Longitudinal, Whole-population Data Examining Pathways of Risk from Conception to Disease: The Western Australian Schizophrenia High-risk e-Cohort

Vera A. Morgan, Giulietta Valuri, Maxine Croft, Sonal Shah, Patsy Di Prinzio, Jenny Griffith, Thomas McNeil and Assen V. Jablensky

1 Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences, Crawley WA 6009, Australia
2 Skane University Hospital, Sweden
3 Centre for Research in Clinical Neurosciences, The University of Western Australia School of Psychiatry & Clinical Neurosciences, Crawley WA 6009, Australia

Keywords: schizophrenia; psychotic disorders; mental illness; pregnancy; obstetric complications; birth defects; intellectual disability; educational achievement; criminality; parental mental illness; familial (genetic) liability; environmental risk factors; childhood abuse; population cohort; Western Australia

**(1) Overview**

**Spatial coverage**

The whole of the State of Western Australia, Australia.

**Temporal coverage**

01/01/1980-31/12/2001 defines the period of birth for the birth cohort. Mental health data for the birth cohort have been collected from 1980 to 2010.

**Species**

Homo sapiens

---

**(2) Methods**

**Steps**

The e-Cohort has been constructed by record linkage across Western Australian administrative health and social services registers. The high risk e-Cohort comprises all children born in Western Australia between 1980 and 2001 to mothers with a severe mental illness (schizophrenia, bipolar disorder, unipolar major depression, and other psychoses). Cases mothers are those who have had an inpatient or outpatient/community mental health contact recorded on the Statewide psychiatric case register between 1966 and 2001. The comparison cohort consists of all children born in Western Australia over the same period whose mothers have not had any record of contact with mental health services. Fathers have been identified and sibships established through record linkage to birth registration and obstetric records.

Data on the psychiatric status of mothers, fathers and children was extracted from the psychiatric case register which uses International Classification of Diseases (ICD) codes to record diagnosis. The International Classification of Diseases, 9th Revision codes [2] that were used to identify specific disorders were: schizophrenia (295.0–295.9), bipolar disorder (296.0, 296.2–5) unipolar major depression (296.1, 296.6, 296.8, 296.9), other psychoses (297.0–297.9, 298.0–298.9). The concurrent validity of register diagnoses of schizophrenia and affective psychoses have been evaluated against an independent sample assessed using a semi-structured diagnostic interview [3] with sensitivity of 0.92 and specificity of 0.88 for schizophrenia and 0.80 and 0.90 respectively for affective psychoses [4].

Detailed data on the pregnancy and perinatal period have been extracted from the Midwives Notification System which includes mandatory, prospectively-collected data on all infants born in Western Australia at 20 weeks gestation or greater or weighing at least 400 grams, including home births, and not restricted to live births [5]. Core data include prospectively recorded data on maternal age, marital status and race, baby’s gestational age and weight, pregnancy complications (e.g., preeclampsia, placenta praevia, abruptio, substance abuse), labour and delivery complications (e.g., cephalopelvic disproportion, prolapsed cord, postpartum haemorrhage) and early neonatal complications (e.g., foetal distress, 5-minute Apgar score). In addition, the database uses a local (Western Australian) version of a well-validated scale for summatting obstetric complications, the McNeil-Sjöström Scale for Obstetric Complications [4,6,7]. This scale was selected because it is underpinned by strong biological considerations, including timing of the complication and likely impact on central nervous system development. Local modifications included the addition of obstetric complication scores for maternal and infant conditions that were not previously recorded in the original McNeil Sjöström Scale Scoring system but were identified in this large population of Southern Hemisphere mothers.

Additional information on all children has been compiled from multiple registers. The registers cover: all hospital admissions including physical and psychiatric morbidity, outpatient/community mental health psychiatric contacts, intellectual disability, birth defects, cerebral palsy, cancer and mortality. Other data sources have been identified in the study protocol and are being added as they become available. Most recently, linkages have been included on school assessment data including literacy and numeracy proficiency, and data from State departments on contacts with child protection services and corrective services. Record linkage was undertaken by the Western Australian Department of Health Data Linkage Branch in a secure setting, using probabilistic data matching protocols [8]. A full description of the database and its elements has been published [1]. As noted above, numbers in this article differ from previously published numbers due to new linkages and updates to the database.

**Sampling strategy**

There was no sampling. This is a whole population birth cohort, born 1980-2001, and grouped into case offspring of mothers with a psychotic disorder and comparison offspring of mothers with no known psychiatric illness.

**Quality control**

Probabilistic record linkage software was used to link records across Western Australian health registers [8]. The accuracy of record linkage using probabilistic data matching is high with the percentage of true matches estimated to be 97-99 percent depending on the number of matching variables used [9]. Validation of the registers involved in the linkages has been published [1].

**Constraints**

One limitation of the record linkage methodology is the inability to identify people receiving psychiatric services from a general medical practitioner or a psychiatrist/psychologist in private practice, as it is not possible to link to these services. However, almost all people with a psychotic disorder make contact with inpatient or outpatient/community mental health services in the course of their illness and will be captured on the service records available for linkage.

The hospital morbidity register is used to capture physical morbidity. Although the hospital morbidity register is comprehensive, all cases reported here are hospitalised cases and therefore represent the more serious end of the spectrum.
Privacy
Records use encrypted identifiers to protect confidentiality. Due to the sensitive nature of the data, access is restricted to people named and approved by the relevant human research ethics committees.

