Clinical Epidemiology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

REVIEW

Taiwan's National Health Insurance Research Database (NHIRD): in the Era of Artificial Intelligence, Causal Inference, and Data Security

Cheng-Yang Hsieh[1,2], Shih-Chieh Shao[2–4], Sheng-Feng Sung[5,6], Miyuki Hsing-Chun Hsieh[2,3], Daniel Hsiang-Te Tsai[2,3], Swu-Jane Lin[7], Edward Chia-Cheng Lai[2,3]

[1]Department of Neurology, Tainan Sin Lau Hospital, Tainan, Taiwan; [2]School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [3]Population Health Data Center, National Cheng Kung University, Tainan, Taiwan; [4]Department of Pharmacy, Keelung Chang Gung Memorial Hospital, Keelung, Taiwan; [5]Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan; [6]Department of Nursing, Fooyin University, Kaohsiung, Taiwan; [7]Department of Pharmacy Systems, Outcomes and Policy, University of Illinois Chicago, Chicago, IL, USA

Correspondence: Edward Chia-Cheng Lai, School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan, Tel +886-6-2353535,ext.6209, Email edward_lai@mail.ncku.edu.tw

Background: Taiwan's National Health Insurance Research Database (NHIRD) has evolved into a cornerstone of real-world evidence generation. As Taiwan's National Health Insurance program reaches its 30th anniversary, a comprehensive reassessment of the NHIRD's development, challenges and future directions is warranted.

Objective: To provide an updated review of the NHIRD.

Methods: We conducted a narrative review of Taiwan's NHIRD, synthesizing published studies, government reports, and policy documents from 2019 through 2024. We summarized developments related to database infrastructure, data collection, validation studies, linkage strategies, governance reforms and interoperability initiatives, with a particular focus on their implications for real-world evidence generation and AI-driven research.

Results: The NHIRD has additionally incorporated structured laboratory results and medical imaging data, significantly broadening its research capabilities. Validation studies have demonstrated the reliability of International Classification of Diseases, 10th Revision, Clinical Modification codes across various conditions, reinforcing the database's applicability for epidemiological research. Integration efforts with national registries, surveys and electronic medical records have further enhanced the depth and accuracy of clinical outcome measurements. Nonetheless, critical challenges persist, including data standardization inconsistencies, cybersecurity vulnerabilities, and heightened scrutiny following constitutional court rulings on data governance and privacy rights. In response, Taiwan's Ministry of Health and Welfare has launched initiatives to address these concerns, notably through the development of a Fast Healthcare Interoperability Resources (FHIR)-based data infrastructure aimed at improving interoperability, data security, and artificial intelligence (AI)-readiness to balance ethical governance with scientific innovation.

Conclusion: The NHIRD's transformation over three decades underscores the importance of continuous investment in data quality, privacy protection, and interoperability. With sustained reforms, the NHIRD will be poised to remain a leading resource for real-world evidence generation and to contribute meaningfully to global health and digital medicine.

Keywords: National Health Insurance Research Database, health claims data, real-world evidence, laboratory data, medical imaging, ICD-10 validation, data linkage

Introduction

Healthcare is exceptionally poised for innovation, driven by its substantial resource demands and the growing burden of aging and expanding populations worldwide.[1] Amid these pressures, recent technological advancements have propelled artificial intelligence (AI) to the forefront of the healthcare transformation.[2] AI's capacity to extract meaningful patterns from large-scale biomedical data offers new opportunities for personalized care, predictive modeling, and improved

Received: 28 July 2025
Accepted: 12 November 2025
Published: 25 November 2025

system efficiency.[3] The transformative potential of medical AI is intrinsically dependent on access to large-scale, high-quality data, which serves as an essential resource in this area.[4] Central to this promise is the availability of structured, population-level health data, often derived from national health insurance claims databases.[5]

In this context, Taiwan's National Health Insurance Research Database (NHIRD), established in conjunction with the single-payer National Health Insurance (NHI) program initiated in 1995, exemplifies this resource-rich ecosystem.[6–8] Taiwan's NHI program provides universal health coverage, ensuring equitable access to healthcare for more than 99% of the population. As of 2024, the NHIRD contains over 70 billion medical records and 3.4 billion medical images,[9] and has supported a total of at least 8,414 peer-reviewed publications.[8] Its comprehensive population coverage, longitudinal follow-up, and integration with diverse clinical datasets make the NHIRD uniquely positioned to support real-world evidence generation and advance health technology research at scale.[6] Moreover, the NHIRD's scale and structure make it not only a critical national asset but also a model for other health systems seeking to build AI-ready, population-level data platforms.

Yet, despite this impressive academic output, the NHIRD is not without its challenges. Heightened scrutiny of privacy protection could hinder its use in academic research.[10] Furthermore, the NHI information system urgently needs upgrades, including improvements to its infrastructure, better data consistency and interoperability, and enhanced cybersecurity measures. At the same time, the escalating frequency and sophistication of cyberattacks targeting healthcare systems worldwide underscore the importance of robust cybersecurity measures, which are essential for safeguarding sensitive population-level data and maintaining public trust in large-scale digital health infrastructures. As Taiwan's NHI marks its 30th anniversary,[11] a timely reassessment of the NHIRD is warranted. This review differs from our previous review[6] by providing the most comprehensive and forward-looking update on the NHIRD to date, incorporating several transformative developments that were not covered previously, including large-scale integration of laboratory and imaging data, rigorous ICD-10 validation studies, linkage with external registries and electronic medical records, and the adoption of interoperability standards designed to support AI applications and real-world evidence generation.

Database Profiles

To fully appreciate the research potential and operational constraints of the NHIRD, it is essential to understand its evolving architecture, governance transitions, and data access mechanisms. The core architecture of the NHIRD has been described in prior literature.[6,12] Briefly, starting in 1998, Taiwan's Bureau of National Health Insurance (BNHI) and its successor, the National Health Insurance Administration (NHIA), have commissioned the National Health Research Institutes (NHRI) to establish and maintain the NHIRD. Since 2000, the NHIRD has been updated annually, incorporating claims data submitted to the NHI by healthcare institutions. Data from the preceding year were initially released at the end of each year and made available to researchers through physical storage devices such as compact discs. However, in June 2016, the Ministry of Health and Welfare (MOHW) terminated the NHIRD services that were previously managed by the NHRI. The database was subsequently placed under the centralized management of the Department of Statistics and the Information Management Center within the MOHW, in compliance with the policy on data protection, whereby all data could only be accessed within the ministry's premises and their removal from the institution was strictly prohibited.[6,12]

With the growing demand for big data research, several challenges have arisen, including a limited number of independent, on-site workstations and difficulties in securing access reservations. In response, a remote virtual desktop infrastructure was established at ten sub-centers across the country in 2016, enabling users to access full-population data files. This system allows researchers to operate remotely and analyze data by connecting via an intranet to the Taipei-based Department of Statistics system.[13]

Currently, each center provides access to at least 65 databases (Supplementary Table 1), which are categorized into three major types:[14]

1. Full-population databases, which include all claims data from Taiwan's NHI system.
2. Longitudinal Health Insurance Database (LHID), a representative 2-million-patient subset.
3. Publicly accessible, de-identified survey datasets.

Full-Population Databases

The full-population datasets of the NHIRD encompass the complete claims data submitted by all healthcare providers contracted under Taiwan's National Health Insurance program. These files include detailed records on ambulatory care, inpatient admissions, pharmacy dispensing, procedures and diagnoses, linked through encrypted personal identifiers. Covering over 99% of the Taiwanese population, these datasets enable comprehensive longitudinal analyses across diverse patient populations, disease areas, and healthcare settings. However, under current regulations, researchers may typically access only up to 10% of the population's data per project, unless granted special authorization. Despite this limitation, the scale and completeness of these datasets remain uniquely valuable for studies requiring national representativeness, rare disease surveillance, or long-term outcome tracking.

