The Journal of Mental Health Policy and Economics
J. Mental Health Policy Econ. **2**, 99–106 (1999)

Rational Decision-Making in Mental Health: the role of systematic reviews

Simon M Gilbody[1]* and Mark Petticrew[2]†

[1]*NHS Centre for Reviews and Dissemination, University of York YO10 5DD, UK*
[2]*MRC Medical Sociology Unit, Glasgow, UK*

Abstract

Background: 'Systematic reviews' have come to be recognized as the most rigorous method of summarizing confusing and often contradictory primary research in a transparent and reproducible manner. Their greatest impact has been in the summarization of epidemiological literature—particularly that relating to clinical effectiveness. Systematic reviews also have a potential to inform rational decision-making in healthcare policy and to form a component of economic evaluation.
Aims of the study: This article aims to introduce the rationale behind systematic reviews and, using examples from mental health, to introduce the strengths and limitations of systematic reviews, particularly in informing mental health policy and economic evaluation.
Methods: Examples are selected from recent controversies surrounding the introduction of new psychiatric drugs (anti-depressants and anti-schizophrenia drugs) and methods of delivering psychiatric care in the community (case management and assertive community treatment). The potential for systematic reviews to (i) produce best estimates of clinical efficacy and effectiveness, (ii) aid economic evaluation and policy decision-making and (iii) highlight gaps in the primary research knowledge base are discussed. Lastly examples are selected from outside mental health to show how systematic reviews have a potential to be explicitly used in economic and health policy evaluation.
Results: Systematic reviews produce the best estimates of clinical efficacy, which can form an important component of economic evaluation. Importantly, serious methodological flaws and areas of uncertainty in the primary research literature are identified within an explicit framework. Summary indices of clinical effectiveness can be produced, but it is difficult to produce such summary indices of cost effectiveness by pooling economic data from primary studies. Modelling is commonly used in economic and policy evaluation. Here, systematic reviews can provide the best estimates of effectiveness and, importantly, highlight areas of uncertainty that can be used in 'sensitivity analysis'.
Discussion: Systematic reviews are an important recent methodological advance, the potential for which has only begun to be realized in mental health. This use of systematic reviews is probably most advanced in producing critical summaries of clinical effectiveness data. Systematic reviews cannot produce valid and believable conclusions when the primary research literature is of poor quality. An important function of systematic reviews will be in highlighting this poor quality research which is of little use in mental health decision making.
Implications for health provision: Health care provision should be both clinically and cost effective. Systematic reviews are a key component in ensuring that this goal is achieved.
Implications for health policies: Systematic reviews have potential to inform health policy. Examples presented show that health policy is often made without due consideration of the research evidence. Systematic reviews can provide robust and believable answers, which can help inform rational decision-making. Importantly, systematic reviews can highlight the need for important primary research and can inform the design of this research such that it provides answers that will help in forming healthcare policy.
Implications for further research: Systematic reviews should precede costly (and often unnecessary) primary research. Many areas of health policy and practice have yet to be evaluated using systematic review methodology. Methods for the summarization of economic data are methodologically complex and deserve further research.

Received 15 February 1999; accepted 1 September 1999.

Introduction

Recent years have seen major changes in the way in which decisions about healthcare are made. In particular, it is now seen as necessary that decisions about the provision of healthcare are 'evidence based', and systematic literature reviews play an increasingly large part in this process, as they represent the best evidence when assessing the effectiveness of an intervention.[1] This paper highlights the contribution of systematic reviews toward rational decision making in mental health.

The particular strength of systematic reviews is their ability to summarize a large body of literature in a critical and replicable fashion, and, in the case of meta-analysis, the ability to increase their power to detect small differences in effectiveness, and to increase the precision of their estimates, by pooling data from many studies. As well as summarizing evidence about clinical effectiveness, systematic reviews also have considerable potential to inform economic and policy evaluations, but there are also limitations. This

*Correspondence to: Dr. Simon M Gilbody, MBChB MMedSc MRCPsych, MRC Fellow in Health Services Research, NHS Centre for Reviews and Dissemination, University of York, YO10 5DD, UK.
Tel: +44 (0)1904 433710.
Fax: +44 (0)1904 433661
Email address: smg5@york.ac.uk
†Formerly Research Fellow, NHS CRD, University of York.
Sources of funding and conflict of interest: The authors are directly funded by the UK Medical Research Council Fellowship Scheme (SG) and the University of York (MP, at the time of writing), and have contributed to the work of the Cochrane Collaboration and the Database of Abstracts of Reviews of Effectiveness (DARE), which are discussed within this article.

CCC 1091–4358/99/030099–08$17.50

paper summarizes their strengths and weaknesses, using examples from several areas of mental health practice and policy, and particularly concentrating on the difficulties in summarizing economic data. Examples of systematic reviews from other areas of healthcare will also be presented, to illustrate how systematic reviews have been explicitly adopted as tools for economic evaluation.

