

THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS
YORK HEALTH ECONOMICS CONSORTIUM
NHS CENTRE FOR REVIEWS & DISSEMINATION

**THE YORK SERIES ON THE NHS WHITE PAPER –
A RESEARCH AGENDA**

**NICEly Does it: Economic
Analysis Within Evidence-Based
Clinical Practice Guidelines**

*James Mason
Martin Eccles
Nick Freemantle
Michael Drummond*

DISCUSSION PAPER 164

**NICEly DOES IT: ECONOMIC ANALYSIS WITHIN
EVIDENCE-BASED CLINICAL PRACTICE GUIDELINES**

James Mason
Centre for Health Economics, University of York

Martin Eccles
Centre for Health Services Research, University of Newcastle upon Tyne

Nick Freemantle
Centre for Health Economics, University of York

Michael Drummond
Centre for Health Economics, University of York

December 1998

SUMMARY

There is increasing professional and policy interest in the role of clinical guidelines for promoting effective and efficient health care. The NHS Health Technology Assessment Programme identified an urgent need, when such guidelines are produced, to develop a framework and methods for incorporating the best evidence of effectiveness, taking into account information on cost-effectiveness. This paper describes the development of recent evidence-based guidelines, for use in primary care, which were the result of recent work by the North of England Guidelines Development Group. Their specific aim was to incorporate economic analysis into the guideline process and treatment recommendations. The introduction of economic data raised some methodological issues, specifically: in providing valid and generalisable cost estimates; in the grading of cost 'evidence'; in finding a presentation helpful to clinicians. The approach used was to help clinicians aggregate the various attributes of treatment to make good treatment recommendations, rather than interpret cost-effectiveness ratios. In none of the guideline areas was there adequate information to estimate a cost per quality-adjusted-life-year. In the light of this research, future areas of work are identified and some recommendations are made for the forthcoming National Institute for Clinical Excellence.

KEYWORDS

Evidence-based medicine, economic evaluation, clinical guidelines, NICE

INTRODUCTION

The recent consultative document 'A first class service: quality in the new NHS' outlines plans for the establishment of a new National Institute for Clinical Excellence (NICE). [1] The objective of NICE is 'to give new coherence and prominence to information about clinical and cost-effectiveness' and to 'produce and disseminate clinical guidelines, based on relevant evidence of clinical and cost-effectiveness'.

Despite the proliferation of clinical guidelines, most do not address issues of cost-effectiveness and indeed, there may be unforeseen problems in introducing economic 'evidence' into guidelines. If recommendations produced by NICE are to lead to more cost-effective healthcare provision, it is important that any potential problems are addressed before NICE begins its work. This paper provides commentary on an attempt to develop evidence-based guidelines that explicitly considered the economic aspects of their recommendations.

A framework for formulating treatment recommendations is presented and the manner in which this differs from a common health-economic policy approach are discussed. The methodology of evidence-based guideline development is described briefly, showing the process required to include cost-effectiveness concepts. Examples of recent guideline work, addressing the use of ACE-inhibitors and NSAIDs in primary care, is provided and also a checklist for health economists becoming involved in guideline work. We will identify the strengths and weaknesses of our approach and move on to consider what questions now need to be answered, in the context of present health service reforms in the United Kingdom.

Why have economics in guidelines?

Within the UK, most guideline development has involved informal consensus processes, focusing on issues of clinical effectiveness. While a number of evidence-based guidelines have been published, these still focus on clinical effectiveness. [2-6] The introduction of cost considerations within guidelines has been argued for [7-9], though it is unclear how the introduction of economic data would influence the recommendations produced by a guideline development group.

The reasons for considering costs are clearly stated by Eddy [7] ; "health interventions are not free, people are not infinitely rich, and the budgets of (health care) programmes are limited. For every dollar's worth of health care that is consumed, a dollar will be paid. Furthermore, the costs will be paid by present and future patients. Such costs will be paid through insurance premiums, smaller employee benefit packages, lower salaries, the cost of commodities or direct or indirect taxation. While these payments can be laundered, disguised or hidden, they will not go away. Therefore, the costs of interventions should be balanced against the health outcomes predicted for that intervention for two reasons. Firstly, at the level of individual patients, failure to make the comparison may cause people to receive (and thus pay for) interventions that they might otherwise have declined, had they been fully

informed. Secondly, at the level of health care systems, costs must be considered if the available resources are to be used efficiently; failure to do this lowers the quality of care and harms the health of patients”.

The Committee on Clinical Practice Guidelines [8] recommends that every set of clinical practice guidelines include information on the cost implications of alternative preventive, diagnostic, and management strategies for the clinical situation in question. Their stated rationale is that this information can help potential users to better evaluate the potential consequences of different practices. Although they acknowledge “the reality is that this recommendation poses major methodological and practical challenges” they suggest that, in the process of considering costs, five theoretical questions that should be considered (Box 1) going on to provide reasons why guideline developers will have difficulty finding the answers to these questions (Box 2).

Box 1: Issues to be addressed in clinical guidelines.

- What evidence suggests that the services are likely to affect outcomes for the condition or intervention being considered?
- What groups at risk are most likely to experience benefits or harms from the proposed course of care and its side effects?
- What is known about the effects of different frequencies, duration, dosages, or other variations in the intensity of the intervention?
- What options in the ways services are organised and provided can affect the benefits, harms and costs of services?
- What benefits, harms and costs can be expected from alternative diagnostic or treatment paths, including watchful waiting or no intervention?

Source: Institute of Medicine (1992) [8]

Box 2: Problems confronting guideline developers.

- Scientific evidence about benefits and harms is incomplete
- Basic, accurate cost data are scarce for the great majority of clinical conditions and services
- While data on charges may be available, significant analytic steps and assumptions are required to treat charge data as cost data
- Techniques for analysing and projecting costs and cost effectiveness are complex and only evolving

Source: Institute of Medicine (1992) [8]

Once cost issues have been considered in the light of these questions and limitations, there is the further question of how to use the data in a guidelines group. Should it be presented alongside recommendations based solely on clinical effectiveness or should it be incorporated into the judgement process of deriving recommendations? Williams [9] argues that guidelines based on effectiveness issues and then costed may differ substantially and be less efficient than guidelines based on cost effectiveness issues. The complexity of this process, and the reactions it evokes, are reflected by the Committee on Clinical Practice Guidelines [8] report of “much debate, and with some vigorous dissent”. There is no widely accepted successful way to incorporate economic considerations into guidelines.

