

Outcomes research in mental health

Systematic review[†]

SIMON M. GILBODY, ALLAN O. HOUSE and TREVOR A. SHELDON

Background Outcomes research involves the secondary analysis of data collected routinely by clinical services, in order to judge the effectiveness of interventions and policy initiatives. It permits the study of large databases of patients who are representative of 'real world' practice. However, there are potential problems with this observational design.

Aims To establish the strengths and limitations of outcomes research when applied in mental health.

Method A systematic review was made of the application of outcomes research in mental health services research.

Results Nine examples of outcomes research in mental health services were found. Those that used insurance claims data have information on large numbers of patients but use surrogate outcomes that are of questionable value to clinicians and patients. Problems arise when attempting to adjust for important confounding variables using routinely collected claims data, making results difficult to interpret.

Conclusions Outcomes research is unlikely to be a quick or cheap means of establishing evidence for the effectiveness of mental health practice and policy.

Declaration of interest S.M.G. is supported by the Medical Research Council Fellowship Programme in Health Services Research.

Randomised controlled trials have generally been accepted as the gold standard when deciding which interventions work in psychiatry (World Health Organization, 1991). Most randomised studies in psychiatry have investigated the effect of drug or psychotherapy interventions in tightly controlled and largely artificial experimental conditions (Hotopf *et al*, 1997; Thornley & Adams, 1998), while patients, clinicians and other decision-makers need to know how treatments work in the real world and whether they are cost-effective under routine conditions (Wells, 1999). Important questions relating to the organisation and delivery of mental health services are also rarely addressed in randomised trials (Gilbody & Whitty, 2002).

The need for research relating to effectiveness (rather than efficacy) has prompted a number of responses. One has been the call to conduct randomised trials in 'real world' settings, using pragmatic designs (Hotopf *et al*, 1999); another has been to synthesise various data sources using decision analysis (Lilford & Royston, 1998). A response that has been influential in the USA in the past decade involves the analysis of large databases of patient information collected in routine care settings – known as outcomes research (Anonymous, 1989; Ellwood, 1988; Wennberg, 1991).

ORIGINS OF OUTCOMES RESEARCH

The 'outcomes' movement emerged as a consequence of rapidly escalating costs, acceleration of the introduction of new health technologies and evidence of massive regional variations in the delivery of health care in the USA (Wennberg, 1990; Thier, 1992; Wennberg *et al*, 1993; Davies & Crombie, 1997). Paul Ellwood, in his 1988 Shattuck lecture (Ellwood, 1988), ushered in the modern outcomes movement and called for the routine

collection of outcome measures by clinicians. He proposed that these records should be assimilated in large databases that would form a resource for clinical and health services research. Such data could eventually be used *inter alia* to compare existing treatments and to evaluate new technologies, thereby avoiding both the expense of clinical trials and the loss of generalisability that results from selective recruitment to conventional efficacy trials.

The Agency for Health Care Policy and Research (AHCPR), now the Agency for Healthcare Research and Quality (AHRQ), was established in the USA under public law in 1989 in order to conduct outcomes research into common medical conditions, with the establishment of patient outcome research teams (PORTs; Wennberg *et al*, 1993). The research programme was allocated US\$6 million in its first year, rising to \$63 million in 1991, with the purpose of using routine outcomes data to determine 'outcomes, effectiveness and appropriateness of treatments' (Anderson, 1994). It was decreed by Congress via the General Accounting Office (1992) that new primary research conducted by the PORTs was not to take the form of the traditional randomised controlled trial; rather, it was to be observational in design, utilising the vast amounts of data routinely collected on US patients. This health research policy produced a new breed of health researchers known as database analysts (Anonymous, 1989, 1992), with the motto 'Happiness is a humongous database' (Smith, 1997).

Outcomes research differs from traditional observational or quasi-experimental research in a number of ways. The key difference is that outcomes research evaluates competing interventions that are already used in routine care settings, using routine data collected by clinicians or by other agencies (such as insurance companies), whereas quasi-experimental studies implement interventions in one setting or in one group of patients, and compare outcomes with patients who have not been subjected to the intervention (Gilbody & Whitty, 2002). Quasi-experimental studies are therefore more like randomised trials and are considered to be clearly different in their approach and ethos to outcomes research (Aday *et al*, 1998). The outcomes that are studied in outcomes research are generally those that are already collected as part of routine care, although there is no reason why these cannot be extended

[†]See editorial, pp. 1–2, this issue.

in the light of the specific question being asked.

The application of outcomes research to UK mental health services has been advocated in psychotherapy (Barkham *et al*, 1998; Mellor-Clarke *et al*, 1999; Guthrie, 2000; Margison *et al*, 2000). Similarly, the pharmaceutical industry is keen to extend the method in the evaluation of new and relatively expensive drug therapies; for example, the Schizophrenia Outpatient Health Outcomes Study (SOHO), funded by Eli Lilly, aims to recruit European collaborators to collect outcomes from patients with schizophrenia who are in receipt of typical and atypical drugs. Others have urged caution (Sheldon, 1994); the principal concerns that have been expressed about outcomes research are their observational (rather than experimental) design; the poor quality of the data that are used; the inability to adjust sufficiently for case mix and confounding; and the absence of clinically meaningful outcomes in routinely collected data (Iezzoni, 1997).

This article presents the first systematic overview of the application of outcomes research in evaluating competing interventions in mental health, and discusses how this approach might meet the needs of clinicians and decision-makers.

