Study Protocol

P5-C1-019

DARWIN EU® - Characterising microbiological pathogen and antibiotic resistance data in DARWIN EU® Data Sources *a priori* mapped to the OMOP CDM

13/10/2026

Version 4.0

Authors: Marzyeh Amini, Guido van Leeuwen, Cesar Barboza, Cristiana Di Tullio, Natasha Yefimenko, Katia Verhamme

Public

This document represents the views of the DARWIN EU® Coordination Centre only and cannot be interpreted as reflecting those of the European Medicines Agency or the European Medicines Regulatory Network.

CONTENTS

LIST OF ABBREVIATIONS **5**
1. DESCRIPTION OF THE STUDY TEAM **7**
2. ABSTRACT **8**
3. AMENDMENTS AND UPDATES **10**
4. MILESTONES **10**
5. RATIONALE AND BACKGROUND **10**
6. RESEARCH QUESTION AND OBJECTIVES **10**
7. RESEARCH METHODS **11**
7.1. Study design 11
Figure 1. Graphical depiction of the study design for Objective 1: assessment of annual rates of recorded bacterial pathogen episodes. 11
Figure 2. Graphical depiction of the study design for Objectives 2a and 2b: characterisation of microbiological data and individuals within bacterial pathogen episodes. 12
Figure 3. Included observation time for the denominator population. 13
7.2. Study population with inclusion and exclusion criteria 13
7.3. Study setting and data sources 14
Table 1. Data sources 14
7.4. Study period 15
7.5. Variables 15
7.5.1. Outcome 15
7.5.2. Covariates, including confounders, effect modifiers, and other variables 16
7.6. Study size 17
7.7. Analysis 17
7.7.1. Federated network analyses 17
7.7.2. Data privacy protection 17
7.7.3. Statistical model specification and assumptions of the analytical approach considered 18
Objective 1 18
Objective 2 18
7.7.4. Output 19
MOCK Figure 1. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, presented by participating data source. 20
MOCK Figure 2. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, stratified by sex and presented by participating data source. 20
MOCK Figure 3. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, stratified by age group and presented by participating data source. 21
MOCK Table 1. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes, presented by participating data source. 21
MOCK Table 2. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes stratified by sex, presented by participating data source. 23
MOCK Table 3. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes stratified by age group, presented by participating data source. 26
MOCK Figure 4. Annual trends in microbiological test recording by test category and participating data source, 2015–2025. 30
MOCK Table 4a. Characteristics associated with pre-selected bacterial pathogen episodes identified through clinical records only, presented by participating data source. 30

MOCK Table 4b. Characteristics associated with pre-selected bacterial pathogen episodes identified through microbiological records only, presented by participating data source. 32
Mock Supplementary Table S1. Annual number of recorded microbiological tests by test category and calendar year in [data source name]. 34
7.8. Evidence synthesis 35

8. STRENGTHS AND LIMITATIONS 35

9. REFERENCES 37

10. ANNEXES 38
ANNEX I. Description of data sources 38
Croatia: Croatian National Public Health Information System (NAJS) 38
Estonia: Estonian Biobank (EBB) 39
Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS) 41
France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux) 42
Spain: Institut Municipal Assistència Sanitària Information System (IMASIS) 43
The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD) 44
The United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR) 46
ANNEX II. Fitness for use assessment 48
ANNEX III. List of stand-alone documents 50
Table S1. Preliminary list of OMOP concept IDs for pathogens identification definitions. 50
Table S2. Preliminary list of OMOP concept IDs for outcomes definitions. 50
Table S3. List of OMOP concept IDs for indications/infections. 66
Table S4. List of OMOP concept IDs for antibiotics classes of interest. 71
ANNEX IV. ENCePP checklist for study protocols 72

Study title	DARWIN EU® - Characterising microbiological pathogen and antibiotic resistance data in DARWIN EU® data sources *a priori* mapped to the OMOP CDM.
Protocol version	V4.0
Date	13/10/2026
EUPAS number	EUPAS1000001118
Research question and objectives	Objectives 1. To estimate the annual rate of recorded pre-selected bacterial pathogen episodes over time, overall and by age and sex. 2. To characterise microbiological data associated with episodes of the pre-selected bacterial pathogens in terms of: a. The proportion of pathogen episodes with recorded Gram staining, microbiological culture, rapid antigen testing, molecular testing, phenotypic characteristics (e.g. haemolysis patterns, catalase, coagulase, oxidase, indole, urease, and motility), and antimicrobial susceptibility and resistance information, overall and by age and sex; b. The characteristics associated with pathogen episodes, including age, sex, type of infection, and prior and post antibiotic use in relation to the pathogen episode index date.
Countries of study	Croatia, Estonia, Finland, France, Spain, the United Kingdom, the United States
Authors	Marzyeh Amini m.amini@darwin-eu.org Guido van Leeuwen G.vanleeuwen@darwin-eu.org Cesar Barboza c.barboza@darwin-eu.org Cristiana Di Tullio c.ditullio@darwin-eu.org Natasha Yefimenko n.yefimenkonosova@darwin-eu.org Katia Verhamme k.verhamme@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
AMR	Antimicrobial resistance
ATC	Anatomical Therapeutic Chemical
AWaRe	Access, Watch, and Reserve
CC	Coordination Centre
CDM	Common Data Model
CDW Bordeaux	Clinical Data Warehouse of Bordeaux University Hospital
CHUBX	Centre Hospitalier Universitaire de Bordeaux
CIPH	Croatian Institute of Public Health
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DTZ	Data Transfer Zone
EBB	Estonian Biobank
ECDC	European Centre for Disease Prevention and Control
EHR	Electronic Health Records
EMA	European Medicines Agency
EMR	Electronic Medical Record
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EARS-Net	European Antimicrobial Resistance Surveillance Network
ETL	Extract, Transform and Load
EU	European Union
EUCAST	European Committee on Antimicrobial Susceptibility Testing
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-HUS	Hospital District of Helsinki and Uusimaa
GLASS	Global Antimicrobial Resistance and Use Surveillance System
GP	General Practitioner
HMA	Heads of Medicines Agencies
IMASIS	Institut Municipal Assistència Sanitària Information System
IQR	Interquartile range
IRB	Institutional Review Board
IQVIA US - AmbEMR	IQVIA US - Ambulatory EMR
LOINC	Logical Observation Identifiers Names and Codes
NAJS	Croatian National Public Health Information System
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PSMar	Consorci Mar Parc de Salut Barcelona

Acronyms/terms	Description
R	R programming language
RxNorm	Medical prescription normalised
SD	Standard deviation
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms
TARTU	University of Tartu
US	United States
WHO	World Health Organization

1. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Marzyeh Amini Guido van Leeuwen	Erasmus MC
Data Scientist	Cesar Barboza Cristiana Di Tullio Ross Williams	Erasmus MC
Clinical Domain Expert	Katia Verhamme	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
Croatian National Public Health Information System (NAJS)	Antea Jezidžić Marko Čavlina Karlo Pintaric Anamaria Jurcevic Jakov Vukovic	Croatian Institute of Public Health (CIPH)
Estonian Biobank (EBB)	Marek Oja Raivo Kolde Ami Sild	University of Tartu (TARTU)
Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)	Marianna Niemi Johanna Niklander Eric Fey Kimmo Porkka	HUS - Helsinki University Hospital, Hospital District of Helsinki and Uusimaa
Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)	Romain Griffer Guillaume Verdy	Centre Hospitalier Universitaire de Bordeaux (CHUBX)
Institut Municipal Assistència Sanitària Information System (IMASIS)	Juan Manuel Ramírez-Anguita Carla Arxé Angela Leis Miguel-Angel Mayer	Consorci Mar Parc de Salut Barcelona (PSMar)
Clinical Practice Research Datalink GOLD (CPRD GOLD)	Antonella Delmestri	Clinical Practice Research Datalink (CPRD)
IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)	Dina Vojinovic Ellen Gerritsen Akram Sharim Mendez Rangel Gargi Jadhav	IQVIA

*Data partners do not have an investigator role. The role of the Data Partners is to execute analytical code and review and approve outputs using their expert knowledge about their respective data sources.

2. ABSTRACT

Title

DARWIN EU® - Characterising microbiological pathogen and antibiotic resistance data in DARWIN EU® data sources *a priori* mapped to the OMOP CDM.

Rationale and background

Antimicrobial resistance (AMR) is a major global public health threat associated with increased morbidity, mortality, and healthcare burden. High-quality Real-World AMR research requires microbiological laboratory data linked to longitudinal healthcare records; however, the availability and granularity of microbiological pathogen and antimicrobial resistance data within the DARWIN EU® Network of data sources mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) have not been systematically characterised.

Research question and objectives

Objectives

1. To estimate the annual rate of recorded pre-selected bacterial pathogen episodes over time, overall and by age and sex.
2. To characterise microbiological data associated with episodes of the pre-selected bacterial pathogens in terms of:
   a. The proportion of pathogen episodes with recorded Gram staining, microbiological culture, rapid antigen testing, molecular testing, phenotypic characteristics (e.g. haemolysis patterns, catalase, coagulase, oxidase, indole, urease, and motility), and antimicrobial susceptibility and resistance information, overall and by age and sex;
   b. The characteristics associated with pathogen episodes, including age, sex, type of infection, and prior and post antibiotic use in relation to the pathogen episode index date.

Methods

Study design

Retrospective descriptive observational study.

Population

The study population will include eligible individuals contributing data between 2015 and 2025, with ≥365 days of prior observation for Objective 1, except for hospital-based data sources where observation may be limited to the duration of hospital encounters. For Objective 2, ≥7 days of observation before and after the pathogen episode index date will be required, with this requirement relaxed for hospital-based data sources.

Variables

Outcome:

Recorded identification of pre-selected bacterial pathogens and the availability and granularity of associated microbiological data, including antimicrobial susceptibility and resistance information.

Relevant covariates:

Sex, age group (0–5, 6–11, 12–18, 19–64, and ≥65 years), calendar year, pre-defined infections/indications for antibiotic use, and prior and post antibiotic use.

Data sources

Croatian National Public Health Information System (NAJS), Estonian Biobank (EBB), Hospital District of Helsinki and Uusimaa (FinOMOP-HUS), Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux), Institut Municipal Assistència Sanitària Information System (IMASIS), Clinical Practice Research Datalink GOLD (CPRD GOLD), IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR).

Statistical analysis

Annual rates of recorded bacterial pathogen episodes will be estimated per 100,000 person-years, overall and stratified by sex and age group. Microbiological data availability and granularity will be summarised using descriptive statistics, overall and stratified by age group and sex. Annual trends in microbiological test recording will also be assessed by test category. Characteristics associated with pathogen episodes will be summarised by pathogen identification source. Analyses will be conducted separately for each participating data source. A minimum cell count of 5 will be applied when reporting results, with smaller counts reported as “<5” and zero counts as “0”.

3. AMENDMENTS AND UPDATES

None.

4. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	September 2026
Creation of Analytical code	October 2026
Execution of Analytical Code on the data	November 2026
Draft Study Report	December 2026
Final Study Report	January 2027

*Planned dates are dependent on obtaining approvals from the internal review boards, execution of the code by the data sources and deliverables review rounds.

5. RATIONALE AND BACKGROUND

Antimicrobial resistance (AMR) is a major global public health threat that undermines the effective prevention and treatment of bacterial infections. The World Health Organization (WHO) recognises AMR as one of the top ten global public health threats, largely driven by the inappropriate use of antimicrobials in human medicine, animal health, and agriculture.[1] In 2021, bacterial AMR was associated with an estimated 4.7 million deaths worldwide, including 1.14 million deaths directly attributable to antimicrobial-resistant bacterial infections.[2]

In Europe, antimicrobial-resistant infections are estimated to cause more than 35,000 deaths annually and impose a substantial clinical and economic burden on healthcare systems.[3, 4] Resistance reduces the effectiveness of antibiotic treatment, leading to prolonged illness, increased risk of complications, longer hospital stays, and higher mortality. The continued emergence and spread of multidrug-resistant pathogens, including *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Acinetobacter* species, *Staphylococcus aureus*, and *Enterococcus* spp., underscore the importance of antimicrobial stewardship and continuous surveillance of resistance patterns.[5, 6]

High-quality epidemiological studies on AMR require access to microbiological laboratory data that can be linked to longitudinal patient healthcare records, including information on pathogen identification and antimicrobial susceptibility testing. Although the DARWIN EU® Network provides access to routinely collected healthcare data across multiple European countries, the availability, completeness, and granularity of microbiological pathogen and antimicrobial resistance data have not yet been systematically characterised. Understanding these data is essential to assess their suitability for future Real-World Evidence studies on infectious diseases, antimicrobial resistance, and antibiotic utilisation.

6. RESEARCH QUESTION AND OBJECTIVES

Objectives

1. To estimate the annual rate of recorded pre-selected bacterial pathogen episodes over time, overall and by age and sex.
2. To characterise microbiological data associated with episodes of the pre-selected bacterial pathogens in terms of:
    a. The proportion of pathogen episodes with recorded Gram staining, microbiological culture, rapid antigen testing, molecular testing, phenotypic characteristics (e.g. haemolysis patterns, catalase,

coagulase, oxidase, indole, urease, and motility), and antimicrobial susceptibility and resistance information, overall and by age and sex;

b. The characteristics associated with pathogen episodes, including age, sex, type of infection, and prior and post antibiotic use in relation to the pathogen episode index date.

7. RESEARCH METHODS

7.1. Study design

This will be a retrospective descriptive observational study (**Figures 1** and **2**).

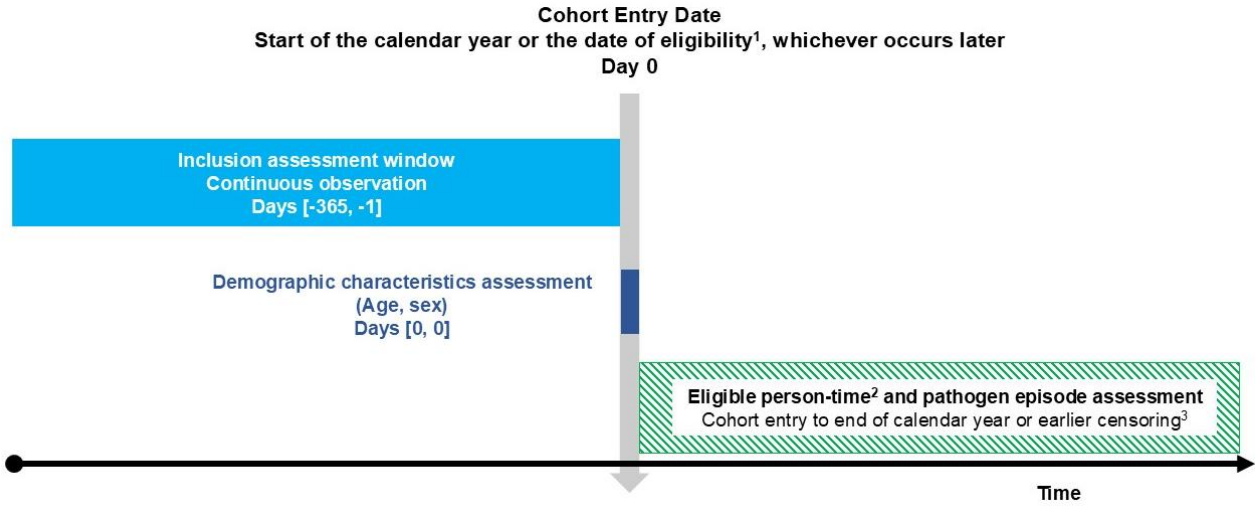


Figure 1. Graphical depiction of the study design for Objective 1: assessment of annual rates of recorded bacterial pathogen episodes.

[1] Eligibility requires ≥365 days of continuous prior observation (except for hospital data). Individuals who meet the eligibility criteria after the start of a calendar year enter on the date the eligibility criteria are first met.
[2] Eligible person-time accrued during this period contributes to the denominator. All pathogen episodes occurring during this period contribute to the numerator; individuals may contribute multiple pathogen episodes.
[3] Eligible person-time and pathogen episodes are assessed from cohort entry until the earliest of death, disenrollment, end of data source availability, end of the respective calendar year, or end of the study period.

