HESG Spring 2008

Presentation of economic information in NICE clinical guidelines: adaptation of the GRADE profile

Joanne Lord and Francis Ruiz

Contact details: Centre for Clinical Practice, National Institute for Health and Clinical Excellence, MidCity Place, 71 High Holborn, London WC1V 6NA
0107 067 5914, joanne.lord@nice.org.uk

Abstract

NICE clinical guidelines give recommendations for the diagnosis, treatment and care of people with specific diseases and conditions within the NHS. The guideline development groups are expected to take account of the impact of interventions on the use of limited healthcare resources, as well as their clinical effects. Guideline economists obtain evidence of costs and cost-effectiveness from the literature, when this is of sufficient quality and relevance, or from their own models.

NICE and its collaborating centres are considering introducing the GRADE system for the presentation of clinical evidence in guidelines. Focused on key outcomes, the GRADE profile table summarises the methodological quality, relevance and results of a body of evidence about the clinical harms and benefits of alternative interventions for a given population. It is thought that this will assist guideline developers in weighing up the body of evidence. It should also help guideline users to understand the clinical evidence base for recommendations.

However, GRADE is not designed for economic information, which is an essential component of decision-making in NICE guidelines. So we propose to adapt the GRADE system for presentation of cost and cost-effectiveness estimates. In this paper, we present a format for an economic evidence profile, including checklists for appraising the methodological quality of economic evaluations and their relevance to the decision-making context. We also present an example of the use of the system and discuss the practical and conceptual challenges of extending its use across the guidelines programme.

1 Introduction

NICE's clinical guidelines cover broad aspects of clinical care and the clinical management of specific conditions. The development processes are underpinned by NICE's key principles of basing recommendations on the best available evidence and involving all stakeholders in a transparent and collaborative manner. Stakeholders include national organisations that represent patients and carers, healthcare professionals and companies that have an interest in the guidance under development

Once a guideline topic is referred to the NICE programme, we commission one of seven national collaborating centres (NCCs) to develop the scope and guideline on our behalf. The NCC convenes a guideline development group (GDG) to produce the guideline. This GDG includes healthcare professionals and patients and carers, as well as technical staff (systematic reviewer, health economist, project manager and information scientist) employed by the NCC. GDGs are expected to take account of the impact of interventions on the use of limited healthcare resources, as well as their clinical effects.

To facilitate the translation of clinical and cost effectiveness evidence to recommendations, and to do this as transparently as possible, NICE and its collaborating centres are considering introducing the GRADE (Grades of Recommendation, Assessment, Development and Evaluation) system for the presentation of clinical evidence in guidelines with the update of the Guidelines Manual in October 2008. Focused on key outcomes, the GRADE profile table summarises the methodological quality, relevance and results of a body of evidence about the clinical harms and benefits of alternative interventions for a given population [18]. GRADE should enable guideline developers to present evidence of health benefits and harms in a more compact and transparent way. Although it is not designed for economic evidence, it may also be adapted for presentation of relevant cost and cost-effectiveness information alongside the clinical evidence.

This paper sets out proposals for the development and testing of a GRADE inspired economic profile for use in NICE guidelines. In Section 2, we give a brief description of the GRADE system for clinical effectiveness evidence, make proposals for how it can be adapted for economic evidence, and present an example of a combined clinical and economic evidence profile for a real guideline question. In section 3 we discuss some potential complications related to the implementation of GRADE and economic evidence profiles. And finally, in section 4 we set out our plans for testing our proposed approach.

2 Format for the economic evidence profile

The GRADE evidence profile table [18] summarises the quality and results of a body of evidence relating to a single clinical question (defined by an intervention, comparator, population and outcomes). It contains a row for each outcome, including harms as well as benefits. If there is more than one study for an outcome, the row may summarise the results of a meta-analysis if available, or otherwise it may show a range of results or those from a single, good-quality study. Each outcome is assigned a quality score (high, moderate, low or very low), based initially on the study designs (RCTs start high and observational studies start low), but then modified for other indicators of bias. For example, an RCT might be downgraded if it is judged to have serious limitations, such as failure of allocation concealment. GRADE users are also expected to rate the relative importance of the different outcomes, using a subjective ordinal scale. The overall quality of the body of evidence is then summarised by the lowest quality score for the outcomes judged to be 'critical' to the decision. The GRADE system also includes a method for classifying the 'strength' of recommendations based on the magnitude and quality of evidence on the trade-offs between benefits, harms and costs.

NICE has decided to make some modifications to this standard GRADE approach for its guidelines programme: dropping the assessment of importance for the outcomes, the overall assessment of quality for the body of evidence; and the GRADE classification of strength of recommendations.

We now discuss how the GRADE profile table may be adapted for the presentation of economic evidence in NICE guidelines. Estimates of resource use obtained directly from observations in clinical trials can be treated like any other outcomes and included in standard GRADE tables. So, here we are concerned with other types of economic evidence, including costs, cost-effectiveness ratios or net benefits estimated using some form of 'modelling'. In addition to formal decision analytic models, this includes economic evaluations conducted alongside clinical trials, since these usually require some external sources of information (e.g. unit costs, health state valuations, long term prognostic data) and estimation procedures to predict long-term costs and outcomes [5]. It also applies to simple cost calculations based on expert judgement or observed resource use and unit cost data.

