Developing a Quality Assessment Scoring System for Economic Evaluations

Francis Pang[1], Keith Tolley[2], Shazad Amin[3]

[1] Centre for Health Economics, University of York, York, UK
[2] Outcomes Research, Central Research, Pfizer Ltd., Sandwich, UK
[3] Department of Psychiatry, University of Nottingham, Nottingham, UK

Address for Correspondence: Francis Pang, Centre for Health Economics, University of York, Heslington, York YO10 5DD Tel: 01904 432697 Fax: 01904 433644 E-mail: fp4@york.ac.uk

Keywords: Economic evaluation; quality; methodology, schizophrenia

Paper presented to the Health Economics Study Group Winter Meeting, University of Birmingham, January 1999

Work in progress: Please do not quote

---

1. Introduction

In recent years, the 'evidence based' approach has been viewed as the way forward for achieving more rational decision-making. In the pursuit of efficiency, there has been growing interest in the economic evaluation of health care technologies and programmes in which their costs are compared with their consequences. In 1993, Elixhauser et al. compiled a bibliography that contained 240 economic studies. A more recent survey listed 3539 studies (Elixhauser et al., 1998). However, evidence from recent surveys (Drummond et al., 1996; Tolley et al.,1997; Pang, 1996) suggests that economic evaluations are having only limited impact on health service decision-making and few decision-making structures have satisfactory mechanisms in place to make best use of the economic data. Furthermore, several review studies have illustrated the varying and generally low quality of previously published economic evaluations (Gerard, 1992; Udvarhelyi et al., 1992; Adams et al., 1992).

Contrary to popular belief, the issue of quality assessment in economic evaluation is not new; in fact, it was first proposed for the assessment of economic study findings over two decades ago (Williams, 1974). More recently, a number of primers, standards and guidelines from the government, the pharmaceutical industry and academia have appeared in the pursuit of promoting 'good practice' and improving the overall quality of economic evaluations. However, to our knowledge, none of these contain a scoring system which summarises objectively the quality of economic studies available.

This paper investigates whether a quality assessment scoring system (as commonly applied to clinical studies) can be developed for published economic evaluations, and discusses how such a system would be useful in a UK context for a variety of end-users including health care decision-makers (e.g. directors of public health, pharmaceutical advisers) and researchers (both academic and industry). In this regard, we first examine whether a quality scoring system is desirable and if so, whether it is technically and practically feasible. Technical feasibility depends upon how comprehensive a scoring system is envisaged, whereas practical feasibility relates to the research costs and benefits of developing and implementing a system. We then explore various clinical quality assessment scoring systems (scales and checklists), and suggest and discuss the characteristics and approaches that a quality assessment scoring system for economic studies might contain. One approach, for example, in developing a system might be to build a quality scoring system into an existing quality checklist such as the British Medical Journal Working Party (1996) guidelines for economic evaluations. We illustrate this approach by applying a scoring system to a previous study quality review of economic studies using the BMJ checklist in the field of schizophrenia by Amin et al. (1998).

2. Is a Quality Assessment Scoring System Needed?

There are several potential uses of a quality assessment scoring system for economic evaluations. The primary reason is that a quality score could facilitate greater acceptance of economic results within the health care community. A quality score would enable decision-makers to rapidly identify the more reliable research results and help increase their confidence in using economic results in resource allocation decision-making. As well as being reliable and valid, any quality score should be simple and informative for health care decision-makers and other researchers in academic settings and pharmaceutical company outcomes research departments.

In recent years, several databases of economic evaluations have appeared, including the OHE and IFPMA Health Economic Evaluations Database (HEED) and the NHS Economic Evaluation Database. The aim of these databases is to provide structured summaries (reviews) of articles appearing in the literature relevant to the economic assessment of health technologies. However they lack a summary measure which reflects the overall quality of the study. These databases would be the best vehicle for adding a quality score and in maximising value for end-users.

A quality assessment scoring system could be of use in an economic meta-analysis (Pang and Drummond, 1998). The quality of the studies could be used as part of the weight assigned to each study in the meta-analysis, as a method to exclude poor quality studies from the meta-analysis or allow the correlation of study outcomes with quality scores to determine the influence of study quality on demonstrated results.

In developing countries, where resources for health services research are limited, decision-makers are likely to need to use the results of economic evaluations performed elsewhere and in this regard, an initial preliminary assessment of the quality of the studies would be beneficial. The Global Health Forum of the World Health Organisation (WHO) is looking at approaches for the international transferability of results and is building a database of standard reference-case economic evaluations from which to generalise to developing countries.

Quality scores could also address the issue surrounding sponsorship bias related to the methodological quality of economic studies funded by different organisations (Hillman et al., 1991; Freemantle et al., 1994). A quality score would facilitate the assessment of the methodological soundness of research, and if economic studies receive a good quality score, there should be little reason for questioning their credibility. Finally quality scores would allow the monitoring of trends in evaluative practice over time and place (e.g. country) and assist the identification of areas where improvement in methodology might be needed.

3. What is Quality?

Quality is complex and difficult to define, because it could encompass the internal validity (design, conduct, analysis) and external validity, as well as reporting of a study (Ioannidis et al. 1998). Internal validity is primarily concerned with minimising bias; i.e. systematic error in the design, conduct or analysis of a study that results in a mistaken estimate of an intervention's effect on the outcome measured, whereas external validity (generalisability) is concerned with the applicability to a broad range of settings, measurements, time periods, and patient populations.

