Developing valid cost effectiveness guidelines: a methodological report from the North of England evidence based guideline development project

Martin Eccles, James Mason, Nick Freemantle

Introduction
Over the last decade clinical practice guidelines have become an increasingly familiar part of clinical care. Defined as “systematically developed statements to assist both practitioner and patient decisions in specific circumstances”, they are viewed as useful tools for making care more consistent and efficient and for closing the gap between what clinicians do and what scientific evidence supports. The broad interest in clinical guidelines is international and has its origin in issues that most healthcare systems face: rising healthcare costs; variations in service delivery with the presumption that at least some of this variation stems from inappropriate care; and the intrinsic desire of healthcare professionals to offer, and patients to receive, the best care possible. Within the UK there is ongoing interest in the development of guidelines and a fast developing clinical effectiveness agenda within which guidelines figure prominently.

During the same 10 year period the methods of developing guidelines have steadily improved, moving from solely consensus methods to methods that take explicit account of relevant evidence. This improvement should make guidelines more valid; guidelines are valid if “when followed, they lead to the improvements in health status and costs predicted by them”. In order to maximise validity, three areas of the guideline development process are important:

1. Identification and synthesis of the evidence should be done using the methods of systematic review to maximise the appropriate identification of evidence;
2. The guideline development group should be appropriately multidisciplinary to ensure full discussion of relevant evidence, associated service delivery issues, and the appropriate construction of recommendations;
3. The recommendations in the guideline should be clearly and explicitly linked to the evidence supporting them.

To date, however, most guidelines have taken a relatively narrow view of evidence, focusing predominantly on effectiveness issues. While such a focus is necessary to define the effect of a drug or surgical intervention, it is not sufficient to provide the breadth of evidence that guideline groups need to derive recommendations for clinical care. This is exemplified by a general practitioner in a guideline group who commented that “many of the drug trials don’t actually measure the sort of end points that you might actually want to know about”. Members of the group wanted evidence on a wider range of end points including, in addition to effectiveness, the potential harms and side effects of treatment and effects on quality of life. However, they fell short of suggesting monetary cost.

The introduction of cost considerations into guidelines has been proposed by a number of workers and is important because re-
sources for health care are not unlimited and, when considering clinical decisions, the effect on resources may influence recommendations. The Committee on Clinical Practice Guidelines in the USA1 recommended that every set of clinical practice guidelines should include information on the cost implications of alternative preventive, diagnostic, and management strategies for the clinical situation in question. They argued that this information could help potential users to evaluate better the potential consequences of different practices. However, they then went on to state: “the reality is that this recommendation poses major methodological and practical challenges” as “scientific evidence about benefits and harms is incomplete and accurate cost data are scarce for the great majority of clinical conditions and services. In addition, while data on charges may be available, significant analytic steps and assumptions are required to treat charge data as cost data and techniques for analysing and projecting costs and cost effectiveness are complex and only evolving”.

We have explored the issues concerning incorporation of costs into guidelines during the development of four evidence based guidelines dealing with drug use in primary care. This work was part of a randomised controlled trial evaluating the effectiveness of outreach visiting within UK primary care.20 Outreach visiting is defined as “the use of a trained person who meets with providers in their practice settings to provide information with the intent of changing the provider’s performance”.21 It has primarily been used to influence prescribing, but little of the evidence on its effectiveness comes from the UK. The four evidence based guidelines were developed to provide the content of the educational messages to be delivered by the outreach visitors. The subject areas of the four guidelines (box 1) were chosen by the research team designing the trial as areas where there was a mismatch between the evidence of effectiveness and current practice. In addition, quality of life, potential harm and drug and service costs were potentially important considerations in deriving recommendations for clinical care. Although they address different clinical areas, the four guidelines addressed the common methodological challenge of incorporating cost issues within their consideration of the evidence, thus making them cost effectiveness guidelines. We have previously described the economic elements of this process.22 In this paper we report on the methodology issues raised in the development of evidence based clinical practice guidelines, including evidence on cost effectiveness.

Guideline development groups
To permit appropriate identification and discussion of the evidence and the appropriate construction of recommendations, guideline development groups should have appropriate multidisciplinary membership.23 Since the subject area of the guidelines was drug use in primary care, we included up to five general practitioners, up to two secondary care specialists in the clinical area relevant to the drugs, a medical or pharmaceutical adviser, and a pharmacist. Consumers were not included. The groups were supported by a health services researcher (NF) and a health economist (JM) and were led by a guideline methodologist (ME).

