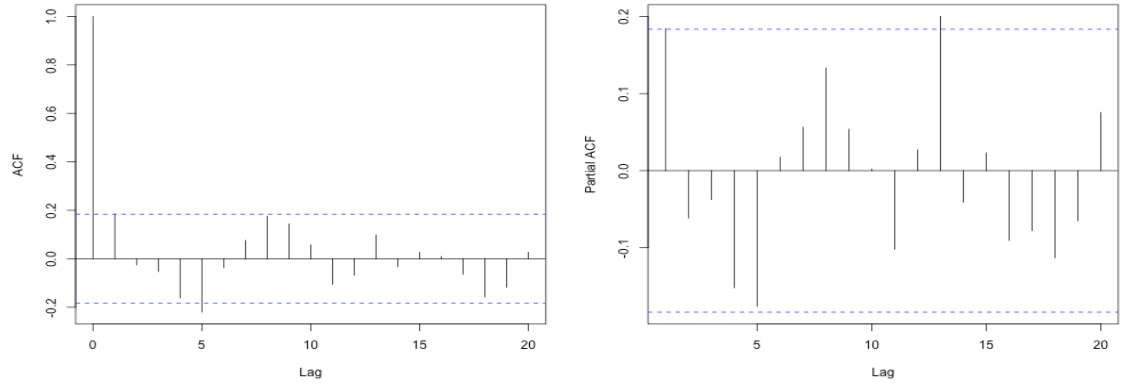
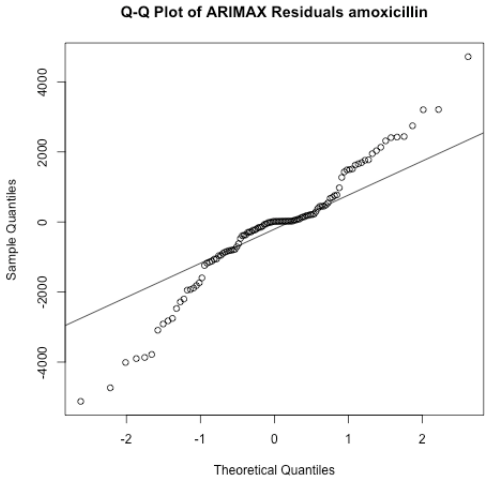
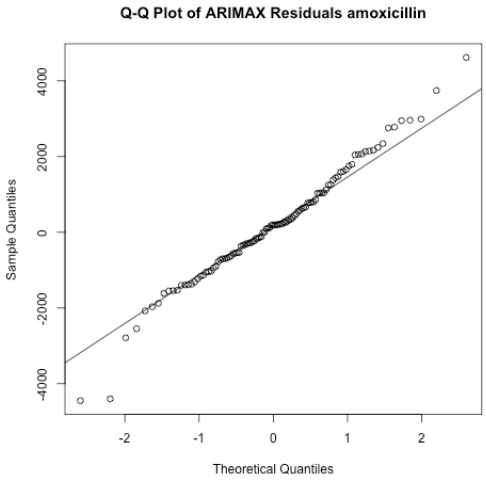


Figure 9. Residual diagnostics plots of autocorrelation function (ACF) and partial ACF in sARIMAX for amoxicillin in each of the databases of the study: A) CPRD, B) IMASIS, C) IQVIA Belgium, D) IQVIA Germany and E). SIDIAP. Blue horizontal dashed lines delimitate the confidence intervals.

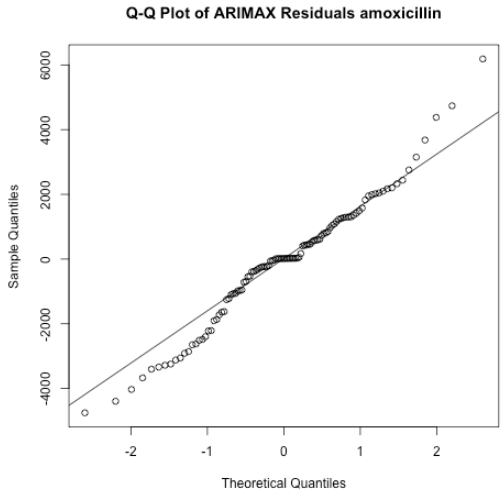
A) CPRD model



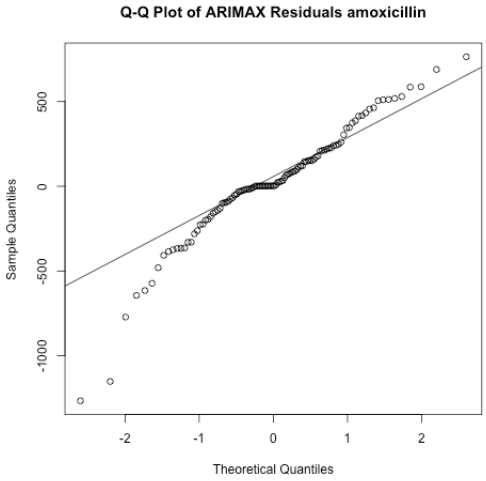
B) IMASIS model



C) IQVIA Belgium model



D) IQVIA Germany model



E) SIDIAP model

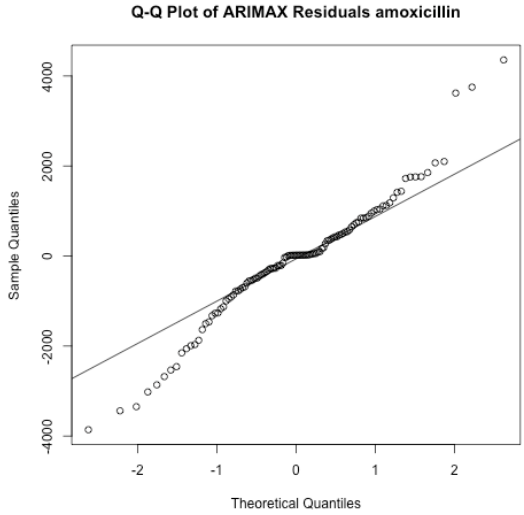
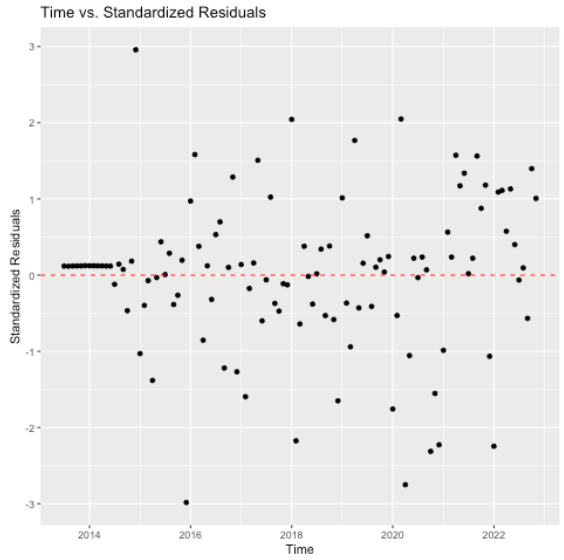
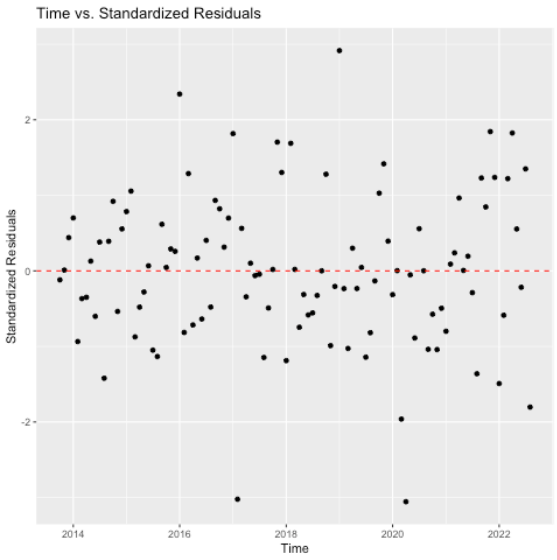


Figure 10. Q-Q plot of sARIMAX residuals for amoxicillin in each of the databases of the study: A) CPRD, B) IMASIS, C) IQVIA Belgium, D) IQVIA Germany and E). SIDIAP.

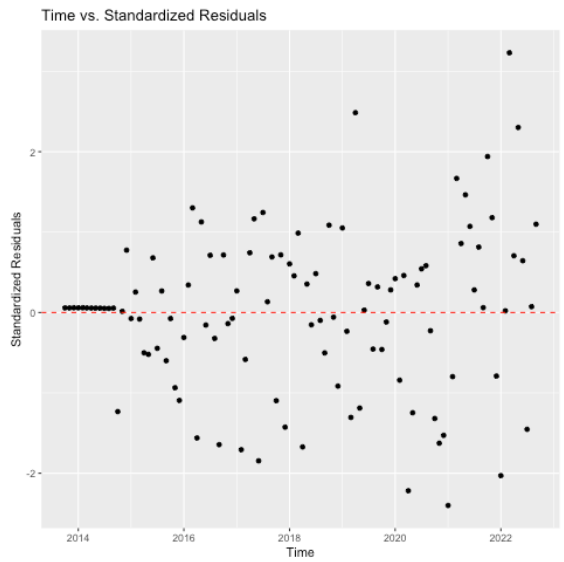
A) CPRD model



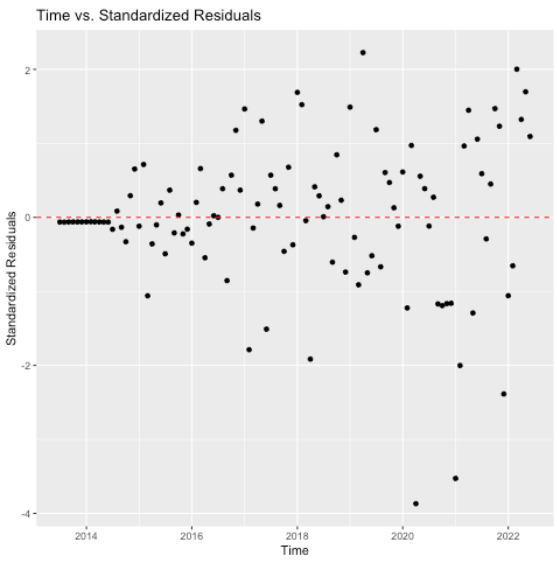
B) IMASIS model



C) IQVIA Belgium model



D) IQVIA Germany model



E) SIDIAP model

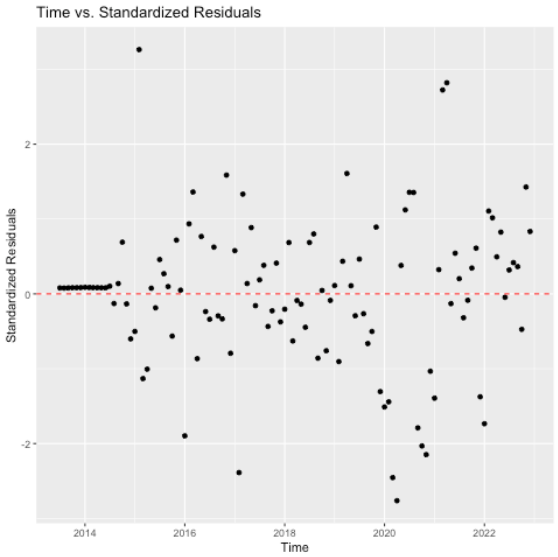


Figure 11. Scatter of sARIMAX residuals for amoxicillin in each of the databases of the study: A) CPRD, B) IMASIS, C) IQVIA Belgium, D) IQVIA Germany and E). SIDIAP

12.4.2.2 sARIMAX models forecasting 6-month of antibiotics use

This section contains the sARIMAX models fitted to the times series data produced in section 12.4.1. Data generated in CPRD and SIDIAP was used until the end of data availability, whilst last 6 months of data generated in IMASIS, IQVIA Belgium and IQVIA Germany was removed due to the artefactual decrease in the denominator as described in 9.9.3.3 Time series modelling section.

