ABSTRACT – **Objective:** Whilst database use for the purposes of audit is well established amongst clinicians, their research potential has yet to be fully realised. This review provides an international perspective on research using databases in psychiatry, considering the advantages and constraints of their use.

**Method:** Large databases used in psychiatric research were identified through a systematic search of MEDLINE and EMBASE databases. The types of research conducted with these databases were then grouped into categories for discussion. A narrative synthesis of results is presented, illustrated with examples.

**Results:** Research identified included pharmacoepidemiology, perinatal, aetiological, suicidology and health service research.

**Conclusions:** Databases have been used to address a range of questions in psychiatry, often by employing linkage between clinical and national administrative databases. The success of such research is not only dependent on appropriate methodology and database design, but also on clinician commitment to data entry. Further exploitation of linkage potential between databases is envisaged, alongside international development of clinical databases designed with the purpose of research in mind.

---

**Introduction**

Patient demographic, clinical and service use information is now routinely collected for the purposes of audit. Some clinicians may view this process with little enthusiasm, seeing it as a bureaucratic task driven by political agendas. It has been argued such scepticism reflects the medical community being slow to appreciate the
research value of databases (Jick 1997, Black 1999).

**Aims of the study**

(i) To identify major databases used in psychiatric research internationally and from this to ascertain the types of research conducted with their use.

(ii) To consider the advantages and constraints of database use in the light of these findings.

**Material and methods**

A selective literature review was conducted to identify large databases used in the routine gathering of patient information that have also been used for the purposes of psychiatric research. MEDLINE (1966-2003) and EMBASE (1980-2003) bibliographic databases were searched with the following strategy: Medical Subject Heading (MeSH) terms ‘mental disorders’ or ‘psychiatry’ combined with the keyword ‘database$’. Databases were not included for consideration if (i.) they were used only for audit purposes without exploration of a research hypothesis or (ii.) the database had been established as part of a specific study. Abstracts were scrutinized and potentially relevant articles obtained. A narrative synthesis of results is presented.

**Results**

MEDLINE and EMBASE searches identified 319 and 254 titles of possible interest respectively. Most studies used essentially administrative databases, containing basic diagnostic, demographic and service use data. These were databases associated with health care plans (e.g. Medicaid), army conscript records, adverse drug reaction monitoring systems or national hospital admissions registers. Databases charting hospital admissions or mental health service use included Western Australia’s Health Services Research Linked Database (WA Linked Database); the Hospital Episode Statistics (HES) in England; psychiatric hospital admission registers for Finland, Denmark, Sweden and Israel. Two databases containing more extensive clinical information were the UK’s General Practice Research Database and the Oxford Monitoring System for Attempted Suicide.

The database studies identified through the search were found to fall into five research categories or areas. These were pharmacoepidemiology; mental illness and pregnancy; aetiology of psychosis; suicidology; health service research. These categories are discussed in turn below with examples of specific studies provided to illustrate the types of research questions that can be addressed and the methods used to answer them. The databases identified and discussed in these categories are listed in Table I.

**Pharmacoepidemiology**

The World Health Organisation runs an international database on adverse drug reactions (ADRs) that has been used to demonstrate an association between clozapine and cardiomyopathy and myocarditis, which also exists with antipsychotics as a group to a lesser degree (Coulter et al. 2001). Drug company databases have also been used to investigate this area e.g. the Novartis cloza-
Table I

