Application of healthcare ‘big data’ in CNS drug research: The example of The Neurological and mental health Global Epidemiology Network (NeuroGEN)

Running heading: The Neurological and mental health Global Epidemiology Network (NeuroGEN)

Authors:
Jenni Ilomäki¹, J Simon Bell¹,², Adrienne YL Chan³, Anna-Maija Tolppanen², Hao Luo⁴, Li Wei⁵, Edward Chia-Cheng Lai⁶, Ju-Young Shin⁷, Giorgia De Paoli⁸, Romin Pajouheshnia⁹, Frederick Ho¹⁰, Lorenna Reynolds¹, Kui Kai Lau¹¹, Stephen Crystal¹², Wallis CY Lau³,⁵, Kenneth KC Man³,⁵, Ruth Brauer⁶, Esther W Chan¹, Chin-Yao Shen⁶, Ju Hwan Kim⁷, Terry YS Lum⁴, Sirpa Hartikainen², Marjaana Koponen², Evelien Rooke⁸, Marloes Bazelier⁹, Olaf Klungel⁹, Soko Setoguchi¹³, Jill P Pell¹⁰, Sharon Cook¹², Ian CK Wong³

1. Centre for Medicine Use and Safety, Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, Melbourne, Victoria, Australia
2. School of Pharmacy, University of Eastern Finland, Kuopio, Finland
3. Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong SAR
4. Department of Social Work and Social Administration & Sau Po Centre on Ageing, The University of Hong Kong, Hong Kong SAR
5. Research Department of Practice and Policy, School of Pharmacy, University College London, London, United Kingdom
6. School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan
7. School of Pharmacy, Sungkyunkwan University, Suwon, Gyeong gi-do, South Korea
8. Medicines Monitoring Unit, Ninewells Hospital and Medical School, University of Dundee, Dundee, United Kingdom
9. Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute of Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, Netherlands
10. Institute of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom
11. Department of Medicine, The University of Hong Kong, Hong Kong SAR
12. Center for Health Services Research, Institute for Health, Health Care Policy, and Aging Research and School of Social Work, Rutgers University, New Brunswick, New Jersey, United States
13. Rutgers Robert Wood Johnson Medical School and School of Public Health and Center for Pharmacoepidemiology and Treatment Sciences, Institute for Health, Rutgers University, New Brunswick, New Jersey, United States
Corresponding author
Jenni Ilomäki
Centre for Medicine Use and Safety
Faculty of Pharmacy and Pharmaceutical Sciences
Monash University
Melbourne
Victoria
Australia
Email: Jenni.Iломаки@monash.edu

Co-corresponding author
Ian CK Wong
Department of Pharmacology and Pharmacy
The University of Hong Kong
Hong Kong SAR
Email: wongick@hku.hk

Declarations and Compliance with Ethical Standards
Funding: This research is supported by Dementia Australia Research Foundation – Yulgilbar Innovation Grant 2018 and The University of Hong Kong-University College London (HKU-UCL) and University College London-Peking University (UCL-PKU) Strategic Partnership Fund.

Conflicts of interest/Competing interests:
J Ilomäki has received consulting fees from AstraZeneca not related to the submitted work. JI has received grant funding from National Health and Medical Research Council (NHMRC), Victorian Government Department of Health and Human Services, Dementia Australia Research Foundation and National Breast Cancer Foundation. JS Bell is supported a NHMRC Boosting Dementia Leadership Fellowship. JS Bell has received grant funding from NHMRC, Victorian Government Department of Health and Human Services, Dementia Australia Research Foundation, GlaxoSmithKline and several aged care provider organizations not related to the submitted work and all paid to the employing institution. AYL Chan reports no receipt of research funding. AM Tolppanen has received grant funding from Academy of Finland. H Luo has received grants from the Research Grant Council (RGC, Hong Kong); as well as consulting fees from the Hong Kong Jockey Club Charities not related to the submitted work. L Wei has received research grants from European Medicines Agency and UK Medical Research Council, National Institute for Health Research and Charities outside of the submitted work. ECC Lai has received research grants from the Ministry of Science and Technology (MOST) of Taiwan, National Health Insurance Administration of Taiwan,
and National Cheng Kung University Hospital, outside the submitted work. **JY Shin** has received research funding from the Ministry of Food and Drug Safety, Ministry of Health and Welfare, National Research Foundation of Republic of Korea, Amgen, Pfizer, Hoffmann-La Roche, Dong-A ST, and Yungjin pharmaceutical outside the submitted work. **G De Paoli** reports no receipt of research funding. **R Pajouheshnia** has no funding or conflicts of interest to declare. **FK Ho** has received research funding from Wellcome Trust Institutional Strategic Support Fund via University of Glasgow outside the submitted work. **L Reynolds** is partly funded by the NHMRC Boosting Dementia Scheme. **KK Lau** has received grants from the Research Fund Secretariat of the Food and Health Bureau HKSAR, Boehringer Ingelheim, Pfizer, Amgen and Sanofi; as well as consulting fees and honorarium from Boehringer Ingelheim and Sanofi; all of which are not related to the submitted work. **S Crystal** is supported by R01-HS026001, DA047347, R18-HS023258, and UL1TR003017. **WCY Lau** has no conflicts of interest to declare. **KKC Man** is supported by the C W Maplethorpe Fellowship and had received consulting fees from IQVIA Inc. not related to the submitted work. **R Brauer** has no conflicts of interest to declare. **EW Chan** has received research grants from the Research Grants Council (RGC, Hong Kong), Narcotics Division of the Security Bureau of the Government of the Hong Kong SAR, Research Fund Secretariat of the Food and Health Bureau, National Natural Science Fund of China, Wellcome Trust, Bayer, Bristol-Myers Squibb, Pfizer, Janssen, Amgen, and Takeda, outside the submitted work. **C-Y Shen** is supported by research grants from the National Cheng Kung University Hospital, Tainan, Taiwan. **JH Kim** report no receipt of research funding. **TYS Lum** has received research grants from the Research Grants Council (RGC, Hong Kong), the Hong Kong Jockey Club Charities, the Social Welfare Department of the HKSAR Government, the Hong Kong Housing Society, the Simon KY Lee Foundation, and the Templeton World Charity Foundation, outside the submitted work. **S Hartikainen** has received lecturing fees from Astellas Pharma not related to the submitted work. **M Koponen** has no conflicts of interest to declare. **E Rooke** has no conflicts of interest to declare. **M Bazelier** has no funding or conflicts of interest to declare. **O Klungel** has no funding or conflicts of S Setoguchi has received research grants from NIH, Cystic Fibrosis Foundation, Pfizer and BMS and served as a consultant for US FDA, PMDA in Japan, Pfizer, Merck, and Medtronic. interest to declare. **JP Pell** has received research funding from the Scottish Government, Health Data Research UK, Medical Research Council, Wellbeing of Women, and National Institute for Health Research outside the submitted work. **S Cook** is supported by R01-HS026001, R18-HS023258, R01-DA045872, R01-DA047347 and UL1TR003017. **ICK Wong** has received research grants from Research Grants Council (RGC, Hong Kong), Innovative Medicines Initiative (IMI), Shire, Janssen-Cilag, Eli-Lilly, Pfizer, Bayer, and grants from European Union FP7 program, outside the submitted work.