METHOD

A search was made for all published examples of outcomes research conducted in psychiatric settings or among psychiatric populations, where two or more competing interventions were compared. (See Appendix for search strategies.)

Inclusion criteria

Reports were included if they fulfilled the following criteria:

- (a) The research was conducted in a setting that was part of usual care in a health-care system
- (b) The outcome data used were those collected routinely for all patients – either for administrative purposes or as a means of monitoring outcomes in the service being evaluated.

Exclusion criteria

We excluded studies that examined only the costs or processes of illness and health care from routinely collected data, with no linkage to the outcomes of care. For

example, primary care prescription databases have been used to conduct research into newer psychotropic drugs (e.g. Donoghue *et al*, 1996), but since they are not linked to patient-level data and outcomes, they cannot be considered as outcomes research.

Also excluded were quasi-experimental or non-randomised evaluations of new technologies, where an intervention was implemented and outcomes measurement systems established only in the course of its evaluation (Cook & Campell, 1979). For example, the PRiSM psychosis study (Thorncroft *et al*, 1998) is an example of a quasi-experimental evaluation of a model of community care for those with severe mental illness, where districts were non-randomly allocated to implement an experimental service, and outcomes were measured under experimental and control conditions as part of the study.

Studies that only examined the relation between patient characteristics and outcome, with no direct comparison between competing treatments or health policy strategies (e.g. Rosenheck *et al*, 1997), were excluded, as were reports of routine outcomes measurement in practice, with no direct report of comparative service or treatment evaluations based on the data.

Data extraction

Data were extracted on the following topics: population; clinical or organisational question being asked; setting; sample size and length of follow-up; outcomes studied and their source; adjustment for case mix and confounding; and results.

RESULTS

Despite the widespread advocacy of outcomes research in health care, only nine published examples were found relating to mental health. Most of these were published in the past 3 years, highlighting an increase in the use of the design. The scope, design and analysis of the studies we identified are summarised in Table 1, and their most important characteristics are reviewed below.

Research questions addressed

Outcomes research has been used broadly in two areas of mental health research.

Evaluation of mental health policy, including aspects of service delivery, organisation and finance

The earliest and perhaps most important example of outcomes research in mental health is the Medical Outcomes Study (MOS) conducted by the RAND Corporation in the USA in the late 1980s (Tarlov *et al*, 1989; Wells *et al*, 1989, 1996). The design and objectives of this study were shaped by US health-care policy debates on the role of financing and reimbursement strategies in private care (fee for service *v.* prepayment) and on the place of speciality (secondary) care.

The researchers justified the use of observational methods in two ways. First, they claimed that the cheaper design and reduced burden on participants could maximise the number and range of collaborators and patients, particularly from non-research settings. Second, they claimed that the specific research questions precluded the use of randomisation, since the very act of randomisation would alter the functioning of existing health-care delivery systems (Wells *et al*, 1996).

Three other studies looked at health policy and organisation questions, such as the consequences of the withdrawal of mental health benefits from insurance plans (Rosenheck *et al*, 1999a), the effectiveness of services directed at homeless people (Lam & Rosenheck, 1999) and the difference in outcome between privately and publicly funded health providers (Leslie & Rosenheck, 2000).

Evaluation of new technologies

Four studies (Hong *et al*, 1998; Melfi *et al*, 1998; Croghan *et al*, 1999; Hylan *et al*, 1999) used an outcomes research design to demonstrate the worth of new antidepressant and antipsychotic medication in routine care settings. One further study (Rosenheck *et al*, 2000) examined the value of an innovative psychosocial intervention for those with war-related post-traumatic stress disorder (PTSD).

Source and choice of cases and outcomes

Outcomes studies can be broadly divided into those that collect data prospectively on a service-wide level, where the choice of outcomes is decided *a priori* and is influenced by the research question or population under examination, and those that