<table>
<thead>
<tr>
<th>Type of research</th>
<th>Databases identified – used alone or in combination</th>
</tr>
</thead>
</table>
| Pharmacoepidemiology                                 | World Health Organisation adverse drug reactions: international
|                                                      | Novartis Clozapine database: international
|                                                      | Medicaid Health Insurance database: patients under scheme in USA
|                                                      | German AMSP database: selected German hospitals
|                                                      | Danish Prescription Database: North Jutland County, Denmark
|                                                      | General Practice Research Database (GPRD): UK, selected practices. |
| Mental illness and pregnancy                         | GPRD: UK, as above
|                                                      | Denmark’s Psychiatric Register: national
|                                                      | Denmark’s Birth & Congenital Malformation Registers: national |
| Aetiology of psychosis                               | Stockholm County In-Patient Register: Sweden, Stockholm only
|                                                      | Sweden’s Birth Register: national
|                                                      | Swedish Army Conscript database: national
|                                                      | Swedish Register of Psychiatric Care: national
|                                                      | Israeli Draft Board Register: national
|                                                      | Israel’s Psychiatric Case Register: national |
| Suicidology                                          | Western Australian Health Service Research Linked Database: regional
|                                                      | Victorian Psychiatric Case Register: regional, Australia
|                                                      | Victorian Dept. of Justice’s Unnatural Deaths Register: regional
|                                                      | Longitudinal Labour Market Register: Denmark, national sample
|                                                      | Danish Medical Register on Vital Statistics: national
|                                                      | Danish Psychiatric Central Register: national
|                                                      | Oxford Monitoring System for Attempted Suicide: UK, regional |
| Health service research                              | Medicaid Health Insurance database: USA, as above
|                                                      | Department of Veterans Affairs database: USA, national
|                                                      | Hospital Episode Statistics database: UK, national
|                                                      | GPRD: UK, as above
|                                                      | Western Australian Health Service Research Linked Database: regional
|                                                      | Finland’s Hospital Discharge Register: national
|                                                      | Finland’s Crime Registers: national |

The Clozapine database covers 2.8 million patient years spanning 27 years. It has been used to study ECG changes (understood as vulnerability markers for fatal arrhythmias with antipsychotic drugs), finding at therapeutic doses the majority of changes were confounded by other medications or physical problems (Warner & Hoffmann 2002).

Databases associated with health care plans, such as Medicaid in America, have also been extensively used in drug safety research using retrospective cohort designs. (Ray et al. 2001) used a cohort of 481,744 patients with 1,282, 996 person-years of follow-up and found a large relative and absolute increase in the risk of sudden cardiac death. (Hennessy et al. 2002) included newer atypical drugs in their analysis and concluded high dose thioridazine may have the highest risk. (Lund et al. 2001) explored the hypothesised link between clozapine and the development of diabetes and hyperlipidemia, concluding clozapine is not an independent risk factor, but it might modify other risk factors such as weight gain in younger patients. Country-limited ADR databases have similar potential. The German AMSP database covers some 29 hospitals and has been used to compare ADRs...
between tricyclic antidepressants (TCAs) and selective serotonin reuptake inhibitors (SSRIs). Similar rates were found for the two groups but with different reaction profiles (Grohmann et al. 1999): toxic delirium was most common with TCAs, whilst non-delirious psychic and neurological symptoms predominated with SSRIs. In Italy researchers using an ADR database for the elderly (covering some 81 hospitals), found moderate alcohol intake to increase the risk of ADRs, especially amongst women (Onder et al. 2002).

Combining national prescribing plus non-psychiatric health care databases has been used in Denmark to look at cancer incidence with antidepressants: amongst the 39,807 adults prescribed antidepressants between 1989 to 1995, users of TCAs were found to have an excess of non-Hodgkin’s lymphoma (Dalton et al. 2000).

The General Practice Research Database (GPRD) in the UK is sourced from over 400 practices. With 35 million patient years of data -including patient diagnoses, prescriptions, hospital referrals and treatment outcomes- it is the world’s largest anonymised clinical database. It is sufficiently complete for psychiatric research purposes (Jick et al. 1991) and diagnostic recording of psychoses has been shown to be accurate (Nazareth et al. 1993). Much of the focus of psychiatric GPRD research has been on antidepressants. (Jick et al. 1995) found risk of suicide was not determined by antidepressant type but with being male, a recent prescription, high doses, use of different antidepressant types previously and a history of feeling suicidal. Other GPRD studies have reached similar conclusions (Derby et al. 1992, Jick et al. 1992, Donovan 1996). De (Abajo et al. 1999) found SSRIs to increase upper gastrointestinal bleeding risk, especially with concurrent non-steroidal anti-inflammatory drugs or aspirin use, though these findings have been subsequently questioned on the basis of possible confounding (Williams 2000, Dickenson et al. 2000).