**Ethics approval:** Not applicable  
**Consent to participate:** Not applicable
Consent for publication: All authors have consented for publication of this article
Availability of data and material: Not applicable
Code availability: Not applicable
Abstract

Neurological and psychiatric (mental health) disorders have a large impact on health burden globally. Cognitive disorders (including dementia) and stroke are leading causes of disability. Mental health disorders including depression contribute up to one third of total years lived with disability. The Neurological and mental health Global Epidemiology Network (NeuroGEN) is an international multi-database network that harnesses administrative and electronic medical records from Australia, Asia, Europe and North America. Using these databases NeuroGEN will investigate medication use and health outcomes in neurological and mental health disorders. A key objective of NeuroGEN is to facilitate high-quality observational studies to address evidence-practice gaps where randomized controlled trials do not provide sufficient information on medication benefits and risks that is specific to vulnerable population groups. International multi-database research facilitates comparisons across geographical areas and jurisdictions, increases statistical power to investigate small sub-populations or rare outcomes, permits early post-approval assessment of safety and effectiveness, and increases generalisability of results. Through bringing together international researchers in pharmacoepidemiology, NeuroGEN has the potential to be paradigm changing for observational research to inform evidence-based prescribing. The first focus of NeuroGEN will be to address evidence-gaps in treatment of chronic comorbidities in people with dementia.

Key points:

- Neurological and mental health disorders have a disproportionately large impact on global disease burden, but people with these disorders are often under-represented in randomized controlled trials (RCTs) and real-world evidence is lacking.
- International multi-database research using administrative data and electronic medical records provides an opportunity to conduct large and generalizable observational studies to generate new evidence to inform prescribing.
- NeuroGEN addresses evidence-gaps in treatment of neurological and mental health disorders by bringing together researchers and data from Australia, Asia, Europe and North America.
1 Introduction

1.1 The global burden of neurological and mental health disorders
Neurological disorders such as cognitive disorders (including dementia), stroke and Parkinson’s disease are leading causes of dependence and disability worldwide [1, 2]. Dementia has a global annual cost of US $818 billion [3]. The prevalence of age-related neurodegenerative disorders, including dementia and Parkinson’s disease, is expected to double over the next 20 years [1]. It was estimated that 43.8 million people were living with dementia in 2016 [4], with 7.7 million new people being diagnosed every year [5]. Over 6 million people worldwide have Parkinson’s disease, and the prevalence has doubled over a generation [6]. The total global burden of stroke is increasing, and close to 6 million people die because of stroke each year [7].

Psychiatric (mental health) disorders affect approximately 4.4% of the world’s population at any one point in time with an estimated 300 million people directly affected by depression in 2015 [8]. It is estimated that mental health disorders may be contributing to one third of total years lived with disability, depression being the most common disorder [9].

Optimizing care and support through appropriate pharmacological and non-pharmacological management can reduce burden in people with neurological and/or mental health disorders, their families, health-care systems and society.