Systematic Reviews in Healthcare

It is now recognized that, in all areas of healthcare, practice and policy should be guided by the highest quality research evidence.[2] In the field of mental health, as in all healthcare, evidence of the effectiveness of interventions is often contradictory, partly because of differences between the studies in terms of methodological rigour, patient populations and interventions. In order to make sense of this disparate and often contradictory literature, practitioners, policy-makers and consumers of healthcare have relied on traditional 'review' articles, which are generally prepared by 'content experts' in a field. Unfortunately, such reviews have been shown to be prone to a number of biases and their conclusions can be just as contradictory as the primary research.[3] For example, content experts may come to a particular field with their own pre-formed opinions and there is a risk that the primary research will be plundered selectively in order to confirm the author's pre-formed opinion (a 'confirmatory bias'), leaving the reader unclear as to how the primary studies have been selected for inclusion or how a particular conclusion has been reached. In the face of growing dissatisfaction with the lack of transparency of methodology and lack of trust in the conclusions of traditional review articles by readers, the *systematic review* article has emerged.

Systematic reviews adopt an explicit methodology in order to limit bias in the search, and selection of studies for review. This takes the form of extensive (including electronic) literature searches, followed by selection of the highest quality studies for review—ideally these should be randomized-controlled trials (RCTs), where these are available or feasible. This evidence is (where appropriate) synthesized in order to produce a clear message or conclusion regarding effectiveness. It may be summarized narratively, or in some cases it is possible to summarize the results of the primary studies quantitatively, in the form of a meta-analysis.[4]

The use of such methods in mental health has a relatively long history. Smith and Glass[5] pioneered the use of meta-analysis in the 1970s in order to synthesize disparate, contradictory and under-powered studies of the effectiveness of psychotherapy. From this early start the methods of systematic review and meta-analysis evolved rapidly, and have been employed to evaluate the effectiveness of a range of interventions in mental health.

Many journals now publish systematic reviews in preference to traditional narrative reviews. Unfortunately the increased acknowledgement of the validity and rigour of systematic reviews has meant that the term has come to be abused. Many authors claim that their reviews are 'systematic', whilst the content and methodology shows them to be far from so, and the reader should critically appraise any review claiming to be systematic, just as they should with any other piece of research.[6] A simple checklist of questions that should guide the critical interpretation of review articles is given in **Table 1**. Two major sources of high quality systematic reviews are the *Cochrane Database of Systematic Reviews* available on CRD-ROM, and the *Database of Abstracts of Reviews of Evaluations* (*DARE*), maintained by the NHS Centre for Reviews and Dissemination at the University of York.[7]

Systematic Reviews in Mental Health Decision Making

It is clear that systematic reviews have the potential to provide important inputs into decision making in mental health practice and policy, and several practical examples are given below. These examples highlight their role in decision-making where the alternatives differ in cost or where there is genuine uncertainty regarding the clinical effectiveness of an intervention. They are chosen to illustrate the diversity of interventions that have been evaluated through systematic review, including pharmacological interventions, psychosocial interventions and innovative modes of delivery of healthcare. It will be clear from these examples that, in mental health, systematic reviews rarely provide the central aspect of any economic or policy evaluation. However, they do help refine an economic discussion and, it will be argued, produce the most believable estimate of clinical effectiveness when this is required as a component of economic evaluation. Moreover, systematic reviews help us to define the limits of our knowledge and help avoid spurious certainty regarding the clinical effectiveness of an intervention.

The Routine Use of New Anti-depressants

The introduction of the serotonin specific reuptake inhibitors (SSRIs) in the late 1980s for the treatment of major depression was heralded as a major innovation. Purported

Table 1. Questions to guide the critical appraisal of a systematic review

Are the results valid, believable and relevant?
Is the research question stated and clearly focused?
Are the inclusion criteria stated explicitly and are they appropriate?
Is the search for relevant studies thorough?
Is the validity of the individual studies assessed?
Were all important outcomes assessed?
Were the results similar from study to study?
Were reasons for differences between studies explored?
What are the overall results of the systematic review?
How precise were the results?
Are the patients and interventions generalizable to real world practice?

Adapted from Oxman *et al.*[48].

improvements in tolerability and acceptability over older (tricyclic) antidepressants were alone felt to justify their routine prescription, despite their much greater unit cost.[8] In the UK, in the absence of any regulatory framework to demand evidence of improved clinical and cost effectiveness, SSRIs have achieved market dominance in a relatively short period of time.[9] It was estimated in the mid-1990s that a wholesale shift in prescribing policy from old to new anti-depressants would increase the direct prescription costs to the UK NHS from £88 million to £250 million.[10] The debate surrounding the relative clinical and cost effectiveness of new versus old anti-depressants continues, and systematic reviews have played an important role in the evolution of this debate in the UK.

In excess of 100 comparative randomized controlled trials (RCTs) of old versus new anti-depressants (4600 individual patients) have now been identified and the question of which is superior will, in part, vary between individual studies. These studies are generally statistically under-powered (median $n = 60$) and therefore lack the ability to detect any real difference between the two classes of compounds.[11] The potential for systematic review to bring some order to this confusing picture was first recognised by Song and colleagues.[12] When studies were pooled, the two classes of drugs were found to be broadly similar in efficacy and tolerability. The reaction to this finding was generally hostile. The general technique of systematic review and meta-analysis was criticized and in particular the pharmaceutical industry criticized the authors for failing to include unpublished data that was held as 'data on file' and was therefore outside the public domain.[13,14] Despite subsequent reviews that have tried to demonstrate the superiority of SSRIs by choosing different inclusion criteria and conducting a series of subgroup analyses,[e.g. 15] the conclusions of Song *et al.*[12] have been validated and replicated in more methodologically sound, comprehensive and up to date reviews.[16,17]