Two guideline groups were convened, each of which developed two guidelines. Four three-hour weekday afternoon meetings were held for each guideline. Whilst it would have been possible to convene four separate groups, it was more efficient to use the same group to develop two guidelines. In the early stages the group members learned about both evidence synthesis and health economics. The early meetings therefore contained common economic components on evidence synthesis and health economics that were relevant to both guideline topics. However, because of the change of clinical area from the first to their second guideline in each group, the secondary care specialists changed after the fourth meeting whilst the rest of the membership stayed constant. The groups handled this change of personnel without any obvious problem.

The groups reached decisions about the interpretation of the evidence by informal consensus. While the use of formal decision making processes within guideline development has been described,24 it is still largely experimental. As the guideline development process allowed explicit definition of certainty and uncertainty, there were no strong disagreements within the process and informal methods were successful.

Defining the area of the guidelines
The general topic areas of the guideline (box 1) were already defined from within the trial protocol. This is a generalisable situation in that guidelines groups will be convened to develop specified guidelines rather than being convened and then asked what guidelines they would like to develop. However, at the first meeting for each guideline the groups were asked to define both the clinical content of the areas of the guideline and the scope of the questions to be answered within it. Thus, in considering the use of ACE inhibitors in the management of patients with heart failure, as well as defining the process of clinical care for patients with heart failure the group also identified the potential relevance of evidence on the quality of life of patients and the impact of knock on consequences of ACE inhibitor use on broader resource use such as general practice consulta-

- The choice of first line antidepressants for the drug treatment of depression
- Aspirin for the secondary prophylaxis of vascular disease
- ACE inhibitors in the management of adults with symptomatic heart failure
- Non-steroidal anti-inflammatory drugs (NSAIDs) versus basic analgesia in the treatment of pain believed to be due to degenerative arthritis

Box 1 Topic areas of the guidelines. All four guidelines were intended for use in primary care.
Developing valid cost effectiveness guidelines

Box 2 Examples of the meta-analytical presentations used within the aspirin guideline

<table>
<thead>
<tr>
<th>Metric</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratio</td>
<td>0.67</td>
</tr>
<tr>
<td>Risk difference</td>
<td>−4.5%</td>
</tr>
<tr>
<td>Incidence risk difference (year)</td>
<td>−0.7%</td>
</tr>
</tbody>
</table>

only with a quantified point estimate of effectiveness and the uncertainty associated with it that other (quantifiable) assumptions could be incorporated into the evidence from which to derive recommendations.

Whilst the group members were all familiar with the general concept of meta-analysis, it was necessary to spend about half of the first meeting explaining the processes involved in handling the data and producing summary statistics of effectiveness. Having the researcher responsible for this within the group enabled the group members to ask specific questions that could be answered either immediately or in subsequent meetings. For example, the group that looked at the use of aspirin as secondary prophylaxis for vascular disease accepted the overall estimate of effect but then wished to know if this effect (or the side effects) was dose related. This analysis was presented to them at the next meeting (box 2).

As well as encouraging such clinical relevance of evidence summaries—for example, effectiveness and tolerability—the interaction within the guidelines group greatly enhanced the practical interpretation of the quantitative summaries. The issue of heterogeneity within included studies was routinely examined and discussed with the group, enabling them to understand the likely limits on the interpretability of the data resulting from problems such as study populations. This involved interpretation of the pooled odds ratio. While pooled odds ratios or standardised effects sizes may be the most statistically robust presentation of the results of a series of trials, they are not necessarily readily interpretable in terms of their clinical meaning. In order to gauge the importance of an intervention, a measure of relative effect such as an odds ratio needs to be considered alongside an indication of absolute effect. However, within the meta-analyses trials were of differing lengths of follow up and the statistical method of incidence risk differences was used to allow an estimate of effect to be expressed over a common time period. Thus, the most statistically robust analyses were presented alongside those that used less statistically robust, but more clinically interpretable, summary values. An example of this for the use of aspirin as an antithrombotic agent in patients with stable angina is shown in table 1. The odds ratio is 0.67 and the risk reduction from the meta-analysis is 4.5%. However, when this is adjusted for a standard time period of one year the incidence risk difference is a smaller reduction of 0.7%.