A bias is any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth (Murphy, 1976). Stages of research in which bias can occur and examples of common biases are shown in box 1.

Box1: Bias

Stages of research in which bias can occur

1. In reading-up on the field
2. In specifying and selecting the study sample
3. In executing the experimental manoeuvre (or exposures)
4. In measuring exposures and outcomes
5. In analysing the data
6. In interpreting the analysis
7. In publishing the results

Some examples of biases

(i) Selection bias

Bias in allocation of subjects to groups is known as selection bias (or assembly bias). Bias arises if subjects are allowed to choose whether they are in the intervention group or the placebo group in a study. Subsequently differences found may be due partly or entirely to differences between the subjects rather to an effect of the intervention. Therefore using an appropriate method that prevents foreknowledge of treatment assignment is crucially important in study design.

(ii) Performance bias

Performance bias arises as a result of systematic differences in care provided to comparison groups other than the intervention of interest. Contamination (provision of the intervention to the control group) and cointervention (provision of unintended additional care to either comparison group) can affect study results. To protect against such unintended differences, blinding is important in protecting against bias.

(iii) Attrition bias

Attrition bias refers to systematic differences between groups in losses of participants from the study. Inadequacies in reporting losses of participants (e.g. withdrawls, dropouts, protocol deviations) and how they are handled has great potential for biasing the analysis.

(iv) Detection bias

Detection bias refers to systematic differences in outcome assessment. Bias can arise as a result of failure to blind outcome assessors.

Sackett et al. (1995) have produced a ranking of the quality of different study designs according to the level of bias inherent within them- see table 1. Randomised controlled trials (RCTs) are considered the gold standard with the lowest threat of bias, followed by cohort studies, case-control studies and case-series.

Table 1: The relationship between levels of evidence and grades of recommendation

Level of evidence	Study design	Grade of recommendation
Level I	Large randomised trials with clear-cut results (and low risk of error)	Grade A
Level II	Small randomised trials with uncertain results (and moderate to high risk of error)	Grade B
Level III	Non-randomised, contemporaneous controls	Grade C
Level IV	Non-randomised, historical controls	Grade D
Level V	No controls, case series only	Grade E

The distinction between the quality of the study and the quality of the reporting, however, is an important one (Moher et al.,1996). The quality of a study provides information on the confidence that the study design, conduct and analyses minimised or avoided biases in its treatment comparisons, whereas the quality of reporting provides information about the

design, conduct, and analysis of the study. Caution should be taken as a study designed with several biases that is well reported can receive a high quality score, while conversely a well-designed and conducted trial that is poorly reported can receive a low quality score. Also information may be reported that is not directly related to validity, such as whether a power calculation was performed (an item that relates more to the precision of results).

4. Quality Assessment Scoring Systems for Clinical Studies

The first step in devising a scoring system for economic evaluations would be to examine the scoring systems currently in operation for clinical studies. Criteria for assessing the quality of clinical trials have been suggested (Chalmers et al., 1987) although no similar list exists for measuring the quality of epidemiologic studies. A number of researchers have also suggested criteria to be considered when evaluating case-control studies (Chalmers et al., 1987) and cohort studies (Feinstein, 1985).

Moher et al. (1995) identified and appraised twenty-five scales and nine checklists for assessing the validity and quality of randomised controlled trials. A scale is a continuum with quantitative units that reflect varying levels of a trait or characteristic to derive an overall summary score (Brown, 1983), whereas a checklist does not contain a quantitative score to each of the questions or attempt to derive an overall quality score. These scoring systems covered four main aspects: study design and analysis; likelihood of bias; randomised controlled trial quality and scientific study quality; and assessed trial features such as:

- Randomisation:
- Double-blinding:
- Dropouts and withdrawls
- Generation of random numbers
- Allocation concealment

In their review, Moher and colleagues evaluated the scales in terms of the type of scale (generic or disease-specific); whether quality was defined; the type of quality assessment (methodological quality or quality of report); the items selected (accepted criteria or pool of items); presence of item on patient assignment; presence of item on masking; presence of item on patient follow-up; presence of item on statistical analysis; whether the scale was rigourously developed; inter-rater reliability; the approximate time in minutes to complete the scoring; the scoring range; and detailed instructions for scoring items. However Moher et al. (1995) discovered major weaknesses in all but one of the scales, criticising them on the grounds of inadequate development and lack of use of standard scale development techniques. Many of the items chosen by the scale developers were based on what authors called 'accepted criteria' from clinical trial textbooks. Although these may be useful, some of them were based on conviction whereas others were based on empirical evidence. The lessons from this review have important implications for developing similar scales for economic evaluations and some of these considerations are summarised in section 7.

5 Quality Assessment Scoring Systems for Economic Studies

Most recent efforts in standardised quality assessment of economic evaluations have employed the checklist approach, predominantly based on the ten-point checklist developed by Drummond et al. (1987) over 10 years ago - see box 2.

Lee and Sanchez (1991) applied an adapted ten-point Drummond et al. checklist to 65 studies asserting the cost-effectiveness of a drug or pharmaceutical services intervention, and concluded that the basic methodological aspects of economic evaluation were frequently overlooked.

