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 Ilomäki[1], J Simon Bell[1,2], Adrienne YL Chan[3], Anna-Maija Tolppanen[2], Hao Luo[4], Li Wei[5], Edward Chia-Cheng Lai[6], Ju-Young Shin[7], Giorgia De Paoli[8], Romin Pajouheshnia[9], Frederick Ho[10], Lorenna Reynolds[1], Kui Kai Lau[11], Stephen Crystal[12], Wallis CY Lau[3,5], Kenneth KC Man[3,5], Ruth Brauer[5], Esther W Chan[3], Chin-Yao Shen[6], Ju Hwan Kim[7], Terry YS Lum[4], Sirpa Hartikainen[2], Marjaana Koponen[2], Evelien Rooke[8], Marloes Bazelier[9], Olaf Klungel[9], Soko Setoguchi[13], Jill P Pell[10], Sharon Cook[12], Ian CK Wong[3]

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.Ilomaki@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-Lily, 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, focussing 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 focusses 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 from1996-2015 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:

1. Fereshtehnejad SM, Vosoughi K, Heydarpour P, Sepanlou SG, Farzadfar F, Tehrani-Banihashemi A, et al. Burden of neurodegenerative diseases in the Eastern Mediterranean Region, 1990-2016: findings from the Global Burden of Disease Study 2016. Eur J Neurol. 2019;26:1252-65.
2. Feigin VL, Roth GA, Naghavi M, Parmar P, Krishnamurthi R, Chugh S, et al. Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet Neurol. 2016;15:913-24.
3. Shah H, Albanese E, Duggan C, Rudan I, Langa KM, Carrillo MC, et al. Research priorities to reduce the global burden of dementia by 2025. Lancet Neurol. 2016;15:1285-94.
4. GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18:88-106.
5. World Health Organization. Dementia: a public health priority. Geneva: World Health Organization; 2012.
6. GBD Parkinson's Disease Collaborators. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2018;17:939-53.
7. GBD Stroke Collaborators. Global, regional, and national burden of stroke, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18:439-58.
8. World Health Organization. Depression and other common mental disorders: Global health estimates. Geneva: World Health Organization; 2017.
9. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry. 2016;3:171-8.
10. Ilomaki J, Lai EC, Bell JS. Using clinical registries, administrative data and electronic medical records to improve medication safety and effectiveness in dementia. Current Opinion in Psychiatry. 2020, in press.
11. Clague F, Mercer SW, McLean G, Reynish E, Guthrie B. Comorbidity and polypharmacy in people with dementia: insights from a large, population-based cross-sectional analysis of primary care data. Age Ageing. 2017;46:33-9.
12. Santos Garcia D, Suarez Castro E, Exposito I, de Deus T, Tunas C, Aneiros A, et al. Comorbid conditions associated with Parkinson's disease: A longitudinal and comparative study with Alzheimer disease and control subjects. J Neurol Sci. 2017;373:210-5.
13. Fanning L, Ryan-Atwood TE, Bell JS, Meretoja A, McNamara KP, Darzins P, et al. Prevalence, Safety, and Effectiveness of Oral Anticoagulant Use in People with and without Dementia or Cognitive Impairment: A Systematic Review and Meta-Analysis. J Alzheimers Dis. 2019;71:1375-8.
14. Leinonen A, Koponen M, Hartikainen S. Systematic Review: Representativeness of Participants in RCTs of Acetylcholinesterase Inhibitors. PLoS One. 2015;10:e0124500.