2.1 Quality Assessment

It is essential that the quality of any economic evidence used to inform guideline recommendations should be assessed. This applies equally to estimates from the published literature, stakeholder submissions, and analyses conducted by the guideline economist.

The quality assessment section of the GRADE evidence profile has columns for four main criteria: study design, limitations, consistency and applicability. It also includes an 'other considerations' column for additional factors that may influence the GRADE quality score, such as sparse or imprecise data. The interpretation of these criteria for economic evidence is discussed below.

2.2 Design

GRADE uses a study-design-based hierarchy of evidence as the starting point for quality assessment, broadly categorising studies as observational or randomised. This reflects the importance of random allocation to the internal validity of effectiveness studies [6].

By extension, we could use the study design for the effectiveness estimate as an indicator of quality for economic evidence. Or a better approach, perhaps, might be to use the study design column to reflect a broader judgement about whether the economic evaluation is consistent with the best available clinical evidence. The effectiveness estimate is an important component of an economic evaluation and it might aid transparency to show explicitly how the economic evidence is linked to the effectiveness evidence.

An alternative way of interpreting the design criterion would be to categorise economic evaluations according to the robustness of their designs. However, there is no obvious or agreed hierarchy for doing this. An analysis based on effectiveness and resource use data from a randomised trial is not necessarily superior to a modelling study using data from various sources. The quality of an economic evaluation depends on the quality of all data inputs and the appropriateness of the modelling assumptions that go into it.

On balance we consider that there would be little or no benefit from retaining a 'design' column for the economic profile. If there are specific issues related to the economic study design these can be noted in the 'other comments' column.

2.3 Limitations

In GRADE this criterion relates to the possibility of bias due to limitations in detailed study methods and execution. The grade of evidence is reduced by one point if 'serious' limitations are identified, or by two for 'very serious' limitations. The GRADE group do not recommend a specific instrument for judging what constitutes a serious or very serious limitation in a clinical study, but refer reviewers to a number of published critical appraisal instruments.

It is also essential that economic evaluations used to inform guideline recommendations are critically appraised, and that any important limitations are identified. The current version of the NICE Guidelines Manual [28] does not recommend use of one specific checklist, but refers to several options [10] [9,31]. We conducted a brief literature review to identify what other instruments are available and whether there is evidence for the reliability or validity of these instruments.

There is an extensive literature on the critical appraisal of economic evaluations. Much of this is derived from a set of principles developed for economic evaluations in the late 1980's, known as the 'Drummond criteria' [11]. These criteria have been adapted into various critical appraisal checklists and scales for assessing the quality of reporting and conduct of published economic evaluations [1-4,7,9,13-17,19,20,22-

24,29,32,33,35,36,38,40]. Additional criteria and instruments have been proposed for assessing decision analytic models [25,31,34,37,39]. These go beyond the basic Drummond criteria to investigate more detailed aspects of the model design, data selection and analysis. Most of the instruments in the literature were developed informally for specific purposes, with selection of items and methods of scoring based on the judgment of a small group of researchers or external experts, and little or no attempt at validation. Some exceptions include the BMJ [9] and 'Gold panel' checklists [36,40], which were developed by working groups of experienced health economists. Two more recent initiatives have also taken a much more formal approach to instrument development and testing: the CHEC [2,12] and QHES projects [7].

Comparing the processes used to develop the CHEC and QHES instruments, they have different strengths and weaknesses. Both started with a systematic review of other published guidelines and checklists to identify an item pool, and consensus techniques with an expert panel to select items for inclusion and to agree wording. The CHEC project team was particularly rigorous in the methods used to agree the items for inclusion in the checklist, with a large, distinguished expert panel and three rounds of DELPHI questionnaires. This lends a degree of face validity to this instrument, although the CHEC team did not attempt to create a weighted scoring system and so far have done little validation of the instrument.

In contrast, QHES was based on a smaller group of experts and less rigorous consensus techniques to select items, but the research team put a lot of effort into developing and testing a weighted scoring system using a conjoint analysis method with a very large panel of health economists. They followed this up with a validation study, involving another large panel of health economists. This effort was reasonably successful, producing a system with good evidence of convergent and discriminant validity. However, the added value of weighted scoring is debateable, since most of the weights are similar, and the weighted score was not significantly better at predicting a global quality rating than an unweighted score. The steering group also chose to retain a question about subgroup analysis, even though this was not a significant predictor of the global quality rating. Users of the system in the validation study appeared ambivalent about the system, with 64% saying that they would use the system or recommend it to others, and 56% rating its value as greater or equal to 3 on a 5 point scale (1="not valuable at all" and 5="extremely valuable")[7,30]. However, this may represent a more generalised uncertainty amongst health economists about the value of checklists and scales for assessing the quality of economic evaluations rather than a specific concern with QHES, since other instruments have not been submitted to this same degree of validation.