Box 2: Checklist for assessing economic evaluations

1. Was a well-defined question posed in answerable form ?
2. Was a comprehensive description of the competing alternatives given ?
3. Was the effectiveness of the programmes or services established ?
4. Were all the important and relevant costs and consequences for each alternative identified?
5. Were costs and consequences measured accurately in appropriate physical units ?
6. Were costs and consequences valued credibly ?
7. Were costs and consequences adjusted for differential timing ?
8. Was an incremental analysis of costs and consequences of alternatives performed ?
9. Was allowance made for uncertainty in the estimates of costs and consequences ?
10. Did the presentation and discussion of study results include all issues of concern to users?

Drummond et al., 1997

Other attempts have included Ganlats and Wong (1991) who rated 47 benefit-cost analyses (BCAs) and cost-effectiveness analyses (CEAs) on seven criteria covering the comprehensiveness of the cost assessment, measurement and valuation of outcomes, use of discounting, sensitivity analysis and appropriateness of conclusions. They awarded 'high' for the satisfaction of 70% or more criteria, 'medium' for 50% to 69%, 'low' for 10% to 49% and '0 for less than 10%.

Also based on Drummond et al. (1987), Adams et al. (1992) applied a checklist to randomised trials identified in a MEDLINE search from January 1966 to June 1988, which considered costs or contained an economic evaluation. A crude attempt was made to produce a quality scoring scale. For those aspects of the checklist seeking to ascertain whether a study had conducted the analysis appropriately e.g. sensitivity analysis, a rating of 0,1 or 2 was awarded corresponding to 'complete and correct sensitivity analysis'; 'partial sensitivity analysis' and 'no sensitivity analysis' respectively. The completeness score was expressed on a 0-1 scale as the proportion of total possible points gained across the checklist. The mean score for completeness of the economic evaluation on the 0-1 scale was 0.52 (range 0.32 to 0.94).

With the aid of published checklists and expert opinion, Gerard (1992) devised a set of 40 characteristics/criteria to evaluate 51 published and unpublished cost-utility studies reporting a cost per QALY ratio between 1980 to 1991. Overall, 46 studies were considered worth undertaking, with 16 technically average, 13 above average and 17 below average.

Udvarhelyi et al. (1992) applied six basic principles to 77 articles published in the periods 1978-1980 and 1985-1987 which were contained 'cost-benefit' or 'cost-effectiveness' in the title or intended to compare costs and effects. They discovered that in general, the literature showed poor compliance with some of the basic principles of economic evaluation. Similarly, Bradley et al. (1995) applied a 12 item checklist of Sacristan et al. (1993) to 90 cost-effectiveness, cost-utility, cost-benefit or cost-minimisation analyses. The checklist contained items concerned with the definition of the study aim, the analysis of the alternatives and the perspective, the measurement of costs and benefits, analysis of results and discussion of the assumptions and limitations of the study. For each item, 4 was allocated for correct, 3 for acceptable, 2 for doubtful, 1 for not reported and 0 for incorrect. To derive the overall score, the sum of individual scores was divided by the number of applicable items and this was compared with a 13th item (overall impression) which was employed as a validity check. 73% of responses across all 12 items were rated adequate (3 or 4) and the mean overall scores improved from 2.5 (1989) to 3.2 (1993).

Finally, the most recent published effort was by Blackmore and Magid (1997) who assessed 44 articles according to six major and four minor methodological principles based on Udvarhelyi et al. (1992). Five studies satisfied all six major criteria, of which three studies also satisfied the minor criteria. On average, studies satisfied a median of three major criteria and one minor criterion.

Although several databases of economic evaluation literature exist, the LMS Alerts Pharmacoeconomic Database (ADIS) is the only one to employ an overall scoring of trial design for economic studies. Elements such as aims, study perspective clarity, relevance of costs and outcomes, appropriate sensitivity analysis performance are considered, scored and graded according to the thresholds in table 2. However, the scoring system does not seem to be validated and ADIS have not explicitly provided details in how they arrive at their scores.

Table 2: ADIS Therapeutic Trial Score Guide

Score	Recommendation
<50	Unacceptable, requires a better designed study
50-70	Fair, some important elements inadequate
71-85	Good to very good, most important elements adequate
86-100	Excellent, highly acceptable

6 Issues and Considerations in Constructing a Scoring System

In this section, we list a range of issues and considerations when devising a scoring system for economic evaluations. These cover issues relevant for considering technical feasibility of a scoring system, and practical feasibility in terms of the costs, benefits and the time of researching, developing a system and implementing it.

(i) Definition of the Quality Construct

A fundamental aspect of any scoring system development is the definition of construct quality. Without such a definition, there is a risk that a system purporting to measure economic study design quality is actually measuring a different construct. For example, some systems may be assessing the methodological quality of a study, whereas others may be assessing the quality of the report or both. The scoring system (scale) by Kleijnen et al. (1991) not only assesses the methodological quality of a trial by asking about the number of patients analysed, but also assesses the quality of the report, by asking whether patient characteristics were adequately described.