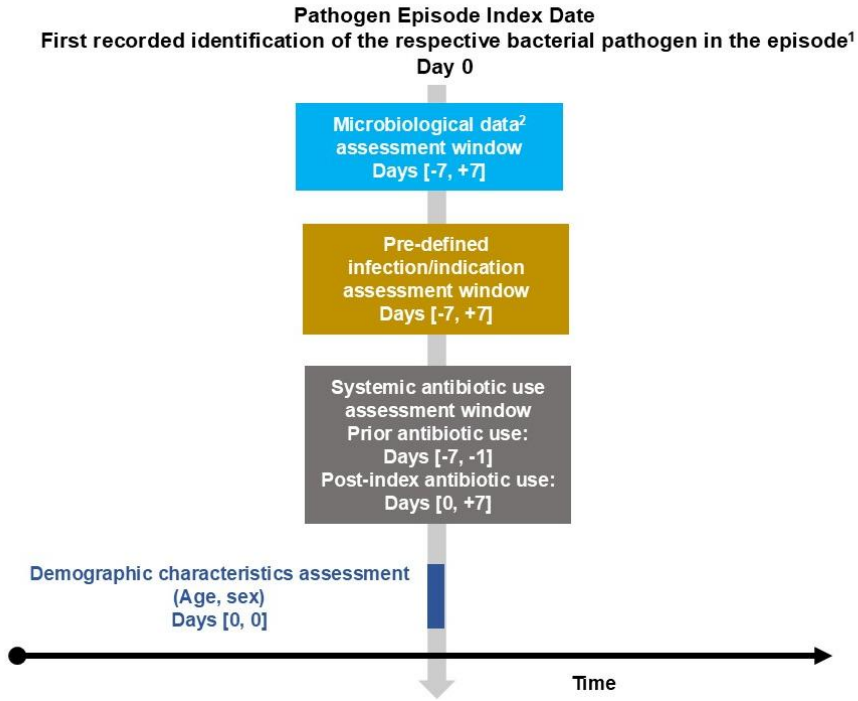


Figure 2. Graphical depiction of the study design for Objectives 2a and 2b: characterisation of microbiological data and individuals within bacterial pathogen episodes.

[1] Pathogen episodes will be defined separately for each pre-selected bacterial pathogen based on relevant records across the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), including condition, observation, and microbiological records. The first recorded identification of the respective pathogen will define the episode index date. Subsequent records of the same pathogen within 30 days of the index date will be considered part of the same episode.
[2] Microbiological data include Gram staining, microbiological culture, rapid antigen testing, molecular/genotyping information, phenotypic characteristics, and antimicrobial susceptibility and resistance results related to the respective bacterial pathogen.

7.2. Follow-up

For Objective 1, within each calendar year, individuals will become eligible to contribute to the denominator population on the latest of (i) the first day of the calendar year, (ii) the study start date (1 January 2015), or (iii) the date on which 365 days of continuous prior observation have been accrued. Because hospital records (i.e. FinOMOP-HUS, CDW Bordeaux, and IMASIS) often cover only the period from the first to the last hospital visit, individuals from hospital data with less than 365 days of prior observation were also included in the analysis. Individuals will remain eligible until the earliest of death, disenrollment, end of data source availability, end of the calendar year, or end of the study period (31 December 2025). All pathogen episodes occurring during eligible person-time will contribute to the numerator. Individuals may contribute multiple pathogen episodes within a calendar year and may contribute person-time to more than one calendar year.

Figure 3 illustrates examples of how person-time will be accrued in the denominator population. Person ID 1 will have sufficient prior history before the study start date and will remain under observation beyond the study end date; therefore, this individual will contribute throughout the study period. Person IDs 2 and 4 will begin contributing once they have accrued 365 days of continuous prior observation. Person ID 3 will stop contributing when the observation period ends. Person ID 5 will have two observation periods and will contribute person-time during both periods once the eligibility criteria have been satisfied.

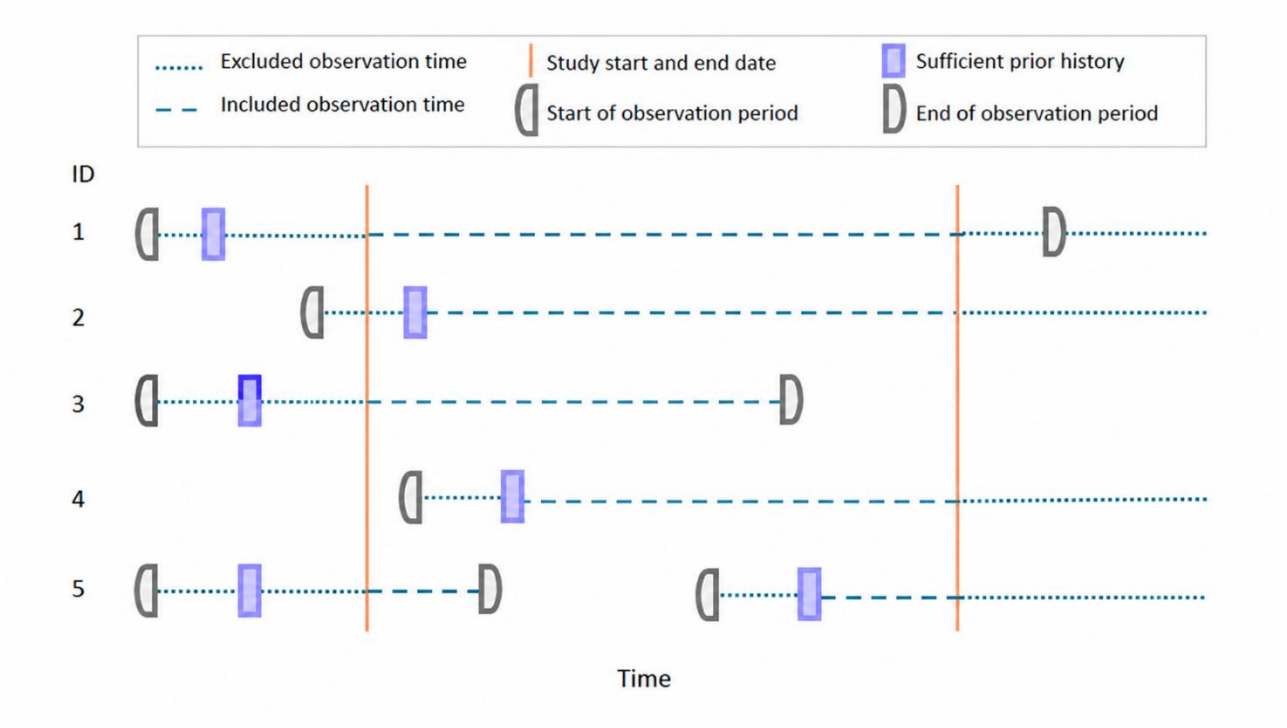


Figure 3. Included observation time for the denominator population.

For Objective 2, pathogen episodes will be created separately for each pre-selected bacterial pathogen to account for repeated microbiological testing and repeated identification of the same pathogen. Pathogen identification will be based on relevant records across the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), including condition, observation, and microbiological records. The first recorded identification of the respective bacterial pathogen will define the index date of an episode. Subsequent records of the same pathogen occurring within 30 days of the index date will be considered part of the same episode. A record of the same pathogen occurring more than 30 days after the index date will define the start of a new episode. Individuals will be required to have continuous observation during the 7 days before and 7 days after the pathogen episode index date. For hospital data (i.e. FinOMOP-HUS, CDW Bordeaux, and IMASIS), where observation may be limited to the duration of the hospital encounter, this requirement will not apply; information within the specified assessment windows will instead be assessed based on the available observation period.

7.2. Study population with inclusion and exclusion criteria

Objective 1

- Minimum 365 days of continuous observation before becoming eligible within each calendar year (except for hospital data).
- Contribution of observation during the study period (01/01/2015 to 31/12/2025).

For each calendar year, eligible individuals with at least one recorded identification of a pre-selected bacterial pathogen will contribute to the numerator for the respective pathogen. Individuals may contribute to more than one calendar year if they continue to meet the eligibility criteria.

Objective 2

Pathogen episodes will be eligible for inclusion if:

- The episode index date occurs during the study period (01/01/2015 to 31/12/2025); and
- The episode relates to one of the pre-selected bacterial pathogens; and
- The individual has continuous observation during the 7 days before and 7 days after the episode index date, except for hospital data.

7.3. Study setting and data sources

This study will use routinely collected healthcare data from 7 participating data sources within the DARWIN EU® Network of Data Partners across six European countries and one from the United States (US). All data have been mapped *a priori* to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Croatia	NAJS	Primary care, outpatient specialist care, and inpatient care	Registries	4.3M	01/01/2014–26/02/2026	Objectives 1 and 2
Estonia	EBB	Primary and secondary care	Biobank	212K	01/01/2004–28/01/2026	Objectives 1 and 2
Finland	FinOMOP-HUS	secondary care – specialists, hospital inpatient care	EHR	810K	01/08/1998–23/02/2026	Objectives 1 and 2
France	CDW Bordeaux	Secondary care	EHR, Claims, Biobank	256K	01/01/2010–01/03/2026	Objectives 1 and 2
Spain	IMASIS	Secondary specialists care	EHR	170K	01/01/1990–20/03/2026	Objectives 1 and 2
The United Kingdom	CPRD Gold	primary care – GPs, secondary care – specialists	EHR	2.4M	01/10/1987–01/03/2026	Objectives 1 and 2
The United States	IQVIA US - AmbEMR	primary care – GPs, secondary care – specialists	EHR	9.1M	01/02/2006–30/11/2025	Objectives 1 and 2

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; EHR=Electronic Health Records; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; GP=General Practitioner; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; K=Thousands; M=Millions; NAJS=Croatian National Public Health Information System.

Data sources selection

As the primary aim of this study is to characterise microbiological pathogen and antimicrobial resistance data *a priori* mapped to the OMOP CDM, participating data sources were selected based on the availability of relevant and already mapped pathogen and microbiological information. Other DARWIN EU® data sources may have relevant microbiological information available in their source data that is not mapped to the OMOP CDM. Information collected through the data partner feasibility questionnaire will be summarised descriptively in the study report to inform potential future mapping and research. The selected data sources meet the DARWIN EU® quality requirements for data quality, timeliness, and representativeness for descriptive epidemiological studies and provide coverage across different countries and healthcare settings (**Annex II**).

7.4. Study period

The study period is from 01/01/2015 to 31/12/2025 or the most recent data available for each contributing data source.

7.5. Variables

7.5.1. Outcome

The pre-selected bacterial pathogens, based on those included in the European Antimicrobial Resistance Surveillance Network (EARS-Net)[4], are *Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, Acinetobacter* species, *Streptococcus pneumoniae, Staphylococcus aureus, Enterococcus faecalis,* and *Enterococcus faecium*. Pathogen identification will be based on relevant records across the OMOP CDM, including condition, observation, and microbiological records. Microbiological information associated with these pathogen episodes will be characterised separately.

For Objective 2a, the frequency of recording of the following microbiological data related to the pre-selected bacterial pathogens will be assessed:

- Gram staining;
- Microbiological culture;
- Rapid antigen testing;
- Molecular testing;
- Phenotypic characteristics, including biochemical identification/profile, catalase, coagulase, oxidase, haemolysis, indole, urease, and motility, where available; and
- Antibiogram or antimicrobial susceptibility/resistance result for an identified microorganism.

Microbiological test and test-related records occurring within 7 days before or after the pathogen episode index date (Days -7 to +7) will be considered associated with that pathogen episode. Where assessment windows overlap across episodes of different pre-selected bacterial pathogens, a microbiological record may contribute to more than one pathogen episode if it falls within the specified assessment window for each episode. No further assignment to a single pathogen episode will be made.

The concept sets used to identify the bacterial pathogens and microbiological outcomes of interest are described in **Annex IV, Tables S1–S2**. These concept sets will use standard OMOP concept IDs from relevant vocabularies, including Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and Logical Observation Identifiers Names and Codes (LOINC), as appropriate for the corresponding OMOP CDM domains.

7.5.2. Covariates, including confounders, effect modifiers, and other variables

- Sex
    - Male/female/unknown: Unknown sex will be retained as a separate category in sex-stratified analyses. Percentages will be calculated using all eligible individuals or pathogen episodes, as applicable, including those with unknown sex, as the denominator;
- Age at the pathogen episode index date, assessed as a continuous variable and categorised as:
    - Children (0–5, 6–11, and 12–18 years);
    - Adults (19–64 years);
    - Older adults (≥65 years);
- Calendar year of cohort entry
- Pre-defined indications for antibiotic use, assessed within a ±7-day window around the pathogen episode index date, will include:
    - Eye infections;
    - Otitis media;
    - Sinusitis;
    - Bronchitis;
    - Pneumonia;
    - Endocarditis;
    - Acute infectious diarrhoea/gastroenteritis;
    - Intra-abdominal infections;
    - Urinary tract infections;
    - Other genito-urinary infections;
    - Meningitis and encephalitis;
    - Sepsis or septic shock;
    - Febrile neutropenia;
    - Osteomyelitis;
    - Septic arthritis;
    - Skin/soft tissue infections;
    - Localized acute bacterial lymphadenitis;

These indications will be identified based on the presence of SNOMED disease codes. The list of codes for identifying the conditions of interest can be found in **Annex IV, Table S3**.

- Prior and post systemic antibiotic use, assessed during the 7 days before (Days -7 to -1) and 7 days from the pathogen episode index date (Days 0 to +7), respectively, according to the pre-defined antibiotic classes.
    - Penicillins;
    - Beta-lactam/beta-lactamase inhibitor combinations;

- Cephalosporins;
- Carbapenems;
- Fluoroquinolones;
- Aminoglycosides;
- Glycopeptides;
- Polymyxins;
- Oxazolidinones;
- Lipopeptides;

The antibiotic classes of interest were selected based on their clinical relevance to the pre-selected bacterial pathogens and their importance in antimicrobial resistance surveillance. The selection includes antibiotic classes for which resistance is of particular public health importance, as reflected in the antimicrobial agent combinations monitored by the EARS-Net, the WHO Bacterial Priority Pathogens List, and the WHO Access, Watch, and Reserve (AWaRe) framework.[4, 7, 8] A list of concept IDs for the antibiotic classes of interest is provided in **Annex IV, Table S4**. Non-systemic antibiotics will be excluded in this study.

The concept sets will be developed and refined, following the standard DARWIN EU® phenotyping process, which includes the review of the code lists by clinical experts and evaluation of the concept sets following their execution in the participating data sources.

7.6. Study size

No sample size calculation will be performed, as this is a descriptive study that does not aim to test a specific hypothesis. All eligible individuals contributing data to the participating data sources during the study period will be included in the analyses.

7.7. Analysis

7.7.1. Federated network analyses

All analyses will be conducted separately for each data source and will be carried out in a federated manner, allowing analyses to be run locally without sharing individuals' data.

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed (see **Annex III**), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

7.7.2. Data privacy protection

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in a secure online repository of the Data Transfer Zone (DTZ). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources will be checked, after which they are made available to the team, and the Study Dissemination Phase can start. All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

7.7.3. Statistical model specification and assumptions of the analytical approach considered

Objective 1

Objective 1 analyses will be performed using the DARWIN EU® *IncidencePrevalence* R package.[9] Annual rates of recorded bacterial pathogen episodes will be estimated overall and stratified by age group and sex, and presented by calendar year and participating data source.

For each calendar year, the annual episode rate will be calculated as the number of pre-selected bacterial pathogen episodes occurring during that year divided by the total eligible person-time contributed during that year. Individuals may contribute more than one pathogen episode within a calendar year.

Annual episode rates will be reported per 100,000 person-years, together with 95% confidence intervals. Analyses will be conducted separately by data source. For each data source, only periods with available data will contribute to the longitudinal analyses. Where data availability ends partway through a calendar year, eligible person-time and pathogen episodes will be included up to the end of data availability. No extrapolation will be performed beyond the available data period.

Objective 2

Objective 2 analyses will be performed using the DARWIN EU® *CohortCharacteristics* (https://cran.r-project.org/web/packages/CohortCharacteristics/index.html) R package.[10] Descriptive statistics will be used to characterise microbiological data related to the pre-selected bacterial pathogens, presented by participating data source.

For Objective 2a, the availability and granularity of microbiological data will be summarised using frequencies and percentages, overall and stratified by age group and sex. For each pre-selected bacterial pathogen, percentages will be calculated using all eligible pathogen episodes as the denominator. The number and proportion of pathogen episodes identified through clinical records only, microbiological records only, or both clinical and microbiological records will also be reported. Clinical records will include pathogen-identifying records from the Condition and Observation domains, while microbiological records will include pathogen-identifying records from the Measurement domain.

To assess the temporal availability of microbiological data, annual numbers of recorded tests will be summarised by data source and by test category (Gram staining, culture, rapid antigen testing, molecular testing, phenotypic characteristics, and antimicrobial susceptibility and resistance testing). All relevant microbiological records will be included, regardless of their association with pre-selected bacterial pathogen episodes, without age or sex stratification. Where available, information from the data sources on the earliest year covered by OMOP CDM mapping and major mapping changes will be considered when interpreting temporal trends.

For Objective 2b, characteristics related to each pathogen episode will be summarised using descriptive statistics, stratified by whether the pathogen was identified through clinical records or microbiological records. Demographic characteristics will be assessed at the pathogen episode index date. Pre-defined infection/indication diagnoses will be assessed within 7 days before and 7 days after the episode index date (Days -7 to +7). Systemic antibiotic use will be assessed separately during the 7 days before the episode index date (Days -7 to -1) and during the 7 days from the episode index date (Days 0 to +7). For hospital data, where observation may be limited to the duration of the hospital encounter, information will be assessed within the specified assessment windows based on the available observation period; therefore, results should be interpreted considering potential incomplete capture of information within these windows. Age, as a continuous variable, will be summarised using the mean and standard deviation (SD) and the median and interquartile range (IQR). Age group, sex, pre-defined infection/indication diagnoses, and systemic antibiotic use by antibiotic class will be summarised using frequencies and percentages.