In addition to effectiveness, most of the included trials reported common outcomes on side effects and withdrawals. However, the use of quality of life measures as study outcomes was less frequent. In the trials examining the effectiveness of antidepressants, measurement
of mental state was one of the main outcomes and other measures of social functioning were often reported. In the aspirin trials quality of life data were rarely reported. In the ACE inhibitor trials the use of quality of life measures as study outcomes was again infrequent but the largest trial, which dominated the effectiveness analysis, did report such data. As the study population was felt to be representative of UK primary care, this study was used as the source of quality of life data for that guideline.

### Summarising evidence on cost

Although in earlier guidelines the guideline development groups had made cost related recommendations, these had not been formally based on the evidence presented. Rather, a health economist reviewed and presented primary health economic studies within the clinical area of the guideline. In the current guidelines the aim was to develop cost effectiveness based recommendations derived from discussion in the group of a profile of the attributes of treatments (box 3). This approach led to simple presentations of costs and consequences that were readily replicable by readers of guidelines (table 2). The data on which these presentations were based were not always particularly robust. However, by explicitly identifying such issues, the evidence more accurately represented its strengths and weaknesses, and end users of the guideline could easily explore alternative values.

**Box 3 Areas for systematic appraisal of alternative treatment options in guidelines.**

As with meta-analysis, members of the guideline development group were familiar with the general concepts of health economics but a proportion of the first meeting was devoted to an explanation of the principles of economic analyses and its supportive (rather than prescriptive) role in guideline development. An advantage of this approach was that economic aspects were developed alongside the evidence of attributes of effectiveness developed for the guideline, and served to pull all the various strands of evidence together in an overall profile to facilitate deriving recommendations.

**Nature of evidence informing recommendations and the influence of costs**

In the guideline groups the recommendations were based on the basis of evidence about effectiveness, quality of life, compliance, safety, delivery issues, resource use, and costs tempered by knowledge of the generalisability of

<p>| Table 2 | Presentation of costs and consequences from the use of ACE inhibitors in the treatment of heart failure. Net cost and benefit per patient of ACE inhibitors for heart failure |</p>
<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Optimistic</th>
<th>Conservative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor</td>
<td>£400</td>
<td>£1400</td>
</tr>
<tr>
<td>£100/year or £340/year for four years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiation of therapy by two GP visits or two outpatient visits</td>
<td>£20</td>
<td>£138</td>
</tr>
<tr>
<td>Reduced hospitalisation or no reduced hospitalisation</td>
<td>£626</td>
<td>£0</td>
</tr>
<tr>
<td>GP visits related to heart failure unchanged or one extra visit/year for four years</td>
<td>£0</td>
<td>£48</td>
</tr>
<tr>
<td>Net cost range</td>
<td>£206</td>
<td>£1578</td>
</tr>
<tr>
<td>Increased life expectancy (based on placebo comparison)</td>
<td>0.203 years</td>
<td>0.203 years</td>
</tr>
<tr>
<td>Incremental cost effectiveness of implementing ACE inhibitor therapy</td>
<td>Small cost saving and health gain</td>
<td>£7770/life year gained</td>
</tr>
</tbody>
</table>

*Costs and benefits are shown which arise from the addition of ACE inhibitors to current care. Diagnosis costs are excluded because of the variation in tests performed, the lack of adequate cost data, and because these costs may occur in any case as part of normal care. Costs presented here are overestimates as withdrawal from treatment has not, for simplicity, been included.

*Cost per:

- GP consultation £10 (excluding prescribing cost).
- Outpatient visit £69.
- Costs of additional blood tests are excluded as no adequate cost data were found.

*Calculation based on:

- Difference in Studies of Left Ventricular Dysfunction (SOLVD) trial treatment and control hospitalisation rates (21.9–15.4%) × 4 years. Inpatient stay of 14.5 days.

*Assumptions:

- Since patients visit their GPs on average once a year in relation to heart failure, it is not plausible to assume an optimistic reduction in GP visits although treatment does delay disease progression and associated morbidity.

- Based on the placebo controlled findings of the SOLVD treatment trial, improved survival was highly statistically significant (p = 0.0036, stratified log rank test). However, the survival gain calculation (using Irwin’s restricted mean) does not provide a useful confidence interval. The point estimate is thus used in optimistic and conservative scenarios. Future costs and benefits are not discounted because of the short four year time frame and because all important costs are distributed along with the benefit in time. Discounting will not substantially alter the cost effectiveness ratios.

- Survival gains are truncated in the SOLVD trial and it is reasonable to presume that, if treatment stopped, there would be some additional benefit after cessation of therapy. However, this is not modelled since it is probable that treatment would continue and so both costs and benefits would occur after four years.

- These data to the British healthcare setting.