(ii) Definition of the Scope of the Scoring System

The scope of the scoring system is important as differences in the scope can lead to discrepancies in how studies are scored. The scope may be concerned with:

- Economic evaluations alongside clinical trials or modelling studies
- Generic or disease-specific
- Published and/or unpublished data

(iii) Specification of Content

The items selected for the scoring system should bear relevance to the objectives. Areas for assessment should be considered carefully otherwise bias could result due to measurement errors with irrelevant information; e.g. if the scoring system contain items not relevant to particular studies. Further attention should be paid to specificity of the assessment. If the scoring system is defined too finely, then a superficial analysis might result. Also the more items there are, the greater the risk of confusing the quality of reporting with the validity of the study. On the other hand, there should be certain criteria that are relevant to all studies ;i.e. some form of core analysis containing an essential set of items common across all studies which allow the possibility for useful comparisons.

(iv) Resource Implications

Systems with multiple scoring and complex scoring take more time to complete than simple approaches. If a 30 item scoring system could generate comparable results to a 5 item one, then resources in terms of costs and time could be reduced, as well as reducing the potential for errors.

(v) Reviewers

A prime consideration is the number of reviewers. Should there be more than one and how will reviewers review the same articles to maximise reliability of mutually exclusive sets of studies to minimise workload? Other considerations include the reviewers' backgrounds and previous training and experience. Would reviewers follow an instruction sheet or have to undergo formal training? Because quality is a subjective process, the potential for bias and error is great. It is recommended that at least two reviewers perform the quality assessment for each study, followed by a consensus meeting to discuss the disagreements might be beneficial.

(vi) Weighting

Most of the clinical scoring systems derive a summary score by adding the scores for each item. If weighting of the items does not occur, then it is implied that all items are of equal importance. Therefore the purpose of weighting is to emphasise the relative value of different dimensions. For example, in an economic evaluation, the specification of the perspective is important as the perspective determines which types of costs are included in the analysis. A consideration in determining which items to weight is related to whether a theoretical approach or empirical approach is adopted. Typically, differential weighting would complicate a scoring system with 40 or more items.

(vii) Threshold Scores

Threshold scores are required to grade studies of differing quality. It is particularly important that the threshold scores are attainable. If threshold scores are used as a measure of quality control to exclude studies , then the setting of threshold scores should be given much consideration.

(viii) Score Adjustment

Scores may need to be adjusted if the maximum attainable total is not a feasible one. For example; compromising a score of 0.375 (on an ordinal scale of zero to one) to 0.5 with a more practical quality benchmark score of 0.75.

(ix) Blind Assessments

To avoid potential bias, it is recommended that identifying information such as authors and institution should be removed from studies before quality assessment. Also reviewers may want to score studies blindly from each other, which would also check reproducibility.

(x) Standardisation

One criticism by Moher and colleagues (1995) of the scales and checklists for randomised controlled trials was the lack of standardisation. Lack of standardisation might be a problem for different databases such as HEED and NHS EED. These have different abstract reviewing processes and each might therefore adopt a different quality scoring process with the result of differing scores for the same study across databases. In this situation, quality scores might serve to confuse the end-user, rather than help.

(xi) Selection of Studies for Review

A consideration would be whether the studies to be scored are randomly sampled or comprehensively retrieved as part of a systematic literature search. If studies from different countries are to be compared in terms of quality for example, then consideration must be given to the assembly of a representative set of studies to be scored.

(xii) Strategies for Scoring

There are several strategies for implementing a scoring system. These might include:

1. Simple reviewer judgement in which the reviewer of the paper assesses the quality of each study with an overall score based on some predetermined criteria

2. A panel of experts, in which a panel comes to a consensus over a quality score. However this is more costly in both time and effort.
3. Either 1 or 2 above, but with more detailed quality scoring for each component of checklist.

(xiii) Value for End-Users

To be most useful for health care decision-makers and researchers, a quality score for economic evaluations would have to be accepted by them and easily accessible. For example; a quality score could be awarded to studies included in an 'accepted' databases such as the NHS Economic Evaluations Database.

7. Case-Study: The Management of Schizophrenia

Although the objectives of the primers, standards and guidelines may differ somewhat, they contain elements of commonality. The main rationale behind academic standards is to evaluate the quality and scientific soundness of the study and to ensure valid and unbiased results are relevant to the study's purposes. Regulatory guidelines (i.e. for reimbursement of pharmaceuticals), on the other hand, have additional concerns such as making the economic analysis results conform to the regulator's own decision-making process. Regulatory guidelines are also concerned with promoting better comparability between studies by recommending a standard methodology, even if it is not necessarily the best method from an academic viewpoint.

A recent attempt to set quality standards has been the work of the BMJ Economic Evaluation Working Party, representing a consensus of European health economists and other healthcare experts. The purpose of this checklist was to help BMJ referees and editors to assess the quality of economic studies submitted to the BMJ.

The case-study is drawn from a previous study of using BMJ guidelines to review quality in schizophrenia (Amin et al., 1998). In this paper (in press), the guidelines were used for quality assessment of a set of economic studies. Each study was assessed to determine how many items in the checklist were covered ('YES'), not covered ('NO'), not applicable ('NA') or not clear. An issue is whether their usefulness could be improved through adding some form of quality scoring. In this case-study, we add a crude quality score to each of the studies reviewed.