It is not clear how UK health care professionals will react to this process. Most health care professionals have a limited knowledge of health economics and economic modelling. Guidelines based on clinical effectiveness could be enhanced or undermined by the incorporation of economic considerations depending on whether they are seen as attempts to achieve cost-effectiveness or cost-containment. Additionally, it is uncertain how the incorporation of economic considerations will affect the use of guidelines with individual patients, although the intention is to encourage a more explicit consideration of cost consequences in each consultation where the guideline is used. UK health professionals are not accustomed to this process at anything other than an implicit level. Furthermore, with the prominence given to systematic reviewing and evidence-based guideline construction [10-12], it is unclear how “evidence” from the methodology of health economics, with its reliance on limited cost data, modelling and assumptions, will sit alongside “evidence” derived from the (perceived) rigour of systematic review. However, the absence of economic data from guidelines may severely limit their usefulness to medicine at both the policy and practice level.

At first sight, the values of clinicians seem very different from those of health economists. Clinicians, as advocates for patients, want to give the best possible treatment in each situation, while economists appear preoccupied with the prudent use of resources. Clinicians, nonetheless, are *de facto* the decision-makers deciding the allocation of resources and, with few exceptions, clinicians may defend their decisions on the basis that they believe they are providing appropriate patient care in each situation and with the information available. There is scant evidence that economic evaluation in any of its forms has made very much impact upon clinical decision-making in the United Kingdom. However, most of the recent reorganisations in the British National Health Service can be considered to be driven (at least in part) by economic principles, e.g. general practitioner fundholding and hospital trust status.

Concepts Underpinning Economic Evaluation

The development of economic evaluation has been driven by concepts of efficiency. A commonly-used objective in publicly provided health care systems is to improve or maintain health, as far as possible, with available resources. To maximise health gains, cost-effective strategies in health care must be identified; these are ones that produce, in a therapeutic area, the maximum output for a given cost (or minimum cost for a given output). This criterion

alone, though, is inadequate, since to maximise health in the population requires the various cost-effective strategies across all therapeutic areas to be implemented in proportions that achieve the socially-optimal allocation of resources. The measurement of health has mushroomed as an academic and clinical pursuit in recent years, with one strand attempting to produce generic measures that reduce patient health status (and its changes over time) to a single index (e.g. the Quality-Adjusted Life-Year or QALY). Hence, theoretically, health gains could be compared across different diseases and patient groups.

One approach to economic evaluation, the ‘decision-making’ approach, recognises that decision-makers have a range of objectives besides efficiency [13-14]. Inputs to decisions may include the decision-makers’ personal values and specific notions of equity. Current thought is to provide an index of output efficiency (the cost/QALY) to contribute to the decision-making process in the hope that decision-makers will give such data a good weight. Literature considering the impact of cost-effectiveness studies shows little impact [15-18] and it is possible that such studies may have been misdirected as to their audience. This may have been no bad thing, since the quality of the studies themselves has often been inadequate [19-24].

The decision-making approach also assumes the existence of an audience of social decision-makers, interested in weighing the costs and benefits of policy changes to all affected parties, and who may apply the results of cost-effectiveness studies. This assumption may be largely invalid in many health care systems. For example, in the United Kingdom, the foremost role of health authorities is (despite a history of reforms) to administer the flow of funds into primary and secondary care. Purchasers contracting for health services (often covering, in one service agreement, provision of a whole medical specialty) may have little use for an economic evaluation of an individual technology. In insurance-based systems, economic evaluations may be useful to health insurers where reimbursement follows (or can be refused) for a specified treatment, but in all health care systems it is the doctor who remains responsible for treatment decisions and ideally should own the conclusions of research findings. Thus, clinical practice guidelines offer the potential to resolve societal and individual patient-based decision-making, since they address the pertinent audience.

The rationale underpinning economic evaluation has been the belief that complex cost and benefit profiles associated with treatments can be aggregated, thus handing ‘an answer’ to aid decision-making (at least with respect to efficiency). This has proved unproductive, in part because the methods and data have not been adequate to provide a simple answer and in part because clinicians (the key audience) do not appear to think of appropriate health care in terms of economic outcomes, such as cost-effectiveness ratios.

The areas in which economic analyses have had some impact (formally or informally) are licensing and reimbursement decisions for new pharmaceuticals and in the introduction of some new ‘big-ticket’ technologies such as heart transplantation [25]. However, there is no evidence of impact upon already established health care technologies. This may generate a distortion in health care policy since new technologies are being evaluated with far greater stringency than existing care. The current evidence-based medicine culture potentially provides an environment for clinicians to review their existing practice. In recognition of the

ubiquitous constraints on resources, it is important that economic analysis takes the opportunity to contribute appropriately to this process.

Guideline Objectives

The guidelines development process recognises the clinician as the decision-maker, acting as the arbiter of appropriate treatment. In making decisions, clinicians balance their own preferences, those of patients and carers, patient specific information, the benefits, side-effects and safety of treatment and, to varying extents (depending on the mode of reimbursement), cost. Consequently, the primary goal of guidelines development is not to derive a cost per quality-adjusted life-year, rather to help the clinician perform the aggregation of attributes of treatment, and develop well-informed social preferences. Such a process still requires the assessment of costs and benefits of treatment to be methodologically sound. The novel aspect is the dynamic use of economic data (rather than as static published studies), alongside traditional clinical inputs, in the development of clinician valuation of treatments and consequent recommendations. In the rest of this chapter, we describe and comment on our initial attempt to introduce economic analysis into the process of developing evidence-based clinical practice guidelines.

DEVELOPING COST CONSCIOUS EVIDENCE-BASED GUIDELINES

The aim of a guideline is to provide recommendations, evidence-based where possible, to inform health care professionals about their use of health technologies. By constructing representative groups to work systematically through available, graded evidence and contextual issues, it may be possible to reach treatment recommendations that achieve the costs and benefits they predict while reflecting remaining uncertainties. A guideline does not replace the responsibility of health care professionals to use general medical knowledge and clinical judgement in consultations. It is recognised that recommendations may not be appropriate for use in all circumstances, and decisions to adopt any particular recommendation must be made by the practitioner in the light of available resources and circumstances presented by individual patients.

