

## Multigenerational health research using population-based linked databases: an international review

Naomi C. Hamm<sup>1</sup>, Amani F. Hamad<sup>1</sup>, Elizabeth Wall-Wieler<sup>1,2</sup>, Leslie L. Roos<sup>1,2</sup>, Oleguer Plana-Ripoll<sup>3</sup>, and Lisa M. Lix<sup>1</sup>

Submission History	
Submitted:	17/06/2021
Accepted:	13/08/2021
Published:	07/10/2021

<sup>1</sup>Department of Community Health Sciences, University of Manitoba, Winnipeg, MB, CANADA, R3E 0W3

<sup>2</sup>Manitoba Centre for Health Policy, University of Manitoba, Winnipeg, MB, CANADA, R3E 3P5

<sup>3</sup>National Centre for Register-based Research, Department of Economics and Business Economics, Aarhus University, Aarhus, DENMARK, 8210

### Abstract

#### Abstract

Family health history is a well-established risk factor for many health conditions but the systematic collection of health histories, particularly for multiple generations and multiple family members, can be challenging. Routinely-collected electronic databases in a select number of sites worldwide offer a powerful tool to conduct multigenerational health research for entire populations. At these sites, administrative and healthcare records are used to construct familial relationships and objectively-measured health histories. We review and synthesize published literature to compare the attributes of routinely-collected, linked databases for three European sites (Denmark, Norway, Sweden) and three non-European sites (Canadian province of Manitoba, Taiwan, Australian state of Western Australia) with the capability to conduct population-based multigenerational health research. Our review found that European sites primarily identified family structures using population registries, whereas non-European sites used health insurance registries (Manitoba and Taiwan) or linked data from multiple sources (Western Australia). Information on familial status was reported to be available as early as 1947 (Sweden); Taiwan had the fewest years of data available (1995 onwards). All centres reported near complete coverage of familial relationships for their population catchment regions. Challenges in working with these data include differentiating biological and legal relationships, establishing accurate familial linkages over time, and accurately identifying health conditions. This review provides important insights about the benefits and challenges of using routinely-collected, population-based linked databases for conducting population-based multigenerational health research, and identifies opportunities for future research within and across the data-intensive environments at these six sites.

#### Keywords

family health history; multigenerational; observational research; population registries; record linkage; routinely-collected data

\*Corresponding Author:

Email Address: [lisa.lix@umanitoba.ca](mailto:lisa.lix@umanitoba.ca) (Lisa M. Lix)



## Introduction

Family health history is a well-established risk factor for many health conditions, including heart disease, Type 2 diabetes, some cancers, and some mental health conditions. Genetics and shared environments are reflected in the family health history, making it a valuable component of multifactorial health risk assessments [1] from both clinical and public health perspectives. Accordingly, systematic collection of family health history information has been recommended [2, 3], despite the associated costs, challenges of capturing accurate and complete information, and privacy considerations.

Linked, routinely-collected electronic databases, which were originally developed for government administration, particularly for healthcare administration and management, are a widely-used resource for population health research [4, 5]. They have also been used in a select number of sites worldwide to construct family health histories for entire populations. Three European sites (Denmark, Norway, Sweden) and three non-European sites (Canadian province of Manitoba, Taiwan, Australian state of Western Australia) have the unique capability to define familial relationships and objectively-measured health histories from administrative and healthcare electronic records. Studies from these sites have made important contributions to understanding associations between familial health histories and individual outcomes for a broad range of health conditions or life events [6–14].

Routinely-collected electronic databases have many advantages for constructing family health histories: they can provide objective measures of health, complete or near-complete coverage of the population of a country or region, and span multiple decades, generations, and family members [5]. The data afford researchers a unique opportunity to investigate potential genetic and environmental effects at a population level. However, these data have some limitations because they were not originally intended to be used for research.

This paper reviews and synthesizes published literature to describe the attributes of routinely-collected, population-based data for multigenerational health histories that are available at these six sites. We examine the types of data available and explore the challenges associated with using routinely-collected, population-based electronic databases for multigenerational health research. We also provide examples of the types of studies and health conditions that have been conducted to date across these sites. We conclude this review by describing potential opportunities for future research within and across the data-intensive environments at these six sites.

## Literature search strategy and included sites

We searched published literature about the data available and the multigenerational health research that has been conducted using the following data sources at these sites: the Civil Registration System in Denmark, the Norwegian Family Based Life Course Linkage, the Multi-generation Register in Sweden, the Manitoba Population Research Data Repository in the Canadian province of Manitoba, Taiwan's National Health Insurance Research Database, and the

Western Australian Family Connections Genealogical Project. We searched PubMed and Google Scholar for articles published up to October 2020. The search was conducted using terms (and their variations) such as “multigenerational”, “parent offspring”, “chronic disease”, “administrative health data” electronic medical records”, “electronic health records”, “data linkage”, and “data centre”. We also used the “similar articles” feature in PubMed and manually searched the reference lists of identified papers to identify additional publications. While we initially focused on papers that described the methods and data sources used to achieve familial linkages at the six sites (12 papers identified in total), we also identified health conditions and events that have been the focus of published studies.

Table 1 provides summary information about the data attributes at the six sites. Key characteristics are described below.

## Identifying family relationships and structures

Population registries and health insurance registries are the primary sources of information about familial structures at these six sites. Denmark, Norway, and Sweden primarily use population registries to determine family structures [15–21]; Manitoba and Taiwan use health insurance registries [8, 22–25] to determine family structures. The Western Australian Family Connections Genealogical Project uses linked birth, marriage, and death registries to determine familial relationships [26].