Modelling has been employed to estimate the cost effectiveness of these drugs, since there are few prospective economic evaluations, conducted alongside rigorous RCTs. One of the most widely quoted models is that by Jönnson and Bebbington.[18] This study demonstrates the cost effectiveness of an SSRI (paroxetine) compared to imipramine (a first generation tricyclic). However, the cost effectiveness of the SSRIs in this model was highly dependent upon high discontinuation rates with imipramine. The primary efficacy data used in this model is a pooled analysis of only six RCTs conducted in the US,[19] which demonstrates differences in patient discontinuation rates that are much larger than those seen in more comprehensive meta-analyses. Woods and Rizzo[20] have reassessed the cost effectiveness of this drug using more precise and representative estimates of efficacy obtained from rigorous meta-analyses, and have reached different conclusions from Jönnson and Bebbington. Several models have now been produced, each of which makes slightly different assumptions and adopts differing perspectives and methods of estimating costs. Some have also employed meta-analysis to produce more precise estimates of efficacy.[e.g. 21–23] Stewart[24] has reviewed the design issues surrounding the evaluation of cost effectiveness of these drugs in detail.

The process of systematic review has been integral both in demonstrating the similar clinical efficacy of old and new drugs and in highlighting many of the methodological problems associated with the available clinical efficacy data. The primary research in this area is not only under-powered, but also methodologically poor.[11] Most studies are short term (median 6 weeks) and compare new anti-depressants with first generation tricyclic drugs, i.e. those with the least acceptable side effects. Clinical outcomes fail to include aspects of quality of life, and important aspects of service use (such as readmission and relapse) are not recorded.

While the relative clinical and cost effectiveness of new and old anti-depressants will probably not be determined by anything other than a well designed, adequately powered RCT with a concurrent economic evaluation, systematic reviews in this area have aided in identifying the dearth of adequate clinical and economic evidence and the questions which future researchers need to address.

New Drugs in the Treatment of Schizophrenia

New anti-schizophrenia drugs ('atypical' anti-psychotics) have a much higher acquisition cost, but are claimed to have a clinical advantage over older drugs in terms of their side-effect profile (less sedation and fewer movement disorders), and an ability to treat negative symptoms (apathy and social withdrawal). Further, some drugs (e.g. clozapine) are claimed to be effective in treating patients who have failed to improve with conventional drugs—so called 'treatment resistant' schizophrenia.[see 25 for review] In the UK, direct drug costs constitute 5% of total care costs, with hospital and social care making up the greatest part of the remainder.[26] The replacement of old drugs with new could potentially improve compliance and reduce service use, thus proving to be cost effective. Such a hypothesis is clearly testable by rigorous economic evaluation, the validity of which will necessarily depend upon good quality clinical data.

Economic evaluations that have been carried out in this area have been limited by the quality of the primary clinical effectiveness data.[27] For example, some authors employ 'mirror' (before and after) designs[e.g. 28] or use 'historical' controls to generate comparative clinical data.[e.g. 29] Such studies are open to numerous biases and are potentially misleading.[30] 'Gold standard' clinical and economic evaluations, i.e. prospective economic analyses alongside well-powered RCTs, are not yet available for many of the newer drugs, although the methodologies of important ongoing studies are reported.[31,32]

In the absence of prospective economic data from RCTs, modelling has played an important part in the economic evaluation of new anti-psychotics. One of the most influential, well constructed and widely quoted models is that of Davies and Drummond,[33] which incorporates (amongst other things) primary efficacy data from RCTs comparing clozapine and haloperidol. The cost effectiveness of clozapine over haloperidol

amongst patients with 'treatment resistant' schizophrenia is demonstrated using this model. However, the validity of the clinical effectiveness data in this and other economic models has been questioned (see Maynard and Bloor[27] for a review). It may therefore be instructive to see what conclusions have been reached when the body of clinical effectiveness research is subjected to the process of systematic review.

A number of systematic reviews have been completed under the auspices of the Cochrane Schizophrenia Group[34]. An example of an abstract from one such review, comparing risperidone with older drugs,[35] is given in **Table 2**. This 'Cochrane' review reveals many limitations of the primary research evidence. The RCTs are generally under-powered and methodologically poor. Most studies make short-term comparisons and fail to record important aspects of service use or 'quality of life'. The chosen comparator drug is generally haloperidol, one of the oldest antipsychotics with the most distressing side effects. Haloperidol does not represent a commonly used treatment alternative in many healthcare systems, so this is not the treatment which risperidone would replace if introduced. Studies are generally conducted on an in-patient basis, which make the results difficult to apply in community settings, where most treatment for schizophrenia takes place. Moreover, risperidone is shown to have *statistically* improved compliance and side effect profile. However the absolute size of this superiority is of questionable *clinical* and *economic* significance. There are no economic data collected or reported in the RCTs included in the review (since none are available) and the process of systematic review allows us only to speculate about the economic meaning of these results.

It is clear therefore, that the most persuasive arguments for the cost effectiveness of atypical anti-psychotics have been put forward using modelling techniques. Modelling is a necessary technique under certain circumstances—particularly early in the life of a new technology and in the absence of prospective economic data or in the absence of studies of longer-term follow-up, when a realistic economic evaluation requires this.[36] Clearly, this is the case with many of the new anti-psychotics. However modelling also has its dangers, in particular the risk of creating spurious certainty and obscuring ignorance when the primary clinical and economic data included in the model are of poor quality.[30] A more recent model has explicitly used estimates of clinical effectiveness from meta-analysis in estimating the cost effectiveness of newer drugs.[37] However, whilst meta-analysis will provide the most statistically precise estimate of effectiveness, it will not remove many of the biases in the primary studies outlined above.