LHID

The LHID was created using household registration records, with insured patients from the NHI system in 2005, 2010 and 2015 as the source population. The dataset was selected through stratified random sampling based on age, sex and residential area. As a preliminary resource, the LHID offers a more accessible option than the full-population health and welfare data. Due to shorter application processing times and lower costs, the LHID is particularly suitable for research projects with limited budgets. Additionally, its smaller file size enables faster data processing and analysis, leading to quicker research outcomes.

Public Survey Datasets

Furthermore, each center provides access to publicly available statistical survey datasets, such as the National Health Interview Survey and the Nutrition and Health Survey in Taiwan, for use outside the center. These datasets have been anonymized using privacy-preserving feature masking techniques (e.g., generalization, suppression or aggregation) to remove or obscure identifiable personal information, ensuring that researchers can access and utilize them outside the data centers while maintaining data privacy and security.

Existing Laboratory Database

The integration of laboratory data into Taiwan's NHIRD represents a pivotal advancement in its research infrastructure. In 2014, the NHIA mandated the collection of laboratory test results and medical examination reports, with access for research purposes commencing in 2017. This expansion initially aimed to support pay-for-performance programs, reduce redundant testing through cross-hospital data sharing, and enhance data richness for clinical research. Before laboratory data integration, the NHIRD contained only billing claims and diagnostic coding data, without access to clinical test results, thereby limiting its ability to assess disease severity and clinical outcomes. Originally intended to support billing and healthcare management programs, the integration of laboratory data into the NHIRD has had an important unintended consequence: it has allowed researchers to access objective clinical measurements, thereby enhancing evaluations of treatment effectiveness and enabling more precise risk stratification.[15]

Content and Structure of the Laboratory Database

Laboratory data in the NHIRD are collected from a broad spectrum of healthcare facilities, including 25 medical centers, 82 regional hospitals, 366 district hospitals, and more than 10,000 primary care clinics. Test results are transmitted in a structured format via a secure virtual private network (VPN) to the National Health Insurance Administration. Initially limited to 72 test items, the scope of data collection had expanded to 641 laboratory tests by 2020. As of August 2020, the database contained more than 5.6 billion test results.[15]

The laboratory database covers a wide range of domains, including hematologic, biochemical, immunological and microbiological assays. Each test record includes 35 standardized fields, such as patient identifiers, test names, numerical results, measurement units, reference ranges, and timestamps. These records are linkable to inpatient and outpatient claims through encrypted identifiers, enabling researchers to integrate laboratory values with diagnostic, procedural and medication data. This integration facilitates disease phenotyping, classification of clinical severity, medication monitoring, and the development of predictive models for clinical outcomes.

Data Quality and Challenges

While the NHIRD laboratory database has greatly expanded the scope of research using real-world data, it faces persistent challenges related to data consistency and interoperability. One of the most pressing issues is the variation in laboratory coding practices across hospitals. Identical tests may be recorded under different codes or names, and the measurement of an analyte may be reported using varying units of measurement. For example, serum creatinine may be expressed in either mg/dL or μmol/L, depending on the institution.

These discrepancies hinder cross-institutional comparability and require researchers to develop and validate customized algorithms to harmonize results of laboratory tests. Standardizing test names, result formats, and units is essential to ensuring data quality and analytical reproducibility. To address these challenges, the NHIA has promoted collaborative solutions, such as encouraging researchers to share harmonization algorithms and reporting methodologies. While these efforts represent important progress, the burden of standardization continues to fall largely on individual research teams, posing barriers to scaling research across institutions or disease domains.

Access and Utilization

Access to the NHIRD laboratory database requires prior approval from an Institutional Review Board and submission of a data request to the Ministry of Health and Welfare. For domestic researchers, data access is provided through the Applied Health Research Data Integration Service, which enables remote analysis via secure virtual desktop infrastructure. International researchers may obtain access through formal collaboration with Taiwanese research institutions.

Data access is fee-based, with costs determined by the number of variables requested and the duration of use. Researchers can request customized datasets by selecting specific laboratory items relevant to their study objectives. This flexibility allows investigators to efficiently tailor the data to their research designs while minimizing unnecessary data exposure.

The availability of laboratory results—linked to claims-based diagnoses, treatments and outcomes—has significantly expanded the scope of real-world evidence studies in Taiwan. For instance, in a nationwide cohort study involving Taiwanese patients with diabetic kidney disease, sodium-glucose co-transporter-2 inhibitors were found to significantly reduce the risk of adverse cardiorenal outcomes, including kidney failure and stroke, compared to dipeptidyl peptidase-4 inhibitors.[16] The study included patients with diabetic kidney disease, and kidney failure outcomes were defined using values of estimated glomerular filtration rate obtained from the outpatient laboratory database.

Medical Imaging

In 2018, the National Health Insurance Administration (NHIA) implemented a policy requiring all contracted hospitals to upload medical imaging data—including both radiology reports and source image files—to the national health information platform. This policy was introduced primarily to support Taiwan's accredited healthcare system by enabling seamless information exchange during referrals and by reducing redundant imaging.

Similarly, an unintended but highly beneficial consequence was that this policy led to the development of Taiwan's largest centralized medical imaging database. The medical imaging dataset includes diverse modalities such as computed tomography, magnetic resonance imaging, X-rays and ultrasound, along with associated narrative reports. These are stored in a centralized system and are accessible under secure, supervised research protocols. Linked with administrative claims, laboratory results, and outcome data, this imaging repository has opened new frontiers for research in health policy evaluation, diagnostic innovation, and artificial intelligence.[17,18]

A growing number of studies have leveraged these imaging data to evaluate both clinical outcomes and digital tools. For example, a nationwide propensity score–matched cohort study assessed the impact of a multidisciplinary post-acute care program for patients with heart failure.[9] To create a comparable control group, the authors extracted left ventricular ejection fraction data from the NHIRD using an artificial intelligence–derived natural language processing algorithm, ensuring that both the intervention and control groups demonstrated comparable severity of heart failure.

In another application, researchers developed a deep learning–based computer-aided detection system for pancreatic cancer using contrast-enhanced CT images.[19] In validation testing, the model demonstrated high sensitivity and specificity even for small tumors (<2 cm), with validation conducted using a nationwide control group extracted from

the NHIRD's imaging dataset. These applications underscore the NHIRD's evolving capacity as a multimodal data platform, supporting scalable imaging research and the development of AI models at a national level.

Validation Studies

In our previous review, we reported modest to high sensitivity and positive predictive values (PPVs) of ICD-9-CM diagnosis codes in Taiwan's NHIRD for the identification of conditions such as epilepsy, ischemic stroke, hypertension, diabetes, hyperlipidemia, atrial fibrillation, and cancer.[6] Since the nationwide transition to ICD-10-CM coding in 2016, a growing body of literature has focused on validating ICD-10-based diagnoses, with most studies published from 2020 onward.[20–34]

These validation studies have assessed a wide spectrum of disease domains, including neurological (e.g., stroke, Guillain-Barré syndrome, cerebral venous sinus thrombosis), cardiovascular (e.g., acute myocardial infarction, atrial fibrillation, out-of-hospital cardiac arrest), immunologic (e.g., anaphylaxis), infectious diseases (e.g., hepatitis B and C), and accident (e.g., carbon monoxide poisoning).[21,23,29,32,33] Across studies, the sensitivity and PPV of ICD-10 codes vary by disease condition and setting, but are generally acceptable for epidemiological research. For example, the PPV for acute ischemic stroke reached 92.7%,[33] while that for glaucoma was 92.2%.[21]

In diseases with both ICD-9 and ICD-10 validation studies—such as stroke, myocardial infarction, glaucoma and out-of-hospital cardiac arrest—diagnostic accuracy has remained comparable across coding systems. This supports the continued use of the NHIRD for longitudinal research despite the coding shift.[35]

A summary of ICD-10 validation studies, including data sources, reference standards, and diagnostic accuracy metrics, is provided in the accompanying Table (Table 1). Collectively, these studies reinforce the reliability of the NHIRD for real-world evidence generation and health services research in the ICD-10 era. Nevertheless, certain disease conditions—particularly rare disorders and mental health diagnoses—remain without formal validation studies in the ICD-10 era, representing an important gap that warrants future methodological research.