Population registries contain information on residents living in a catchment area. While the data in these registries are collected and maintained for administrative purposes such as tax collection, the registries also contain information on familial relationships. For example, the Danish Civil Registration System contains information on all permanent residents of Denmark as well as children born in Denmark [15, 16]. Individuals are assigned a unique identification number (known as a CPR-number), which can be used to identify family structures. Each individual's registration record contains the following information: sex; date of birth; legal parents' CPR-numbers, if applicable; spouse's CPR-number; place of birth; and place of residence, emigration, immigration, or date of disappearance (i.e., current residence unknown) [15]. Sibling information is not directly available, but linkages can be established based on maternal and paternal CPR-numbers [15]. Additional registries, such as the adoption register and birth register can be used to ascertain parent information, although these data sources do not have the same temporal coverage as the Danish Civil Registration System [17–19, 27].

In Norway's National Persons Registry, permanent residents are identified via a personal identity code [20]; this code can also be used to establish linkages between parents and offspring. The Norwegian Family Based Life Course Linkage also uses census data, the Educational Registry, and the Cause of Death Registry to identify familial relationships and structures missing in the National Personal Registry (i.e., in older cohorts before the personal identity code was implemented) [20]. However, these familial relationships are

Table 1: Characteristics of electronic health data for population-based multigenerational health research at six sites

Site	Name of database or registry	Source data	Number of generations	Family relationships	Methods for determining family structures and relationships	Linkable databases
<i>European Sites</i>						
Denmark	Danish Civil Registration System (CRS)	CRS registration	Not reported	Legal parent & child Siblings (inferred) Spouses	Previously: families residing at same residence (1968-1978) 1978 onwards - legal parental status	Linkable to Danish National Health Service Register, Danish Cancer Registry, Danish Register of Causes of Death, Danish National Patient Registry, National Prescription Database, Pathology Database, Western Denmark Heart Registry, Danish Stroke Registry, Medical Birth Register, and Adoption Register
Norway	Norwegian Family Based Life Course Linkage	Statistics Norway National Personal Registry Educational Registry Cause of Death Registry	Not reported	Parent & child Spouses	Household information from census for those where relationships were not indicated in the National Persons Registry	Linkable to Norway Cause of Death Registry, Medical Birth Registry of Norway, Sickness and Disability Registry, Cohort of Norway
Sweden	Swedish Multi-generation Register	National registration number Total Population Register Statistics Sweden Medical Birth Register	Not reported	Biological parent & child Adoptive parent & child Siblings	For married or recently widowed mothers, husband is seen as father. Otherwise paternity is established by acknowledgment or by court order	Linkable to Swedish Medical Birth Register, Swedish National Inpatient Register, Swedish Prescribed Drug Register, Swedish Cancer Register, Swedish Cause of Death Register
<i>Non-European Sites</i>						
Manitoba, Canada	Manitoba Population Research Data Repository	Vital Statistics Birth records Health insurance registrations	Up to three generations	Parent & child Siblings	Family unit registration Birth records  Vital Statistics	Linkable to hospital, physician, nursing home, home care, vital statistics, prescription drug, cancer registry, education, family services, income assistance, social, and justice data

(Continued)

Table 1: Continued

Site	Name of database or registry	Source data	Number of generations	Family relation-ships	Methods for determining family structures and relationships	Linkable databases
Taiwan	National Health Insurance Research Database	Taiwan Birth Registry  Health insurance registrations	Not reported	Biological parent & child Siblings Spouses	Identifiers in database for the insured individuals and their dependents Children must have birth certificate or DNA testing to be considered dependent	Linkable to ambulatory care visits, inpatient visits, prescription data, medical personnel databases, use of medical facilities, health screening data, welfare/society data, birth, death, and maternal data
Western Australia, Australia	Western Australian Family Connections Genealogical Project	Birth, death, and marriage registrations Midwife records Hospital records	Up to three generations A limited number of four-generation linkages are available	Biological parent & child Siblings Spouses (not divorces)	Primarily birth registrations supplemented with information from death, marriage, and midwife registrations	Birth, death and marriage registrations, electoral roll, hospital morbidity, emergency department presentations, mental health information, midwives notifications, and cancers

primarily inferred based on household information, rather than direct relationship linkages [20].

In Sweden, residents are assigned a national registration number, which is maintained in the Total Population Register. The Multi-generation Register uses information from the Total Population Register, Statistics Sweden, as well as information from additional studies undertaken to improve parent-offspring linkages [21], in order to identify family relationships. The Multi-generation Register contains the following data on registered individuals: registration number, sex, country of birth, biological parents' registration number, biological parents' date of birth, and biological parents' country of birth. As well, this register contains the adoptive parents' registration number, date of birth, country of birth, date of immigration, and date of adoption [21]. The total number of children linked to the registered mother and father is also available, as well as the registered individual's position in the mother's family (e.g., first child, second child) [21]. In non-adoptive cases, paternity is determined through the national register, where if a mother is married or recently widowed at the time of birth, the husband is determined to be the father [21]. For other cases, paternity is determined through acknowledgement or via court order.

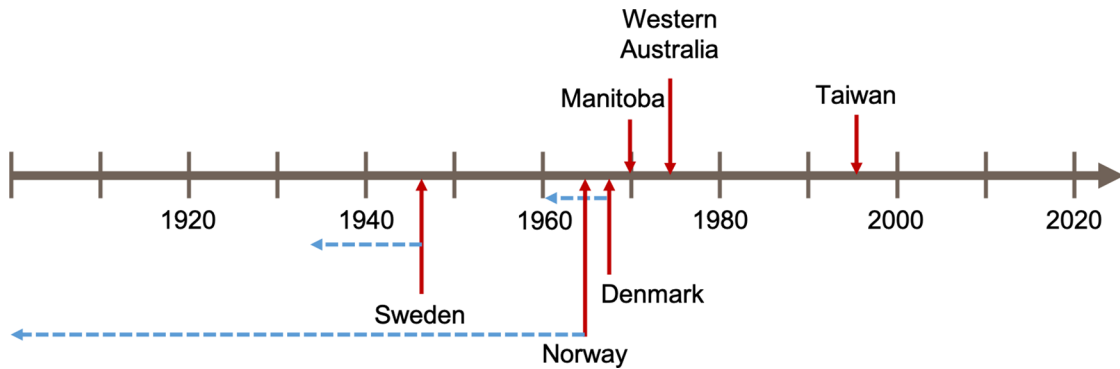
Health insurance registries contain information on individuals eligible to receive healthcare coverage, often from publicly funded sources. In Taiwan, spouses and blood relatives such as parents, grandparents, children, and grandchildren can be claimed as dependents of an insured person [8, 22]. Thus, sibling and parent-offspring linkages can be determined from

