Big Data in Healthcare from Aragon

First published: 01/02/2024

Last updated: 17/10/2024

Data source | Human | Administrative healthcare records (e.g., claims) | Other | Pharmacy dispensing records

Administrative details

Administrative details

Data source ID

1111124

Data source acronym

BiGAN

Data holder

Instituto Aragonés de Ciencias de la Salud (IACS)

Data source type

Administrative healthcare records (e.g., claims)

Other

Pharmacy dispensing records

Data source type, other

Electronic health records, Regional wide health data, including Primary Care, hospital, emergency, pharmacy and biobank

Main financial support

National, regional, or municipal public funding

Care setting

Hospital inpatient care

Hospital outpatient care

Primary care – GP, community pharmacist level

Primary care – specialist level (e.g. paediatricians)

Secondary care – specialist level (ambulatory)

Data source qualification

If the data source has successfully undergone a formal qualification process (e.g., from the EMA, ISO or other certifications), this should be described.

No

Data source website

https://bigan.iacs.es

Contact details

Carlos Tellería Orriols bigan.iacs@aragon.es

Main

bigan.iacs@aragon.es

Data source regions and languages

Data source countries

Spain

Data source languages

Spanish

Data source regions

Aragón

Data source establishment

Data source established

15/06/2018

Data source time span

First collection: 01/01/1996

The date when data started to be collected or extracted.

Publications

Data source publications

Changes in severity, mortality, and virus genome among a Spanish cohort of patients hospitalized with SARS-CoV-2

Construction of Empirical Care Pathways Process Models From Multiple Real-World Datasets

Analysis of the impact of social determinants and primary care morbidity on population health outcomes by combining big data: A research protocol

Analysis of Clinical Parameters, Drug Consumption and Use of Health Resources in a Southern European Population with Alcohol Abuse Disorder during COVID-19 Pandemic

Data elements collected

Disease information

Does the data source collect information with a focus on a specific disease? This might be a patient registry or other similar initiatives.

Yes

Disease details

Colon cancer

Disease details (other)

Chronic Kidney Disease

Rare diseases

Are rare diseases captured? In the European Union a rare disease is one that affects no more than 5 people in 10,000.

Yes

Pregnancy and/or neonates

Does the data source collect information on pregnant women and/or neonatal subpopulation (under 28 days of age)?

Yes

Hospital admission and/or discharge

Yes

ICU admission

Is information on intensive care unit admission available?

Yes

Cause of death

Not Captured

Prescriptions of medicines

Captured

Prescriptions vocabulary

ATC

other

Dispensing of medicines

Captured

Dispensing vocabulary

ATC

other

Advanced therapy medicinal products (ATMP)

Is information on advanced therapy medicinal products included? A medicinal product for human use that is either a gene therapy medicinal product, a somatic cell therapy product or a tissue engineered products as defined in Regulation (EC) No 1394/2007 [Reg (EC) No 1394/2007 Art 1(1)].

No

Contraception

Is information on the use of any type of contraception (oral, injectable, devices etc.) available?

No

Indication for use

Does the data source capture information on the therapeutic indication for the use of medicinal products?

Captured

Indication vocabulary

ICD-10

Medical devices

Is information on medicinal devices (e.g., pens, syringes, inhalers) available?

Yes

Administration of vaccines

Yes

Procedures

Does the data source capture information on procedures (e.g., diagnostic tests, therapeutic, surgical interventions)?

Captured

Procedures vocabulary

ICD-10

SNOMED CT

Healthcare provider

Is information on the person providing healthcare (e.g., physician, pharmacist, specialist) available? The healthcare provider refers to individual health professionals or a health facility organisation licensed to provide health care diagnosis and treatment services including medication, surgery and medical devices.

Yes

Clinical measurements

Is information on clinical measurements (e.g., BMI, blood pressure, height) available?

Yes

Genetic data

Are data related to genotyping, genome sequencing available?

Not Captured

Biomarker data

Does the data source capture biomarker information? The term “biomarker” refers to a broad subcategory of medical signs (objective indications of medical state observed from outside the patient), which can be measured accurately and reproducibly. For example, haematological assays, infectious disease markers or metabolomic biomarkers.

Captured

Patient-reported outcomes

Is information on patient-reported outcomes (e.g., quality of life) available?

No

Patient-generated data

Is patient-generated information (e.g., from wearable devices) available?

No

Units of healthcare utilisation

Are units of healthcare utilisation (e.g., number of visits to GP per year, number of hospital days) available or can they be derived? Units of healthcare utilisation refer to the quantification of the use of services for the purpose of preventing or curing health problems.

Yes

Unique identifier for persons

Are patients uniquely identified in the data source?