Integration of the NHIRD with Other Databases

Linking the NHIRD with other datasets enables more comprehensive and clinically relevant research by addressing the inherent limitations of claims data—such as the lack of direct clinical measurements, lifestyle information, and patient-reported outcomes. Integration occurs primarily through two mechanisms: direct data linkage via encrypted personal identifiers and indirect triangulation strategies involving external datasets or federated network approach (Figure 1). Table 2 presents the summary of direct and indirect data linkage.

Direct Data Linkage

Direct linkage leverages unique encrypted personal IDs to integrate the NHIRD with more than 40 other national datasets managed by Taiwan's Health and Welfare Data Science Center.[6,36] These include vital registries (e.g., birth, death, cancer), administrative datasets (e.g., catastrophic illness certificates), and health services databases (e.g., Adult Preventive Health Services, the Artificial Reproduction Database, and the National Immunization Information System). Linkage with major national surveys—such as the National Health Interview Survey and Nutrition and Health Survey—further enhances analytical breadth.[6,36] By linking national databases, researchers can conduct studies with more comprehensive follow-up and outcome capture, particularly for critical endpoints such as mortality.[37]

Several studies have demonstrated the value of NHIRD direct linkage to the aforementioned databases. For example, Meng et al[38] and Chuang et al[39] linked the NHIRD with the Birth Reporting Database and Maternal and Child Health Database to investigate the association between maternal psychotropic medication use and adverse pregnancy and neonatal outcomes. Another study[40] combined the NHIRD with the Taiwan Cancer Registry and the Taiwan Cause of Death Database to recalibrate a Western lung cancer risk model using Taiwan's population data. Additionally, research[41] linking the NHIRD with the Nutrition and Health Survey found that optimal dietary intake was associated with lower medical service utilization among preschoolers in Taiwan.

The NHIRD data can also be directly linked to electronic medical records (EMR) to provide more in-depth healthcare data for patients of interest. Like other administrative datasets, the NHIRD has inherent limitations, particularly the lack of essential clinical outcomes such as bone mineral density, or stroke severity. These markers are essential for defining

Table 1 Summary of Validation Studies Assessing the Validity of ICD-10 Diagnosis Codes in Taiwan's NHIRD*

Diseases	Data Source	ICD-10 Codes	Reference Standard	Sensitivity, %	PPV %	References
Guillain-Barré syndrome[20]	Chang Gung Research Database during 2017–2022	G61.0	Chart reviews based on the diagnostic criteria established by the National Institute of Neurological Disorders and Stroke	Not reported	76.0 (76.3–83.6)	Clin Epidemiol. 2024:16:733-742
Glaucoma[21]	Chang Gung Research Database during 2011- 2020	H4010X, H4011X, H4012X, H4020X, H4021X, H4022X, H4023X, H4024X	Chart reviews based on visual field examinations	87.0 (81.1.-91.6)	92.2 (87.1–95.8)	Clin Epidemiol. 2024:16:227-234
Ankylosing spondylitis[22]	Chung Shan University Hospital	M45	Chart reviews based on the diagnostic criteria of 2009 ASAS guideline	Not reported	72.8 (66.8–78.8)	Int J Rheum Dis. 2024;27:e15041
STEMI and NSTEMI[23]	Chia-Yi Christian Hospital during 2006–2022	STEMI: I20.0, I21.1, I21.2, I21.3, I21.9 NSTEMI: I21.4	Chart reviews based on the diagnostic criteria of clinical practice guidelines	STEMI: 95.7 (94.0–97.0); NSTEMI: 82.6 (80.7–84.4)	STEMI: 86.6 (84.2–88.7); NSTEMI: 96.5 (95.3–97.4)	Clin Epidemiol. 2023:15:1027-1039
Myocarditis[24]	Chang Gung Research Database during 2017-March 31st, 2022	A381, A3952, B2682, B3320, B3322, B3324, B5881, D8685, I012, I090, I400, I401, I408, I409, I41, I514, J1082, J1182	Chart reviews based on diagnostic criteria from European Society of Cardiology	Not reported	73.5 (69.6–77.4)	Clin Epidemiol. 2023:15:459-468
Anaphylaxis[25]	Chang Gung Research Database during 2017-2021	T78.00XA, T78.01XA, T78.02XA, T78.03XA, T78.04XA, T78.05XA, T78.06XA, T78.07XA, T78.08XA, T78.09XA, T80.51XA, T80.52XA, T80.59XA, T88.6XXA, T78.2XXA	Chart reviews based on clinical criteria for the diagnosis of anaphylaxis, as proposed at the Second National Institute of Allergy and Infectious Disease / Food Allergy and Anaphylaxis Network	Not reported	79.4% (73.3–85.5)	J Med Syst. 2023;47(1):97
Incident hip fracture[26]	National Taiwan University Hospital-Integrated Medical Database during 2007–2020	S72.0, S72.1, S72.2	Chart reviews based on admission notes, surgical notes as well as X-ray, CT, and MRI	97.2 (94.9–99.4)	98.6 (97.0–100.0)	J Formos Med Assoc. 2023:122 Suppl 1:S82-S91
Cerebral venous sinus thrombosis[27]	Chang Gung Memorial Hospitals during 2017-2020	G08, I636, or I676	Chart reviews according to the definition of the Taiwan Stroke Registry program	Not reported	G08: 88.2 (78.1–94.8) I636: 100.0 (78.2–100.0) I676: 91.3 (84.1–95.9) G08, I636 or I676: 90.7 (85.5–94.5)	Clin Epidemiol. 2022:14 1–7
Carbon monoxide poisoning[28]	Four Chang Gung Memorial Hospitals during 2011–2020	T58	Chart reviews based on patients' exposure history and clinical signs or symptoms, in addition to the COHb levels	94.9 (89.2–98.1)	76.6 (68.8–83.2)	Front Pharmacol. 2022;13:882632
Out-of-hospital cardiac arrest[29]	Ditmanson Medical Foundation Chia-Yi Christian Hospital during 2010–2020	R99, I46, I46.2, I46.8, I46.9	Chart reviews based on clinical criteria of Utstein criteria	87.8 (85.9–89.7)	88.5 (86.4–90.3)	Clin Epidemiol. 2022;14:721-730
HBV infection[30]	Chi Mei health care system during 2008-2019	B16.0, B16.1, B16.2, B16.9, B17.0, B18.0, B18.1, B19.1, Z22.51	Laboratory test or textual diagnosis recorded in electrical medical records	56.0 (52.4–59.6)	88.5 (85.6–91.4)	BMC Infect Dis. 2022;22:222
HCV infection[30]		B17.1, B18.2, B19.2, Z22.52		61.7 (57.6–65.9)	96.1 (94.1–98.2)	

https://doi.org/10.2147/CLEP.S553894

Hypertension[31]	Ditmanson Medical Foundation Chia-Yi Christian Hospital and the E-Da Hospital during 2018-2020	I10, I11, I12, I13, I14, I15	The clinical diagnosis in the stroke registry	85.7 (84.5–86.8)	91.9 (90.9–92.8)	Clin Epidemiol. 2022;14:327-335
Diabetes[31]		E10, E11, E12, E13, E14		89.9 (88.4–91.2)	87.3 (85.7–88.7)	
Hyperlipidemia[31]		E78		86.9 (85.4–88.3)	69.9 (67.8–71.3)	
Atrial fibrillation[31]		I48		73.5 (69.9–77.1)	85.6 (82.1–88.7)	
Ischemic heart disease[31]		I20, I21, I22, I25		64.4 (58.9–69.6)	49.0 (44.2–53.9)	
Acute hemorrhagic Stroke[32]	Ditmanson Medical Foundation Chia-Yi Christian Hospital during 2018–2019	Acute hemorrhagic stroke: I60-I61 SAH: I60 ICH: I61	The clinical diagnosis in the stroke registry	Acute hemorrhagic stroke: 98.6 (97.6–99.3) SAH: 94.7 (89.8–97.7) ICH: 99.2 (98.4–99.7)	Acute hemorrhagic stroke: 88.6 (86.4–90.5) SAH: 88.2 (82.2–92.7) ICH: 90.8 (88.6–92.6)	Clin Epidemiol. 2021;13:43-51
Acute ischemic stroke[33]	E-Da Hospital during 2018–2019	I63	The clinical diagnosis in the stroke registry	99.4 (98.8–99.7)	92.7 (91.2–94.0)	Clin Epidemiol. 2020;12:1007-1013
HBV infection[34]	NHIRD during January – March, 2018	B180, B181, B191	National Health Insurance Lab & Exam Dataset	46.3 (not reported)	44.6 (not reported)	Clin Epidemiol. 2020;12:185–192
HCV infection[34]	NHIRD during January – March, 2018	B182, B192	National Health Insurance Lab & Exam Dataset	46.9 (not reported)	80.8 (not reported)	Clin Epidemiol. 2020;12:185-192