these health insurance registration files. In the province of Manitoba in Canada, a health registration number is given to all immediate family members residing in the province, including the registrant, spouse (or common-law spouse), and all children who are dependent on the registrant, including child, step-child, incapacitated child (i.e., dependent beyond the age of 18), and grandchild [25]. Familial relationship information is also supplemented using birth records [24].

In contrast to the databases described above, the Western Australian Family Connections Genealogical Project determines family relationships and structures by linking and cross-referencing birth, death, and marriage registrations, as well as midwife and hospital records [26]. Biological relationships and degree of relatedness can be determined from these data sources.

## Data coverage for family relationships and structures

Based on our literature review, we ascertained that all six sites have linkable records that extend at least 25 years. A timeline of data coverage across sites is provided in Figure 1. Denmark's Civil Registration System was first introduced in 1968 [15]. However, data related to parental links in the system are considered to be virtually correct since 1960 [15, 16]. Norway's personal identity code was introduced in 1964; however, the Norwegian Family Based Life Course Linkage also

Figure 1: Temporal coverage of linked electronic health data for population-based multigenerational health research at six sites<sup>a,b</sup>

<sup>a</sup>Red arrows indicate start of data temporal coverage

<sup>b</sup>Blue arrows indicate that the site reports having data coverage extending prior to the collection of familial relationships

uses information from additional data sources (i.e., census, Educational Registry, and Cause of Death Registry) and therefore contains information on individuals born as early as 1900 [20]. The national registration number in the Swedish Multi-generation Register was implemented in 1947 [21]. With supplemental data sources, records are available for individuals born from 1932 onwards and those alive on January 1, 1961 [21]. Taiwan's National Health Insurance Research Database has the most limited temporal coverage with data extending back to 1995 [22]. The Manitoba Population Research Data Repository contains data from 1970 onward [25], and the Western Australian Family Connections Genealogical Project contains data from 1974 onward [26].

Temporal coverage is crucial for identifying multiple generations. Western Australia reported being able to identify up to three generations in their database, and up to four generations for some individuals [26]. Manitoba reported identifying up to three generations [24]. Given the years of data available, it is reasonable to assume that two generations (i.e., parent-offspring) are identifiable in Taiwan's data, and three or four generations are identifiable in the Nordic countries, although we did not find any studies that explicitly stated the number of identifiable generations in these sites.

All sites have near complete capture of residents in their catchment area. In Denmark, Norway, and Sweden, registration is required by law, ensuring all residents (past and present) are captured in the data [15, 20, 21]. In Manitoba and Taiwan, registration is required to receive health insurance coverage. These sites have universal and publically funded healthcare systems; therefore, most residents are contained within the database (i.e., >99%) [22]. Coverage in Western Australia is slightly lower as individuals are required to have healthcare records in at least one of the linked databases to be included. These limitations appear minimal however, with about 94% of the population having familial links [26].

## Identifying health histories via data linkage

Based on our literature review, numerous routinely-collected health databases can be used to construct family health histories in the six sites. The International Classification of

Diseases (ICD) is used to record diagnoses in many of these health databases.

Multiple healthcare registers are available in Denmark, Norway, and Sweden for linkage to familial information. In Denmark, these include the Danish National Health Service Register, Danish Cancer Registry, the Danish Register of Causes of Death, Danish National Patient Registry, National Prescription Database, Pathology Database, Western Denmark Heart Registry, the Danish Stroke Registry, the Danish Medical Birth Register and the Danish Adoption Register [17, 18, 28–30]. Data in the Norwegian Family Based Life Course Linkage can be linked to the Norway Cause of Death Registry, the Medical Birth Registry of Norway, the Sickness and Disability Registry and the Cohort of Norway, which is a collection of health surveys [20]. Linkage to census data is also possible. The Swedish Multi-generation Register can be linked to Sweden's Medical Birth Register [31], National Inpatient Register [32], Prescribed Drug Register [33], Cancer Register [34], and Cause of Death Register [35]. In addition, linkage to more than 100 Swedish healthcare quality registers is described in published literature [36], although the data in each of these registers do not cover the entire population of Sweden. Multiple versions of ICD are used to capture diagnosis information in these European databases, given that they span multiple years of administrative health and clinical data.

In the Canadian province of Manitoba, linkable databases include physician billing claims, hospital discharge records, emergency department records, prescription medication dispensations, electronic medical records, cancer registry data, home care data, and long term care (i.e., nursing home) data. Social databases, which capture information on income assistance for low-income individuals and child welfare data are also available, as are justice data. Physician billing claims and hospital discharge abstracts extend back to 1970, but other data have a shorter period of availability. A complete list of databases is provided online ([http://umanitoba.ca/faculties/health\\_sciences/medicine/units/chs/departamental\\_units/mchp/](http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departamental_units/mchp/)). Hospital data are coded using three different ICD versions (ICDA-8, 1970-1979; ICD-9-CM, 1979-2004; ICD-10-CA, 2004 onward) [25]. Physician data are coded using two ICD versions (ICDA-8, 1970-1979; ICD-9-CM, 1979 onward) [25]. Prescription medication dispensations are coded using the Anatomical



Therapeutic Chemical (ATC) Classification System developed by the World Health Organization [25].