Systematic reviews have not played a large part in the debate surrounding the clinical and cost effectiveness of atypical anti-psychotics, but the production of rigorous systematic reviews, such as that outlined above, again alerts us to the fact that the primary clinical data is poor. In many cases we know very little about the 'real world' effectiveness of many of the new anti-psychotics. Much of the clinical data is limited to short-term studies which make unhelpful comparisons of new anti-psychotics with very old drugs,

Table 2. Cochrane Review of Risperidone versus 'conventional' antipsychotic medication for schizophrenia[35]

Objectives: To evaluate the effectiveness of risperidone for schizophrenia in comparison to 'conventional' neuroleptic drugs.
Search strategy: Electronic searches of Biological Abstracts, Cochrane Schizophrenia Group's Register, The Cochrane Library, EMBASE, MEDLINE, PsycLIT, and SCISEARCH were undertaken. References of all identified studies were searched for further trial citations. Pharmaceutical companies and authors of trials were contacted.
Selection criteria: All randomized trials comparing risperidone to any 'conventional' neuroleptic treatment for those with schizophrenia or other serious mental illnesses.
Data collection and analysis: Citations and, where possible, abstracts were independently inspected by reviewers, papers ordered, re-inspected and quality assessed. Data were also independently extracted. Sensitivity analyses on dose of risperidone, haloperidol and duration of illness were undertaken for the primary outcomes of clinical improvement, side effects (movement disorders) and acceptability of treatment. For homogeneous dichotomous data the Odds Ratio (OR), 95% Confidence Interval (CI) and, where appropriate, the Number Needed to Treat (NNT) were calculated on an intention-to-treat basis.
Main results: Twelve short-term studies provided data on 2533 people. This review provides no evidence relating to the effect of risperidone on cognitive or social functioning, quality of life, employment status, discharge from hospital and relapse rates. Risperidone increases the odds of moderate clinical improvement (OR 0.76, CI 0.63–0.93, NNT = 17). It appears to have little or no additional effect on the positive and negative symptoms of schizophrenia but did have less tendency to cause movement disorders, largely in comparison with haloperidol (OR 0.43, CI 0.34–0.55, NNT = 4.5 for use of antiparkinsonian medication). Risperidone seems to be more acceptable to those with schizophrenia (OR 0.73, CI 0.59–0.90, NNT = 20, 30% baseline risk of dropping out). Those taking risperidone are also marginally less likely to experience somnolence (OR 0.78, CI 0.61–0.99, NNT = 100). Weight gain is however more likely with risperidone (OR 1.51, CI 1.14–2.00, NNT = 10). Funnel plots show that smaller studies generally show greater benefit for risperidone than larger studies. A publication bias in favour of risperidone amongst the included studies may explain this effect. Sensitivity analyses on dose of risperidone (excluding those receiving 1 or 2 mg) did not materially change the results for the principal outcomes. Excluding data from those on higher doses of haloperidol (>10 mg/day) does marginally change the results. Risperidone is less effective in achieving clinical improvement and preventing dropout but outcomes relating to movement disorders change little.
Conclusions: Little can be concluded about the long term effects of risperidone and generalizing results beyond a comparison with haloperidol would be imprudent. Risperidone may be more acceptable to those with schizophrenia and have marginal benefits in terms of limited clinical improvement and side effect profile compared to haloperidol. The superiority of risperidone in these respects may have been overestimated due a possible publication bias in favour of risperidone. Any marginal benefit has to be balanced against the greater cost of the drug and its increased tendency to cause other side effects such as weight gain. Long term, well conducted and reported trials are needed. Further studies are required to establish the cause of the observed funnel plot asymmetry—particularly the presence of publication bias.

given in doses and by a route which is not typical of routine care in many countries. Even given these limitations, it is clear that the superiority of new anti-psychotics over older drugs is relatively small where systematic reviews have been carried out. Rigorous prospective economic evaluations, alongside well designed clinical trials, which are representative of routine care are needed.

Assertive Community Treatment and Case Management for Severe Mental Disorders

In the United Kingdom, as in many healthcare systems, community care for the those with severe mental illness has been perceived by some as a failure, and concerns have been raised that mental health policy has been formulated and implemented without any explicit consideration of cost and clinical effectiveness.[38] Two systematic reviews by Marshall and colleagues, published and updated in the *Cochrane Library*, evaluate the clinical effectiveness of two important alternative models of community care—case management (CM) and assertive community treatment (ACT).[39,40] These two reviews illustrate the potential for systematic reviews to produce valid estimates of clinical effectiveness to help inform policy decisions.