Research using the GPRD to study the suggested link between the MMR (measles, mumps & rubella) vaccination, gastrointestinal disease and autism has been widely publicised due to the levels of public interest. Firstly a time-trend analysis found no correlation between the prevalence of MMR vaccination and the increase in autism (Kaye 2001). Secondly a nested case-control study found children with autism were no more likely to have had gastrointestinal disorders before their diagnosis of autism, and no temporal association existed between the vaccine and onset of gastrointestinal symptoms in children with the disorder (Black et al. 2002).

Mental illness and pregnancy

The GPRD has been used to investigate the fertility rate in women with psychotic disorders (Howard et al. 2002), which remains markedly lower in women with schizophrenia and related disorders compared with the general population. These pregnancies are also at high risk of stillbirths and subsequent perinatal deaths (Howard et al. 2003), which may be partly due to decreased attention to routine antenatal care, as alcohol and smoking consumption was less likely to be recorded in women with psychotic disorders compared with controls (Howard et al. 2003). These findings highlighted the importance of good primary and secondary health care liaison for optimal obstetric care.

Denmark’s Psychiatric and Birth and Congenital Malformation Registers have
also been used in conjunction to study pregnancy outcome in women with schizophrenia. Children of women with schizophrenia have been found to be at increased risk of preterm birth, low birth-weight and being small for gestational age (Bennedsen et al. 1999), with an increased risk of sudden infant death syndrome (Bennedsen et al. 2001).

Aetiology of psychosis

Sweden’s Stockholm County In-Patient Register has been used alongside Swedish birth records to study obstetric complications in relation to schizophrenia. 648 people diagnosed between 1971 and 1994 were identified, their birth records traced and the frequency of obstetric complications calculated in comparison to 1,043 controls. Signs of asphyxia were associated with an increased risk of schizophrenia (Dalman et al. 2001).

Swedish army conscript data has also been used to explore risk factors for developing schizophrenia. Questionnaires administered to a cohort of 50,087 men at conscription between 1969 and 1970 provided information on family background, social adjustment, relationships and drug use. This data was linked to the Swedish National Register of Psychiatric Care and individuals subsequently diagnosed with schizophrenia between 1973 and 1983 identified. Early problems with interpersonal relationships were more common in those who went on to develop schizophrenia, although these were common in the cohort overall (Malmberg et al. 1998). Israel’s army conscript data has also been used to research premorbid intellectual, language and behavioural functioning in schizophrenia. Merging the Israeli Draft Board Registry with Israel’s psychiatric case register, (Reichenberg et al. 2002) demonstrated that subjects who developed schizophrenia showed significant premorbid deficits on all these domains, unlike those with nonpsychotic bipolar disorder or schizoaffective disorder.

Suicidology

The Western Australian Health Services Research Linked Database (WA Linked Database) records birth information, inpatient hospital morbidity, cancer registrations, outpatient mental health service use and death data for the 1.7 million people in Western Australia (Holman et al. 1999). Suicide research using this resource found the risk for psychiatric patients was higher than the general population, and it was highest in the first week after discharge, decreasing exponentially over time, whilst remaining higher for at least 5 to 10 years (Lawrence et al. 2001). In addition, suicide rates overall were found to have increased between 1980-98. Linking databases of deaths and psychiatric service use in Australia’s Victoria, cases of suicide and controls could be distinguished only on the basis of being more likely to be male, not-working and to have had recent contact with psychiatric services (Pirkis et al. 2002). Similarly, combining three Danish databases (Longitudinal Labour Market Register; Danish Medical Register on Vital Statistics; Danish Psychiatric Central Register) 811 suicide cases and 79,871 controls were compared to determine suicide risk predictors (Qin et al. 2000). A history of hospitalised mental illness was found to be the strongest risk factor and differences between the genders were evident: unemployment, single status and sickness absence were risk factors for men, whilst
having a child under 2 years was protective for women.