Health databases available in Taiwan encompass ambulatory care visits, inpatient care visits, prescription dispensations at contracted pharmacies, health service utilization at medical facilities, medical personnel registries, a medical facilities registry, and a registry for catastrophic illnesses. Screening data, birth and death records, disease and injury databases, and social data are also available. A listing of databases is provided by Hsieh et al. [22]. In Taiwan, health data before 2016 are coded using ICD-9-CM, whereas data from 2016 onward are coded using ICD-10 [22]. Prescription medications are coded using a system developed specifically for Taiwan; however, researchers have mapped these codes to ATC system codes [22].

Nine different datasets are linkable in the Western Australia Data Linkage System: birth, death and marriage registrations, electoral roll, hospital morbidity, emergency department presentations, mental healthcare contacts, midwife notifications and cancer registrations [26]. Additional datasets are listed as linkable on the linkage system website, including data from the health and wellbeing surveillance system, the monitoring of drug dependence system, the infectious disease database and registers for development anomalies. ICD codes are used to record diagnoses in health databases. For example, in the hospital mortality database, four ICD versions are used (ICD-8, 1970-1978; ICD-9, 1979-1987; ICD-9-CM, 1988-1999; ICD-10-AM, 1999 onward). Data dictionaries for the databases are available online: <https://www.datalinkage-wa.org.au/resources/dataset-information/>.

## Data quality

Misclassification of family relationships was identified in the published literature as a common concern in several sites. For example, parental linkages in Danish data are based on legal status and not on biological relationships [16]. In contrast, dependents identified in Taiwan's health insurance registry must be blood related, as determined by a birth certificate or DNA test [8]. While maternal relationships can be determined based on birth certificates and health records, assumptions may be made about paternal relationships. As the traditional family unit becomes less prevalent in all countries, the ability to identify paternal relationships may not be consistent over time [24]. Moreover, differentiating between half-siblings and full siblings and ensuring proper handling of sibling data in analyses (i.e., accounting for different degrees of relatedness) presents additional challenges, especially as sibling-comparison designs can help reduce confounding in observational research [37].

Across the Nordic registers, information about data linkage accuracy is noted to be limited for early years of data. In the Danish Civil Registration System, the accuracy of parental links is somewhat variable over time due to changes in methodology. Family linkages were originally determined based on residence (i.e., those living at the same residence were considered a family unit) [16, 27]. Linkages were then removed once an individual moved away, or a child in the home had a child of their own [16]. In 1978, parental linkages were established based on legal status and retained permanently;

previously-removed linkages have since been added and verified [16]. The estimated accuracy for linkages of children born in Denmark to legal mothers and legal fathers is 98.7% and 95.7%, respectively for data from 1960 to 1968. However, linkages from 1969 onward are considered to be 100% accurate [15, 16, 27]. Given that parental links are based on legal status [15], this may pose problems for researchers interested in studies about disease heritability. Biological parentage can be determined using Denmark's Medical Birth Register; however, linkages for biological mothers and biological fathers are only available from 1973 and 1991 onward, respectively [17]. It is estimated that complete sibling linkage is possible for children of women born after 1935 [15].

For Sweden, linkage to parents varies by year, with linkage being more variable for those born outside of Sweden. In 2005, it was estimated that 97% and 95% of individuals in the database who were born in Sweden could be linked to their mothers and fathers, respectively [21]. Information on both biological and adoptive parents is available. With respect to biological fathers, it is assumed that the husband of a married or recently-widowed mother is the father. Paternity can also be established by acknowledgment or a court order. For individuals with parents born in Sweden in 1915 or later, information on sibling relationships is considered to be of good quality [21]. The quality of the linkage is lower for those who were born outside of Sweden. Only 27% and 22% of individuals in the database who were born outside of Sweden have information on their mothers and fathers, respectively [21]. Sibling linkage is also less complete among individuals whose parents were born outside of Sweden [21]. Thus, both migrations out of and into the catchment area can influence the ability to ascertain family relationships.

A similar phenomenon exists for Norway, where familial links for those born outside of the country are missing at a higher rate than for those born in Norway [20]. For the latter, linkage is generally of high quality, with 100% of individuals born in Norway after 1952 having links to their mothers [20]. Supplementing the registry information with census data increased linkage in older generations by a noticeable margin [20]. Information about the quality of paternal linkage over time was not reported in the published literature that we examined.

The Canadian province of Manitoba reports accurate linkage to family members, with individuals ages 35 to 39 years having linkage to an average of 5 family members (including siblings, children, parents, and spouse) in 2014 [24]. However, similar to other data sites, the quality of familial linkage appears to vary over time. Ability to link to siblings decreases in older age groups compared to younger age groups; paternal linkages are more difficult to ascertain for younger generations [24]. We did not identify any studies that described the accuracy of familial linkages for Taiwan or Western Australia.

## Examples of multigenerational studies using routinely-collected data

Our review of published literature identified numerous multigenerational studies using the routinely-collected

population-based data from the six sites. These studies had two main themes: (a) family health histories or family relationships as predictors of a health condition, and (b) health trajectories after a critical health event or diagnosis within a family.

### Family health histories or relationships as predictors

In one study we identified, Rom et al. [38] used multiple Danish registries to test the association between autism spectrum disorder (ASD) in offspring and maternal and paternal rheumatoid arthritis. Linkages amongst family members were established using CPR-numbers. The Danish Medical Birth Register was used to create the initial cohort of children. Parental rheumatoid arthritis diagnosis was determined using the Danish National Patient Registry, and offspring ASD diagnosis was determined using both the Danish National Patient Registry and the Danish Psychiatric Central Research Registry. Close to two million children were included in the cohort [38]. Maternal rheumatoid arthritis diagnosis was associated with an increased risk of offspring ASD diagnosis; for paternal rheumatoid arthritis diagnosis, the relationship was not statistically significant although the risk was elevated (maternal: hazard ratio [HR] = 1.31; 95% confidence interval [CI] = 1.06–1.63; paternal: HR = 1.33; 95% CI = 0.97–1.82).

