

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

Daniel Hsiang-Te Tsai^{1,2}, J Simon Bell³, Shahab Abtahi⁴, Brenda N Baak⁵, Marloes T Bazelier⁴, Ruth Brauer⁶, Adrienne YL Chan⁶⁻⁹, Esther W Chan⁹⁻¹², Haoqian Chen¹³, Celine SL Chui^{9,14,15}, Sharon Cook¹⁶, Stephen Crystal¹⁶, Poonam Gandhi¹³, Sirpa Hartikainen¹⁷, Frederick K Ho¹⁸, Shao-Ti Hsu¹, Jenni Ilomäki³, Ju Hwan Kim¹⁹, Olaf H Klungel⁴, Marjaana Koponen¹⁷, Wallis CY Lau^{6,8,9}, Kui Kai Lau^{20,21}, Terry YS Lum²², Hao Luo²², Kenneth KC Man^{6,8,9}, Jill P Pell¹⁸, Soko Setoguchi^{13,23}, Shih-Chieh Shao^{1,24}, Chin-Yao Shen¹, Ju-Young Shin^{19,25,26}, Patrick C Sovereign⁴, Anna-Maija Tolppanen¹⁷, Li Wei^{6,9}, Ian CK Wong^{6,8,9,27}, Edward Chia-Cheng Lai¹

¹School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan; ²Centre for Neonatal and Paediatric Infection, St George's University of London, London, UK; ³Centre for Medicine Use and Safety, Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, Melbourne, Australia; ⁴Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands; ⁵PHARMO Institute for Drug Outcomes Research, Utrecht, the Netherlands; ⁶Research Department of Practice and Policy, UCL School of Pharmacy, London, UK; ⁷Groningen Research Institute of Pharmacy, Unit of Pharmacotherapy, -Epidemiology and -Economics, University of Groningen, Groningen, the Netherlands; ⁸Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ⁹Laboratory of Data Discovery for Health (D²4H), Hong Kong Science Park, Hong Kong, Special Administrative Region, People's Republic of China; ¹⁰Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ¹¹Department of Pharmacy, the University of Hong Kong-Shenzhen Hospital, Shenzhen, People's Republic of China; ¹²The University of Hong Kong Shenzhen Institute of Research and Innovation, Shenzhen, People's Republic of China; ¹³Center for Pharmacoepidemiology and Treatment Science (PETS), Institute for Health, Rutgers University, New Brunswick, NJ, USA; ¹⁴School of Nursing, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ¹⁵School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ¹⁶Center for Health Services Research, Rutgers University, New Brunswick, NJ, USA; ¹⁷Kuopio Research Centre of Geriatric Care and School of Pharmacy, University of Eastern Finland, Kuopio, Finland; ¹⁸School of Health and Wellbeing, University of Glasgow, Glasgow, UK; ¹⁹School of Pharmacy, Sungkyunkwan University, Suwon, South Korea; ²⁰Division of Neurology, Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ²¹State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ²²Department of Social Work and Social Administration, University of Hong Kong, Hong Kong, Special Administrative Region, People's Republic of China; ²³Department of Medicine, Rutgers Robert Wood Johnson Medical School and Pharmacoepidemiology and Treatments Science, Institute for Health, Rutgers University, New Brunswick, NJ, USA; ²⁴Department of Pharmacy, Keelung Chang Gung Memorial Hospital, Keelung, Taiwan; ²⁵Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, South Korea; ²⁶Department of Biohealth Regulatory Science, Sungkyunkwan University, Seoul, South Korea; ²⁷Aston Pharmacy School, Aston University, Birmingham, UK

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

Purpose: To describe and categorize detailed components of databases in the Neurological and Mental Health Global Epidemiology Network (NeuroGEN).

Methods: An online 132-item questionnaire was sent to key researchers and data custodians of NeuroGEN in North America, Europe, Asia and Oceania. From the responses, we assessed data characteristics including population coverage, data follow-up, clinical information, validity of diagnoses, medication use and data latency. We also evaluated the possibility of conversion into a common data model (CDM) to implement a federated network approach. Moreover, we used radar charts to visualize the data capacity assessments, based on different perspectives.

Results: The results indicated that the 15 databases covered approximately 320 million individuals, included in 7 nationwide claims databases from Australia, Finland, South Korea, Taiwan and the US, 6 population-based electronic health record databases from Hong Kong, Scotland, Taiwan, the Netherlands and the UK, and 2 biomedical databases from Taiwan and the UK.

Conclusion: The 15 databases showed good potential for a federated network approach using a common data model. Our study provided publicly accessible information on these databases for those seeking to employ real-world data to facilitate current assessment and future development of treatments for neurological and mental disorders.

Keywords: meta-data, data repository, Neurological and Mental Health Global Epidemiology Network, NeuroGEN

Introduction

The world is now facing an increasing incidence and prevalence of neurological and mental disorders.^{1,2} Mental disorders are globally widespread, impacting individuals from every corner of the world (lifetime prevalence: 29.2%; 25.9–32.6%).³ In 2010, it was estimated that 35.6 million people globally had some form of dementia. This number is projected to nearly double every 20 years, reaching 65.7 million by 2030 and 115.4 million by 2050.⁴ Neurological disorders tend to be the biggest contributor to disability-adjusted life years (DALYs) and the second largest group cause of deaths in the world, accounting for 11.6% of global DALYs and 16.5% of all-cause mortality.² In addition, non-pharmacological intervention and medication management of neurological and mental health disorders are often sub-optimal, including both potentially inappropriate prescribing and under-prescribing of clinically appropriate treatment.^{5,6} Furthermore, treatment of neurological and mental disorders remains constrained due to lack of new chemical entities, patient susceptibility to adverse drug events and high rates of intervention or medication non-adherence.^{7–9} Some available pharmacological treatments are associated with safety issues. For example, in older patients, the use of antipsychotics has been associated with mortality risk^{10,11} and psychotropic medications have been associated with sedation, falls, arrhythmia, metabolic syndromes and extrapyramidal symptoms.^{12–14} In pregnant women, the use of some anticonvulsants such as valproates has been associated with an increased risk of congenital and neurodevelopmental disorders in the offspring.¹⁵ These issues underscore the importance of proper assessment of safety and effectiveness of new and existing treatments for neurological and mental disorders.