Table 1 Examples of outcomes research in mental health

Study	Clinical problem, population and setting	Clinical or organisational question or hypothesis being examined	Source of outcomes data and sample size	Outcomes studied	Methods used in adjusting for case mix	Results
Medical Outcomes Study (Wells et al, 1989)	Depression (major depression, dysthymic disorder and sub-threshold) being managed in family practices and specialist health care providers	<ol style="list-style-type: none"> How does treatment for depression differ by speciality and payment system? How does outcome for depression differ by speciality and payment system? How can care for depression become more cost-effective? 	Data routinely collected by clinicians and research workers during the course of the study on 1772 patients	<p>Detection of depression by physicians</p> <p>Adequacy of treatment</p> <p>Depressive symptoms (including HRSD scores)</p> <p>Health status (including SF36)</p>	Baseline demographic data and case mix measured and adjusted for (including medical comorbidity, psychiatric comorbidity, past history of depressive episodes)	<p>Depression is generally under-recognised, inadequately treated and associated with a poor level of functioning</p> <p>Depression is associated with poorer quality treatment and outcome when a prepayment plan is in place, rather than a fee for service</p>
Lam & Rosenheck (1999)	Severe mental illness among homeless people contacted through street outreach	Is case management as effective for homeless people contacted on the streets as for those contacted through shelters and other service agencies?	Routinely collected data from a 5-year, 18-site demonstration project which established and sought to evaluate outreach services for the homeless mentally ill; n=5431 (Randolph et al, 1997)	<p>Depressive and psychotic symptoms</p> <p>Alcohol and drug misuse</p> <p>Housing</p> <p>Paid employment</p> <p>Social support</p> <p>Quality of life and service use</p>	Those in receipt of street outreach (n=434) were compared with those receiving conventional outreach after adjusting for baseline socio-demographic differences, and baseline differences in psychosis and substance misuse	Assertive outreach resulted in client improvement in 14 out of 20 outcome indicators. These benefits persisted and were similar to conventional outreach, following adjustment for case mix and confounding
Rosenheck et al (1999a)	Mental health service use among those enrolled in a health insurance plan following mental health spending cutbacks	Do cutbacks of mental health coverage by an insurer result in increased non-mental health service utilisation and reduced productivity?	Employees work records and health care claims data relating to 20 814 employees in a single US corporation	<p>Mental health and non-mental health service use (number of days of in-patient and out-patient health care)</p> <p>Health-care costs</p> <p>Days absent from work</p> <p>Length of stay</p> <p>Readmission rates (14, 30 and 180 days after discharge)</p> <p>Proportion receiving out-patient care</p>	Baseline differences between years in terms of socio-demographic factors, employment, income and state of employment	Reduction in mental health care utilisation was accompanied by a marked increase in non-mental health care service use and costs, and sick time
Leslie & Rosenheck (2000)	Individuals in receipt of US public sector (VA) and privately insured in-patient mental health care followed up for 6 months following discharge	Is publicly insured health care of lower quality and associated with poorer outcome compared with privately insured health care?	Routinely collected VA outcomes data were available on 180 000 in-patient episodes	<p>Readmission rates (14, 30 and 180 days after discharge)</p> <p>Proportion receiving out-patient care</p> <p>Quality of life and service use</p>	Adjustment made for known and measured confounders (age, gender, diagnostic category and psychiatric comorbidity)	<p>The VA patients were older and more prone to psychiatric illness</p> <p>Quality indicators and outcome were poorer for VA care than privately insured care</p> <p>The results are largely impossible to interpret, given that the observed difference may be real, or an artefact of case mix</p>
			Routinely collected data from 7 million privately insured lives were available on a commercially available database – 6000 in-patient episodes were selected		No data available on important confounders, including socio-economic status, employment, homelessness, health status and level of disability	

Table 1 (continued)

<p>Rosenheck <i>et al</i> (2000)</p>	<p>US patients with chronic war-related PTSD being treated in VA in-patient programmes Followed up for 4 months</p>	<p>Is an innovative psychosocial treatment – compensated work programme (CWP) – effective in routine care settings?</p>	<p>Routine data for all patients in receipt of VA in-patient mental health care Supplemented by disease-specific measures collected for all patients in receipt of care for PTSD Complete data on 542 patients in receipt of CWP, with 542 matched controls in receipt of routine or standard care for PTSD Compliance with treatment guidelines operationally defined as having made a claim for four or more antidepressant prescriptions over a 6-month period following initiation of medication 4052 patients classified into one of three groups, according to whether they met this criterion from Medicaid claims records</p>	<p>PTSD symptoms Substance misuse Violent behaviour Employment and medical status</p>	<p>Matching patients to controls by selecting characteristics that predict participation in the intervention condition (propensity scoring; Rubin, 1997) Logistic regression of baseline differences on PTSD symptom scores between CWP patients and controls</p>	<p>CWP has no impact on any of the outcomes measured, compared with controls, when adjusted analyses were conducted The treatment is likely to be clinically and cost ineffective. A formal randomised trial is not justified on the basis of this observational study</p>
<p>Melfi <i>et al</i> (1998)</p>	<p>US patients in receipt of antidepressant medication for depressive disorders</p>	<p>Does adherence to antidepressant treatment guidelines prevent the relapse and recurrence of depression?</p>	<p>Relapse or recurrence during an 18-month follow-up period was defined as the initiation of a new antidepressant prescription, or by evidence of a suicide attempt, hospitalisation, mental health-related emergency room visit, or receipt of electroconvulsive therapy</p>	<p>A series of general comorbidity adjustments were made using hospitalisation for any other physical disorder, together with demographic variables Severity of depression was controlled for using proxy measures, including whether an individual was seen by a mental health specialist</p>	<p>Patients with four or more prescriptions of antidepressants were less likely to relapse</p>	
<p>Croghan <i>et al</i> (1999)</p>	<p>Depression being managed in primary care</p>	<p>Does specialist referral for psychotherapy improve compliance with antidepressant therapy, compared with those managed exclusively in a primary care setting?</p>	<p>Use of antidepressants ascertained from claims Continuous medication use over 6 months is taken to be a proxy measure of effective antidepressant therapy and good outcome (Agency for Health Care Policy Research, 1993) Total health care costs were also measured from cost claims data</p>	<p>There were substantial differences between those in receipt of care in primary and specialist settings in terms of age, gender and previous history of depression Previous claims, hospitalisations and diagnoses of depression were used to adjust, using logistic regression</p>	<p>Referral to a specialist increases the chance of receiving continuous antidepressant therapy by 11% in adjusted analyses. The authors calculate cost-effectiveness ratios to achieve this benefit and conclude that continuous medication is likely to be a good proxy measure of improved outcome</p>	

Table 1 (continued)

