National Health Insurance Research Database

First published: 23/10/2025

Last updated: 24/10/2025

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

Administrative details

Administrative details

Data source ID

1000000784

Data source acronym

NHIRD

Data holder

Population Health Data Center (PHDc), National Cheng Kung University (NCKU)

Data source type

Administrative healthcare records (e.g., claims)

Main financial support

Funding from industry or contract research

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.

Yes

Description of the qualification

All datasets are encrypted and verified by the Department of Statistics before being transferred to the MOHW Data Science Center for research use. Researchers must obtain approval and conduct analyses only at certified on-site Research Centers that comply with ISO 27001 information security management standards. The certification ensures a robust information security management system encompassing data encryption, user authentication, access control, audit logging, and continuous risk assessment, with regular third-party audits to ensure compliance. All data are de-identified, securely stored, and accessed within a controlled environment to uphold confidentiality, integrity, and accountability.

Data source website

National Health Insurance Research Database

Contact details

Edward Chia-Cheng Lai edward_lai@mail.ncku.edu.tw

Main

edward_lai@mail.ncku.edu.tw

Daniel Tsai danielhttsai@gmail.com

Alternate

danielhttsai@gmail.com

Data source regions and languages

Data source countries

Taiwan

Data source languages

Chinese

English

Data source establishment

Data source established

01/01/2000

Data source time span

First collection: 01/01/2000

The date when data started to be collected or extracted.

Publications

Data source publications

Taiwan’s National Health Insurance Research Database: past and future

Cross-Regional Data Initiative for the Assessment and Development of Treatment for Neurological and Mental Disorders

Brief report: databases in the Asia-Pacific region: the potential for a distributed network approach

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

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)].

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-CM

ICD-9-CM

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

Domestic codes

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?

No

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.

Not 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-CM

ICD-9-CM

Medicinal product information

Captured

Medicinal product information collected

Active ingredient(s)

Brand name

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

Living in rural area

Sex

Socioeconomic status

Type of residency

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 (≥18 years)

Adults (18 to < 65 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

99.9%

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)

<0.1%

Family linkage

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

Permanently

Family linkage available between the following persons

Father-child

Mother-child

Sibling

Population

Population size

23000000

Active population size

23000000

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).

https://www.apre.mohw.gov.tw/

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

The Ministry of Health and Welfare (MOHW) in Taiwan integrates health-related data from several agencies, including the Health Promotion Administration,

Centers for Disease Control, National Health Insurance Administration, and Social and Family Affairs Administration. These datasets are encrypted and verified by the Department of Statistics before being transferred to the MOHW Data Science Center for research use.

Researchers must submit applications for approval, and analyses can only be conducted at certified on-site Research Centers that comply with ISO 27001 information security standards. All data are de-identified, securely stored, and accessed within a controlled environment to ensure privacy protection and data integrity.

Event triggering registration

Event triggering registration of a person in the data source

Birth

Insurance coverage start

Residency obtained

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

Death

Emigration

Insurance coverage end

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

In NHIRD, multiple nationwide registries provide high-quality data for causal inference and pharmacoepidemiologic research, including the Birth Certificate Application Database (mother-baby linkage), Taiwan Cancer Registry Database, Rare Disease Database, COVID-19 Vaccine and Record Database, Assisted Reproduction Database, Registry for Catastrophic Illness Patients, Death Registry, Traditional Chinese Medicine Database, National Health Interview Survey. Each registry contains standardized, disease- or treatment-specific variables and is pre-linked to the NHIRD through encrypted personal identifiers.

Linkage description, possible linkage

The Taiwan Biobank (TWB) is a large-scale national biomedical repository integrating genetic, clinical, and lifestyle information from both community-based and hospital-based cohorts. It contains demographic data, medical histories, biospecimens (blood, urine, saliva), biochemical markers (e.g., lipids, HbA1c, liver and kidney function), and whole-genome sequencing data. TWB's design enables potential linkage—with appropriate ethical approval—to other insurance or health registries, facilitating real-world genomic and precision medicine studies.

Domain-specific databases such as the Institute of Forensic Medicine Autopsy Database and the Taiwan Illicit Drug Issue Database (TIDID) provide high-quality data on cause of death, toxicology, and substance use patterns. The autopsy database includes demographic and toxicological profiles of individuals undergoing forensic examination, while the TIDID and its complementary Administrative Penalty System records document arrests and verified cases of illicit substance use.

Data management specifications that apply for the data source

Data source refresh

Yearly

Informed consent for use of data for research

Waiver

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

Data source preservation length (years)

2.5 years

Approval for publication

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

Yes

Data source last refresh

01/01/2025

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 status

In progress

CDM name

Sentinel

CDM website

https://www.sentinelinitiative.org/methods-Data-tools/sentinel-common-Data-model

Data source ETL status

In progress