ACT was introduced in the early 1970s to address the difficulties of caring for severely mentally ill patients in the community. A key component of ACT is its team-based approach. Typically, a multi-disciplinary team, which includes social workers, nurses and psychiatrists, shares exclusive responsibility for a defined set of patients. The team seeks to meet all the social and psychiatric needs of their clients, rather than referring to outside agencies. In contrast to this, case management consists of community care given by a single autonomous individual with an exclusive 'caseload' of patients with severe mental illness. Although superficially similar to ACT, case management is globally the most widely practised and the cheapest (in terms of direct costs).[40]

The systematic review of the effectiveness of case management pools data from the relevant RCTs and shows this intervention to have two main effects. Firstly, it ensures that patients with severe mental illness are more likely to remain in contact with psychiatric services than those who receive 'standard care'. Secondly, patients in receipt of case management are more likely to be readmitted to hospital and hospital length of stay is likely to be longer. There is no evidence that case management improves the clinical or social outcome of severe mental illness. By way of contrast, the systematic review of ACT shows that hospital admissions are fewer and shorter than for those receiving both standard community care and hospital based rehabilitation services. Unfortunately, there are few studies that allow direct comparison between ACT and case management, but the two reviews taken together suggest the clinical superiority of ACT over case management.

Both these reviews are important in that they provide compelling evidence for the relative efficacy of important alternative methods of the delivery of mental health care, which has the potential to inform mental healthcare organization and policy. The direct costs of ACT are likely to be significantly higher than standard forms of community care. However, the reduced relapse rate and reduced service could potentially offset these higher direct costs, making ACT a cost effective policy initiative. In contrast to the systematic reviews outlined in the previous examples, many of the primary studies of case management and ACT incorporate both an economic evaluation and a clinical evaluation. Pooling of the clinical data is a logistically complex process, but is achievable and summary statistics are provided by the reviewers to quantify the degree of benefit and level of certainty surrounding the pooled estimates of clinical effectiveness. For example, when trials comparing ACT versus standard care are pooled, it becomes clear that ten patients need to receive ACT to prevent one hospital admission during the study period (OR 0.59, 99% CI 0.41–0.85, NNT = 10). However, despite the relatively high proportion of studies reporting economic data, the authors are unable to provide such clear and precise summaries of the cost effectiveness of ACT.

Although clinical and service use endpoints are reported and summarized in this systematic review, difficulties arise when the authors attempt to synthesize the economic data that are analysed in the primary studies. The authors of the review try to extract data relating to three aspects of costs from the primary studies:

(a) *costs of psychiatric in-patient care*,
(b) *costs of all health care* (including the above plus the costs of all other medical and psychiatric care such as out-patient care and assertive community treatment) and,
(c) *total costs* (including types of cost above plus the costs of accommodation and transfer payments and minus benefits, such as earnings).

In the case of studies comparing ACT with standard care, only five of the 14 studies provide cost data. Where cost data are reported, these relate to 'cost of inpatient care' and 'costs of all health care'. No studies report 'total costs', as defined above. The authors conclude that the interpretation of the cost data from primary studies is difficult, since statistical difference was either not reported or was subject to incorrect statistical analysis by the authors of primary studies (generally by the application of parametric tests to skewed data). Despite these limitations of the data, the authors provide tentative conclusions that the higher direct costs of ACT might be offset through reduced service use. More complex or rigorous economic evaluation is not possible from this limited data set and summary economic indices (such as incremental cost effectiveness ratios) cannot be produced in the same way as summary estimates of clinical effectiveness—such as 'numbers needed to treat'.[41]

Evaluating Cost Effectiveness From Systematic Reviews

It should be clear from the preceding discussion that systematic reviews have a great deal to contribute to the

evaluation of both the clinical and cost effectiveness of a wide variety of interventions in mental health. However, the greatest strengths of systematic reviews lie in providing the most precise and robust estimates of clinical effectiveness and in highlighting gaps in the epidemiological knowledge base. The contribution to estimates of cost effectiveness is less direct and reflects the epidemiological bias of the primary research literature. No meaningful economic evaluation can be made in the absence of valid estimates of clinical effectiveness and this is perhaps where the greatest potential for systematic reviews in informing mental healthcare practice and policy lies. However, it is instructive to look at a scenario, from outside mental health, where systemic reviews have been used explicitly to estimate cost effectiveness. These have used two approaches: the first appropriates the methods of systematic review to appraise and synthesize economic evaluations[42] and the second uses a rigorous systematic review of clinical effectiveness data (meta-analysis) as the basis of a rigorous economic model.[43]

In the first of these, Morris *et al.*[42] undertook a systematic review of the cost effectiveness studies of all cholesterol lowering interventions. Economic data were extracted from included primary studies in a standardized way using a checklist (see **Table 3**) and standardization between currencies and over time was carried out. A rigorous search of the literature identified 38 relevant studies. These used various economic methodologies and reported outcome in various ways, including a variety of incremental cost effectiveness ratios such as cost per life year saved, cost per QALY, cost per percentage reduction in cholesterol and cost per coronary heart disease case prevented. All of the individual economic evaluations based their economic data on single clinical studies, with the majority using modelling techniques based upon already published epidemiological literature. A variety of costing methods were reported. Generally, individual economic evaluations collected and reported direct costs, taking the perspective of payers of services and only counting drug costs, whilst ignoring the cost of managing the side effects of treatment. Costs avoided through effective treatment were rarely reported or measured in an explicit way.