7.7.4. Output

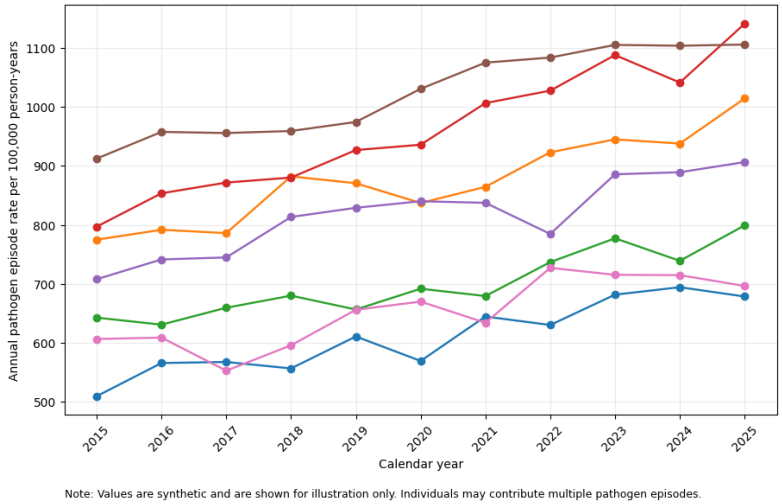
The output will consist of a PDF report, including an executive summary and the following tables and figures.

- Figure 1. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, presented by participating data source.
- Figure 2. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, stratified by sex and presented by participating data source.
- Figure 3. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, stratified by age group and presented by participating data source.
- Table 1. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes, presented by participating data source.
- Table 2. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes stratified by sex, presented by participating data source.
- Table 3. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes stratified by age group, presented by participating data source.
- Figure 4. Annual trends in microbiological test recording by test category and participating data source, 2015–2025.
- Table 4a. Characteristics associated with pre-selected bacterial pathogen episodes identified through clinical records only, presented by participating data source.
- Table 4b. Characteristics associated with pre-selected bacterial pathogen episodes identified through microbiological records only, presented by participating data source.
- Supplementary Table S1. Annual number of recorded microbiological tests by test category and calendar year in [data source name].

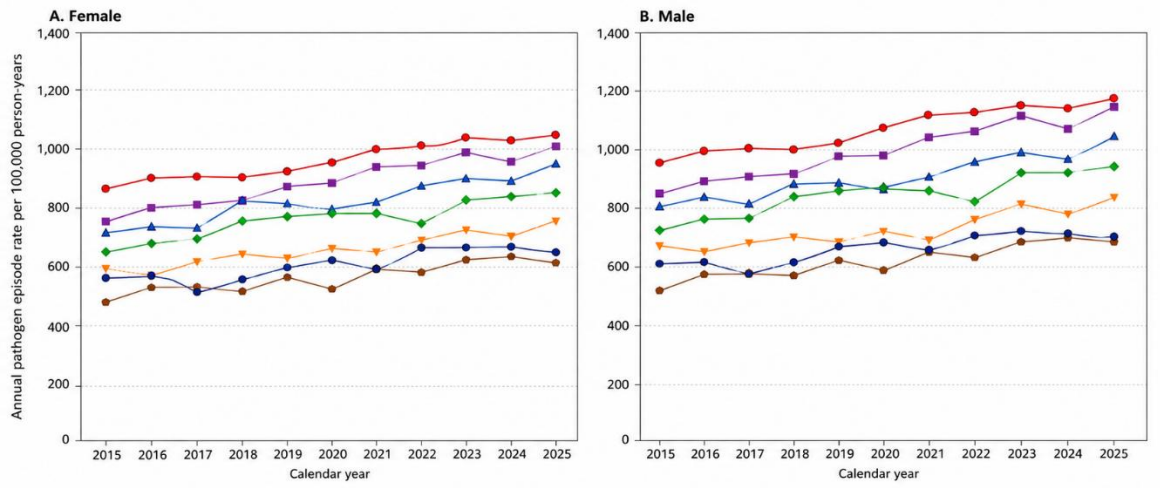
An interactive dashboard will be generated by incorporating all the results (tables and figures) included in the PDF report mentioned above.

Objective 1

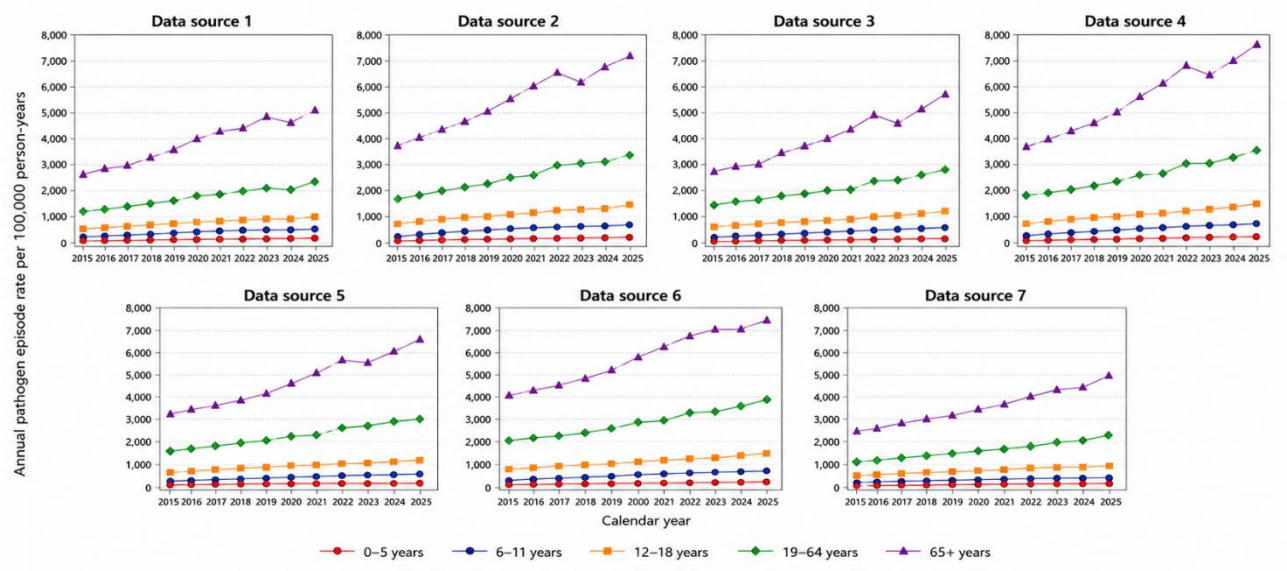
The following plots will be generated for each pre-selected bacterial pathogen. Corresponding annual rate estimates will be presented in tabular format in the Shiny app.



MOCK Figure 1. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, presented by participating data source.



MOCK Figure 2. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, stratified by sex and presented by participating data source.



MOCK Figure 3. Annual rates of recorded pre-selected bacterial pathogen episodes per 100,000 person-years, stratified by age group and presented by participating data source.

Objective 2

MOCK Table 1. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes, presented by participating data source.

Pathogen and microbiological characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
Escherichia coli pathogen episodes; N							
Escherichia coli pathogen episodes identified through clinical records; N (%)							
Escherichia coli pathogen episodes identified through microbiological records; N (%)							
Escherichia coli pathogen episodes identified through both clinical and microbiological records; N (%)							
Microbiological test type related to *Escherichia coli* pathogen episodes; N (%)							
Gram stain							
Culture							
Rapid antigen test							
Molecular test							
Phenotypic characteristics related to *Escherichia coli* pathogen episodes; N (%)							
Catalase							
Coagulase							
Oxidase							

Pathogen and microbiological characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
Haemolysis							
Indole							
Urease							
Motility							
Biochemical identification/profile							
Antimicrobial susceptibility/resistance related to *Escherichia coli* pathogen episodes; N (%)							
Klebsiella pneumoniae pathogen episodes; N							
Klebsiella pneumoniae pathogen episodes identified through clinical records; N (%)							
Klebsiella pneumoniae pathogen episodes identified through microbiological records; N (%)							
Klebsiella pneumoniae pathogen episodes identified through both clinical and microbiological records; N (%)							
Microbiological test type related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
Gram stain							
Culture							
Rapid antigen test							
Molecular test							
Phenotypic characteristics related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
Catalase							
Coagulase							
Oxidase							
Haemolysis							
Indole							
Urease							
Motility							
Biochemical identification/profile							
Antimicrobial susceptibility/resistance related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
Other pathogens….							

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; NAJS=Croatian National Public Health Information System.
Note: Percentages will be calculated using all eligible episodes of the respective bacterial pathogen as the denominator. Clinical records include pathogen-identifying records from the Condition and Observation domains, while microbiological records include pathogen-identifying records from the Measurement domain. Episodes with pathogen-identifying records from both categories will

be classified as both clinical and microbiological records. Microbiological variables will be assessed within 7 days before and 7 days after the pathogen episode index date. More than one microbiological test or characteristic may be recorded within an episode; therefore, percentages across microbiological test and characteristic categories may not sum to 100%.

MOCK Table 2. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes stratified by sex, presented by participating data source.

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
Female	*Escherichia coli* pathogen episodes; N							
	Escherichia coli pathogen episodes identified through clinical records; N (%)							
	Escherichia coli pathogen episodes identified through microbiological records; N (%)							
	Escherichia coli pathogen episodes identified through both clinical and microbiological records; N (%)							
	Microbiological test type related to *Escherichia coli* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							
	Phenotypic characteristics related to *Escherichia coli* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Escherichia coli* pathogen episodes; N (%)							
	Klebsiella pneumoniae pathogen episodes; N							
	Klebsiella pneumoniae pathogen episodes identified through clinical records; N (%)							

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Klebsiella pneumoniae pathogen episodes identified through microbiological records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through both clinical and microbiological records; N (%)							
	Microbiological test type related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							
	Phenotypic characteristics related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Other pathogens….							
Male	*Escherichia coli* pathogen episodes; N							
	Escherichia coli pathogen episodes identified through clinical records; N (%)							
	Escherichia coli pathogen episodes identified through microbiological records; N (%)							
	Escherichia coli pathogen episodes identified through both clinical and microbiological records; N (%)							

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Microbiological test type related to *Escherichia coli* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							
	Phenotypic characteristics related to *Escherichia coli* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Escherichia coli* pathogen episodes; N (%)							
	Klebsiella pneumoniae pathogen episodes; N							
	Klebsiella pneumoniae pathogen episodes identified through clinical records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through microbiological records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through both clinical and microbiological records; N (%)							
	Microbiological test type related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Phenotypic characteristics related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Other pathogens...							

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; NAJS=Croatian National Public Health Information System.
Note: Percentages will be calculated using all eligible episodes of the respective bacterial pathogen within each sex category as the denominator. Clinical records include pathogen-identifying records from the Condition and Observation domains, while microbiological records include pathogen-identifying records from the Measurement domain. Episodes with pathogen-identifying records from both categories will be classified as both clinical and microbiological records. Microbiological variables will be assessed within 7 days before and 7 days after the pathogen episode index date. More than one microbiological test or characteristic may be recorded within an episode; therefore, percentages across microbiological test and characteristic categories may not sum to 100%.

MOCK Table 3. Availability and granularity of microbiological data related to pre-selected bacterial pathogen episodes stratified by age group, presented by participating data source.

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
0–5 years	*Escherichia coli* pathogen episodes; N							
	Escherichia coli pathogen episodes identified through clinical records; N (%)							
	Escherichia coli pathogen episodes identified through microbiological records; N (%)							
	Escherichia coli pathogen episodes identified through both clinical and microbiological records; N (%)							

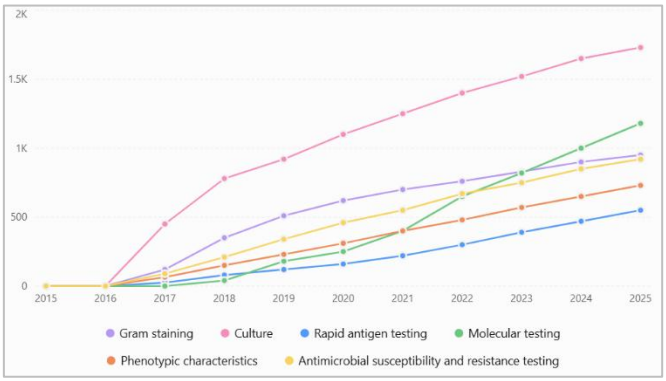
	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Microbiological test type related to *Escherichia coli* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							
	Phenotypic characteristics related to *Escherichia coli* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Escherichia coli* pathogen episodes; N (%)							
	Klebsiella pneumoniae pathogen episodes; N							
	Klebsiella pneumoniae pathogen episodes identified through clinical records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through microbiological records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through both clinical and microbiological records; N (%)							
	Microbiological test type related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Phenotypic characteristics related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Other pathogens...							
6–11 years	*Escherichia coli* pathogen episodes; N							
	Escherichia coli pathogen episodes identified through clinical records; N (%)							
	Escherichia coli pathogen episodes identified through microbiological records; N (%)							
	Escherichia coli pathogen episodes identified through both clinical and microbiological records; N (%)							
	Microbiological test type related to *Escherichia coli* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							
	Phenotypic characteristics related to *Escherichia coli* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Escherichia coli* pathogen episodes; N (%)							
	Klebsiella pneumoniae pathogen episodes; N							
	Klebsiella pneumoniae pathogen episodes identified through clinical records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through microbiological records; N (%)							
	Klebsiella pneumoniae pathogen episodes identified through both clinical and microbiological records; N (%)							
	Microbiological test type related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Gram stain							
	Culture							
	Rapid antigen test							
	Molecular test							
	Phenotypic characteristics related to *Klebsiella pneumoniae* pathogen episodes; N (%)							
	Catalase							
	Coagulase							
	Oxidase							
	Haemolysis							
	Indole							
	Urease							
	Motility							
	Biochemical identification/profile							
	Antimicrobial susceptibility/resistance related to *Klebsiella pneumoniae* pathogen episodes; N (%)							

	Microbiological variable	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
	Other pathogens...							
Other age group								

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; NAJS=Croatian National Public Health Information System.
Note: Percentages will be calculated using all eligible episodes of the respective bacterial pathogen within each age group as the denominator. Clinical records include pathogen-identifying records from the Condition and Observation domains, while microbiological records include pathogen-identifying records from the Measurement domain. Episodes with pathogen-identifying records from both categories will be classified as both clinical and microbiological records. Microbiological variables will be assessed within 7 days before and 7 days after the pathogen episode index date. More than one microbiological test or characteristic may be recorded within an episode; therefore, percentages across microbiological test and characteristic categories may not sum to 100%.



MOCK Figure 4. Annual trends in microbiological test recording by test category and participating data source, 2015–2025.

MOCK Table 4a. Characteristics associated with pre-selected bacterial pathogen episodes identified through clinical records only, presented by participating data source.

Characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
All pathogen episodes; N							
Pathogen episodes identified through clinical records only; N (%)							
Age at episode index date, years							
Mean (SD)							
Median (IQR)							
Age group at episode index date; N (%)							

Characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
0–5 years							
6–11 years							
12–18 years							
19–64 years							
≥65 years							
Sex; N (%)							
Male							
Female							
Unknown							
Pre-defined infection/indication within Days -7 to +7; N (%)							
Eye infections							
Otitis media							
Sinusitis							
Bronchitis							
Pneumonia							
Endocarditis							
Acute infectious diarrhoea/gastroenteritis							
Intra-abdominal infections							
Urinary tract infections							
Other genito-urinary infections							
Meningitis and encephalitis							
Sepsis or septic shock							
Febrile neutropenia							
Osteomyelitis							
Septic arthritis							
Skin/soft tissue infections							
Localized acute bacterial lymphadenitis							
Systemic antibiotic use during Days -7 to -1; N (%)							
Penicillins							
Beta-lactam/beta-lactamase inhibitor combinations							
Cephalosporins							

Characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
Carbapenems							
Fluoroquinolones							
Aminoglycosides							
Glycopeptides							
Polymyxins							
Oxazolidinones							
Lipopeptides							
Systemic antibiotic use during Days 0 to +7; N (%)							
Penicillins							
Beta-lactam/beta-lactamase inhibitor combinations							
Cephalosporins							
Carbapenems							
Fluoroquinolones							
Aminoglycosides							
Glycopeptides							
Polymyxins							
Oxazolidinones							
Lipopeptides							

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; NAJS=Croatian National Public Health Information System.
Note: Percentages will be calculated using pathogen episodes identified through clinical records only within each data source as the denominator. Age and sex will be assessed at the pathogen episode index date. Pre-defined infection/indication diagnoses will be assessed from 7 days before to 7 days after the pathogen episode index date (Days -7 to +7). Systemic antibiotic use will be assessed separately during the 7 days before the pathogen episode index date (Days -7 to -1) and during the 7 days from the pathogen episode index date (Days 0 to +7). Individuals may contribute more than one pathogen episode.