The plots shown in this section display the times series data generated up to the date displayed in the column "Last month of observed data used in ARIMA(X)" from **Table 4** using a dashed green line (i.e., the observed values); the train and tune values using an orange line (i.e. expected values); and the values forecasted by the model for the next 6 months using a purple line. Vertical back dashed lines indicate the start of the tune and forecast periods.

.

CPRD data:

Models in CPRD showed in general MAPE values <30% (ranged from 8.1% in amoxicillin/clavulanate to 29.7% in ceftriaxone), except for cefuroxime, which was 66.7%. Cefuroxime use decreased over time (observed IR [95%CI] per 100,000 person-years was 24.4 [20.4-28.9] in July 2023 vs fewer than 5 in March 2023, 2.1 [0.7-4.9] in April 2023, and 4.4 [2.2-7.9] in May 2023). (**Figure 12**)

Peaks (IR [95%CI]) were observed in December 2022 for: amoxicillin (IR [95%CI] observed: 31,203 [30,993-31,414], vs expected: 19,090 [15,599- 22,580]); azithromycin (IR [95%CI] observed: 1,475 [1,428-1,523] , vs expected: 986 [895-1,077]); clarithromycin (IR [95%CI] observed: 5,938 [5,844-6,034] , vs expected: 3,514 [2,862-4,166]); amoxicillin/clavulanate (IR [95%CI] observed: 2,488 [2,427-2,550], vs expected: 2,072 [1,840-2,304]); and phenoxymethylpenicillin (IR [95%CI] observed: 12,016 [11,881-12,151], vs expected: 6,009 [5,296-6,722]).

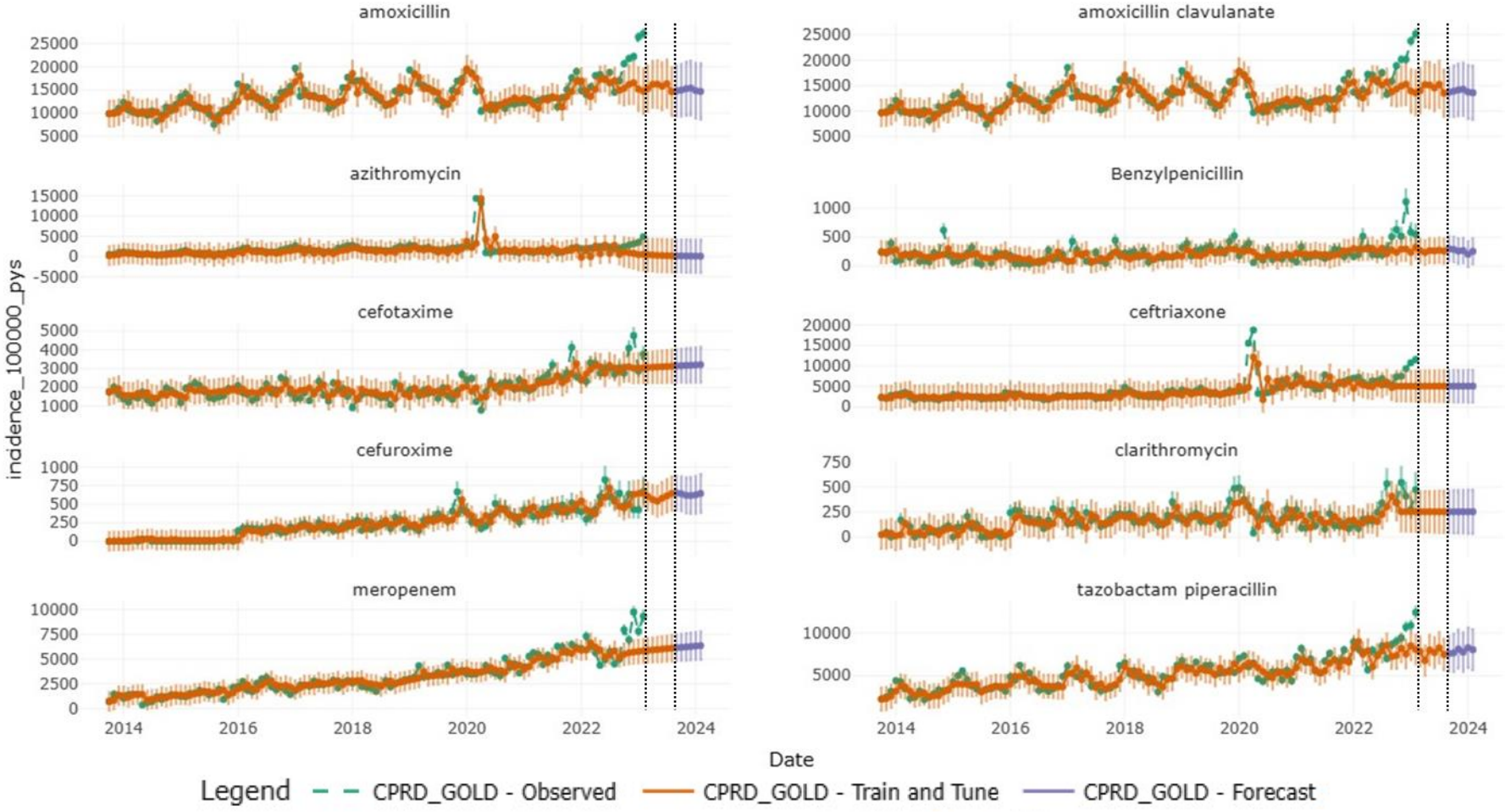


Figure 12: CPRD models. Y-axis display the IR [95%CI] per 100,000 person-years. Abbreviations: CI, confidence interval; IR, incidence rates.

IMASIS data:

Models in IMASIS showed MAPE values <40% (ranged from 17.5% in cefotaxime to 38.3% in ceftriaxone), except for benzylpenicillin (54%), azithromycin (74%) and clarithromycin (153%).

Observed use of all antibiotics of interest was increasingly higher than expected (i.e., train and tune values) from mid-2022-onwards (**Figure 13**). For instance, differences for observed and expected use of amoxicillin or meropenem in October 2022 (IR[95%CI] per 100,000 person-years of amoxicillin use observed: 20,668 [19,832-21,531] vs expected: 15,374 [11,427-19,321]; meropenem use observed: 7,978 [7,453-8,530] vs expected: 5,486 [4,296-6,677]) kept increasing, as values from February 2023 showed (IR[95%CI] per 100,000 person-years of amoxicillin use observed: 27,192 [26,090-28,329] vs expected: 14,748 [9,634-19,862]; meropenem use observed: 9,339 [8,686-10,027] vs expected: 5,833 [4,510-7,156]). These outliers in the observed use of antibiotics might be related with the artefactual decrease in the denominator as described in *9.9.3.3 Time series modelling* section.

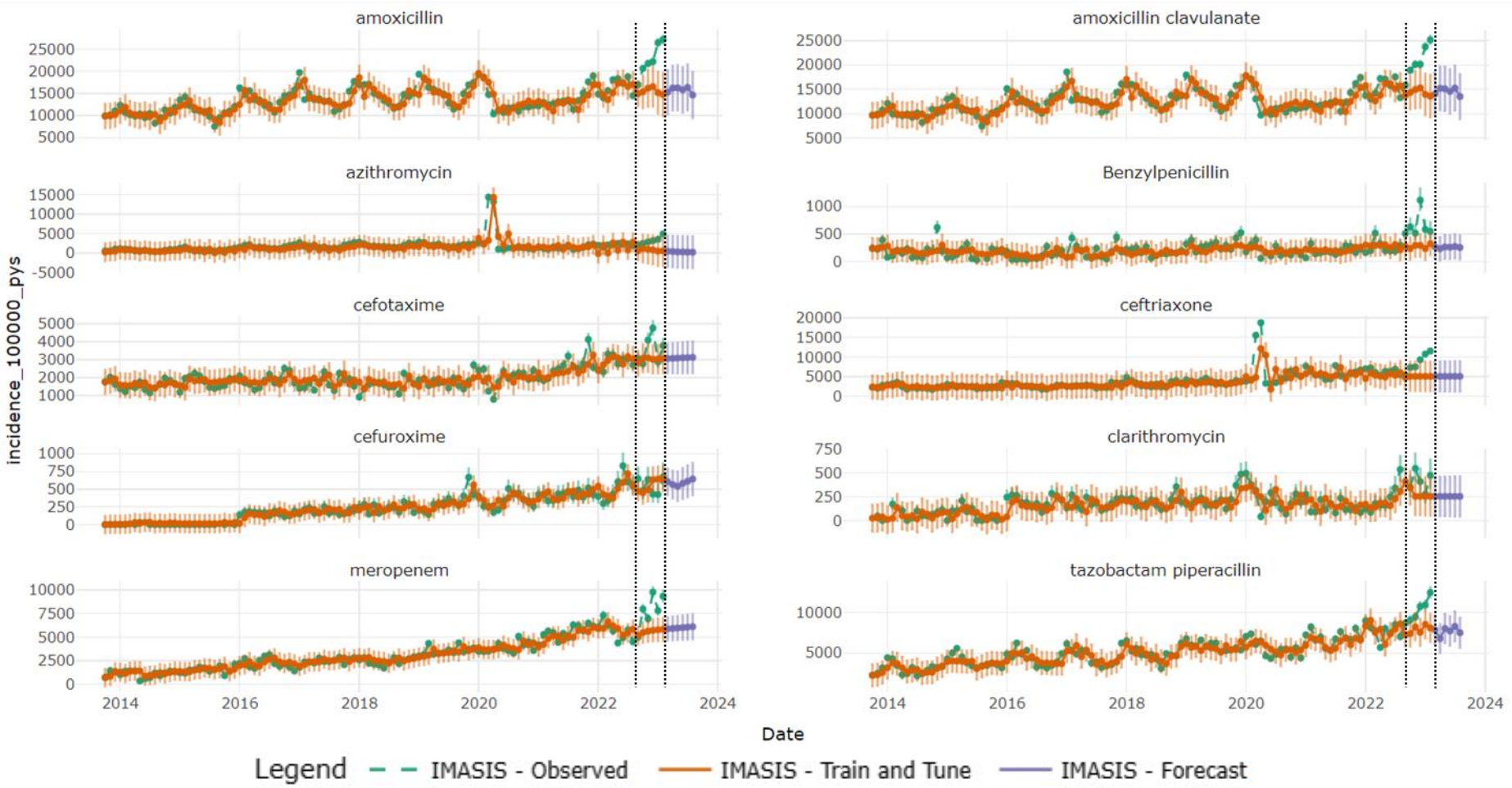


Figure 13: IMASIS models. Y-axis display the IR [95%CI] per 100,000 person/month. Abbreviations: CI, confidence interval; IR, incidence rates.