Table 3. Data extracted by Morris *et al.*[42] from primary cost-effectiveness studies (originally adapted from[47])

Author(s)
Year of publication
Year used for cost valuation
Country where analysis occurred
Currency used for cost valuation
Alternatives considered
Cost effectiveness measure
Patient population
Effectiveness data sources
Cost elements
Cost data sources
Time horizon
Discount rate
Variables considered in the sensitivity analysis
Baseline results
Results from sensitivity analysis
Author(s) conclusions

Given such a disparate primary data set, Morris *et al.*[42] generally limited their analysis to commenting on the results of the individual studies, drawing attention to their heterogeneity of methods and results. Importantly, no attempt was made to categorize the methodological validity of the primary economic studies with reference to the quality of the primary clinical effectiveness data, for example by using RCTs in preference to observational studies. Rather than providing some summary (pooled) cost effectiveness ratio, the authors conclude by making specific suggestions regarding the robustness of the economic models used and by making specific recommendations regarding the variables that might be usefully included as a sensitivity analyses to produce more believable economic evaluations.

In the second example by comparison, a rigorous systematic review of different approaches to cholesterol reduction and coronary heart disease (CHD) prevention is used as the basis of an economic model to the estimate the cost effectiveness of these interventions.[43] The authors of the economic model combined their best estimates of clinical effectiveness from meta-analysis with lifetable actuarial data, to estimate 'life years gained'. Total costs were generally estimated from a societal perspective, and included estimates of direct costs and health service savings on avoided admissions and treatments such as coronary bypasses. Summary clinical effectiveness statistics were presented as 'numbers needed to treat' for five years in order to avoid one CHD event. Similarly, summary cost effectiveness statistics were presented as 'costs per life year gained' together with their 95% confidence intervals. This systematic review allows the reader to directly compare various approaches more readily and gives clear guidance regarding the level of uncertainty surrounding an individual estimate.

These two examples demonstrate two different approaches to the estimation of cost effectiveness, each of which explicitly adopts the methodology of systematic reviews. The first example highlights many of the methodological shortcomings of existing economic evaluations. In contrast, the second approach is able to use high quality epidemiological data from meta-analysis and high quality costing data specific to a healthcare system, to produce directly comparable incremental cost effectiveness ratios. No one approach is inherently superior and each has its strengths and weaknesses. Both reviews demonstrate that primary data rarely provide a clear answer regarding clinical and cost effectiveness. By introducing a systematic review methodology, a confusing research literature is summarized in an explicit way and a summary statistic is obtained that can be used to inform rational decision-making.

The Limitations of Systematic Reviews

The results of systematic reviews will only be as good as the primary data that are included. Where the internal validity of the study providing the primary data is low, then

we cannot expect systematic reviews to overcome this limitation, though they can help highlight it. One major strength of systematic reviews is an ability to produce power and precision where the primary data set does not show this. However, the uncritical use of techniques associated with systematic review, such as meta-analysis, will produce precise but nonetheless biased estimates of clinical effectiveness. Bias can be introduced from several sources, other than from the poor methodological quality of the primary studies. 'Publication bias' is one such example, where selective submission by authors of primary studies with positive findings and the non-publication of those with negative results can lead to spuriously optimistic estimates of clinical effectiveness.[44] Techniques for identifying such biases such as 'funnel plots' exist, and have shown their use in the systematic review comparing risperidone with older anti-psychotics outlined above (**Table 2**).

The examples outlined above highlight the difficulties in extracting clear estimates of cost effectiveness from the primary data. This should not surprise us for a number of reasons. Firstly, the quality of much of the primary economic data has been shown to be poor in many important areas of clinical practice and policy.[45] Secondly, heterogeneity in the methods by which cost effectiveness is measured, such as differences in evaluative designs, perspectives and methods of obtaining costs, make it difficult to combine individual studies in any sensible way. Thirdly, many economic evaluations are specific to a country or healthcare system and do not present disaggregated data or units of cost and resource use in sufficient detail to allow standardization. The methodologies of systematically reviewing and pooling cost data are not well advanced and this is an area that deserves further research.[46]

It is likely that, as the use of systematic reviews in mental health becomes more widespread, the limitations of the primary data set will also become more apparent. There are more questions regarding clinical practice and policy than there are systematic reviews to answer them. Unfortunately, the extent of this mismatch is not the same in all areas of practice and policy, and it often reflects the inherent biases of the primary research literature. Many systematic reviews in the area of mental health, such as those included in DARE and the Cochrane Library have been carried out on the effectiveness of pharmacological interventions rather than behavioural, organizational or psychological interventions. This reflects the biases in the primary research literature. RCTs are time consuming and costly to carry out and consequently are likely to be undertaken by agencies who have the resources and legislative requirement to evaluate their products in this way, such as pharmaceutical companies. Psycho-social and organizational interventions are unlikely to be evaluated to the same extent, and consequently (with notable exceptions) systematic reviews of these interventions may be unable to provide convincing evidence of their effectiveness. Similarly, drug companies are rarely compelled to produce evidence of cost effectiveness in the process of product legislation and, consequently, cost effectiveness research is not likely to be as prevalent as clinical effectiveness data.

The Future of Systematic Reviews

Systematic review is an imperfect method for synthesizing clinical and cost effectiveness research, but it is much more rigorous, reliable and scientific than that it seeks to replace—the traditional review article.[3] Recent methodological advances also mean that estimates derived from meta-analysis can provide valid and believable estimates for use in economic evaluations. Where uncertainty is demonstrated using systematic review, then this uncertainty can be recognized and incorporated into economic models, though this possibility often goes unrecognized. Many economic models make fundamental errors by using invalid or unrepresentative clinical data, and they often fail to recognize or incorporate areas of clinical uncertainty or to use these as 'sensitivity analyses' by which to test the robustness of their results.[30] Systematic reviews have a great potential to inform the design of robust and believable economic models where these are required.