Notes: *We searched PubMed on May 1, 2025, using the terms *(validation OR validity OR positive predictive value OR PPV OR accuracy) AND (*"ICD-10" *OR* "ICD-10-CM"*) AND (*"National Health Insurance" *OR* "National Health Insurance Research Database" *OR* "NHIRD" *OR claims) AND Taiwan* to identify validation studies assessing the accuracy of ICD-10 diagnosis codes in Taiwan's National Health Insurance Research Database. **Abbreviations**: ICD-10, International Classification of Diseases, 10th Revision; NHIRD, National Health Insurance Research Database; PPV, Positive predictive value; CT, Computed tomography; MRI, Magnetic resonance imaging; COHb, Carboxyhemoglobin; HBV, Hepatitis B virus; HCV, Hepatitis C virus; ICH, Intracerebral hemorrhage; SAH, Subarachnoid hemorrhage; STEMI, ST-elevation myocardial infarction; NSTEMI, Non-ST-elevation myocardial infarction.

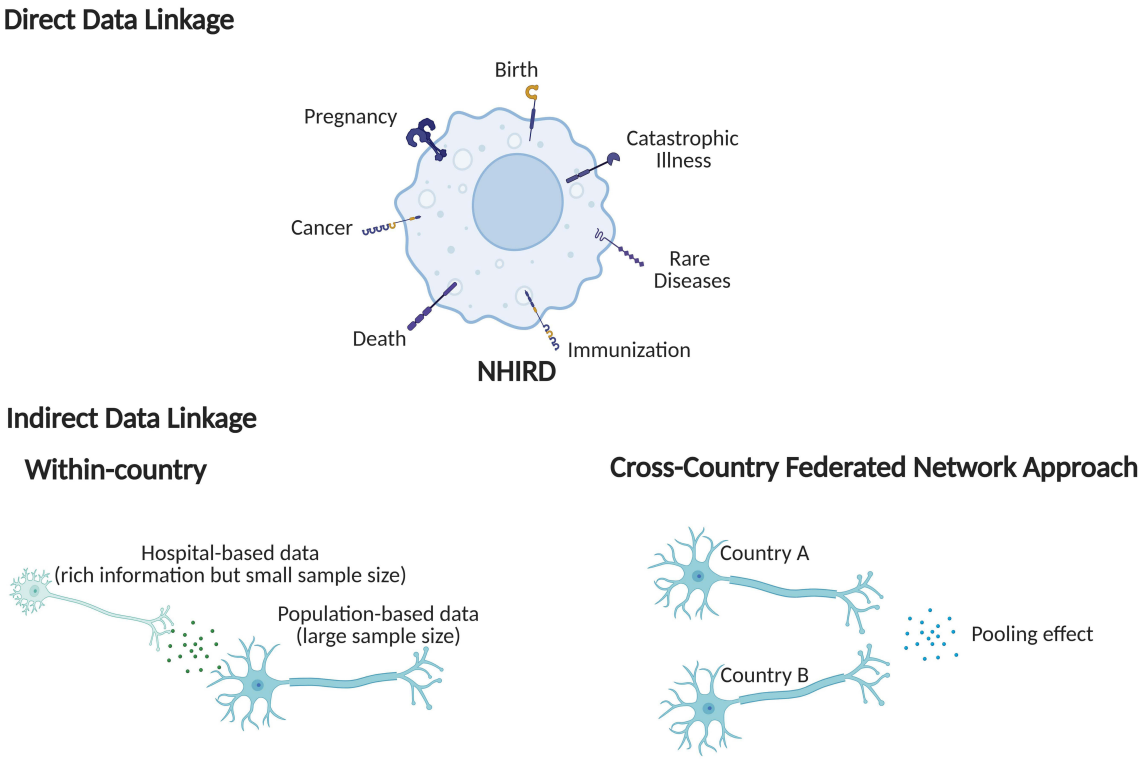

Figure 1 Overview of Integration Mechanisms for Enhancing the Research Utility of the National Health Insurance Research Database (NHIRD). Note: Created in BioRender. Tsai, (**D**) (2025) https://BioRender.com/7kvwv0z.

study populations, particularly when the exposures or outcomes of interest rely on laboratory values for ascertainment. Moreover, such data are critical for evaluating the effects of specific exposures on patient outcomes and for characterizing disease severity. By integrating the NHIRD with EMR data, researchers can enhance the validity of exposure and outcome assessments and adjust for important confounders that are often unmeasured in the NHIRD.[42]

Despite the available infrastructure, integration of the NHIRD with EMR remains limited. A study by Hsu et al[43] found that only 31% of patients in a linked EMR-NHIRD dataset had continuous follow-up over 30 months, and overall EMR completeness was 52%, highlighting key limitations of relying solely on hospital-based EMRs. These findings underscore the advantage of NHIRD linkage, which enhances data continuity and longitudinal validity. Currently, the Chang Gung Research Database (CGRD) accounts for approximately 10% of NHIRD enrollees and represents the most comprehensive EMR-linked resource.[44] However, researchers must remain mindful of selection bias and heterogeneous follow-up when interpreting results from integrated datasets.

Although direct linkage between EMR and the NHIRD is currently limited—likely due to challenges related to data privacy, standardization and technical barriers—some studies have employed alternative approaches to integrate information from both sources. For example, one study[45] cross-linked patient characteristics between hospital EMR and the NHIRD, despite the absence of a common ID, to validate in-hospital mortality data among patients with acute myocardial infarction or stroke. Alternative methodologies without direct linkage demonstrate the potential of leveraging both the NHIRD and EMR data to enhance the quality and depth of healthcare research, even in the absence of a fully integrated system.

Indirect Data Triangulation

While direct linkage involves merging internal datasets using unique identifiers, indirect data triangulation provides an alternative by integrating external data sources and conducting multinational studies to facilitate the research process. This approach is particularly valuable when privacy restrictions or structural differences in healthcare systems are challenging. Indirect data triangulation consists of two main strategies: external data integration and federated network approach. Using the appropriate study designs or statistical models, these data triangulations enable control of

Table 2 Summary of Direct and Indirect Data Linkage

	Direct Data Linkage	**Indirect Data Triangulation**	
		External Data Integration (Within-Country*)	**Federated Network Approach (Cross-Country**)**
Mechanisms	Link to internal data: other national registries; healthcare facility records such as EMR	Indirect link to external data: healthcare facility records such as EMR; clinical trial data	Integrate data with other international databases through privacy-protected methods.
Objectives	1) Enhance data capture for specific research needs. 2) Improve the validity of exposure and outcome assessments. 3) Control for unmeasured confounders.	(Same objectives as direct linkage)	1) Enable cross-region comparisons. 2) Increase statistical power. 3) Assess generalizability and reproducibility.
Examples	Linkages to registries for death, cancer, birth, catastrophic illnesses, rare diseases, and immunization.	Propensity score calibration / validation / external control	Analysis of prescribing patterns, drug effectiveness and safety, and policy research.