While randomized controlled trials (RCTs) are considered the gold standard for evaluating drug efficacy, some vulnerable populations such as older patients and pregnant women or children are often excluded from RCTs due to ethical considerations.^{16,17} When investigating adverse events associated with medications, it is often neither ethical nor feasible to run an RCT. As a result, real-world data has become increasingly sought after, to identify the unmet needs and service-use patterns of patients with neurological and mental disorders, or to evaluate the outcomes of interventions for future research and development strategies.^{13,18} In addition, international multi-database pharmacoepidemiologic studies have become broadly accessible with the growth of information technology, healthcare databases and analytic tools, making it much easier to collect large datasets across heterogeneous healthcare systems, raising the possibility of studying rare outcome measures and including various races and ethnicities.^{19–21} Furthermore, cross-regional epidemiologic research provides opportunities to investigate differences among healthcare systems worldwide.

The Neurological and Mental Health Global Epidemiology Network (NeuroGEN) is a world-wide research initiative that aims to develop a platform for cross-country collaboration using real-world data to facilitate research and development of treatment for neurological and mental illnesses.²² The importance of cross-country collaboration and multi-national studies lies in the potential to increase sample size, which can improve statistical power for rare diseases, provide the opportunity for international comparisons across races and ethnicities, and foster exchanges of techniques and opinions. The NeuroGEN comprises researchers from North America, Europe, Asia and Oceania, who together have access to 15 databases in these regions. Several ongoing projects within NeuroGEN are described elsewhere.^{22–24} However, detailed information about these 15 databases and their available information have been scant. A better understanding of the data capacity of these databases with regard to neurological and mental disorders will allow for more accurate, comparative assessments of the effectiveness of treatments across healthcare systems, and support research and development of related treatments.²⁵ To this end, this study aimed to describe and categorize the detailed components of the databases available to NeuroGEN, including patient demographics, diagnoses, treatments, laboratory examinations and healthcare claims details. The goal was to triangulate different data sources and to report “metadata” that could provide information about other datasets for researchers specializing in neurological and mental illnesses. Furthermore, we investigated the databases’ potential for the development of a federated network approach which would enable collaboration across countries.

Methods

We designed an online questionnaire to collect information about the databases available to NeuroGEN researchers. The survey included 132 questions covering the following categories: (1) database characteristics (4 questions), (2) accessibility to the participating databases (7 questions), (3) patient information (29 questions), (4) healthcare facility visit details (5 questions), (5) diagnosis details (7 questions), (6) drug details (16 questions), (7) procedure details (5 questions), (8) laboratory examination details (5 questions), (9) claims details (9 questions), (10) information on alternative medicine (17 questions), (11) hospital details (9 questions), (12) physician details (10 questions) and (13) details on other healthcare professionals (9 questions). The survey was emailed to key researchers and data custodians in each NeuroGEN member organization. An email reminder was sent to those researchers and data custodians who did not respond to the original email. Researchers and data custodians were also given the option of nominating a colleague to complete the survey. After completion of the questionnaire, our coordinating center distributed the results of the questionnaire to all participants in order to confirm all information provided was correct, complete and up to date at the time of study completion.

Population coverage, follow-up data, clinical information, validity of diagnosis, medication use and data latency were examined using radar charts and a points system to visualize the assessments,²⁶ with 5 points assigned for very good, 4 points for good, 3 points for satisfactory, 2 points for poor and 1 point for very poor capacity. The assessments were performed independently by two investigators, whereby differences in interpretation were resolved through discussion with a third investigator. For example, Taiwan's NHIRD with 99.9% coverage of the population received 5 points (ie, very good) for population coverage and follow-up, but only 2 points (ie, poor) for clinical information.²⁷

The Common Data Model (CDM) concept means that all data partners convert their native databases to follow standardized data structures and terminologies, allowing the coordinating centre to generate a common analytic program that can be applied to all converted databases.²⁶ We assessed the essential data in the databases, such as enrollment period, patient characteristics, healthcare facility visit details, diagnosis details and drug details for their convertibility to a general common data model. We classified the data components by three categories and visualized them using three colors, as follows: Green to indicate that the data were available from the database and could be simply converted into a global CDM that could be applied for routine pharmacoepidemiology study without additional data, yellow to indicate that the data could be captured from additional data or by using proxy measures, and red to indicate that the data were not available in the database and it therefore could not be converted using a CDM.

Results

The 15 databases covered a total of approximately 320 million individuals from 9 countries/regions in 2020. [Table 1](#) presents the participating database characteristics. Among the 15 databases, 7 were claims databases from Australia (Pharmaceutical Benefits Scheme 10% sample dataset [PBS] and Victorian Linked Health Data [VLHD]), South Korea (National Health Insurance Service-National Health Insurance Database [NHIS-NHID]), Taiwan (Taiwan's National Health Insurance Research Database [NHIRD]), US (Medicaid [Medicaid] and 20% sample of Medicare [Medicare] databases) and Finland (The Finnish healthcare registers capturing the population of Finland [FinReg]); and 6 databases were electronic health records (EHR) databases from Hong Kong (Clinical Data Analysis and Reporting System [CDARS]), Scotland (Public Health Scotland [ISD]), Taiwan (Chang Gung Research Database [CGRD]), the Netherlands (PHARMO database network [PHARMO]) and the United Kingdom (Clinical Practice Research Datalink [CPRD] and The Health Improvement Network [THIN]). We also included 2 large scale biomedical databases from Taiwan (Taiwan Biobank [TWB]) and the UK (UK Biobank [UKB]). [Figure 1](#) presents the start year, lag times and numbers of individuals for each database. The response rate of the online questionnaire from NeuroGEN researchers was 100%.