IQVIA Belgium data:

Models in IQVIA Belgium showed a MAPE values ranged from 22.3% (benzylpenicillin) to 52.4% (clarithromycin), except for phenoxymethylpenicillin, with MAPE value not calculated due to observed IR values at the tune period dropping to 0 (**Figure 14**).

A higher use of selected antibiotics of interest (except for phenoxymethylpenicillin) in the observed estimates when compared to the expected use (i.e., tune period) was also seen in this data base from November 2022-onwards. For instance, IR[95%CI] per 100,000 person-years in the observed vs expected use of amoxicillin November 2022: 22,684 [22,090-23,291] and 15,414 [9,985-20,844], respectively; and its use in March 2023: 27,987 [27,245-28,744] and 18,318 [12,249-24,386] for observed and expected, respectively. These outliers in the observed use of antibiotics might be related with the artefactual decrease in the denominator as described in *9.9.3.3 Time series modelling* section.

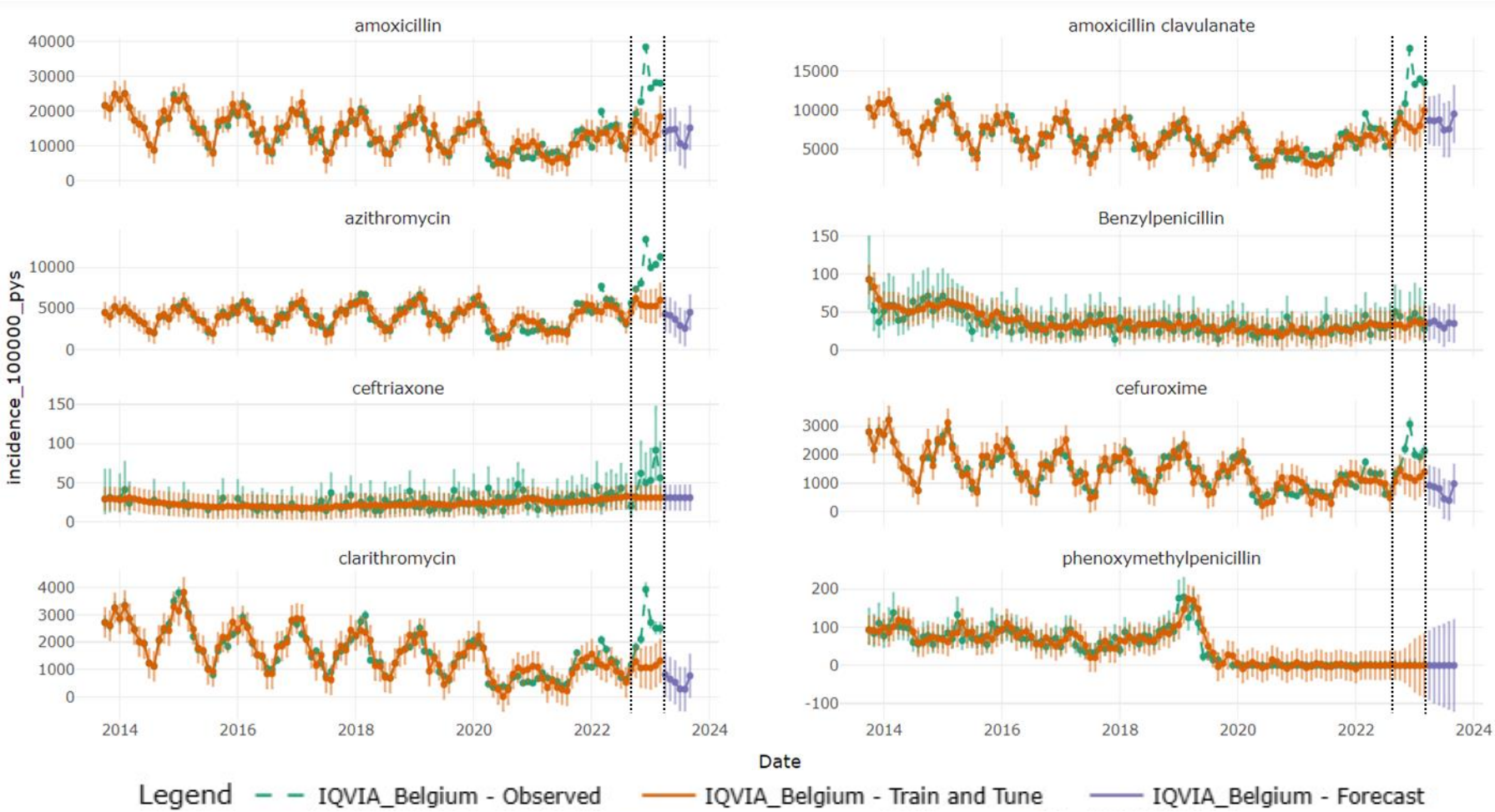


Figure 14: IQVIA Belgium models. Y-axis display the IR [95%CI] per 100,000 person-years. Abbreviations: CI, confidence interval; IR, incidence rates.

IQVIA Germany data:

Models in IQVIA Germany showed MAPE values <35% (ranged from 8.7% in benzylpenicillin to 34.5% in amoxicillin).

A higher observed use of selected antibiotics of interest, except for benzylpenicillin and ceftriaxone, when compared to the expected use (i.e., tune period) was also seen in this data base from the last months of 2022-onwards (**Figure 15**). For instance, the increase in the observed vs expected use of amoxicillin: IR [95%CI] per 100,000 person-years in September 2022 were 3,495 [3,442-3,548] and 2,527 [1,322-3,732], respectively; and its use in December 2022 were 8,434 [8,347-8,522] and 2,396 [1,079-3,714] for observed and expected, respectively. These outliers in the observed use of antibiotics might be related with the artefactual decrease in the denominator as described in *9.9.3.3 Time series modelling* section.

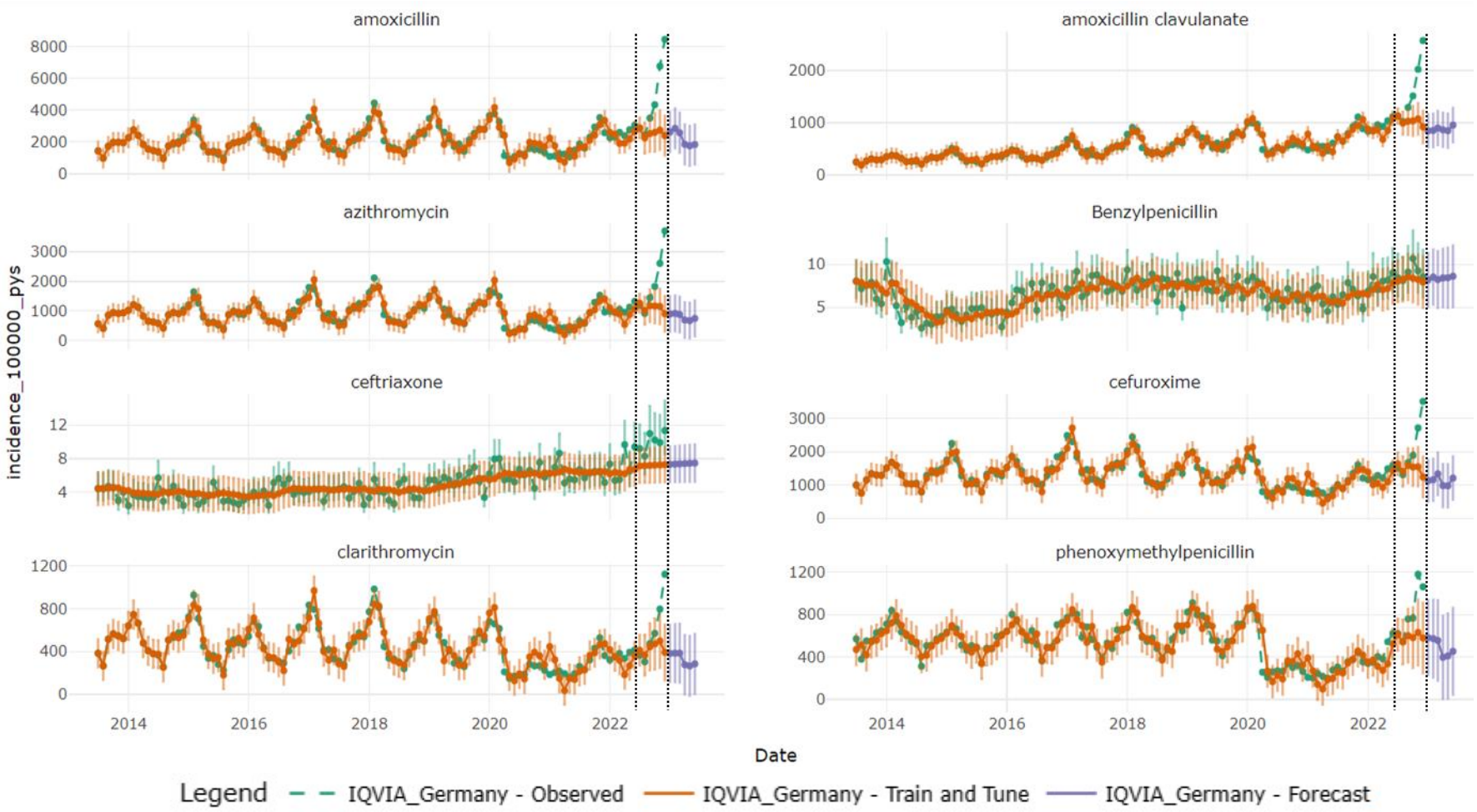


Figure 15: IQVIA Germany models. Y-axis display the IR [95%CI] per 100,000 person-years. Abbreviations: CI, confidence interval; IR, incidence rates.

SIDIAP data:

Models in SIDIAP showed MAPE values <20% (ranged from 4.3% in cefuroxime to 19.6% in ceftriaxone), except for azithromycin (40.3%) and benzylpenicillin (41.7%).

Observed use of azithromycin was lower than expected (i.e., tune period) at January 2023 (IR [95%CI] per 100,000 person-years observed: 4,330 [4,273-4,388], vs expected: 5,413 [4,631-6,196]) and for observations from February 2023 until the end of data availability (e.g., observed vs expected use in April 2023: 2,517 [2,473-2,561] and 4,127 [2,806-5,449], respectively) (**Figure 16**). Observed use of benzylpenicillin decreased from January 2014 to June 2023 (IR [95%CI] per 100,000 person-years: 21.8 [17.9-26.3] to 3.9 [2.3-6.1], respectively). Although the expected value for the use of benzylpenicillin at the end of the tune period was lower than the observed, their confidence intervals overlapped (IR [95%CI] value expected in June 2023: 1.3 [-5.2-7.9]).