Some organizations (such as the UK Medical Research Council or the UK NHS Health Technology Assessment Programme) now require that a systematic review be undertaken prior to the funding of costly primary research. This can help to avoid funding unnecessary primary research, where a systematic literature review can already provide the answer. Similarly, systematic reviews have the potential to ensure that future primary research asks the right research questions and, as a consequence, will hopefully produce results that will be of relevance to real clinical practice and policy. Systematic reviews are often costly and time consuming, but, if unnecessary primary research is avoided, can potentially be cost effective. This and the other benefits that have been outlined should ensure that funding of systematic reviews will be given a high priority.

In conclusion, this paper has shown that systematic reviews have considerable potential to inform policy and practice and can readily be incorporated into assessments of the cost effectiveness of interventions. It is important however that those who carry out and those who use systematic reviews are aware of the importance of methodological rigour in such reviews. Systematic reviews, like all research, can be biased and misleading, and it is imperative that decision-makers can readily differentiate between good and poor quality systematic reviews. The main potential sources of bias are now well known, and basic questions can be used to help differentiate rigorous from poor quality systematic reviews (see **Table 1**). The use of only methodologically sound systematic reviews as sources of evidence will however ensure that decision-making in mental health is rational and scientific, and this is surely a laudable goal for mental health policy and practice.

Acknowledgements

We are grateful to Professor Trevor Sheldon, Dr Mark Sculpher and Ms Alison Eastwood for comments on an earlier draft of this paper.