In terms of content, there is a lot of overlap between QHES and CHEC, which clearly show a common inheritance from the Drummond criteria. However, there are some differences. CHEC was developed to provide a 'minimum standard' for assessing the quality of full economic evaluations based on effectiveness studies (RCTs and observational studies). The authors note that other methodological criteria are required for modelling studies. Thus

CHEC would have to be supplemented with additional questions for evaluation of modelling studies, taken, for example, from the HTA checklist for good practice in decision-analytic modelling [31]. The QHES team had rather broader objectives, aiming to produce a system for appraising the quality of any economic evaluations. As a result, their instrument includes items relating to the selection of data inputs and methods of modelling studies, although the level of detail covered is much lower than that in the HTA modelling checklist.

The wording of CHEC is generally simpler and less open to misunderstanding than that of QHES, which includes some double-barrelled questions: e.g. "were the perspective of the analysis and the reasons for its selection stated?" The CHEC-list avoids such ambiguous questions and includes notes to help users to interpret them in a consistent way. These notes are useful, although they might not be entirely appropriate for a review intended to inform NICE guidance – for example, they indicate a preference for a societal perspective.

To summarise, the methods used to develop CHEC and QHES were more rigorous than those used for other available critical appraisal instruments for economic evaluations. However, neither system is ideal for use in NICE guidelines. CHEC is clear and simple for users and has good face validity, but some of the questions and user notes are not appropriate for the context of NICE decision-making, and it is not suitable on its own for assessment of modelling studies. QHES is a little less easy for users to apply, with some ambiguous questions and no user notes. Although it contains key questions relevant to models, it does not give a detailed assessment of model quality. The QHES scoring system may be seen as an advantage, although it is not apparent that this adds much value over an unweighted score[17].

On balance we feel that use of a common checklist for critical appraisal of economic evaluations would facilitate consistency across the NICE clinical guidelines programme. It is proposed that the CHEC-list be modified and supplemented by key questions relating to the quality of modelling studies from QHES (see Table 1 for suggested checklist). When necessary, the HTA checklist for decision-analytic models[31] may be used to give a more detailed assessment of methodological quality for modelling studies.

In considering methodological quality, the aim for the appraiser is to assess the extent to which an individual evaluation succeeds in fulfilling its stated objective. An individual study may not be directly applicable to the relevant guideline question however, for example if the evaluation was undertaken in a non-UK setting (see 2.5, p9).

We also suggest that the CHEC-list could be further simplified by removing items that are not directly relevant for our purposes (removing items 16, 17 and 19 from the original CHEC-list). The instrument is designed for appraisal of published studies in the context of a literature review. It thus includes questions relating to the thoroughness of reporting and the authors' interpretation of results – for example, item 19 asks if the authors have discussed ethical and distributional issues. However, the objective for critically appraising economic evaluations in the context of a clinical guideline

is not to assess how good the authors are at writing papers, but to determine whether the results are sufficiently reliable to inform recommendations. This should be made explicit to readers of the guideline in the methods section to avoid confusion.

The judgements that an individual economist makes against the modified CHEC list (and HTA modelling checklist if appropriate) should be recorded and presented in an appendix to the guideline. The 'other comments' column should be used to record reasons for these judgements and additional details where necessary. It is also essential for economists to continue to report the methods and results of economic evaluations conducted for the guideline in full.

It is proposed that economists should summarise the methodological quality of cost, cost-effectiveness or net benefit estimates in the economic profile table as:

- **Very serious limitations** – the study fails to meet one or more quality criteria and this is very likely to change the conclusions about cost-effectiveness. Studies with very serious limitations would usually be excluded from the economic profile table.

- **Potentially serious limitations** – the study fails to meet one or more quality criteria, and this could change the conclusions about cost-effectiveness

- **Minor limitations** – the study meets all quality criteria, or the study fails to meet one or more quality criteria, but this is unlikely to change the conclusions about cost-effectiveness

The robustness of the study results to methodological limitations may be demonstrated through sensitivity analysis. In the absence of relevant sensitivity analyses, however, judgement will be required to assess whether a limitation is likely to change the results.

The nature of any limitations identified should be briefly outlined in footnotes to the profile table, as in the standard GRADE profile.

Table 1 . Modified CHEC-list.
The QHES has been used to adapt the CHEC-list (highlighted items) allowing users to consider modelling studies within a single checklist while preserving syntactic simplicity.