The Oxford Monitoring System for Attempted Suicide database in the UK is an on-going record of research-relevant demographic and clinical data on all patients presenting after an episode of deliberate self-harm (DSH) to the John Radcliffe hospital in Oxford. Over 70 published studies have utilised this resource to address a range of research questions, as the following three examples illustrate. The commonest diagnosis amongst patients who self-harm was found to be depression (70.7%), raising the question whether depressive disorders in this group are under-recognised and under-treated (Haw et al. 2001). The same study found 45.9% of this group to have a personality disorder, with 10.7% having an eating disorder. In 1998 legislation was introduced in the UK to limit the number of paracetamol or salicylates tablets sold per pack in an attempt to reduce fatalities from overdoses. Analysing annual death rates from paracetamol and salicylate poisoning, (Hawton et al. 2001) provided evidence that this legislation has been of benefit, with reductions in both the annual number of deaths (21% and 48% reductions respectively) and liver transplant rates after paracetamol poisoning (66%). Another study drew attention to the effect of media representations of suicide, demonstrating a 17% rise in self-poisoning the first week and 9% the second week after a paracetamol overdose was depicted on the popular UK television drama Casualty (Hawton et al. 1999).

Health service research

The Hospital Episode Statistics (HES) database records all National Health Service (NHS) admissions in England to chart health care needs across the country and measure variations in effectiveness between hospitals (Registrar-General, 1953). The earliest epidemiological research using the HES explored patterns of mental illness across social class, finding admissions for most types of illness (with the exception of alcohol related problems) were more common amongst those from a lower socio-economic group (Registrar-General, 1958). Geographic variations in service use have also been demonstrated (Jarman et al. 1992), models constructed to account for this observation (Thornicroft 1991, Smith et. al. 1996) and a usable needs index developed (Glover 1998). The HES data is however limited to basic demographic details, diagnostic information and length of admission. Its use is therefore limited by its lack of information on major confounders. In addition, it is only 67% complete according to the most recent figures (NHS Executive 1999).

Inequalities in health between different socio-economic groups in the UK have been demonstrated using the GPRD. (Moser 2001) found strong deprivation gradients for treated depression, anxiety, schizophrenia as well as coronary heart disease and non-insulin treated diabetes. Register research also allows comparisons to be made between prevalence rates and service use not only across different geographical areas but also between different countries. The Andalusian Register of Schizophrenia in Granada, southern Spain, is one such register currently being developed and utilised in this way (Moreno Kustner et al. 2003).

The GPRD has also been used to identify areas of sub-optimal care for people with mental health problems. (Nazareth et al. 1993) looked at patients with a diagnosis of schizophrenia and found they consulted their general practitioner with similar fre-
quency to patients with chronic physical disorders, but the index disorder was more often assessed in the latter group. Patients with schizophrenia therefore appeared to receive little specific management in general practice, suggesting a more structured approach is required. In the USA, the Department of Veterans Affairs database has also been used to study health care use amongst people with mental health problems e.g. (Cradock et al. 2002) found patients with psychiatric diagnoses had fewer medical visits than controls. This difference was especially high for young adults with schizophrenia and posttraumatic stress disorder, and adults of all ages with bipolar disorder.

The suggestion of under-use of health care resources amongst people with mental health problems is of further concern in the light of research into levels of need amongst this group. For example the Saskatchewan Health database in Canada has been used to study cardiovascular disease in patients with schizophrenia, finding increased rates of arrhythmia, syncope, heart failure, stroke, transient cerebral ischemia, and diabetes in relation to controls (Curkendall et al. 2004).

In the USA some small studies had tentatively suggested disparities in antipsychotic prescribing for patients of different ethnic groups. This hypothesis was explored further using the Medicaid health insurance database; it was found African-American subjects were indeed less likely than Caucasian subjects to receive clozapine or risperidone and more likely to receive depot medication (Kuno & Rothbard 2002).

