Reappraisal

From inventory to benchmark: quality of psychiatric case registers in research

Vera A. Morgan and Assen V. Jablensky

Summary

In recent years, there has been a marked increase in the use of psychiatric case registers for research purposes. Registers are a valuable data asset but there are no standard guidelines for evaluating their use in research. It is becoming increasingly important for researchers who use register data, and journal editors who publish their work, to set benchmarks to assess the quality of a register and its research application. Several criteria that could form the basis of such an evaluative framework are discussed. The discussion is illustrated using a Western Australian e-cohort of half a million children for whom we have assembled comprehensive data cross-linked across a number of administrative registers.

Declaration of interest
None.

Two articles on the use of psychiatric case registers in research were published recently: an editorial in the September issue of the *British Journal of Psychiatry*[1] closely followed by a paper on the Nordic registers in *Acta Psychiatrica Scandinavica*.[2] Their publication is timely and welcome. The longitudinal nature of registers, their size and coverage of defined populations make them an important research asset. Recent years have seen an increase in the use of the psychiatric case register for research purposes, including linkage across diverse health and other population databases such as criminological databases. It is over 20 years since the publication of Ten Horn's comprehensive inventory of the psychiatric case register and its use in research.[3] An update on the current status of psychiatric registers worldwide is well overdue. It is also an opportune time to move from the inventory to the benchmark, and discuss guidelines for the appropriate use of registers in research, including a re-examination of their limitations and a dialogue about suitable methods for dealing with these challenges.

A framework for evaluating registers

Mortensen[4] provides an excellent framework for evaluating the use of registers for epidemiological research by utilising four objectives of epidemiological research against which a study's methodology can be assessed:

(a) maximisation of the precision of the disease estimate

(b) minimisation of selection bias

(c) minimisation of information bias

(d) control of confounders.

The aim of this paper is to expand on these objectives. Where appropriate, we illustrate our points using a Western Australian e-cohort of half a million children for whom we have assembled comprehensive data cross-linked across a number of administrative registers including the Western Australian Mental Health Information System, which is the psychiatric case register for a State of 2.5 million square kilometres and a population of 2.2 million. This provides a concrete demonstration of both the advantages and limitations of the use of psychiatric case registers and record linkage in psychiatric research.

The first objective, precision, refers to the reliability of a parameter estimate and is reflected in the width of its confidence interval. Sample size is an important determinant of precision. The larger the sample, the narrower the confidence interval, the greater the precision and the less likelihood of a type II error (false negative). An advantage of register-based studies is that they are a cost-effective means of collecting data on large samples, especially for the study of rare outcomes. However, this advantage is offset by the time needed to clean and validate data, and build meaningful constructs using a set of variables collected primarily for administrative purposes.

Mortensen's second objective relates to selection bias affecting a register-based study when the sample it covers is not representative of the total population of interest. Different forms of selection bias include: geographical bias (e.g. a sample may only include persons living in the metropolitan area); treatment bias (e.g. a sample drawn solely from an in-patient hospital register may include more severe cases only); social bias (e.g. a sample may cover public facilities only and exclude private facilities); and selective survival bias (with differential survival skewing the sample profile). For each research undertaking, the coverage of the psychiatric case register must be examined and evaluated against the study's aim, to determine its adequacy for the purpose. For example, the Western Australian Mental Health Information System is a whole-of-population register of long standing (since 1966) with no periods of discontinuity. It covers all public and private in-patient admissions, as well as public out-patient and ambulatory care contacts with mental health services across the State.[5] As such, it is an excellent resource for the study of low-prevalence disorders such as the psychoses where the vast majority of affected individuals have had contact with either in-patient or out-patient services. However, as it misses contacts with general practitioners and psychologists/psychiatrists working in private practice, its use in the study of high-prevalence disorders more commonly seen outside of in-patient and out-patient services is limited. Such use calls for justification and explicit statement of the limitations.

The third objective, minimising information bias, is concerned with reducing bias associated with: loss to follow-up due to death or movement out of the area of the register; lack of information prior to first entry on the register due to migration into the region covered by the register (in-migration); failed linkages across registers; and periods of discontinuity of the collection of data on the register. In Western Australia, linkage failures due to name changes and spelling variants are reduced through the use of aliases and phonetic spelling in probabilistic linkage protocols. Cross-linkage with a range of other state-wide registers, updated regularly, further aids the identification of identity errors. Loss to follow-up due to death is reduced by linkage to mortality data

which include date and cause of death. Loss to follow-up through migration out of the region covered by the register (out-migration) is minimised owing to the State's geographical isolation and its economic advantages that give it one of the highest rates of Australian interstate in-migration in the past decade, and one of the lowest levels of out-migration. Where in-migration is an issue, we have limited case-finding to people born in Western Australia, a variable recorded on the psychiatric case register (e.g. Morgan *et al*[6]). One advantage of registers is that, as a result of prospective recording of information, they avoid recall bias, a form of information bias that may affect the retrospective collection of data.