1.2 Evidence-gaps in the treatment of people with neurological and mental health disorders
Reducing the social and economic burden of neurological and mental health disorders, particularly dementia, is a global health priority [3]. The World Health Organization (WHO) Ministerial Conference on Global Action Against Dementia highlighted the need for research to determine and ensure the optimal use of pharmacological treatments for symptoms of dementia [3]. There are currently clear evidence gaps affecting the quality of medication use in certain vulnerable populations, such as those with dementia. For example, participants included in randomized controlled trials (RCTs) do not necessarily represent the characteristics of people prescribed medications in routine clinical practice. Older people with neurodegenerative disorders are often excluded from RCTs [10] resulting in a lack of evidence for medication safety and effectiveness. This is despite people with neurodegenerative disorders often experiencing high rates of multimorbidity and treatment with multiple medications [11, 12]. For example, few people with dementia were eligible to participate in the pivotal direct oral anticoagulant (DOAC) RCTs [13], despite a high prevalence of cardiovascular and cerebrovascular disease in this population [11]. In RCTs of acetylcholinesterase inhibitors, participants have been notably younger than the real life population with Alzheimer’s disease [14].
Specific evidence regarding the benefits and risks of medications in people with dementia is lacking [10], yet results of a recent nationwide study demonstrated that people with dementia were more likely to be exposed to polypharmacy (dispensed five or more medications) than people without dementia [15]. Insufficient evidence may lead to reliance on evidence extrapolated from other populations or settings, or prescribing decisions based on assumed benefits and risks. This could compound prescribing uncertainty, or lead to inappropriate prescription of guideline recommended medications for comorbid conditions. The United Kingdom (UK) primary care data suggest comorbid depression is diagnosed in 17%, 21%, 18% and 32% of people with coronary heart disease, stroke, diabetes and dementia, respectively [16]. Despite being highly prevalent, people with diagnosed depression are often excluded from RCTs related to management of these conditions.

1.3 The role of administrative claims and electronic medical record data in central nervous system drug research

The rapid increase in the availability of administrative and electronic medical record (EMR) data has resulted in new potential for ‘big data’ research in medication safety and effectiveness [10]. These data are collected from hospitals, primary care medical practices and pharmacies. Clinical registries have also been established in primary and secondary care, often with linkage to administrative data. High quality multi-database observational studies of such ‘big data’ enables comparisons across geographical areas and jurisdictions, an increase in statistical power to investigate small sub-populations, rare outcomes, early post-approval assessment of safety and effectiveness, and an increased generalisability of findings.

2 The Neurological and Mental Health Global Epidemiology Network (NeuroGEN)

2.1 Description of NeuroGEN

NeuroGEN (https://www.neurogen.hku.hk/) is an international ‘big data’ collaboration platform established at a multidisciplinary meeting of 30 researchers from eight international geographical regions in Hong Kong in October 2018 [17]. NeuroGEN evolved out of the PharmAlliance collaboration in pharmacoepidemiology between Monash University, University College London (UCL) and University of North Carolina at Chapel Hill (UNC). PharmAlliance is a strategic partnership for staff and students in the Universities’ Pharmacy Schools: Monash University Faculty of Pharmacy and Pharmaceutical Sciences, UCL School of Pharmacy and UNC Eshelman School of Pharmacy. PharmAlliance provides strategic seed funding for multi-institutional initiatives in research, practice and education. The funds are an investment from each School to support PharmAlliance activities. The purpose of the October 2018 meeting was to explore data available in different jurisdictions, identify the breadth of clinical and methodological expertise, and to set research priorities. Research priority setting involved identifying 29 topics, of which six were
prioritized highest. A second multidisciplinary meeting was held in London in August 2019 which included new member institutions and researchers. At this meeting respective research groups presented and discussed their progress in relation to the existing and new topics. A map of current member regions is presented in Figure 1. Initial seed funding provided by PharmAlliance has been supplemented by grants from the Victorian Medical Research Acceleration Fund, University College London (UCL) - Peking University Strategic Partnership Grant, and University of Hong Kong - UCL Strategic Partnership Grant and Research Grant Council of Hong Kong. Dementia Australia Research Foundation – Yulgilbar Innovation Grant was received to investigate guideline-recommended medication use in people with dementia and chronic comorbidities. This Four Continents For Dementia (4C4D) program involves Australia, Hong Kong, UK and United States (US).

2.2 NeuroGEN member institutions and databases
NeuroGEN facilitates access to a global network of administrative and medical record data for the purpose of conducting multi-database observational research with a focus on neurological and mental health. Collectively, data are available for an estimated 100 million people with and without neurological and mental health disorders. There are significant variations of local regulatory and ethical framework between different parts of the world, therefore, each NeuroGEN partner works with the relevant data custodians and ethics committees to comply with the local legal and ethical requirements.

Data included in each of the databases are described in Table 1.

2.2.1 Monash University, Australia
Monash University’s Centre for Medicine Use and Safety (CMUS) comprises investigators in pharmacoepidemiology and clinical pharmacy. One of CMUS research priorities is to use administrative claims data to improve optimal use of medications, focusing on dementia and cardiovascular diseases [18, 19]. The primary data source is a 10% random sample of national dispensing data from Australia’s Pharmaceutical Benefits Scheme (PBS) covering 10% (≈2.5 million) of Australia’s population. This dataset is provided in a standard de-identified form by Services Australia by application. A newly established data source is Victorian-wide hospitalization database linked to the PBS, general medical practitioner data obtained through the Medicare Benefits Schedule (MBS), emergency department visits and mortality data. Victoria is the second most populous state in Australia with a population of 6.6 million. The linked Victorian data has been approved by all data custodians with a waiver of informed consent due to retrospective use of the data and that consent would not have been feasible to obtain.