15. Kristensen RU, Norgaard A, Jensen-Dahm C, Gasse C, Wimberley T, Waldemar G. Polypharmacy and Potentially Inappropriate Medication in People with Dementia: A Nationwide Study. J Alzheimers Dis. 2018;63:383-94.
16. Guthrie B, Payne K, Alderson P, McMurdo ME, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. Bmj. 2012;345:e6341.
17. NeuroGEN. Neurological and mental health Global Epidmeiology Network (NeuroGEN). 2020 [cited 2020 March 4]; Available from: https://www.neurogen.hku.hk/
18. Ilomäki J, Fanning L, Keen C, Sluggett JK, Page AT, Korhonen MJ, et al. Trends and Predictors of Oral Anticoagulant Use in People with Alzheimer's Disease and the General Population in Australia. J Alzheimers Dis. 2019;70:733-45.
19. Ofori-Asenso R, Liew D, Lalic S, Mazidi M, Magliano DJ, Ademi Z, et al. Adherence, Persistence, and Switching Among People Prescribed Sodium Glucose Co-transporter 2 Inhibitors: A Nationwide Retrospective Cohort Study. Adv Ther. 2019;36:3265-78.
20. Raman SR, Man KKC, Bahmanyar S, Berard A, Bilder S, Boukhris T, et al. Trends in attention-deficit hyperactivity disorder medication use: a retrospective observational study using population-based databases. Lancet Psychiatry. 2018;5:824-35.
21. Man KKC, Chan EW, Ip P, Coghill D, Simonoff E, Chan PKL, et al. Prenatal antidepressant use and risk of attention-deficit/hyperactivity disorder in offspring: population based cohort study. Bmj. 2017;357:j2350.
22. Man KKC, Coghill D, Chan EW, Lau WCY, Hollis C, Liddle E, et al. Association of Risk of Suicide Attempts With Methylphenidate Treatment. JAMA Psychiatry. 2017;74:1048-55.
23. Brauer R, Lau WCY, Hayes JF, Man KKC, Osborn DPJ, Howard R, et al. Trazodone use and risk of dementia: A population-based cohort study. PLoS Med. 2019;16:e1002728.
24. Wei L, Lai EC, Kao-Yang YH, Walker BR, MacDonald TM, Andrew R. Incidence of type 2 diabetes mellitus in men receiving steroid 5alpha-reductase inhibitors: population based cohort study. Bmj. 2019;365:l1204.
25. George J, Majeed W, Mackenzie IS, Macdonald TM, Wei L. Association between cardiovascular events and sodium-containing effervescent, dispersible, and soluble drugs: nested case-control study. Bmj. 2013;347:f6954.
26. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Inform Prim Care. 2011;19:251-5.
27. Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. Pharmacoepidemiol Drug Saf. 2007;16:393-401.
28. Ho FK, Gray SR, Welsh P, Petermann-Rocha F, Foster H, Waddell H. Nonlinear associations of fat and carbohydrate intake with cardiovascular disease and mortality: prospective cohort study of UK Biobank participants. BMJ. 2020;in Press.
29. Welsh C, Welsh P, Celis-Morales CA, Mark PB, Mackay D, Ghouri N, et al. Glycated Hemoglobin, Prediabetes, and the Links to Cardiovascular Disease: Data From UK Biobank. Diabetes Care. 2020;43:440-5.
30. Sillars A, Ho FK, Pell GP, Gill JMR, Sattar N, Gray S, et al. Sex differences in the association of risk factors for heart failure incidence and mortality. Heart. 2020;106:203-12.
31. Gilsenan A, Fortuny J, Cainzos-Achirica M, Cantero OF, Flynn RWV, Garcia-Rodriguez L, et al. Cardiovascular Safety of Prucalopride in Patients with Chronic Constipation: A Multinational Population-Based Cohort Study. Drug Saf. 2019;42:1179-90.
32. Mackenzie IS, Morant SV, Wei L, Thompson AM, MacDonald TM. Spironolactone use and risk of incident cancers: a retrospective, matched cohort study. British journal of clinical pharmacology. 2017;83:653-63.