[Table 2](#) presents the accessibility to the participating databases. Specific policies and protocol approvals for the use of the database or a review by an Institutional Review Board (IRB) were required for all. Validation studies were conducted in six of the claims databases (ie, NHIRD, NHIS-NHID, Medicaid, Medicare, VLHD, FinReg) and six of the EHR databases (ie, CDARS, CGRD, CPRD, THIN, PHARMO, ISD). The average costs of access to the databases for

Table 1 Database Characteristics

Database Abbreviation	Database Full Name	Country	Source Type
PBS	Pharmaceutical Benefits Scheme 10% Sample Dataset	Australia	National Claims Database
VLHD	Victorian Linked Health Data	Australia	Claims Database
CDARS	Clinical Data Analysis and Reporting System	Hong Kong	Multi-center EHR [†]
NHIS-NHID	National Health Insurance Service-National Health Insurance Database	Republic of Korea	National Claims database
CGRD	Chang Gung Research Database	Taiwan	7-hospital EHR
NHIRD	Taiwan's National Health Insurance Research Database	Taiwan	National Claims Database
TWB	Taiwan Biobank	Taiwan	Interview Survey Data
FinReg	Finnish National Healthcare Registers	Finland	National Healthcare Registers
PHARMO	PHARMO Data Network	The Netherlands	Multi-center EHR [#]
ISD	Public Health Scotland	Scotland	National EHR
CPRD	Clinical Practice Research Datalink	United Kingdom	EHR [‡]
THIN	The Health Improvement Network	United Kingdom	EHR
UKB	United Kingdom Biobank	United Kingdom	Interview Survey Data [*]
Medicaid	Medicaid & CHIP Research Data	United States	Claims Database [*]
Medicare	20% Sample of Medicare	United States	National Claims Database

Notes: [†]Public healthcare sector in Hong Kong includes all general out-patient clinics, specialist out-patient clinics, and hospitals managed by Hospital Authority. [#]A total of 50 geographically defined areas in the Netherlands which comprises pharmacy dispensing information (4.2 million active patients), with possible linkage to a nationwide hospitalization database, in-patient hospital pharmacy database (2.0 million patients), general practice database (2.5 million patients), clinical laboratory database (1.2 million patients), and nationwide cancer, pathology, and perinatal registries. [‡]There are 14.2 million active (alive, currently registered) patients in UK CPRD meeting quality criteria. ^{*}With linked EHR. ^{*}Alabama, Alaska, Arkansas, California, Colorado, Connecticut, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming.

Abbreviation: EHR, electronic health records.

academic institutions varied from 0 to 96,000 USD for individual projects or per data year, depending on the corresponding length of access to the database – varying from one day to two years.

[Supplementary Tables 1–11](#) present the database information concerning diagnoses, prescriptions, procedures, health expenses and coding systems, as below. Specifically, unique patient identifiers and demographic characteristics such as sex, age and birth and death information were available in thirteen databases, excepting PBS and TWB ([Supplementary Table 1](#)). Information on race and health information was available in most of the EHR databases, but not in the claims databases. Thirteen databases contained the visit type and date of visit ([Supplementary Table 2](#)). The reasons for the visit or discharge were only available in EHR databases and some of the claims databases, except for NHIRD, PBS, THIN, TWB. Eleven databases used ICD-9 or ICD-10 as the diagnostic codes ([Supplementary Table 3](#)), while the other four databases, including PBS, THIN, CPRD and TWB, used domestic codes that could be successfully mapped to the international codes. Domestic codes were commonly used for drug details, and some could be matched to international codes (ie, ATC codes) ([Supplementary Table 4](#)). As for procedure details, the NHIRD, Medicaid, Medicare, VLHD, CDARS and ISD used ICD-9/10 procedure codes ([Supplementary Table 5](#)). Most of the EHR databases, NHIS-NHID, FinReg (since 2014), and TWB contained clinical values for laboratory testing results, but the other claims databases, and UKB did not ([Supplementary Table 6](#)). However, data on the type of test were available in most of the databases. NHIRD, NHIS-NHID, PBS provided health claims details, although others did not or provided limited information ([Supplementary Table 7](#)). Four databases (ie, NHIRD, NHIS-NHID, CGRD, PHARMO) contained longitudinal dispensing data for alternative medicine (AM) using a domestic coding system ([Supplementary Table 8](#)). Other information about the hospitals ([Supplementary Table 9](#)), physicians ([Supplementary Table 10](#)) and other healthcare providers ([Supplementary Table 11](#)) was also investigated, although this information was limited.

The radar charts quantify strengths in the features of databases ([Figure 2](#)). Specifically, NHIRD, NHIS-NHID, CDARS, ISD, Medicare, Medicaid, and FinReg scored high in population coverage. However, we should note the heterogeneity among these databases. The Medicare program covers all retirees, ie, mostly older people in the US, while