MOCK Table 4b. Characteristics associated with pre-selected bacterial pathogen episodes identified through microbiological records only, presented by participating data source.

Characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
All pathogen episodes; N							
Pathogen episodes identified through clinical records only; N (%)							
Age at episode index date, years							
Mean (SD)							

Characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
Median (IQR)							
Age group at episode index date; N (%)							
0–5 years							
6–11 years							
12–18 years							
19–64 years							
≥65 years							
Sex; N (%)							
Male							
Female							
Unknown							
Pre-defined infection/indication within Days -7 to +7; N (%)							
Eye infections							
Otitis media							
Sinusitis							
Bronchitis							
Pneumonia							
Endocarditis							
Acute infectious diarrhoea/gastroenteritis							
Intra-abdominal infections							
Urinary tract infections							
Other genito-urinary infections							
Meningitis and encephalitis							
Sepsis or septic shock							
Febrile neutropenia							
Osteomyelitis							
Septic arthritis							
Skin/soft tissue infections							
Localized acute bacterial lymphadenitis							
Systemic antibiotic use during Days -7 to -1; N (%)							
Penicillins							

Characteristic	NAJS	EBB	FinOMOP-HUS	CDW Bordeaux	IMASIS	CPRD GOLD	IQVIA US - AmbEMR
Beta-lactam/beta-lactamase inhibitor combinations							
Cephalosporins							
Carbapenems							
Fluoroquinolones							
Aminoglycosides							
Glycopeptides							
Polymyxins							
Oxazolidinones							
Lipopeptides							
Systemic antibiotic use during Days 0 to +7; N (%)							
Penicillins							
Beta-lactam/beta-lactamase inhibitor combinations							
Cephalosporins							
Carbapenems							
Fluoroquinolones							
Aminoglycosides							
Glycopeptides							
Polymyxins							
Oxazolidinones							
Lipopeptides							

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; NAJS=Croatian National Public Health Information System.
Note: Percentages will be calculated using pathogen episodes identified through microbiological records only within each data source as the denominator. Age and sex will be assessed at the pathogen episode index date. Pre-defined infection/indication diagnoses will be assessed from 7 days before to 7 days after the pathogen episode index date (Days -7 to +7). Systemic antibiotic use will be assessed separately during the 7 days before the pathogen episode index date (Days -7 to -1) and during the 7 days from the pathogen episode index date (Days 0 to +7). Individuals may contribute more than one pathogen episode.

Mock Supplementary Table S1. Annual number of recorded microbiological tests by test category and calendar year in [data source name].

Calendar year	Gram staining	Culture	Rapid antigen testing	Molecular testing	Phenotypic characteristics	Antimicrobial susceptibility and resistance testing
2015						
2016						
2017						

Calendar year	Gram staining	Culture	Rapid antigen testing	Molecular testing	Phenotypic characteristics	Antimicrobial susceptibility and resistance testing
2018						
2019						
2020						
2021						
2022						
2023						
2024						
2025						

Note: An individual may contribute multiple records within a calendar year, and different test categories are not mutually exclusive.

7.8. Evidence synthesis

Results from analyses described in **Section 7.8** will be presented separately for each data source. No meta-analysis of results will be conducted.

8. STRENGTHS AND LIMITATIONS

This study is based on routinely collected healthcare data from multiple data sources within the DARWIN EU® Network, including hospital, primary care, and registry data sources from European countries. The inclusion of multiple data sources enables a comprehensive assessment of the availability and recording of microbiological pathogen identification and antimicrobial resistance data across diverse healthcare settings. This will provide insights into the relevance of data sources for future studies with a focus on these outcomes, as well as a better understanding of the limitations of the underlying data and where attention should be paid for increasing data curation, data source linkages, use of free-text, and standardisation (mapping).

However, several limitations should be considered. First, microbiological data and antimicrobial susceptibility results are not uniformly available across participating data sources, and their completeness and granularity may vary considerably. Some data sources may require additional local mapping of microbiological data before analysis, potentially resulting in differences in data completeness across data sources. Importantly, the absence of a microbiological record in the analysed data does not necessarily indicate that testing was not performed, as information may not be captured, linked, mapped, or available in structured fields.

Second, recording practices for microbiological testing, pathogen identification, and antimicrobial susceptibility may differ between healthcare settings and countries because of variations in clinical practice, laboratory information systems, coding standards, and local implementation of the OMOP CDM. These differences may limit direct comparability of estimates across data sources. Third, the absence of a standard OMOP convention for representing detailed microbiological and antimicrobial susceptibility data may lead to heterogeneity in how these data are mapped and interpreted. Finally, temporal changes in laboratory methods, coding practices, and antimicrobial susceptibility interpretation standards, such as updates to European Committee on Antimicrobial Susceptibility Testing (EUCAST) breakpoints, may influence observed trends over time independently of changes in the underlying epidemiology of antimicrobial resistance.

Consequently, the findings should primarily be interpreted as reflecting the availability and recording of microbiological and antimicrobial resistance data within the participating DARWIN EU® Data Sources rather than the true burden of antimicrobial resistance in the underlying populations.

9. REFERENCES

1. World Health Organization. Antimicrobial resistance – Fact sheet. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

2. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet. 2024;404(10459):1199-226.

3. European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-Net) – Annual Epidemiological Report for 2024 [Internet]. Stockholm: ECDC; 2025 [cited 2026 Jul 31]. Available from: https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-eueea-ears-net-annual-epidemiological-report-2024.

4. European Centre for Disease Prevention and Control. European Antimicrobial Resistance Surveillance Network (EARS-Net) [Internet]. Stockholm: ECDC; [cited 2026 Sep 11]. Available from: https://www.ecdc.europa.eu/en/about-us/networks/disease-networks-and-laboratory-networks/ears-net-data.

5. World Health Organization. Global Antimicrobial Resistance and Use Surveillance System (GLASS) [Internet]. Geneva: World Health Organization; [cited 2026 Jul 31]. Available from: https://www.who.int/initiatives/glass.

6. European Centre for Disease Prevention and Control. European Antimicrobial Resistance Surveillance Network (EARS-Net) [Internet]. Stockholm: ECDC; [cited 2026 Jul 31]. Available from: https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/ears-net.

7. World Health Organization. WHO bacterial priority pathogens list, 2024: bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Geneva: WHO; 2024. https://www.who.int/publications/i/item/9789240093461.

8. World Health Organization. AWaRe classification of antibiotics for evaluation and monitoring of use, 2023. Geneva: World Health Organization; 2023. https://www.who.int/publications/i/item/WHO-MHP-HPS-EML-2023.04.

9. Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. Pharmacoepidemiol Drug Saf. 2024;33(1):e5717.

10. OHDSI. CohortCharacteristics: Characterise Cohorts in OMOP Common Data Model [Internet]. R package version. Available from: https://cran.r-project.org/package=CohortCharacteristics.

10. ANNEXES

ANNEX I. Description of data sources

Croatia: Croatian National Public Health Information System (NAJS)

#	Section	Description
1	Data source identification and country	NAJS (Croatian National Public Health Information System) Croatia
2	Data partner information section	Croatian Institute of Public Health (CIPH) Department of Data Science and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. Geographic coverage covers whole Croatia, with various levels of resolution for different registries. Current estimates for the population in Croatia will be available at: https://podaci.dzs.hr/hr/podaci/stanovnistvo/procjena-stanovnistva/ for each year. The total and active person count in the NAJS data is larger than the current population of Croatia. This explained by: a) the person table included deceased and all previously insured people and b) there is no information about insurance ending, c) healthcare is also used by people with dual citizenship from neighbouring countries It is known that a lot of people emigrated (300k-400k) and weren't included in the last population census but still are in the NAJS database. There is also an influx of immigrant workers that are insured and registered but weren't included in the census.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. For both inpatient and outpatient setting diagnoses, medication, procedures, and measurements are captured. The year of availability of information depends on the setting: • 2014-2025 for biochemical lab tests in primary care from EHR patients records (measurements with results) • 2015-2025 for primary care data from EHR patient records (conditions, procedures, and drug prescriptions) • 2015- 2024 for inpatient hospital data from EHR administrative records (conditions, procedures, measurements without results and drug administrations) • 2016-2025 for health risk assessment data entered by GPs (measurements with results - height, weight…) • 2016-2022 for secondary conciliatory care data from EHR administrative records (conditions, procedures, measurements without results and drug administrations) • 2016-2022 for emergency care data from EHR patient records (conditions) • 2017-2025 for hospital records from registry data (conditions and procedures) • 2020-2025 for vaccination data from EHR patient records
5	Data collection process	Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries, and Other. Data is entered by clinicians at healthcare contact, then combined by CIPH into the NAJS database and integrated with registries for public health purposes.
6	General representativeness	The data is collected from the evidence of public health records collected for public health purposes, as the majority of health care in Croatia is public and under single health insurance provider. Personal details are collected to a better extent for insured individuals compared to uninsured patients, who are excluded in the ETL process.
7	Data content / source coding	Medication prescriptions are recorded with ATC codes with an additional 3 digit code denoting the package. Diagnoses with ICD10 codes (Australian modification). Procedures with local source codes. Lab results with local source codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Records from 2017 include insured patients with reliable IDs. Uninsured patients do not have reliable IDs. For example, if a patient changed her status from insured to uninsured, or vice versa, she could be counted several times, as could tracking records from before 2017 and after. By using the unique personal identifier for Croatian citizens, it can be checked and verified.

#	Section	Description
9	Quality control (data source specific)	There is a network of registry personnel (leaders, administrators, coders, sources) working on data coverage and other quality dimensions. An analytical team routinely checks for erroneous entries in hospital records, removing double entries, false dates, and overlapping stays. Entries without enough data or with obviously erroneous dates from primary care analysis are being excluded.
10	Linkage	The national death registry is updated yearly, with one year lag, but the fact of someone's death (just the date) is updated daily, without the cause of death or any other additional details. Primary care is updated weekly and hospital level care monthly. Specific registries are included in NAJS (e.g. diabetes registry), where inclusion criteria vary across these registries.
11	Vital status	NAJS is linked to the national death registry.
12	Limitations	Hospital data is available from 2017 onwards. This is often used as start of data collection, while laboratory and GP data is captured before that (since 2014 and 2015 respectively). Drug duration is often not available, and set to 1 day for administration and 30 days for prescription. Hospital discharge summaries are currently not captured in NAJS. Hospital drug administration data is less reliable than prescription data from primary care, with some drugs (monoclonal antibodies / precision medicine drugs) that require additional approval not being recorded at all.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111155 Website: https://www.hzjz.hr/nacionalni-javnozdravstveni-informacijski-sustav-najs/

Estonia: Estonian Biobank (EBB)

#	Section	Description
1	Data source identification and country	EBB (Estonian Biobank) Estonia
2	Data partner information section	University of Tartu (TARTU) Institute of Computer Science
3	Coverage and timespan	Data collection since: 2004 Extent: Nation-wide. EBB is a nation-wide database containing records from 2004 onwards. Estonian population-based cohort size of 211,800 participants (01/01/2024) aged 18 years and older recruited at GP offices, private practices, and hospitals or in the recruitment offices of the Estonian Genome Center.
4	Healthcare setting / type of data	Primary care – General Practitioner, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Healthcare providers must report the data to national registries. The data includes claims, prescriptions, EHR and two registries - the cause of death and cancer registry. The claims data covers diagnosis, procedures and services performed. Claims data contains information about services paid by the Estonian Health Insurance Fund and not the services paid by the person themself. The prescription data contains information prescribed and dispensed prescription drugs. It does not contain information about the over-the-counter drugs. The EHR data includes diagnosis, results of the procedures and lab measurements and case summaries. The cause of death and cancer registry data is checked by the registry before it is entered to the registry. Also, Estonian Biobank, has collected genetic and questionnaire data when the person gives consent. Questionnaire data includes personal data (e.g. nationality), genealogical data (family history of medical conditions), educational history, lifestyle data (e.g. smoking, alcohol consumption etc).
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Biobank, and Other. Data is retrieved by Estonian Biobank once a year from national registries. The insurance claims are requested from Estonian Health Insurance Fund. The inpatient and outpatient electronic

#	Section	Description
		health records are requested from National Health Information System. The cancer registry and cause of death registry information is requested from The National Institute for Health Development. The data is sent to the national registry by the healthcare providers.
6	General representativeness	The age, sex, and geographical distribution closely reflect those of the Estonian adult population and encompass approximately 20% of adult population. Female participants are over-represented in EBB. Older people tend to participate less frequently, however, all age groups are well represented.
7	Data content / source coding	All participants have undergone a standardized health assessment, including provision of blood samples for purification of DNA, white blood cells, and plasma, and completed a questionnaire covering various health-related topics, such as lifestyle, diet, and clinical diagnoses. Diseases and health problems are recorded as ICD-10 codes and prescribed medicine according to the ATC classification and local package codes. Procedures and services are coded with NOMESCO classifier and local service codes. Lab measurements are coded with local codes and in later years with LOINC codes. In cancer registry also ICD-O-3 codes are used. The data is available with different granularity. For claims data we only now that procedure or service is done. The result of the procedure or measurement is recorded in EHR. Usually the result is in unstructured text (except lab measurements) and needs specific pipelines for extraction.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. There is one national identifier that allows linking together all encounters across databases.
9	Quality control (data source specific)	The quality control procedures in the Estonian Biobank aim to remove the most obvious mistakes in the data, misspellings, impossible dates, duplicates. Before performing the ETL, several problems are fixed on the source data. Since the ETL procedures are used for a number of different datasets (from the same national sources), we have a growing number of pre-processing steps that correspond to the issues we have discovered previously in the data, such as checking for the presence of critical values, harmonizing date and unit of measurement formats, checking the validity of certain entries against classifiers, etc.
10	Linkage	Follow-up data are available via linkage with national health-related registries and via re-examination of participants. Furthermore, electronic health records are updated for phenotypic outcome information every year. The EBB database is regularly linked with national registries, hospital databases, and the databases of the Estonian Health Insurance Fund (EHIF) and the National Health Information System (NHIS).
11	Vital status	Vital status (death date and causes of death) are obtained from the Causes of Death Registry.
12	Limitations	Participation in EBB cohort is voluntary, therefore the biobank does not represent a random sample and could be subject to recruitment bias. Although recruitment was open to everyone, there is a disproportion between ethnic Estonians and ethnic Russians in the biobank, with Estonians being overrepresented.
13	Main references	Milani L, Alver M,Laur S,Reisberg S,Haller T,Aasmets O,Abner E,Alavere H,Allik A,Annilo T,Fischer K,Hofmeister R,Hudjashov G,Jõeloo M,Kals M,Karo-Astover L,Kasela S,Kolde A,Krebs K,Krigul KL,Kronberg J,Kruusmaa K,Kukuškina V,Kõiv K,Lehto K,Leitsalu L,Lind S,Luitva LB,Läll K,Lüll K,Metsalu K,Metspalu M,Mõttus R,Nelis M,Nikopensius T,Nurm M,Nõukas M,Oja M,Org E,Palover M,Palta P,Pankratov V,Pantiukh K,Pervjakova N,Pujol-Gualdo N,Reigo A,Reimann E,Smit S,Rogozina D,Särg D,Taba N,Talvik HA,Teder-Laving M,Tõnisson N,Vaht M,Vainik U,Võsa U,Yelmen B,Esko T,Kolde R,Mägi R,Vilo J,Laisk T,Metspalu A "The Estonian Biobank's journey from biobanking to personalized medicine." Nature communications (2025): 40188112
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111114 Website: https://genomics.ut.ee/en/content/estonian-biobank

Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)

#	Section	Description
1	Data source identification and country	FinOMOP-HUS (Hospital District of Helsinki and Uusimaa) Uusimaa, Finland
2	Data partner information section	FinOMOP (FinOMOP) HUS Tietohallinto (Information Management)
3	Coverage and timespan	Data collection since: 2004 Extent: Regional. HUS is responsible for specialized healthcare in Finland's Uusimaa region and treatment of many rare and severe diseases that are nationally centralized to HUS.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. All visits, examinations, laboratory test, procedures, and treatments are recoded in the HUS IT systems and integrated into the data lake. The data lake stores decades of clinical information in digital format, and data from both past and current source systems are available. Systems providing data into the data lake include: CGI Uranus and Epic Apotti (EHR: visits, diagnoses, medication, etc.), Opera (procedure records), Kemokur and Beacon (cancer-specific medications), Marela (hospital pharmacy), Multilab/Mylab+ (laboratory system), Qpati/Mylab+ (pathology records system). MerlotMedi (prehospital emergency care), Disease specific quality registers (disease specific). Obstetrix (birth register).
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. All visits, all procedures, and all given treatments have been recorded systematically in the electronic format. We use more than hundred different operational IT systems.
6	General representativeness	The data reflects patients treated at the hospital and is thus not a generic population.
7	Data content / source coding	Drugs: ATC, generic name, brand name + dosage and strength information in the data + local route mapping Nordic product number (VNR). Procedures: Nomesco NCSP. Genetic: special mappings to OMOP Genomics. Diagnosis: ICD10 + Finnish adaptation “ICD10fi”, ICD9fi, ICD8fi Laboratory tests: National laboratory test numbers (LABfi kuntaliitto) + local HUS variants National microbe listings, local “additional information” -coding Various observations during care: LOINC and local coding Pathology: SNOMED2
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. All patients are identified with a unique, national social security number, which is a permanent person identifier. Based on this, the derived patient identifiers (pseudonyms) in the HUS data lake also are unique.
9	Quality control (data source specific)	The HUS data lake and IT management is ISO 13485:2016 and ISO 9001:2015 certified and has capabilities in developing, validating, and CE-marking medical devices and software. Data passes through three layers of validations and transformations (bronze, silver, cold i.e. medallion lakehouse architecture) before being stored in a layout optimized for efficient analytics. The data is monitored for data freshness and data anomaly. OMOP data is converted from silver level. DARWIN EU® CC data quality dashboard SQL queries are implemented HUS ETL pipeline in addition of the standard quality control that is being done together with the DARWIN EU® CC upon onboarding.
10	Linkage	HUS is linked with Digital and Population Data Services Agency from where national death registry, socicaal security number and address are imported. Data from many nation-wide registries can be combined at HUS, which is subject to obtaining special data permits.
11	Vital status	The source of vital status (date of death) is the Digital and Population Data Services Agency.