33. Pottegard A, Pedersen SA, Schmidt SAJ, Lee CN, Hsu CK, Liao TC, et al. Use of hydrochlorothiazide and risk of skin cancer: a nationwide Taiwanese case-control study. Br J Cancer. 2019;121:973-8.
34. Hsieh CY, Su CC, Shao SC, Sung SF, Lin SJ, Kao Yang YH, et al. Taiwan's National Health Insurance Research Database: past and future. Clin Epidemiol. 2019;11:349-58.
35. Shao SC, Chan YY, Kao Yang YH, Lin SJ, Hung MJ, Chien RN, 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;28:593-600.
36. Shao SC, Chang KC, Hung MJ, Yang NI, Chan YY, Chen HY, et al. Comparative risk evaluation for cardiovascular events associated with dapagliflozin vs. empagliflozin in real-world type 2 diabetes patients: a multi-institutional cohort study. Cardiovasc Diabetol. 2019;18:120.
37. Shao SC, Lin YH, Chang KC, Chan YY, Hung MJ, Kao Yang YH, 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.
38. Won DY, Byun SJ, Jeong JS, Shin JY. Association Between Acetylcholinesterase Inhibitors and Osteoporotic Fractures in Older Persons With Alzheimer's Disease. J Am Med Dir Assoc. 2020.
39. Joung KI, Shin JY, Kim S, Cho SI. Anticholinergic Use Among the Elderly With Alzheimer Disease in South Korea: A Population-based Study. Alzheimer Dis Assoc Disord. 2020.
40. Taipale H, Rysa J, Hukkanen J, Koponen M, Tanskanen A, Tiihonen J, et al. Long-term thiazide use and risk of low-energy fractures among persons with Alzheimer's disease-nested case-control study. Osteoporos Int. 2019;30:1481-9.
41. Tapiainen V, Lavikainen P, Koponen M, Taipale H, Tanskanen A, Tiihonen J, et al. The Risk of Head Injuries Associated With Antipsychotic Use Among Persons With Alzheimer's disease. J Am Geriatr Soc. 2020.
42. Tolppanen AM, Taipale H, Koponen M, Lavikainen P, Tanskanen A, Tiihonen J, et al. Cohort profile: the Finnish Medication and Alzheimer's disease (MEDALZ) study. BMJ open. 2016;6:e012100.
43. Paakinaho A, Karttunen N, Koponen M, Taipale H, Tolppanen AM, Hartikainen S, et al. Incidence of muscle relaxant use in relation to diagnosis of Parkinson's disease. Int J Clin Pharm. 2020.
44. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44:827-36.
45. Platt RW, Platt R, Brown JS, Henry DA, Klungel OH, Suissa S. How pharmacoepidemiology networks can manage distributed analyses to improve replicability and transparency and minimize bias. Pharmacoepidemiol Drug Saf. 2019.
46. Ibanez L, Sabate M, Vidal X, Ballarin E, Rottenkolber M, Schmiedl S, et al. Incidence of direct oral anticoagulant use in patients with nonvalvular atrial fibrillation and characteristics of users in 6 European countries (2008-2015): A cross-national drug utilization study. Br J Clin Pharmacol. 2019;85:2524-39.
47. Bazelier MT, Van Staa TP, Bentzen J, Vestergaard P, Uitdehaag BMJ, Leufkens HGM, et al. Multiple sclerosis and fracture risk: traditional meta-analysis versus mega-analysis of individual patient data. OA Epidemiology. 2013;1:1-9.
48. Williams AR, Samples H, Crystal S, Olfson M. Acute Care, Prescription Opioid Use, and Overdose Following Discontinuation of Long-Term Buprenorphine Treatment for Opioid Use Disorder. Am J Psychiatry. 2020;177:117-24.
49. Stroup TS, Gerhard T, Crystal S, Huang C, Tan Z, Wall MM, et al. Comparative Effectiveness of Adjunctive Psychotropic Medications in Patients With Schizophrenia. JAMA Psychiatry. 2019;76:508-15.
50. Birks J, Grimley Evans J, Iakovidou V, Tsolaki M, Holt FE. Rivastigmine for Alzheimer's disease. The Cochrane database of systematic reviews. 2009:CD001191.

51. Birks JS, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. The Cochrane database of systematic reviews. 2018;6:CD001190.
52. Bond M, Rogers G, Peters J, Anderson R, Hoyle M, Miners A, et al. The effectiveness and cost-effectiveness of donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (review of Technology Appraisal No. 111): a systematic review and economic model. Health Technol Assess. 2012;16:1-470.
53. O'Regan J, Lanctot KL, Mazereeuw G, Herrmann N. Cholinesterase inhibitor discontinuation in patients with Alzheimer's disease: a meta-analysis of randomized controlled trials. J Clin Psychiatry. 2015;76:e1424-31.
54. Rossor M, Knapp M. Can we model a cognitive footprint of interventions and policies to help to meet the global challenge of dementia? Lancet. 2015;386:1008-10.
55. Macleod AD, Taylor KS, Counsell CE. Mortality in Parkinson's disease: a systematic review and meta-analysis. Mov Disord. 2014;29:1615-22.
56. Fanning L, Lau WCY, Mongkhon P, Man KKC, Bell JS, Ilomaki J, et al. Safety and Effectiveness of Direct Oral Anticoagulants vs Warfarin in People With Atrial Fibrillation and Dementia. J Am Med Dir Assoc. 2020.
57. Mongkhon P, Fanning L, Lau WCY, Tse G, Lau KK, Wei L, et al. Oral anticoagulant and reduced risk of dementia in patients with atrial fibrillation: A population-based cohort study. Heart Rhythm. 2020.
58. Mongkhon P, Naser AY, Fanning L, Tse G, Lau WCY, Wong ICK, et al. Oral anticoagulants and risk of dementia: A systematic review and meta-analysis of observational studies and randomized controlled trials. Neurosci Biobehav Rev. 2019;96:1-9.
59. Alsharif AA, Wei L, Ma T, Man KKC, Lau WCY, Brauer R, et al. Prevalence and Incidence of Dementia in People with Diabetes Mellitus. J Alzheimers Dis. 2020.
60. Sluggett JK, Koponen M, Bell JS, Taipale H, Tanskanen A, Tiihonen J, et al. Metformin and risk of Alzheimer's disease among community-dwelling people with diabetes: a national case-control study. J Clin Endocrinol Metab. 2019.
61. Lai EC, Ryan P, Zhang Y, Schuemie M, Hardy NC, Kamijima Y, et al. Applying a common data model to Asian databases for multinational pharmacoepidemiologic studies: opportunities and challenges. Clin Epidemiol. 2018;10:875-85.
62. Platt RW, Henry DA, Suissa S. The Canadian Network for Observational Drug Effect Studies (CNODES): Reflections on the first eight years, and a look to the future. Pharmacoepidemiol Drug Saf. 2020;29 Suppl 1:103-7.
63. FitzHenry F, Resnic FS, Robbins SL, Denton J, Nookala L, Meeker D, et al. Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. Appl Clin Inform. 2015;6:536-47.
64. Brauer R, Wong ICK, Man KK, Pratt NL, Park RW, Cho SY, et al. Application of a Common Data Model (CDM) to rank the paediatric user and prescription prevalence of 15 different drug classes in South Korea, Hong Kong, Taiwan, Japan and Australia: an observational, descriptive study. BMJ open. 2020;10:e032426.