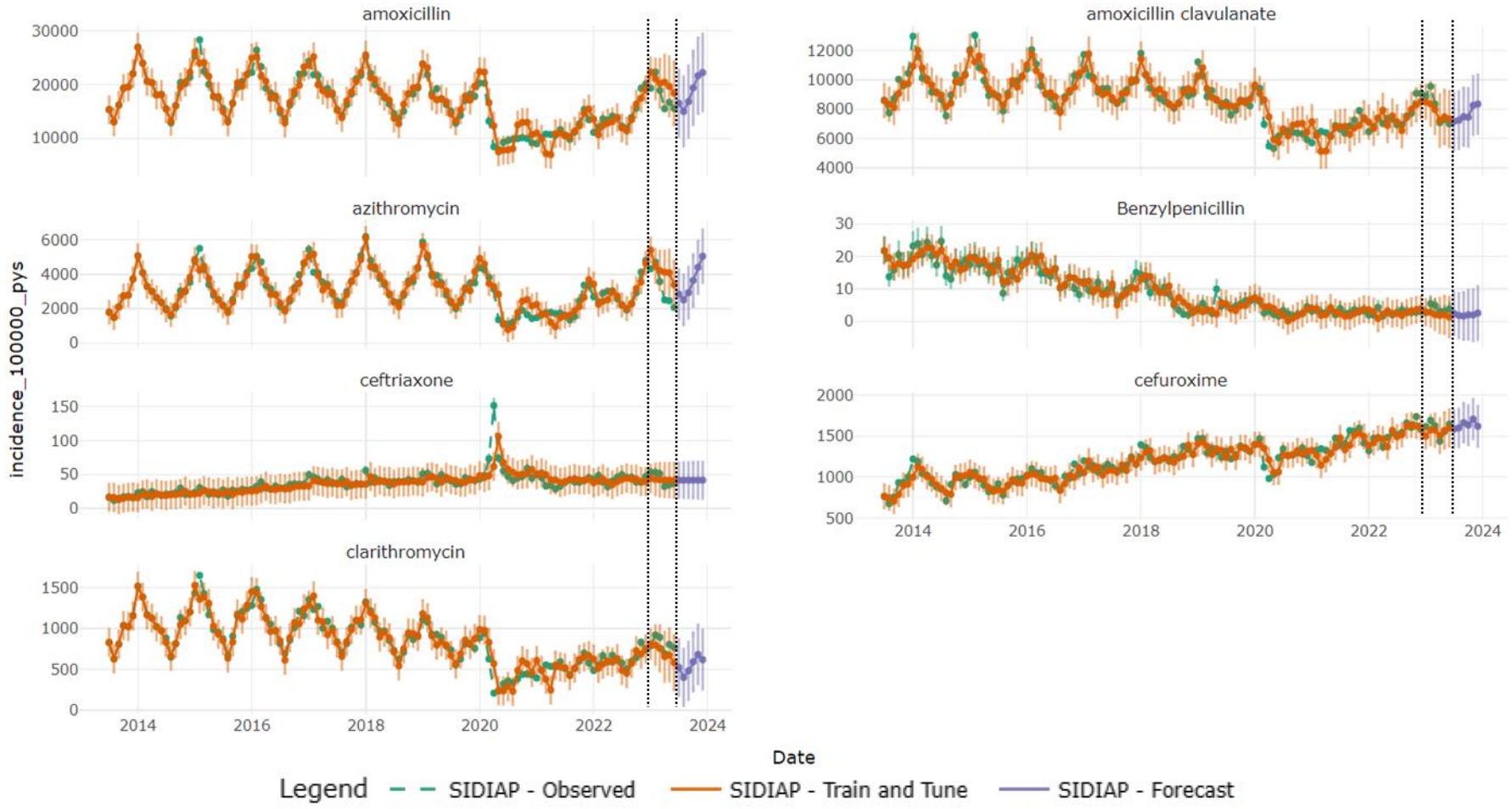


Figure 16: SIDIAP models. Y-axis display the IR [95%CI] per 100,000 person-years. Abbreviations: CI, confidence interval; IR, incidence rates.

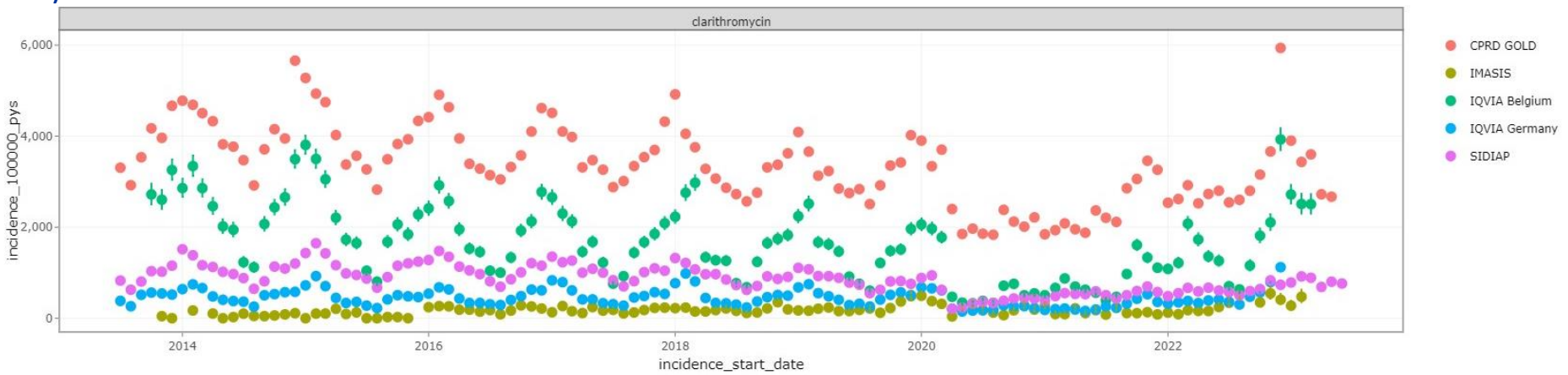
12.5 Sensitivity analyses

12.5.1 New users cohort

Sensitivity analyses for new user cohorts are available in the shiny app https://data-dev.darwin-eu.org/connect/#/apps/71c3f554-ef71-4ccd-b135-4a5d34809a63/accesshttps://data-dev.darwin-eu.org/connect/, within the 'Population-level Drug Utilization' tab.

When restricting the analysis for new users, most of the results were consistent, presenting similar trends over time (**Figure 17**), with the exception of amoxicillin, and amoxicillin and clavulanate for IMASIS. When restricting users by not been exposed to the drug of interest for at least 30 days prior the current use, incidence of and amoxicillin alone or combined with clavulanate decreased in IMASIS. (**Figure 18**).

A) Users



B) New users

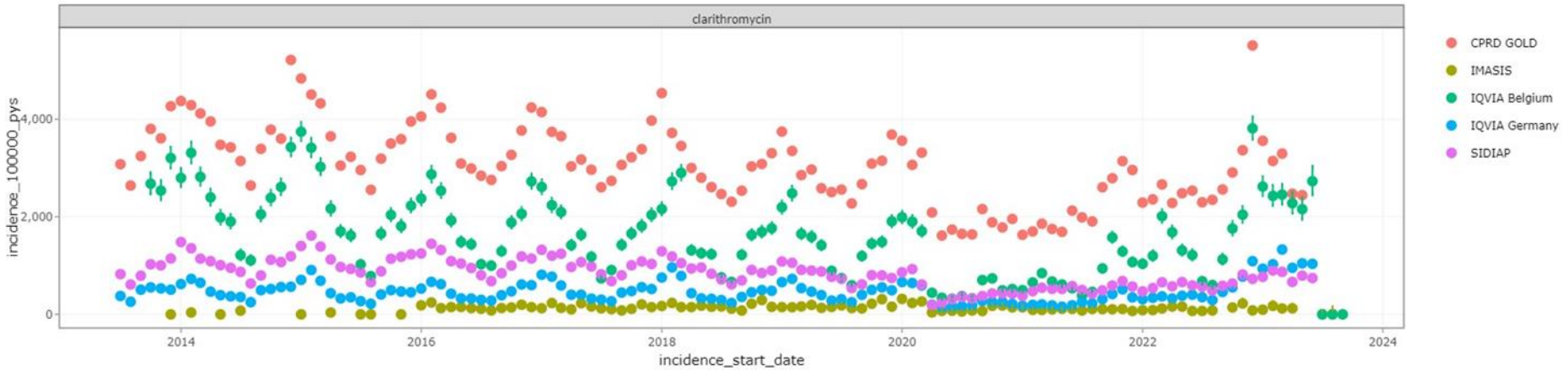


Figure 17: Incidence rates of use of clarithromycin per 100,000 person-years in the overall population for A) users and B) new users (i.e., not have been exposed to the drug of interest for at least 30 days prior the current use).

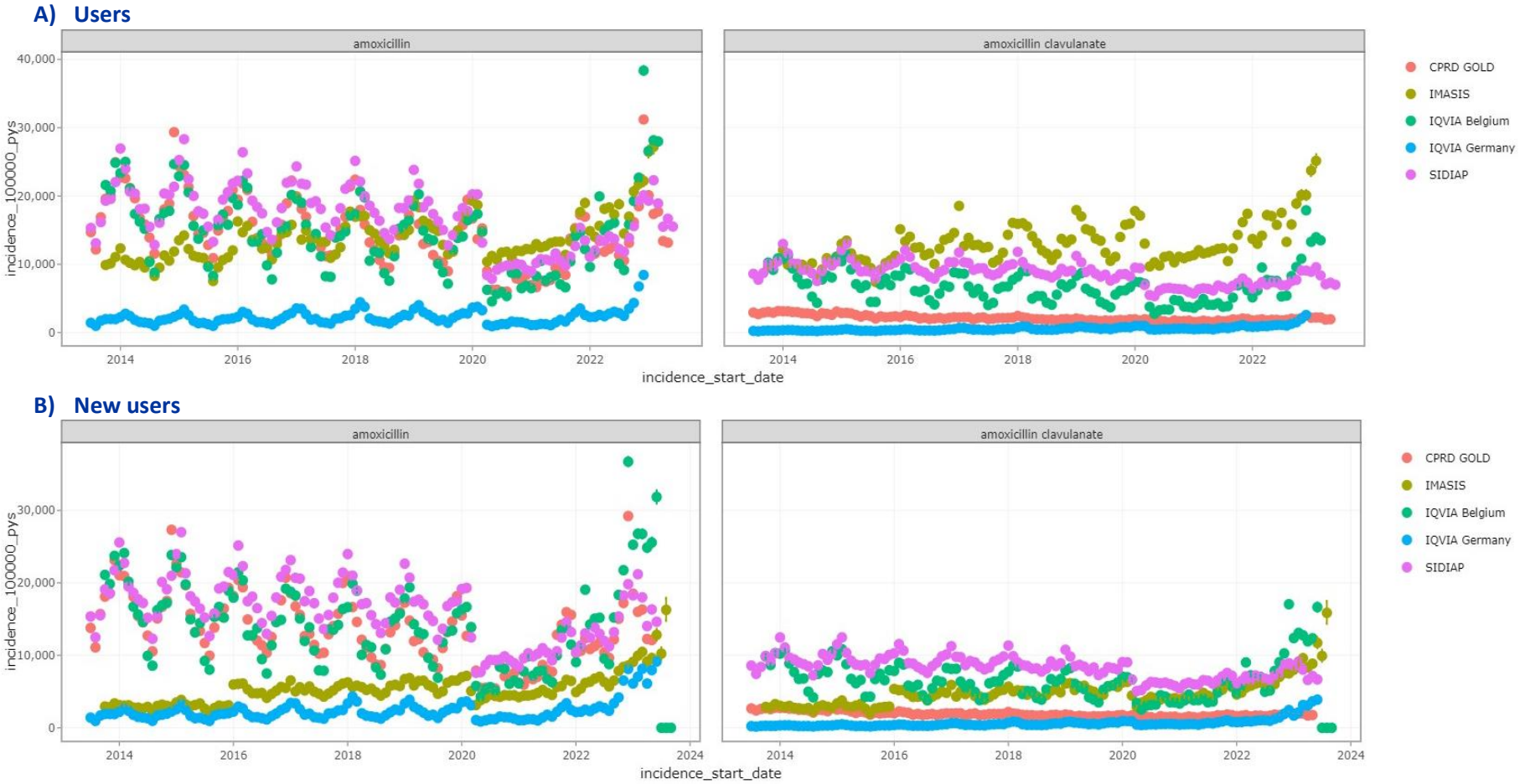


Figure 18: Incidence rates of use of amoxicillin and, amoxicillin and clavulanate per 100,000 person-years in the overall population for A) users and B) new users (i.e., not have been exposed to the drug of interest for at least 30 days prior the current use).