Background

Schizophrenia is probably the most 'costly' mental illness, both in terms of its financial and social burden. Schizophrenia is one of the most common chronic psychiatric disorders, having an incidence of approximately 15/100,000 per year and a prevalence of 0.5 – 1%. Since the most common age of first presentation is in the 25-35 age group, the disorder produces clinical and social disability leading to maximum disruption of the patients and families; lives. Lifetime health and social care costs per person have been estimated to be £1,175 to £315,776 according to severity of illness, with at least 60% of the cost associated with hospital and community residential care. Total NHS direct care costs in England in 1992/1993 have been estimated at £810 million, representing 2.8% of all NHS costs.

Methodology

MEDLINE, OHE, NEED, BIDS, COCHRANE databases searches were conducted in the period 1966-1997, using the keywords cost and cost analysis (all subheadings), health economics, economic evaluation, schiz*, psych* and ment*. Search coverage was extended to bibliographies of retrieved articles and handsearches of journals.

The inclusion criteria were (i) the study considered both the costs, and cost or health consequences of alternative health programmes for schizophrenia. Direct cost studies, which are used to evaluate the costs of treatment and care for a disease or condition without measuring outcomes were excluded; (ii) the study was published in English; (iii) the majority (i.e. >50%) of the patients sampled had a diagnosis of schizophrenia or related psychoses. In

practice, the latter term describes the non-affective psychoses; F20-F29 in the ICD-10 classification of Mental & Behavioural Disorders.

The studies were quality reviewed using the BMJ guidelines by 2 authors (SA, KT) blind from each other. In principle, the maximum total of checklist items which could be achieved for each study, was 35. This ranged from 21 to 29 for each specific study, reflecting the wide range of types of studies reviewed, and the number of items that were not applicable. For example items 12 and 13 relate to cost-utility studies of which there were none in the review. Although caution should be used when comparing studies in this way, one of the main reasons for using a standard checklist is to introduce a degree of objectivity when assessing methodological quality. We have thus reported the % of relevant items scoring positively in the final column: % positive score (Yes/Max total), and used this as a crude scoring system. This provides a basis as to whether it is worth developing a more complex quality scoring system for these studies (which could be incorporated into the BMJ checklist).

Results

Thirty studies met the inclusion criteria. The studies either considered alternative methods of service delivery (mainly community based) or the use of novel antipsychotic drugs such as clozapine or risperidone. 17 studies were performed in the US, 7 in the UK, and 6 elsewhere.

Table 3 presents the results of the review using the standards set out in the BMJ checklist and the yes% quality assessment.

Table3: Schizophrenia Studies Reviewed Using BMJ Checklist

	Study						Quality score
		YES	NO	NOT CLEAR	N/A	MAX	%
1	Endicott et al. 1978	15	7	2	11	24	63%
2	Bedell et al. 1989	10	8	5	12	23	43%
3	McCrone et al. 1994	16	6	2	11	24	67%
4	Honingfield et al. 1990	10	9	3	13	22	45%
5	Fenton et al. 1984	15	5	2	13	22	68%
6	Matheson et al. 1994	23	4	1	7	28	82%
7	Hale et al. 1996	14	6	7	8	27	52%
8	Reid et al. 1994	15	8	1	11	24	63%
9	Glazier et al. 1996	17	5	3	10	25	68%
10	Rosenbeck et al. 1995	19	6	0	10	25	76%
11	Quinvalan et al. 1995	13	8	2	12	23	57%
12	Rappaport et al. 1987	16	6	2	11	24	67%
13	Hu et al. 1991	12	8	6	9	26	46%
14	Hoult et al. 1984	15	8	0	12	23	65%
15	Wherley et al. 1987	14	7	2	12	23	61%
16	Revicki et al. 1990	16	5	1	13	22	73%
17	Melzer et al. 1993	19	2	3	11	24	79%
18	Wiersma et al. 1991	16	4	2	13	22	73%
19	Wiersma et al. 1995	15	7	1	12	23	65%
20	Linn et al. 1979	14	6	1	14	21	67%
21	Levenson et al. 1977	9	9	3	14	21	43%
22	Rund et al. 1994	16	5	3	11	24	67%
23	Newton et al. 1983	13	5	4	13	22	59%
24	Guest et al. 1996	22	3	3	7	28	79%
25	Hyde et al. 1987	17	3	2	13	22	77%
26	Jones et al. 1980	21	3	0	11	24	88%
27	Dickey et al. 1986	17	4	0	14	21	81%
28	Weisbrod et al. 1980	20	5	0	10	25	80%
29	Jonsson et al. 1995	18	3	2	12	23	78%
30	Davies et al. 1993	23	1	5	6	29	79%

The checklist contained 35 items covering study content, design of economic evaluation, analysis and reporting of results. Studies were assessed blind by SA and KT to determine the number of items covered ('YES'), not covered ('NO'), not applicable ('NA') or not clear. Agreement on YES/NO/NA were made by consensus of the two reviewers (the NOT CLEAR category reflected where consensus could not be reached).

8. Discussion and Conclusions

Although efforts to assess quality in the economic literature can be traced as far back as 1974 when Williams listed the essential elements that should be present in economic evaluations, quality assessment scoring is still relatively new. Quality assessment scoring is important in that it would allow health care decision-makers and researchers to recognise reliable economic evidence for use in the allocation of scarce resources.