Quality criteria	YES/ NO/ NA	Comments
1. Is the study population clearly described?		
2. Are competing alternatives clearly described?		
3. Is a well-defined research question posed in answerable form?		
4. Is the economic study design appropriate to the stated in the objective?		
5. Is the chosen time horizon appropriate in order to include relevant costs and consequences?		
6. Was the perspective of the analysis (societal, third-party payer, etc) clearly stated?		
7. Were the parameter estimates used in the analysis from the best available source?		
8. Are all important and relevant costs for each alternative identified?		
9. Are all costs measured appropriately in physical units?		
10. Are costs valued appropriately?		
11. Are all important and relevant outcomes for each alternative identified?		
12. Are all outcomes measured appropriately?		
13. Are outcomes valued appropriately?		
14. Is an incremental analysis of costs and outcomes of alternatives performed?		
15. Are all future costs and outcomes discounted appropriately?		
16. Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis?		
17. Was the modelling strategy appropriate given the research question?		
18. Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?		
Overall assessment: Very serious limitations/ Serious limitations/ Minor limitations		
Other comments:		

2.4 Consistency

In GRADE, this criterion relates to the similarity of estimates of effect across studies. When a meta-analysis has been conducted, a significant measure of heterogeneity that cannot be explained by appropriate subgroup analyses might indicate an important inconsistency in the evidence. Or in the absence of meta-analysis, the degree of consistency might be indicated by a wide range of estimates obtained from the available studies. In the presence of 'important inconsistency', the grade of evidence should be reduced by one.

The issue of inconsistency may also be important for economic evidence. However, In contrast to effectiveness evidence, pooling of cost effectiveness evidence from separate studies is not likely to be appropriate given the methodological differences between economic evaluations. It would be possible to summarise several economic evaluations on the same subject, by presenting a range of cost or cost-effectiveness estimates in a single row. However, it would be difficult to summarise the quality of such a body of economic evidence.

To aid transparency it is therefore proposed that under most circumstances only one economic study should be presented per row of the evidence profile. This is feasible for economic studies as there are usually only a few economic evaluations of sufficiently high quality to justify presentation in the profile. If a large number of relevant economic studies were to be identified for a guideline question, studies of lower quality could be excluded. Alternatively, a single row could be used to summarise a number of studies based on shared characteristics: these should be explicitly justified in a footnote.

We therefore do not consider it necessary for there to be a column within the economic profile considering 'consistency'. Instead, inconsistency between the results of economic studies will be shown by differences between rows of the economic evidence profile. For example, one might have quite different cost-effectiveness estimates from different industry or academic models. The implication of any such inconsistency depends on the circumstances, however. If two good-quality models were to give different results, then that might increase uncertainty about the cost-effectiveness of the intervention, whereas differences between estimates are not necessarily important if they derive from studies of variable quality or intended for different contexts.

2.5 Applicability (called 'directness' in GRADE)

In GRADE, this criterion reflects the degree to which the study results are appropriate for the decision-making situation. This entails a judgement about the relevance of the study populations, interventions and outcome measures. The grade of evidence is reduced by one point if there is 'some uncertainty' about directness, or by two points if there is 'major uncertainty' about directness.

The concept of applicability is also very important for economic evidence. The results of an economic evaluation conducted in another health care system

may not be applicable for the NHS. The appropriateness of the patient population and the nature of the intervention should also be considered. Thus 'applicability' relates to both the relevance of the study results to the guideline question under consideration (defined by population, interventions and outcome measures), as well as the general framework underpinning NICE decision making. The latter is defined in the NICE 'reference case' for economic evaluations [26], as well as the broader considerations in the NICE Social Value Judgements document [27].

It is proposed that 'Applicability' should be assessed using the checklist in table 2. An overall judgement on applicability should be summarised in the evidence profile by categorising the study as:

- **Directly applicable** – the applicability criteria are met, or one or more criteria are not met but this is not likely to change the cost-effectiveness conclusions

- **Partially applicable** – one or more of the applicability criteria are not met, and this might possibly change the cost-effectiveness conclusions

- **Not applicable** - one or more of the applicability criteria are not met, and this is likely to change the cost-effectiveness conclusions

2.6 Summary of findings

The results section of the GRADE profile table also needs adaptation for economic evidence. This could include, for example, columns for incremental cost (£), incremental effects (measured in QALYs when available), and the ICER where relevant.

The results of sensitivity analyses should also be shown in a column for 'uncertainty'. Uncertainty could be reported in various ways, depending on the type of sensitivity analysis used. For example, the results of a probabilistic sensitivity analysis could be reported as an interval for the ICER or as the probability of cost-effectiveness (for a given cost-effectiveness threshold, say £20,000 per QALY). For deterministic sensitivity analyses a range for the ICER could be given.