Notes: * For example, one study enhanced Taiwan's National Health Insurance Research Database (NHIRD) by integrating it with a multi-center stroke registry, allowing the researchers to use additional variables such as smoking status, BMI, and stroke severity. This integration enabled the use of propensity score calibration to adjust for unmeasured confounders, improving the validity of outcome assessments in the absence of detailed clinical data. **For instance, the Asian Pharmacoepidemiology Network (AsPEN) conducted a study on denosumab use in patients with bone metastases across Taiwan, Hong Kong, and South Korea. By comparing treatment persistence and re-initiation patterns across countries with different healthcare systems and prescribing habits, the study demonstrated the reproducibility and generalizability of findings across diverse Asian populations.

unmeasured confounders and allow for cross-national comparisons of drug effects and disease outcomes to strengthen causal inference. Additionally, they enable researchers to explore racial and cultural differences in health outcomes.

One mechanism of indirect data triangulation is the integration of external datasets from different sources within-country. Since claims data often lack some detailed clinical information, integrating them with EHRs can enhance the accuracy of exposure and outcome assessments.[46] For example, propensity score calibration can help account for unmeasured confounders by utilizing a cross-sectional validation study, even when the study does not include data on disease severity or other health factors.[47] One previous study employed propensity score calibration to integrate external data from a multi-center stroke registry with the NHIRD, incorporating additional confounders such as smoking status, body mass index, stroke severity, and disability.[48] Statisticians can apply regression models to calibrate the systematic difference in confounder measurement between the two datasets. Additionally, researchers have validated findings by using the CGRD to enhance findings from the NHIRD with clinical data.[44,49–51] This approach better adjusted for confounders and improved the validity of their analysis. Future work may include using the NHIRD as a source of external control, providing a comparison of treatments in the absence of a comparator group.[52] As single-arm trials play an increasingly significant role in regulatory approvals, integrating real-world data with trial data, particularly in oncology and rare diseases, can enhance statistical power and promote efficiency of clinical trials.

Another mechanism of indirect data triangulation is multinational studies, which involve collaborations across different sites to analyze treatment effects and disease epidemiology on a global scale. Multinational studies incorporate variations in prescribing patterns and reproduce risk estimates in diverse settings, thereby strengthening the assessment of generalizability and reproducibility. They also enhance statistical power in studies of rare events or adverse drug reactions.[46] For example, a study involving the Neurological and mental health Global Epidemiology Network (NeuroGEN) initiative, and including 3.6 million mother-child pairs, suggested that an association between gestational diabetes mellitus and ADHD was unlikely due to a direct causal effect.[53] Another initiative, involving the Asian Pharmacoepidemiology Network (AsPEN), investigated the prescribing patterns of denosumab across Taiwan, Hong Kong and South Korea and elucidated treatment persistence and re-initiation among patients with bone metastases.[54,55]

In an effort to address transparency issues, analytic consistency and compliance with data privacy regulations, these initiatives have also contributed to publicly available information and sustainability efforts with regard to databases covering topics such as vaccines, neurological disorders, and mental illnesses.[46,56,57] To protect patient confidentiality, direct individual data sharing is often not feasible in the conduct of multinational studies. The common data model

(CDM) framework addresses this challenge by enabling data partners to convert their local databases into a standardized format through an extract, transform, and load process. This allows the coordinating center to run a unified analytic script across all converted CDM tables, ensuring compatibility while only collecting summary results from each site, without accessing individual-level data.[46] The AsPEN has successfully converted its participating databases to adopt the OMOP CDM, including Taiwan's NHIRD, Hong Kong's CDARS, the UK's THIN and CPRD, and the United States' Medicare, facilitating large-scale, international collaborative studies.[58,59] Due to the NHIRD's comprehensive coverage and longitudinal follow-up, Taiwan's healthcare database has played a crucial role in these efforts, further demonstrating Taiwan's contributions to global epidemiologic research.[60]

Beyond individual studies, multinational studies play a critical role in shaping public health policies and regulatory decisions. Findings from multinational comparisons inform drug regulatory agencies, such as the US Food and Drug Administration (FDA), the European Medicines Agency EMA, and Taiwan's Food and Drug Administration (TFDA), about variations in drug safety across different ethnic and racial settings. For example, following the results of a Danish study suggesting that hydrochlorothiazide may be associated with an increased risk of skin cancer,[61] regulatory authorities, including the US FDA and TFDA, issued timely safety warnings. Researchers from Taiwan and Denmark subsequently collaborated and used the NHIRD to assess the risk of such adverse events in patients receiving hydrochlorothiazide,[60,62] ultimately reaching the conclusion that skin cancer is not a major concern when prescribing antihypertensives for Asians. These multinational studies will further inform future meta-analyses comparing such risks among different races and ethnicities.[63] Such research not only helps to assuage public concern about the use of the drugs, but also clarifies to what extent the observed adverse effects may be influenced by racial or ethnic differences.

Together, these integration strategies expand the scientific utility of the NHIRD beyond a standalone claims database —transforming it into a flexible, federated health research platform of both national and global relevance.

Limitations and Perspectives

Despite its remarkable breadth and extensive adoption in academic research, the NHIRD of Taiwan faces several important limitations. First and foremost, concerns about data privacy, cybersecurity and potential data misuse have come under intense scrutiny. We herewith summarize the recent progress and future perspectives regarding these issues.

Constitutional Court Judgement

In March 2012, several human rights groups in Taiwan launched a campaign urging the public to send official letters to the BNHI (the predecessor of the NHIA), refusing to allow their personal health insurance data to be released to third parties. After the BNHI rejected their requests, the applicants filed a series of administrative lawsuits. The Supreme Administrative Court dismissed these lawsuits in January 2017, prompting a request for constitutional interpretation.

In August 2022, the Constitutional Court issued Judgment No. 13 of the Year 2022.[64] This ruling addressed the NHIA's practice of providing personal data from the health insurance database to third parties for secondary use, beyond the original purpose of NHI.[65] Going forward, this landmark judgment may significantly impact personal data protection legislation and practices in Taiwan, potentially expediting amendments to the Personal Data Protection Act (PDPA).

Key points of this Constitutional Court Judgment were:

1. The PDPA's provisions on using medical data for statistical or academic research are constitutional.
2. The lack of an independent supervision mechanism for personal data protection risks unconstitutionality.
3. Regulations regarding the storage, processing, external transmission, and provision of health insurance data are unconstitutionally ambiguous.
4. The absence of opt-out rights for individuals regarding the use of their health insurance data is unconstitutional.

Meanwhile, the Constitutional Court mandated relevant authorities to establish or amend laws within three years. These laws should specify subjects, reasons, procedures and effects regarding requests for cessation and exceptions (the right to opt out). If not amended within this period, individuals may request to stop the use of their data for purposes beyond the original collection.

In response to the Constitutional Court's requirements, Taiwan's MOHW proposed a draft "Health and Welfare Data Management Act" for legislative review in March 2024.[66] This draft aims to: (1) establish a legal foundation for the NHIRD; (2) elevate the process of applying for data usage beyond the original purposes to a legal level; (3) create a mechanism for individuals to request cessation of data usage beyond the original purposes; and (4) expand the regulations' scope to cover all health and welfare data managed by the MOHW. These measures are designed to streamline data re-use while addressing legal and ethical concerns. If implemented effectively, the Health and Welfare Data Management Act may serve as a model framework balancing individual privacy and data-driven research innovation.

Cybersecurity

Beyond legal and ethical scrutiny, practical challenges such as cybersecurity threats and data inconsistency further complicate the use of the NHIRD. These issues collectively highlight the urgent need for a more resilient and transparent data infrastructure. Recent investigative reports and expert commentary have highlighted the NHIRD as a potential national security vulnerability, particularly in the context of escalating geopolitical tensions and cyberattacks targeting sensitive health information.[67] The centralized design of the NHIRD, while beneficial for data integration, renders it a high-value target for malicious actors. Taiwan's National Security Council reports that the country experiences over five million cyberattacks daily, with more than half of all malware incidents in the Asia–Pacific region directed at Taiwan — underscoring the scale and persistence of the threat.[68,69] Healthcare and insurance databases are among the most attractive targets for espionage and disinformation campaigns, given their potential for strategic intelligence exploitation.