Hitherto, checklists have been the norm for appraising the quality of the economic literature, despite the fact that checklists can suffer from many problems. These include checklist developers not providing details on how and why their items were selected for inclusion and the variation in the number of items between checklists. Checklists are probably most useful as guidance for which items authors should include when reporting their studies.

Perhaps akin to the development of clinical quality assessment scoring systems in which checklists preceded scales, more attention will now be paid to economic evaluation scales and checklists. Future efforts in developing scales will require appropriate rigour and in this regard, developing a scale to assess quality should be considered similar to developing any other instrument. From our case-study of schizophrenia , comparing studies using checklist scores such as based on the BMJ economic evaluation guidelines may be one reasonable method of summarising methodological quality, and provide a basis for adding an overall quality assessment.

However, several formidable obstacles to quality assessment for economic evaluations still remain. Economic evaluation is an inexact science and there is a lack of methodological standardisation, with many methodological issues remaining unresolved. Also recommendations for improved reporting have largely been ignored and the transparency of economic evaluations tends to be poor. Without adequate reporting, it is simply not possible in many cases to assess the quality of studies. These obstacles may need to be adequately addressed before more attention can be given to the next step of quantitative quality scoring.

It is hypothesised that poor economic evaluations may have negative effects on the allocation of scarce resources, can be potentially misleading and represent an obstacle in generalising across economic settings (BMJ Working Party, 1996), but this is largely untested. Research needs to be performed not only on the development of economic quality assessment methods, but also the link between the quality of economic information and their impact on health care decision-making.

To conclude, assessing quality and scoring it is difficult and potentially very subjective. Is it worth pursuing further and why?

Within the spirit of the HESG, this is very much work in progress. The primary purpose of this short paper is to stimulate discussion on the possibility of quality assessment scoring in economic evaluation.

Possible discussion points:

1. *Is a quality assessment scoring system desirable or technically and practically feasible? How do you implement it?*
2. *What characteristics should a quality assessment scoring system possess? (e.g. simple or complex; costs vs benefits)*
3. *Is a scoring system useful as an add-on to a checklist such as a BMJ one?*
4. *How can quality scores be made more useful for health care decision-makers? Will it make them use economic evaluation results more in their decision-making and priority-setting?*

9. References

Elixhauser, A. et al. Healthcare CBA/CEA: an update on the growth and composition of the literature. *Medical Care* 1990; 31:JS1-JS11

Elixhauser, A., Halpern, M., Schmier, J., Luce, B.R. Health Care CBA and CEA from 1991 to 1996: An updated bibliography. Medical Care 1998;36(5):MS1-MS9

Drummond, M., Cooke, J., Walley, T. Economic evaluation in health care decision-making: evidence from the UK. York: University of York, *Centre for Health Economics Discussion Paper* 1996

Tolley, K., Obermann, K. Setting priorities. York: University of York, *Centre for Health Economics Discussion Paper* 1997

Pang, F. Uses of economic evaluation by decision-makers in resource allocation decision-making. *MSc Thesis* 1996. unpublished.

Gerard, K. Cost-utility in practice: a policy-maker's guide to the state of the art. *Health Policy* 1992; 21:249-279

Udvarhelyi, S., Colditz, G.A., Rai, A., Epstein, A.M. Cost-effectiveness and cost-benefit analysis in the medical literature: are the methods being used correctly? *Annals of Internal Medicine* 1992; 116:238-244

Adams, M.E., McCall, N.T., Gray, D.T., Orza, M.J., Chalmers, T.C. Economic analysis in randomized controlled trials. *Medical Care* 1992; 30(3):231-243

Williams, A. The cost-benefit approach. *British Medical Bulletin* 1974; 30:252-256

BMJ Working Party on Economic Evaluation. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal* 1996; 313:275-283

Shazad, A., Tolley, K., Harrison, G. Improving quality of economic evaluations of the management of schizophrenia. *British Journal of Medical Economics* 1998; in press

NHS Centre for Reviews and Dissemination. *NHS Economic Evaluation Database (NHS EED).* University of York 1998

Office of Health Economics (OHE)/ International Federation of Pharmaceutical Manufacturers Association (IFPMA). *Health Economic Evaluation Database (HEED)*, OHE, London, 1998

Freemantle, N., Maynard, A. Something rotten in the state of clinical and economic evaluations ? *Health Economics* 1994;3:63-67

Hillman, A.L., et al. Avoiding bias in the conduct and reporting of cost-effectiveness research published by pharmaceutical companies. *New England Journal of Medicine* 1991;324:1362-1365

Murphy, E.A., *The Logic of Medicine.* Baltimore: John Hopkins University Press 1976

Moher D., Jadad, A. et al. Assessing the quality of randomised controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995; 16:62-73

Chalmers, T.C., Levin, H., Sacks, H.S., Reitman, D., Berrier, J., Nagalingam, R. Meta-analysis of clinical trials as a scientific discipline I. Control of bias and comparison with large co-operative trials. *Statistics in Medicine* 1987;6:315-325

Chalmers, T.C., Berrier, J., Sacks, H.S., Levin, H., Reitman, D. Nagalingam, R. Meta-analysis of clinical trials as a scientific discipline II. Replicate variability and comparison of studies that agree and disagree. *Statistics in Medicine* 1987;6:733-744