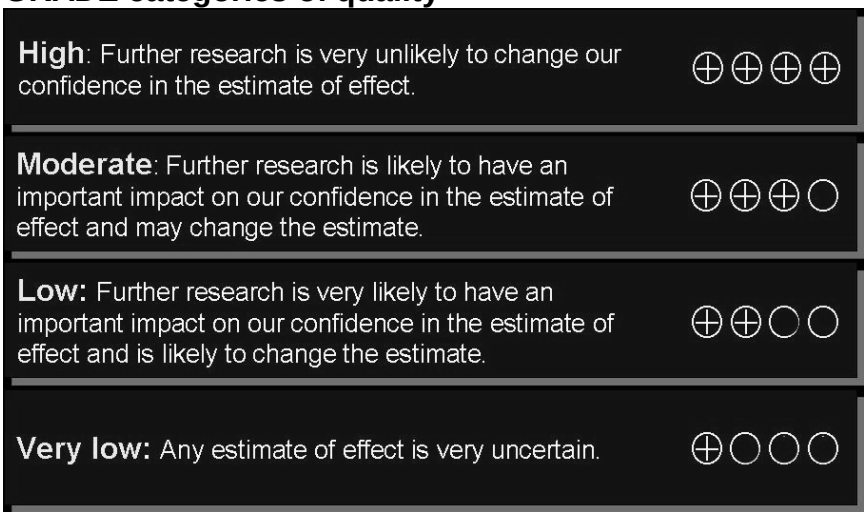
Table 2. Applicability checklist

Applicability criteria	Yes/ No/NA	Comments
Population – is the patient population covered in the economic analysis fully in line with the population specified in the guideline question?		
Intervention – is the intervention considered by the economic study the same as that specified in the guideline question?		
Comparators – are the comparators in this study entirely relevant for the guideline question under consideration?		
Outcomes – are all important health effects appropriately considered by the study?		
BASED ON 'NICE REFERENCE CASE' As detailed in the 'Guide to the Methods of Technology Appraisal' (April 2004), Box 5.1 (page 21). Note: this Guide is currently under review. As section 5.3.1.3 of the Guide states: "It is recognised that in some instances, data required to present reference case results are not available. Similarly, there may be important barriers to applying reference case methods". Under these circumstances deviations from this reference case should be clearly specified.		
Perspective – In the base case analysis, has a UK NHS and PSS perspective been taken?		
Discounting – Are both costs and benefits discounted at 3.5%?		
Measure of health benefits – are QALYs used and presented?		
Description of health states for calculating QALYs – have they been described using a standardised generic instrument?		
Method of preference elicitation for health state valuation – has a choice-based method been used?		
Source of preference data – has this been provided from a representative sample of the public?		
Overall judgement: Directly applicable/ Partially applicable/ Not applicable		
Other comments:		

2.7 Overall quality rating

The final column of the GRADE profile shows the quality rating for each outcome, measured using a four point scale (see figure below). This is calculated by combining the four criteria of study design, limitations, consistency and directness (applicability).

GRADE categories of quality



A similar approach could possibly be used for economic studies; however it is not apparent how the criteria described in section 2.3 could be summarised into a single quality rating. It would be possible to make a subjective judgement, using an adaptation of the GRADE quality definitions above. This rating should depend on an overall assessment of the impact of any limitations in study quality, applicability, the extent of uncertainty, and consistency in results across a number of studies if applicable.

The current proposal is not to apply an overall quality rating to the economic evidence.

Table 3: Example of evidence profile
Clinical question: Omega 3 acid ethyl ester supplements vs. control in people within 3 months of an acute MI

Quality assessment							Summary of findings				
Clinical evidence											
Outcome	*No. of studies*	*Design*	*Limitations*	*Inconsistency*	*Directness*	*Other considerations*	*Intervention*	*Control*	*Relative Risks*	*Risk Difference*	*Quality*
All cause mortality	3	RCT	Serious [a]	No important inconsistency	Some uncertainty [b]	None	581/6830	755/6830	0.83 (0.75 to 0.93)	-0.02 (-0.03 to -0.01)	Low
Combined cardiovascular events	3	RCT	Serious [a]	No important inconsistency	Some uncertainty [b]	None	755/6830	839/6826	0.90 (0.82 to 0.99)	-0.01 (-0.02 to 0.00)	Low
Cancers	3	RCT	Serious [a]	No important inconsistency	Some uncertainty [b]	None	150/6830	138/6826	1.09 (0.86 to 1.36)	0.00 (0.00 to 0.01)	Low
Economic Evidence											
Study	*Limitations*	*Applicability*	*Other comments*				*Incremental cost (2006 £)*	*Incremental effects*	*ICER*	*Uncertainty*	
Franzosi 2001	Potentially serious limitations [c]	Partially applicable [d]	Based only on measured resource use and survival in 3.5 years follow-up in GISSI-P.				£871 [e]	0.0332 LYs	£26,243 per LY gained	£16,769 to £56,025 per LY gained (best / worst case)	
Lamotte 2006	Very serious limitations [f]	Partially applicable [g]	Based on measured resource use and survival over 3.5 years in GISSI-P, plus longer term survival benefits attributed to nonfatal events using Canadian database. Belgian results presented.				£ 1,090 [h]	0.282 LYs	£3,860 per LY gained	>98% probability ICER less than €20,000 per QALY gained	
NCC analysis	Minor limitations [i]	Directly applicable [j]	Based on morbidity and mortality estimated from Markov model using pooled effectiveness data from GISSI-P and DART. Results were sensitive to the size of treatment effects and over their assumed duration.				£1,073	0.09 QALYs	£12,480 per QALY gained	£3,912 to £130,705 per QALY gained (Range in one-way sensitivity analyses)	