Study	Clinical problem, population and setting	Clinical or organisational question or hypothesis being examined	Source of outcomes data and sample size	Outcomes studied	Methods used in adjusting for case mix	Results
Hylan <i>et al</i> (1999)	Patients in receipt of pharmacotherapy for depression in primary care settings	Is there a difference between different SSRI antidepressants in terms of patient compliance?	A commercially available medical insurance database of linked pharmacy and medical claims data on 750 000 individuals Complete episodes available on 1034 patients in receipt of a new SSRI prescription	Continuous prescription of the same antidepressant, without dosage change, or switch between different drugs or drug classes over 6 months, was taken as a proxy measure of a successful initial choice of antidepressant	Logistic regression of available confounders included demographic details; severity of depression from ICD codes; comorbid drug and alcohol problems; comorbid physical disorder (counts of other ICD codes); provider characteristics (primary care or specialist)	Patients in receipt of fluoxetine were more likely to receive continuous prescriptions over a 6-month period, when compared with sertraline or paroxetine. The authors conclude that fluoxetine is better-tolerated than either sertraline or paroxetine The annual hospital readmission rate was 50% A prospective randomised trial would need 192 patients per arm in order to detect a 15% reduction in readmission with 80% power
Hong <i>et al</i> (1998)	US patients with relapsing schizophrenia and high levels of health care resource use	Is a newer antipsychotic (quetiapine) associated with better compliance, and therefore lower rates of rehospitalisation, when compared with conventional treatment?	Those with schizophrenia ($n=1400$) selected from a commercially available insurance claims database, coupled with a Medicaid claims file, providing detailed health care costs and resource use on 5% of the 5 million Californian Medicaid population	Hospital readmission rates and the prevalence of high service utilisation were calculated A power calculation was used to design a prospective, randomised trial	Not applicable	

HRSD, Hamilton Rating Scale for Depression; ICD, International Classification of Diseases; SSRI, selective serotonin reuptake inhibitor; VA, Veterans Affairs.

use existing outcomes data, collected for other purposes.

The MOS is the best-known example of prospective outcomes research. The authors set out to measure patient-centred outcomes, in addition to clinician-rated depressive symptoms within existing health care services. The enduring legacy of the MOS is the fact that patient-centred measures of health status were developed for the study and eventually evolved into the Short Form 36 (SF36) (Stewart & Ware, 1992) – now the most commonly used generic measure of health-related quality of life.

A further study (Rosenheck *et al*, 2000) measured a number of outcomes, including disease-specific measures relating to the underlying condition (PTSD), measures of social function, health-related quality of life, and service use. This study used a large, existing data-set describing all of the 600 000 patients in receipt of mental health care under the US Veterans Affairs (National Committee on Quality Assurance, 1995). It was supplemented with routinely collected disease-specific patient outcomes measures collected for all patients in receipt of care for PTSD (Rosenheck, 1996).

All the other studies that we identified used existing outcomes already entered on large administrative databases, studying a much more limited range of outcomes. For example, studies examining the value of new antidepressant drugs in routine care settings used a commercially available medical insurance database of linked pharmacy and medical claims data on 750 000 individuals (Melfi *et al*, 1998; Croghan *et al*, 1999; Hylan *et al*, 1999). Cases of depression were identified retrospectively, either from a reimbursement claim for antidepressant medication or by the presence of one of six ICD codes indicative of depression (World Health Organization, 1992). This approach is hampered by the fact that antidepressant drugs are commonly prescribed for a number of conditions other than depression (Streator & Moss, 1997). Similarly, depression is consistently underidentified by clinicians (Jencks, 1985) and mislabelled or underreported, in part as a consequence of the stigma of mental illness (Rost *et al*, 1994).

Commercially available administrative databases also hold no direct information about disease severity, such as scores on symptom rating scales. Disease progression, relapse or remission cannot be directly measured, and database studies are forced

to use alternatives. For example, Hylan *et al* (1999) used continuous 6-month claims for refills of prescriptions as a proxy measure of acceptable pharmacotherapy and therefore good outcome, ignoring the fact that patients discontinue medications for a whole host of reasons other than treatment failure.

Sample size and length of follow-up

Sample size was generally much greater than that achieved in the traditional randomised trial, with a median sample size of 2678 (range 1034 to 20 814). Studies that recruited subjects prospectively, such as the MOS (Wells *et al*, 1989), achieved smaller sample sizes ($n=1772$) than those selecting subjects retrospectively from large, existing data-sets (Croghan *et al*, 1999; Rosenheck *et al*, 1999a) (median $n=4052$). Periods of follow-up were of median 6 months (range 4 to 48 months).

Adjustment for confounding and case mix

All studies made some attempt to describe and adjust for confounding factors, typically using some form of regression analysis or propensity scoring (Rubin, 1997). Authors rarely reported each of the potentially confounding factors that were entered into their analysis – often restricting reports to those that were positive and related to outcome. However, it was clear that the ability of studies to adjust for confounding was determined by the collection or availability of suitable measures. Two studies serve to illustrate the contrast between limited and more complete adjustment for confounding.