In Australia the WA Linked Database (as described above) has been used to explore the associations of mental illness with cancer. Little difference in incidence rates between psychiatric patients and the general community were found yet mortality rates amongst psychiatric patients were 39% higher in males and 24% higher in females (Lawrence et al. 2000). Examining hospital treatment for ischaemic heart disease, rates of admissions were similar amongst the two groups but revascularisation procedures were performed less often for patients with psychiatric illness, suggesting their cardiac treatment may be sub-optimal (Lawrence et al. 2003).

Psychiatric service use data amongst offenders in Finland has been used to investigate hypothesised relationships between mental illness and criminality. Rates of admissions amongst male offenders were calculated using the Northern Finland 1966 Birth Cohort in conjunction with the Finnish Hospital Discharge Register and national crime registers. Psychiatric admission rates for both violent and non-violent male offenders were significantly higher than males with no criminal history. Amongst violent males, only half of inpatient days due to psychiatric illness actually occurred in psychiatric hospitals, suggesting this group may often be treated at an inappropriate health care level (Timonen et al. 2000).

Limitations

Firstly, we have sought to provide an international outline of this area, rather than an exhaustive (and unfeasible) review of all the databases ever used in psychiatric research. As our search strategy was necessarily broad across all psychiatric disciplines and research areas, we acknowledge that some databases familiar to researchers in specific areas will not have been identified. Secondly, this review does not address...
database validity. We are currently in the process of conducting a systematic review of studies that validate databases for research purposes.

Discussion

The types of studies described above illustrate many of the potential advantages and disadvantages of database use in research. It is of note that many of the databases identified were linked with other health care, birth or criminal records for the purposes of addressing specific research questions. Linkage techniques allow data about an individual recorded in different databases at different times to be brought together. This is particularly useful in psychiatric research where information relating to risk, illness course and service use is often held by disparate agencies (Smith 1997).

The advantages of database use for psychiatric research highlighted here include:

i) Large databases provide statistical power. Rare important outcomes for psychiatry such as suicide or autism can therefore be studied.

ii) Large databases can be representative.

iii) Database research can be done relatively quickly. The time-consuming task of collecting data de novo for individual studies can be bypassed with a pre-existing dataset.

iv) Database research can be relatively inexpensive. Costs can be spread over the database’s administrative, audit and research uses (Black 1997).

v) Potentially large database research could complement individual randomised controlled trials (RCTs). RCTs may establish intervention efficacy, but the patients involved or the treatment context may be atypical. Databases can be used to calculate the theoretical applicability of RCT findings for the wider population of interest by comparing study patients’ characteristics to those in a representative ‘real-world’ patient database. This approach, still relatively novel, is illustrated by an example from the medical intensive care research literature. The PROWESS RCT provided evidence for the use of recombinant activated protein C in severe sepsis. By comparing trial patients to those in the national intensive care database, differences could be adjusted for and the benefits and likely costs of routinely using this intervention calculated (Padkin et al. 2001). In addition, databases can be used after RCTs for non-randomised analyses of outcome amongst large numbers of patients in ordinary clinical conditions. This has the added advantage that a longer longitudinal perspective can be taken than that usually afforded to individual studies.

Disadvantages of database use for psychiatric research:

i) Data may be limited in scope. The information available will reflect the origins of the database and the priorities that underpinned its formation. Proxy measures used as a consequence may be misleading. For example, studies that use first admission rates as an approximation of incidence assume everyone with the condition under study will be admitted, and their diagnosis at that time will be the final one (Glover 2003). Certainly this has not been found to be the case for schizophrenia (Goldacre et
al. 1994). Less obviously a proxy measure, prescribing data is often implicitly assumed to be synonymous with medication use, despite widespread agreement that around 40-60% of patients are non-adherent (Nichol et al. 1999). With limited data controlling for confounders may also be difficult. Basic questions regarding disorder prevalence and incidence may therefore be imprecise.