Moreover, multiple breaches involving unauthorized access and the dissemination of personal health data—including those of intelligence officers—have raised questions about the adequacy of current oversight mechanisms[70]. The lack of a robust opt-out mechanism (the insured person can choose not to reveal their data) and limited independent data governance have been declared unconstitutional by Taiwan's judiciary, with reforms mandated within a three-year timeframe. Until such changes are implemented, researchers—particularly those working with high-risk subpopulations or training AI algorithms— must acknowledge the possibility of systemic data exposure risk, which could undermine public trust and limit the long-term viability of the database.[17,71]

In addition to governance concerns, issues related to data consistency, interoperability, and access barriers remain. Variability in coding practices across institutions, particularly in the newer laboratory and imaging datasets, poses challenges to cross-hospital standardization. Although the NHIA has encouraged algorithm sharing and infrastructure expansion through remote data centers, the analytical burden still falls heavily on individual research teams, which may limit reproducibility and slow scientific progress. These limitations necessitate heightened vigilance in research design, particularly in studies relying on longitudinal follow-up, diagnostic accuracy, or cross-institutional comparability.

Taken together, the long-term value of the NHIRD will depend on ethical safeguards, transparent governance, and ongoing investment in secure, interoperable infrastructure. Strengthening national cybersecurity capacity, enhancing cross-border collaboration, and integrating real-time threat intelligence into data governance strategies will be essential to protect this critical research resource in an increasingly adversarial digital landscape.[72]

Fast Healthcare Interoperability Resources (FHIR)

To address this challenge and enable seamless cross-institutional data exchange, the MOHW launched the National Medical Data Standards Platform in March 2025.[73] At its core, this initiative adopts the international FHIR (Fast Healthcare Interoperability Resources) standard to unify the structure and format of healthcare data across Taiwan. FHIR is often likened to the HTML of the internet: an essential common language for enabling health data interoperability across different systems. The new architecture introduces three levels of standardization:

1. Data standardization via FHIR, enhanced with support for SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms), LOINC (Logical Observation Identifiers Names and Codes), and RxNorm (Prescription Normalized Names);
2. Rule standardization via CQL (Clinical Quality Language), which streamlines NHI reimbursement rules and quality evaluation;

3. Application standardization through SMART (Substitutable Medical Applications, Reusable Technologies) on FHIR, facilitating modular development of AI tools and medical applications, similar to a healthcare-specific app ecosystem.[74]

To accelerate implementation, MOHW is establishing three regional FHIR centers tasked with building cross-institutional "data middle platforms".[75] These platforms serve as standardized interfaces for EMRs, enabling the transformation, storage, and visualization of data in FHIR format. Hospitals involved must pass rigorous certification based on Taiwan's TW Core Implementation Guide, covering over 100 essential clinical data items (e.g., admission notes, lab results, prescriptions).

By integrating the NHIRD's claims and clinical data with the FHIR-based ecosystem, Taiwan aims to create a national health data infrastructure that is not only standardized and secure but also AI-ready. The long-term vision is to support real-time data sharing, decision support systems, and personalized medicine across the healthcare continuum. As Taiwan stands at the intersection of health innovation and ethical governance, the evolution of the NHIRD will serve as a critical bellwether for balancing national health priorities with global standards in data stewardship. The next decade will determine whether the NHIRD can sustain its pioneering role in population health research amidst mounting legal, technological and ethical demands. Importantly, FHIR standardization also lays the groundwork for future integration of the NHIRD with hospital EMR, enabling real-time clinical decision support and expanding the database's utility in translational and implementation research.

Conclusion

As Taiwan's NHIRD enters its fourth decade, this review synthesizes three decades of evolution and demonstrates how the database's transformation into a multimodal, AI-ready platform is reshaping the landscape of real-world evidence generation in Taiwan and beyond. Once limited to administrative claims, the NHIRD now integrates laboratory results, medical imaging, and validation studies, while also expanding its governance, legal frameworks, and data linkage capabilities. These advancements significantly enhance its utility for epidemiological research, regulatory science, and health policy decision-making.

Looking ahead, sustaining the NHIRD's scientific and societal impact will require continued investment in several key areas: strengthening cybersecurity defenses against increasingly sophisticated threats, improving interoperability and data harmonization to support cross-institutional and international collaboration, and ensuring transparent governance that balances accessibility with public trust. As artificial intelligence, precision medicine, and federated data networks redefine the future of healthcare, the NHIRD's comprehensive population coverage and evolving infrastructure position it as a vital resource for innovation.

With these measures in place, the NHIRD is poised not only to remain a cornerstone of Taiwan's health research ecosystem but also to serve as a global model for secure, ethically governed, and interoperable health data platforms — advancing clinical discovery, informing regulatory decisions, and shaping evidence-based healthcare in the decades to come.

Data Sharing Statement

Not applicable. As this article does not involve primary data collection, patient recruitment, or analysis of identifiable personal information, formal ethical approval was not required.

Author Contributions

All authors made a significant contribution to the work reported; be that in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all of these areas; they all took part in drafting, revising or critically reviewing the article, gave final approval of the version to be published, agreed on the journal to which the article has been submitted, and agree to be held accountable for all aspects of the work.

Funding

This study was supported in part by grants from the National Science and Technology Council of Taiwan [NSTC 114-2628-B-006-005-] and the National Health Research Institutes of Taiwan [NHRI-11A1-CG-CO-04-2225-1, NHRI-12A1-CG-CO-04-2225-1, NHRI-13A1-CG-CO-04-2225-1 and NHRI-14A1-CG-CO-04-2225-1] to ECCL. The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Disclosure

All authors declare no conflict of interest with regard to this work.