The authors of the MOS prospectively measured a broad range of case-mix variables, including disease severity and comorbidity, in addition to traditional demographic characteristics such as age, gender and socio-economic status. This is especially important in the MOS since the type of health care provider is inexorably linked to disease severity, making unadjusted comparisons of outcome impossible to interpret. One of the more unexpected results of the MOS demonstrates the limitation of an observational approach and the need to measure and adjust for case mix and confounding. In unadjusted samples, the receipt of any treatment (antidepressant medication or counselling) was associated with a much worse 2-year outcome than the receipt of

no treatment. In analyses that adjusted for baseline health differences, treated and untreated patients had a comparable 2-year outcome. In a subgroup analysis, designed to minimise unmeasured biases by restricting the analysis to those with the most severe depression, treatment was in fact associated with a significantly better 2-year outcome (Wells *et al*, 1996; Wells, 1999).

In contrast, outcomes studies based on administrative data are much more limited in their ability to measure and adjust for confounding. For example, in retrospective database studies of new antidepressant drugs (Melfi *et al*, 1998; Hylan *et al*, 1999) disease severity could not be measured since these data were not directly included in administrative data and could only be crudely inferred from the setting in which care was given (primary *v.* secondary care).

DISCUSSION

Despite the enthusiasm with which outcomes research was adopted and funded in the USA, by the 1990s its value was being called into question. The US Office of Health Technology Assessment (1994) offered a stinging appraisal: ‘Contrary to the expectations expressed in the legislation establishing the AHCPR . . . administrative databases have generally not proved useful in answering questions about the comparative effectiveness of alternative medical treatments.’ Clearly, the superficially appealing opportunity to generate large-scale studies from readily available and existing data sources should be approached with caution. This review highlights both the strengths and the limitations of outcomes research as a method for evaluating mental health services.

Strengths of outcomes research

The criticism is often made that randomised trials are undermined by the fact that the participants form a highly selected and homogeneous group, and their health care and follow-up are different from that received by the majority of patients (Anonymous, 1994). The consequence is that it is not always possible to apply the results in clinical practice – in other words, trials lack external validity (Naylor, 1995).

One potential advantage of outcomes research is that observational data are routinely collected for all patients and the results can therefore be applied more

generally. Further, data are generated in routine health-care services, rather than in artificially constructed trials. Lastly, outcomes research might be able to deliver answers to some questions quickly, cheaply and with greater statistical power, and without the need to seek ethical approval and individual patient consent, compared with the time-consuming and costly randomised trial. This review suggests that outcomes research in mental health has indeed realised these advantages – incorporating large numbers of subjects from real-life clinical populations and following them up for clinically meaningful periods of time.

Weaknesses of outcomes research

Ellwood's original vision of outcomes research required that a rich and clinically meaningful set of outcomes would be collected for all patients during their routine care (Ellwood, 1988). However, the feasibility and cost of such data collection has meant that the building blocks of much outcomes research (with notable exceptions) have been data that are collected as part of the administrative process (Iezzoni, 1997). These administrative data (produced by federal health providers, state governments and private insurers) contain the minimum amount of information required to fulfil an administrative function, particularly billing. They generally include little more than routine demographic data, ICD-9 diagnostic codes, details of interventions received during a hospital episode, length of stay and mortality during a hospital episode. The fundamental problem with research using these data is that the outcomes available are generally not those that we would like to study. Research becomes driven by the availability of data rather than by the need to answer specific questions, as acknowledged by one outcomes researcher: 'I utilise data that are available. I do not start with "what is the problem and what is the outcome?" I say, "given these data, what can I do with them?"' (Blumberg, 1991).

The other major problem with outcomes research, as with all observational research, is the problem of confounding and selection bias (Cook & Campbell, 1979; Iezzoni, 1997). The treatment that a patient receives will often be determined by a number of factors that are related to outcome, such as disease severity. Thus patients will differ in many ways other

than the treatment they receive, and it is therefore difficult to attribute any differences in outcome to the treatment itself (Green & Byar, 1984).

Our review suggests that, in mental health, large-scale studies using 'homogenous databases' are largely achieved at the expense of clinically meaningful outcomes and limited opportunities to adjust for confounding. Only two studies stand out as having collected a broad range of clinically important outcomes and case-mix variables, reflecting not just disease severity but the facets of service use and health-related quality of life – the MOS (Wells *et al*, 1989) and Rosenheck's study of PTSD (Rosenheck *et al*, 1999b).

Can outcomes research ever be useful in the UK?

Professor Nick Black has recently called for the establishment of large-scale, high-quality clinical databases across all disciplines in the UK (Black, 1999). The most ambitious example of this work in the UK has been in the field of intensive care (Rowan, 1994). According to Black, such databases need not be seen as an alternative to the randomised trial, but rather as a complement. The attractions for researchers include the possibility of generating large samples from many participating centres, and of including clinically important subgroups of patients who might be excluded from traditional trials. Outcomes research can also be used to promote rather than replace randomised trials in a number of ways. First, raising the level of uncertainty among clinicians as to the effectiveness of established interventions might increase clinicians' likelihood of participating in a randomised trial. Second, it could provide a permanent infrastructure for mounting multi-centre trials. Finally, the adoption of such databases means that research would no longer be the preserve of a minority of clinicians working in specialist centres, thus enhancing the generalisability of the results.

How feasible are such developments in mental health research in the UK?

The absence of a centralised administrative data-collection system in the UK has meant that the building blocks of outcomes research have never developed to the extent that they have in the USA. Initiatives to ensure that uniform outcomes are collected

for all patients, such as the Health of the Nation Outcome Scales (Wing, 1994), have been proposed but have not so far been adopted in routine practice (Slade *et al*, 1999). Consequently, the adoption of routine outcomes monitoring will entail substantial effort.