[a] Increase in statin use over follow-up in GISSI-P differed between the groups (from 4.4% to 46.0% in the omega-3 group and from 5.1% to 44.4% in the control group).
[b] High baseline rate of fish consumption in GISSI (more than 70%).
[c] This study is relatively conservative, as it does not impute any quality of life or longer-term survival benefit to supplements. Although, conversely it omits GI side effects.
[d] Some uncertainty over the applicability of Italian trial data to UK. May be differences in population risk and diet as well as health care use and unit costs.
[e] Converted from 1999 Italian Euros using a PPP exchange rate of 0.797 (http://www.oecd.org/std/ppp) then uprated by inflation factor of 133.8% (http://www.pssru.ac.uk/pdf/uc/uc2006/uc2006.pdf).
[f] Methods and data used to estimate life expectancy are questionable, and were not subjected to sensitivity analysis. This is likely to have biased the results.
[g] Some uncertainty over the applicability of Italian trial data to UK. May be differences in population risk and diet as well as health care use. Unit costs may also differ for UK.
[h] Converted from 2004 Belgian Euros using a PPP conversion rate of 0.706 (http://www.oecd.org/std/ppp) then uprated by inflation factor of 107.3% (http://www.pssru.ac.uk/pdf/uc/uc2006/uc2006.pdf)
[i] Some limitations in reporting (eg for inputs taken from NICE statins appraisal). However, analysis is based on best-available effectiveness estimates and follows NICE methodological guidance. The robustness of results is also well tested through sensitivity analysis and comparison with other study results.
[j] Some uncertainty over applicability of trial data to UK due to differences in population risk and diet. However, resource use and unit costs are UK-specific and the perspective and discount rates follow the NICE reference case.

2.8 Example: Omega 3 supplements for secondary prevention of MI

An example of a combined clinical and economic evidence profile is presented in Table 3. This relates to a case from the NICE guideline on secondary prevention of myocardial infarction (MI) [8]. The guideline group reviewed the clinical evidence on advice to eat oily fish, and concluded that there was sufficient evidence to recommend this in patients who had recently had an acute MI (within the last three months). This led to the question of whether patients who cannot achieve the recommended dietary intake should be given supplements, and whether these supplements would be a cost-effective use of NHS resources. The table above shows a simplified summary of the evidence available to the GDG for this question.

The clinical evidence is taken from the Cochrane review for the three RCTs which included use of supplements in patients with recent MI [21]. Here we have presented the results for only three primary outcomes, although the guideline group did consider others. A longer-term follow-up study (DART2) was excluded as the GDG felt that it had significant limitations. The guideline group had information from two published economic evaluations, and from two confidential reports submitted by manufacturers of the supplements (not presented in the table). However, they felt that these studies were insufficient and conducted their own analysis for the guideline.

The guideline group concluded that there was evidence of a significant reduction in mortality in this subgroup of patients, and that this was unlikely to be outweighed by any harms of treatment. The three economic studies were of variable quality and applicability for the NHS. However, although the results differed, the models gave similar results when assumptions about the size and duration of survival benefits were the same. The GDG concluded that supplements were likely to be cost-effective and recommended their use for post-MI patients with insufficient dietary intake of omega-3 oils.

3 Discussion

There are a number of other practical issues related to the implementation of our proposed approach for presenting economic evidence alongside GRADE profiles for clinical evidence:

3.1 Absence of cost-effectiveness estimate

For some guideline questions there might not be an estimate of cost-effectiveness. This will happen when there is no adequate published estimate and the topic is not judged to be of sufficient priority for *de novo* modelling by the guideline economist. In this situation, the GDG is still expected to make a judgement on the broad balance of costs, benefits and harms – a form of cost-consequence analysis. In such cases an estimate of cost should be obtained, and included in the economic section of the GRADE profile.

There may also be some guideline questions, where the cost implications of the question are considered to be negligible. In the absence of any cost or cost-effectiveness information, it would aid transparency if this were noted in the economic section of the GRADE profile, which would otherwise be left blank.

3.2 Subgroup analyses

There may be complications where results are available for patient subgroups. This could be handled by presenting the results for each subgroup in a separate row or table. When different estimates of treatment effect are available for the subgroups, it may be better to present both clinical and economic estimates for each subgroup in a separate table. However, sometimes there may be differences in estimated costs or cost-effectiveness by subgroup despite there being only a single estimate of treatment effect. This could arise, for example, due to differences in baseline risk or longevity by age. In such cases it might be better to present a single GRADE table for the clinical data, with a separate row for each subgroup in the economic section. Where results are calculated for a continuum of risk, it may be simpler to present the risk threshold at which treatment becomes cost-effective (i.e. where the ICER falls below £20,000 per QALY).

