

# Lithuanian National Data Lake

**First published:** 14/08/2025

**Last updated:** 28/05/2026

**Data source**

Human

Administrative healthcare records (e.g., claims)

Birth registry

Death registry

Diagnostic tests or procedures reimbursement

Drug registry

Hospital discharge records

Hospital inpatient records

Hospital outpatient visit records

Other

Pharmacy dispensing records

Population registry

Primary care medical records

Registration with healthcare system

Specialist ambulatory care records

Vaccination registry

## Administrative details

### Administrative details

#### Data source ID

1000000683

#### Data holder

[State Data Agency \(SDA\)](#)

#### Data source type

Administrative healthcare records (e.g., claims)

Birth registry

Death registry

Diagnostic tests or procedures reimbursement

Drug registry

Hospital discharge records

Hospital inpatient records

Hospital outpatient visit records

Other

Pharmacy dispensing records

Population registry

Primary care medical records

Registration with healthcare system

Specialist ambulatory care records

Vaccination registry

---

### **Data source type, other**

Various data collected by the state can be linked (demographics, socio-economics, education etc.)

---

### **Main financial support**

European public funding

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

[Health Data website \(State Data Agency\)](#)

## Contact details

Julius Juodakis [julius.juodakis@stat.gov.lt](mailto:julius.juodakis@stat.gov.lt)

Main

[julius.juodakis@stat.gov.lt](mailto:julius.juodakis@stat.gov.lt)

## Data source regions and languages

### Data source countries

Lithuania

---

### Data source languages

Lithuanian

## Data source establishment

### Data source established

01/07/2022

---

### Data source time span

**First collection:** 01/01/2015

The date when data started to be collected or extracted.

# Publications

## Data source publications

[Data Resource Profile: The Lithuanian health data reuse pathway](#)

## Data elements collected

### The data source contains the following information

#### **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.

No

---

#### **Disease details (other)**

undefined

---

#### **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**

Captured

---

### **Cause of death vocabulary**

ICD-10

Other

---

### **Cause of death vocabulary, other**

ICD-10-AM (Australian modification); should be largely similar to ICD-10

---

### **Prescriptions of medicines**

Captured

---

### **Prescriptions vocabulary**

ATC

---

### **Dispensing of medicines**

Captured

---

### **Dispensing vocabulary**

ATC

---

### **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)].

Yes

---

## **Contraception**

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

Yes

---

## **Indication for use**

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

Captured

---

## **Indication vocabulary**

ICD-10

Other

---

## **Indication vocabulary, other**

ICD-10-AM

---

## **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**

Other

---

## **Procedures vocabulary, other**

ACHI and national vocabularies

---

## **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

Other

---

### **Diagnosis / medical event vocabulary, other**

ICD-10-AM

---

### **Medicinal product information**

Captured

---

### **Medicinal product information collected**

Active ingredient(s)

Brand name

Dosage regime

Dose

Formulation

Package size

Route of administration

Strength

---

### **Medicinal product vocabulary**

ATC

---

### **Quality of life measurements**

Not Captured

---

### **Lifestyle factors**

Not Captured

---

### **Sociodemographic information**

Captured

---

### **Sociodemographic information collected**

Age

Country of origin

Education level

Living in rural area

Marital status

Other

Pharmaceutical copayment

Sex

---

### **Sociodemographic information other**

various demographics from national data sources

## Quantitative descriptors

### Population Qualitative Data

#### **Population age groups**

All

In utero

Paediatric Population (< 18 years)

Neonate

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)

Adult and elderly population ( $\geq 18$  years)

Adults (18 to < 65 years)

Adults (18 to < 46 years)

Adults (46 to < 65 years)

Elderly ( $\geq 65$  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**

Near-complete, but varies by dataset (e.g. almost all hospitalizations, reimbursed prescriptions, but not all elective procedures)

---

**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)**

Varies by dataset. Most prescribed drugs and reimbursed procedures should be covered. See accompanying publication for details

## Family linkage

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

Ad hoc

## Population

**Population size**

3700000

---

**Active population size**

2900000

## Data flows and management

## Access and validation

**Governance details**

Documents or webpages that describe the overall governance of the data source and processes and procedures for data capture and management, data quality check and validation results (governing data access or utilisation for research purposes).

[About health data at the State Data Agency](#)

[Catalog of health data available at the State Data Agency \(in Lithuanian\)](#)

### **Biospecimen access**

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

No

---

### **Access to subject details**

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

No

---

### **Description of data collection**

All data available in national IT systems and registries of Lithuania can be retrieved, linked and provided for use, following the national Law for Health Data Reuse.

## **Event triggering registration**

### **Event triggering registration of a person in the data source**

Birth

Disease diagnosis

Immigration

Insurance coverage start

Other

Practice registration

Residency obtained

Start of treatment

---

**Event triggering registration of a person in the data source, other**  
varies by dataset

---

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

Death

Emigration

End of treatment

Insurance coverage end

Other

Practice deregistration

---

**Event triggering de-registration of a person in the data source, other**  
varies by dataset

---

**Event triggering creation of a record in the data source**  
varies by dataset

## 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

Individuals are identified by their personal codes and can be tracked throughout various data tables. Personal codes are pseudonymized before data provision.

## Data management specifications that apply for the data source

**Data source refresh**

Monthly

---

**Informed consent for use of data for research**

Not Required

---

**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?

No

## Common Data Model (CDM) mapping

**CDM mapping**

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

No