

Brameld

The Western Australian Family Connections Genealogical Project: detection of familial occurrences of single gene and chromosomal disorders.

Kate J. Brameld, PhD ^{1,2,3}

Danielle E. Dye, PhD ⁴

Susannah Maxwell, MPH ²

Joanna M. Brisbane, BSc GradDipGenetCounsell, MPH ²

Emma J. Glasson, PhD ^{3,5}

Jack Goldblatt, MBChB MD ^{6, 7}

Peter O'Leary, PhD ^{2, 8, 9}

¹ Neuropsychiatric Epidemiology Research Unit, School of Psychiatry and Clinical Neurosciences, The University of Western Australia.

² Centre for Population Health Research, Curtin University, Bentley, Western Australia.

³ School of Population Health, The University of Western Australia, Crawley, Western Australia.

⁴ School of Biomedical Sciences, Curtin Health Innovation Research Institute (CHIRI), Curtin University, Bentley, Western Australia

⁵ Telethon Institute for Child Health Research, Centre for Child Health Research, The University of Western Australia, Subiaco, Western Australia

⁶ Genetic Services of Western Australia, Subiaco, Western Australia.

⁷ School of Paediatrics and Child Health, The University of Western Australia, Crawley, Western Australia.

⁸ School of Pathology and Laboratory Medicine, The University of Western Australia, Crawley, Western Australia

⁹ School of Women's and Infants' Health, The University of Western Australia, Crawley, Western Australia.

Running Title: The WA Family Connections Genealogical Project

Address for Correspondence:

Professor Peter O'Leary, Centre for Population Health Research, Curtin University, Bentley, Western Australia, Australia. Email: peter.oleary@curtin.edu.au

Brameld

Abstract

Aim

To investigate using a Western Australian (WA) genealogical database for the identification of single gene and chromosome disorders among families.

Method

Hospital admissions for single gene and chromosome disorders recorded during 2000-2006 were identified from the WA Hospital Morbidity Data System. The proportion of these conditions occurring in family groups were then identified using genealogical links created through the WA Family Connections Genealogical Project.

Results

There were 216 family clusters among 11,303 people who were recorded as having a genetic or chromosomal disorder on their hospital admission record. The most common single gene conditions found to occur in multiple family members included blood clotting disorders such as Factor VIII deficiency and Von Willebrand's disease, followed by cystic fibrosis, myotonic dystrophies, neurofibromatosis, tuberous sclerosis and osteogenesis imperfecta.

Discussion

Single gene disorders most commonly occurring in multiple family members have been identified using the WA Family Connections Genealogical Project. These disorders reflect the most common single gene disorders requiring hospital admission, but which are not fatal prior to reproductive age and do not result in loss of fertility. They are also restricted to disorders with earlier onset, as the WA Family Connections Genealogical Project currently covers 2-3 of the most recent

Brameld

generations. This study demonstrates the utility of record linkage genealogies to identify kindreds with genetic disorders, offering a rich resource of information for focussed genetic epidemiological research.

Introduction

The Western Australian (WA) Family Connections Genealogical Project is a system of genealogical links designed to be used in conjunction with health data to support family-based epidemiological or genetic health research (Glasson *et al.*, 2008). The WA Family Connections Genealogical Project is supplementary to the WA Data Linkage System (WADLS) which is a system of regularly linked population-based data sets, including birth and death registrations, midwife records, hospital morbidity data, mental health information and cancer records (Holman *et al.*, 2008).

Genealogy for the population is constructed using information from birth registrations that identify parent-child relationships and unions between parents. Through the WADLS, health data can be extracted for the individuals contained in the genealogical matrix so that patterns of disease inheritance can be studied.