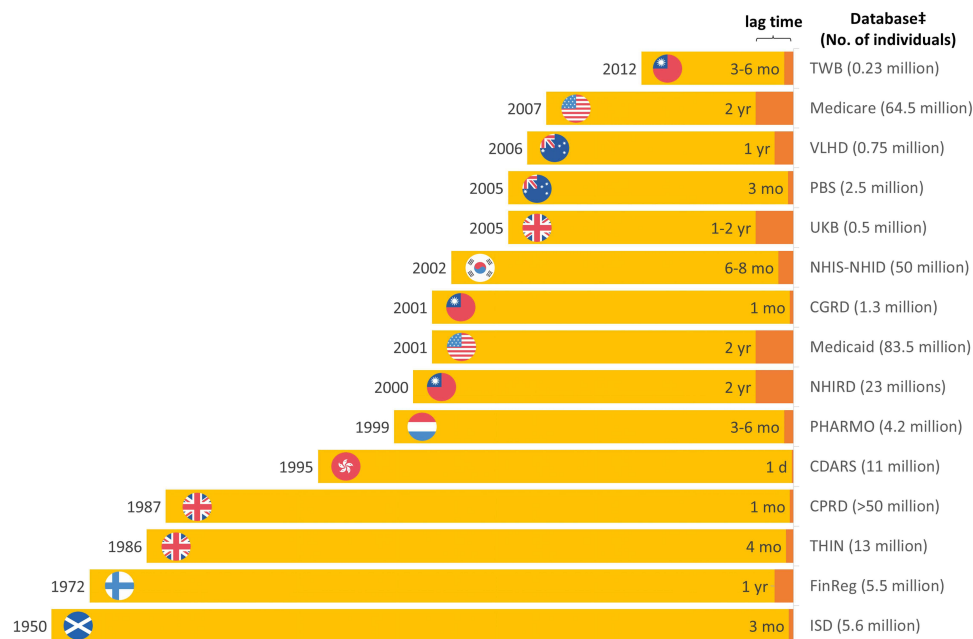


Figure 1 NeuroGEN databases.

‡**Abbreviations:** PBS, Pharmaceutical Benefits Scheme 10% sample dataset; VLHD, Victorian Linked Health Data; FinReg, Finnish national healthcare registers; CDARS, Clinical Data Analysis and Reporting System; NHIS-NHID, National Health Insurance Service-National Health Insurance Database; ISD, Public Health Scotland; CGRD, Chang Gung Research Database; NHIRD, Taiwan's National Health Insurance Research Database; TWB, Taiwan Biobank; PHARMO, PHARMO Data Network; CPRD, Clinical Practice Research Datalink; THIN, The Health Improvement Network; UKB, United Kingdom Biobank; Medicaid, Medicaid & CHIP Research data; Medicare, 20% sample of Medicare.

the Medicaid program is available to low-income people in the US. However, while eligibility for the Medicare program is standardized throughout the US, eligibility for Medicaid varies state by state. Moreover, the NHIRD, NHIS-NHID, and CDARS provided the best follow-up data for researchers to study long-term outcomes. The EHR databases, including CGRD, CDARS, CPRD, THIN, ISD, and PHARMO scored well for clinical information and data latency, enabling timely assessments.

Figure 3 presents the assessments of possibility for conversion to a federated network approach using a CDM. Seven claims databases, six EHR databases and two biomedical databases were included. Most of the databases provided sufficient information for CDM conversion to support routine pharmacoepidemiologic and health services research studies, including the observation period, patient characteristics, visits, diagnoses, drug exposures and drug strengths. Specifically, the PBS lacked diagnosis information that could be converted to CDM, which may pose challenges for multinational studies where diagnosis information is required.

Discussion

This study established a metadata framework and provided detailed information on the 15 databases available to NeuroGEN researchers. Most of these databases included information on patients' demographics, diagnoses, prescriptions, procedures and claims details, offering opportunities for large-scale investigation to study neurological and mental disorders, including understanding the burden of diseases, drug safety and clinical outcomes, and their healthcare utility. Accessibility varied among the different databases in terms of the length of time for application approval, IRB review requirements and cost of access to the database. Most of the databases provided structured information on diagnoses, prescriptions, procedures and health expenses that could be easily converted to a common data format. However, the databases used a variety of international and domestic coding systems that may require good mapping procedures to conform to a common terminology. Our study provided publicly accessible information on these databases for those seeking to employ real-world data to facilitate current assessment and future development of treatments for neurological

Table 2 Accessibility of Available Databases

Source Type Database	Claims Databases							Electronic Health Records						Others	
	NHIRD	NHIS-NHID	PBS	Medicaid	Medicare	VLHD	FinReg	CDARS	CGRD	THIN	CPRD	ISD	PHARMO	TWB	UKB
Specific policy for using	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Validation study	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓ [‡]	✓		
Average costs (USD)	15,000	600 per mo	350 per 3-mo [¶]	35,000 per data yr	96,000 per data yr	20,000	Depending on cohort size	0	0	0	19,000*	20,000	NA	4,000	7,000
Period of obtaining access to the database	6 mo	5–6 mo	4 wk	6 mo	3–6 mo	1.5–2 yr	Unknown	Depending on cohort size	Depending on cohort size	– [§]	1 day (with license) [§]	3 mo	1 mo	6 mo	1–2 mo
Require IRB approval before application	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Waiting time for IRB approval	4–6 wk	2–3 wk	2 wk	3–6 wk [¶]	3–6 wk [¶]	2 mo	6–16 mo	2–3 wk	3 wk	2–3 mo	1 mo	1–2 mo [#]	1 mo	4–6 wk	–
IRB application fee (USD)	170	0	0	Varies by IRB and request	Varies by IRB and request	100	1,000–3,000	0	0 [‡]	0	0	0	0	170	–

Notes: [‡]1% audit of all data (validation of diagnosis paper submitted for publication). [¶]\$950 for initial supply, \$350 per 3-month data update plus \$460 for each project approval. *GP data only, for academia. [§]The full dataset is available within UCL School of Pharmacy. [§]Without a license: unknown. [¶]Medicare/Medicaid requires additional approval from Centers for Medicare and Medicaid Services (CMS) Privacy Board, which takes 8–16 weeks. [#]Depends on urgency. [‡]If applied by the healthcare personnel from CGMH.

Abbreviations: NA, not applicable; mo, month(s); wk, week(s); yr, year(s).



Figure 2 Database features.

and mental disorders. This will also serve to raise public awareness of neurological and mental illness research to further facilitate treatment and non-pharmacological intervention.