3.3 The relationship between clinical questions, evidence & recommendations

There is not always a simple one-to-one correspondence between clinical questions, a body of evidence, and a set of recommendations. This can pose challenges for the organisation of the full guideline document. It may also be a particular problem for economic analyses, which often combine information on diagnosis or assessment, effectiveness and prognosis from several clinical questions. GRADE tables will only be available for effectiveness questions, however. Thus a single economic analysis might relate to no, one or several clinical GRADE profiles. This suggests that some flexibility will be required when deciding where best to present economic evidence.

3.4 Multiple comparisons

Each comparison needs a separate GRADE table, which becomes cumbersome with several treatment options. NICE is considering use of a top-level summary table to show key elements of the GRADE profile for all relevant comparisons in one place.

There are two additional considerations for economic evidence:

- It will not always be appropriate to present comparative economic results (e.g. an incremental cost-effectiveness ratio) for all pairs of interventions for which there is direct comparative clinical data – each option should only be compared with the next most expensive, non-dominated option.

- Conversely, it may sometimes be appropriate to present comparative economic results in the absence of direct comparative clinical data. This will arise, for example, when a mixed treatment comparison has been used to obtain the effectiveness estimates in an economic model.

The summary table might thus have blank cells to show where clinical or economic evidence is missing, but it might not always be feasible to show all direct and indirect comparisons within a single table.

3.5 Multiple outcomes

When there is proliferation of outcome measures in the clinical literature, the GDG will have to be selective about which outcomes to include in order to prevent GRADE evidence profiles becoming excessively long. This decision should always be made before viewing study results to avoid bias, but GDGs do often find this difficult.

The presence of multiple clinical indicators is not a problem for economic evidence when health outcomes have been summarised in a single measure (such as QALYs, a monetary estimate of 'willingness-to-pay' or some key clinical indicator such as 'life years saved'). However, this is not always possible. An alternative might be to present an estimate of cost alongside the clinical measures in the GRADE profile (cost-consequence analysis). In such cases, the GDG might find it harder to make meaningful decisions if there are multiple outcome measures to be traded off.

3.6 Missing outcomes

Clinical evidence of sufficient quality might not always be available for every outcome measure. When this happens should a blank row be retained in the GRADE profile to indicate missing data for the outcome? This has the advantage of transparency, although it could be cumbersome for some guidelines.

By analogy, a blank row could be inserted in an economic evidence profile when no estimates of cost, cost-effectiveness or net benefit are available. This row could contain a brief statement noting whether economic evidence was searched for, or whether it was not thought to be relevant or of sufficient importance for the question.

4 Future work on developing the GRADE economic profile

The proposals set out in this paper are currently being piloted for their feasibility and usefulness before an economic profile approach is implemented across the NICE guidelines programme.

In the first instance NCC economists have been asked to apply the critical appraisal checklist and economic evidence profile to three published economic evaluations, selected from studies that they are reviewing in their ongoing guidelines. Participants in this feasibility study are asked to complete a questionnaire on their perceptions of the usefulness of the proposed checklists and economic evidence profile and to make suggestions for improvement.

The approach recommended in this paper will also be tested prospectively on the economic evidence for some guidelines chosen for the pilot of the GRADE system. This will test the use of economic evidence profiles alongside GRADE clinical evidence profiles in the context of a full guideline development process. For the pilot guidelines all economic evaluations used to inform guideline recommendations will be critically appraised with the proposed checklist, and summarised and presented to the GDG in the proposed economic evidence profile. The economists working on these guidelines will be asked to document any problems experienced and any proposed modifications. The results from this assessment will not be available until the guidelines are near publication.

More formal, testing of the validity or reliability of the instrument could be conducted if considered necessary. This could include a test for convergent and discriminant validity on a larger sample of economic evaluations, using methods analogous to those used in the validation study for the QHES instrument [7,30].

Acknowledgements: We would like to thank the following people who have participated in discussions and commented on drafts of this paper: Raquel Aguiar-Ibañez, Phil Alderson, Françoise Cluzeau, Kalipso Chalkidou, Stefanie Kuntze, Ifigeneia Mavranezouli, Jeff Round, Sarah Willis, David Wonderling and other members of the Health Economists in Guidelines group.

References

1. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Medical Care* 1992;**30**:231-43.

2. Ament A, Evers S, Goossens M. Criteria list for conducting systematic reviews based on economic evaluation studies - the CHEC project. In Donaldson C, Mugford M, Vale L, eds. *Evidence-based health economics* , London: BMJ Books, 2002.

3. Blackmore CC,.Magid DJ. Methodologic evaluation of the radiology cost-effectiveness literature. *Radiology* 1997;**203** :87-91.

4. Bradley CA, Iskedjian M, Lanetot KL. Quality assessment of economic evaluations in selected pharmacy, medical and health economics journals. *Annals of Pharmacotherapy* 1995;**29**:689.

5. Buxton MJ, Drummond MF, van Hout BA, Prince RL, Sheldon TA, Szucs T *et al*. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997;**6**:217-27.

6. Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. CRD Report 4 (second edition). 2001. University of York.

7. Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD *et al*. Development and validation of a grading system for the quality of cost-effectiveness studies. *Medical Care* 2003;**41** :32-44.