Research initiatives are under way; for example, the Centre for Outcomes Research and Effectiveness (CORE) has been established under the auspices of the British Psychological Society (Clifford, 1998) in order to generate 'practice-based evidence' of effectiveness framed within routine services (Marginson *et al*, 2000). At this juncture, it would be timely to learn from the examples of outcomes research in the USA, and to recognise the limitations and potential of the approach.

Rosenheck *et al* (1999b), who provided one of the more rigorous examples of outcomes research, outlined several ingredients of a successful clinical database, capable of producing rigorous and informative research. Outcomes databases should:

- (a) include large numbers of subjects
- (b) use standardised instruments that are appropriate for the clinical condition being treated
- (c) measure outcomes in multiple relevant domains
- (d) include extensive data in addition to outcomes measures, in order to support matching
- (e) collect data at standardised intervals after a sentinel event such as entry to or discharge from hospital
- (f) take aggressive steps to achieve the highest possible follow-up rates.

Data should also be collected prospectively if they are to meet these aims.

Such databases are going to require substantial time, effort and expense to establish, making outcomes research far from the quick and cheap research option that was envisaged. For example, the whole MOS cost US\$12 million, and the depression component cost about US\$4 million (Wells *et al*, 1996). Outcomes research requires resolution of the practical and ethical problems of using clinical data for study purposes, as highlighted in recent debates about the Data Protection Act, the European Human Rights Act and the Health and Social Care Bill (Al-Shahi & Warlow, 2000; Medical Research Council, 2000; Anderson, 2001; Kmietowicz, 2001).

The pharmaceutical industry is especially keen to use outcomes research to examine the effectiveness of its products. This review highlights the fact that, so far, outcomes studies conducted by the pharmaceutical industry have been generally of poor quality and do not adhere to the sensible recommendations outlined by Rosenheck *et al* (1999b). The use of this method has clear advantages for the pharmaceutical industry – particularly in terms of cost. In conducting such research, the industry can claim that expensive (pragmatic) randomised trials are no longer needed in order to examine clinical and economic effectiveness in routine care settings; neither will they have to provide and dispense the drugs for the many thousands of patients who are included in these studies. Informed consent and ethical approval may no longer be required, since treatment is as received, as part of usual care, and outcomes are those that are collected anyway. Large-scale outcomes studies that are currently in progress – such as the SOHO Study – will need to demonstrate that they are methodologically robust and that their results are believable.

Mental health researchers must give clear thought as to how outcomes databases should be constructed, how resources might be put in place, and to what extent informed consent is required for research conducted using these data. Outcomes research should not be seen as an alternative to randomised controlled trials, but rather as a complement. Clinicians do not generally like collected standardised data for each and every patient (Walter *et al*, 1996a,b; Slade *et al*, 1999). It would be unfortunate if outcomes research was simply to be regarded as a quick and flawed solution to the many political and clinical problems in mental health.

ACKNOWLEDGEMENTS

The authors are grateful to Kate Misso for performing all literature searches.

APPENDIX

Search terms

The following bibliographic databases were searched: Medline (1966–2000); EMBASE (1981–2000); Cinahl (1982–2000); PsycLit (to 2000). In addition, we hand-searched a number of key journals and scrutinised reference lists for additional studies; we contacted key authors in the field. Our search included the following terms.

“HEALTH STATUS-INDICATORS”; “OUTCOME-AND-PROCESS-ASSESSMENT-(HEALTH-CARE)” / all subheadings; “OUTCOME-ASSESSMENT-(HEALTH-CARE)” / all subheadings; (OUTCOME MEASURE*) in ti,ab; (HEALTH OUTCOME*) in ti,ab; (QUALITY OF LIFE) in ti,ab; MEASURE* in ti,ab; ASSESS* in ti,ab; (SCORE* or SCORING) in ti,ab; INDEX in ti,ab; “OUTCOMES-RESEARCH” / all subheadings; HEALTH OUTCOME* in ti,ab; SCALE* in ti,ab; MONITOR* in ti,ab; ASSESS* in ti,ab; OUTCOME* in ti,ab; explode “TREATMENT-OUTCOMES”; explode “PSYCHOLOGICAL-ASSESSMENT”; “QUALITY-OF-LIFE”; (OUTCOME* or PROCESS*) near3 ASSESSMENT*; HEALTH STATUS INDICATOR*; HEALTH STATUS; HEALTH OUTCOME* in ti,ab; QUALITY OF LIFE in ti,ab