Yes

Diagnostic codes

Captured

Diagnosis / medical event vocabulary

ICD-10

ICPC-1

SNOMED CT

Medicinal product information

Captured

Medicinal product information collected

Active ingredient(s)

Brand name

Dosage regime

Dose

Formulation

Medicinal product vocabulary

ATC

Other

Quality of life measurements

Captured

Quality of life measurements vocabulary

Not coded (Free text)

Lifestyle factors

Captured

Lifestyle factors

Alcohol use

Frequency of exercise

Tobacco use

Sociodemographic information

Captured

Sociodemographic information collected

Age

Country of origin

Deprivation index

Gender

Health area

Living in rural area

Pharmaceutical copayment

Quantitative descriptors

Population Qualitative Data

Population age groups

Paediatric Population (< 18 years)

Preterm newborn infants (0 – 27 days)

Term newborn infants (0 – 27 days)

Infants and toddlers (28 days – 23 months)

Children (2 to < 12 years)

Adolescents (12 to < 18 years)

Adults (18 to < 46 years)

Adults (46 to < 65 years)

Elderly (≥ 65 years)

Adults (65 to < 75 years)

Adults (75 to < 85 years)

Adults (85 years and over)

Estimated percentage of the population covered by the data source in the catchment area

99%

Description of the population covered by the data source in the catchment area whose data are not collected (e.g., people who are registered only for private care)

The data source covers all the population covered by Public Health System (virtually, all the population), and all the activities in public centers, plus hospitalizations in private hospitals

Family linkage

Family linkage available in the data source permanently or can be created on an ad hoc basis

Ad hoc

Population

Population size

2244921

Active population size

2133897

Population by age group

Age group	Population size	Active population size
Paediatric Population (< 18 years)	255001	225927
Preterm newborn infants (0 – 27 days)	17	17
Term newborn infants (0 – 27 days)	204	204
Infants and toddlers (28 days – 23 months)	17079	16910
Children (2 to < 12 years)	130597	121180
Adolescents (12 to < 18 years)	107104	87633
Adults (18 to < 46 years)	660027	471363
Adults (46 to < 65 years)	553164	408568
Elderly (≥ 65 years)	753046	307171
Adults (65 to < 75 years)	204384	148500
Adults (75 to < 85 years)	165027	100386
Adults (85 years and over)	407318	58285

Median observation time

Median time (years) between first and last available records for unique individuals captured in the data source

52.00

Median time (years) between first and last available records for unique active individuals (alive and currently registered) capt

46.00

Data flows and management

Biospecimen access

Are biospecimens available in the data source (e.g., tissue samples)?

Yes

Biospecimen access conditions

https://www.iacs.es/servicios/biobanco/biobanco-solicitar-muestras-y-servicios/

Access to subject details

Can individual patients/practitioners/practices included in the data source be contacted?

No

Description of data collection

A collection of orchestred ETL processes are executed daily against Clinical Information Systems in Aragon Public Health Service. Data is pseudonymised in origin, and staged in BIGAN's staging repository. Afterwards, a series of recoding, data curation, data linkage and quality improvement processes are executed to final storage in a Common Data Model

Event triggering registration

Event triggering registration of a person in the data source

Birth

Immigration

Residency obtained

Event triggering de-registration of a person in the data source

Death

Emigration

Loss to follow up

Other

Event triggering de-registration of a person in the data source, other

Records are not deleted from repository. When a patient dies or emigrates, its personal record is marked as not-active, but is preserved for retrospective analysis

Event triggering creation of a record in the data source

Records are created in Clinical Information Systems during daily clinical practice and health administrative processes. There is no process to create records directly in BIGAN repository.

Data source linkage

Linkage

Is the data source described created by the linkage of other data sources (prelinked data source) and/or can the data source be linked to other data source on an ad-hoc basis?

Yes

Linkage description, pre-linked

All the information systems share a unique person code for each citizen. BIGAN data source generates a pseudonym code by encrypting the citizen's ID.

Linked data sources

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Chronic Kidney Disease register

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Colon Cancer Register

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

e-Prescription

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Laboratory IS (from 10 public hospitals)

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Newborn screening register

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Regional Electronic Health Record

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Regional Radiology Ring

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Pre linked

Is the data source described created by the linkage of other data sources?

Yes

Data source, other

Users Database (health system cover-up)

Linkage variable

Citizen's ID

Linkage completeness

Almost complete, except some private hopital patients not correctly identified

Data management specifications that apply for the data source

Data source refresh

Monthly

Informed consent for use of data for research

Other

Possibility of data validation

Can validity of the data in the data source be verified (e.g., access to original medical charts)?

Yes

Data source preservation

Are records preserved in the data source indefinitely?

Yes

Approval for publication

Is an approval needed for publishing the results of a study using the data source?

Yes

Informed consent, other

Required for studies with small number of patients or if aditional information from EHR is needed

Data source last refresh

12/04/2023

Common Data Model (CDM) mapping

CDM mapping

Has the data source been converted (ETL-ed) to a common data model?

Yes

CDM Mappings

CDM name

OMOP

CDM website

https://www.ohdsi.org/Data-standardization/

Data source ETL CDM version

5.3

Data source ETL frequency

1,00 month

Data source ETL status

Completed