Figure legends:

Fig 1. Map of the NeuroGEN sites

Fig 2. The proposed common data model structure for the NeuroGEN

Table 1. Summary of the content of the databases and their geographical locations

Database name	Region	Size	Years of data	Demographic variables	Source for medication use	Source for medical conditions	Source for laboratory results	Availability of other clinical data	Availability of lifestyle information	Date and cause of death
Victorian Linked health data (Cohort extracted from a state-wide Victorian Admitted Episodes Dataset)	Australia	615 000 people hospitalised for myocardial infarction, ischaemic stroke, diabetes or hip fracture	2006-2018	Age, sex, ethnicity, language spoken, geographic area, marital status	Dispensed reimbursed medications, PBS item code, date of dispensing, quantity, strength	Hospital diagnoses and procedures (ICD-10-AM)	Referrals only (Medicare)	N/A	N/A	Month, year and cause of death
10% random sample of national PBS dispensing data	Australia	2.5 million	2005-2019	Year of birth, sex, state	Dispensed reimbursed medications, PBS item code, authority code where relevant, date of dispensing, quantity, strength, prescriber	Selected medical conditions can be inferred from medication dispensings using the Rx-Risk Index tool and prescriptions requiring specific diagnosis for reimbursement (authority)	N/A	N/A	N/A	Year of death
MEDALZ (Cohort and data linkages extracted from nationwide registers)	Finland	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	Incident AD diagnoses from 2005-2011	Age, sex, hospital district, occupational social class (since 1970)	Reimbursed dispensings of prescription medications (1995 onwards) data on e.g. ATC codes, medication names, pack size, dispensed amount	-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	N/A	institutionalizations, required level of assistance at hospital discharge	N/A	Date and cause of death

					in defined daily doses, strength, formulation, costs	chronic conditions (since 1972, national criteria consistent with international guidelines) -Detailed data on cancer from the cancer register				
FinPark (Cohort and data linkages extracted from nationwide registers)	Finland	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	Incident diagnoses from 1996-2015	Same as MEDALZ	Same as MEDALZ	Same as MEDALZ	N/A	Same as MEDALZ	N/A	Date and cause of death
Hospital Authority's Clinical Data Analysis and Reporting System (CDARS)	Hong Kong	>7 million active, >11 million total	1995 - 2019	Sex, Year of birth, Month of birth, Race, Ethnicity, Location of patient	Prescription and dispensing information including date, dispensing status, quantity, duration, daily dose)	Hospital diagnoses and procedure ICD-9-CM/ICD-10	Laboratory test orders, laboratory test results	Diagnosis, inpatient, outpatient, accident and emergency department admissions and discharges records, payment method	Via linkage to Family Medicine records	Date and cause of death
National Health Insurance system (NHIS) Database	Korea	50 million	2003 - 2018	Age, sex, geographic area, insurance type, income level	Hospital medication order, Pharmacy claims	Hospital diagnoses and procedure (ICD-10-CD)	N/A	N/A	Via linkage to the national health screening program database	Month and year of death