We have previously used the WADLS to study the occurrence of single gene and chromosome disorders at a population level in WA (Dye *et al.*, 2011a,b) by extracting data on individuals recorded as having these conditions between 2000-2006. The purpose of the current study was to use the WA Family Connections Genealogical Project to assess the possibility of ascertaining multiple family members with the same disorder. Such information could potentially be used to detect family members at risk of inherited disease for genetic counselling purposes, as a resource for health professionals to conduct genetic epidemiological research, to facilitate collecting family health history information and to inform population genetic policy.

A recent review of published literature on the use of genealogy databases for risk assessment in genetic health services (Stefansdottir *et al.*, 2012) found descriptions

Brameld

of potential uses, but little evidence of application in genetic health services. This study reports an application of the WA Family Connections Genealogical Project by linking health administrative data collections to validate the identification of multiple affected family members with these disorders.

Methods

The WA Hospital Morbidity Data System is a statutory collection of all inpatient admissions to public, private and freestanding day hospitals in WA since 1970. For this study, 296 diagnosis codes specific to single gene and chromosomal disorders were identified from the International Statistical Classification of Diseases, Tenth Revision, Australian Modification (ICD-10-AM) following a review of all potential codes (Dye *et al.*, 2011b). The WA Hospital Morbidity Data System was searched for any individual who had at least one of these ICD10 codes recorded in the diagnosis fields between 2000 and 2006. During 2000-2006, there were 4,575,700 hospital admissions, ranging from 319,228 patients in 2000 to 365,674 in 2006.

At the time of the study, the Family Connections Project included familial data on the 1 million births that had occurred in WA since 1974. (Glasson *et al.*, 2008). Thus genealogical links were available if one or more family member was born or had a child since 1974. At the time of data extraction, up to 4-generation pedigrees were available within the genealogical matrix for the population. The entire genealogy exists as a population matrix and families are linked to other families via sibling marriages and linear generations such that a complete pedigree for an individual could range from a 2-person family unit to an extended genealogy of several thousand people or more.

The protocols of the WA Family Connections Project are designed to maximise the protection of individual privacy while providing information for approved research projects (Kelman *et al.*, 2002). The genealogical relationships are represented by

Brameld

unique identifiers that represent links between other records, and no actual data are stored with the genealogical indices. Research extractions have encrypted identifiers to make them meaningless outside the research context.

Familial clustering of the conditions was ascertained from linkage of the hospital data to the WA Family Connections Genealogical Project. Relatives were compared using their three-digit ICD-10-AM codes and were considered to constitute a familial case if the three-digit code ICD-10-AM for any of their admissions was identical to that of their relative. Both the index case and their relatives were counted as familial cases. Index cases were assigned arbitrarily as there was no information to indicate which case was identified first. For the purposes of this study, a relative was a person who had any degree of relatedness to each of the probands. This included a first, second or third degree relative.

The proportion of familial cases for each disease was calculated by dividing the number of familial cases by the total number of persons admitted to hospital for each unique three-digit ICD code.

The project received ethics approval from The University of Western Australia Human Research Ethics Committee and the Department of Health (Western Australia) Human Research Ethics Committee.

Results

During 2000-2006, 11,303 people admitted to hospital were recorded with a single gene or chromosome disorder, within which 216 familial clusters were identified where relatives shared the same ICD10-AM diagnosis code. The clusters comprised 463 relatives; of these 216 probands. 218 first-degree relatives (parent, child or sibling), 27 second degree relatives (grandparent, grandchild, aunt/uncle, niece/nephew or half-sibling) and 2 third degree (cousins). The mean number of people per cluster was 2.14 (range of 2 to 4). The familial clusters accounted for 4.1% of all patients with single gene and chromosomal disorders admitted to hospital.

The International Classification of Diseases, Tenth Revision, Australian Modification (ICD-10-AM; ICD-10) (2004) was used to identify 296 diagnosis codes specific to single gene and chromosomal disorders. A limitation of the ICD classification system is that many ICD codes include multiple gene abnormalities. Of the 296 diseases studied, 48 (16%) occurred in multiple family members. The disorders with the highest numbers of patients with affected family members included blood clotting disorders such as Factor VIII deficiency and Von Willebrand's disease, followed by cystic fibrosis, hereditary spherocytosis, osteogenesis imperfecta and myotonic dystrophies. The proportion of familial cases for each disease is listed in Table 1.