Cross-country studies offer the advantage of evaluating the heterogeneity of healthcare systems from different countries in a real-world setting, while also increasing the study sample size to facilitate the study of rare neurological and mental disorders, or the monitoring of adverse drug reactions. This is especially important in the case of neurological and mental disorders where the prevalence rate of diseases is sometimes low, and the treatment preferences are varied, and affected by the characteristics of different healthcare systems. For example, a study by Raman et al evaluated the trends of medication use in patients with attention-deficit hyperactivity disorder (ADHD) in 15 countries and found large variations in ADHD medication use across multiple regions.²⁸ The NeuroGEN databases constitute a platform to make international comparisons across North America, Europe, Asia and Oceania. We found that most of the databases provided good longitudinal follow-up of patients for

	Component/Database	Claims databases								EHR databases						Biomedical databases	
		NHIRD	NHIS-NHID	VLHD	PBS	Medicaid	Medicare	FinReg	CDARS	CGRD	THIN	CPRD	PHARMO	ISD	TWB	UKB	
Enrollment Period	Date of enrollment																
	Date of disenrollment																
Patient Information	Person ID																
	Sex																
	Year of birth																
	Month of birth																
	Day of birth																
	Death date																
Healthcare Institute Visit Details	Type of healthcare visits																
	Start date of the visit																
	End date of the visit																
Diagnosis Details	Diagnosis code																
	Diagnosis date																
	Order of diagnosis code																
Drug Details	Start date of the drug use																
	End date of the drug use																
	Quantity of drug																
	Number of days																
	Drug administration route																
	Drug domestic code																
	Drug amount																
Drug unit																	

Figure 3 Assessment for common data model conversion.

Notes: Green indicates that the data were available from the database and could be converted into CDM without additional data. Yellow indicates that the data could be captured by additional data or use of a proxy. Red indicates that the data were not available in the database, and it therefore could not be converted to CDM.

the assessment of long-term outcomes of neurological and mental disorder treatments. Several examples are available from the literature review.^{29–32} Moreover, one of the strengths of the NeuroGEN databases was that they covered a large variety of populations, thus providing the opportunity to evaluate racial and ethnic differences in the responses to medical products across countries, and especially for populations that are under-represented in clinical trials. Another great strength of real-world data is that it usually covers a large population, which could reduce random errors in the analysis. However, real-world data may be subject to systematic errors since the treatments are not randomized. While the databases provide the sources for real-world analyses, we should carefully examine data completeness to avoid possible selection bias, ensure the accuracy of data to avoid misclassification bias, and consider potential confounding factors because treatments were not randomized.

Compared to the unrepresentative sampling and limited geographic coverage of registries, the NeuroGEN databases contain more complete and accurate information.³³ Taking Taiwan as an example, the accuracy of records of diagnoses or interventions is associated with the government’s reimbursement policy. It is important to recognize that various databases may have distinct characteristics. For example, the population from military veterans’ databases may include a higher proportion of older adults than the general population, and hence we may risk overestimating the disease prevalence if age adjustment is not considered.³⁴ Additionally, the seven EHR databases of NeuroGEN can complement some information not available in the claims databases, including patients’ lifestyle factors (eg, body weights and height), self-paid medications or examinations, the values of laboratory data and pathology reports and images.³⁵ These EHRs are important to extend the potential scope of study topics or to serve as external datasets to validate records of diagnosis or for dealing with unmeasured confounding.^{13,36} Another feature of EHR databases is that the lag time for updates is short. We found that the lag time for EHR databases was between one day and six months, enabling timely assessment of emerging treatments or conditions for patients with neurological and mental disorders.³⁷ Some databases may overlap with regard to the population they cover, eg, CPRD and THIN from the UK, or NHIRD and CGRD from Taiwan. Although patient identifiers are encrypted by each database independently, some approaches are available that can identify duplicates in overlapping databases, which could be considered.³⁸ The TWB and UKB are a unique database that can provide genomic information to study racial or ethnic effects in conjunction with the use of medical products. This database also allows the extension of study to translational research.^{39,40} However, more databases are required to cover a wider range of genetic information since TWB and UKB are the only two databases with genomic data currently included in NeuroGEN.

The federated network approach with CDM is crucial for multiple database study in order to maintain data privacy and ensure the consistency of the analysis.^{19,41} The concept of the common data model means that all data partners convert their local databases following a standardized and harmonized extract, transform and load process which allows the coordinating centre to run a common analytic script on the converted CDM tables to produce mutually compatible results.^{26,42} As a result, the coordinator only needs to collect summary results from each data partner without accessing individual level data. Some global CDMs are currently available that can be applied in most routine pharmacoepidemiologic and health services research,

including the Observational Medical Outcomes Partnership (OMOP) CDM, Sentinel CDM and the National Patient-Centered Clinical Research Network (PCORnet) CDM, and ConcePTION.^{22,43–45} Some databases have already been converted to a CDM. For example, another initiative, the Asian Pharmacoepidemiology Network (AsPEN), has converted its participating databases into the OMOP CDM,²⁶ including Taiwan's NHIRD, Hong Kong's CDARS, the UK's THIN, CPRD and the United States' Medicare.³²