Mortensen's final objective, control of confounders, is critical. A well-designed prospective or retrospective study collects data on variables that previous research or theory suggest may confound or modify the relationship between the exposure and outcome of interest. In register-based studies, the researcher is restricted to variables available on the register. Since many of these are collected for administrative rather than research purposes, important confounders and modifiers may be missing. There are a number of ways to deal with their absence. Linking several different registers covering the same population increases the chance that variables not available on one database may be available on another. For example, we linked women with psychosis on the psychiatric case register to their obstetric data on the midwives register to examine familial and environmental (obstetric) risk factors for adverse neuropsychiatric outcomes in their children.[5,7] We also included hospital morbidity, mortality, birth defects and intellectual disability registers, among others, to increase the data sources available for just under half a million offspring (Fig. 1). For some missing data, there may be

Psychiatric in- and out-patient case register
25 887 women with psychosis (ICD–9 295–298)

Midwives register 472 722 births 1980–2001
Women with no in-patient/ out-patient psychiatric history

9750 Case mothers: 20 209 births

239 366 Comparison mothers: 452 513 births

Obstetric complications (midwives register)

Fathers identified/siblings established (birth registration)

Maternal/paternal mental and physical morbidity (psychiatric, hospital morbidity registers)

Additional linkages of offspring to other population registers

Cerebral palsy	Psychiatric
Birth defects	Cancer
Intellectual disability	Mortality
School testing	Corrective services
Hospital morbidity	Child protection

Fig. 1 Western Australian registers linked for the population-based study of obstetric, developmental and neuropsychiatric outcomes in the offspring of women with severe mental disorders.

adequate proxies. For instance, census-based area-level data derived from address fields can be proxies for missing socio-economic data.[8] Where important fields are not available, the statistical impact of the missing variables may be modelled using data from other sources. Alternatively, it may be possible to nest a smaller clinical case–control study into the design of the larger register-based study or to link, with consent, register data for an individual to their survey data. Regardless of the approach taken, discussion of the potential impact on the study of missing confounders and effect modifiers, and justification of the strategy employed, are essential.

Building constructs using register data

A fifth key objective can be added to Mortensen's list: development of proxies and constructs from register data fields so that they closely approximate their primary form in the clinical setting. In psychiatric case register research there are three fields in particular that require careful scrutiny: the allocation of a diagnosis; the timing of onset of illness; and the measurement of illness severity. Assigning a diagnosis on the basis of longitudinal register data is particularly complex and will be used here to illustrate the difficulties.

First, the reliability of the diagnosis as recorded by a clinician and coded by the data clerk for administrative purposes will vary by register and time period. For each register, some evidence of the concurrent validity of the register diagnosis against a clinical interview-based diagnosis is essential. However, the review of validity studies by Byrne *et al* suggests that the number of validation studies is low and their quality variable.[9] Second, a register is dynamic and captures clinical and other data at key administrative time points during each episode of illness for an individual who may have had many contacts with services. The diagnoses recorded at these time points may change owing to the acquisition of new information by the diagnostician or changes in a patient's symptomatology over time. Researchers need to be explicit about how they handle multiple diagnoses. Some of the strategies that researchers employ include: application of a diagnostic hierarchy,[10–12] at least one discharge with the diagnosis of interest,[5] the last diagnosis recorded on the register,[13] or a calculation of the most frequent diagnosis, so-called diagnostic dominance.[14] Researchers using the Swedish psychiatric case register (which until recently only had access to in-patient data) have set a criterion of at least two in-patient admissions with the discharge diagnosis of interest.[15] In Western Australia, the register has long included out-patient contacts, and such information should not be lost. To assign a diagnosis of schizophrenia or affective psychosis, we have developed and validated an iterative algorithm based on the last diagnosis recorded at specific time points.[5] Third, changes in the official classification and diagnostic criteria over time also influence the recording of diagnoses. Researchers relying on longitudinal data affected by multiple revisions within a diagnostic classification need adequate correspondence between revisions. Deficiencies in correspondence may be subtle but critical. For example, most concordances assume that major differences between ICD–9[16] and ICD–9–CM[17] occur at the five-digit level only; in fact, coding at the four-digit level has a major impact on differentiating unipolar from bipolar diagnoses with categories reversed for some codes. Moreover, with each new revision, one generally finds a finer level of detail of diagnostic categorisation and a decision needs to be made whether to code up to the new classification (despite the splitting of categories) or down.