Discussion

In this study, we linked data from the WA Hospital Morbidity Data System with the WA Family Connections Genealogical Project to identify related patients with single gene and chromosomal disorders who were admitted to hospital. Overall, we found

Brameld

that 4.1% of patients had family members affected by the same disorder and this involved 16% of the genetic and chromosomal disorders included in the study. Of the 46 disorders identified with at least two affected relatives, the majority were autosomal recessive conditions (33%), followed by autosomal dominant conditions (29%), X-linked (9%), chromosomal (2%) and 27% showed variable patterns of inheritance.

The ability of the WA Family Connections Genealogical Project to identify genetic disorders in related patients may be affected by the mode of inheritance, prevalence, age of onset and clinical severity of the disorders. For example, autosomal dominant disorders have a 50% chance of expression in offspring if only one parent is heterozygous for the condition whereas autosomal recessive conditions only have a 25% chance of expression if both parents are heterozygous for the condition. Single gene conditions may also arise spontaneously and the extent to which they are inherited or arise from a de novo mutation varies according to the condition. For example, 90% of people with polycystic kidney disease have an affected parent whereas 80% of cases of achondroplasia result from a new mutation (National Genetics Education and Development Centre). Chromosomal disorders may also arise spontaneously or be inherited from a parent with a balanced chromosomal rearrangement, although these disorders are frequently associated with reduced fertility and therefore most cases are not inherited.

The conditions most likely to be identified in families using the WA Family Connections Genealogical Project are those that are prevalent in the population, are severe enough to warrant periodic hospital admission but have a limited impact on

Brameld

fertility. In this study, genetic disorders identified as occurring in families reflect commonly occurring earlier onset conditions (Dye *et al.*, 2011a,b). For example, osteogenesis imperfecta (OI) is an early-onset autosomal dominant disorder with a prevalence of approximately 1/20 000 in Caucasian populations. Although there are some rare severe forms, most patients with type 1 OI survive to adulthood, and reproduce, with periodic hospitalization to treat fractures. We identified 101 patients in WA between 2000 and 2006, of whom 22 had an affected relative. In contrast, Huntington's disease (HD), an autosomal dominant disorder with a late onset phenotype and prevalence of 1/10 000 was identified in 77 patients in our WA Family Connections Genealogical Project, yet only three of these had an affected relative (4%) documented in the database. This is a significant underestimation of the familial nature of this disorder, and reflects a limitation of the WA Family Connections Genealogical Project to identify late onset conditions at the time of data extraction and to include older family members. Currently, genealogical data from earlier historical records (1945-1973) are being incorporated into the system and this will allow a greater number of familial links for all probands, include individuals with late-onset disease, and enhance the identification of second and third degree relatives through the detection of common ancestors.

The power of the WA Family Connections Genealogical Project is also limited by the make-up of the resident population. For example, at 30 June 2006, 29.9% of WA's total population of 2.0 million was born overseas (Australian Bureau of Statistics, 1998). In these cases, complete family pedigrees cannot be identified from the data. However, data on country of birth are available on the WA Hospital Morbidity Data

Brameld

System and could be used to estimate the proportion of cases where further data are unlikely to be available.