2.2.2 University of Hong Kong (HKU), Hong Kong
The HKU team has conducted multi-database pharmacoepidemiological studies using EMRs [20-22]. The primary source of data is the Clinical Data Analysis and Reporting System (CDARS) managed by the Hospital Authority (HA) in Hong Kong. The HA is the sole public-funded health care provider, whose primary, secondary and tertiary care services are accessible to all Hong Kong residents (>7 million people). The CDARS includes records from all public hospitals, outpatient clinics and institutions under the HA. Research proposals are approved by the Research Ethics Committee under the HA. Informed patient consent is waived as the CDARS data used are de-identified.

2.2.3 University College London (UCL), United Kingdom
The UCL School of Pharmacy team’s research focuses on neurodegenerative and cardiovascular diseases, diabetes, child health and pregnancy [23-25]. The main source of data is the Health Improvement Network (THIN). THIN is a nationwide database that contains electronic primary care records from UK general practices for 15 million individuals [26]. THIN covers a 6% representative sample of the UK population. Multiple diagnoses and lifestyle variables recorded in THIN database including cardiovascular diseases, diabetes, obesity, smoking have been used and validated for pharmacoepidemiological research [27]. THIN is subject to the UK Data Protection Act 2018 and EU General Data Protection Regulation (GDPR). Data obtained has been anonymised and consent was previously collected by the general practices where patients can opt-out.

2.2.4 University of Glasgow, United Kingdom
The University of Glasgow has expertise on vascular neurological and cardiometabolic diseases [28-30]. The primary source of data is the UK Biobank. The UK Biobank recruited 502,536 participants aged 39–72 years from the general population between 2007-2010. Participants attended one of 22 assessment centres across England, Scotland, and Wales where they completed a self-administered questionnaire and face-to-face interview, and trained staff took a series of measurements including height, weight, and blood pressure. Mortality, hospitalization, and primary care consultations are available through data linkage. The UK Biobank has acquired explicit informed consent from all participants.

2.2.5 University of Dundee, Scotland
The Medicines Monitoring Unit (MEMO) Research group at the University of Dundee and Ninewells Hospital conducts observational studies [31, 32] and large decentralized clinical trials. MEMO currently have approximately 40,000 patients randomized into clinical trials. For pharmacoepidemiological studies MEMO researchers use data from the Information Services Division (ISD) of National Services Scotland, which is part of the public National Health Service (NHS). ISD provides health information, health intelligence, statistical services and advice that support the NHS with the goal to improve Scotland's health. The Service holds health-related data
which in some cases cover an individual from before birth (with the mother’s antenatal records) to their death.

2.2.6 National Cheng Kung University (NCKU), Taiwan

NCKU focuses on pharmacoepidemiology and big data research using claims data based on the National Health Insurance program in Taiwan [33, 34]. The National Health Insurance Database (NHID) was launched in 1995. The program covers over 99% of Taiwan’s population (25 million people) and enrolled more than 90% of hospitals and clinics. Ministry of Health and Welfare (MOHW) established a Health and Welfare Data Centre (HWDC), a data repository site that centralizes the NHID. The NHID includes medications, medical visits and procedures recorded in ambulatory, in-patient and emergency services. In addition, a multi-institutional electronic medical records database, the Chang Gung Research Database (CGRD) [35] containing clinical data such as pathological and laboratory results, is available to serve as external validation data for the NHID. The CGRD includes 1.3 million outpatients and 0.2 million inpatients in Taiwan [36, 37]. Due to retrospective nature of the provided data, informed consent is not required for either the NHID or the CGRD.

2.2.7 SungKyunKwan University (SKKU), South Korea

The Korean team focus on analyses of the National Health Insurance System (NHIS) claims database [38, 39] and multi-database studies. The NHIS in South Korea achieved universal coverage of the entire population in 1989. The database contains diagnostic and prescribing data for approximately 50 million Koreans. The claims database includes data on each individual’s age, sex, diagnoses (ICD-10) and prescription medications. Information on prescription medications includes generic name, the date of prescription, duration, and route of administration. Due to retrospective nature of the provided data, informed consent is not required.

2.2.8 University of Eastern Finland (UEF), Finland

The Kuopio Research Centre of Geriatric Care focuses on pharmacoepidemiology in people with Alzheimer’s disease and Parkinson’s disease. This includes aetiological research, drug utilization studies and outcome studies [40, 41]. Primary sources of data are the nationwide MEDication use and ALZheimer’s disease (MEDALZ) study [42] on people with Alzheimer’s disease and the Finnish Medication and Parkinson’s disease (FINPARK) study on people with Parkinson’s disease [43]. Both studies include a matched cohort to facilitate comparisons to persons without these conditions and are derived from Finnish nation-wide databases including medication dispensing data, hospital discharge data and mortality data. The MEDALZ cohort includes incident cases of Alzheimer’s disease diagnosed from 2005-2011 and the FINPARK study includes incident cases of Parkinson’s disease diagnosed from 1996-2015 with ongoing follow-up. Both MEDALZ and FINPARK data are used in
pseudonymised form. The research proposals were approved by the data custodians and according to Finnish legislation, other approvals or informed consent are not needed as the study is based on pseudonymized register data, and the participants are not contacted.