#	Section	Description
12	Limitations	Drugs are mapped on ingredient level, making this data infeasible for studies requiring dosage information.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111134 Website: https://www.hus.fi/en

France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

#	Section	Description
1	Data source identification and country	CDW Bordeaux (Clinical Data Warehouse of Bordeaux University Hospital) Nouvelle-Aquitaine, France
2	Data partner information section	Centre Hospitalier Universitaire de Bordeaux (CHUBX)
3	Coverage and timespan	Data collection since: 2005 Extent: Regional. It covers the population of Bordeaux metropolitan area, and possibly beyond, as the health care centre for referrals and expertise for the Nouvelle Aquitaine region. The database contains data from 2005 onwards.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The database currently captures information about patient demographics, visit details, conditions, procedures, drugs, measurements, and mortality.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Biobank. The integrated data is extracted from the hospital production information system via a real-time research database. The data is then processed and quality controlled by a team dedicated to maintaining the database. Internal evaluations were carried out to ensure consistency between the research database and the patient bedside software.
6	General representativeness	This is the 6th largest metropolitan area in France, and CHUBX is the largest hospital in the region. More than 75% of the patients administered to Bordeaux university hospital reside in the Gironde departments, with almost 50% coming directly from the Bordeaux metropolitan area. The hospital also captures additional cases from Nouvelle-Aquitaine region.
7	Data content / source coding	Diagnosis source data is coded in ICD-10 terminology. Procedures are coded in CCAM (French terminology). Laboratory measurements are coded in local terminology and partially mapped to LOINC. Drugs are coded through a local terminology and then mapped to UCD (French terminology), as well as ATC codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. We use the hospital's unique identifier to generate the patient identifier in OMOP. If two identities are merged at the hospital, the merge is taken into account in the CDW. An automatic (hourly) detection of suspected duplicated identities has been implemented at the hospital since 2020, with merging of duplicated identities by a specialized team. Identities since 2015 were processed retrospectively. Thus, the rate of identity duplication in the database is low, especially since 2015.
9	Quality control (data source specific)	- The integrated data comes from the hospital production information system through a real-time replicated database. Consistency evaluations between the replicated database and the production system are performed by the technical team in charge of maintaining the replicated database. - In the same way, consistency checks are performed between the replicated database and the data integrated into the i2b2 CDW. In addition, dashboards enable monitoring the data

#	Section	Description
		integrated into the i2B2 CDW, in particular by controlling the amount of data available over time, and its evolution, according to the various data sources. - An internal evaluation was carried out to ensure the consistency between the data integrated into i2b2 and the data available in the software used at the patient's bedside. In addition, many use cases were performed on the i2b2 CDW, with return to the patient chart and comparison of the data integrated into i2b2 and the data available in the care file.
10	Linkage	Death certificates (without the cause of death).
11	Vital status	For mortality there are two data sources: hospital data, which is included in the source data, and the national death register, for which an algorithm with matching rules is run to obtain out-of-hospital mortality. This algorithm does not run regularly, and can have a delay of more than a year.
12	Limitations	CDW Bordeaux is limited to events captured in the hospital setting and thus does not include patient events not treated by the hospital (e.g. rare cancers). Patient events that are not included in CDW Bordeaux are rare disease treatments or specialist events that occur outside of CHUBX.
13	Main references	Cossin S, Diouf S,Griffier R,Le Barrois d'Orgeval P,Diallo G,Jouhet V "Linkage of Hospital Records and Death Certificates by a Search Engine and Machine Learning." JAMIA open (2021): 33709061
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111112 Website: https://www.chu-bordeaux.fr/

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

#	Section	Description
1	Data source identification and country	IMASIS (Institut Municipal Assistència Sanitària Information System) Catalunya, Spain
2	Data partner information section	Parc de Salut Mar Barcelona (PSMAR) Direcció de Control de Gestió (Management and Control Department)
3	Coverage and timespan	Data collection since: 1990 Extent: Regional. The primary catchment area covers the Barcelona city area, in Catalonia (Spain). However, all recorded visits are included, which may also include patients residing outside this area.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The available information includes diagnoses, procedures, drug administrations, dispensations and prescriptions, laboratory tests, visit types, cancer data, and vaccinations. These data are linked to primary care centers, which currently contribute information on diagnoses, and treatments, as well as to mortality data.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. The data collection process begins at the point of care, when a patient is attended in any facility within our hospital network or associated primary care centers. Clinical information is recorded directly into the electronic health records (EHRs) and subsequently complemented with data from external sources, including the regional vaccine registry, the cancer registry, the national mortality registry, and outpatient treatment records. All information is consolidated into a relational database using standardized formats and subjected to multiple quality control procedures to ensure completeness and consistency. The resulting datasets are then extracted, transformed, and loaded (ETL) into an OMOP Common Data Model (CDM) database on a quarterly to semi-annual basis.

#	Section	Description
6	General representativeness	The data source includes patients across all socioeconomic levels, ages, and genders, suggesting that the dataset is representative of the population of Barcelona requiring secondary (hospital-based) healthcare at any given time.
7	Data content / source coding	Health outcomes, diagnoses, and procedures are recorded using ICD-9-CM and ICD-10-CM classification. Laboratory tests are recorded in the source data using LOINC terminology and drugs using the national codes classification. Cancer registry is based on ICD-O-3 classification. Additionally, a set of demographics data are available such as age and sex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient cannot be registered under different ID numbers. They are registered under the same ID, even when changing practice of re-registration or admission.
9	Quality control (data source specific)	Several quality control procedures have been implemented during the creation and update of IMASIS. The process includes a series of filters that exclude any data whose format or content do not match the specifications of the target variables. Some examples of these actions include the following. - Records containing codes that do not belong to the standardized vocabularies used for encoding are reviewed within their respective domains and, if necessary, discarded. - Records with implausible values, such as dates preceding a patient's birthdate, are identified and corrected or removed. - Before each IMASIS update, the number of records in the source files is verified. Record counts from the new IMASIS instance are compared with both the source files and previous IMASIS versions to detect potential inconsistencies or anomalies in the source data. - In addition, if the Data Quality Dashboard (DQD) package identifies quality issues in the OMOP instance, for example diagnoses recorded for patients with an implausible gender, these discrepancies are reviewed and corrected in the source data.
10	Linkage	Information of diagnosis and treatments received in outpatient settings and in primary care as well as national mortality registry, cancer registry and regional vaccination registry.
11	Vital status	The hospital's death registry records the age at death for each patient when the event occurs within the hospital. In addition, deaths occurring outside the hospital are captured through linkage with the national mortality registry; however, in these cases, the specific cause of death is not currently available.
12	Limitations	No database-specific limitations are present but information coming from primary care are not complete. General limitations related to Real World Data can be considered.
13	Main references	Mayer MA, Furlong LI,Torre P,Planas I,Cots F,Izquierdo E,Portabella J,Rovira J,Gutierrez-Sacristan A,Sanz F "Reuse of EHRs to Support Clinical Research in a Hospital of Reference." Studies in health technology and informatics (2015): 25991136
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104316 Website: https://www.parcdesalutmar.cat/en/

The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Data source identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) the United Kingdom
2	Data partner information section	University of Oxford - Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences (Oxford NDORMS) NDORMS

#	Section	Description
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data are entered by clinicians into the EHR. Data is processed by CPRD that provides data releases for research.
6	General representativeness	In the last 10 years, the CPRD GOLD regional distribution of currently contributing general practitioner (GP) practices has significantly shifted, resulting in many new practices joining from Scotland, Wales, and Northern Ireland, and fewer participating from England. These changes have affected the CPRD GOLD population size, regional coverage, and eligibility for data linkages. CPRD GOLD January 2024 contains >21.3 million historical and current patients (12.9 in England, 3.1 in Wales, 4.7 in Scotland, 0.7 in Northern Ireland). Of these, nearly 3 million are currently registered in a GP practice and represent ~4.3% of the estimated current UK population (0.1% in England, 32.3% in Wales, 28.6% in Scotland, 16.2% in Northern Ireland). Patients currently registered in CPRD GOLD January 2024 are broadly representative of the UK population with respect to age and sex. Reference: https://doi.org/10.1093/ije/dyaf077
7	Data content / source coding	Gemscript, Read, dm+d
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747

#	Section	Description
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL,Dedman D,Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-gold

The United States: IQVIA US - Ambulatory EMR (IQVIA US - AmbEMR)

#	Section	Description
1	Data source identification and country	IQVIA US - AmbEMR (IQVIA US - Ambulatory EMR) the United States
2	Data partner information section	IQVIA (IQVIA) -
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. Over 100 million patients, with data sourced from more than 800 ambulatory practices and 100,000+ physicians across the United States.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians). The IQVIA Ambulatory EMR-US dataset primarily covers ambulatory care settings, which include: Outpatient physician practices, Primary care and specialty clinics, Independent and group practices, and non-hospital-based care environments.
5	Data collection process	Outpatient electronic health records. Anonymized patient records collected from patient management software used by GPs and selected specialists to document patients' clinical encounters. These records are sent to a central location then processed and aggregated.
6	General representativeness	The data is representative of the US population, with a variety of socioeconomic status patients that is not limited to specific insurances.
7	Data content / source coding	The following source vocabularies are used: CPT, HCPCX, Dx, NDC.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. IQVIA's patient tokenization process generates a consistent, anonymous patient identifier by applying multi-layered, irreversible hashing to encrypted patient attributes, enabling privacy-protected linkage of de-identified data across care settings without exposing personally identifiable information. This enables longitudinal tracking of patients.
9	Quality control (data source specific)	Integrity constraints including • 230+ IQVIA Internal QC checks • DQD & Achilles review • IQVIA internal business validation check
10	Linkage	Linkage to other IQVIA datasets is possible, e.g. PharmetricsPlus, but has to be done on a project basis.
11	Vital status	There is no linkage to a death registry.
12	Limitations	The main limitation is that no mortality information is available and there is no possibility to link a death registry. Drug dosages cannot always be calculated as for half the drug exposures the quantity is missing.

#	Section	Description
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111118 Website: https://www.iqvia.com/locations/united-states/library/fact-sheets/iqvia-ambulatory-emr-us

ANNEX II. Fitness for use assessment

The table below summarises the availability of information relevant to the planned study outcomes across the seven participating data sources, based on information provided by data partners during the feasibility assessment and additional searches in the DARWIN EU® Data Portal. Availability reflects the information identified at the time of assessment.

Data source	Gram staining	Microbiological culture	Rapid antigen / molecular testing	Pathogen identification	Antibiogram	Antimicrobial susceptibility / resistance	Comments	IRB timelines
NAJS	Mapped data identified*	Mapped data identified	Limited mapped data identified	Mapped data identified	Mapped data identified	Mapped data identified	Information based on mapped data identified through searches in the DARWIN EU® Data Portal; not confirmed through the Fitness for Use questionnaire.	<1 month
EBB	Only in source data	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	-	<1 month
CDW Bordeaux	Only in source data	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Only in source data	Only in source data	Mapping may be non-exhaustive and vary by organism	<2 weeks
FinOMOP-HUS	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Only in source data	Only in source data	Susceptibility/resistance not currently mapped to OMOP CDM	<3 months
IMASIS	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Mapped to OMOP CDM	Only in source data	Susceptibility/resistance can potentially be mapped	<3 months
CPRD GOLD	Mapped to OMOP CDM	Mapped to OMOP CDM	Only in source data	Mapped to OMOP CDM (limited to SARS-CoV-2 species detect test codes)	Only in source data	Only in source data	Only records that testing was performed are generally available; laboratory results are largely unavailable. Antigen tests exist but are not identified as rapid.	<1 month
IQVIA US - AmbEMR	Mapped data identified	Mapped data identified	Mapped data identified	Mapped data identified	Mapped data identified	Mapped data identified	Information based on mapped data identified through searches in the DARWIN EU® Data Portal; not confirmed through the Fitness for Use questionnaire.	<2 weeks

CDW Bordeaux=Clinical Data Warehouse of Bordeaux University Hospital; CPRD GOLD=Clinical Practice Research Datalink GOLD; EBB=Estonian Biobank; FinOMOP-HUS=Hospital District of Helsinki and Uusimaa; IMASIS=Institut Municipal Assistència Sanitària Information System; IQVIA US - AmbEMR=IQVIA US - Ambulatory EMR; NAJS=Croatian National Public Health Information System.
*Mapped data identified: Relevant mapped concepts with non-zero record counts were identified in the DARWIN EU® Data Portal.

ANNEX III. List of stand-alone documents

Table S1. Preliminary list of OMOP concept IDs for pathogens identification definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Escherichia coli	Escherichia coli	4011683	-	SNOMED
Klebsiella pneumoniae	Klebsiella pneumoniae	4209452	-	SNOMED
Pseudomonas aeruginosa	Pseudomonas aeruginosa	4198675	-	SNOMED
Acinetobacter species	Acinetobacter	4300777	-	SNOMED
Streptococcus pneumoniae	Streptococcus pneumoniae	4321152	-	SNOMED
Staphylococcus aureus	Staphylococcus aureus	4149419	-	SNOMED
Enterococcus faecalis	Enterococcus faecalis	4302157	-	SNOMED
Enterococcus faecium	Enterococcus faecium	4234555	-	SNOMED

Table S2. Preliminary list of OMOP concept IDs for outcomes definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Gram staining	Gram stain microscopy, CSF gram stain method, Gram stain microscopy, Identification of organism on gram stain of sputum, Identification of organism in smear by Gram stain, Microscopic observation [Identifier] in Aspirate by Gram stain, Microscopic observation [Identifier] in Eye by Gram stain, Microscopic observation [Identifier] in Stool by Gram stain, Microscopic observation [Identifier] in Cervix by Gram stain, Microscopic observation [Identifier] in Sputum by Gram stain, Microscopic observation [Identifier] in Throat by Gram stain, Microscopic observation [Identifier] in Urine by Gram stain, Microscopic observation [Identifier] in Wound by Gram stain, Microscopic observation [Identifier] in Tissue by Gram stain, Microscopic observation [Identifier] in Pus by Gram stain, Microscopic observation [Identifier] in Ear by Gram stain, Microscopic observation [Identifier] in Nose by Gram stain, Microscopic observation [Identifier] in Urethra by Gram stain, Microscopic observation [Identifier] in Specimen by Gram stain, Identification of organism in respiratory smear by Gram stain, Microscopic observation [Identifier] in Skin by Gram stain, Microscopic observation [Identifier] in Vomitus by Gram stain, Microscopic observation [Identifier] in Nasopharynx by Gram stain, Microscopic	4266852, 4307315, 4014505, 37392885, 4057614, 1246793, 3000556, 3001296, 3001426, 3001529, 3002516, 3002946, 3003745, 3007654, 3011565, 3019029, 3019385, 3024877, 3037143, 3037167, 1246799, 1176003, 36303397, 36303890, 36304197, 36304872, 36304944, 36305203, 36305247, 36305851, 43055039, 3000855, 3001057, 3002389, 3005384, 3011875, 3025346, 3027003, 3033336, 3036272, 3048285, 1091073, 1175677, 1176187, 1176326, 36303710, 36304247, 36304954, 36305102, 36305799, 3001886, 1091823, 1092076, 36304605, 36305614, 36305634, 36305744, 37021523, 1091754, 1092457	-	SNOMED, LOINC