ii) Data may be unreliable. Errors may stem from any of the four stages of data collection and recording i.e. from the patient, recorder, data-collection procedure or system design (Baldwin et al. 1965). For example, diagnoses might be inaccurate if operationalised diagnostic criteria are not used.

iii) Databases may be incomplete. For example, whilst the General Practice Research Database (GPRD, discussed further below) is 99% complete for prescribing and adverse drug event data, lifestyle data fares poorly in comparison e.g. only 15% of records contain information on alcohol use in women of childbearing age (Howard et al. 2002). This may reflect clinician bias in focusing on interventions and outcomes (as opposed to cause or prevention) in their patient consultations and / or data entry.

iv) Data on outcomes may not reflect the priorities of service users or their carers. Mental health research has only relatively recently begun to assimilate their perspectives into the evaluation of interventions (Trivedi & Wykes 2002). Individual studies are shaped by the prevailing research climate; large databases conceived years ago can be monolithic in comparison and updating them a time-consuming and expensive task potentially hampered by out-of-date computer software.

v) Research using databases is often opportunistic. Both hypothesis generation and investigation can be constricted by the data available, e.g. socio-economic data may not be available and therefore cannot be controlled for in any analysis. This may have far-reaching implications in what it known about a topic, or what is thought to be important. Research findings are assimilated into the literature and used in turn to generate further hypotheses for investigation; such conceptual biases may therefore be self-perpetuating.

vi) A dataset is only as useful as the skill of those handling it allows. Recent controversy surrounding the postulated association between venous thromboembolism and the third generation oral contraceptive pill illustrates the potential pitfalls of using databases in pharmacoepidemiology research (Skegg 2000). Essentially two research teams used different methods to investigate this question and drew opposite conclusions from the same dataset - in this case the GPRD. (Farmer et al. 2000) found the incidence of venous thromboembolism did not fall following the 1995 ‘pill scare’ and subsequent reduction in pill use, whilst (Jick et al. 2000) re-analysed the data and contested that the risk with third generation preparations was around twice that of those containing levonorgestrel. Whilst Farmer et al. (2000) used a time correlation study, Jick et al. (2000) used a cohort plus nested case-control designs, controlling extensively for potential confounders and including a time correlation approach for comparison. Such contrasting conclusions from the same dataset should not be viewed entirely with pessimism. It is the public availability of such a database that allows different groups of researchers to revisit the data, reanalyse each other’s work and engage in debates.
about research methodology to a degree not previously possible (Glover 2003).

Conclusion

The variety of databases identified and the examples of studies described demonstrate the potential for databases to be used across a range of psychiatric research, from pharmacoepidemiology to health service research. The large numbers involved can offset the limitations of a dataset (Yates 1982), provided researchers are trained in its use and have the necessary epidemiological skills (Jick 1997). In addition, it is essential that there is continued monitoring of the quality, detail, accuracy and completeness of data entry, which requires adequate funding (Jick 1997, Black 1998).

In future the research potential from linkage between existing datasets is likely to be further exploited. In addition, specialities could potentially establish comprehensive national clinical databases that combine both administrative and more detailed clinical data (Black 1997). New databases need to be designed with the demands of research, and the priorities of service users and carers borne in mind. Data collected should include all patient/illness characteristics known to affect outcome, with standardised definitions of conditions and outcomes used (Black 1997). An important development in this direction in England is the new Mental Health Minimum Data Set (Glover 2000). Ultimately the success of database research depends on the commitment of individuals to on-going data collection. It is essential clinicians have a sense of ownership of such projects, and understand their potential as valuable resources for research.

References


Adress of correspondence:
Dr. Nicola Higgins, Specialist Registrar, South London and Maudsley NHS Trust
Section of Community Psychiatry (PriSM), Health Services Research Department, Institute of Psychiatry PO29, De Crespigny Park, LONDON SE5 8AF Fax: 020 7277 1462 E-mail: n.higgins@iop.kcl.ac.uk