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

Daniel Hsiang-Te Tsai1,2, J Simon Bell3, Shahab Abtahi4, Brenda N Baak5, Marloes T Bazelier4, Ruth Brauer6, AdrienneYL Chan6–9, Esther W Chan9–12, Haoqian Chen13, Celine SL Chui9,14,15, Sharon Cook16, Stephen Crystal16, Poornam Gandhi13, Sirpa Hartikainen17, Frederick K Ho18, Shao-Ti Hsu1, Jenni Ilomäki18,2, Ju Hwan Kim19,10, Olaf H Klungel4, Marjaana Koponen17, Wallis CY Lau6,8,9, Kui Kai Lai20,21, Terry YS Lum22, Hao Luo22, Kenneth KC Man6,8,9, Jill P Pell18, Soko Setoguchi13,23, Shih-Chieh Shao1,24, Chin-Yao Shen1, Ju-Young Shin19,25,26, Patrick C Souverein4, Anna-Maija Tolppanen16,17, Li Wei6,9, Ian CK Wong6,8,9,27, 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 Pharmacoeconomics 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 Pharmacoepidemiology, -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 (D4H), 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. Mental disorders are globally widespread, impacting individuals from every corner of the world (lifetime prevalence: 29.2%, 25.9–32.6%). 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. 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. 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. 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. Some available pharmacoceutical treatments are associated with safety issues. For example, in older patients, the use of antipsychotics has been associated with mortality risk and psychotic medications have been associated with sedation, falls, arrhythmia, metabolic syndromes and extrapyramidal symptoms. 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. 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. 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. 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. 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. The importance of cross-country collaboration and multinational 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. 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. 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, 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.

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. 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 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

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

Notes: *With linked EHR. †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, Arizona, 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 NHRID, 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 NHRID, 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. NHRID, NHIS-NHID, PBS provided health claims details, although others did not or provided limited information (Supplementary Table 7). Four databases (ie, NHRID, 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, NHRID, 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...
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 disorders.
Table 2: Accessibility of Available Databases

<table>
<thead>
<tr>
<th>Source Type</th>
<th>Claims Databases</th>
<th>Electronic Health Records</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NHIRD</td>
<td>NHIS-NHID</td>
<td>PBS</td>
</tr>
<tr>
<td>Database</td>
<td></td>
<td></td>
<td>Medicaid</td>
</tr>
<tr>
<td>Specific policy for using</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Validation study</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Average costs (USD)</td>
<td>15,000</td>
<td>600</td>
<td>350</td>
</tr>
<tr>
<td>Period of obtaining access to the database</td>
<td>6 mo</td>
<td>5–6 mo</td>
<td>4 wk</td>
</tr>
<tr>
<td>Require IRB approval before application</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Waiting time for IRB approval</td>
<td>4–6 wk</td>
<td>2–3 wk</td>
<td>2 wk</td>
</tr>
<tr>
<td>IRB application fee (USD)</td>
<td>170</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

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).
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. 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

Figure 2 Database features.
the assessment of long-term outcomes of neurological and mental disorder treatments. Several examples are available from the literature review.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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,\textsuperscript{26} including Taiwan’s NHIRD, Hong Kong’s CDARS, the UK’s THIN, CPRD and the United States’ Medicare.\textsuperscript{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.\textsuperscript{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.\textsuperscript{46} Third, NeuroGEN can be used to expand the evidence base for traditional medications for the management of dementia or psychiatric disorders.\textsuperscript{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.\textsuperscript{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 fees 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**