Each guideline involves a systematic appraisal of a medical intervention in terms of the areas shown in Box 3. This being the most current, pertinent and complete data available, each guideline sets out, or profiles, these attributes of treatment and, where appropriate, attempts a robust presentation showing the possible bounds of cost-effectiveness that might result. The range of values used to generate low and high cost-effectiveness estimates reflects available evidence and the concerns of the guideline development group. Simple and transparent presentation also permits reworking with different values from the ones used by the guideline group.

Box 3: Areas for systematic appraisal in guidelines.

- effectiveness
- quality-of-life
- compliance
- safety
- health service delivery issues
- resource use
- costs in the British health care setting

Our guidelines groups did not feel it helpful to consider a synthesis of previously published economic analyses which adopted a variety of differing perspectives, analytic techniques and baseline data. However, the economic literature was reviewed to position guideline findings against representative published economic analyses and interpret differences in findings when the group found it beneficial. We are unaware of any previous attempt to make formal inclusion of recommendations for the efficient use of health service resources in an evidence-based guideline. Hence, the approach used is recognised to be developmental.

Guideline Development

North of England Guidelines Development Groups were convened during 1996-7 to develop evidence-based guidelines for primary care prescribing decisions in four clinical areas. These guidelines were developed for use within a trial evaluating the effectiveness and efficiency of evidence-based outreach visits. Guideline topics were chosen for treatment areas where there was perceived to be scope for appropriate change and where guideline messages were likely to vary in the scale of changes of costs and benefits involved. The four treatments addressed were: ACE-inhibitors in the primary care management of adults with symptomatic heart failure; aspirin for the secondary prophylaxis of vascular disease in primary care; the choice of antidepressants for depression in primary care; and non-steroidal anti-inflammatory drugs (NSAIDs) versus basic analgesia in the treatment of osteoarthritis. [2, 3, 26, 27]

Guideline development groups were composed of general practitioners (GPs), secondary care physicians, local authority medical and pharmaceutical advisers, pharmacists and a research team consisting of a guideline development methodologist (ME, who acted as the group leader), a systematic reviewer (NF), and a health economist (JM). Clinical members brought to the group their knowledge of the disease, experiences in treating patients, and understanding of the practicalities of the health care system. The research team (ME, NF, JM) were responsible for reviewing and summarising the evidence relating to treatments and interactively feeding this information back to the group to allow the development of recommendations. Additionally, the research team were responsible for the drafting of the guidelines and the resourcing of the guideline development group. The working of the group can be characterised as a dynamic process where an understanding of the pros and cons of

treatment emerges and is refined, questions are responded to with available evidence and uncertainties are assessed. A detailed report of the working of the guidelines groups has been published. [28]

Levels of Evidence and Strength of Recommendation

To make a critical assessment of information on clinical effectiveness for evidence-based guidelines, reviewers follow a process of establishing the level of evidence that individual studies provide. Papers are categorised according to study design reflecting susceptibility to bias, and questions are answered using the best evidence available. Evidence categories, shown in Table 1, are adapted from the US Agency for Health Care Policy and Research Classification. [29]

Table 1: Categories of Evidence

-
- | | |
|------|---|
| Ia: | evidence from meta-analysis of randomised controlled trials |
| Ib: | evidence from at least one randomised controlled trial |
| IIa: | evidence from at least one controlled study without randomisation |
| IIb: | evidence from at least one other type of quasi-experimental study |
| III: | evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case-control studies |
| IV: | evidence from expert committee reports or opinions and/or clinical experience of respected authorities |
-

When considering a question of the effect of an intervention, if the question can be answered by category I evidence provided by a meta analysis or randomised controlled trial then studies of weaker design (e.g. controlled studies without randomisation) are not reviewed.

Recommendations are graded A to D as shown in Table 2, though categories of evidence do not always map easily onto a certain strength of recommendation. First, it is possible to have methodologically sound (category I) evidence about an area of practice that is clinically irrelevant or has such a small effect that it is of little practical importance and therefore attracts a lower strength of recommendation. Second, a statement of evidence may only cover one part of an area in which a recommendation has to be made, or evidence of similar quality may be contradictory. To produce comprehensive recommendations, a group has to extrapolate from the available evidence and this sometimes leads to lower strength recommendations based upon category I evidence. [4] In addition to the strength of clinical evidence, recommendations reflect the applicability of the evidence to the population of interest; economic considerations; guideline developers' awareness of practical issues; and (inevitably) guideline developers' societal values. [29]

Table 2: Strength of Recommendation

A	directly based on category I evidence
B	directly based on category II evidence or extrapolated recommendation from category I evidence
C	directly based on category III evidence or extrapolated recommendation from category I or II evidence
D	directly based on category IV evidence or extrapolated recommendation from category I, II or III evidence

To apply a strength of recommendation to cost information presents some difficulties. For example, it is possible that a large, well-conducted trial may estimate some overall resource savings for a new treatment, but unless these findings are generalisable to normal care they may not reflect the best evidence possible. Moreover, resource measurement taken from other sources, for example insurance claims databases, may be subject to unknowable influences, particularly selection biases, and similarly not provide a reliable view. Commonly, in health care, alternative treatment strategies feature small differences in outcome. Consequently, precise and internally valid trials are required not just to achieve a reliable measure of differences in health outcome, but also (correlated) differences in resource consequences. The approach adopted by the guidelines group for cost recommendations was to apply the same categories of evidence used for effectiveness to resource use and additionally, to establish the generalisability and relevance of findings by mapping their consequences onto current national patterns of resource use. For example, the SOLVD trial, for the treatment of heart failure with an ACE-inhibitor, [30] reported rates of hospitalisation for heart failure in the placebo group consistent with rates reported nationally for England. Hence, the reduction in hospitalisation in the active treatment group was consistent with improved health outcomes reported in the trial and plausible in the English setting. In the guidelines groups, strengths of recommendations for cost-effectiveness were, therefore, determined by the categories of evidence for both resource use and health outcome, as well as the generalisability of those data. The grade attached was determined by the overall quality of evidence as interpreted by the group.