Our survey suggested that most of the databases in NeuroGEN contain key information components ready for conversion into a common structure, including diagnoses, prescriptions, procedures and health expenses; however, the conversion to common terminologies will require careful consideration because different countries may use differing or local, domestic terminological codes.³² Common terminology helps to maintain consistency of analysis and interpretation of the results from multi-national studies. Careful mapping between the different terminologies may be more required for diagnoses than for drugs or procedures since most of the databases use international codes such as ICD-9 or ICD-10. However, some databases include some unique codes such as HPCPS and CPT in the US Medicaid and Medicare databases, OPCS 4 in the ISD, ICPC in the PHARMO and SNOMED in the THIN or CPRD. Because coding systems have different hierarchies and structures, they sometimes cannot be mapped 1:1, which can lead to a loss of information during the mapping procedure. The easiest way to conduct mapping of drug codes is to transfer the domestic codes, such as READ or BNF codes in the THIN, to an international coding system, such as WHO ATC codes. Some of the NeuroGEN databases have included ATC in their data and some have completed the mapping from domestic codes to ATC, such as the NHIRD. These offer a good foundation for conversion. Some global CDMs such as the OMOP CDM incorporate unique standard terminologies for drugs at product levels, which can preserve more detailed information than ATC, which is at ingredient level. Great care must be exercised during the mapping procedure to minimize the loss of information, especially for some unique products that are only available in specific countries without standardized terminology. A minimum requirement for the quality of conversion is the completeness of data conversion. Based on a scan of the converted CDM, we can calculate the frequencies of the codes and compare the results with the corresponding frequencies in non-converted data. Moreover, the demographics and diagnoses by calendar years can be checked for consistency after conversion to CDM. The heterogeneity of healthcare systems and related database formats, software for data storage and analysis, techniques, languages and time differences all present challenges for conversion. The heterogeneity of healthcare systems and related database formats, software for data storage and analysis, techniques, languages and time differences all present challenges for conversion. Good communication between sites will be the cornerstone for conversion to ensure consistency.

Future Direction

Several directions may be considered for the development of further research on neurological and psychiatric diseases. First, NeuroGEN could work with other initiatives such as AsPEN, NorPEN or EU PE&PV to expand the capacity and diversity of its databases and to facilitate the federated network approach using a CDM to integrate the experience gained from previous projects. Experience gleaned in mapping the terminological codes by those initiatives is especially important for NeuroGEN members. Second, the effectiveness and safety of currently available and new drugs can be evaluated. For example, the effectiveness of a recently approved drug for Alzheimer's disease by the US Food and Drug Administration, aducanumab, has not been evaluated using real-world data, despite reports of adverse events such as increased risk of vascular edema and hemorrhage, which require further evaluation to ascertain the causality.⁴⁶ Third, NeuroGEN can be used to expand the evidence base for traditional medications for the management of dementia or psychiatric disorders.^{47–49} Some countries and databases contain data on such traditional medications (eg, alternative medicines). Fourth, data may also be used to identify and validate targets for drug repurposing.^{50–52} Fifth, the available databases have great potential to advance treatment other than drug prescriptions, such as many evidence-based non-pharmacological interventions for mental disorders. Finally, high-quality training of pharmacoepidemiologists and statisticians through teaching programs also forms a cornerstone. Routine educational courses, workshops and conferences could be considered to improve investigators' understanding of the databases and to share analytical skills among countries.

Conclusion

We have established a publicly accessible metadata framework of the 15 databases available to NeuroGEN researchers across North America, Europe, Asia and Oceania, covering approximately 320 million individuals, to facilitate the use of

real-world data for the assessment of disease burden and the development of current and future treatments for neurological and mental disorders. We provided detailed information on the participating databases to assess the accessibility of their data and the feasibility of future investigations. Moreover, we found that most of the databases included structured information on patients' demographics, diagnoses, prescriptions, procedures and healthcare expenditures, and offered great potential for a federated network approach after conversion to a CDM.

Data Sharing Statement

All data were available upon reasonable request by contacting Edward Chia-Cheng Lai.

Author Contributions

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work.

Disclosure

ECCL reports research funding outside the submitted work from Amgen, Pfizer, Sanofi, Takeda, Roche, IQVIA, the Taiwan National Science and Technology Council (ID:111-2628-B-006-007-) and the Taiwan National Health Research Institutes. EWC reports grants from Research Grants Council (RGC, Hong Kong), Research Fund Secretariat of the Food and Health Bureau, National Natural Science Fund of China, Wellcome Trust, Bayer, Bristol-Myers Squibb, Pfizer, Janssen, Amgen, Takeda and the Narcotics Division of the Security Bureau of the Hong Kong Special Administrative Region; a honorarium from the Hospital Authority; outside the submitted work. CSLC has received grants from the Food and Health Bureau of the Hong Kong Government, Hong Kong Research Grant Council, Hong Kong Innovation and Technology Commission, Pfizer, IQVIA, MSD and Amgen; and personal fees from PrimeVigilance; outside the submitted work. WL reports a research grant from the AIR@InnoHK administered by the Innovation and Technology Commission outside the submitted work. HL has received grants from the Research Grants Council of Hong Kong, outside the submitted work. JSB is supported by a National Health and Medical Research Council (NHMRC) Boosting Dementia Research Leadership Fellowship and has received grant funding or consulting funds from the NHMRC, Medical Research Future Fund, Victorian Government Department of Health and Human Services, Dementia Australia Research Foundation, Yulgilbar Foundation, Aged Care Quality and Safety Commission, Dementia Centre for Research Collaboration, Pharmaceutical Society of Australia, Society of Hospital Pharmacists of Australia, GlaxoSmithKline Supported Studies Programme, Amgen and several aged care provider organizations unrelated to this work. All grants and consulting funds were paid to the employing institution. AMT reports research funding outside the submitted work from Amgen. ICKW reports research funding outside the submitted work from Amgen, Bristol-Myers Squibb, Pfizer, Janssen, Bayer, GSK, Novartis, the Hong Kong Research Grants Council, the Food and Health Bureau of the Government of the Hong Kong Special Administrative Region, National Institute for Health Research in England, European Commission and the National Health and Medical Research Council in Australia; and is a non-executive director of Jacobson Medical in Hong Kong and a consultant to the World Health Organization. The authors report no other conflicts of interest in this work.