Cheng et al. [39] tested the association between a diagnosis of schizophrenia in any first degree relative and an individual's risk of a psychiatric disorder (including schizophrenia, bipolar disorder, major depressive disorder, ASD, and attention-deficit hyperactivity disorder [ADHD]). Administrative health data from the Taiwan National Health Insurance program were used to determine individual and relatives' health status, as well as demographic characteristics. The family relationship groups included in the study were parents, offspring, siblings, and twins. Blood relatives were identified using the National Health Insurance Data. Of the total Taiwanese population (more than 23 million), 227,967 individuals were identified as having a first-degree relative diagnosed with schizophrenia. Having a first-degree relative with schizophrenia was associated with an increased risk of a psychiatric disorder; the strength of this association was greatest for schizophrenia (relative risk [RR] 4.76; 95% CI = 4.65–4.88). There was also a dose-response relationship; the risk of a psychiatric disorder was greater when two first-degree relatives were diagnosed with schizophrenia, compared to when only one first-degree relative was diagnosed with schizophrenia.

Solberg et al. [40] used Norway's population-based registries to examine the risk of ADHD amongst offspring of parents with ADHD, bipolar disorder, schizophrenia spectrum disorder, or major depressive disorder. Data from the Medical Birth Registry were used to build the cohort, while the Norwegian Prescription Database and the Norwegian Patient Registry were used to determine parental and offspring health status, and the National Education Database and Statistics Norway were used to identify demographic characteristics. More than 2.4 million offspring were identified during the study period (1967–2011); 79,719 were diagnosed with ADHD. The RR of offspring ADHD was greatest when both parents

had ADHD (11.7; 95% CI = 11.0–12.5); maternal ADHD diagnosis appeared to have a stronger association with offspring ADHD diagnosis compared with paternal ADHD diagnosis (maternal RR: 8.4; 95% CI = 8.2–8.6; paternal RR: 6.2; 95% CI = 6.0–6.4). Parental diagnosis of bipolar disorder, schizophrenia spectrum disorder, and major depressive disorder were also associated with an increased risk of offspring ADHD diagnosis, although the strength of the association was lower.

### Health trajectories in families

Taipale et al. [41] used Swedish data to examine the association of antipsychotic drug exposure in offspring who had schizophrenia with subsequent parental health and work disability. Data from the National Patients Registers and Prescribed Drug Register were used to determine offspring and parental health status; data from Statistics Sweden, the Micro-data for Analyses of the Social Insurance Register, and Cause of Death Register were used to ascertain demographic factors, work sickness absences and disability, and cause of death, respectively. Familial links were determined using the Multi-generation Register. Overall, 10,883 offspring with schizophrenia and 18,215 parents were identified. Offspring exposure to first- and second-generation oral antipsychotic medications was associated with an increased risk of parental psychiatric healthcare use (RR range 1.10–1.29). Use of oral medications was associated with an increased risk of parental long term sickness absences from work; long-acting injection medications were associated with a decreased risk.

Bolton et al. [42] used Manitoba data to examine the mental, physical, and social outcomes of parents who had an offspring die in a motor vehicle collision. Data from the Vital Statistics Registry was used to identify deceased offspring. The Manitoban Health Insurance Registry was used to link parents and offspring, while hospital records and physician billing claims were used to identify health outcomes. Census data were used to identify socio-demographic characteristics. In total, 1,458 bereaved parents were identified and matched to the same number of non-bereaved parents. The risks of depression and anxiety disorders were higher in bereaved parents than in non-bereaved parents in the two years following the death of an offspring (depression prevalence ratio: 2.85; 95% CI = 2.44–3.33; anxiety prevalence ratio: 1.45 95% CI = 1.26–1.67). There was also an increased risk of cancer and hypertension. Compared to non-bereaved parents, bereaved parents also had higher risks of marital break-up and outpatient physician mental illness visits.

Morris et al. [43] characterized adolescent and young adult (ages 12–24 years) offspring in Western Australia between 1982 and 2015 whose had at least one parent with a cancer diagnosis during the offspring's lifetime. Data from the Cancer Registry were used to identify parents with an incident malignant cancer diagnosis as well as parents' demographic data, cancer information, and cancer-related death data. Offspring were identified using the Family Connections database. Information on offspring demographic factors and mortality came from birth registrations, the Midwives Notification System, and the mortality registry. In total, 57,708 offspring were linked to 34,600 parents who had an incident cancer diagnosis. The

authors found that 0.46% (95% CI = 0.43–0.49) of offspring in Western Australia were affected by an incident parental cancer diagnosis on an annual basis.

## Discussion

This review paper provides an overview of routinely-collected data available at six international sites that have the capability to conduct multigenerational health research by defining familial relationships and objectively-measured health histories from linkable administrative and healthcare electronic records. Our review found that multiple sources of health data, as well as data from other sectors such as education and employment, have been linked to information about family relationships and structures.

In the European sites included in our review, the population register is central to the identification of family relationships and structures. Previous research describes the establishment of population registries in multiple European countries their importance for describing the demographic characteristics of a country's population [44]. We did not identify any other European countries that had linked their population registries with routinely-collected healthcare data to conduct multigenerational studies. However, for Finland, we did find studies describing the use of routinely-collected data for birth cohort studies [45] and the Finnish Twin Cohort Study [46, 47]; these cohorts use population registration data and a variety of healthcare administrative databases, as well as follow-up with primary data collection.