Handling costs

While a social perspective in economic evaluation is both desirable and formally correct, in practice, due to the (un)availability of data, analyses of cost were limited to those borne by the NHS. Cost data used in guidelines were those in the public domain; it was beyond the scope of the guidelines development process to conduct new costing studies.

The approach is incrementalist, thinking of the net costs and consequences of changes in practice. Costs were calculated by attaching published average national unit costs (with the exception of drugs: see below) to resource items. Economists often argue that, for decision-making purposes, marginal costs are preferable to average costs. [31] While the problems

associated with average unit costs are recognised, there is no generally valid or accepted method for presenting marginal costs on items or procedures: these will vary from locality to locality. The simple presentation of analyses permits decision-makers to apply different unit costs where such information is locally available. Reflating of unit costs from different years of origin, to a common year, to adjust for health care cost changes over time, was not conducted. In general, there was no more than a two year gap between the oldest and newest values. Additionally, reflating is an ambiguous practice for certain items such as drug costs where, under UK reimbursement, the price tends to remain fixed over substantial periods of time.

Data on drugs

It is not the role of a guideline group to conduct original research, rather to synthesise obtainable information. However, in three of the four areas (ie, excluding ACE-inhibitors), Prescribing Analysis and Cost (PACT) data were obtained from the Prescription Pricing Authority (PPA). These data provide the total number of prescriptions reimbursed and their cost for each drug. Because prescriptions may not adequately describe the volume of use of drugs, [32] quantity data were adjusted using World Health Organisation tables of defined daily doses, to provide a measure of the patient years of prescribing reimbursed. [33] These are not necessarily the doses recommended in the British National Formulary, [34] but are intended to reflect the average maintenance dose in adults. Where WHO values were not available, the standard lower maintenance dose listed in the British National Formulary was used. Hence, it was possible to calculate the volume of use of each drug as well as comparative acquisition costs of drugs per patient for a set period of treatment (reflecting the mix of forms of each drug currently prescribed).

Examples of the guidelines

The full guidelines give a systematic presentation of the available evidence in each treatment area and the treatment recommendations derived by the guidelines groups. Here we present shortened summaries of two of the guidelines alongside selected recommendations, to illustrate issues for the inclusion of cost-effectiveness information in guideline development.

Guideline 1: ACE-inhibitors for Heart Failure

Introduction

Heart failure is a common, chronic condition with a very poor prognosis: as much as half of patients with a diagnosis of heart failure may die within 4 years. However, only 20-30% of patients assessed by British general practitioners (GPs) as having heart failure are prescribed an ACE-inhibitor. Most patients who are investigated for heart failure receive chest x-ray and electrocardiography but only about a third undergo echocardiography. Diagnosis by clinical assessment has been estimated to be correct in about half of cases when confirmed by echo-

cardiogram. While it is possible that some patients are being inappropriately treated, it is likely that ACE-inhibitors are considerably under-utilised in patients who might benefit from them.

Effectiveness data

Trials of ACE-inhibitors use left ventricular ejection fraction measured by an echocardiogram to detect heart failure. A meta-analysis of 39 studies found that the pooled relative risk of mortality using a fixed effects model was 0.83 (95% CI 0.76 to 0.90) when taking an ACE-inhibitor compared to placebo. The findings of the meta-analysis are consistent with the largest trial, SOLVD, which randomised 2569 patients with overt but stabilised heart failure and a left ventricular ejection fraction of 35% or less to enalapril or placebo with average follow-up of about 41 months. [30] The average benefit from ACE-inhibitors estimated during the SOLVD trial was 2.4 months of extended life.

Quality of life data

Narang and colleagues (1996) [35] reviewed aspects of quality-of-life reported in 35 double-blind studies, including 3411 symptomatic patients, comparing the effect of ACE-inhibitors and placebo. Exercise duration improved in 23 of 35 (66%) studies while symptoms improved in 25 of 33 (76%) studies. All 9 studies with a sample size of more than 50, a follow-up of 3-6 months and using a treadmill exercise test, showed improved exercise capacity as well as symptoms.

The single largest and most general assessment of patient quality of life comes from a subsidiary analysis of the SOLVD treatment and prevention trials. [36] This found interim improvements in self-assessed dyspnoea and social functioning in those patients treated with ACE-inhibitors that were statistically significant, although statistical significance did not persist for the full 2 years of follow-up. Another analysis of SOLVD symptomatic patient data using observed frequency of dyspnoea showed a statistically significant reduction achieved and maintained beyond 2 years for those patients treated with enalapril compared to those treated with placebo. [37]

Resource and cost data

Trials consistently show a reduction in hospitalisation for progressive heart disease when on ACE-inhibitor therapy. It is unclear whether these are lasting reductions or simply reflect a 'time-window' effect; with more patients on ACE-inhibitors completing the trial follow-up period without their heart disease progressing, but deteriorating in following years. The data do not suggest that greater hospitalisation for other reasons offsets reduced heart failure hospitalisation; ACE-inhibitors seem to have a positive effect on other-cause hospitalisation in symptomatic patients. It is not generally safe to assume that hospitalisation rates found in trials will be matched in clinical practice. As already noted, the hospitalisation rate in the control arm of the SOLVD treatment trial matches the average rate reported nationally for England. Each GP could expect, on average, four inpatient cases with heart failure each year and SOLVD trial data suggest that ACE-inhibition might prevent (or delay) one of these hospitalisations.

The annual cost of purchasing ACE-inhibitors (at maintenance doses) ranges from £100-£340 per year, considering doses reported in the British National Formulary. The incremental cost per patient using ACE-inhibitors in primary care may vary from a small cost saving through to a net cost of nearly £1600 over four years (see Table 3). In cost-effectiveness terms, it is likely that ACE-inhibitors for heart failure fall in the approximate range £0 - £10,000 per life year gained, given the range of assumptions listed and the remaining uncertainties. The important variables are the cost of the ACE-inhibitor itself and hospitalisation savings. It is not possible in this simple model to explore the influence of compliance with therapy on the cost-effectiveness estimates presented. The trial data, analysed on an intention-to-treat basis, reflect the level of compliance achieved in the SOLVD trial; the degree to which this represents experience in general practice in the UK is uncertain. Where non-compliance involves withdrawal of prescription then both costs and benefits are foregone and the overall cost-effectiveness ratios are not significantly altered. Substantial cross-over to ACE-inhibitor therapy in the placebo group in the SOLVD trial may mean the attributable benefits are underestimated.