When examining trends of antibiotics on new users, use of clarithromycin was highest in CPRD, followed by IQVIA Belgium (**Figure 17B**); whilst use of azithromycin was highest in IQVIA Belgium and SIDIAP when compared to the other databases (**Figure 19**).

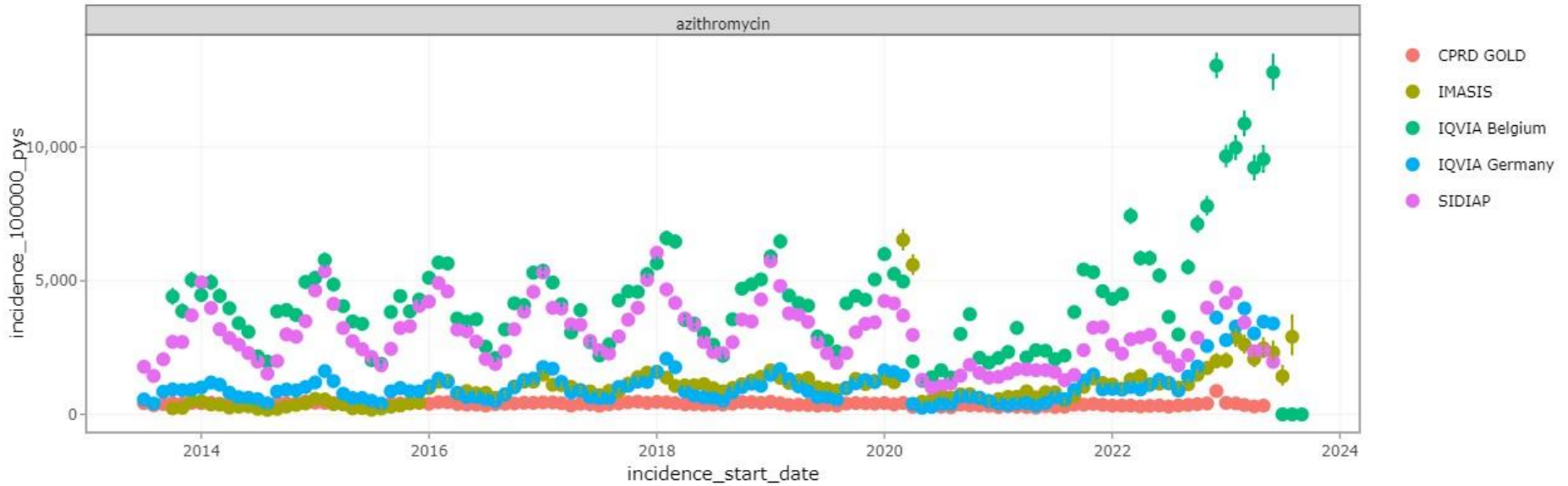


Figure 19. Incidence rates of use of azithromycin per 100,000 person-years in the overall population for new users (i.e., not have been exposed to the drug of interest for at least 30 days prior the current use).

IQVIA Belgium showed the higher use of cefuroxime in 2014, but it continuously decreased until August 2022. Conversely, its use in IQVIA Germany and SIDIAP moderately increased from 2013, and its use in IMASIS started in 2016 and increased over time (**Figure 20**).

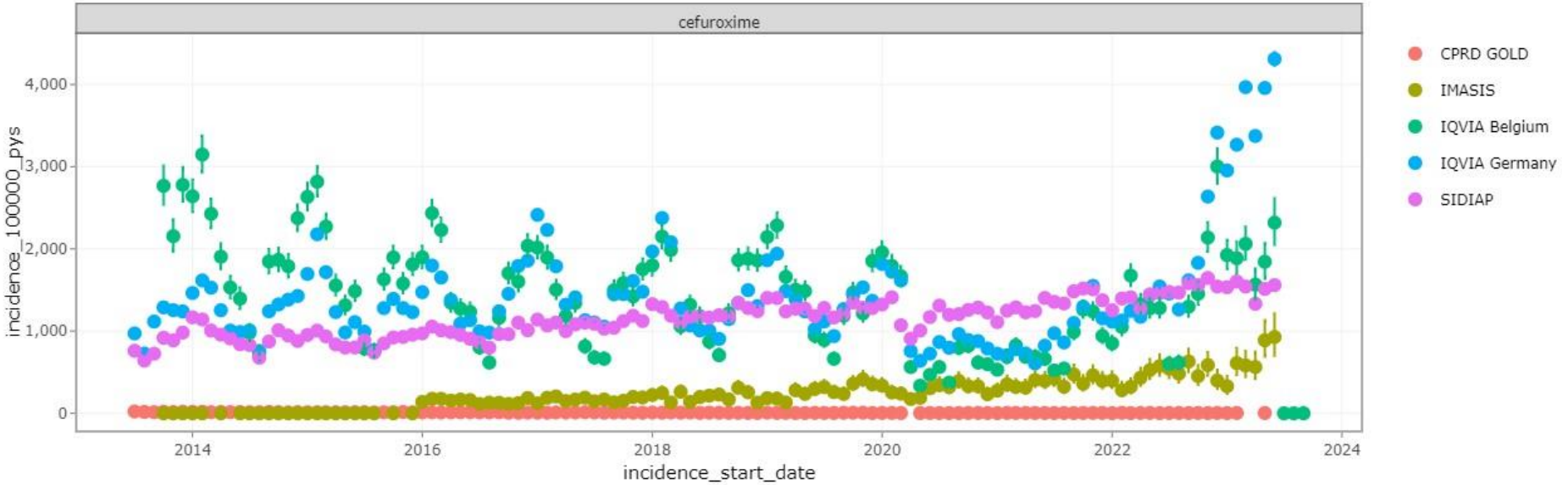


Figure 20. Incidence rates of use of cefuroxime per 100,000 person-years in the overall population for new users (i.e., not have been exposed to the drug of interest for at least 30 days prior the current use).

12.5.2 Forecasting 12-month of data

Additionally, an extended forecasting of 12 month rather than 6 months was performed. Full results can be explored in the following shiny app (which differs from the one containing the main results of the analysis): https://data-dev.darwin-eu.org/content/78820a82-7117-460a-a12a-398c9da7415f .

12.5.2.1 Model diagnostics

The 12-month forecast has been based in the same models reported in section 12.4.2. Thus, the coefficients and the model diagnostics for the 12-month forecast are identical to the ones reported for the 6-month forecast.

12.5.2.2 sARIMAX models forecasting 12-month of antibiotics use

CPRD data:

The extended forecasted period (including 6- additional months, corresponding from 01-06-2023 to 01-12-2023) suggest use of cefuroxime continue decreasing whilst use of azithromycin continue increasing in CPRD. Last IR ([95%CI] per 100,000 person-years) value forecasted for these drugs were 0.6 [-3.6-5.0] and 1,026 [891-1162], respectively. (**Figure 21**)

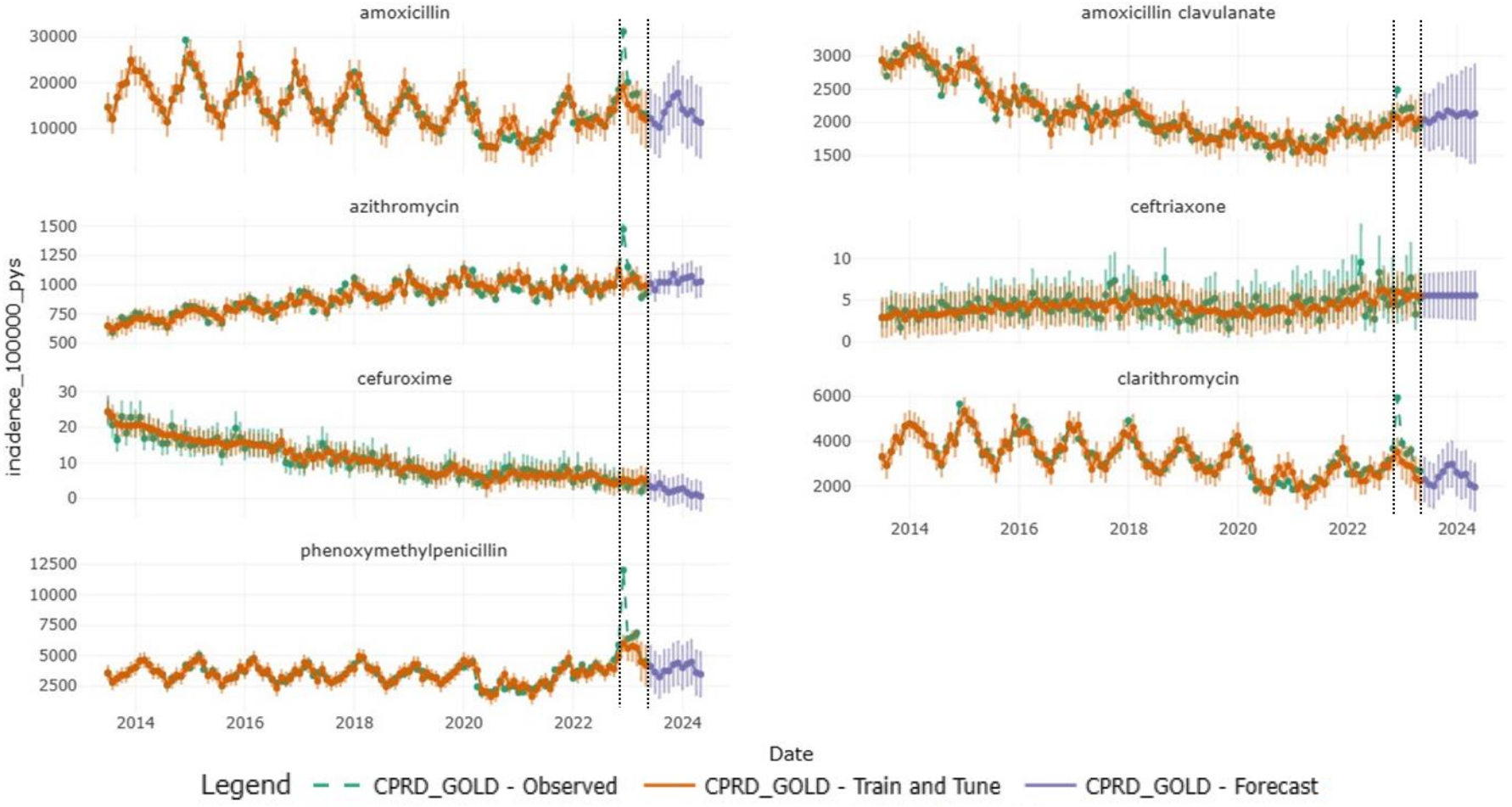


Figure 21. CPRD models forecasting 12-months. Y-axis display the IR [95%CI] per 100,000 person/month. Abbreviations: CI, confidence interval; IR, incidence rates.