References

1. Bero LA, Jadad AR. How consumers and policymakers can use systematic reviews for decision making. In *Systematic Reviews: Synthesis of Best Evidence for Health Care Decisions*. Mulrow C, Cook D, eds. Philadelphia, PA: ACP, 1998. 29–38.
2. Muir Gray JA. *Evidence-based Healthcare: how to make health policy and management decisions*. 1997, London: Churchill Livingstone.
3. Mulrow CD. Rationale for systematic reviews. *Br. Med. J.* 1994; **309**: 597–599.
4. NHS Centre for Reviews and Dissemination. *Undertaking Systematic Reviews of Research on Effectiveness: CRD guidelines for those carrying out or commissioning reviews. CRD Report 4*. 1996, York: University of York.
5. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am. Psychol.* 1977; **32**: 752–760.
6. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature: IV. How to use an overview. *J. Am. Med. Assoc.* 1994; **272**(17): 1367–71.
7. Sheldon T, Chalmers I. The UK Cochrane Centre and the NHS Centre for reviews and dissemination: respective roles within the information systems strategy of the NHS R&D programme, co-ordination and principles underlying collaboration. *Health Econ.* 1994; **3**: 201–203.
8. Edwards JG. Selective serotonin reuptake inhibitors. *Br. Med. J.* 1992; **304**(6843): 1644–6.
9. Donoghue J, Tylee A, Wildgust H. Cross sectional database analysis of antidepressant prescribing in general practice in the United Kingdom, 1993–5. *Br. Med. J.* 1996; **313**: 861–2.
10. Hotopf M, Lewis G, Normand C. Are SSRIs a cost-effective alternative to tricyclics? *Br. J. Psychiatry* 1996; **168**: 404–409.
11. Hotopf M, Lewis G, Normand C. Putting trials on trial—the costs and consequences of small trials in depression: a systematic review of methodology. *J. Epidemiol. Community Health* 1997; **51**: 354–8.
12. Song F, Freemantle N, Sheldon TA, House A, Watson P, Long A *et al*. Selective serotonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *Br. Med. J.* 1993; **306**: 683–7.
13. Cialdella P. Effective and acceptable treatment for depression. *Br. Med. J.* 1993; **306**: 1125.
14. Foulds RA. Effective and acceptable treatment for depression. *Br. Med. J.* 1993; **306**: 1125.
15. Montgomery SA, Henry J, McDonald G, Dinan T, Lader M, Hindmarch I *et al.* Selective serotonin reuptake inhibitors: meta-analysis of discontinuation rates. *Int. Clin. Psychopharmacol.* 1994; **9**: 47–53.
16. Anderson IM, Tomenson BM. Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: a meta-analysis. *Br. Med. J.* 1995; **310**: 1433–8.
17. Hotopf M, Hardy R, Lewis G. Discontinuation rates of SSRIs and tricyclic antidepressants: a meta-analysis and investigation of heterogeneity. *Br. J. Psychiatry* 1997; **170**: 120–7.
18. Jönsson B, Bebbington PE. What price depression? The cost of depression and the cost-effectiveness of pharmacological treatment. *Br. J. Psychiatry* 1994; **164**: 665–73.
19. Dunbar GC, Cohn JB, Fabre LF, Feighner JP, Fieve RR, Mendels J *et al.* A comparison of paroxetine, imipramine and placebo in depressed out-patients. *Br. J. Psychiatry* 1991; **159**: 394–8.
20. Woods SW, Rizzo JA. Cost-effectiveness of antidepressant treatment reassessed. *Br. J. Psychiatry* 1997; **170**: 257–63.
21. Stewart A. Antidepressant pharmacotherapy: cost comparison of SSRIs and TCAs. *Br. J. Med. Econ.* 1994; **7**: 67–79.
22. Le Pen C, Levy E, Ravily V, Beuzen JN, Meurgey F. The cost of treatment dropout in depression—a cost–benefit analysis of fluoxetine. *J. Affective Disord.* 1994; **31**: 1–18.
23. Kind P, Sorensen J. Modelling the cost effectiveness of the prophylactic use of SSRIs in the treatment of depression. *Int. Clin. Psychopharmacol.* 1995; **10** (suppl. 1): 41–48.
24. Stewart A. Cost effectiveness of SSRIs: a European perspective. *J. Mental Health Policy Econ.* 1998; **1**: 41–50.
25. Thomas CS, Lewis S. Which atypical antipsychotic. *Br. J. Psychiatry* 1998; **172**: 106–9.
26. Knapp MRJ. Costs of schizophrenia. *Br. J. Psychiatry* 1997; **171**: 509–18.
27. Maynard A, Bloor K. Building castles on sands or quicksands? The strengths and weaknesses of economic evaluation in pharmaceuticals. *Br. J. Psychiatry* 1998; **173** (suppl. 12): 12–18.
28. Aitchison KJ, Kerwin RW. Cost effectiveness of clozapine: a UK clinic based study. *Br. J. Psychiatry* 1997; **171**: 125–130.
29. Revicki DA, Luce BR, Weschler JM, Brown RE, Adler MA. Cost-effectiveness of clozapine for treatment-resistant schizophrenic patients. *Hosp. Community Psychiatry* 1990; **41**(8): 850–4.
30. Sheldon TA. Problems in using modelling in the economic evaluation of healthcare. *Health Econ.* 1996; **5**: 1–11.
31. Drummond MF, Knapp MRJ, Burns TP, Miller KD, Shadwell P. Issues in the design of studies for the economic evaluation of new atypical antipsychotics: the ESTO study. *J. Mental Health Policy Econ.* 1998; **1**: 15–22.
32. Glazer WM, Johnstone BM. Pharmacoeconomic evaluation of antipsychotic therapy for schizophrenia. *J. Clin. Psychiatry* 1997; **58** (suppl. 10): 50–54.
33. Davies LM, Drummond MF. Assessment of costs and benefits of drug therapy for treatment resistant schizophrenia in the United Kingdom. *Br. J. Psychiatry* 1993; **162**: 38–42.
34. Adams CE, Soares K, Ahmed I. Clinical dilemmas and the Cochrane Collaboration. *Psychiatr. Bull.* 1998; **22: 467–69**: 467–469.
35. Kennedy E, Song F, Hunter R, Gilbody S. *Risperidone versus 'conventional' antipsychotic medication for schizophrenia (Cochrane Review*), in *The Cochrane Library, Issue 3*. Oxford: Update Software, 1998.
36. Buxton MJ, Drummond MF, Van Hoult BA, Prince RL, Sheldon TA, Szucs T *et al*. Modelling in economic evaluation: An unavoidable fact of life. *Health Econ*. 1997; **6**: 217–227.
37. Oh P, Einarson TR, Iskedjian M, Addis A, Lanctôt K. *Pharmacoeconomic evaluation of risperidone and clozapine in chronic and treatment-resistant schizophrenia.* 1996, Ottawa: Canadian Co-ordinating Office for Health Technology Assessment (CCOHTA).
38. Tyrer P. Cost-effective or profligate community psychiatry? *Br. J. Psychiatry* 1998; **172**: 1–3.
39. Marshall M Lockwood A. Assertive community treatment for people with severe mental disorders (*Cochrane Review*), in *The Cochrane Library, Issue 3*. Oxford: Update Software, 1998.
40. Marshall M, Gray A, Lockwood A, Green R. Case management for people with severe mental disorders *(Cochrane Review)*, in *The Cochrane Library, Issue 3*. Oxford: Update Software, 1998.
41. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann. Intern. Med.* 1997; **126**: 712–20.
42. Morris S, McGuire A, Caro J, Pettitt D. Strategies for the management of hypercholesterolaemia: a systematic review of the cost effectiveness literature. *J. Health Serv. Res. Policy* 1997; **2**: 231–250.
43. NHS CRD. Cholesterol and coronary heart disease: screening and treatment. *Effective Health Care Bull.* 1998; **4**(1).
44. Egger M, Smith GD. Misleading meta-analysis. Lessons from an 'effective, safe, simple' intervention that wasn't. *Br. Med. J.* 1995; **310**: 752–754.
45. Sheldon TA, Vanoli A. *Providing research intelligence to the NHS: the role of the NHS Centre for Reviews and Dissemination*, in *Guidelines for the Economic Evaluation of pharmaceuticals: can the UK learn from Australia and Canada*?, Towse, A. Editor. 1997, Office of Health Economics: London. 53–63.
46. Jefferson T, Mugford M, Gray A, Demicheli V. An exercise on the feasibility of carrying out secondary economic analyses. *Health Econ.* 1996; **5**: 155–65.
47. NHS Centre for Reviews and Dissemination. *Making Cost Effectiveness Information Accessible: the NHS Economic Evaluations Database. CRD Report 6*. 1996, York: University of York.
48. Oxman AD, Cook DJ, Guyatt GH. Users' guide to the medical literature, VI. How to use an overview. Evidence-Based Medicine Working Group. *Am. Med. Assoc.* 1994; **272**: 1367–1371.