2.2.9 Utrecht University (UU), the Netherlands
UU’s Pharmacoepidemiology and Clinical Pharmacy group has a clinical, policy and methodological focus. The UU group has methodological expertise in methods to prevent and/or control for confounding, analysis of effect modification and conducting multi-database analysis. The primary data sources used for large pharmacoepidemiological studies include the Dutch PHARMO database (www.pharmo.nl) and the UK Clinical Practice Research Datalink (CPRD) [44]. CPRD is subject to the UK Data Protection Act 2018 and EU GDPR. Data obtained has been anonymised and consent was previously collected by the general practices where patients can opt-out. In PHARMO, patient information is deidentified and the requirement for individual consent is waived unless an intervention is planned. All use of the data requires approval by the independent Compliance Committee STIZON / PHARMO Institute, in compliance with the Netherlands Personal Data Protection Act and Medical Treatment Contract Act. Data access is funded by the Utrecht Institute for Pharmaceutical Sciences. The UU group coordinate the European Research Network of Pharmacovigilance and Pharmacoepidemiology (EU PE&PV) and have developed novel methodologies for the conduct of multi-country, multi-database studies on variability of medication use and health outcomes [45-47].

2.2.10 Rutgers University, United States
The Center for Health Services Research and Center of the Pharmacoepidemiology and Treatment Sciences at Rutgers’ Institute for Health are interdisciplinary groups with research focussing on the use and outcomes of medications across large, diverse usual-care populations in the US and other countries [48, 49]. Researchers at Rutgers have worked on studies particularly on the use and outcomes of central nervous system (CNS) drugs including opioid use disorders; use and outcomes of antipsychotics; treatment of adults with severe mental illness; use and safety of selective serotonin reuptake inhibitors in pregnant women; and psychotropic treatment for children. Main data sources include health insurance data from Center for Medicare and Medicaid services, which is updated annually: a 20% sample of Medicare patients representative of the US older people and people with end-stage renal diseases, and 45 State Medicaid Analytic Extracts (MAX) representative of low income population including pregnant women and children. According to the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rules, informed consent was not required as the data are collected originally for insurance purposes, and secondary use of data for researchers is conducted without person identifiers.

2.3 Ongoing case studies and initiatives
2.3.1 Case Study 1: Adherence and persistence to acetylcholinesterase inhibitors

Acetylcholinesterase inhibitors (AChEIs) are the most widely prescribed medications for dementia, although efficacy [50, 51] and cost-effectiveness [51, 52] are modest. Non-adherence and non-persistence reduce potential benefits, with a systematic review of five RCTs reporting discontinuation is associated with a significant decline in cognition and worsening of neuropsychiatric symptoms [53]. This highlights the importance of persistence in maximising benefit. This study will investigate adherence and persistence to AChEIs across the NeuroGEN partners. Australia, South Korea and Taiwan have analysed their respective data using a common study protocol. The study will utilize proportion of days covered to estimate adherence from medication dispensing and prescribing databases. Persistence will be estimated using a pre-specified gap of no dispensing or prescribing. This study will permit a comparison adherence and persistence using standardized definitions and methodology. This program of work is funded through the NHMRC Boosting Dementia Leadership Fellowship Scheme.

2.3.2 Case Study 2: Predicting dementia and survival from cognitive footprints of electronic health records using machine learning

Based on the ‘cognitive footprint’ of medical history, this population-based case-control study will aim to develop and validate an algorithm for predicting dementia using machine learning [54]. The algorithm will be trained using territory-wide EMRs from the CDARS in Hong Kong, and tested both locally and externally in other databases (e.g. the UK THIN and the Finnish MEDALZ). The CDARS currently hosts records from more than 70,000 people with dementia diagnoses between 2001 and 2018. Potential protective/risk factors, which will be selected based on the cognitive footprint theory, will be modelled holistically. It is anticipated that the modelling will include analyses of diagnostic data, laboratory test results and the prescription of antidepressants, antipsychotics, statins and polypharmacy. Other than a set of Hong Kong-specific factors, a set of common factors that are shared by other databases will be identified to maximize interoperability. The subsequent common algorithm, to be derived from real-world data in Hong Kong, may then be suitable for embedding into other health information systems. Patients with a high risk or likelihood of dementia can be efficiently identified to permit targeting of risk-reduction programs. A secondary objective of this project is to estimate survival from the point of recorded diagnosis of dementia in Hong Kong, Canada, Finland, Germany, Korea, Taiwan, UK and US. This project will aggregate large population-scale data from different geographical regions. The project is ongoing and expected to complete by June, 2022. This project is funded by Research Grant Council of Hong Kong under the Early Career Scheme.