IMASIS data:

The extended forecasted period (including 6- additional months, corresponding from 01-08-2023 to 01-02-2024) suggest an increment in the use of cefotaxime, cefuroxime, meropenem, tazobactam/piperacillin in IMASIS. Last IR ([95%CI] per 100,000 person-years) value forecasted for these drugs were 3,204 [2,197-4,212], 644 [368-919], 6,370 [4,827-7,914], and 8,060 [5,522-10,599], respectively. (**Figure 22**)

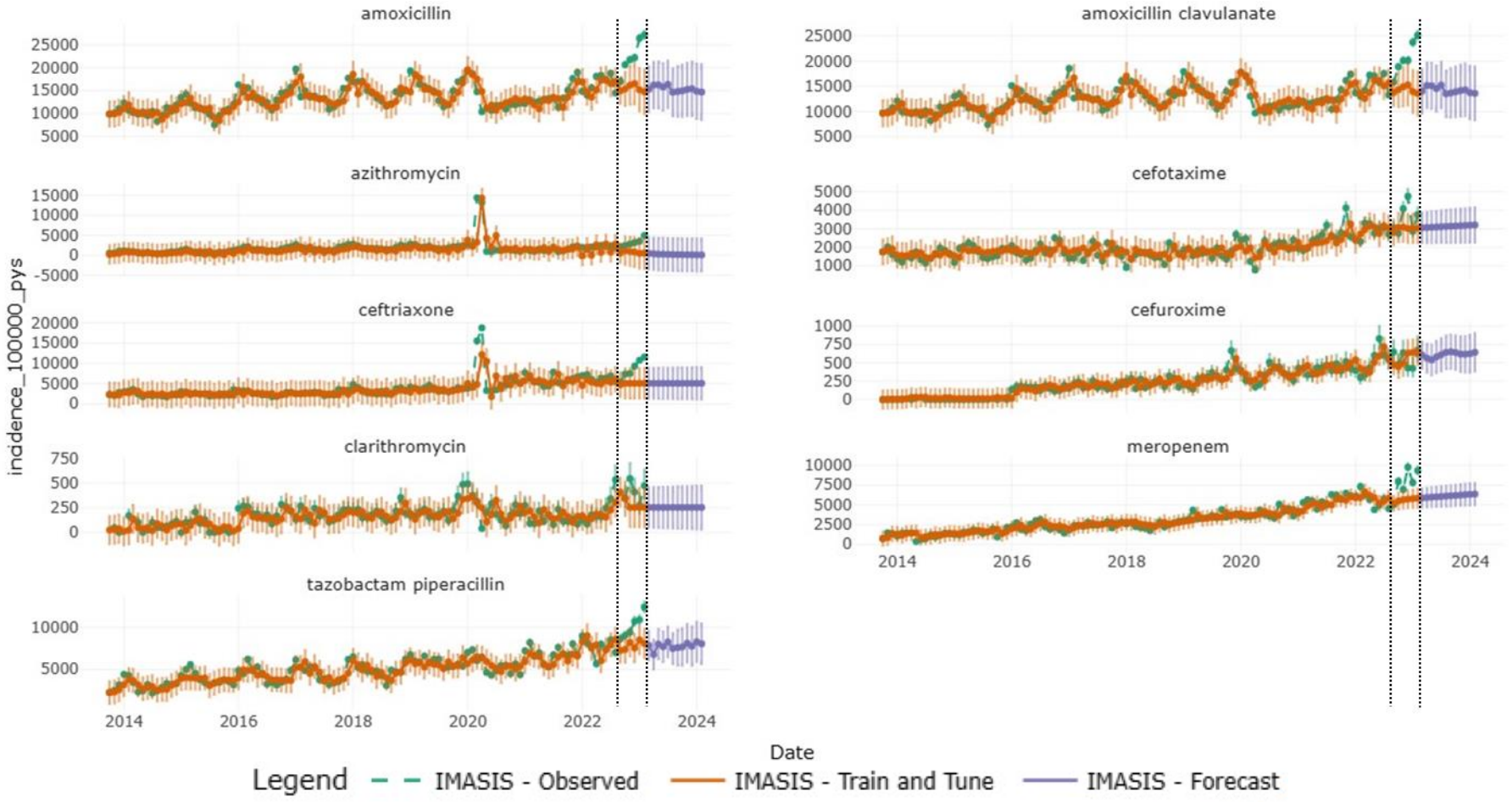


Figure 22. IMASIS models forecasting 12-months. Y-axis display the IR [95%CI] per 100,000 person/month. Abbreviations: CI, confidence interval; IR, incidence rates.

IQVIA Belgium data:

The extended forecasted period (including 6- additional months, corresponding from 01-09-2023 to 01-03-2024) suggest an increment in the use of amoxicillin, alone or combined, in IQVIA Belgium. Last IR ([95%CI] per 100,000 person-years) value forecasted for these drugs were 19,634 [11,830- 27,438] and 12,707 [7,530-17,884], respectively.

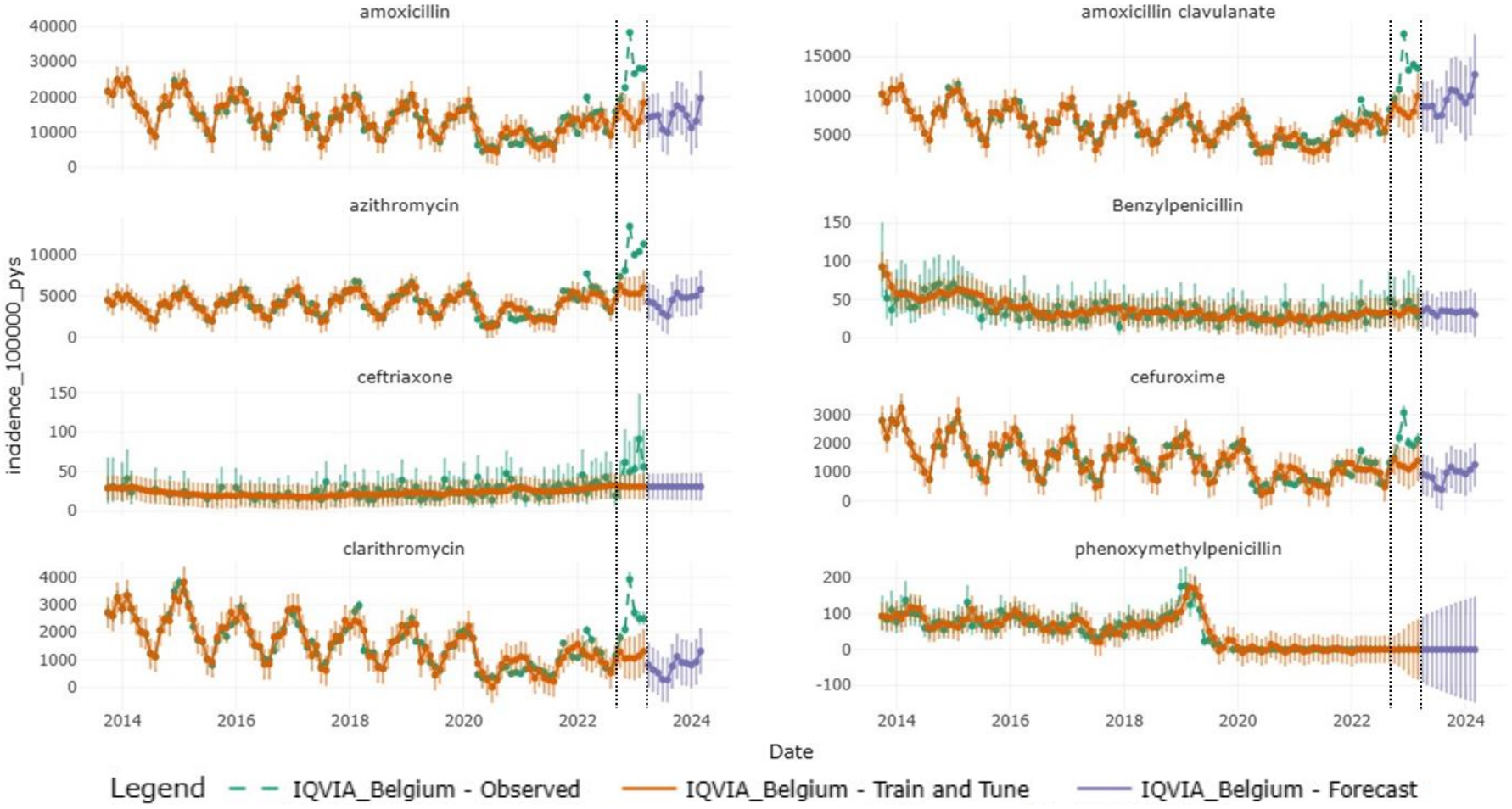


Figure 23. IQVIA Belgium models forecasting 12-months. Y-axis display the IR [95%CI] per 100,000 person/month. Abbreviations: CI, confidence interval; IR, incidence rates.

IQVIA Germany data:

The extended forecasted period (including 6- additional months, corresponding from 01-06-2023 to 01-12-2023) suggest an increment in the use of amoxicillin/clavulanate and ceftriaxone in IQVIA Germany. Last IR ([95%CI] per 100,000 person-years) value forecasted for these drugs were 1,001 [613-1,389] and 8 [5-10], respectively. (**Figure 24**)