Table 3: Net cost and benefit per patient of ACE-i for Heart Failure.

Assumptions (optimistic or conservative) [1]:	Optimistic	Conservative
ACE-inhibitor £100/year or 340/year for four years	£400	£1400
Initiation of therapy by 2 GP visits or 2 outpatient visits [2]	£20	£138
Reduced hospitalisation or no reduced hospitalisation [3]	-£626	£0
GP visits related to heart failure unchanged or 1 extra visit/year for four years [4]	£0	£40
Net cost range	-£206	£1578
Increased life expectancy (based on placebo comparison) [5]	0.203 years	0.203 years
Incremental cost-effectiveness of implementing ACE-inhibitor therapy [6]	Small cost saving and health gain	£7770 / life year gained

- [1] Costs and benefits are shown which arise from the addition of ACE-i 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. For simplicity of presentation the consequences of treatment withdrawal are not modelled, hence drug costs are likely to be over-estimated.
- [2] Cost per:
- GP consultation £10 (excluding prescribing cost): Netten and Dennett, 1997.[38]
 Outpatient visit £69: CIPFA/HFM, 1997 [39]
- Costs of additional blood tests are excluded as no adequate cost data was found.
- [3] Calculation based on:
- Difference in SOLVD trial treatment and control hospitalisation rates (21.9%-15.4%) x 4 years;
 Inpatient stay of 14.5 days (McMurray 1993) [40];
 Cost of an inpatient day £166: CIPFA/HFM, 1997 [39]
- [4] 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.
- [5] Based on the placebo-controlled findings of the SOLVD treatment trial, improved survival was highly statistically significant ($p = 0.0036$ by 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.
- [6] 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. This is not modelled, since it is probable that therapy would continue and so both costs and benefits would occur after 4 years.

Commentary

Within the effectiveness data, no extrapolation beyond the period of the trials has been attempted. The presentation of available evidence is relatively free of assumptions, and it is easy to explore values different from the ones used. The values presented reflected those felt appropriate by the group in discussion. It would, technically, have been better to make a fractional reduction in the drug costs over four years to reflect mortality. However, such an adjustment made no difference to the substantive findings and the group preferred the simplicity of the presentation shown. The purpose of the analysis was to confirm to the guidelines group that not only were the individual attributes of treatment favourable, but also the 'ball-park' cost-effectiveness. If treatment stopped at four years, there would be some additional benefit after cessation of therapy and the findings presented would be conservative, though it is probable that therapy would continue, and that both costs and benefits would occur after four years. Future costs and benefits were not discounted because of the short, four year time frame and because all important costs are distributed along with the benefit in time. Neither extrapolation nor discounting will substantially alter the cost-effectiveness ratios presented.

The strength of recommendation (Box 4) for the cost-effectiveness of ACE-inhibitors for heart failure reflects both that worthwhile health benefits have been established in trials as well as the net impact upon resources (from the SOLVD trial) which appears generalisable to the English context.

Box 4: Selected recommendations from the heart failure guideline.

- All patients with symptomatic heart failure and evidence of impaired left ventricular function should be treated with an ACE-inhibitor (A).
- Treatment of heart failure with ACE-inhibitors is cost-effective (A).
- As there is no good evidence of clinically important differences in the effectiveness of available ACE-inhibitors, patients should be treated with the cheapest drug that they can use effectively (B).

Guideline 2: NSAIDs for Osteoarthritis

Introduction

On average, approximately 3% of patients on general practice lists will be recorded as suffering from osteoarthritis, but osteoarthritis being one of a continuum of connective tissue disorders, the extent to which these inter-relate and share common treatment is uncertain. The prevalence of all such conditions (ICD 710-739) amounts to nearly 19% of patients.

Currently, GPs in England spend approximately £150 million each year on Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) for musculoskeletal and joint disorders. Annually, this corresponds to nearly 1.5 million person-years of treatment, with ibuprofen and diclofenac

constituting respectively 26% and 37% of prescribed usage. (Prescribed use of ibuprofen is likely to be an underestimate of overall use due to its over-the-counter availability.) Although they enjoy very extensive use in UK general practice, attempts to derive an evidence-based rationale to choose between NSAIDs has not proven fruitful. This is because of the poor quality of trials comparing different NSAIDs, which exhibit many biases. [41, 42] The guideline addressed a more fundamental question, raised by the group of general practitioners: when is it appropriate to use NSAIDs in preference to simple analgesia? Thus, the guideline addresses the decision to manage pharmacologically painful joints believed to be due to degenerative arthritis.

Effectiveness data

There are only three adequately designed trials [43-45] which may, in combination, provide useful information about the appropriate place of NSAID use in primary care. These studies examine the relative short term effectiveness (4-6 weeks) of the NSAIDs Ibuprofen and Naproxen against paracetamol in the treatment of osteoarthritis of the knee [43, 44] or the NSAID diclofenac versus co-proxamol in the treatment of non traumatic pain in one or two of the hip, knee, ankle or wrist joints. [45] In addition, the Bradley study compares ibuprofen at 'analgesic dose' or at 'anti-inflammatory dose' with paracetamol.

These studies indicate small benefits for NSAIDs over simple analgesia for pain at rest and pain in motion, using a visual analogue scale, but no significant change in time to walk 50 feet. One study, Parr et al [45] used the Nottingham Health Profile to describe different elements of the comparison between diclofenac and co-proxamol on broader health outcomes; the results suggest that there are no substantial differences in outcome for simple analgesia versus NSAIDs.

Compliance and safety data

Comparing NSAIDs and simple analgesia, treatment discontinuation was slightly less common for patients taking an NSAID, although the difference was not statistically significant. Overall, 3.3% (95% CI -1.2% to 7.7%) fewer patients dropped out of NSAID therapy over an average of 4.5 months of treatment.