Brown, S.A. Measurement of quality of primary studies for meta-analysis. *Nursing Research* 1991;40:352-355

Drummond, M.F., O'Brien, B.J., Stoddart, G.L. and Torrance, G.W. *Methods for the economic evaluation of health care programmes.* 2nd Edition. Oxford: Oxford University Press, 1997

Lee, J.T., Sanchez, L.A. Interpretation of 'cost effective' and soundness of economic evaluations in the pharmacy literature. *American Journal of Hospital Pharmacy* 1991;48:2622-2627

Ganiats, T,G., Wong, A.F. Evaluation of cost-effectiveness research: a survey of recent publications. *Family Medicine* 23; 457-462

Blackmore, C., Magid, D.J. Methodologic evaluation of the radiology cost-effectiveness literature. *Radiology* 1997;203:87-91

Kleijnen, J., Knipschild, P., ter Riet, G. Clinical trials of homeopathy. *British Medical Journal* 1991; 302:316-323

Lumley, J., Bastian, H. Competing or complementary? Ethical considerations and the quality of randomised controlled trials. *International Journal of Technology Assessment in Health Care* 1996; 12:247-263

Schultz, K., Chalmers et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled clinical trials. *Journal of the American Medical Association* 1995; 273:408-412

Moher D, Jadad, A., Tugwell, P. Assessing the quality of randomised controlled trials. *International Journal of Technology Assessment in Health Care* 1996; 12:195-208

Gotzsche, P. Methodology and over and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clinical Trials* 1989; 10:31-56

Dickerson, K. Herxheimer. Introduction. The quality of the medical evidence: Is it good enugh? *International Journal of Technology Assessment in Health Care* 1996; 12:187-189

Sackett, D. Bias in analytic research. *Journal of Chronic Diseases* 1979; 32:51-63

Bero, L., Drummond, R. Influences on the quality of published drug studies. *International Journal of Technology Assessment in Health Care* 1996; 12:209-237

Stewart, L., Parmar, M. Bias in the analysis and reporting of randomised controlled trials. *International Journal of Technology Assessment in Health Care* 1996; 12:264-275

Jacobs, P., Bachynsky, J., Baladi, J. A comparative review of pharmacoeconomic guidelines. *PharmacoEconomics* 1996; 8; 182-189

Schizophrenia Studies

Endicott J., Hertz, M., Gibbon, M. Brief versus standard hospitalisation: the differential costs. *American Journal of Psychiatry* 1978;135(6):707-712

Bedell, J. Ward, J. An intensive community-based treatment alternative to state hospitalisation. *Hospital Community Psychiatry* 1989;40(5):533-535

McCrone, P., Beecham, J., Knapp, M. Community psychiatric nurse teams: cost-effectiveness of intensive support versus generic care. *British Journal of Psychiatry* 1994; 165:218-221

Honigfield, G., Patin, J. A two-year clinical and economic follow-up of patients on clozapine. *Hospital Community Psychiatry* 1990; 41(8):882-885

Fenton, F., Tessier, L., Struening, E., Smith, F., Benoit, C., Contandriopoulus, A. et al. A two-year follow-up of comparative trial of the cost-effectiveness of home and hospital psychiatric treatment. *Canadian Journal of Psychiatry* 1984; 29:205-211

Matheson, L., Cook, H., McKenna, P., Bosanquet, N. Value for money care for patients with schizophrenia. *British Journal of Medical Economics* 1994; 7:25-34

Hale, A., Wood, C. Comparison of direct costs for schizophrenia using oral or depot neuroleptics: a pharmacoeconomic analysis. *British Journal of Medical Economics* 1996; 10:37-45

Reid, W., Mason, M., Toprac, M. Savings in hospital bed-days related to treatment with clozapine. *Hospital Community Psychiatry* 1994; 45(3):261-264

Glazer, M., Ereshefsky, L. A pharmacoeconomic model of outpatient antipsychotic therapy in 'revolving door' schizophrenic patients. *Journal of Clinical Psychiatry* 1996; 57:337-345

Rosenheck, R., Neale, M., Leaf, P., Milstein, R., Frisman, L. Multisite experimental cost study of intensive psychiatric community care. *Schizophrenia Bulletin* 1995;21(1):129-140

Quinlivan, R., Hough, R., Crowell, A., Beach, C., Hofsetter, R., Kenworthy, K., Service utilisation and costs of care for severly mentally ill clients in an intensive case management program. *Psychiatic Services* 1995; 46(4):361-371

Rappaport, M. Goldman, H., Thornton, P., Stegner, B., Moltzen, S., Hall, K., et al. A method of comparing two systems of acute 24-hour psychiatric care. *Hospital Community* 1987; 38:1091-1095

Hu, T., Jerrell, J. Cost-effectiveness of alternative approaches in treating severly mentally ill in California. *Schizophrenia Bulletin* 1991; 17(3):461-468

Hoult, J., Rosen, A., Reynolds, I. Community orientated treatment compared to psychiatry hospital orientated treatment. *Social Science in Medicine* 1984; 18:11:1005-1110

Whereley, M., Bisgaard, S. Beyond model programs: evaluation of a countrywide system of residential treatmengt programs. *Hospital Community Psychiatry* 1987;38(8):852-857