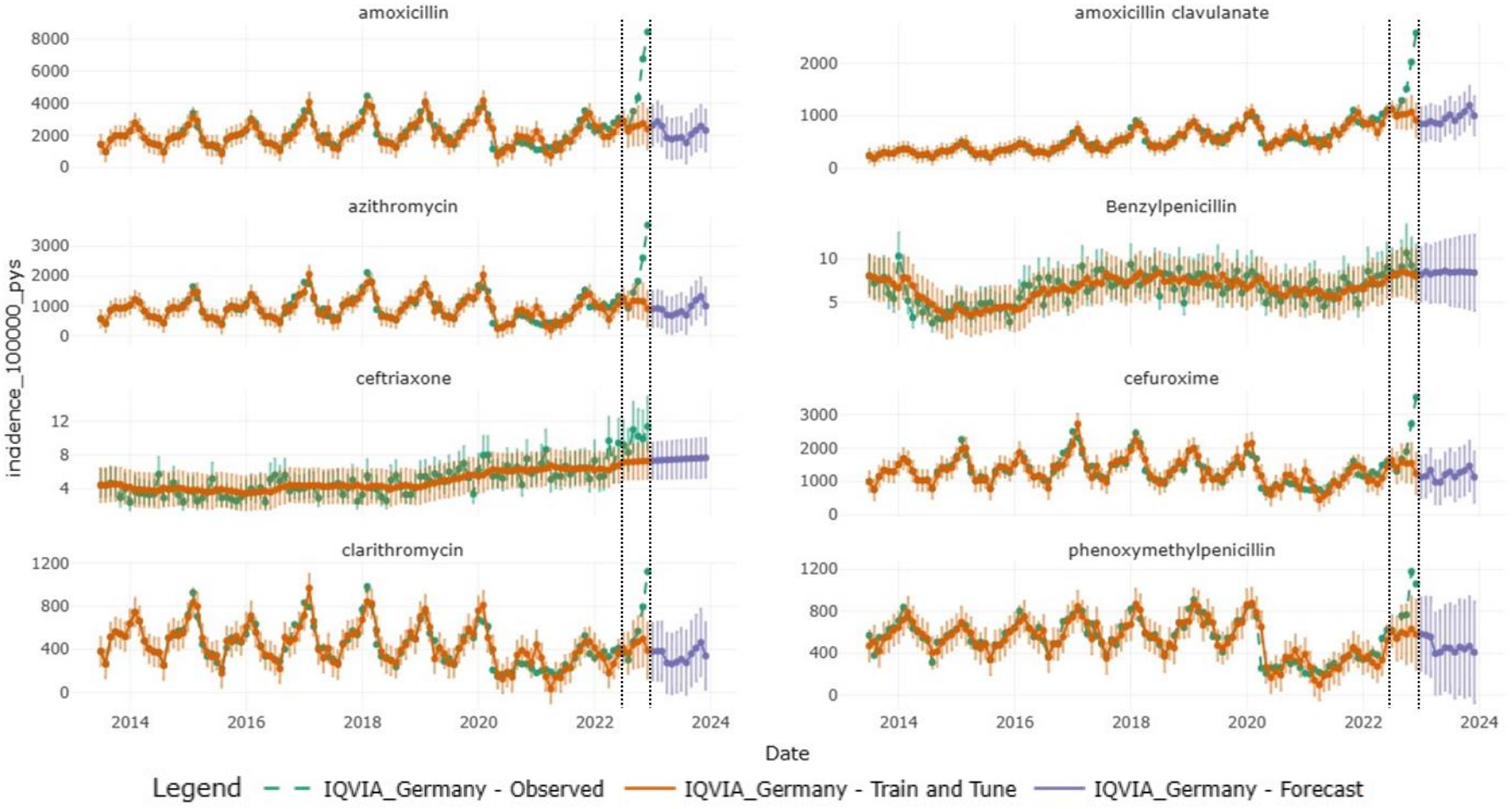


Figure 24. IQVIA Germany models forecasting 12-months. Y-axis display the IR [95%CI] per 100,000 person/month. Abbreviations: CI, confidence interval; IR, incidence rates.

SIDIAP data:

The extended forecasted period (including 6- additional months, corresponding from 01-12-2023 to 01-06-2024) suggest an increment in the use of cefuroxime in SIDIAP. Last IR ([95%CI] per 100,000 person-years) value forecasted for this drug was 1,630 [1,281-1,980]. (**Figure 25**)

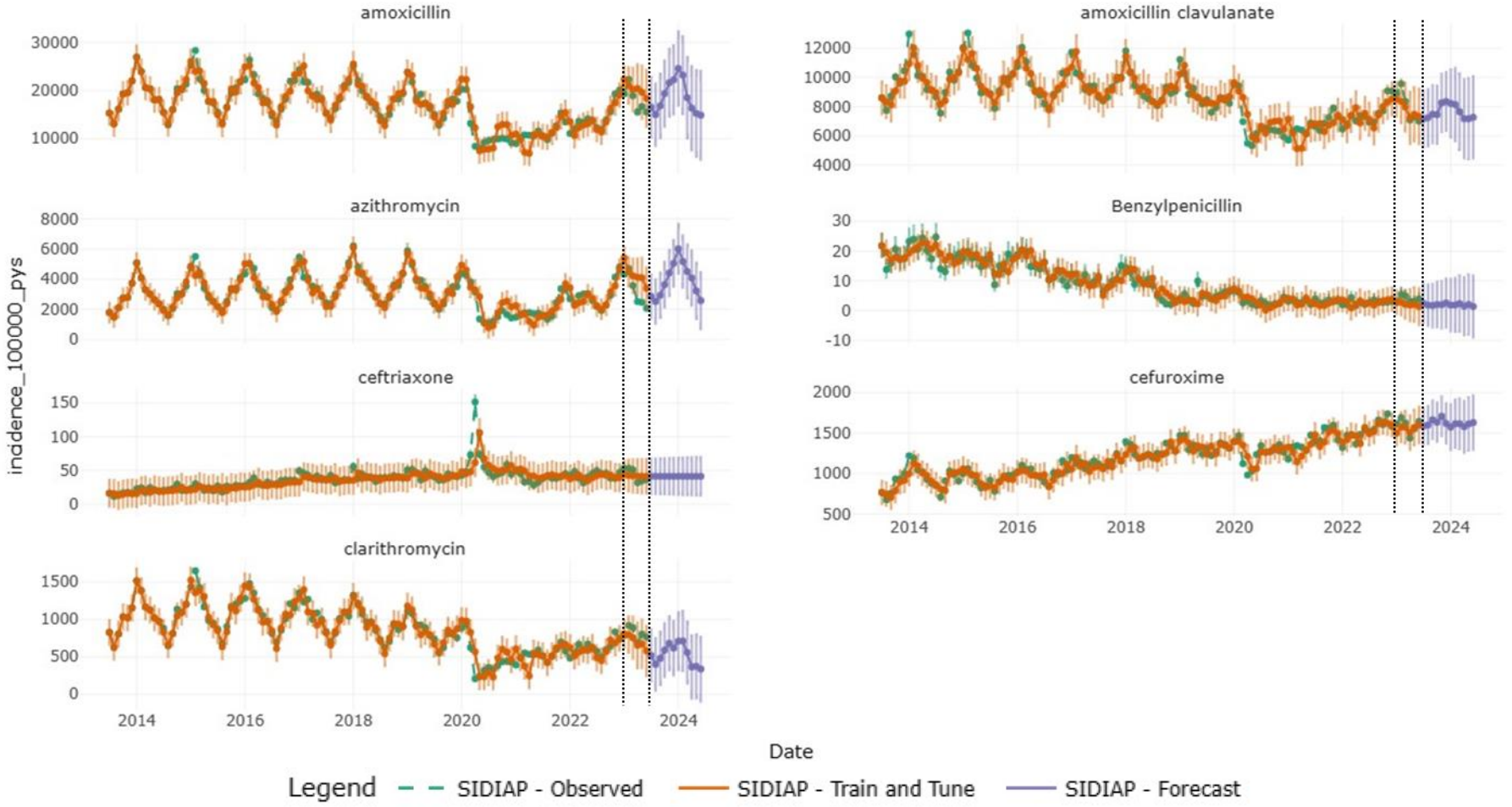


Figure 25. SIDIAP models forecasting 12-months. Y-axis display the IR [95%CI] per 100,000 person/month. Abbreviations: CI, confidence interval; IR, incidence rates.

13 MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions were not collected or analysed as part of this evaluation. The nature of this non-interventional evaluation, through the use of secondary data, does not fulfil the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf).

Only in case of prospective data collection, there is a need to describe the procedures for the collection, management and reporting of individual cases of adverse events/adverse reactions.

14 DISCUSSION

14.1 Key Results

Population-level drug utilisation

Overall, the results showed a seasonality pattern with peaks around November-March months in antibiotics used typically for bacterial infections frequent in winter, such as amoxicillin or azithromycin, across the 5 databases. The age-stratified analysis showed the use of antibiotics was generally higher in individuals aged 65 or older. When stratified by sex, women had a higher incidence in primary care; whilst men generally presented larger use in hospital and outpatient data. Additionally, a consistent pattern across the 5 databases where use of most of the antibiotics of interest experienced a substantial drop after the start of the COVID-19 pandemic (March 2020).

Despite these overall trends, there were specific drugs presenting different trends. For instance, cefuroxime use decreased over time in CPRD, benzylpenicillin decreased in SIDIAP, and phenoxymethylpenicillin in IQVIA Belgium was discontinued after September 2019. Observed incidence rates higher than forecasted were observed in the use of amoxicillin, azithromycin, clarithromycin, amoxicillin/clavulanate and phenoxymethylpenicillin during December 2022 in CPRD.

Crude number of antibiotic users in IMASIS, IQVIA Belgium and IQVIA Germany were increased during winter 2022 when compared to previous months. However, IR produced in the last 6 months of data available in these three databases (which contained the winter months from 2022), were inaccurate due to artefactual increases in rates of antibiotic use associated with the incomplete computation of the observation period.

IMASIS, IQVIA Belgium and IQVIA Germany uses the last observation/interaction of the individual with their health care system to determine the end of the observation period. When the observation period is based on the last observation or interaction, infrequent users have shorter follow-up, which translates in a decreasing of the time at risk (i.e., the denominator) in the IR estimates. However, users of antibiotics had, per definition, a contact with their health care system and therefore number of users of antibiotics (i.e., the numerator) are fully captured until the end the study period. Thus, during the latest IR estimated in IMASIS, IQVIA Belgium and IQVIA Germany, the numerator (i.e., number of antibiotic users) was stable whilst the

denominator decreased, overestimating the IR observed in the last 6 months of available data. Consequently, these periods were excluded from the sARIMA(X) modelling in these three databases.

Time series modelling

Based on the model diagnostics, sARIMAX presented a marginal better fit than sARIMA models. sARIMAX models demonstrate adequate levels of goodness of fit, which validates the reliability of their respective coefficients. Additionally, sARIMAX successfully modelled the trend changes due to the impact of the COVID-19 pandemic.

A number of observations were out of the 95% predictive intervals of the train and tune values based on sARIMAX, including the use of amoxicillin, azithromycin, clarithromycin, amoxicillin/clavulanate, phenoxymethylpenicillin in December 2022 in CPRD when compared with the expected use.

Despite the exclusion of the latest 6 months of data available in IMASIS, IQVIA Belgium and IQVIA Germany, observed values after mid-2022 for IMASIS, and end-2022 for both IQVIA databases may still contain artifacts related with the artefactual decrease in the denominator as described in *9.9.3.3 Time series modelling* section. The observed IR from these periods (from the latest 6 to 12 months of available data) were used for tuning the sARIMA(X) models. The values forecasted for the tune periods followed a reasonable trend despite the differences with the observed values. IQVIA Germany seemed the less affected of these three databases (i.e., the one whose difference between observed vs predicted in the tune period was the lowest). The forecasted values for IMASIS and IQVIA databases need to be interpreted with caution, since the forecasted usage may be slightly overestimated, despite the predictive intervals, given that the observed IR used for the forecast may also be slightly overestimated. On the other hand, we provided an extension of the forecasted values (from 6 months to 12 months) as a sensitivity analysis. This extension aims to compensate for the months excluded in IMASIS and IQVIA databases by providing an insight in the trends of use of antibiotics over a more relevant and recent period of time.

14.2 Limitations of the research methods

General limitations:

This study includes routinely collected health care data and so data quality issues must be considered. In particular, a recording of a use (prescription or dispensation) does not mean that the patient actually took the drug. In addition, assumptions around the duration of drug use are unavoidable.

Moreover, forecasting with sARIMA(X) models has several limitations that must be also taken into account.

sARIMA(X) assumes homoscedasticity of the data (variance is constant over time) which might not hold true even after integration of the time series. To overcome this limitation, we evaluated the potential presence of heteroscedasticity in the data, and we explored whether applying a logarithmic transformation of the data prior to fitting the sARIMA(X) model was necessary.

Additionally, we successfully captured and modelled the effect of COVID-19 pandemic as an exogenous variable, but we must acknowledge the possibility of the existence of other unknown exogenous variables that may explain extreme observations, such as the high incidence rates in the prescription of certain drugs during December 2022 in the UK data (i.e., CPRD). These outliers may negatively impact the estimation and

forecasting but have the potential to be used as a valuable source to flag plausible risks (and indicative estimates) of increased demand and possible sources of shortages.