The absolute level of risk associated with individual NSAIDs and their safety relative to paracetamol is unknown. In a large US trial, the control group of patients took a variety of different NSAIDs for rheumatoid arthritis. [46] In this cohort, the number needed to treat for a 6 month period to expect one serious gastrointestinal event was 105 (95% CI 81 to 151), though it is unclear to what extent this risk is attributable to NSAIDs. Placebo controlled studies of aspirin in heart disease indicate that, while gastrointestinal disease is common in patients, the additional disease attributable to NSAIDs may be relatively small. [3] Comparing the use of ibuprofen to no NSAID use, various case-control studies have estimated the rate of serious gastrointestinal damage to vary from no additional risk to a relative risk of two. [47]

The guideline group were unable to uncover any evidence to indicate the absolute risk of attributable side effects of paracetamol-based analgesics (compared to no analgesia use) in primary care patients with osteoarthritic pain.

The relative risk of serious gastrointestinal complications with individual NSAIDs is reviewed systematically by Henry and colleagues [48] who examine the relationship between drug use and admission to hospital. The review identifies 12 epidemiological controlled studies examining 14 NSAIDs for which comparison with ibuprofen can be made. Ibuprofen presents the lowest risk, and is used as a baseline to rank the relative safety of the other NSAIDs, although most differences are non-significant and the findings vulnerable to a range of biases. The review also suggests that the risk of gastrointestinal injury increases for higher doses of the same NSAID.

Resource and cost data

Paracetamol appears a cost-effective alternative to any NSAID because of lower acquisition cost and relative absence of gastrointestinal toxicity, while often providing adequate symptomatic relief and displaying similar levels of patient withdrawal from treatment. For the three NSAIDs for which randomised control trial data comparing with simple analgesia are available, the Henry [48] study suggests an ordering, on safety grounds, of ibuprofen, and then diclofenac or naproxen. While diclofenac and naproxen are similarly priced, ibuprofen is three to four times cheaper, given the forms in which these drugs are currently prescribed. Therefore, in the likely event that ibuprofen results in lower gastrointestinal injury and symptomatology, and without clear evidence of a general therapeutic advantage for naproxen or diclofenac, ibuprofen is the most cost-effective first-line NSAID.

There is wide variation in the purchase costs of different preparations of the same NSAID available on the NHS. There is no evidence to support the use of more expensive preparations over cheaper ones or the use of the modified release preparations.

In those patients requiring NSAID treatment it is important to consider what strategies may be available to minimise the risk of gastrointestinal injury. Such preventative strategies should not be confused with treatment of (common) dyspepsia where prescription or over-the-counter purchase of antacids may be considered when NSAID treatment cannot be modified.

Preventing NSAID induced gastro-intestinal injury

Concerns about the high risk of NSAID-induced gastrointestinal injury have led to a number of trials of H₂-receptor antagonists (such as cimetidine and ranitidine), and misoprostol. A recent meta analysis of trials focusing upon endoscopic assessment of lesions demonstrated a statistically significant reduction in the number of NSAID induced gastric ulcers in patients randomised to take misoprostol, but not those taking H₂ blockers. [49] Both H₂ blockers and misoprostol reduced the risk of duodenal ulcers in long term but not short term administration. However, it is recognized that endoscopically detected lesions may overestimate clinically important injury.

A recent large, double-blind trial randomised primary and secondary care patients, with rheumatoid arthritis and receiving NSAIDs, to receive misoprostol or placebo. [46] The trial assessed the development of serious upper gastrointestinal complications detected by clinical symptoms or findings (rather than scheduled endoscopy) and found a small but borderline

significant reduction in favour of the use of misoprostol. Twenty-five of 4404 patients receiving misoprostol, and 42 of 4439 patients receiving placebo suffered a serious upper gastrointestinal complication (odds ratio: 0.60; 95% CI 0.35 to 1.00 by Gart exact method) during six months of follow-up. The number needed to treat to prevent one serious gastrointestinal complication in this period is 264 (95% CI 132 to 5703). In the first month of the study 5% more patients taking misoprostol withdrew, primarily because of diarrhoea and other side effects.

A rate of serious gastrointestinal events necessitating hospitalisation for rheumatoid arthritis patients, associated with NSAID therapy, and preventable by misoprostol, can be estimated (Table 4).

Table 4: Rates of serious gastrointestinal events with and without misoprostol in rheumatoid arthritis patients taking NSAIDs (derived from Silverstein, 1995) [46]

Treatment	No of patients	Follow-up	Person-years on drug	Events*	Event rate/year
misoprostol+	4404	6 months	2202	25	0.0114
placebo	4439	6 months	2220	42	0.0189

* Serious GI events definitely attributable to NSAID use

+ Average daily dose 680 μ g

Hence, for 1000 patient years of treatment, 7.6 events will be prevented (95% CI 0.4-15.1), at a purchase cost of misoprostol of £230,000 (using the average dose reported in the trial). The net cost/event prevented is calculated, and the confidence interval of events prevented is used to provide low and high estimates (Table 5).

Table 5: Net cost and serious gastrointestinal events prevented with misoprostol prophylaxis (for 1000 patients)

Scenario	Events avoided	Cost of misoprostol	Savings in reduced hospitalisation*	Net cost	Cost/event avoided
High	0.4	£230,000	£1,200	£228,800	£572,000
Best guess	7.6	£230,000	£22,800	£207,200	27,300
Low	15.1	£230,000	£45,300	£184,700	£12,200

*The average cost of inpatient hospitalisation across all specialities was £3000/episode for Scotland in 1995/6. [50]

The rate of serious gastrointestinal complications in the control group in the Silverstein trial was 1.9% per person-year of treatment. Extrapolation using the number of person-years of

treatment currently prescribed in England implies 30,000 hospitalisations for NSAID-associated gastrointestinal injury per year. Thus, half of the 60,000 annual gastrointestinal ulcer/bleed associated hospitalisations in England [51] might be estimated to be NSAID-associated. Inclusion of over-the-counter NSAID use might suggest a bigger proportion. There were 4304 ulcer-associated deaths (ICD 531-3) in England in 1991. Assuming that chance of fatality following hospitalisation is independent of the underlying reason for gastrointestinal injury, then 2150 deaths per year can be attributed to NSAID-associated injury, or 1.38 deaths per year in 1000 patients taking an NSAID for one year. This suggests that nearly one in 10 serious gastrointestinal complications are fatal. The use of misoprostol led to a 40% relative reduction in serious events (95% CI 64%-2%) and so might be argued to lead to a 40% reduction in the average fatality rate. These figures are used to estimate the cost-effectiveness of a general policy of misoprostol prophylaxis in terms of the cost per life saved (Table 6). High and low estimates are derived assuming lives saved are a constant fraction of serious events avoided.