PHARMO database network	The Netherlands	4.2 million active (prior to linkage)	Pre 2000-2019	Age, sex, geographic area	Pharmacy dispensing data (sample of in-hospital treatments available) (date of treatment, quantity, duration, daily dose)	Linkage to nationwide hospitalization database, in-patient hospital pharmacy database, (2 million), GP database (2.5 million)	Linkage to clinical laboratory database (1.2 million)	Linkage to nationwide registries (cancer, pathology, perinatal)	N/A	Date of death
Information Services Division (ISD) Scotland	Scotland	5 million	2000-2019	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 & 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	Date and cause of death

Chang Gung Research Database (CGRD)	Taiwan	1.3 million outpatients and 0.2 million inpatients	2008-most updated	Age, sex, year of birth, ethnicity, language spoken, marital status, socio-economic status	Prescription information, including medication code, strength, dose frequency, quantity, date of supply	Hospital and clinic diagnoses and procedure ICD-9-CM/ ICD-10	All laboratory data	N/A	Smoking, BMI, alcohol consumption	Date of death
National Health Insurance Database (NHID)	Taiwan	23 million	2003 - 2017	Age, sex, date of birth, geographic area	Prescription information, including medication code, strength, dose frequency, quantity, date of supply	Hospital and clinic diagnoses and procedure ICD-9-CM/ ICD-10	N/A	N/A	N/A	Date of death
The Health Improvement Network (THIN)	UK	>4 million active, 13 million in total	1990-2018 (best quality data 2004-2018)	Age, sex, year of birth, registration status, transfer out date, region, ethnicity, language spoken, marital status, socio-economic status	Primary care prescriptions (ATC codes/BNF product codes) (date, quantity, duration, daily dose)	Primary care clinical data, referral data, immunisation data (READ codes)	Test results (Type of test, result, normal range of result, unit of measure)	Possible linkage to HES	GP recorded BMI, smoking, alcohol consumption	Date of death
UK Biobank	UK	0.5 million	2007-2010 for baseline data	Age, sex, ethnicity, area-based deprivation index, education, income, and occupation	Self-report at baseline and linked primary care data on prescription	Self-report at baseline and linked mortality, hospitalization, and primary care data	Majority of participants at baseline; n=18k repeated measures in 2012-13; additional	Brain and heart MRI data from subset of participants; cognitive tests,	Self-report lifestyle at baseline; small subset in repeated measurements	Date and cause of death

							from linked primary care data	genetic / genomic data		
Clinical Practice Research Datalink (CPRD)	UK	4.4 million active, >11.3 million total patients meeting quality criteria	Pre-2000 - 2019	Age, sex, month and year of birth, registration status, transfer out date, region, ethnicity, deprivation index (linked)	Primary care prescriptions (Gemscript/BNF product codes) (date of treatment, quantity, duration, daily dose)	Primary care clinical data, referral data, immunization data (READ codes)	Test results (Type of test, result, normal range of result, unit of measure)	Possible linkage to HES, ONS, National Cancer Registry	GP recorded BMI, smoking, alcohol consumption	Date of death (possible linkage to ONS for death statistics/ cause)
20% sample of the Medicare	US	10 million per year	2007-2017	Date of birth, sex, county of residence, race, enrolment information	Outpatient dispensings including dates of dispensing, National Drug Codes, strength, quantity dispensed, days supply	Inpatient data ICD-9-CM/ ICD-10 codes for diagnoses and procedures	Laboratory tests ordered	Medical equipment, home care, long-term care	N/A	Date and cause of death
Medicaid Analytic Abstracts (MAX), 45 states	US	>152 million	2001-2012; 2013 (26 states); 2014 (14 states)	Date of birth, sex, state and county of residence, race/ethnicity, enrolment information (e.g., basis of eligibility, dual Medicare status)	Paid prescription drug claims, including national drug codes, dispense dates, quantity dispensed and days supplied	Inpatient, outpatient and long-term care claims with ICD-9-CM/ICD-10-CM codes for diagnoses and procedures	Inpatient, outpatient and long-term care claims with ICD-9-CM/ICD-10-CM, HCPCS, CPT procedure codes	Long-term care and diagnostic codes for palliative care, drug overdose, emergency visits	N/A	National Death Index date and ICD-10 cause of death codes for 2001-2007

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

Index; PACS picture archiving and communication system; BMI Body Mass Index; UK United Kingdom; BNF British National Formulary; MRI Magnetic resonance imaging; HES Hospital Episode Statistics; ONS Office for National Statistics; US United States; HCPCS Healthcare Common Procedure Coding System; CPT Current Procedural Terminology