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	observation [Identifier] in Semen by Gram stain, Microscopic observation [Identifier] in Sinus by Gram stain, Microscopic observation [Identifier] in Placenta by Gram stain, Microscopic observation [Identifier] in Blood by Gram stain, Microscopic observation [Identifier] in Isolate by Gram stain, Microscopic observation [Identifier] in Burn by Gram stain, Microscopic observation [Identifier] in Anal by Gram stain, Microscopic observation [Identifier] in Vaginal fluid by Gram stain, Microscopic observation [Identifier] in Genital mucus by Gram stain, Microscopic observation [Identifier] in Synovial fluid by Gram stain, Microscopic observation [Identifier] in Pleural fluid by Gram stain, Microscopic observation [Identifier] in Dialysis fluid by Gram stain, Microscopic observation [Identifier] in Bronchial specimen by Gram stain, Microscopic observation [Identifier] in Body fluid by Gram stain, Microscopic observation [Identifier] in Peritoneal fluid by Gram stain, Microscopic observation [Identifier] in Genital fluid by Gram stain, Microscopic observation [Identifier] in Genital specimen by Gram stain, Microscopic observation [Identifier] in Bronchoalveolar lavage by Gram stain, Microscopic observation [Identifier] in Catheter tip by Gram stain, Microscopic observation [Identifier] in Bile fluid by Gram stain, Microscopic observation [Identifier] in Implanted device by Gram stain, Microscopic observation [Identifier] in Pericardial fluid by Gram stain, Microscopic observation [Identifier] in Amniotic fluid by Gram stain, Microscopic observation [Identifier] in Bone marrow by Gram stain, Microscopic observation [Identifier] in Ocular fluid by Gram stain, Microscopic observation [Identifier] in Contact lens by Gram stain, Microscopic observation [Identifier] in Cerebral spinal fluid by Gram stain, Microscopic observation [Identifier] in Blood by Gram stain --3rd specimen, Microscopic observation [Identifier] in Blood by Gram stain --2nd specimen, Microscopic observation [Identifier] in Cornea or Conjunctiva by Gram stain, Microscopic observation [Identifier] in Lower respiratory specimen by Gram stain, Microscopic observation [Identifier] in			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Saliva (oral fluid) by Gram stain, Microscopic observation [Identifier] in Peritoneal dialysis fluid by Gram stain, Microscopic observation [Identifier] in Upper respiratory specimen by Gram stain, Microscopic observation [Identifier] in Catheter tip by Gram stain --2nd specimen, Microscopic observation [Identifier] in Catheter tip by Gram stain --3rd specimen			
Microbiological culture	Microbial culture, Aerobic culture, Anaerobic culture, Microorganism preliminary growth [Presence] in Blood by Anaerobic culture with antimicrobial neutralizer, Microorganism preliminary growth [Presence] in Blood by Aerobic culture with antimicrobial neutralizer, Microorganism preliminary growth [Presence] in Body fluid by Anaerobic culture with antimicrobial neutralizer, Microorganism preliminary growth [Presence] in Body fluid by Aerobic culture with antimicrobial neutralizer, Bacteria Identification [Presence] in Isolate by Culture, Bacteria identified in Isolate by Culture, Bacterial biochemical profile [Identifier] in Isolate by Culture, Microscopic examination of specimen from operative wound, culture and sensitivity, Genital microscopy, culture and sensitivities, Bacteria identified in Genital specimen by Aerobe culture, Bacteria identified in Genital specimen by Anaerobe+Aerobe culture	4189544, 37399191, 37399248, 1175303, 1175589, 1175387, 1176478, 3044054, 3042936, 3042872, 2007399, 4268449, 3018069, 44817158,	4006847, 4015193, 4015310, 42573066, 42599552, 42600441, 42599554, 42599555, 4130349, 4098036, 2213097, 4219447, 2213256, 2108787	SNOMED, LOINC
Rapid antigen testing	Rapid antigen detection immunoassay technique, Streptococcus pneumoniae Ag [Presence] in Cerebral spinal fluid by Rapid immunoassay, Streptococcus pneumoniae Ag [Presence] in Urine by Rapid immunoassay, Streptococcus pneumoniae antigen test in serum qualitative result	603155, 36204309, 46236341, 37155967	-	SNOMED, LOINC
Molecular testing	Acinetobacter baumannii DNA [Presence] in Specimen by NAA with probe detection, Acinetobacter baumannii DNA [Presence] in Urine by NAA with probe detection, Acinetobacter sp identified [Type] in Specimen by Sequencing, Bacteria identified based on 16S rRNA gene [Identifier] in Specimen by Sequencing, Bacterial 16S ribosomal ribonucleic acid detection and sequencing by nucleic acid amplification test, Bacterial 16S rRNA [#/mass] in XXX.tissue by NAA with probe detection,	40764121, 1988769, 1259897, 40769539, 44808606, 46234891, 46234951, 46234881, 36659879, 4039859, 36031282, 37019777, 1469977, 40764123, 1616399, 1988961, 1260074, 3966610, 36031423, 37021132, 1469475, 40764158, 1616505, 1988629, 46234895, 46234955, 46234885, 4046135, 1091580, 37021208, 36031720, 37020216, 36203576, 1001636, 1469612, 36660155, 40764159, 1617148, 1989450,	-	SNOMED, LOINC

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Bacterial 16S rRNA [#/volume] in XXX.body fluid by NAA with probe detection, Bacterial 16S rRNA [Presence] in Specimen by NAA with probe detection, Bacterial 16S rRNA gene [Nucleotide sequence] in Isolate by Sequencing, Bacterial nucleic acid assay, Enterococcus faecalis DNA [Presence] by NAA with non-probe detection in Positive blood culture, Enterococcus faecalis DNA [Presence] by NAA with probe detection in Positive blood culture, Enterococcus faecalis DNA [Presence] in Body fluid by NAA with non-probe detection, Enterococcus faecalis DNA [Presence] in Specimen by NAA with probe detection, Enterococcus faecalis DNA [Presence] in Synovial fluid by NAA with non-probe detection, Enterococcus faecalis DNA [Presence] in Urine by NAA with probe detection, Enterococcus faecalis+Enterococcus faecium DNA [Presence] in Specimen by NAA with probe detection, Enterococcus faecalis+Enterococcus faecium DNA [Presence] in Wound by NAA with probe detection, Enterococcus faecium DNA [Presence] by NAA with non-probe detection in Positive blood culture, Enterococcus faecium DNA [Presence] by NAA with probe detection in Positive blood culture, Enterococcus faecium DNA [Presence] in Body fluid by NAA with non-probe detection, Enterococcus faecium DNA [Presence] in Specimen by NAA with probe detection, Enterococcus faecium DNA [Presence] in Synovial fluid by NAA with non-probe detection, Enterococcus faecium DNA [Presence] in Urine by NAA with probe detection, Enterococcus faecium sodA gene [#/mass] in XXX.tissue by NAA with probe detection, Enterococcus faecium sodA gene [#/volume] in XXX.body fluid by NAA with probe detection, Enterococcus faecium sodA gene [Presence] in Swab specimen by NAA with probe detection, Enterococcus species rRNA assay, Escherichia coli DNA [Cycle Threshold #] in Positive blood culture by NAA with probe detection, Escherichia coli DNA [NCncRange] in Bronchoalveolar lavage by NAA with non-probe detection, Escherichia coli DNA [NCncRange] in Lower respiratory specimen by NAA with non-probe detection, Escherichia coli DNA [NCncRange] in Sputum by NAA	1616645, 3966743, 3964935, 1616308, 3966631, 1617228, 3966221, 36304996, 21493348, 1988643, 21493472, 1617150, 3043572, 3040868, 1260039, 2213167, 36204423, 1002284, 1469846, 36659962, 3964818, 40764160, 1989313, 46234946, 46235092, 46235773, 36305068, 37021268, 36032283, 37020863, 36031722, 1617677, 1092179, 37020889, 36031744, 37020756, 36203227, 1001898, 1469537, 36659735, 40764162, 1617258, 1989428, 46234894, 46234954, 1091969, 46234884, 3049510, 36659811, 1616902, 4046126, 36203554, 36203553, 36203558, 36203567, 36203552, 36203551, 36203557, 36203550, 36203549, 36203548, 36203547, 36203566, 36203565, 36203564, 36203563, 36203556, 36203562, 36203561, 36203560, 36203555, 36203559, 1092413, 37020323, 36031517, 37021074, 36203571, 37020488, 3038874, 1469938, 1470022, 1092348, 21493353, 1988899, 36659904, 1091086, 37020998, 3032699, 1617553, 3038743, 40765229, 36203507, 46234888, 46234948, 46234878, 37173626, 3012359, 4046127, 36203340		

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	with non-probe detection, Escherichia coli DNA [Presence] by NAA with non-probe detection in Positive blood culture, Escherichia coli DNA [Presence] by NAA with probe detection in Positive blood culture, Escherichia coli DNA [Presence] in Body fluid by NAA with non-probe detection, Escherichia coli DNA [Presence] in Lower respiratory specimen by NAA with probe detection, Escherichia coli DNA [Presence] in Specimen by NAA with probe detection, Escherichia coli DNA [Presence] in Synovial fluid by NAA with non-probe detection, Escherichia coli DNA [Presence] in Urine by NAA with probe detection, Escherichia coli enteroaggregative DNA [Presence] in Stool by NAA with probe detection, Escherichia coli enteroinvasive DNA [Presence] in Stool by NAA with probe detection, Escherichia coli enteroinvasive DNA [Presence] in Wound by NAA with probe detection, Escherichia coli enteropathogenic DNA [Presence] in Stool by NAA with probe detection, Escherichia coli enteropathogenic DNA [Presence] in Wound by NAA with probe detection, Escherichia coli enterotoxigenic DNA [Presence] in Stool by NAA with probe detection, Escherichia coli enterotoxigenic DNA [Presence] in Wound by NAA with probe detection, Escherichia coli gene panel - Specimen by NAA with probe detection, Escherichia coli K1 DNA [Presence] in Cerebral spinal fluid by NAA with non-probe detection, Escherichia coli K1 DNA [Presence] in Cerebral spinal fluid by NAA with probe detection, Escherichia coli O157 DNA [Presence] in Stool by NAA with non-probe detection, Escherichia coli O157 DNA [Presence] in Stool by NAA with probe detection, Escherichia coli O157:H7 DNA [Identifier] in Specimen by NAA with probe detection, Escherichia coli O157:H7 DNA [Presence] in Specimen by NAA with probe detection, Escherichia coli O157:H7 DNA [Presence] in Wound by NAA with probe detection, Infectious agent detection by nucleic acid (DNA or RNA); Staphylococcus aureus, amplified probe technique, Klebsiella pneumoniae DNA [Presence] by NAA with non-probe detection in Positive blood culture, Klebsiella pneumoniae DNA [Presence] by NAA with probe detection in Positive blood culture,			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Klebsiella pneumoniae DNA [Presence] in Body fluid by NAA with non-probe detection, Klebsiella pneumoniae DNA [Presence] in Lower respiratory specimen by NAA with probe detection, Klebsiella pneumoniae DNA [Presence] in Respiratory system specimen by NAA with probe detection, Klebsiella pneumoniae DNA [Presence] in Specimen by NAA with probe detection, Klebsiella pneumoniae DNA [Presence] in Urine by NAA with probe detection, Klebsiella pneumoniae phoE gene [#/mass] in XXX.tissue by NAA with probe detection, Klebsiella pneumoniae phoE gene [#/volume] in XXX.body fluid by NAA with probe detection, Klebsiella pneumoniae phoE gene [Presence] in Swab specimen by NAA with probe detection, Klebsiella pneumoniae yggE gene [Presence] by Probe in Positive blood culture, Klebsiella pneumoniae+Klebsiella variicola+Klebsiella quasipneumoniae DNA [NCncRange] in Bronchoalveolar lavage by NAA with non-probe detection, Klebsiella pneumoniae+Klebsiella variicola+Klebsiella quasipneumoniae DNA [NCncRange] in Lower respiratory specimen by NAA with non-probe detection, Klebsiella pneumoniae+Klebsiella variicola+Klebsiella quasipneumoniae DNA [NCncRange] in Sputum by NAA with non-probe detection, Klebsiella pneumoniae+Klebsiella variicola+Klebsiella quasipneumoniae DNA [Presence] by NAA with non-probe detection in Positive blood culture, Klebsiella pneumoniae+Klebsiella variicola+Klebsiella quasipneumoniae DNA [Presence] in Synovial fluid by NAA with non-probe detection, Pseudomonas aeruginosa DNA [Cycle Threshold #] in Positive blood culture by NAA with probe detection, Pseudomonas aeruginosa DNA [NCncRange] in Bronchoalveolar lavage by NAA with non-probe detection, Pseudomonas aeruginosa DNA [NCncRange] in Lower respiratory specimen by NAA with non-probe detection, Pseudomonas aeruginosa DNA [NCncRange] in Sputum by NAA with non-probe detection, Pseudomonas aeruginosa DNA [Presence] by NAA with non-probe detection in Positive blood culture, Pseudomonas aeruginosa DNA [Presence] by NAA with probe detection			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	in Positive blood culture, Pseudomonas aeruginosa DNA [Presence] in Body fluid by NAA with non-probe detection, Pseudomonas aeruginosa DNA [Presence] in Lower respiratory specimen by NAA with probe detection, Pseudomonas aeruginosa DNA [Presence] in Specimen by NAA with probe detection, Pseudomonas aeruginosa DNA [Presence] in Synovial fluid by NAA with non-probe detection, Pseudomonas aeruginosa DNA [Presence] in Urine by NAA with probe detection, Pseudomonas aeruginosa regA gene [#/mass] in XXX.tissue by NAA with probe detection, Pseudomonas aeruginosa regA gene [#/volume] in XXX.body fluid by NAA with probe detection, Pseudomonas aeruginosa regA gene [Cycle Threshold #] in Positive blood culture by NAA with probe detection, Pseudomonas aeruginosa regA gene [Presence] in Swab specimen by NAA with probe detection, Pseudomonas sp DNA [Identifier] in Specimen by NAA with probe detection, Staphylococcus aureus and Methicillin Resistant Staphylococcus aureus identified in Isolate or Specimen by Molecular genetics method, Staphylococcus aureus identification and resistance panel by Molecular genetics method, Staphylococcus aureus rRNA assay, Streptococcus pneumoniae Danish serotype 1 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 14 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 16F DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 19A DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 19F DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 2 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 23A DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 23F DNA [Cycle Threshold #] in Specimen by NAA with probe detection,			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Streptococcus pneumoniae Danish serotype 3 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 4 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotype 5 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 11A+11D DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 12A+12F+44+46 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 15A+15F DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 18A+18B+18C+18F DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 22A+22F DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 33A+33F+37 DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 6A+6B+6C+6D DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 6C+6D DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 7A+7F DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae Danish serotypes 9A+9V DNA [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae DNA [Cycle Threshold #] in Positive blood culture by NAA with probe detection, Streptococcus pneumoniae DNA [NCncRange] in Bronchoalveolar lavage by NAA with non-probe detection, Streptococcus pneumoniae DNA [NCncRange] in Lower respiratory specimen by NAA with non-probe detection, Streptococcus pneumoniae DNA [NCncRange] in Sputum by NAA with non-probe detection, Streptococcus pneumoniae DNA [Presence] by NAA with non-probe detection in Positive blood culture, Streptococcus			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	pneumoniae DNA [Presence] by NAA with probe detection in Positive blood culture, Streptococcus pneumoniae DNA [Presence] in Blood by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Body fluid by NAA with non-probe detection, Streptococcus pneumoniae DNA [Presence] in Bronchial specimen by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Bronchoalveolar lavage by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Cerebral spinal fluid by NAA with non-probe detection, Streptococcus pneumoniae DNA [Presence] in Cerebral spinal fluid by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Lower respiratory specimen by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Nasopharynx by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Respiratory system specimen by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Specimen by NAA with probe detection, Streptococcus pneumoniae DNA [Presence] in Synovial fluid by NAA with non-probe detection, Streptococcus pneumoniae DNA and Haemophilus influenza DNA and Pseudomonas aeruginosa DNA [Identifier] in Unspecified specimen by NAA with probe detection, Streptococcus pneumoniae DNA and Haemophilus influenza DNA and Streptococcus agalactiae DNA [Identifier] in Unspecified specimen by NAA with probe detection, Streptococcus pneumoniae lytA gene [Cycle Threshold #] in Specimen by NAA with probe detection, Streptococcus pneumoniae nanA gene [#/mass] in XXX.tissue by NAA with probe detection, Streptococcus pneumoniae nanA gene [#/volume] in XXX.body fluid by NAA with probe detection, Streptococcus pneumoniae nanA gene [Presence] in Swab specimen by NAA with probe detection, Streptococcus pneumoniae RNA nucleic acid amplification qualitative result, Streptococcus pneumoniae rRNA [Presence] in Specimen by Probe, Streptococcus pneumoniae rRNA assay, Streptococcus pneumoniae serotype [Identifier] in Specimen by NAA with probe detection.			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
Antibiogram or antimicrobial susceptibility/resistance result	Acinetobacter sp multidrug resistant identified in Specimen by Organism specific culture, Amikacin [Susceptibility] by Genotype method, Azithromycin [Susceptibility] by Genotype method, Carbapenem resistance bla OXA-48-like gene [Presence] by Molecular method, Carbapenem resistance blaGES gene [Presence] by Molecular method, Carbapenem resistance blaGIM gene [Presence] by Molecular method, Carbapenem resistance blaIMP gene [Presence] by Molecular method, Carbapenem resistance blaIMP+blaVIM genes [Presence] by Molecular method, Carbapenem resistance blaKPC gene [Presence] by Molecular method, Carbapenem resistance blaNDM gene [Presence] by Molecular method, Carbapenem resistance blaOXA gene [Presence] by Molecular method, Carbapenem resistance blaOXA-134-like gene [Presence] by Molecular method, Carbapenem resistance blaOXA-23-like gene [Presence] by Molecular method, Carbapenem resistance blaOXA-23-like+blaOXA-48-like genes [Presence] in Isolate or Specimen by Molecular method, Carbapenem resistance blaOXA-24-like gene [Presence] by Molecular method, Carbapenem resistance blaOXA-48 gene [Presence] by Molecular method, Carbapenem resistance blaOXA-51-like gene [Presence] by Molecular method, Carbapenem resistance blaOXA-58-like gene [Presence] by Molecular method, Carbapenem resistance blaSPM gene [Presence] by Molecular method, Carbapenem resistance blaVIM gene [Presence] by Molecular method, Carbapenem resistant Acinetobacter, Carbapenem resistant Acinetobacter baumannii, Carbapenem resistant bacteria identification and resistance panel by Molecular genetics method, Carbapenem resistant Escherichia coli, Carbapenem resistant Klebsiella pneumoniae, Carbapenem resistant Pseudomonas aeruginosa, Carbapenem-resistant Acinetobacter baumannii-calcoaceticus complex, Carbapenemase [Presence] in Isolate, Carbapenemase enzyme panel - Isolate by Rapid immunoassay, Carbapenemase resistance genes panel by Molecular genetics method, Carbapenemase-producing Escherichia coli,	11. 3038950, 36303883, 36305864, 36203627, 36203623, 36203668, 36203342, 36031900, 3033281, 44786640, 36203643, 36204213, 36203625, 1001767, 36203626, 36203347, 36204214, 36203628, 36203664, 36203345, 40492250, 37395666, 1616850, 37395753, 40489864, 37111504, 37311317, 36304196, 3965928, 36203346, 42537865, 42539719, 1762661, 1762685, 1762629, 1762596, 1762626, 1762655, 45769172, 45878777, 45878204, 3037064, 3035814, 4249829, 36309501, 1074765, 4232583, 3966742, 1552104, 724545, 2213121, 1762643, 1762594, 3966111, 3009196, 3001052, 3007224, 36305194, 36203279, 3050167, 36203840, 4019195, 1175793, 3033966, 1091253, 3050898, 3039355, 46235760, 1091581, 3019902, 1092926, 1002846, 1002599, 1003779, 1092729, 1002612, 3172535, 4046599, 44791899, 4207305, 37068492, 37064865, 44805759, 440940, 4261071, 44805508, 4258130, 3170862, 763925, 4296650, 37161103, 618568, 37079272, 37160702, 37162497, 37162496, 45773609, 37019940, 37021005, 37020956, 37020595, 37020700, 37020566, 37020491, 37020269, 37172439, 37397888, 4201099, 40486501, 37155971, 46272035, 37396650, 37395792, 37017619, 37017940, 37017620, 4257549, 3965178, 37396649, 1091767, 1091544, 3965235, 4253018, 36311052, 40766210, 37078013, 1470666, 36659811, 21492812, 37059304, 42527877, 37067149, 42527878, 42527819, 1618298, 1616902, 43534061, 37033874, 36660488, 46236372, 36660997, 37024475, 45879792, 46236973, 3032254, 3037971, 3005004, 4229388, 4249133, 1244063, 46274007, 4225532, 4232584, 44784188, 3965833, 4220083, 4032801, 4211363, 4098501, 4234899, 4014622, 4013177, 4015325, 4231889, 4016374, 4312086, 4269484, 4337366, 40488332, 40489303,	-	LOINC, SNOMED, CPT4, Nebraska Lexicon