Table 6: Modelled cost-effectiveness of misoprostol prophylaxis (for 1000 treated patients)

Scenario	Net cost	Lives saved*	Cost/life-saved
High	228,800	0.0276	£8,290,000
Best guess	207,200	0.552	£375,000
Low	184,700	0.883	£209,000

* [High, best guess, low] estimate calculated as [2%, 40%, 64%] of 1.38 respectively (see text).

The above estimates must be viewed as tentative, given the assumptions required to reach them. For example, hospitalisation rates are generally higher in the US than in the UK, and the relationship between hospitalisation and mortality with ulcers of different underlying cause is unknown; notably there is currently no direct evidence of gastrointestinal injury-associated death being prevented by misoprostol prophylaxis. The purchase cost of misoprostol reflects additional separate prescription of misoprostol tablets. Combined NSAID and misoprostol tablets are available: subtracting the cost of non-proprietary forms of these NSAIDs, the combined drugs are generally a more expensive way of providing misoprostol, and if used will increase the net cost.

The mean ages of death in men and women due to ulcer, haemorrhage and perforation (ICD-9: 531-3) are 76 and 81 respectively. The average life-expectancy in the normal population for both genders at this age is about eight years, [52] so a crude calculation of cost per year of life gained is possible using estimates in Table 4. The age distribution of ulcer fatalities presented in national statistics aggregates those ulcers caused by NSAIDs, those related to *Helicobacter Pylori*, and those due to other causes. These causes would need to be separated,

and other modelling assumptions validated, before formal calculation of years-of-life saved is possible.

On the available evidence extrapolated from a trial of rheumatoid arthritis patients, it is not demonstrated that a strategy of routine and unselected misoprostol prophylaxis for patients taking NSAID therapy is cost-effective. Patient review and sequential therapy selection beginning with simple analgesia, is likely to minimise adverse event rates in the general patient group.

It is possible, although not demonstrated, that misoprostol prophylaxis may be more cost-effective in a high risk group for which current NSAID therapy has to be maintained. The Silverstein trial of rheumatoid arthritis patients suggested greater relative risks of serious gastrointestinal injury for patients with age >75 (Odds ratio [OR]=2.48), history of peptic ulcer (OR=2.29), gastrointestinal bleeding (OR=2.56) and heart disease (OR=1.84). These risks factors have been presented in such a manner that it is not possible to calculate absolute reductions in the rates of serious events for each high risk group, and the numbers of events in each instance are small. Without these, costs per life saved for high risk groups cannot be estimated, but none of the risk factors individually appear very important, and the cost effectiveness of misoprostol prophylaxis in high risk NSAID-user groups remains undemonstrated.

Although it appears likely that omeprazole may be similar in effectiveness to misoprostol in NSAID-induced ulcer prophylaxis and healing, [53] and better tolerated (although purchase costs are also higher), trials that rely on ulcers detected through endoscopy overestimate the effectiveness (and hence value) of protective agents in practice. No large, pragmatically designed trials with serious gastrointestinal events as primary outcome are available for omeprazole. Without such data it is not possible to recommend the routine use of omeprazole prophylaxis as an evidence-based strategy.

Commentary:

The guideline group were not using the available evidence to rule out any drug given for osteoarthritis, since the evidence regarding any of the consequences of these treatments is not compelling. Instead, on the basis of broad arguments about superiority or (near) equivalence of attributes, they established a sequencing of treatment:

- 1 Paracetamol is cheaper than ibuprofen, has most of the efficacy, is nearly as well tolerated, and is safer at therapeutic dose.
- 2 Ibuprofen is cheaper than diclofenac or naproxen, is not known to be less efficacious, and is probably safer.
- 3 Other NSAIDs are similarly more expensive and less safe than ibuprofen, and their relative efficacy to analgesia is not known.

The underlying quality of evidence is reflected in the strengths of recommendations in Box 5.

Box 5: Selected recommendations from the osteoarthritis guideline.

- In terms of cost-effectiveness, patients presenting with painful joints believed to be due to degenerative arthritis should initially be treated with paracetamol. If inadequate symptomatic relief is obtained, then ibuprofen is the most cost effective alternative (C).
- Modified release preparations are relatively expensive, while no evidence demonstrates that they are more effective than standard therapy; therefore they should not be used (D).
- Prophylactic protective therapy (with misoprostol or proton pump inhibitors) should not be used routinely as it is not cost effective for the reduction of serious gastric events (D).
- There are a group of patients who are at higher risk of upper gastrointestinal bleeding or perforation for whom prophylaxis may be cost-effective but further evidence is required (D).

The group also wished to explore the appropriateness of prophylactic treatment in patients who were at risk of gastrointestinal injury and for whom symptomatic relief could only be obtained with an NSAID. Given the lack of direct evidence, the group requested a modelling exercise to explore the likely value of treatment and having been led through the model, with all its uncertainties, the group felt confident that the case for a general policy of prophylaxis had not been established.

RESEARCH ISSUES

The four guidelines developed have some important similarities. Each addresses a highly focused question on the use of one, or several, classes of drug for the treatment of one condition in primary care. Each guideline assesses the available evidence using quantitative techniques appropriate to the task of informing the guidelines group. These factors may limit the value of the methodology presented in this paper in other contexts.

Certain areas of medicine are not amenable to quantitative analysis and may only be summarised by *narrative review*. An example may be where individual trials for a medical condition use different intermediate endpoints or outcomes. The need for narrative review will also increase as the focus shifts from treatment of one defined condition to the management of a whole disease. In these cases, the health economic component may be restricted to equivalence and dominance messages and not be able to offer clear advice where an alternative achieves (possibly conflicting signs of) a marginally greater effect at additional cost. Where the benefits of diagnostic procedures and treatments are patchy or unclear, it may prove difficult to set these against the costs of implementation.