The grades attached were determined by the overall quality of evidence as interpreted by the group. The consideration of a range of evidence that went beyond efficacy allowed the issues underlying the recommendations to be more clearly defined.

In the guideline on the use of ACE inhibitors in patients with heart failure the drugs were stated to be effective in both prolonging life and improving heart failure symptoms but did not have a demonstrable effect on quality of life. The recommendations were therefore
derived in the light of the question “should increased costs be incurred (and if so how much) to extend life and symptomatically improve the life of patients with heart failure?”

In the group developing the guideline for the first line choice of antidepressants the differences between two major drug groups were being examined. The two groups were shown to have equivalent efficacy, small differences in side effects, large differences in toxicity in overdose, and large differences in drug acquisition costs. Thus, the underlying question informing the recommendations became “how much should be paid to avoid accidental or non-accidental fatal overdoses with antidepressants?”

Although the broad scope of the guidelines was defined during the initial meetings of the guideline groups, the precise nature of these questions was not apparent at the outset. It was only possible to pose these much more focused questions once the evidence review had been completed and the issues being considered had been refined.

**Discussion**

We have described a method of evidence based guideline development that involves more sophisticated methods of considering evidence than has previously been achieved. This has included explicit consideration of a broad range of harm and cost issues and summarising the effectiveness using meta-analysis.

We made extensive use of cost data and developed a health economic methodology for use within the guidelines. The method has relied upon being able to set quantitative evidence of effectiveness alongside data on harm and cost. This has allowed an explicit value to be attached to a profile of relative benefits of alternative treatment strategies.

Categories of evidence have been developed on the basis of the susceptibility of differing trial designs to a range of biases. What constitutes best evidence in resource usage (and thus cost) is debatable. The current taxonomies for evidence, suited to evaluating treatments, only take account of experimental studies of various sorts of professional opinion. However, it is inevitable that, as what is considered as evidence moves from estimates of effectiveness derived from randomised controlled trials to resource use data, the quality of the data will tend to fall. New taxonomies need to be developed that not only adequately categorise data from other study designs—for example, findings from prognostic or diagnostic studies—but also take account of evidence on issues such as harm, quality of life, and cost. This will be important to avoid the risk of inappropriate downgrading of recommendations.

We explicitly excluded consumer representation from the guideline group. Finding appropriate means to take consumer views into account is recognised to be an unresolved methodological issue.

Our experience and that of other groups has been disappointing when trying to include a “token” patient representative. Instead, we assumed that the objective of guidelines produced was to inform the doctor-patient relationship about the best available evidence on the various attributes of treatment, at which point patient specific information must always be introduced into the decision making process. This may be a valid stance when considering the role of individual drugs but may be less supportable when considering packages of care for “whole” diseases. However, in these circumstances, rather than direct group involvement, it may be more productive for patients to contribute to the content of a guideline through carefully constituted focus groups or some similar forum. This would entail patients (or their representatives) exploring the available evidence with a trained facilitator, with comments fed back to the guideline process. This type of activity was outside the scope of our work but may be adopted in ongoing national guidelines.

The use of meta-analysis worked well within a focused clinical area where the evidence base for effectiveness is entirely composed of randomised controlled trials that allow the identification and extraction of common end points. It is unclear whether or not these methods can be applied across broader clinical areas (such as diabetes or ischaemic heart disease) where such homogeneity of study methods, patient groups, and outcomes may be absent.

We have further developed the methods of developing evidence based guidelines. However, if guideline development methods are to improve along the lines that we have used, there will be a need for training of group members and for resources to be made available to guideline development groups. Training of members of the guideline development group in meta-analysis and health economics was addressed by setting aside time within the guideline development process, both initially and on an ongoing basis as the work proceeded. Without that it would not have been possible for the group to understand the issues they were being asked to deal with. In addition, the research skills that were required to support the guideline development process may not be widely available. These encompassed both technical competence in the areas of meta-analysis and health economics and also the ability to interact with practising clinicians. Such skills require appropriate resources to be made available to allow guideline development to continue to improve. Finally, it is not yet clear how applicable these methods are beyond this rather focused area of application. The challenge now becomes to apply the rigour of these methods to broader clinical areas, where the evidence summarising tasks will be that much more complex. However, such challenges have to be met if guideline development methods are to keep pace with the increasingly sophisticated demands of their potential users.

29 Netten A, Dennett J. Unit costs of health and social care. Personal Social Services Research Unit (PSSRU), University of Cambridge, 1997.