REFERENCES

- Aday, L. A., Begley, C. E., Lairson, D. R. & Slater, C. H. (1998)** *Evaluating Healthcare System: Effectiveness, Efficiency and Equity* (2nd edn). Chicago: AHSR.
- Agency for Health Care Policy Research (1993)** *Depression in Primary Care*. Washington, DC: US Department of Health and Human Services.
- Al-Shahi, R. & Warlow, C. (2000)** Using patient identifiable data for observational research and audit: overprotection could damage the public interest. *BMJ*, **321**, 1031–1032.
- Anderson, C. (1994)** Measuring what works in health care. *Science*, **263**, 1080–1082.
- Anderson, R. (2001)** Undermining data privacy in health information: new powers to control patient information contribute nothing to health. *BMJ*, **322**, 442–443.
- Anonymous (1989)** Databases for healthcare outcomes. *Lancet*, **396**, 195–196.
- (1992) Cross design synthesis: a new strategy for studying medical outcomes. *Lancet*, **340**, 944–946.
- (1994) From research to practice. *Lancet*, **344**, 417–418.
- Barkham, M., Evans, C., Marginson, F., et al (1998)** The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Mental Health*, **7**, 35–47.
- Black, N. A. (1999)** High quality clinical databases: breaking down barriers. *Lancet*, **353**, 1205–1206.
- Blumberg, M. S. (1991)** Potentials and limitations of database research illustrated by the QMMPAMI Medicare mortality study. *Statistics in Medicine*, **10**, 637–646.
- Clifford, P. (1998)** M is for Outcome: the CORE outcomes initiative. *Journal of Mental Health*, **317**, 1167–1168.
- Cook, T. D. & Campbell, D. T. (1979)** *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Croghan, T., Melfi, C., Dobrez, D., et al (1999)** Effect of mental health speciality care on antidepressant length of therapy. *Medical Care*, **37** (suppl), AS20–23.
- Davies, H. T. O. & Crombie, I. K. (1997)** Interpreting health outcomes. *Journal of Evaluation in Clinical Practice*, **3**, 187–199.
- Donoghue, J., Tylee, A. & Wildgust, H. (1996)** Cross sectional database analysis of antidepressant prescribing in general practice in the United Kingdom, 1993–5. *BMJ*, **313**, 861–862.
- Ellwood, P. M. (1988)** Shattuck lecture – outcomes management. A technology of patient experience. *New England Journal of Medicine*, **318**, 1549–1556.
- General Accounting Office (1992)** *Cross Design Synthesis: A New Strategy for Medical Effectiveness Research*. Washington, DC: GAO.
- Gilbody, S. & Whitty, P. (2002)** Improving the delivery and organisation of mental health services: beyond the conventional randomised controlled trial. *British Journal of Psychiatry*, **180**, 101–103.
- Green, S. B. & Byar, D. P. (1984)** Using observational data from registries to compare treatments. *Statistics in Medicine*, **3**, 351–370.
- Guthrie, E. (2000)** Psychotherapy for patients with complex disorders and chronic symptoms: The need for a new research paradigm. *British Journal of Psychiatry*, **177**, 131–137.
- Hong, W. W., Rak, I. W., Ciuryla, V. T., et al (1998)** Medical-claims databases in the design of a health-outcomes comparison of quetiapine (‘Seroquel’) and usual-care antipsychotic medication. *Schizophrenia Research*, **32**, 51–58.
- Hotopf, M., Lewis, G. & Normand, C. (1997)** Putting trials on trial – the costs and consequences of small trials in depression: a systematic review of methodology. *Journal of Epidemiology and Community Health*, **51**, 354–358.
- , Churchill, R. & Lewis, G. (1999) Pragmatic randomised trials in psychiatry. *British Journal of Psychiatry*, **175**, 217–223.
- Hylan, T., Crown, W., Meneades, L., et al (1999)** SSRI antidepressant drug use patterns in the naturalistic setting: a multivariate analysis. *Medical Care*, **37** (suppl. 4), AS36–44.
- Iezzoni, L. I. (1997)** Assessing quality using administrative data. *Annals of Internal Medicine*, **127**, 666–674.
- Jencks, S. F. (1985)** The recognition of mental distress and diagnosis of mental disorder in primary care. *JAMA*, **253**, 1903–1907.
- Kmietowicz, Z. (2001)** Registries will have to apply for right to collect patients’ data without consent. *BMJ*, **322**, 1199.
- Lam, J. A. & Rosenheck, R. (1999)** Street outreach for homeless persons with serious mental illness: is it effective? *Medical Care*, **37**, 894–907.
- Leslie, D. L. & Rosenheck, R. A. (2000)** Comparing quality of mental health care for public sector and privately insured populations. *Psychiatric Services*, **51**, 650–655.
- Lilford, R. & Royston, G. (1998)** Decision analysis in the selection, design and application of clinical and health services research. *Journal of Health Services Research*, **3**, 159–166.
- Margison, F. R., Barkham, M., Evans, C., et al (2000)** Measurement and psychotherapy: evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, **177**, 123–130.
- Medical Research Council (2000)** *Personal Information in Medical Research*. London: MRC.
- Melfi, C., Chawla, A., Croghan, T., et al (1998)** The effects of adherence to antidepressant treatment guidelines on relapse and recurrence of depression. *Archives of General Psychiatry*, **55**, 1128–1132.
- Mellor-Clarke, J., Barkham, M., Connell, J., et al (1999)** Practice based evidence and the need for a

standardised evaluation system: informing the design of the CORE system. *European Journal of Psychotherapy, Counselling and Health*, **3**, 357–374.

National Committee on Quality Assurance (1995) *National Committee on Quality Assurance Report Card Pilot Project*. Washington, DC: NCQA.

Naylor, C. D. (1995) Grey zones of clinical practice: some limits to evidence based medicine. *Lancet*, **345**, 840–842.

Office of Health Technology Assessment [US Congress] (1994) *Identifying Health Technologies that Work: Searching for Evidence*. OTA-H-608. Washington, DC: US Government Printing Office.

Randolph, F. L., Blasinsky, M., Leginski, W., et al (1997) Creating integrated service systems for homeless persons with mental illness: the ACCESS programme. Access to Community Care and Effective Services and Supports. *Psychiatric Services*, **48**, 369–374.