2.3.3 Case study 3: Mortality of people with Parkinson’s disease across geographical areas
In a meta-analysis of inception cohorts, Parkinson’s disease was associated with 1.5 times higher mortality [55]. The same meta-analysis demonstrated major heterogeneity in mortality ratios stratified by sex, and identified a need for further high-quality studies of mortality in Parkinson’s disease. Specifically, there is a lack of large-scale population-based inception cohorts with long-term follow-up. This study will investigate survival of people with Parkinson’s disease following diagnosis, as well as possible geographical differences in mortality ratios and factors that predict higher mortality. The project is in its initiation phase. This project will be coordinated from Finland, and data from Finland, Hong Kong, Korea, Australia and the UK will be utilized. Additional countries will be included once confirmed with the corresponding investigators. Funding for this project has been applied. Once secured, the development of the common study protocol will commence.

2.3.4 Case Study 4: Capacity building

One of the objectives of NeuroGEN is capacity building and training the next generation of pharmacoepidemiologists. This is being achieved by providing opportunities to early career researchers, including PhD candidates and post-doctoral researchers. For example, PhD students from Monash University, Naresuan University (Thailand) and Princess Norah Bint Abdul Rahman University (Saudi Arabia) have conducted exchanges to UCL to conduct a pharmacoepidemiological studies [56-59]. Similarly, a PhD student from Monash University has conducted an exchange to HKU, and researchers from Utrecht University and UCL have conducted exchanges to Monash University. A bi-lateral exchange of post-doctoral researchers from University of Eastern Finland and Monash University has taken place [60]. These exchanges have been funded through Royal Golden Jubilee Ph.D. Program (Thailand) Newton Fund (UK), Saudi Arabian Ministry of Higher Education, the Australian Government Endeavour Fellowship Scheme, Monash Doctoral Program and NHMRC Boosting Dementia Leadership Scheme.

3 Discussion and Future Directions

3.1 Discussion

Multi-national collaboration with data from multiple regions globally is a growing opportunity to conduct large generalizable observational studies that address research questions with international relevance. Use of a common protocol approach (CPA) and common data models (CDM) can facilitate large multi-database studies that address topics of international public health importance. NeuroGEN is currently using both the CPA and CDM. Although the CPA is more straightforward to implement, it requires close communication between investigators to ensure that all analyses are conducted consistently. A CDM is a sophisticated data platform supporting secondary use of data across multiple databases. The major advantage of CDM is that analyses are controlled by the use of standardized data structure, terminology, variable definitions and analytical program. Such distributed network
approach in which data partners maintain physical and operational control over the data in their existing environments also addresses data privacy issues across jurisdictions, because data are not shared. However, establishing the CDM requires a considerable investment of time and resources to convert native databases into the CDM.

Existing examples of consortia with similar approach include the Asian Pharmacoepidemiology Network (AsPEN). AsPEN uses modified distributed networks with a common data structure across databases to allow single analytic programs to be used in each site [61]. Some NeuroGEN investigators are part of AsPEN and have established skills in multi-database studies. Another similar example is the Canadian Network for Observational Drug Effect Studies (CNODES) where databases across provinces in are analyses with the same approach [62].

NeuroGEN investigators have created a simplified CDM based on the Observational Medical Outcomes Partnership (OMOP) CDM [63], containing all relevant information to conduct analyses for ongoing projects. A standalone analysis programme for each study will be developed based on the NeuroGEN CDM. Because the data structure and terminologies are identical among the converted databases, the analyses can be conducted in each home institution. Each site will generate a standardized results file which will then be collected by the coordinating site. Figure 2 presents the structure of the NeuroGEN CDM. Previous applications of similar conversions include for example paediatric use of prescription medications [64].

Dementia Australia and Yulgilbar Foundation have funded the development of the CDM for four databases focussing on dementia research. The databases include the Australian linked health data, the US Medicare data, the UK THIN data and the Hong Kong CDARS data. The respective investigators are currently working together synchronising the databases into the CDM format to investigate the use of guideline-recommended medications for chronic comorbidities in people with and without dementia.

3.2 Future directions

The third NeuroGEN investigator meeting will be conducted in conjunction with the Asian Conference on Pharmacoepidemiology in Seoul, Korea in October, 2020. NeuroGEN is currently in discussion with partner research groups in other geographical regions, including Oceania and South America. The collaboration will continue to seek to address topics of global importance to better management of neurological and mental health disorders.

3.3 Conclusions
NeuroGEN is a recent initiative addressing medication use and outcomes in people with neurological and mental health disorders. NeuroGEN uses similar approach to other multi-database initiatives such as AsPEN and CNODES. However, NeuroGEN is the only global multi-database network addressing specifically issues arising in neurological and mental health field, and more widely in psychopharmacology. This will address significant evidence-gaps in this under-researched field.