Of note, particularly in relation to the last two objectives, is the capacity to link records between psychiatric and other
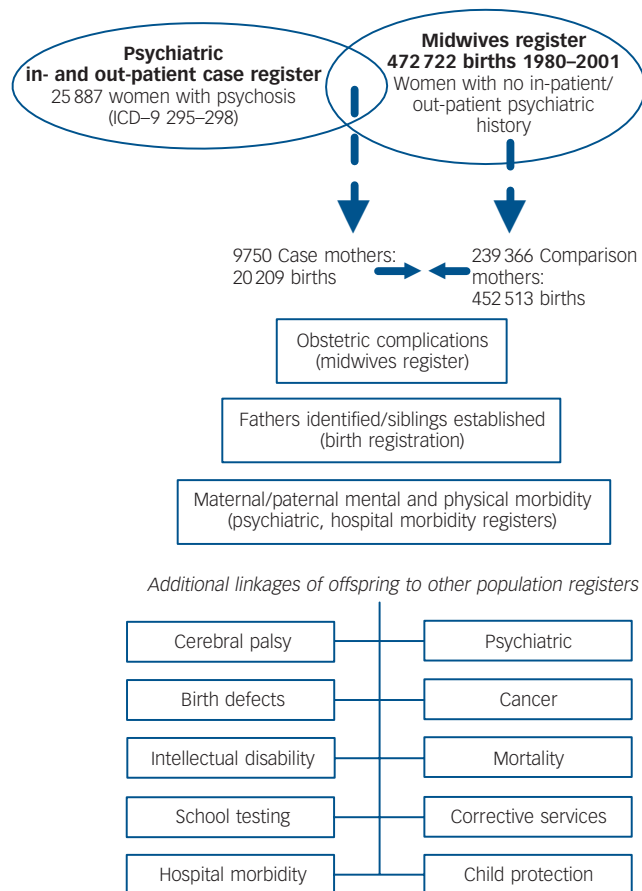
registers. Record linkage does not, of itself, validate the psychiatric case register but it does provide important additional material for analysis, subject to the validation of the data on the linked source. This includes data on confounders, risk modifiers, other exposures and other outcomes. The development of an integrated system of linkages across a jurisdiction to capture multiple sources of longitudinal data is a long-term project. The strategy employed in Western Australia has been described in some detail.[18] Of recent interest is the capacity to link to biosamples such as archived neonatal dried blood spot samples[19] and DNA samples,[18] thereby bridging an important gap between epidemiological and biological research. This is of particular importance in the area of genetics research.

Privacy and scientific accountability

Finally, although psychiatric case registers are an efficient and cost-effective data source, their use in large-scale epidemiological studies raises complex issues related to consent and privacy. Sibthorpe *et al*[20] make an important contribution to the discussion by differentiating the role of individual identity depending on the use of a register. For although the identity of the individual is critical to the use of the register as an administrative management and care planning tool, in research the identity of specific individuals is of no intrinsic consequence, except in the interim to enable cross-linkage. Notwithstanding that, access to registers is a privilege that comes with a large onus on researchers. The onus is not only to preserve confidentiality but also to be accountable to the people whose data they utilise. They need to ensure that the investigations undertaken are well-designed and of scientific merit, so that the findings are of benefit to the broader community, be it directly or indirectly, in the short-term or the long-term. Part of that scientific accountability includes ensuring that the use of register data meets appropriate standards.

Future directions

In other research modalities, we have standard psychometric measures to assess the reliability and validity of scales developed for use in research, while consortia and reference groups have proposed benchmarks for assessing the quality of: clinical trials,[21] systematic reviews and meta-analyses,[22] reviews of health care interventions[23] and qualitative research.[24] As more and more psychiatric case registers are employed for research purposes, it is becoming increasingly important for the community of researchers who use register data to set agreed benchmarks to assess the quality of a register and its research application. Some refinement of Mortensen's criteria (precision, selection bias, information bias, confounders),[4] integrating Allebeck's overlapping characteristics (coverage, attrition, representativeness and validity)[2] is a starting point. In turn, publications relying on register data need to provide sufficient details for reviewers and readers to assess how well the study meets these standards.