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Carbapenemase-producing Pseudomonas aeruginosa, E. coli carbapenemase producing, E. coli colistin resistant, E. coli ESBL, E. coli OXA-48 carbapenemase producing, Enterococcus faecalis (vanB), Enterococcus faecium (vanA), Enterococcus faecium genotype vanA, Enterococcus faecium.hsp60 gene and vanA genes (VRE), Enterococcus faecium.hsp60 gene and vanB genes (VRE), Ertapenem [Susceptibility], Ertapenem [Susceptibility] by Minimum inhibitory concentration (MIC), ESBL Escherichia coli, Escherichia coli extended-spectrum beta-lactamases (ESBL) isolated, Extensively drug resistant Pseudomonas aeruginosa, Glycopeptide resistant Staphylococcus aureus culture, IMP Carbapenemase [Presence] in Isolate by Rapid immunoassay, Infectious agent detection by nucleic acid (DNA or RNA), genitourinary pathogens, identification of 20 bacterial and fungal organisms and identification of 16 associated antibiotic-resistance genes, multiplex amplified probe technique, Infectious agent detection by nucleic acid (DNA or RNA), genitourinary pathogens, identification of 21 bacterial and fungal organisms and identification of 21 associated antibiotic-resistance genes, multiplex amplified probe technique, urine, Infectious agent detection by nucleic acid (DNA or RNA); vancomycin resistance (eg, enterococcus species van A, van B), amplified probe technique, K. pneumoniae carbapenemase producing, K. pneumoniae colistin resistant, KPC carbapenemase [Presence] in Isolate by Rapid immunoassay, Meropenem [Susceptibility], Meropenem [Susceptibility] by Gradient strip, Meropenem [Susceptibility] by Minimum inhibitory concentration (MIC), Meropenem+Vaborbactam [Susceptibility], Meropenem+Vaborbactam [Susceptibility] by Minimum inhibitory concentration (MIC), Methicillin resistance mecA gene [Presence] by Molecular method, Methicillin resistance mecC gene [Presence] by Molecular method, Methicillin resistant Staphylococcus aureus, Methicillin resistant Staphylococcus aureus (MRSA) DNA [Presence] in Nose by NAA with probe detection, Methicillin resistant	724584, 745516, 745515, 3172535, 44791899, 44805759, 440940, 3170862, 37161103, 37160702, 37162497, 37162496, 4224948, 37017620, 4266368		

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Staphylococcus aureus (MRSA) DNA [Presence] in Specimen by NAA with probe detection, Methicillin resistant Staphylococcus aureus [Presence] in Axilla by Organism specific culture, Methicillin resistant Staphylococcus aureus [Presence] in Genital specimen by Organism specific culture, Methicillin resistant Staphylococcus aureus [Presence] in Nose by Organism specific culture, Methicillin resistant Staphylococcus aureus [Presence] in Pharynx by Organism specific culture, Methicillin resistant Staphylococcus aureus [Presence] in Skin by Organism specific culture, Methicillin resistant Staphylococcus aureus [Presence] in Specimen by Organism specific culture, Methicillin resistant Staphylococcus aureus | Axilla | Microbiology, Methicillin resistant Staphylococcus aureus | Genital | Microbiology, Methicillin resistant Staphylococcus aureus | Nose | Microbiology, Methicillin resistant Staphylococcus aureus | Pharynx | Microbiology, Methicillin resistant Staphylococcus aureus | Skin | Microbiology, Methicillin resistant Staphylococcus aureus | XXX | Microbiology, Methicillin resistant Staphylococcus aureus cellulitis of left foot, Methicillin resistant Staphylococcus aureus culture, Methicillin resistant staphylococcus aureus decontamination therapy, Methicillin resistant Staphylococcus aureus detected, Methicillin resistant Staphylococcus aureus DNA | Nose | Microbiology, Methicillin resistant Staphylococcus aureus DNA | XXX | Antibiotic susceptibilities, Methicillin resistant Staphylococcus aureus eradication therapy, Methicillin resistant Staphylococcus aureus infection, Methicillin resistant Staphylococcus aureus isolated by culture, Methicillin resistant staphylococcus aureus negative, Methicillin resistant Staphylococcus aureus not detected by screening, Methicillin resistant Staphylococcus aureus pneumonia, Methicillin resistant Staphylococcus aureus screen positive, Methicillin resistant Staphylococcus aureus screening test, Methicillin resistant Staphylococcus aureus skin infection, Methicillin resistant Staphylococcus aureus swab, Methicillin-			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	resistant Staphylococcus aureus (MRSA), Methicillin-resistant Staphylococcus aureus colitis, Methicillin-resistant Staphylococcus aureus enteritis of intestine, Methicillin-resistant Staphylococcus aureus enteritis of small intestine, Meticillin-resistant Staphylococcus aureus targeted genetic test, Microorganism identification and resistance pattern determination panel by Molecular genetics method, Microorganism molecular resistance panel by Molecular genetics method, Microorganism resistance mutation amino acid change by Molecular method, Microorganism resistance mutation amino acid change type by Molecular method, Microorganism resistance mutation detected [Presence] by Molecular method, Microorganism resistance mutation nucleic acid change by Molecular method, Microorganism resistance mutation nucleic acid change type by Molecular method, Microorganism resistance mutation tested for by Molecular method, MRSA (methicillin resistant Staphylococcus aureus) DNA nucleic acid amplification qualitative result, MRSA (methicillin resistant Staphylococcus aureus) screening test, Multi-resistant staphylococcus aureus screening, Multidrug-resistant Acinetobacter, Multidrug-resistant Acinetobacter baumannii-calcoaceticus complex, Multidrug-resistant Pseudomonas aeruginosa, Multiple drug resistant methicillin resistant Staphylococcus aureus, Multiple drug-resistant Acinetobacter baumannii, Multiple drug-resistant Escherichia coli, Multiple drug-resistant Klebsiella pneumoniae, Multiple drug-resistant Klebsiella pneumoniae subsp. ozaenae, Multiple drug-resistant Streptococcus pneumoniae, NDM Carbapenemase [Presence] in Isolate by Rapid immunoassay, Non-multiple drug resistant methicillin resistant Staphylococcus aureus, OXA-23 carbapenemase [Presence] by Rapid immunoassay, OXA-40 and OXA-58 carbapenemase [Presence] by Rapid immunoassay, OXA-48-like carbapenemase [Presence] in Isolate by Rapid immunoassay, Penicillin resistant Streptococcus pneumoniae, Pseudomonas aeruginosa multidrug			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	resistant isolated, Pseudomonas aeruginosa.multidrug resistant isolate [Presence] in Specimen by Organism specific culture, Pseudomonas aeruginosa.multidrug resistant isolate | XXX | Microbiology, Staphylococcus aureus and Methicillin Resistant Staphylococcus aureus | Isolate or Specimen | Microbiology, Staphylococcus aureus and Methicillin Resistant Staphylococcus aureus identified in Isolate or Specimen by Molecular genetics method, Staphylococcus aureus and Methicillin-resistant Staphylococcus aureus panel - Nose by NAA with probe detection, Staphylococcus aureus and Methicillin-resistant Staphylococcus aureus panel | Nose | Microbiology Panels, Staphylococcus aureus glycopeptide resistance panel, Staphylococcus aureus glycopeptide resistance panel | Isolate | Antibiotic Susceptibility Panels, Staphylococcus aureus glycopeptide resistant based on vancomycin and teicoplanin IC [Identifier] by ARD gradient strip, Staphylococcus aureus glycopeptide resistant identified [Type] in Isolate, Staphylococcus aureus identification and resistance panel | Isolate or Specimen | Antibiotic Susceptibility Panels, Staphylococcus aureus identification and resistance panel by Molecular genetics method, Staphylococcus aureus methicillin resistance SCCmec [Presence] in Nose by NAA with probe detection, Staphylococcus aureus methicillin resistance SCCmec | Nose | Antibiotic susceptibilities, Staphylococcus aureus methicillin resistance SCCmec+orfX junction [Presence] by Molecular method, Staphylococcus aureus methicillin resistance SCCmec+orfX junction [Presence] in Nose by NAA with probe detection, Staphylococcus aureus methicillin resistance SCCmec+orfX junction | Isolate or Specimen | Antibiotic susceptibilities, Staphylococcus aureus methicillin resistance SCCmec+orfX junction | Nose | Antibiotic susceptibilities, Staphylococcus aureus.tuf and mecA genes (MRSE), Staphylococcus aureus.vancomycin resistance [Susceptibility] Qualitative, Tigecycline [Susceptibility], Tigecycline [Susceptibility] by Minimum inhibitory			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	concentration (MIC), Vancomycin [Susceptibility] by Gradient strip, Vancomycin intermediate Staphylococcus aureus, Vancomycin intermediate/resistant Staphylococcus aureus, Vancomycin resistant Enterococcus faecalis genotype vanB, Vancomycin resistant Enterococcus faecium, Vancomycin resistant Staphylococcus aureus, Vancomycin resistant Staphylococcus aureus culture, Vancomycin susceptible Staphylococcus aureus, VIM Carbapenemase [Presence] in Isolate by Rapid immunoassay, Bacterial antibiogram analysis, Antimicrobial susceptibility test, Bacterial culture and sensitivity, Culture and susceptibility, Antibiotic sensitivity, disk method, Antibiotic sensitivity, agar diffusion method, Agar disk elution susceptibility test for Mycobacteria, Antibiotic sensitivity, tubercle bacillus, Antifungal susceptibility test, Antiviral susceptibility test, BACTEC susceptibility test for slow growing Mycobacteria, Broth microdilution susceptibility test for Mycobacteria, Broth microdilution susceptibility test for Nocardia, HIV 1 phenotyping, HIV coreceptor tropism assay, Infectious agent detection by nucleic acid (DNA), genitourinary pathogens, identification of 20 bacterial and fungal organisms, including identification of 20 associated antibiotic-resistance genes, if performed, multiplex amplified probe ... (Deprecated), Infectious disease (Mycoplasma genitalium), macrolide sensitivity (23S rRNA point mutation), oral, rectal, or vaginal swab, algorithm reported as probability of macrolide resistance, Infectious disease (Neisseria gonorrhoeae), sensitivity, ciprofloxacin resistance (gyrA S91F point mutation), oral, rectal, or vaginal swab, algorithm reported as probability of fluoroquinolone resistance, Methicillin resistant Staphylococcus aureus cellulitis of left foot, Methicillin resistant staphylococcus aureus decontamination therapy, Methicillin resistant Staphylococcus aureus eradication therapy, Methicillin resistant Staphylococcus aureus infection, Methicillin resistant Staphylococcus aureus pneumonia, Methicillin resistant Staphylococcus aureus skin infection,			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Methicillin-resistant Staphylococcus aureus colitis, Methicillin-resistant Staphylococcus aureus enteritis of intestine, Methicillin-resistant Staphylococcus aureus enteritis of small intestine, Modified proportion agar dilution test for slow growing Mycobacteria, Multiple drug-resistant Klebsiella pneumoniae subsp. ozaenae, Radiometric test for slow growing Mycobacteria.			
Haemolysis patterns	Hemolysis on sheep blood agar test, Bacteria producing hemolysis [Presence] in Isolate, Bacterial colony hemolysis, Bacterial colony hemolysis, gamma, Bacterial colony hemolysis, alpha, Bacterial colony hemolysis, beta, Bacterial colony hemolysis, alpha prime	4142347, 3047289, 4292085, 4167006, 4215615, 4312209, 4009929	-	SNOMED, LOINC
Catalase	Catalase test, Catalase test, heat-stable, Catalase test, semi-quantitative, Bacteria producing catalase [Presence] in Isolate, Gram positive catalase producing cocci identified in Isolate by Organism specific culture, Gram positive non-catalase producing cocci identified in Isolate by Organism specific culture	4051794, 4272017, 4338218, 3045324, 37021288, 37019980	-	SNOMED, LOINC
Coagulase	Coagulase test, Bacteria producing coagulase and protein A [Presence] in Isolate by Latex agglutination	4214131, 43533695	-	SNOMED, LOINC
Oxidase	Spot oxidase test, Bacteria producing oxidase [Presence] in Isolate	4307948, 3044985	-	SNOMED, LOINC
Indole	Spot indole test, Bacteria producing indole [Presence] in Isolate	4172509, 3045945	-	SNOMED, LOINC
Urease	Urease test, Urease test, Wayne method, Bacteria producing urease [Presence] in Isolate, Helicobacter pylori detected by rapid urease test, Helicobacter pylori not detected by rapid urease test	4215465, 4097695, 3046685, 4056394, 4058632	-	SNOMED, LOINC, CPT4
Motility	Motility test, Microorganism motility	4281838, 40480051	-	SNOMED, CPT4

Table S3. List of OMOP concept IDs for indications/infections.