Ethics
Data collection was based on the use of routinely collected information on administrative records. All records were de-identified before they were provided to the research team. Consent was not sought for access to these de-identified participants’ records. Exemption was given by the relevant ethics committees on the grounds that the development of the database was in the public interest, and that the public interest in our project substantially outweighed the public interest in the protection of privacy. Consideration was given to the fact that the advantage of population-based epidemiological studies using registers is their ability to study whole populations in an efficient and cost-effective manner. Seeking individual consent in our project would have been prohibitive in terms of cost and time and, because of the age of many records, would have resulted in large amounts of missing data. These factors would have negated the benefits of using register data.

The project was approved by: The University of Western Australia Human Research Ethics Committee; Western Australian Department of Health Human Research Ethics Committee.

(3) Dataset description

Object name
The Western Australian e-Cohort at High Risk for Schizophrenia and Other Psychotic Disorders.

Data type
Processed data are in multiple tables within an Ingres relational database on a dedicated server running the Linux operating system with the unique person identifier as the key for retrieval (see Figure 1). A Structured Query Language is used to access data for analysis within any of a number of statistical packages such as R, SPSS, SAS and Stata. Due to the number of records in the Ingres database - over 8,000,000 most analyses are performed directly against the database using SQL to limit the number of records accessed. Only if a statistical package cannot access Ingres, are subsets of data extracted from the database. Data extracts are based on the specification of dependent and independent variables required for a specific series of analyses.


Ontologies
N/A

Figure 1: Relational data model for children only for selected base tables. * Indicates the ID linking tables.

Format names and versions
- Program files have .sql and osq extensions
- Ingres II for Linux version 10.1.0
- Redhat Enterprise Linux Server release 5.4

Creation dates
The creation of the database began in 2003 and is still ongoing in 2014 as new linkages become available.

Dataset creators
Also listed as authors:
- Vera A. Morgan, (overall data coordination); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Giulietta Valuri, (database manager; psychiatric, criminological data); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Maxine Croft, (maternal morbidity, obstetric data); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Sonal Shah, (mortality, childhood protection data); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Patsy Di Prinzio, (early neurodevelopmental, socio-structural data); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Jenny Griffith, (casenotes review); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Thomas McNeil (McNeil-Sjöström Scale for Obstetric Complications and diagnostic review); Skane University Hospital, Sweden
- Assen Jablensky (conceptualisation, diagnostic review and validation); The University of Western Australia School of Psychiatry & Clinical Neurosciences
Not listed as authors:

- Deidra Young* (educational achievement and area-level socioeconomic data); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Nadia Badawi (Western Australia modification of the McNeil-Sjöström Scale for Obstetric Complications); Grace Centre for Newborn Care, The Children’s Hospital Westmead, University of Sydney / Notre Dame University / Cerebral Palsy Alliance
- John Keogh (Western Australia modification of the McNeil-Sjöström Scale for Obstetric Complications); University of Sydney / Sydney Adventist Hospital
- John Dean* (medications data review); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- Peterina Duggan*, Cate Simpson*, Carmel Ryan*, Anna Waterreus (case notes data extraction); Neuropsychiatric Epidemiology Research Unit, The University of Western Australia School of Psychiatry & Clinical Neurosciences
- The Western Australian Data Linkage Branch (record linkage); Western Australian Department of Health

* affiliation is at the time of contribution, not current affiliation

Language
English

Programming language
SQL

Licence
N/A

Accessibility criteria
Current access to the data is restricted to researchers named and approved by the relevant human research ethics committee, due to the sensitive nature of the data collection. It is anticipated that, once key constructs are fully operationalized in the next 12-18 months, other researchers will be welcome to work with senior investigators on the study, subject to ethics approval and behind approved firewalls. Please refer to details published in Morgan et al. [1]. Further information is available from Prof Vera Morgan: vera.morgan@uwa.edu.au.

Repository location
N/A

Publication date
N/A

(4) Reuse potential
The database can be used to:

- Test hypotheses about environmental risks for psychotic illness and their interaction with familial liability
- Examine specificity across a spectrum of psychiatric outcomes
- Identify specific subgroups for case-control analysis, including analysis of sibships within the data

Using emergent findings, it will be possible to:

- Generate hypotheses for follow-up studies in clinical settings, where more detailed information can be obtained directly from the participant without relying solely on administrative records

Publication of results based on a subsample of children (N=6303) within the current database [4,10] has already seen translation into practice. Our paper in the American Journal of Psychiatry on reproductive pathology in women with severe mental illness found a significantly increased risk of adverse obstetric outcomes in women [4]. Many of these were related to psychosocial risk factors and behavioural concomitants of psychosis. In order to disseminate our findings widely among clinicians, we published them in Directions in Psychiatry (warded “Exemplary Status” by the Accreditation Council for Continuing Medical Education) [11]. We also worked with mental health services to develop a care coordination package for these women in pregnancy, resulting in the implementation of a State care-coordination program for pregnant women with severe mental illness, the first of its kind in Australia [12].

References
health research database for research and policy evaluation in Western Australia. *Seminars in Neonatology*, 2(3):195-201. DOI: http://dx.doi.org/10.1016/S1084-2756(97)80015-2