Rosenheck, R. A. (1996) *Department of Veterans Affairs national mental health programme performance monitoring system: fiscal 1995 report*. West Haven, CT: Northeast Programme Evaluation Centre.

—, **Leda, C., Frisman, L., et al (1997)** Homeless mentally ill veterans: race, service use, and treatment outcomes. *American Journal of Orthopsychiatry*, **67**, 632–638.

—, **Druss, B., Stolar, M., et al (1999a)** Effect of declining mental health service use on employees of a large corporation. *Health Affairs*, **18**, 193–203.

—, **Fontana, A. & Stolar, M. (1999b)** Assessing quality of care: administrative indicators and clinical outcomes in post traumatic stress disorder. *Medical Care*, **37**, 180–188.

—, **Stolar, M. & Fontana, A. (2000)** Outcomes monitoring and the testing of new psychiatric treatments: work therapy in the treatment of chronic post-traumatic stress disorder. *Health Services Research*, **35**, 133–152.

Rost, K., Smith, G. R., Matthews, D. B., et al (1994) The deliberate misdiagnosis of major depression in primary care. *Archives of Family Medicine*, **3**, 333–342.

Rowan, K. M. (1994) Intensive Care National Audit and Research Centre: past, present and future. *Care of the Critically Ill*, **10**, 148–149.

Rubin, D. B. (1997) Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, **127**, 757–763.

Sheldon, T. A. (1994) Please bypass the PORT. *BMJ*, **309**, 142–143.

Slade, M., Thornicroft, G. & Glover, G. S. O. (1999) The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology*, **34**, 243–249.

Smith, D. M. (1997) Database research: is happiness a humongous database? *Annals of Internal Medicine*, **127**, 725–756.

Stewart, A. L. & Ware, J. E. (eds) (1992) *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press.

Streator, S. E. & Moss, J. T. (1997) Non approved usage identification of off label anti-depressant use and cost in a network model HMO. *Drug Benefit Trends*, **9**, 42–47.

Tarlov, A. R., Ware, J. E., Greenfield, S., et al (1989) The Medical Outcomes Study. An application of

CLINICAL IMPLICATIONS

- Robust evidence is needed of the effectiveness of new and existing treatments, interventions and policy initiatives in mental health.
- Randomised trials have formed the 'gold standard' of this evidence but are subject to many limitations.
- Outcomes research has the potential to provide 'real world' evidence of clinical and economic effectiveness, relatively quickly and cheaply, using routinely collected data from clinical services.

LIMITATIONS

- Outcomes research uses an observational design and is subject to many limitations – principally bias and confounding.
- The quality of the data upon which outcomes research is based is often poor.
- Successful outcomes research depends upon the routine collection of diverse and clinically meaningful outcomes, which requires substantial effort and cost.

SIMON M. GILBODY, DPhil, ALLAN O. HOUSE, MRCPsych, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds; TREVOR A. SHELDON, DSc, Department of Health Studies, University of York, UK

Correspondence: Dr Simon M. Gilbody, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds, LS2 9LT, UK. Tel: +44 (0)113 233 1899; fax: +44 (0)113 243 3719; e-mail: s.m.gilbody@leeds.ac.uk

(First received 14 February 2001, final revision 14 June 2001, accepted 14 June 2001)

methods for monitoring the results of medical care. *JAMA*, **262**, 925–930.

Thier, S. O. (1992) Forces motivating the use of health status assessment measures in clinical settings and related clinical research. *Medical Care*, **30** (suppl. 5), M515–22.

Thornicroft, G., Strathdee, G., Phelan, M., et al (1998) Rationale and design. PRISM psychosis study I. *British Journal of Psychiatry*, **173**, 363–370.

Thornley, B. & Adams, C. E. (1998) Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*, **317**, 1181–1184.

Walter, G., Cleary, M. & Rej, J. (1996a) Attitudes of mental health personnel towards rating outcome. *Journal of Quality Clinical Practice*, **18**, 109–115.

—, **Kirkby, K. & Marks, I. (1996b)** Outcome measurement: sharing experiences in Australia. *Australasian Psychiatry*, **4**, 316–318.

Wells, K. B. (1999) Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry*, **156**, 5–10.

—, **Stewart, A., Hays, R. D., et al (1989)** The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *JAMA*, **262**, 914–919.

—, **Sturm, R., Sherbourne, C. D., et al (1996)** *Caring for Depression*. Cambridge: Harvard University Press.

Wennberg, J. E. (1990) Outcomes research, cost containment, and the fear of health care rationing. *New England Journal of Medicine*, **323**, 1202–1204.

— (1991) What is outcomes research? In *Health Services Research: Key to Health Policy* (ed. E. Ginzberg), pp. 33–46. Cambridge: Harvard University Press.

—, **Barry, M. J., Fowler, F. J., et al (1993)** Outcomes research, PORTs, and health care reform. *Annals of the New York Academy of Sciences*, **703**, 52–62.

Wing, J. (1994) Measuring mental health outcomes: a perspective from the Royal College of Psychiatrists. In *Outcomes into Clinical Practice* (ed. T. Delamonte), pp. 147–153. London: BMJ Publishing.

World Health Organization (1991) *Evaluation of Methods for the Treatment of Mental Disorders*. Geneva: WHO.

— (1992) *The Tenth Revision of the International Classification of Diseases and Related Health Problems*. Geneva: WHO.