Database-specific limitations:

IMASIS, IQVIA Belgium and IQVIA Germany:

The observation period of the patients in these databases was calculated based on the last visit, observation or interaction of the patient with the health care system. This methodology severely impacts the individuals considered "at risk" for the different antibiotics of the study (i.e., the individuals included in the denominator populations) during the latest months of data, where healthy and/or non-frequent users of the health care system were not considered active. Consequently, the denominators used to calculate the incident use of drugs in the population artefactually decreased whilst the incident users remained, increasing the incidence ratios. Since such artifacts become more relevant when reaching the last months before the instance date lock, and therefore the corresponding values were not reliable, the IR estimates corresponding to the latest 6-months of available data in IMASIS and both IQVIA databases were excluded from the ARIMA modelling. Despite the exclusion of the latest 6 months of data available in IMASIS, IQVIA Belgium and IQVIA Germany, observed values after mid-2022 for IMASIS, and end-2022 for both IQVIA databases may still contain artifacts related with the artefactual decrease in the denominator (i.e. overestimate the IRs).

In future studies, assuming a stability in capturing the patients using the drugs of interest, instead of forecasting the incidence of prescription or dispensation, it can be considered to forecast the crude number of antibiotic users in these three databases, which would not require a denominator.

IQVIA Belgium:

Phenoxymethylpenicillin has been discontinued from the Belgium Market since May 2019, which explains the stop in its use during this same year.(7) However, we observed use of phenoxymethylpenicillin in IQVIA Belgium until September 2019. The observed delay (i.e., the prescriptions observed after May) can be interpreted as use of the last stocks available in the country or driven by GPs persisting in prescribing phenoxymethylpenicillin until they became accustomed to prescribing other alternatives. However, since IQVIA Belgium data on drug use is based in prescriptions, it is more likely a delay in GPs prescribing other alternatives.

SIDIAP:

Phenoxymethylpenicillin records were incomplete in SIDIAP and could not be incorporated in this report.

14.3 Interpretation

Population-level drug utilisation

In terms of the broader trends of antibiotic use in Europe, observed seasonality patterns with usage peaks in winter months for antibiotics with indications for frequent winter infections, such as respiratory infections, are well known.(8, 9) When exploring sex and age difference in antibiotics prescription, *Schröder et al.* meta-analysis found a linear trend where antibiotic prescription increases with age, supporting the higher use observed in individuals aged 65 and older from our results.(10) Similarly, this was previously

observed in a past DARWIN EU study on 141 antibiotics using SIDIAP, CPRD, IMASIS and IQVIA Germany, among other databases (EUPAS103381).(11) Additionally, *Schröder et al.* report larger number of prescriptions of cephalosporins and macrolides in women than men in primary care, which align with larger use of antibiotics observed in women from our databases containing only primary care data (i.e., CPRD, IQVIA Belgium and SIDIAP).(10)

COVID-19 drastically altered the patterns of usage of medicines. The decrease in the use of systemic antibiotics during the COVID-19 pandemic in Europe, attributed to infection control measures and reduced contact with the health service, has been previously described.(12) The exceptions observed in the usage of azithromycin and ceftriaxone in IMASIS can be explained by the attempt of finding effective treatments to reduce the risk of death due to COVID-19 in hospitals,(13, 14) as well as, perhaps, to treat secondary bacterial infections such as pneumonia.(15)

On the other hand, peaks observed in amoxicillin, azithromycin, clarithromycin, amoxicillin/clavulanate and phenoxymethylpenicillin use during December 2022 in CPRD were previously documented. Every month, the National Health Service in England publishes anonymised data about the drugs prescribed by GPs, which showed a marked increase in the use of antibacterial drugs during December 2022.(16) Moreover, a study from *Wrenn et al.*, reported a surge of lower respiratory tract infections by group A *Streptococcus* and other bacteria during winter 2022 in the UK, which might have led to the increased incidence of antibiotic use.(17)

Time series modelling

ARIMA models has been widely used for testing temporal trends and seasonality of antibiotic use, and its association with antimicrobial resistance, (18-20) but its usage to forecast use of antibiotics in the upcoming months is limited to one study.(21) *Xie et al.* used ARIMA to predict the trends of antibacterial use rate for a five months period, and concluded that short-term prediction could be applied to reasonably estimate the antibacterials use to facilitate informed decision-making in healthcare management.(21) Considering that uncertainty of predicted values increases with the distance between the time of the last observation and the timepoint of the predicted value, the six-months prediction is more reliable than the 12-month prediction included in the sensitivity analysis section.

Generally, spikes in time series analysis may be explained by exogenous variables not captured in the model. While sARIMA(X) models could not capture these extreme observations, identification of spikes can still be informative, for instance, by providing estimates of increases in demand that can occur naturally.(22)
The divergence between the observed use and forecasted use under 'normal circumstances' can be used to flag potential risks of shortages. In this regard, the peaks observed in the use of amoxicillin, azithromycin, clarithromycin, amoxicillin/clavulanate, phenoxymethylpenicillin in December 2022 in CPRD when compared with the predicted use could be an example. As described in the interpretation of the population-level data, these peaks in December 2022 observed in CPRD were confirmed by UK data on high-level prescribing trends of antibacterial drugs.(16)

14.4 Generalisability

The study comprised all individuals at risk of using the antibiotics of interest present in 6 databases from 5 different European countries. These databases covered a varied range of health care settings to observe the medications that, due to their indications, may be administrated in primary or secondary care: CPRD, IQVIA Belgium and SIDIAP as primary care databases, IMASIS as hospital data, and IQVIA Germany containing primary care and hospital data. While we consider the results representative for the study population in the respective countries, the results should not be generalised to other countries or databases but only reflect the situation in the specific region and setting covered by the respective database.

15 CONCLUSION

sARIMA(X) models fitted to time series data on prescription rates enabled forecasting future incidence rates. The models produced more accurate predictions when adding an indicator term to account for the impact of the COVID-19 pandemic between January 2020 and December 2021 (i.e., sARIMAX models). Since the selected databases are representative for the study population in the respective countries, the fitted models can be used to anticipate the incidence rates of antibiotic use in the four European countries of the study. The observation of drug usage beyond the sARIMAX-based forecast could be useful for the management of potential shortages. The forecasted values for IMASIS, IQVIA Belgium and IQVIA Germany databases need to be interpreted with caution, as the forecasted usage may be slightly overestimated.

16 REFERENCES

1. Aronson JK, Heneghan C, Ferner RE. Drug shortages. Part 1. Definitions and harms. Br J Clin Pharmacol. 2023;89(10):2950-6.
2. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.
3. Reyes C, Pistillo A, Fernandez-Bertolin S, Recalde M, Roel E, Puente D, et al. Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study. BMJ Open. 2021;11(12):e057632.
4. Rathmann W, Bongaerts B, Carius HJ, Kruppert S, Kostev K. Basic characteristics and representativeness of the German Disease Analyzer database. Int J Clin Pharmacol Ther. 2018;56(10):459-66.
5. Recalde M, Rodriguez C, Burn E, Far M, Garcia D, Carrere-Molina J, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). Int J Epidemiol. 2022;51(6):e324-e36.
6. Raventos B, Fernandez-Bertolin S, Aragon M, Voss EA, Blacketer C, Mendez-Boo L, et al. Transforming the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model and Its Use for COVID-19 Research. Clin Epidemiol. 2023;15:969-86.
7. BCFI. GECOMMENTARIEERD GENEESMIDDELEN REPERTORIUM - 11.1.1.1. Penicillines [Available from: https://www.bcfi.be/nl/chapters/12?frag=8902470.
8. Goossens H, Ferech M, Vander Stichele R, Elseviers M, Group EP. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet. 2005;365(9459):579-87.
9. Elseviers MM, Ferech M, Vander Stichele RH, Goossens H, group Ep. Antibiotic use in ambulatory care in Europe (ESAC data 1997-2002): trends, regional differences and seasonal fluctuations. Pharmacoepidemiol Drug Saf. 2007;16(1):115-23.
10. Schroder W, Sommer H, Gladstone BP, Foschi F, Hellman J, Evengard B, et al. Gender differences in antibiotic prescribing in the community: a systematic review and meta-analysis. J Antimicrob Chemother. 2016;71(7):1800-6.

11. EMA | RWD Catalogues. DARWIN EU® DUS of Antibiotics in the 'Watch' category of the WHO AWaRe classification of antibiotics for evaluation and monitoring of use (EUPAS103381) [Available from: https://catalogues.ema.europa.eu/node/3659/.
12. Blix HS, Hoye S. Use of antibiotics during the COVID-19 pandemic. Tidsskr Nor Laegeforen. 2021;141(4).
13. Popp M, Stegemann M, Riemer M, Metzendorf MI, Romero CS, Mikolajewska A, et al. Antibiotics for the treatment of COVID-19. Cochrane Database Syst Rev. 2021;10(10):CD015025.
14. Thapa B, Pathak SB, Jha N, Sijapati MJ, Shankar PR. Antibiotics Use in Hospitalised COVID-19 Patients in a Tertiary Care Centre: A Descriptive Cross-sectional Study. JNMA J Nepal Med Assoc. 2022;60(251):625-30.
15. DeVoe C, Segal MR, Wang L, Stanley K, Madera S, Fan J, et al. Increased rates of secondary bacterial infections, including Enterococcus bacteremia, in patients hospitalized with coronavirus disease 2019 (COVID-19). Infect Control Hosp Epidemiol. 2022;43(10):1416-23.
16. OpenPrescribing.net, Bennett-Institute-for-Applied Data Science, University of Oxford, 2024. 5.1: Antibacterial drugs [Available from: https://openprescribing.net/bnf/0501/.
17. Wrenn K, Blomquist PB, Inzoungou-Massanga C, Olufon O, Guy RL, Hatziioanou D, et al. Surge of lower respiratory tract group A streptococcal infections in England in winter 2022: epidemiology and clinical profile. Lancet. 2023;402 Suppl 1:S93.
18. Martinez EP, van Rosmalen J, Bustillos R, Natsch S, Mouton JW, Verbon A. Trends, seasonality and the association between outpatient antibiotic use and antimicrobial resistance among urinary bacteria in the Netherlands. J Antimicrob Chemother. 2020;75(8):2314-25.
19. Trinh NTH, Chalumeau M, Bruckner TA, Levy C, Bessou A, Milic D, et al. Monitoring outpatient antibiotic utilization using reimbursement and retail sales data: a population-based comparison in France, 2012-17. J Antimicrob Chemother. 2021;76(9):2446-52.
20. Zeng S, Xu Z, Wang X, Liu W, Qian L, Chen X, et al. Time series analysis of antibacterial usage and bacterial resistance in China: observations from a tertiary hospital from 2014 to 2018. Infect Drug Resist. 2019;12:2683-91.
21. Xie J, Kanghuai Z, Cai Y, Wang N, Wang H, Zhao Y, et al. Application of ARIMA model in monitoring the use rate of antibiotics in outpatients in 2010 - 2018[P SEP]Int J Clin Pharmacol Ther. 2020;58(5):282-8.
22. Goin DE, Ahern J. Identification of Spikes in Time Series. Epidemiologic Methods. 2019;8(1).

17 ANNEXES

Appendix I: List of Stand-Alone documents

Medication concept sets

Concepts sets whose administration rote was non-systemic (i.e., topical) were previously removed from the analysis. Full list of concept sets (total of 45,247) for the 11 antibiotics of interest included in this study is available at:

https://github.com/darwin-eu-studies/P2-C1-012-DrugShortages/tree/main/Diagnostics/ConceptSets