In the non-European sites included in our review, health insurance registration files from universal healthcare systems were central to constructing family relationships and structures in both the Canadian province of Manitoba and Taiwan for their entire populations. While many other countries or jurisdictions have universal healthcare, not all have the capability to identify family members through registration numbers [48].

We note that several prospective cohort studies have been used to conduct multigenerational studies, including the Framingham Heart Study [49, 50], Norway's HUNT study [51, 52], and the Victorian Family Heart Study [53]. These studies have relied primarily on survey data to compile family health histories, which are prone to recall and self-report biases. However, cohort studies that rely on primary data collection can provide information that routinely-collected data cannot, including biological and health behavior information. As such, multigeneration studies using primary data sources have also made considerable contributions to understanding the impact of the family on health outcomes [3].

Our review of the literature did not identify any multigenerational health research involving two or more sites. Distributed networks that have recently been established to conduct pharmacoepidemiology studies [54] could be used as a model for multi-site multigenerational studies. In a distributed network, the same analysis is conducted at multiple sites; this approach does not involve data sharing across sites, which helps to ensure data privacy. Benefits of distributed networks include the ability to conduct studies on very large populations, which can benefit the investigation of rare health conditions

or events and the ability to assess the generalizability of single-site research findings, which can reduce bias.

## Conclusion

In conclusion, while the literature we reviewed demonstrated the extensive work that has been undertaken to establish familial linkages, additional research opportunities exist. To address the current challenges and limitations of multigenerational health research, future research could validate methods for identifying familial structures and explore the use of new data sources, such as electronic medical records, for constructing health histories of family members. Polubriaginof et al. [55] identified familial structures using emergency contact information in electronic medical records, but no formal validation of the method was undertaken. In a recently-funded study in Denmark, church records will be used to build a multigenerational registry with family relationships for those born as early as 1920 [56]. Methods involving the use of church records to identify familial relationships could be validated in other sites where linkage of these data with population registries and health insurance registries is possible. Such linkages may make it possible to increase the number of generations that can be identified in routinely-collected data and to ensure the accuracy of identification of family units. As well, linkage of new data sources will increase the range of research questions that can be asked and contribute to the generation of new knowledge about family health histories. Finally, establishing a research network would help to facilitate multi-site multigenerational studies and sharing of methodological expertise.

## Acknowledgments

We acknowledge funding provided by the Canadian Institutes of Health Research (PCS-168200). LML is supported by a Canada Research Chair in Methods for Electronic Health Data Quality. EWW is supported by a Canada Research Chair in Population Data Analytics and Data Curation. OPR has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 837180.

## Statement on conflicts of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

## Ethics statement

Ethics approval was not required for this work as it is a review of published literature and did not involve collecting data on human or animal subjects.