References:


**Figure legends:**

Fig 1. Map of the NeuroGEN sites

Fig 2. The proposed common data model structure for the NeuroGEN
<table>
<thead>
<tr>
<th>Database name</th>
<th>Region</th>
<th>Size</th>
<th>Years of data</th>
<th>Demographic variables</th>
<th>Source for medication use</th>
<th>Source for medical conditions</th>
<th>Source for laboratory results</th>
<th>Availability of other clinical data</th>
<th>Availability of lifestyle information</th>
<th>Date and cause of death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victorian Linked health data (Cohort extracted from a state-wide Victorian Admitted Episodes Dataset)</td>
<td>Australia</td>
<td>615 000 people hospitalised for myocardial infarction, ischaemic stroke, diabetes or hip fracture</td>
<td>2006-2018</td>
<td>Age, sex, ethnicity, language spoken, geographic area, marital status</td>
<td>Dispensed reimbursed medications, PBS item code, date of dispensing, quantity, strength</td>
<td>Hospital diagnoses and procedures (ICD-10-AM)</td>
<td>Referrals only (Medicare)</td>
<td>N/A</td>
<td>N/A</td>
<td>Month, year and cause of death</td>
</tr>
<tr>
<td>10% random sample of national PBS dispensing data</td>
<td>Australia</td>
<td>2.5 million</td>
<td>2005-2019</td>
<td>Year of birth, sex, state</td>
<td>Dispensed reimbursed medications, PBS item code, authority code where relevant, date of dispensing, quantity, strength, prescriber</td>
<td>Selected medical conditions can be inferred from medication dispensings using the Rx-Risk Index tool and prescriptions requiring specific diagnosis for reimbursement (authority)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Year of death</td>
</tr>
<tr>
<td>MEDALZ (Cohort and data linkages extracted from nationwide registers)</td>
<td>Finland</td>
<td>70,719 persons with AD and 282,862 comparison persons without AD (all community dwelling at the time of diagnosis); data linkage currently until 2015</td>
<td>Incident AD diagnoses from 2005-2011</td>
<td>Age, sex, hospital district, occupational social class (since 1970)</td>
<td>Reimbursed dispensings of prescription medications (1995 onwards) data on e.g. ATC codes, medication names, pack size, dispensed amount</td>
<td>-Hospital stays from 1972 onwards: diagnoses (ICD-8 until 1986, ICD-9 until 1995, ICD-10) -Procedure codes (Sairaalaliitto until 1995, NOMESCO since 1996) -Entitlement for special reimbursements for</td>
<td>N/A</td>
<td>institutionalizations, required level of assistance at hospital discharge</td>
<td>N/A</td>
<td>Date and cause of death</td>
</tr>
<tr>
<td>Source Description</td>
<td>Country</td>
<td>Sample Size / Data Details</td>
<td>Incident Diagnoses from/or Period</td>
<td>Intake methods</td>
<td>Data Linkage Currently Until</td>
<td>Date and cause of death</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FinPark (Cohort and data linkages extracted from nationwide registers)</td>
<td>Finland</td>
<td>21,683 persons with PD and 146,306 comparison persons without PD (all community dwelling at the time of diagnosis); data linkage currently until 2016</td>
<td>1972, national criteria consistent with international guidelines</td>
<td>Same as MEDALZ</td>
<td>N/A</td>
<td>Date and cause of death</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital Authority’s Clinical Data Analysis and Reporting System (CDARS)</td>
<td>Hong Kong</td>
<td>&gt;7 million active, &gt;11 million total</td>
<td>1995 - 2019</td>
<td>Prescription and dispensing information including date, dispensing status, quantity, duration, daily dose)</td>
<td>Same as MEDALZ</td>
<td>N/A</td>
<td>Date and cause of death</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Health Insurance system (NHIS) Database</td>
<td>Korea</td>
<td>50 million</td>
<td>2003 - 2018</td>
<td>Hospital medication order, Pharmacy claims</td>
<td>Same as MEDALZ</td>
<td>Via linkage to Family Medicine records</td>
<td>Date and cause of death</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database Network</td>
<td>Country</td>
<td>Active (prior to linkage)</td>
<td>Time Period</td>
<td>Data Elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHARMO database network</td>
<td>The Netherlands</td>
<td>4.2 million</td>
<td>Pre 2000-2019</td>
<td>Age, sex, geographic area, Pharmacy dispensing data (sample of in-hospital treatments available), Linkage to nationwide hospitalization database, in-patient hospital pharmacy database, (2 million), GP database (2.5 million)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Services Division (ISD)</td>
<td>Scotland</td>
<td>5 million</td>
<td>2000-2019</td>
<td>Date of Birth, Gender, Ethnic Group, Marital Status, GMC No. of referring Dr/Dentist/Nurse, Allied Healthcare Professional, GP Practice Code, NHS Number, Postcode, UCPN, PIS; Dispensing System, Prescription and dose duration are decoded, Hospital episodes of care data for acute conditions consisting of ICD and OPCS codes. There are also databases on maternity/birth record/child health/cancer registration/office for National Statistics death certification; data by ICD codes, birth, death, marriage. Via a separate system (Albasoft) access to GP data across nearly all practices in Scotland, A series of regional databases called SCI-Store containing all labs data linked to the CHI number, PACS system, a Scottish wide record of all imaging in Scotland; Accident and Emergency attendance data; Vaccination records; Ambulance calls databases. Via linkage to the National datasets for Lifestyle Alcohol Brief Interventions, Drug &amp; Alcohol Treatment Waiting Times, Drug Prevalence Estimates, National Drug Related Deaths Database, National Sexual Health System, Scottish Drug Misuse Database, Scottish School Adolescent Lifestyle and Substance Use Survey, Smoking Cessation Database; Social Deprivation data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database</td>