References

1. Pedersen CB, Mors O, Bertelsen A, et al. A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. *JAMA Psychiatry*. 2014;71(5):573–581. doi:10.1001/jamapsychiatry.2014.16
2. Feigin VL, Nichols E, Alam T, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*. 2019;18(5):459–480. doi:10.1016/S1474-4422(18)30499-X
3. Steel Z, Marnane C, Iranpour C, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *Int J Epidemiol*. 2014;43(2):476–493. doi:10.1093/ije/dyu038
4. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers Dement*. 2013;9(1):63–75.e2. doi:10.1016/j.jalz.2012.11.007
5. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med*. 2007;146(5):317–325. doi:10.7326/0003-4819-146-5-200703060-00004
6. Allan CE, Valkanova V, Ebmeier KP. Depression in older people is underdiagnosed. *Practitioner*. 2014;258(1771):19–22, 2–3.

7. Wood T, Nance E. Disease-directed engineering for physiology-driven treatment interventions in neurological disorders. *APL Bioeng.* 2019;3(4):040901. doi:10.1063/1.5117299
8. El Abdellati K, De Picker L, Morrens M. Antipsychotic treatment failure: a systematic review on risk factors and interventions for treatment adherence in psychosis. *Front Neurosci.* 2020;14. doi:10.3389/fnins.2020.531763
9. Stevović LI, Repišti S, Radojčić T, et al. Non-pharmacological interventions for schizophrenia-analysis of treatment guidelines and implementation in 12 Southeast European countries. *Schizophrenia (Heidelb).* 2022;8(1):10. doi:10.1038/s41537-022-00226-y
10. Kheirbek RE, Fokar A, Little JT, et al. Association between antipsychotics and all-cause mortality among community-dwelling older adults. *The Journals of Gerontology: Series A.* 2019;74(12):1916–1921.
11. Tsai DHT, Chang WH, Lin HW, Lin SJ, Shao SC, Lai ECC. Post-discharge use of antipsychotics in patients with hospital-acquired delirium and associated risk of mortality - A population-based nested case-control study. *Asian J Psychiatry.* 2023;83:103533. doi:10.1016/j.ajp.2023.103533
12. Man KKC, Shao SC, Chang YC, et al. Cardiovascular and metabolic risk of antipsychotics in children and young adults: a multinational self-controlled case series study. *Epidemiol Psychiatr Sci.* 2021;30. doi:10.1017/S2045796021000494
13. 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
14. Huang AR, Mallet L, Rochefort CM, Egualé T, Buckeridge DL, Tamblin R. Medication-related falls in the elderly: causative factors and preventive strategies. *Drugs Aging.* 2012;29(5):359–376. doi:10.2165/11599460-000000000-00000
15. Weston J, Bromley R, Jackson CF, et al. Monotherapy treatment of epilepsy in pregnancy: congenital malformation outcomes in the child. *Cochrane Database Syst Rev.* 2016;11(11):Cd010224. doi:10.1002/14651858.CD010224.pub2
16. Oude Voshaar RC, Dhondt TDF, Fluiter M, et al. Study design of the Routine Outcome Monitoring for Geriatric Psychiatry & Science (ROM-GPS) project; a cohort study of older patients with affective disorders referred for specialised geriatric mental health care. *BMC Psychiatry.* 2019;19(1):182. doi:10.1186/s12888-019-2176-6
17. Shao SC, Lin YH, Chang KC, et al. Sodium glucose co-transporter 2 inhibitors and cardiovascular event protections: how applicable are clinical trials and observational studies to real-world patients? *BMJ Open Diabetes Res Care.* 2019;7(1). doi:10.1136/bmjdr-2019-000742
18. Reutfors J, Cesta CE, Cohen JM, et al. Antipsychotic drug use in pregnancy: a multinational study from ten countries. *Schizophr Res.* 2020;220:106–115. doi:10.1016/j.schres.2020.03.048
19. Lai ECC, Man KKC, 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
20. Lai ECC, Stang P, Kao Yang YH, Kubota K, Wong ICK, Setoguchi S. International multi-database pharmacoepidemiology: potentials and pitfalls. *Curr Epidemiol Rep.* 2015;2(4):229–238. doi:10.1007/s40471-015-0059-z
21. Man KKC, Shao SC, Chaiyakunapruk N, et al. Metabolic events associated with the use of antipsychotics in children, adolescents and young adults: a multinational sequence symmetry study. *Eur Child Adolesc Psychiatry.* 2022;31(1):99–120. doi:10.1007/s00787-020-01674-6
22. Ilomäki J, Bell JS, Chan AYL, et al. Application of healthcare ‘Big Data’ in CNS drug research: the example of the Neurological and mental health Global Epidemiology Network (NeuroGEN). *CNS Drugs.* 2020;34(9):897–913. doi:10.1007/s40263-020-00742-4
23. Leung TYM, Chan AYL, Chan EW, et al. Short- and potential long-term adverse health outcomes of COVID-19: a rapid review. *Emerg Microbes Infect.* 2020;9(1):2190–2199. doi:10.1080/22221751.2020.1825914
24. Man KKC, Chan EW, Ip P, et al. Prenatal antidepressant use and risk of attention-deficit/hyperactivity disorder in offspring: population based cohort study. *BMJ.* 2017;357:j2350. doi:10.1136/bmj.j2350
25. Ilomäki J, Lai ECC, Bell JS. Using clinical registries, administrative data and electronic medical records to improve medication safety and effectiveness in dementia. *Curr Opin Psychiatry.* 2020;33(2):163–169. doi:10.1097/YCO.0000000000000579
26. 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
27. Huang WC, Yang ASH, Tsai DHT, Shao SC, Lin SJ, Lai ECC. Association between recently raised anticholinergic burden and risk of acute cardiovascular events: nation-wide case-case-time-control study. *BMJ.* 2023;382:e076045. doi:10.1136/bmj-2023-076045
28. Raman SR, Man KKC, Bahmanyar S, et al. Trends in attention-deficit hyperactivity disorder medication use: a retrospective observational study using population-based databases. *Lancet Psychiatry.* 2018;5(10):824–835. doi:10.1016/S2215-0366(18)30293-1
29. 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
30. Lai ECC, Chang CH, Kao Yang YH, Lin SJ, Lin CY. Effectiveness of sulpiride in adult patients with schizophrenia. *Schizophr Bull.* 2013;39(3):673–683. doi:10.1093/schbul/sbs002
31. Wang Z, Chan AYL, Coghill D, et al. Association between prenatal exposure to antipsychotics and attention-deficit/hyperactivity disorder, autism spectrum disorder, preterm birth, and small for gestational age. *JAMA Intern Med.* 2021;181. doi:10.1001/jamainternmed.2021.4571
32. 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
33. Krysinska K, Sachdev PS, Breitner J, Kivipelto M, Kukull W, Brodaty H. Dementia registries around the globe and their applications: a systematic review. *Alzheimers Dement.* 2017;13(9):1031–1047. doi:10.1016/j.jalz.2017.04.005
34. Washington DL, Sun S, Canning M. Creating a sampling frame for population-based veteran research: representativeness and overlap of VA and department of defense databases. *J Rehabil Res Dev.* 2010;47(8):763–771. doi:10.1682/JRRD.2009.08.0127
35. 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
36. Sung SF, Chen SC, Hsieh CY, Li CY, Lai ECC, Hu YH. A comparison of stroke severity proxy measures for claims data research: a population-based cohort study. *Pharmacoepidemiol Drug Saf.* 2016;25(4):438–443. doi:10.1002/pds.3944
37. Shao SC, Lai ECC, Chen YH, Chan YY, Chen HY. Management of irrational self-purchase of hydroxychloroquine during the COVID-19 pandemic: experiences from the largest healthcare system in Taiwan. *J Patient Saf.* 2021;17(1):e43–e4. doi:10.1097/PTS.0000000000000795
38. Fortuny J, Gilseman A, Cainzos-Achirica M, et al. Study design and cohort comparability in a study of major cardiovascular events in new users of prucalopride versus polyethylene glycol 3350. *Drug Saf.* 2019;42(10):1167–1177. doi:10.1007/s40264-019-00836-z