Phenotypes	Concept IDs (including descendants) *	Concept names	Exclude concept IDs	Vocabulary
Eye infections	378425, 438411, 4134613, 379019	Blepharitis, Endophthalmitis, Eye infection, Conjunctivitis	42572963, 42598813, 42598791, 4220916, 42593602, 42593599, 42573137	SNOMED
Otitis media	372328	Otitis media	42598429	SNOMED
Sinusitis	4283893	Sinusitis	42573224	SNOMED
Bronchitis	4105601, 255841, 4279922, 4233924, 260139	Purulent bronchitis, Chronic bronchitis, Septic bronchitis, Pneumococcal bronchitis, Acute bronchitis	765431, 261889, 37206130, 4230358, 4172303	SNOMED
Pneumonia	4050869, 255848	Atypical pneumonia, Pneumonia	42598655, 42573181, 42572881, 42599199, 42599060, 42598979, 42573179, 42573178, 42573020, 42572644, 42599561, 42598991, 42598908, 42593423, 42600053, 42573349, 42573218, 42600167	SNOMED
Endocarditis	441589	Endocarditis	42597024, 42597025, 42597026, 42574210	SNOMED
Acute infectious diarrhoea/ gastroenteritis	197484, 198337, 4101468, 4167082, 193688	Diarrhea of presumed infectious origin, Infectious diarrheal disease, Gastroenteritis, Bacterial gastroenteritis, Clostridioides difficile infection, Purulent bronchitis, Diarrhea of presumed infectious origin, Infectious diarrheal disease, Gastroenteritis, Bacterial gastroenteritis, Clostridioides difficile infection	42598601, 42594405, 42598710, 42595096, 4146457, 196895, 42572970, 42572684, 42597110, 42573165, 42599178, 42599162, 42599037, 42572975, 42572901, 42599097, 42572974, 42598915, 42572896, 42599123, 42598906, 42573209	SNOMED
Intra-abdominal infections	4141224, 4178110, 4276784, 4340497, 4192640, 4171019, 4088114, 3187323, 3176245, 195856, 196152, 192956, 42574016, 201901	Diverticulitis, Traumatic perforation of large intestine, Traumatic perforation of small intestine, Infected pancreatic necrosis, Pancreatitis, Abdominal visceral abscess, Postoperative intra-abdominal abscess, Right upper abdominal abscess, Right lower quadrant abdominal abscess, Cholangitis, Peritonitis, Cholecystitis, Clostridial colitis, Abscess of liver	42573163, 42573162, 42599237, 42595100, 42573164, 4342877, 4341774, 42599158, 42574230, 42597029, 42597105, 42597359, 42599235, 42597055, 42573348, 954117	SNOMED

Phenotypes	Concept IDs (including descendants) *	Concept names	Exclude concept IDs	Vocabulary
Urinary tract infections	195862, 81902, 198199, 195588, 77340, 4142194, 198192	Urethritis, Urinary tract infectious disease, Pyelonephritis, Cystitis, Genitourinary tract infection in pregnancy, Renal abscess, Renal and perinephric abscess	42600293, 42598600, 42593547, 42598968, 42573007, 42600389, 42572888, 3194810, 4056622, 4047937, 607157, 4032450, 42600089, 4100954, 606982, 4239213, 4030507, 37017585, 4128374, 37017217, 954561	SNOMED
Other genito-urinary infection	194997, 78193, 199067, 4312329, 4149084, 196163, 4171915, 4150043	Prostatitis, Orchitis and epididymitis, Female pelvic inflammatory disease, Perimetritis, Vaginitis, Cervicitis and endocervicitis, Orchitis, Epididymitis	42599058, 42597060, 954079, 42598794, 42573188, 42599971, 42573334, 42599837	SNOMED
Meningitis and encephalitis	435785, 378143	Meningitis, Encephalitis	42593067, 42599865, 42599239, 42596612, 42574231, 42596575, 42596610, 42596611, 42596666, 42596655, 42596657, 42599838, 42572953, 42598680, 42572787, 42573325, 42597997, 42599011, 42599000, 42596629, 42572712, 42598867, 42597146, 42598804	SNOMED
Sepsis or septic shock	132797, 196236	Sepsis, Septic shock	-	SNOMED
Febrile neutropenia	4250734	Febrile neutropenia	-	SNOMED
Osteomyelitis	141663	Osteomyelitis	42598482, 42597272, 42597177, 42597205, 42574042, 42597284, 42597165, 42597202, 42597228, 42597288, 42597289, 42597232, 42574047, 42597209, 42597155	SNOMED
Septic arthritis	4180167	Infectious disorder of joint	42598859, 42599075, 42598482, 42597189, 42598680, 42573193, 42598483, 42597162, 42597218, 42597169, 42574034, 42599633, 42597196, 42597291, 42597269, 42598723, 42599281, 42597284, 42574047, 42575811	SNOMED
Skin and soft tissue infections	435740, 442013, 4029483, 4030849, 4046789, 4090679, 4096471, 4096472, 4096474, 4141909, 4151842, 4262920, 4159742, 4153877, 4183970, 4211967, 4237450, 4297984, 36715557, 44806474,	Tetanus, Burn, Anaplasmosis, Wound abscess, Penetrating wound, Pasteurella septic infection (cat or dog bite, Deep wound, Penetrating wound, Deep avulsion wound, Bite wound, Bite - wound, Skin ulcer, Diabetic foot ulcer, Post-traumatic wound infection, Wound dirty, Deep wound, Postoperative wound infection, Local	4335890, 42596707, 42599845, 42596392, 42599652, 42574046, 46270494, 3657767, 3657768, 4161830, 3657864, 36714565, 4146111, 380944, 36714497, 4028254, 37209315, 608803, 37209444, 608804, 37108901, 4287838, 372550, 442552, 3654936, 375187, 378352, 4244336, 4231740, 4345914, 761333, 761332, 765217, 37209445,	SNOMED

Phenotypes	Concept IDs (including descendants) *	Concept names	Exclude concept IDs	Vocabulary
	44806641, 440638, 435613, 4200289, 4300385, 4320030	infection of wound, Pyogenic infection of skin and subcutaneous tissues caused by bacterium, O/E - wound necrotic, Infection associated with wound, Lyme disease, Cellulitis, Preseptal cellulitis, Vulval cellulitis, Soft tissue infection	765280, 765071, 37310072, 37309999, 37309947, 37310064, 37310055, 37310010, 37309965, 37310053, 37309987, 37310051, 36683380, 4335889, 372842, 4253626, 4218443, 4225881, 4222588, 4224271, 4222419, 4294573, 4222420, 4222421, 4224272, 4096836, 4294574, 4300324, 4300325, 4293288, 4294575, 4293289, 4220840, 4298991, 4300326, 4294576, 4305732, 4298994, 4133972, 4082049, 4078538, 4167083, 80946, 4222430, 4226003, 4222431, 4222597, 4300328, 4226004, 134865, 4300335, 4294703, 4163280, 133141, 4183870, 4294841, 4094067, 4294578, 4298995, 4293290, 4060657, 4078931, 4300215, 4144953, 4028324, 40490394, 40491348, 40490302, 40489357, 4291603, 4051339, 192361, 4110039, 4137452, 3654661, 4026017, 198076, 197790, 4342876, 4342877, 4144034, 196345, 443796, 4341774, 4198021, 199863, 198189, 44808741, 44808743, 198345, 194268, 37311895, 37311635, 3176778, 36683379, 37311636, 37310070, 37310004, 37309946, 37310063, 37310054, 37310009, 37309964, 42599710, 42598686, 42572881, 42598451, 42598458, 42596424, 42592702, 42572887, 42599711, 42599271, 42592701, 42574134, 42600167, 42598622, 42598579, 42572604, 42574135, 42598463, 42598796, 42596302, 42599966, 42573149, 954118, 4239213, 4338751, 4344052, 37161097, 42536747, 37017777, 4298735, 4082050, 3655666, 3655670, 607399, 4122211, 4300224, 4301039, 4294449, 4112116, 134074, 4124365, 4080460, 4124508, 3657965, 4122755, 4122209, 4124509, 4122759, 4122756, 4122760, 4095882, 4096496, 4122757, 4122758, 4056925, 4122210, 4122751, 4126928, 761331, 37209443, 761784, 761330, 761334, 761335, 761336, 4103893, 4265858, 42599147, 42573184, 42572969, 42572859, 42573180,	

Phenotypes	Concept IDs (including descendants) *	Concept names	Exclude concept IDs	Vocabulary
			42593793, 42597178, 42597044, 4253627, 376028, 4209134, 4159754, 4103653, 4089879, 4265439, 4163288, 761329, 761328, 4106896, 4259503, 442757, 4312036, 4265742, 4238336, 376422, 765528, 764828, 765739, 4171248, 4084947, 36717459, 373870, 4146848, 4132505, 4171250, 378753, 4306683, 4300208, 4348434, 4302049, 4098881, 4009923, 4088086, 380038, 4093292, 137213, 4141481, 140027, 132835, 4080331, 4160328, 137785, 4212782, 4239344, 4243325, 760174, 4217694, 4249876, 4202370, 4201753, 619343, 4084948, 4291601, 760929, 4084816, 4080330, 4347555, 4177636, 4148102, 4345473, 4084817, 4080770, 4291602, 4081909, 4142828, 4185025, 4182586, 760906, 760907, 44810378, 4345474, 4294439, 4300214, 4294440, 3656108, 4111926, 4124510, 4200244, 4080745, 4172458, 45757655, 4234876, 4122212, 4080744, 4270604, 4070648, 4298988, 4084967, 4142837, 4082052, 4270605, 4080315, 609055, 36687211, 609054, 4163426, 4067324, 4183059, 4174374, 4188361, 4010883, 4204375, 4321597, 4182398, 4185834, 4225882, 37310052, 37309986, 37310050, 37310012, 4052375, 4343215, 4155027, 4152958, 4181310, 4151521, 4155028, 4152956, 4151520, 40483694, 137441, 4052230, 4052650, 40408433, 436859, 4171380, 443761, 4171058, 4221875, 4219819, 4032628, 4178891, 4308445, 4210750, 4065912, 80883, 444103, 142016, 4140541, 4213341, 442640, 4093263, 4183151, 132483, 194799, 4198074, 760765, 4246106, 42709771, 443594, 439197, 4195716, 4148533, 4149402, 760764, 4031829, 4187237, 4300986, 442351, 4066406, 137425, 4262214, 444436, 42596301, 42573954, 42597130, 42597138, 42572756, 42598859, 42599243, 42599065, 42572696, 42572847, 42598627, 42599516, 42599028,	

Phenotypes	Concept IDs (including descendants) *	Concept names	Exclude concept IDs	Vocabulary
			42572939, 42599020, 42596299, 42592867, 42596077, 42573163, 42573187, 42572772, 42572968, 42573162, 42599237, 42572450, 42572963, 42574126, 42598916, 42598956, 42598813, 42572960, 42574044, 42572894, 42599106, 42600083, 42573137, 42599273, 42596405, 42596392, 42599606, 42573164, 42598791, 42598582, 42593338, 42593339, 42599712, 42593342, 42572950, 42599373, 42599866, 42599502, 42598589, 42599043, 42596300, 42572953, 42598732, 42597046, 42597037, 42597036, 705175, 4136355	
Localized acute bacterial lymphadenitis	316084	Lymphadenitis	313989, 4081069, 4210780, 4141212, 435234, 4256233, 40487345, 4212138, 42598890, 42572895, 42599003	SNOMED

*Concept IDs included descendants unless highlighted as being excluded. By OMOP standards descendants automatically include the ancestor.

Table S4. List of OMOP concept IDs for antibiotics classes of interest.

	Antibiotics class	ATC code	Concept ID (including descendants)	Vocabulary
1	Penicillins	J01CA, J01CE, J01CF	21602819, 21602840, 21602852	RxNorm
2	Beta-lactam/beta-lactamase inhibitor combinations	J01CR	21602861	RxNorm
3	Cephalosporins	J01DB, J01DC, J01DD, J01DE, J01DI	21602869, 21602882, 21602894, 21602914, 21602927	RxNorm
4	Carbapenems	J01DH	21602920	RxNorm
5	Fluoroquinolones	J01MA	21603007	RxNorm
6	Aminoglycosides	J01GA, J01GB	21602991, 21602994	RxNorm
7	Glycopeptides	J01XA	21603042	RxNorm
8	Polymyxins	J01XB	21603048	RxNorm
9	Oxazolidinones	J01XX08, J01XX11	21603068, 1588637	RxNorm
10	Lipopeptides	J01XX09	21603069	RxNorm

ANNEX IV. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Characterising microbiological pathogen and antibiotic resistance data in DARWIN EU® data sources *a priori* mapped to the OMOP CDM

EU PAS Register® number: EUPAS1000001118.
Study reference number (if applicable): P5-C1-019

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	4
1.1.2 End of data collection[2]	☒	☐	☐	4
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☐	☐	☒	
1.1.5 Registration in the EU PAS Register®	☐	☒	☐	Not registered yet
1.1.6 Final report of study results.	☒	☐	☐	4

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	6
2.1.1 Why the study is conducted? (e.g., to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	5
2.1.2 The objective(s) of the study?	☒	☐	☐	6
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	7.3
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☒	☐	☐	7.7

Comments:

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Section 3: Study design		**Yes**	**No**	**N/A**	**Section Number**
3.1	Is the study design described? (e.g., cohort, case-control, cross-sectional, other design)	☒	☐	☐	7.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	7.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	7.8
3.4	Does the protocol specify measure(s) of association? (e.g., risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g., adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	7.3, 7.4
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	☒	☐	☐	7.5
	4.2.2 Age and sex	☒	☐	☐	7.3
	4.2.3 Country of origin	☒	☐	☐	7.4
	4.2.4 Disease/indication	☒	☐	☐	7.3
	4.2.5 Duration of follow-up	☒	☐	☐	7.2
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g., event or inclusion/exclusion criteria)	☒	☐	☐	7.3

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g., operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☒	
5.2	Does the protocol address the validity of the exposure measurement? (e.g., precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3	Is exposure categorised according to time windows?	☐	☐	☒	

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.4 Is intensity of exposure addressed? (e.g., dose, duration)	☐	☐	☒	
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6 Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	7.6.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	7.6.2
6.3 Does the protocol address the validity of outcome measurement? (e.g., precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☒	☐	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g., HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☒	☐	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g., confounding by indication)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g., healthy user/adherer bias)	☒	☐	☐	8
7.3 Does the protocol address information bias? (e.g., misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	8

Comments:

Section 8: Effect measure modification	Yes	No	N/A	Section Number
8.1 Does the protocol address effect modifiers? (e.g., collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☐	☐	☒	

Comments:

Section 9: Data sources	Yes	No	N/A	Section Number
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g., pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	☒	
9.1.2 Outcomes? (e.g., clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	7.6.2
9.1.3 Covariates and other characteristics?	☒	☐	☐	7.6.3
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g., date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☒	
9.2.2 Outcomes? (e.g., date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	7.6.2
9.2.3 Covariates and other characteristics? (e.g., age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	7.6.3
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g., WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☒	
9.3.2 Outcomes? (e.g., International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	7.6.2/Annex IV
9.3.3 Covariates and other characteristics?	☒	☐	☐	7.6.3/Annex IV
9.4 Is a linkage method between data sources described? (e.g., based on a unique identifier or other)	☐	☐	☒	

Comments:

Section 10: Analysis plan	Yes	No	N/A	Section Number
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	7.8
10.2 Is study size and/or statistical precision estimated?	☐	☐	☒	

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.3 Are descriptive analyses included?	☒	☐	☐	7.8
10.4 Are stratified analyses included?	☒	☐	☐	7.8
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7 Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8 Are relevant sensitivity analyses described?	☐	☐	☒	

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g., software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	Annex III
11.2 Are methods of quality assurance described?	☒	☐	☐	Annex III
11.3 Is there a system in place for independent review of study results?	☒	☐	☐	Annex III

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☐	☐	☒	
12.1.2 Information bias?	☒	☐	☐	8
12.1.3 Residual/unmeasured confounding? (e.g., anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☐	☒	
12.2 Does the protocol discuss study feasibility? (e.g., study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	Annex II

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	Annex II
13.2 Has any outcome of an ethical review procedure been addressed?	☒	☐	☐	Annex II
13.3 Have data protection requirements been described?	☒	☐	☐	Annex III

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	3

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g., to regulatory authorities)?	☒	☐	☐	Annex III
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	Annex III

Comments:

Name of the main author of the protocol: Marzyeh Amini

Date: 07/09/2026

Signature: *Marzyeh Amini*