## References

1. Berg AO, Baird MA, Botkin JR, Driscoll DA, Fishman PA, Guarino PD, et al. National Institutes of Health



- state-of-the-science conference statement: family history and improving health. *Ann Intern Med.* 2009 Dec 15;151(12):872–7.
2. Ginsburg GS, Wu RR, Orlando LA. Family health history: underused for actionable risk assessment. *The Lancet.* 2019 Aug 17;394(10198):596–603. [https://doi.org/10.1016/S0140-6736\(19\)31275-9](https://doi.org/10.1016/S0140-6736(19)31275-9)
  3. Valdez R, Yoon PW, Qureshi N, Green RF, Khoury MJ. Family history in public health practice: a genomic tool for disease prevention and health promotion. *Annu Rev Public Health.* 2010;31:69–87. <https://doi.org/10.1146/annurev.publhealth.012809.103621>
  4. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, et al. Electronic health records: new opportunities for clinical research. *J Intern Med.* 2013;274(6):547–60. <https://doi.org/10.1111/joim.12119>
  5. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. *J Epidemiol Community Health.* 2014 Mar 1;68(3):283–7. <http://dx.doi.org/10.1136/jech-2013-202744>
  6. Lin H-T, Liu F-C, Lin S-F, Kuo C-F, Chen Y-Y, Yu H-P. Familial aggregation and heritability of nonmedullary thyroid cancer in an Asian population: a nationwide cohort study. *J Clin Endocrinol Metab.* 2020 Jul 1;105(7):e2521–30. <https://doi.org/10.1210/clinem/dgaa191>
  7. Wang C-L, Kuo C-F, Yeh Y-H, Hsieh M-Y, Kuo C-T, Chang S-H. Familial aggregation of myocardial infarction and coaggregation of myocardial infarction and autoimmune disease: a nationwide population-based cross-sectional study in Taiwan. *BMJ Open* 2019;9. <https://doi.org/10.1136/bmjopen-2018-023614>
  8. Wu HH, Kuo CF, Li IJ, Weng CH, Lee CC, Tu KH, et al. Family aggregation and heritability of ESRD in Taiwan: a population-based study. *Am J Kidney Dis* 2017;70:619–26. <https://doi.org/10.1053/j.ajkd.2017.05.007>
  9. Lindgren MP, Ji J, Smith JG, Sundquist J, Sundquist K, Zöller B. Mortality risks associated with sibling heart failure. *Int J Cardiol.* 2020 May 15;307:114–8. <https://doi.org/10.1016/j.ijcard.2019.10.022>
  10. Orholm M, Fonager K, Sørensen HT. Risk of ulcerative colitis and Crohn's disease among offspring of patients with chronic inflammatory bowel disease. *Am J Gastroenterol.* 1999 Nov;94(11):3236–8. <https://doi.org/10.1111/j.1572-0241.1999.01526.x>
  11. Hemminki K, Czene K. Attributable risks of familial cancer from the Family-Cancer Database. *Cancer Epidemiol Biomark Prev.* 2002 Dec;11(12):1638–44.
  12. Yang S, Leslie WD, Walld R, Roos LL, Morin SN, Majumdar SR, et al. Objectively-verified parental non-hip major osteoporotic fractures and offspring osteoporotic fracture risk: a population-based familial linkage study. *J Bone Miner Res* 2017;32:716–21. <https://doi.org/10.1002/jbmr.3035>
  13. Yang S, Lix LM, Yan L, Walld R, Roos LL, Goguen S, et al. Parental cardiorespiratory conditions and offspring fracture: a population-based familial linkage study. *Bone.* 2020 Oct;139:115557. <https://doi.org/10.1016/j.bone.2020.115557>
  14. Morgan VA, Di Prinzio P, Valuri G, Croft M, McNeil T, Jablensky A. Are familial liability for schizophrenia and obstetric complications independently associated with risk of psychotic illness, after adjusting for other environmental stressors in childhood? *Aust N Z J Psychiatry.* 2019;53(11):1105–15. <https://doi.org/10.1177/0004867419864427>
  15. Pedersen CB. The Danish civil registration system. *Scand J Public Health.* 2011 Jul;39(7 Suppl):22–5. <https://doi.org/10.1177/1403494810387965>
  16. Pedersen CB, Gøtzsche H, Møller JO, Mortensen PB. The Danish civil registration system: a cohort of eight million persons. *Dan Med Bull.* 2006 Nov;53(4):441–9.
  17. Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The Danish medical birth register. *Eur J Epidemiol* 2018;33:27–36. <https://doi.org/10.1007/s10654-018-0356-1>
  18. Petersen L, Sørensen TIA. The Danish adoption register. *Scand J Public Health* 2011;39:83–6. <https://doi.org/10.1177/1403494810394714>
  19. Skytthe A, Kyvik KO, Holm NV, Christensen K. The Danish twin registry. *Scand J Public Health* 2011;39:75–8. <https://doi.org/10.1177/1403494810387966>
  20. Næss Ø, Hoff DA. The Norwegian family based life course (NFLC) study: data structure and potential for public health research. *Int J Public Health* 2013;58:57–64. <https://doi.org/10.1007/s00038-012-0379-4>
  21. Ekblom A. The Swedish multi-generation register. *Methods Mol Biol Clifton NJ* 2011;675:215–20. [https://doi.org/10.1007/978-1-59745-423-0\\_10](https://doi.org/10.1007/978-1-59745-423-0_10)
  22. Hsieh C-Y, Su C-C, Shao S-C, Sung S-F, Lin S-J, Kao Yang Y-H, et al. Taiwan's national health insurance research database: past and future. *Clin Epidemiol* 2019;11:349–58. <https://doi.org/10.2147/CLEP.S196293>
  23. Lin C-M, Lee P-C, Teng S-W, Lu T-H, Mao I-F, Li C-Y. Validation of the Taiwan birth registry using obstetric records. *J Formos Med Assoc Taiwan Yi Zhi* 2004;103:297–301. <https://doi.org/10.29828/JFMA.200404.0008>
  24. Roos L, Walld R, Burchill C, Nickel N, Roos NP. Linkable administrative files: family information and existing data. *Longitud Life Course Stud* 2017;8:264–80. <https://doi.org/10.14301/llic.v8i3.406>

25. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository - Overview. Univ Manito 2020. [http://umanitoba.ca/faculties/health\\_sciences/medicine/units/chs/departamental\\_units/mchp/resources/repository/index.html](http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departamental_units/mchp/resources/repository/index.html) (accessed October 27, 2020).
26. Glasson EJ, de Klerk NH, Bass AJ, Rosman DL, Palmer LJ, Holman CDJ. Cohort profile: The Western Australian Family Connections Genealogical Project. *Int J Epidemiol*. 2008 Feb;37(1):30–5. <https://doi.org/10.1093/ije/dym136>
27. Erlangsen A, Fedyszyn I. Danish nationwide registers for public health and health-related research. *Scand J Public Health*. 2015 Jun;43(4):333–9. <https://doi.org/10.1177/1403494815575193>
28. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol*. 2014 Aug;29(8):541–9. <https://doi.org/10.1007/s10654-014-9930-3>
29. Andersen JS, Olivarius NDF, Krasnik A. The Danish national health service register. *Scand J Public Health* 2011;39:34–7. <https://doi.org/10.1177/1403494810394718>
30. Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, et al. The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol*. 2019 Jul 12;11:563–91. <https://doi.org/10.2147/CLEP.S179083>
31. Axelsson O. The Swedish medical birth register. *Acta Obstet Gynecol Scand*. 2003 Jun;82(6):491–2. <https://doi.org/10.1034/j.1600-0412.2003.00172.x>
32. Ludvigsson JF, Andersson E, Ekblom A, Feychting M, Kim J-L, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011 Jun 9;11:450. <https://doi.org/10.1186/1471-2458-11-450>
33. Wettermark B, Hammar N, Fored CM, MichaelFored C, Leimanis A, Otterblad Olausson P, et al. The new Swedish prescribed drug register-opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007;16:726–35. <https://doi.org/10.1002/pds.1294>
34. Barlow L, Westergren K, Holmberg L, Talbäck M. The completeness of the Swedish cancer register: a sample survey for year 1998. *Acta Oncol Stockh Swed* 2009;48:27–33. <https://doi.org/10.1080/02841860802247664>
35. Brooke HL, Talbäck M, Hörnblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish cause of death register. *Eur J Epidemiol*. 2017;32(9):765–73. <https://doi.org/10.1007/s10654-017-0316-1>
36. Emilsson L, Lindahl B, Köster M, Lambe M, Ludvigsson JF. Review of 103 Swedish healthcare quality registries. *J Intern Med*. 2015;277(1):94–136. <https://doi.org/10.1111/joim.12303>
37. D'Onofrio BM, Lahey BB, Turkheimer E, Lichtenstein P. Critical need for family-based, quasi-experimental designs in integrating genetic and social science research. *Am J Public Health*. 2013 Oct;103 Suppl 1:S46-55. <https://doi.org/10.2105/AJPH.2013.301252>
38. Rom AL, Wu CS, Olsen J, Jawaheer D, Hetland ML, Mørch LS. Parental rheumatoid arthritis and autism spectrum disorders in offspring: a Danish nationwide cohort study. *J Am Acad Child Adolesc Psychiatry* 2018;57:28-32.e1. <https://doi.org/10.1016/j.jaac.2017.10.002>
39. Cheng C-M, Chang W-H, Chen M-H, Tsai C-F, Su T-P, Li C-T, et al. Co-aggregation of major psychiatric disorders in individuals with first-degree relatives with schizophrenia: a nationwide population-based study. *Mol Psychiatry*. 2018;23(8):1756–63. <https://doi.org/10.1038/mp.2017.217>
40. Solberg BS, Hegvik T-A, Halmøy A, Skjaerven R, Engeland A, Haavik J, et al. Sex differences in parent-offspring recurrence of attention-deficit/hyperactivity disorder. *J Child Psychol Psychiatry* 2020. <https://doi.org/10.1111/jcpp.13368>
41. Taipale H, Rahman S, Tanskanen A, Mehtälä J, Hoti F, Jedenius E, et al. Health and work disability outcomes in parents of patients with schizophrenia associated with antipsychotic exposure by the offspring. *Sci Rep* 2020;10. <https://doi.org/10.1038/s41598-020-58078-4>
42. Bolton JM, Au W, Walld R, Chateau D, Martens PJ, Leslie WD, et al. Parental bereavement after the death of an offspring in a motor vehicle collision: a population-based study. *Am J Epidemiol* 2014;179:177–85. <https://doi.org/10.1093/aje/kwt247>
43. Morris JN, Zajac I, Turnbull D, Preen D, Patterson P, Martini A. A longitudinal investigation of Western Australian families impacted by parental cancer with adolescent and young adult offspring. *Aust N Z J Public Health*. 2019;43(3):261–6. <https://doi.org/10.1111/1753-6405.12885>
44. Poulain M, Herm A, Depledge R. Central population registers as a source of demographic statistics in Europe. *Population*. 2013 Oct 2;Vol. 68(2):183–212. <https://doi.org/10.3917/popu.1302.0215>
45. Isohanni M, Lauronen E, Moilanen K, Isohanni I, Kempainen L, Koponen H, et al. Predictors of schizophrenia: evidence from the Northern Finland 1966 Birth Cohort and other sources. *Br J Psychiatry Suppl*. 2005 Aug;48:s4–7. <https://doi.org/10.1192/bjpp.187.48.s4>
46. Kaprio J. The Finnish twin cohort study: an update. *Twin Res Hum Genet Off J Int Soc Twin Stud* 2013;16:157–62. <https://doi.org/10.1017/thg.2012.142>

47. Kaprio J, Koskenvuo M, Rose RJ. Population-based twin registries: illustrative applications in genetic epidemiology and behavioral genetics from the Finnish twin cohort study. *Acta Genet Medicae Gemellol Twin Res* 1990;39:427–39. <https://doi.org/10.1017/S000156600003652>
48. Hamm NC, Robitaille C, Ellison J, O'Donnell S, McRae L, Hutchings K, et al. Population coverage of the Canadian Chronic Disease Surveillance System: a survey of the contents of health insurance registries across Canada. *Health Promot Chronic Dis Prev Can Res Policy Pract*. 2021 Jul-Aug;40(7-8):230-241. <https://doi.org/10.24095/hpcdp.41.7/8.04>
49. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular diseases: a historical perspective. *Lancet*. 2014 Mar 15;383(9921):999–1008. [https://doi.org/10.1016/S0140-6736\(13\)61752-3](https://doi.org/10.1016/S0140-6736(13)61752-3)
50. Tsao CW, Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol*. 2015 Dec;44(6):1800–13. <https://doi.org/10.1093/ije/dyv337>
51. Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort profile: the HUNT Study, Norway. *Int J Epidemiol*. 2013 Aug;42(4):968–77. <https://doi.org/10.1093/ije/dys095>
52. Holmen TL, Bratberg G, Krokstad S, Langhammer A, Hveem K, Midthjell K, et al. Cohort profile of the Young-HUNT Study, Norway: a population-based study of adolescents. *Int J Epidemiol*. 2014 Apr;43(2):536–44. <https://doi.org/10.1093/ije/dys232>
53. Harrap SB, Stebbing M, Hopper JL, Hoang HN, Giles GG. Familial patterns of covariation for cardiovascular risk factors in adults: the Victorian Family Heart Study. *Am J Epidemiol*. 2000 Oct 15;152(8):704–15. <https://doi.org/10.1093/aje/152.8.704>
54. 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*. 2020;29(S1):3–7. <https://doi.org/10.1002/pds.4722>
55. Polubriaginof FCG, Vanguri R, Quinnes K, Belbin GM, Yahi A, Salmasian H, et al. Disease heritability inferred from familial relationships reported in medical records. *Cell* 2018;173:1692-1704.e11. <https://doi.org/10.1016/j.cell.2018.04.032>
56. Novo Nordisk Fonden. Artificial intelligence will transcribe the family relationships of Danes and strengthen research. Novo Nordisk Fonden. n.d. <https://novonordiskfonden.dk/en/news/kunstig-intelligens-skal-kortlaegge-danskernes-stam-trae-og-styrke-forskning/>

## Abbreviations

ICD:	International Classification of Diseases
ATC:	Anatomical Therapeutic Chemical
ASD:	Autism Spectrum Disorder
ADHD:	Attentional Deficit Hyperactivity Disorder
RR:	Risk Ratio
HR:	Hazard Ratio
CI:	Confidence Interval