39. Li Z, Kormilitzin A, Fernandes M, et al. Validation of UK Biobank data for mental health outcomes: a pilot study using secondary care electronic health records. *Int J Med Inform.* 2022;160:104704. doi:10.1016/j.ijmedinf.2022.104704
40. Socrates A, Maxwell J, Glanville KP, et al. Investigating the effects of genetic risk of schizophrenia on behavioural traits. *NPJ Schizophrenia.* 2021;7(1):. doi:10.1038/s41537-020-00131-2
41. Trifirò G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med.* 2014;275(6):551–561. doi:10.1111/joim.12159
42. Gini R, Sturkenboom MCJ, Sultana J, et al. Different strategies to execute multi-database studies for medicines surveillance in real-world setting: a reflection on the European model. *Clin Pharmacol Ther.* 2020;108(2):228–235. doi:10.1002/cpt.1833
43. FitzHenry F, Resnic FS, Robbins SL, et al. Creating a common data model for comparative effectiveness with the observational medical outcomes partnership. *Appl Clin Inform.* 2015;6(3):536–547. doi:10.4338/ACI-2014-12-CR-0121
44. Weeks J, Pardee R. Learning to share health care data: a brief timeline of influential common data models and distributed health data networks in U.S. health care research. *EGEMS (Washington, DC).* 2019;7:4. doi:10.5334/egems.279
45. Thurin NH, Pajouheshnia R, Roberto G, et al. From inception to ConcePTION: genesis of a network to support better monitoring and communication of medication safety during pregnancy and breastfeeding. *Clin Pharmacol Ther.* 2022;111(1):321–331. doi:10.1002/cpt.2476
46. Anderson TS, Ayanian JZ, Souza J, Landon BE. Representativeness of participants eligible to be enrolled in clinical trials of aducanumab for Alzheimer disease compared with medicare beneficiaries with Alzheimer disease and mild cognitive impairment. *JAMA.* 2021;326(16):1627–1629. doi:10.1001/jama.2021.15286
47. Weinmann S, Roll S, Schwarzbach C, Vauth C, Willich SN. Effects of Ginkgo biloba in dementia: systematic review and meta-analysis. *BMC Geriatr.* 2010;10:.. doi:10.1186/1471-2318-10-14
48. Yuan Q, Wang CW, Shi J, Lin ZX. Effects of Ginkgo biloba on dementia: an overview of systematic reviews. *J Ethnopharmacol.* 2017;195:1–9. doi:10.1016/j.jep.2016.12.005
49. Shao SC, Lai ECC, Huang TH, et al. The Chang Gung Research Database: multi-institutional real-world data source for traditional Chinese medicine in Taiwan. *Pharmacoepidemiol Drug Saf.* 2021;30(5):652–660. doi:10.1002/pds.5208
50. Koponen M, Paakinaho A, Lin J, Hartikainen S, Tolppanen AM. Identification of drugs associated with lower risk of Parkinson's disease using a systematic screening approach in a nationwide nested case-control study. *Clin Epidemiol.* 2022;14:1217–1227. doi:10.2147/CLEP.S381289
51. Paakinaho A, Koponen M, Tiihonen M, Kauppi M, Hartikainen S, Tolppanen A-M. Disease-modifying antirheumatic drugs and risk of Parkinson disease. *Nested Case-Control Study People Rheumatoid Arthritis.* 2022;98(12):e1273–e81.
52. Sunnarborg K, Tiihonen M, Huovinen M, Koponen M, Hartikainen S, Tolppanen AM. Association between different diabetes medication classes and risk of Parkinson's disease in people with diabetes. *Pharmacoepidemiol Drug Saf.* 2022;31(8):875–882. doi:10.1002/pds.5448

Clinical Epidemiology

Dovepress

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>