References

1. Oderanti FO, Li F, Cubric M, Shi X. Business models for sustainable commercialisation of digital healthcare (eHealth) innovations for an increasingly ageing population. *Technol Forecasting Social Change*. 2021;171:120969. doi:10.1016/j.techfore.2021.120969
2. Briganti G, Le Moine O. Artificial Intelligence in Medicine: today and Tomorrow. *Front Med Lausanne*. 2020;7:27. doi:10.3389/fmed.2020.00027
3. Akyüz K, Cano Abadía M, Goisauf M, Mayrhofer MT. Unlocking the potential of big data and AI in medicine: insights from biobanking. *Front Med*. 2024;11. doi:10.3389/fmed.2024.1336588
4. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31–38. doi:10.1038/s41591-021-01614-0
5. Näher AF, Vorisek CN, Klopfenstein SAI, et al. Secondary data for global health digitalisation. *Lancet Digit Health*. 2023;5(2):e93–e101. doi:10.1016/S2589-7500(22)00195-9
6. Hsieh CY, Su CC, Shao SC, et al. Taiwan's national health insurance research database: past and future. *Clin Epidemiol*. 2019;11:349–358. doi:10.2147/CLEP.S196293
7. Sung SF, Hsieh CY, Hu YH. Two decades of research using Taiwan's national health insurance claims data: bibliometric and text mining analysis on pubmed. *J Med Internet Res*. 2020;22(6):e18457. doi:10.2196/18457
8. Wang TH, Tsai YT, Lee PC. Health big data in Taiwan: a national health insurance research database. *J Formos Med Assoc*. 2023;122(4):296–298. doi:10.1016/j.jfma.2022.12.016
9. Lee PC, Ho CH, Wang JTH. Digital transformation of big data. In: Lee PC, Wang JTH, Chen TY, Peng CH, editors. *Digital Health Care in Taiwan: Innovations of National Health Insurance*. Cham: Springer International Publishing; 2022:219–228.
10. Lin ZY, Hsu CC. Constitutional court judgement on research use of National Health Insurance database. *JFormos Med Assoc*. 2023;122(6):440–441. doi:10.1016/j.jfma.2023.03.019
11. Tsuei SH. Reflection on 30 years of Taiwanese national health insurance: analysis of Taiwanese health system progress, challenges, and opportunities. *J Formos Med Assoc*. 2024;123(Suppl 3):S180–s187.
12. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the national health insurance research database (NHIRD). *Epidemiol Health*. 2018;40:e2018062. doi:10.4178/epih.e2018062
13. About the Ministry of Health and Welfare Data Center. 2025. Available from: https://dep.mohw.gov.tw/DOS/cp-5119-59201-113.html. Accessed April 20, 2025.
14. List of Files. Available from: https://dep.mohw.gov.tw/DOS/lp-2503-113-xCat-DOS_dc004.html. Accessed April 20, 2025.
15. Lee PC, Kao FY, Liang FW, Lee YC, Li ST, Lu TH. Existing data sources in clinical epidemiology: the Taiwan national health insurance laboratory databases. *Clin Epidemiol*. 2021;13:175–181. doi:10.2147/CLEP.S286572
16. Chen HL, Wang IT, Tsai YW, et al. Superior benefits of sodium-glucose co-transporter-2 inhibitors compared with dipeptidyl peptidase-4 inhibitors for diabetic kidney disease: a cohort study. *Diabetes Obes Metab*. 2025;27(1):174–183. doi:10.1111/dom.15998
17. Lu JR, Liang LL. The role of digital health under Taiwan's national health insurance system: progress and challenges. *Health Syst Reform*. 2024;10(2):2375433. doi:10.1080/23288604.2024.2375433
18. Digital health care in Taiwan: innovations of national health insurance. 2023. Available from: https://www.nhi.gov.tw/en/cp-1111-156fd-200-2.html. Accessed April 20, 2025.
19. Chen PT, Wu T, Wang P, et al. Pancreatic cancer detection on CT scans with deep learning: a nationwide population-based study. *Radiology*. 2023;306(1):172–182. doi:10.1148/radiol.220152
20. Hsieh CY, Chen PT, Shao SC, Lin SJ, Liao SC, Lai ECC. Validating ICD-10 diagnosis codes for Guillain-Barré syndrome in Taiwan's national health insurance claims database. *Clin Epidemiol*. 2024;16:733–742. doi:10.2147/CLEP.S485953
21. Lu PT, Tsai TH, Lai CC, Chuang LH, Shao SC. Validation of diagnostic codes to identify glaucoma in Taiwan's claims data: a multi-institutional study. *Clin Epidemiol*. 2024;16:227–234. doi:10.2147/CLEP.S443872
22. Gau SY, Tsai HE, Wang YH, Wei JC. Validation of diagnosis algorithms for ankylosing spondylitis in claim-based database. *Int J Rheum Dis*. 2024;27(1):e15041. doi:10.1111/1756-185X.15041
23. Tsai TY, Lin JF, Tu YK, et al. Validation of ICD-10-CM diagnostic codes for identifying patients with ST-elevation and non-ST-elevation myocardial infarction in a national health insurance claims database. *Clin Epidemiol*. 2023;15:1027–1039. doi:10.2147/CLEP.S431231
24. Wu LY, Shao SC, Liao SC. Positive predictive value of ICD-10-CM codes for myocarditis in claims data: a multi-institutional study in Taiwan. *Clin Epidemiol*. 2023;15:459–468. doi:10.2147/CLEP.S405660
25. Chang C, Liao SC, Shao SC. Positive predictive values of anaphylaxis diagnosis in claims data: a multi-institutional study in Taiwan. *J Med Syst*. 2023;47(1):97. doi:10.1007/s10916-023-01989-2
26. Fu SH, Yu PY, Li CY, et al. Diagnostic accuracy of algorithms to define incident and second Hip fractures: a Taiwan validation study. *J Formos Med Assoc*. 2023;122(Suppl 1):S82–s91.
27. Liao SC, Shao SC, Lai ECC, Lin SJ, Huang WI, Hsieh CY. Positive predictive value of ICD-10 codes for cerebral venous sinus thrombosis in Taiwan's national health insurance claims database. *Clin Epidemiol*. 2022;14:1–7. doi:10.2147/CLEP.S335517
28. Chiang MY, Shao SC, Liao SC. Validation of diagnostic codes to identify carbon monoxide poisoning in Taiwan's claims data. *Front Pharmacol*. 2022;13:882632. doi:10.3389/fphar.2022.882632
29. Tsai MJ, Tsai CH, Pan RC, Hsu CF, Sung SF. Validation of ICD-9-CM and ICD-10-CM diagnostic codes for identifying patients with out-of-hospital cardiac arrest in a national health insurance claims database. *Clin Epidemiol*. 2022;14:721–730. doi:10.2147/CLEP.S366874
30. Sheu MJ, Chin TW, Ku FP, Li CY, Li ST, Lu TH. Validation of coding algorithms for identifying people with viral hepatitis using claims data according to different standard references. *BMC Infect Dis*. 2022;22(1):222. doi:10.1186/s12879-022-07212-w
31. Hsieh MT, Hsieh CY, Tsai TT, Sung SF. Validation of stroke risk factors in patients with acute ischemic stroke. *Transient Ischemic Attack, or Intracerebral Hemorrhage on Taiwan's National Health Insurance Claims Data Clin Epidemiol*. 2022;14:327–335.