8. Cooper, A., Skinner, J., Nherera, L., Feder, G., Ritchie, G., Kathoria, M., Turnbull, N., Shaw, G., MacDermott, K., Minhas, R., Packham, C., Squires, H., Thomson, D., Timmis, A., Walsh, J., Williams, H., and White, A. Clinical Guidelines and Evidence Review for Post Myocardial Infarction: Secondary prevention in primary and secondary care for patients following a myocardial infarction. 2007. London, National Collaborating Centre for Primary Care and Royal College of General Practitioners .

9. Drummond MF, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* 1996;**313**:275-83.

10. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1997.

11. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1987.

12. Evers SMAA, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care* 2005;**21** :240-5.

13. Evers, S. M. A. A., Wijk, A. S., and Ament, A. J. H. A. Economic evaluations of mental health care interventions. A review. 1994. University of Maastricht, Department of Health Economics .

14. Ganiats TG,.Wong AF. Evaluation of cost-effectiveness research. A survey of recent publications. *Family Medicine* 1991;**23**:457-61.

15. Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. *Health Policy* 1992;**21**:249-79.

16. Gerard K, Seymour J, Smoker I. A tool to improve quality of reporting published economic analyses. *Int J Technol Assess Health Care* 2000;**16** :100-10.

17. Gonzales-Perez J. Developing a scoring system to quality assess economic evaluations. *European Journal of Health Economics* 2002;**3** :131-6.

18. GRADE working group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490-8.

19. Heyland DK, Kernerman P, Gafni A, Cook DJ. Economic evaluations in the critical care literature: do they help us improve the efficiency of our unit? *Critical Care Medicine* 1996;**24**:1591-8.

20. Holloway RG,.Benesch CG. Systematic review of cost-effectiveness research of stroke evaluation and treatment. *Stroke* 1999;**30**:1340-9.

21. Hooper L, Thompson RL, Harrison RA, Summerbell CD, Moore H, Worthington HV *et al*. Omega 3 fatty acids for prevention and treatment of cardiovascular disease. *Cochrane Database of Systematic Reviews* 2004;**4**.

22. Huijsman R. Economic evaluation of care for the chronically ill. A literature review. *EJPH* 1995;**5**:8-19.

23. Iskedjian M, Trakas K, Bradley CA, Addis A, Lanctot K, Kruk D *et al*. Quality assessment of economic evaluations published in PharmacoEconomics. The first four years (1992 to 1995). *Pharmacoeconomics* 1997;**12**:685-94.

24. Maetzel A, Ferraz MB, Bombardier C. A review of cost-effectiveness analyses in rheumatology and related disciplines. *Current Opinion in Rheumatology* 1998;**10**:136-40.

25. McCabe C,.Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics* 2000;**17**:501-13.

26. National Institute for Clinical Excellence. Guidance for Manufacturers and Sponsors. 1-6-2001. Technology Appraisals Process Series No 5, NICE, London, available online from http://www.nice.org.uk/pdf/technicalguidanceformanufacturersandsponsors.pdf; accessed 3 October 2002.

27. National Institute for Clinical Excellence. Social Value Judgement - Principles for the development of NICE guidance. 2005. London.

28. National Institute for Clinical Excellence. The Guidelines Manual. 2007. London.

29. Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976-1997. *Ann Intern Med* 2000;**132**:964-72.

30. Ofman JJ, Sullivan SD, Neumann PJ, Chiou CF, Henning JM, Wade SW *et al*. Examing the value and quality of health economic analyses: implications of utilizing the QHES. *Journal of Managed Care Pharmacy* 2003;**9** :53-61.

31. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R *et al*. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004;**8**.

32. Sacristan JA, Soto J, Galende I. Evaluation of pharmcoeconomic studies: utilization of a checklist. *Ann Pharmacother* 1993;**27**:1126-33.

33. Sanchez LA. Evaluating the quality of published pharmacoeconomic evaluations. *Hospital Pharmacy* 1995;**30**:146.

34. Sculpher MJ, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models - A suggested framework and example of application. *Pharmacoeconomics* 2000;**17**:461-77.

35. Severens JL,.van der Wilt GJ. Economic evaluation of diagnostic tests. A review of published studies. *Int J Technol Assess Health Care* 1999;**15**:480-96.

36. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. *JAMA* 1996;**276**:1339-41.

37. Soto J. Health economic evaluations using decision analytic modelling: principles and practices - utilization of a checklist to their development and appraisal. *Int J Technol Assess Health Care* 2002;**18**:94-111.

38. Udvarhelyi IS, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature. Are the methods being used correctly? *Ann Intern Med* 1992;**116**:238-44.

39. Weinstein MC, O'Brien BJ, Hornberger J, Jackson J, Johannesson M, McCabe C *et al*. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices-Modeling Studies. *Value Health* 2003;**6**:9-17.

40. Weinstein MC, Siegel JE, Gold MR, Kamlet MS. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA* 1996;**276**:1253-8.