<td>Country</td>
<td>Sample Size</td>
<td>Time Period</td>
<td>Data Elements</td>
<td>Source Data</td>
<td>Smoking, BMI, alcohol consumption</td>
<td>Date of death</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------</td>
<td>----------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chang Gung Research Database (CGRD)</td>
<td>Taiwan</td>
<td>1.3 million outpatients and 0.2 million inpatients</td>
<td>2008 - most updated</td>
<td>Age, sex, year of birth, ethnicity, language spoken, marital status, socio-economic status</td>
<td>Prescription information, including medication code, strength, dose frequency, quantity, date of supply</td>
<td>Hospital and clinic diagnoses and procedure ICD-9-CM/ ICD-10</td>
<td>All laboratory data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Health Insurance Database (NHID)</td>
<td>Taiwan</td>
<td>23 million</td>
<td>2003 - 2017</td>
<td>Age, sex, date of birth, geographic area</td>
<td>Prescription information, including medication code, strength, dose frequency, quantity, date of supply</td>
<td>Hospital and clinic diagnoses and procedure ICD-9-CM/ ICD-10</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Health Improvement Network (THIN)</td>
<td>UK</td>
<td>&gt;4 million active, 13 million in total</td>
<td>1990-2018 (best quality data 2004-2018)</td>
<td>Age, sex, year of birth, registration status, transfer out date, region, ethnicity, language spoken, marital status, socio-economic status</td>
<td>Primary care prescriptions (ATC codes/BNF product codes) (date, quantity, duration, daily dose)</td>
<td>Primary care clinical data, referral data, immunisation data (READ codes)</td>
<td>Test results (Type of test, result, normal range of result, unit of measure)</td>
<td>Possible linkage to HES</td>
<td>GP recorded BMI, smoking, alcohol consumption</td>
<td>Date of death</td>
</tr>
<tr>
<td>UK Biobank</td>
<td>UK</td>
<td>0.5 million</td>
<td>2007-2010 for baseline data</td>
<td>Age, sex, ethnicity, area-based deprivation index, education, income, and occupation</td>
<td>Self-report at baseline and linked primary care data on prescription</td>
<td>Self-report at baseline and linked mortality, hospitalization, and primary care data</td>
<td>Majority of participants at baseline; n=18k repeated measures in 2012-13; additional</td>
<td>Brain and heart MRI data from subset of participants; cognitive tests,</td>
<td>Self-report lifestyle at baseline; small subset in repeated measurements</td>
<td>Date and cause of death</td>
</tr>
<tr>
<td>Clinical Practice Research Datalink (CPRD)</td>
<td>UK</td>
<td>4.4 million active, &gt;11.3 million total patients meeting quality criteria</td>
<td>Pre-2000 - 2019</td>
<td>Age, sex, month and year of birth, registration status, transfer out date, region, ethnicity, deprivation index (linked)</td>
<td>Primary care prescriptions (Gemscript/BNF product codes) (date of treatment, quantity, duration, daily dose)</td>
<td>Primary care clinical data, referral data, immunization data (READ codes)</td>
<td>Test results (Type of test, result, normal range of result, unit of measure)</td>
<td>genetic / genomic data</td>
<td>from linked primary care data</td>
<td>Possible linkage to HES, ONS, National Cancer Registry</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>20% sample of the Medicare</td>
<td>US</td>
<td>10 million per year</td>
<td>2007-2017</td>
<td>Date of birth, sex, county of residence, race, enrolment information</td>
<td>Outpatient dispensings including dates of dispensing, National Drug Codes, strength, quantity dispensed, days supply</td>
<td>Inpatient data ICD-9-CM/ICD-10 codes for diagnoses and procedures</td>
<td>Laboratory tests ordered</td>
<td>Medical equipment, home care, long-term care</td>
<td>N/A</td>
<td>Date and cause of death</td>
</tr>
<tr>
<td>Medicaid Analytic Abstracts (MAX), 45 states</td>
<td>US</td>
<td>&gt;152 million</td>
<td>2001-2012; 2013 (26 states); 2014 (14 states)</td>
<td>Date of birth, sex, state and county of residence, race/ethnicity, enrolment information (e.g., basis of eligibility, dual Medicare status)</td>
<td>Paid prescription drug claims, including national drug codes, dispense dates, quantity dispensed and days supplied</td>
<td>Inpatient, outpatient and long-term care claims with ICD-9-CM/ICD-10-CM codes for diagnoses and procedures</td>
<td>Inpatient, outpatient and long-term care claims with ICD-9-CM/ICD-10-CM, HCPCS, CPT procedure codes</td>
<td>Long-term care and diagnostic codes for palliative care, drug overdose, emergency visits</td>
<td>N/A</td>
<td>National Death Index date and ICD-10 cause of death codes for 2001-2007</td>
</tr>
</tbody>
</table>

Abbreviations: PBS Pharmaceutical Benefits Scheme; ICD International Classification of Diseases; N/A not applicable; MEDALZ MEDication use and ALZheimer's disease; AD Alzheimer's disease; ATC Anatomical Therapeutic Classification; NOMESCO Nordic Medico-Statistical Committee; FINPARK Finnish Medication and Parkinson's disease; PD Parkinson's disease; GP general practitioner; GMC General Medical Council; NHS National Health Service; UCPN Unique Care Pathway Number; PIS Prescribing Information System; OPCS Office of Population Censuses and Surveys; SCI Scottish Care Information; CHI Community Health