Vera A. Morgan, PhD, MSocSc, Neuropsychiatric Epidemiology Research Unit, School of Psychiatry and Clinical Neurosciences, The University of Western Australia; **Assen V. Jablensky**, MD, DMSc, FRCPsych, FRANZCP, Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Australia

Correspondence: Professor Vera A. Morgan, The University of Western Australia School of Psychiatry and Clinical Neurosciences, Level 3 Medical Research Foundation Building, Rear 50 Murray Street, Perth, Western Australia, Australia 6000. Email: vmorgan@cyllene.uwa.edu.au

First received 15 Feb 2010, final revision 18 Feb 2010, accepted 10 Mar 2010

References

1 Perera G, Soremekun M, Breen G, Stewart R. The psychiatric case register: noble past, challenging present, but exciting future. *Br J Psychiatry* 2009; **195**: 191–3.

2 Allebeck P. The use of population based registers in psychiatric research. *Acta Psychiatr Scand* 2009; **120**: 386–91.

3 Ten Horn G, Giel R, Gulbinat W, Henderson J. *Psychiatric Case Registers in Public Health*. Elsevier, 1986.

4 Mortensen P. The untapped potential of case registers and record linkage studies in psychiatric epidemiology. *Epidemiol Rev* 1995; **17**: 205–9.

5 Jablensky A, Morgan VA, Zubrick S, Bower C, Yellachich L. Pregnancy, delivery, and neonatal complications in a population cohort of women with schizophrenia and major affective disorders. *Am J Psychiatry* 2005; **162**: 79–91.

6 Morgan VA, Castle D, Page A, Fazio S, Gurrin L, Burton P, et al. Influenza epidemics and the incidence of schizophrenia, affective disorders and mental retardation in Western Australia: no evidence of a major effect. *Schizophr Res* 1997; **26**: 25–39.

7 Valuri G, Croft M, Morgan VA, Jablensky A. The high risk children of mothers with schizophrenia and other severe mental illness: modelling risk from conception to disease. *Schizophr Res* 2008; **102** (suppl 2): 169.

8 Australian Bureau of Statistics. *1996 Census of Population and Housing. Socio-Economic Indexes for Areas. Catalogue No. 2039.0*. ABS, 1998.

9 Byrne N, Regan C, Howard L. Administrative registers in psychiatric research: a systematic review of validity studies. *Acta Psychiatr Scand* 2005; **112**: 409–14.

10 Ruschena D, Mullen P, Burgess P, Cordner S, Barry-Walsh J, Drummer O, et al. Sudden death in psychiatric patients. *BMJ* 1998; **172**: 331–6.

11 Wallace C, Mullen P, Burgess P, Palmer S, Ruschena D, Browne C. Serious criminal offending and mental disorder. Case linkage study. *Br J Psychiatry* 1998; **172**: 477–84.

12 Sigurdsson E, Fombonne E, Sayal K, Checkley S. Neurodevelopmental antecedents of early-onset bipolar affective disorder. *Br J Psychiatry* 1999; **174**: 121–7.

13 Morgan VA, Jablensky A, Castle D. Season of birth in schizophrenia and affective psychoses in Western Australia 1916–1961. *Acta Psychiatr Scand* 2001; **104**: 138–47.

14 Kirkby KC, Hay DA, Daniels BA, Jones IH, Mowry BJ. Comparison between register and structured interview diagnoses of schizophrenia: a case for longitudinal diagnostic profiles. *Aust N Z J Psychiatry* 1998; **32**: 333–7.

15 Lichtenstein P, Björk C, Hultman CM, Solnick E, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish National Cohort. *Psychol Med* 2006; **36**: 1417–25.

16 World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, Ninth Revision* (ICD–9). WHO, 1977.

17 World Health Organization. *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD–9–CM). WHO, 1979.

18 Holman CDAJ, Bass AJ, Rosman D, Smith M, Semmens J, Glasson E, et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev* 2008; **32**: 766–77.

19 Korzeniewski SJ, Grigorescu V, Copeland G, Gu G, Thoburn KK, Rogers JD, et al. Methodological innovations in data gathering: newborn screening linkage with live births records, Michigan, 1/2007–3/2008. *Matern Child Health J* 2009; Apr 8 [Epub ahead of print].

20 Sibthorpe B, Kliewer E, Smith L. Record linkage in Australian epidemiological research: health benefits, privacy guards and future potential. *Aust J Public Health* 1995; **19**: 250–6.

21 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; **134**: 663–94.

22 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

23 Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6*. The Cochrane Collaboration, 2006 (http://www2.cochrane.org/resources/handbook/Handbook4.2.6Sep2006.pdf?q=resources/handbook/Handbook4.2.6Sep2006.pdf).

24 Miles MB, Huberman AM. *Qualitative Data Analysis* (2nd edn). Sage Publications, 1994.