32. Hsieh MT, Huang KC, Hsieh CY, Tsai TT, Chen LC, Sung SF. Validation of ICD-10-CM diagnosis codes for identification of patients with acute hemorrhagic stroke in a national health insurance claims database. *Clin Epidemiol*. 2021;13:43–51. doi:10.2147/CLEP.S288518
33. Hsieh MT, Hsieh CY, Tsai TT, Wang YC, Sung SF. Performance of ICD-10-CM diagnosis codes for identifying acute ischemic stroke in a national health insurance claims database. *Clin Epidemiol*. 2020;12:1007–1013. doi:10.2147/CLEP.S273853
34. Sheu MJ, Liang FW, Li ST, Li CY, Lu TH. Validity of ICD-10-CM codes used to identify patients with chronic hepatitis B and C Virus infection in administrative claims data from the Taiwan national health insurance outpatient claims dataset. *Clin Epidemiol*. 2020;12:185–192. doi:10.2147/CLEP.S236823
35. Hsu MC, Wang CC, Huang LY, Lin CY, Lin FJ, Toh S. Effect of ICD-9-CM to ICD-10-CM coding system transition on identification of common conditions: an interrupted time series analysis. *Pharmacoepidemiol Drug Saf*. 2021;30(12):1653–1674. doi:10.1002/pds.5330
36. Ministry of Health and Welfare. Introduction of databases governed by the health and welfare data science center. Available from: https://dep.mohw.gov.tw/DOS/lp-2503-113-xCat-DOS_dc002.html. Accessed February 21, 2025.
37. Huang WC, Yang ASH, Tsai DHT, Shao SC, Lin SJ, Lai ECC. Association between recently raised anticholinergic burden and risk of acute cardiovascular events: nationwide case-case-time-control study. *BMJ*. 2023;382:e076045. doi:10.1136/bmj-2023-076045
38. Meng LC, Lin CW, Lin YC, et al. Association between maternal benzodiazepine or Z-hypnotic use in early pregnancy and the risk of stillbirth, preterm birth, and small for gestational age: a nationwide, population-based cohort study in Taiwan. *Lancet Psychiatry*. 2023;10(7):499–508. doi:10.1016/S2215-0366(23)00148-7
39. Chuang HM, Meng LC, Lin CW, et al. Concomitant use of antidepressants and benzodiazepines during pregnancy and associated risk of congenital malformations: a population-based cohort study in Taiwan. *Lancet Psychiatry*. 2024;11(8):601–610. doi:10.1016/S2215-0366(24)00176-7
40. Chien LH, Chen TY, Chen CH, et al. Recalibrating risk prediction models by synthesizing data sources: adapting the lung cancer PLCO model for Taiwan. *Cancer Epidemiol Biomarkers Prev*. 2022;31(12):2208–2218. doi:10.1158/1055-9965.EPI-22-0281
41. Chen YC, Lo YC, Wu HY, Huang YC. Adherence to dietary guidelines associated with lower medical service utilization in preschoolers: a longitudinal study. *Nutr Diabetes*. 2024;14(1):11. doi:10.1038/s41387-024-00270-w
42. Lee JT, Huang N, Majeed A. The need for better linkage between administrative data and clinical datasets. *BMJ*. 2015;351:h5816. doi:10.1136/bmj.h5816
43. Hsu CN, Huang K, Lin FJ, et al. Continuity and completeness of electronic health record data for patients treated with oral hypoglycemic agents: findings from healthcare delivery systems in Taiwan. *Front Pharmacol*. 2022;13. doi:10.3389/fphar.2022.845949
44. Shao SC, Chan YY, Kao Yang YH, et al. The Chang Gung research database-A multi-institutional electronic medical records database for real-world epidemiological studies in Taiwan. *Pharmacoepidemiol Drug Saf*. 2019;28(5):593–600. doi:10.1002/pds.4713
45. Cheng CL, Chien HC, Lee CH, Lin SJ, Yang YH. Validity of in-hospital mortality data among patients with acute myocardial infarction or stroke in National Health Insurance Research Database in Taiwan. *Int J Cardiol*. 2015;201:96–101. doi:10.1016/j.ijcard.2015.07.075
46. Tsai DHT, Bell JS, Abtahi S, et al. Cross-regional data initiative for the assessment and development of treatment for neurological and mental disorders. *Clin Epidemiol*. 2023;15:1241–1252. doi:10.2147/CLEP.S426485
47. Lunt M, Glynn RJ, Rothman KJ, Avorn J, Stürmer T. Propensity score calibration in the absence of surrogacy. *Am J Epidemiol*. 2012;175(12):1294–1302. doi:10.1093/aje/kwr463
48. Su CC, Yang YK, Lai EC, et al. Comparative safety of antipsychotic medications in elderly stroke survivors: a nationwide claim data and stroke registry linkage cohort study. *J Psychiatr Res*. 2021;139:159–166. doi:10.1016/j.jpsychires.2021.05.025
49. Yang ASH, Tsai, DHT, Chen, LW, et al. Epidemiology and Burden of Illness of Lennox-Gastaut Syndrome in Taiwan: A Retrospective Cohort Study. Risk Manag Healthc Policy, 2025;18:3153–3166. doi:10.2147/RMHP.S519367
50. Chung YS, Shao SC, Chi MH, et al. Comparative cardiometabolic risk of antipsychotics in children, adolescents and young adults. *Eur Child Adolesc Psychiatry*. 2021;30(5):769–783. doi:10.1007/s00787-020-01560-1
51. Pan HC, Chen JY, Chen HY, et al. GLP-1 receptor agonists' impact on cardio-renal outcomes and mortality in T2D with acute kidney disease. *Nat Commun*. 2024;15(1):5912. doi:10.1038/s41467-024-50199-y
52. Park JE, Campbell H, Towle K, et al. Unanchored population-adjusted indirect comparison methods for time-to-event outcomes using inverse odds weighting, regression adjustment, and doubly robust methods with either individual patient or aggregate data. *Value Health*. 2024;27(3):278–286. doi:10.1016/j.jval.2023.11.011
53. Chan AYL, Gao L, Hsieh MHC, et al. Maternal diabetes and risk of attention-deficit/hyperactivity disorder in offspring in a multinational cohort of 3.6 million mother–child pairs. *Nat Med*. 2024;30(5):1416–1423. doi:10.1038/s41591-024-02917-8
54. Shen CY, Au PC, Baek YH, et al. Comparative treatment persistence with bone-targeting agents among asian patients with bone metastases from solid tumors: a multinational retrospective cohort study. *BioDrugs*. 2022;36(3):381–392. doi:10.1007/s40259-022-00528-8
55. Kim JH, Shen CY, Au PC, et al. Bone-targeting agents in major solid tumour metastases: a multinational cohort study. *BMJ Support Palliat Care*. 2024;13(e3):e1064–e1073. doi:10.1136/bmjspcare-2021-003062
56. Duszynski KM, Stark JH, Cohet C, et al. Suitability of databases in the Asia-Pacific for collaborative monitoring of vaccine safety. *Pharmacoepidemiol Drug Saf*. 2021;30(7):843–857. doi:10.1002/pds.5214
57. Lai ECC, Man KK, Chaiyakunapruk N, et al. Brief report: databases in the Asia-Pacific Region: the potential for a distributed network approach. *Epidemiology*. 2015;26(6):815–820. doi:10.1097/EDE.0000000000000325
58. Lai ECC, Ryan P, Zhang Y, et al. Applying a common data model to Asian databases for multinational pharmacoepidemiologic studies: opportunities and challenges. *Clin Epidemiol*. 2018;10:875–885. doi:10.2147/CLEP.S149961
59. Su CC, Lai ECC, Kao Yang YH, et al. Incidence, prevalence and prescription patterns of antipsychotic medications use in Asia and US: a cross-nation comparison with common data model. *J Psychiatr Res*. 2020;131:77–84. doi:10.1016/j.jpsychires.2020.08.025
60. Yang ASH, Djebarri L, Lee CN, et al. Hydrochlorothiazide use and risk of skin cancer: a population-based retrospective cohort study. *Pharmacoepidemiol Drug Saf*. 2024;33(11):e70027. doi:10.1002/pds.70027
61. Pedersen SA, Gaist D, Schmidt SAJ, Hölmich LR, Friis S, Pottegård A. Hydrochlorothiazide use and risk of nonmelanoma skin cancer: a nationwide case-control study from Denmark. *J Am Acad Dermatol*. 2018;78(4):673–681.e679. doi:10.1016/j.jaad.2017.11.042
62. Pottegård A, Pedersen SA, Schmidt SAJ, et al. Use of hydrochlorothiazide and risk of skin cancer: a nationwide Taiwanese case-control study. *Br J Cancer*. 2019;121(11):973–978. doi:10.1038/s41416-019-0613-4
63. Shao SC, Lai CC, Chen YH, Lai ECC, Hung MJ, Chi CC. Associations of thiazide use with skin cancers: a systematic review and meta-analysis. *BMC Med*. 2022;20(1):228. doi:10.1186/s12916-022-02419-9

64. Judicial Interpretation No. 13 of the Year 111. 2022. Available from: https://cons.judicial.gov.tw/docdata.aspx?fid=38&id=309956. Accessed April 25, 2025.
65. Lin ZY, Hsu CC. Constitutional court judgement on research use of national health insurance database. *J Formos Med Assoc*. 2023;122(6):440–441.
66. Draft Proposal for the "Regulations on the Management of Health and Welfare Data". 2024. Available from: https://www.mohw.gov.tw/cp-18-77823-1.html. Accessed April 20, 2025.
67. Tai HY, Wu SH. Infrastructure of the medical information system. In: Lee PC, Wang JTH, Chen TY, Peng CH, editors. *Digital Health Care in Taiwan: Innovations of National Health Insurance*. Cham: Springer International Publishing; 2022:111–128.
68. Huang H. A collaborative battle in cybersecurity? Threats and opportunities for Taiwan. *Asia Policy*. 2020;15(2):101–106. doi:10.1353/asp.2020.0015
69. Yau HM. A critical strategy for Taiwan's cybersecurity: a perspective from critical security studies. *Journal of Cyber Policy*. 2019;4(1):35–55. doi:10.1080/23738871.2019.1604782
70. Liu CY, Li WP, Tu YP. Privacy perils of open data and data sharing: a case study of Taiwan's open data policy and practices. *Wash Int'l LJ*. 2020;30:545.
71. Ho CH. Secondary use of health data for medical AI: a cross-regional examination of Taiwan and the EU. *Asian Bioeth Rev*. 2024;16(3):407–422. doi:10.1007/s41649-024-00279-4
72. Jing B, Wu T. UK-Taiwan cooperation in cyber security: challenges and opportunities. 2025.
73. The Ministry of Health and Welfare Officially Launches the Taiwan Medical Information Standards Platform to Promote the Standardization of Medical Data. 2025. Available from: https://www.informationsecurity.com.tw/article/article_detail.aspx?aid=11726. Accessed April 20, 2025.
74. Taiwan Medical Information Standards Platform. Available from: https://medstandard.mohw.gov.tw/. Accessed April 20, 2025.
75. The Ministry of Health and Welfare establishes three FHIR centers with the goal of building a cross-hospital data platform. 2024. Available from: https://www.cio.com.tw/wfhk-to-build-three-major-fhir-centers-to-build-cross-yard-data-center/. Accessed April 20, 2025.