The conditions identified in this study are those that require hospitalization and can be accurately identified using ICD-10-AM codes. The accuracy of the ICD-10-AM coding on the WA Hospital Morbidity Data System has been demonstrated on a number of occasions (Brameld *et al.*, 1999; Brameld *et al.*, 2003; Teng *et al.*, 2008; Morgan *et al.*, 2011) and while some errors in coding may be present we are unable to provide specific statistics on accuracy for the chromosomal codes used for our study. Conditions for which there is no specific ICD-10-AM code or those not generally requiring hospitalization have been excluded. The ICD-10-AM coding system itself has some limitations. Previously, we reported that the ICD-10-AM code for muscular dystrophy (G710) includes early and late onset disorders, with variable modes of inheritance (Dye *et al.*, 2011b). In addition, the data linkage methods for this study only detected cases where both family members were hospitalized during the study period and the condition of interest was documented on all admission records. Therefore, conditions requiring hospitalization only at the time of diagnosis, and those requiring infrequent or no hospitalization, are unlikely to be identified, whereas those requiring repeated hospitalization (for example, transfusions for blood clotting disorders, therapeutic venesection for haemochromatosis) are much more likely to be identified.

The Family Connections system links individuals together using information recorded on original birth registrations. It was not possible to validate family links because the pedigrees identified in this data linkage study are not recorded on any other

Brameld

database. However, children born in WA since 1980 and their mothers are automatically linked together within the WADLS through a mandatory Midwives notification system of all WA births and fathers (and mothers prior to 1980) have been linked using the Family Connections Project. The WADLS linkage strategies are tuned to balance false positives and false negatives, both of which are estimated to be very small. The accuracy of linkage from a subset of the WA Hospital Morbidity Data System records was once estimated at 99.9% (Holman et al, 1999).

Benefits of the WA Family Connections Project

The WA Family Connections Genealogical Project facilitates the construction of pedigrees for specific individuals based on genealogy ascertained from birth registrations and the linkage of health information to these family trees (Glasson *et al.*, 2008). We anticipate that this will have both clinical and research applications in the future. For example, for the purposes of genetic counselling, the system could provide genetic health professionals with information that complements the family history provided by the patient. Many people are unable to recall all of their family health history, particularly for second degree and more distant relatives (Molster *et al.*, 2011). This is particularly relevant with genetic disorders, where disclosure of genetic information within families can be challenging due to family dynamics and the ability of family members to both explain and understand genetic concepts (Forrest *et al.*, 2007; Maxwell *et al.*, 2009). By having access to more complete familial health information, genetic counsellors will be better able to provide informed genetic counselling to patients. In this instance, we recognise that privacy aspects of probands and family members will need to be considered before allowing these data to be available for counselling purposes.

In addition, the widespread uptake of prenatal fetal anomaly screening and diagnostic procedures in recent years has generated increasingly sophisticated information to predict recurrence risk and pregnancy outcome (Brameld *et al.*, 2008). Evaluations of population screening programs that determine uptake, access and social attitudes (Maxwell *et al.*, 2011a,b) will be improved by incorporating genealogy data. Where common autosomal and x-linked recessive conditions are detected, parents and first degree relatives can access genetic testing to assist future reproductive decisions (Maxwell *et al.*, 2011b). For the majority of inherited conditions that manifest later in life, suitable screening tests are not yet available. However the situation will change with the advent of next generation sequencing when it is applied to personalised medicine in a population health setting (Burke *et al.*, 2011; Goldstein, 2011). As technological advances enable a greater range of predictive testing based on familial conditions, the WA Family Connections Genealogical Project can support genetic research as we have described, to map family disease pedigrees.

The information initially derived from administrative health databases and enhanced by links to the WA Family Connections Genealogical Project provides additional insights into the impact of common familial conditions on health services. This study confirms the utility of the WA Family Connections Genealogical Project to identify conditions occurring in families from data covering a relatively short timeframe. This could be enhanced in the future with a longer study timeframe, an extended genealogical matrix and the addition of other health data. Data from such studies are vital to inform genetic health services and policy development.

Brameld

Author Disclosure Statement

No competing financial interests exist.