Revicki, D., Luce, B., Weschler, M., Brown, R., Adler, M. Cost-effectiveness of clozapine for treatment-resistant schizophrenic patients. *Hospital Community Psychiatry* 1990;41(8):850-854

Melzer, H., Cola, P., Way. L., Thompson, P., Bastani, B., Davies, M. et al. Cost-effectiveness of clozapine in neuroleptic resistant schizophrenia. *American Journal of Psychiatry* 1993; 150(11):1630-1638

Wiersma, D., Kluiter, H., Nienhuis, F.J., Ruphan, M., Giel, R. Costs and benefits of day treatment with community care for schizophrenic patients. *Schizophrenia Bulletin* 1991; 17(3):411-419

Wiersma, D., Kluiter, H., Nienhuis, J., Ruphan, M., Giel, R. Costs and benefits of hospital and day treatment with community care of affective and schizophrenic disorders. *British Journal of Psychiatry* 1995;.166(suppl. 27):52-59

Linn, M., Caffey, E., Klett, J., Hogarty, G., Lamb, L. Day treatment and psychotropic drugs in the after care of schizophrenic patients. *Archives of General Psychiatry* 1979; 36:1055-1066

Levenson, A., Lord, C., Sermas, C., Thornby, J., Sullender, W., Comstock, B. Acute schizophrenia: an efficacious outpatient treatment approach as an alternative to full-time hospitalisation. *Disorders of Nervous System* 1977; 38:242-245

Rund, B., Moe, L., Sollien, T., Fjell, A., Borchgrevnik, T., Hallert, M., Naess, P. The psychosis project: outcome and cost-effectiveness of a psychoeducational treatment programme for schizophrenic adolescents. *Acta Psychiatric Scandinavia* 1994; 89:211-218

Newton, P. An evaluation of the cost effectiveness of day hospitalisation for black male schizophrenics. *Journal of National Medical Association* 1983; 75(3):273-285

Guest, J., Hart, W., Cookson, R., Lindstrom, E. Pharmacoeconomic evaluation of long-term treatment with Risperidone for patients with chronic schizophrenia. *British Journal of Medical Economics* 1996; 10:59-67

Hyde, C., Bridges, K., Goldberg, D., Lowson, K. Sterling, C., Faragher, B. The evaluation of a hostel ward. A controlled study using modified cost-benefit analysis. *British Journal of Psychiatry* 1987:151:805-812

Jones, R., Goldberg, D., Hughes, B. A comparison of two different services treating schizophrenia: a cost-benefit approach. *Psychological Medicine* 1980;10:493-505

Dickey, D., Cannon, N., McGuire, T., Gudeman, J. The quarterway house: a two year cost study of an experimental residential program. *Hospital Community Psychiatry* 1986; 37(11):1136-1143

Weisbrod, B., Test, M., Stein, L. Alternative to mental hospital treatment – II Economic benefit-cost analysis. *Archives General Psychiatry* 1980; 37:400-405

Jonsson, D., Walinder, J. Cost-effectiveness of clozapine treatment in therapy-refractory schizophrenia. *Acta Psychologica Scandinavia* 1995; 92:199-201

Davies, L., Drummond, M. Assessment of costs and benefits of drug therapy for treatment resistant schizophrenia in the United Kingdom. *British Journal of Psychiatry* 1993;162:38-42

Appendices

Appendix 1: The BMJ Checklist

Item No.	Checklist Item
<Study Design>	
1	The research question is stated.
2	The economic importance of the research question is stated.
3	The viewpoint(s) of the analysis are clearly stated and justified.
4	The rationale for choosing the alternative programmes or interventions being compared is stated.
5	The alternatives being compared are clearly described.
6	The form of economic evaluation used is stated.
7	The choice of form of economic evaluation is justified in relation to the questions addressed.
<Data Collection>	
8	The source(s) of effectiveness estimates used are stated.
9	Details of the design and results of effectiveness study are given (if based on a single study)
10	Details of the methods of synthesis or meta-analysis of estimates are given (if based on a number of effectiveness studies)
11	The primary outcome measure(s) for the economic evaluation are clearly stated.
12	Methods to value health states and other benefits are stated.
13	Details of the subjects from whom valuations were obtained are given.
14	Productivity changes (if included) are reported separately.
15	The relevance of productivity changes to the study question is discussed.
16	Quantities of resources are reported separately from their unit costs.
17	Methods for the estimation of quantities and unit costs is described.
18	Currency and price date are recorded.
19	Details of currency or price adjustments for inflation or currency conversion are given.
20	Details of any model used are given.
21	The choice of model used and the key parameters on which it is based are justified.
<Analysis and Interpretation of Results>	
22	Time horizon of costs and benefits.
23	The discount rate(s) is stated.
24	The choice of rate(s) is justified.
25	An explanation is given if costs or benefits are not discounted.
26	An explanation is given if costs or benefits are given for stochastic data.
27	The approach to sensitivity analysis is given.
28	The choice of variables for sensitivity analysis is justified.
29	The ranges over which the variables are varied are stated.
30	Relevant alternatives are compared.
31	Incremental analysis is reported.
32	Major outcomes are presented in a disaggregated as well as aggregated form.
33	The answer to the study question is given.
34	Conclusions follow from the data reported.
35	Conclusions are accompanied by the appropriate caveats.

Each checklist item rated YES, NO, NOT CLEAR OR N/A.