References

- Glasson EJ, de Klerk NH, Bass AJ, *et al.* (2008) Cohort profile: the Western Australian Family Connections Genealogical Project. *Int J Epidemiol* 37:30-35.
- Holman CD, Bass AJ, Rosman DL, *et al.* (2008) A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian Health Review* 32:766-777.
- Dye D, Brameld KJ, Maxwell S, *et al.* (2011a) The impact of single gene and chromosomal disorders on hospital admissions of children and adolescents: a population-based study. *Public Health Genomics* 14:153-161.
- Dye D, Brameld KJ, Maxwell S, *et al.* (2011b) The impact of single gene and chromosomal disorders on hospital admissions in an adult population. *Journal of Community Genetics* 2:81-90.
- Holman CDJ, Bass AJ, Rouse IR, Hobbs MST. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Australia and New Zealand Journal of Public Health*. 1999; 23:453-459.
- Stefansdottir V, Johannsson OT, Skirton H, *et al.* (2012) The use of genealogy databases for risk assessment in genetic health service: a systematic review. *J Community Genet* 4:1-7.
- ICD-10-AM. The University of Sydney (2004). The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification. 4th ed. Sydney, Australia: National Centre of Classification in Health.
- Kelman CW, Bass AJ, Holman CDJ (2002). Research use of linked health data - a 'best practice' protocol. *Aust NZ J Public Health* 26:251-5.

Brameld

National Genetics Education and Development Centre: Genetics and Genomics for healthcare: Genetic Conditions; National Health Service

<http://www.geneticseducation.nhs.uk/genetic-conditions-54> accessed August 1 2013.

Australian Bureau of Statistics. National nutrition survey: Nutrient intakes and physical measurements. 1998. Available from:

<http://www.Abs.Gov.Au/ausstats/abs@.Nsf/latestproducts/4805.0main%20features11995?Opendocument&tabname=summary&prodno=4805.0&issue=1995&num=&view=> accessed August 1 2013.

Brameld K, Thomas MAB, Holman CDJ, *et al.* (1999) Validation of linked administrative data on end-stage renal failure: application of record linkage to a 'clinical base population'. *Aust NZ J Public Health* 23:465-467.

Brameld K, Holman CD, Lawrence DM, *et al.* (2003) Improved methods for estimating incidence from linked hospital morbidity data. *Int J Epidemiol* 32:617-624.

Teng, T, Finn J, Hung, *et al.* (2008) A validation study: how effective is the Hospital Morbidity Data as a surveillance tool. *Aust NZ J Public Health* 32:405-407.

Morgan V, Valuri GM, Croft ML, *et al.* (2011) Cohort Profile: Pathways of risk from conception to disease: the Western Australian schizophrenia high-risk e-cohort. *Int J Epidemiol* 40:1477-1485.

Molster C, Kyne G, O'Leary P. (2011) Motivating intentions to adopt risk-reducing behaviours for chronic diseases: impact of a public health tool for collecting family health histories. *Health Prom J Aust* 22:57-62.

Brameld

Forrest LE, Delatycki MB, Skene L, *et al.* (2007) Communicating genetic information in families - a review of guidelines and position papers. *Eur J Hum Genet* 15:612-618.

Maxwell S, Molster C, Poke SJ, *et al.* (2009) Communicating genetic information within families. *Genet Test Mol Biomarkers* 13:301-306.

Brameld K, Dickinson J, Bower C *et al.* (2008) First trimester predictors of adverse pregnancy outcomes. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2008;48:529-35.

Maxwell S, Brameld K, Bower C *et al.* (2011a) Socio-demographic disparities in the uptake of prenatal screening in Western Australia. *Aust NZ J Obstet Gynaecol* 5:9-16.

Maxwell S, Kyne G, Molster C *et al.* (2011b) Perceptions of population cystic fibrosis prenatal and preconception carrier screening among individuals with cystic fibrosis and their family members. *Gene Test Mol Biomarkers* 15:1-6.

Burke W, Tarini B, Press NA *et al.* (2011) Genetic screening. *Epidemiologic Reviews* 33:148-64.

Goldstein DB (2011) Growth of genome screening needs debate. *Nature* 476:27-28.