The process of grading recommendations of costs alongside health consequences raised some interesting issues in the guideline groups, particularly in areas of uncertainty or broad equivalence of effect. An example came in the overall consideration of the relative value of Selective Serotonin Reuptake Inhibitor (SSRI) or tricyclic treatment in first line treatment of depression. The effects of the treatments are broadly equi-poised with the evidence

suggesting tricyclics are slightly more efficacious and SSRIs slightly better tolerated. A range of secondary claims concerning the advantages of SSRIs were explored (for example, they are safer in overdose than some but not all tricyclics) and were thought to be of minor worth, except where overdose was a genuine concern. In the light of inconclusive differences, it would be intuitive to suppose that any treatment recommendation favouring (if at all) either SSRIs or tricyclics would attract a low strength of recommendation. However, when the group had considered the net cost implications of using SSRIs instead of tricyclics they were certain that the additional costs were not worthwhile in the light of such uncertain benefits. Should the cost-effectiveness grading reflect the uncertainty of the treatment effect, the certainty of the cost consequence, or an amalgam? Ultimately, the strength of recommendation reflected not just the quality of evidence but the guideline group's understanding of the importance of the message. An alternative would be to treat costs and effects separately and develop different grading systems for cost evidence statements and recommendations. Then two grades would be attached to any cost-effectiveness recommendation. This would be at the expense of the simplicity of the method presented here, where a recommendation encodes the overall importance of a message.

The levels of evidence attached to the resource consequences of treatments mirrored those used for clinical effectiveness. The validity, susceptibility to bias, and generalisability of different sources of resource data is, ultimately, an empirical question and different viewpoints are possible. Adoption of a different system would again suggest movement towards separate grading of costs and treatment effects.

Many clinical guidelines currently in the public domain do not address health economic concepts. It is possible to envisage retrospective inclusion of economic analyses in such guidelines and this raises the issue of whether the health economist needs to be present at all at the time of guideline development. A concern with retrospective 'bolting-on' of economic analyses is the (often) complex issues of managing patients and delivering health care alongside the technical nature of the treatments themselves. Health economists need to understand these issues and reflect them in their analyses; additionally, it would appear to be a retrograde step to pass off an important opportunity to work so closely with clinicians and other health care professionals.

DISCUSSION

Incorporating economic data into evidence-based guidelines introduces some methodological challenges: specifically in providing valid, generalisable cost estimates, in the grading of cost 'evidence', and in finding a presentation helpful to clinicians. The guideline development process reported here presents an attempt to address these issues.

In recognition of the disparate level and nature of information available for each of the treatment areas, a profile approach was adopted to set out the attributes of treatment. Evidence concerning the effectiveness, quality-of-life, compliance, safety and health service delivery and resource use associated with treatment options was presented and discussed in the guideline groups. A recurring theme in these discussions was GPs attempts to derive

useful statements of how treatments would affect patients. In the case of ACE-inhibitors, the pooled relative risk of survival, reported in placebo-controlled trials, conveyed little and the average increase in life-expectancy was more helpful, though even this caused some debate, since the average increase masked a wide variation in benefit for individual patients. It is possible that clinicians find not just the results of economic analyses but clinical trials themselves difficult to put into the context of treating individual patients. Interestingly, the GPs in the group seemed comfortable moving from the very solid evidence presented on the value of ACE-inhibitors to the very speculative exploration of prophylaxis against NSAID-induced ulcer, reported here, when they themselves were participants in the steps and assumptions. A concern would be whether other clinicians outside the group could make such an investment to understand the issues.

Traditional summary presentations of cost-effectiveness were made, where available data permitted, but these provided only a partial view of the value of treatment. For example, the analysis of ACE-inhibitors in symptomatic heart failure suggests between £0 and £10,000 per life-year gained. These values do not inform on the quality of extended life or how the quality of existing life of patients might be improved when receiving an ACE-inhibitor. Clinicians asserted that the provision of simple presentations of the attributes of treatment and how these were aggregated was most useful: results could be unpicked permitting other clinicians to apply their own values.

When reaching recommendations, the clinicians appeared to require treatments to leap two broad hurdles: first did the treatment option really work (i.e. did changes in health and risks of consequent adverse events support the use of a treatment and could these benefits be delivered in practice); and then second, was the purchase worthwhile (for the benefits involved, what was the net requirement in health resource costs, and where did the costs and savings occur)? Additionally, they wanted an independent appraisal of the likely bounds of uncertainty for the various attributes of treatment (e.g. the likely range of the size of clinical effect and the likely range of the cost of achieving this effect).

The *raison d'être* of economic evaluation has (largely) been to achieve greater efficiency in health care delivery. In none of the areas examined was there adequate data to make an informative estimation of a cost/QALY. Even if such calculation had been possible, it is unclear how much additional value would have been attached to this information. Instead, the clinicians approached decisions by thinking about treatment attributes and by using the two-hurdle approach described above. This may reflect an appropriate response to the disparate effects of treatment, some good - some bad, requiring a different cognitive model. Seldom can a treatment's value be adequately captured by a simple, cost-effectiveness construct and it is apparent that the general practitioners in the group were not working with a pre-defined notion of 'worthwhile' in the way that health-economists often approach concepts of efficiency. By contributing to the 'clinical group consciousness' of a guideline development group, it may be possible to influence treatment recommendations to appropriately consider concepts of cost-effectiveness, consistent with clinical decision-making processes. An overview is provided in Table 7 both for health economists becoming involved in guidelines work, and for guideline developers considering including a health economist.

The definition of a valid guideline is one that, when followed, leads to the costs and benefits it predicts. This emphasises the need to use the best evidence available to produce the most valid guidelines and a variety of effective strategies to help clinicians to implement them. By these means, if economic issues are appropriately incorporated, then they will maximise cost-effectiveness in the limited way that current information permits.

Table 7: An overview of guidelines development for health economists

-
- It is important to be clear about the process and objectives of guidelines work, the conduct of group meetings and the role of each of the group members.
- Objective (if probabilistic) attributes of treatment decisions include effectiveness, side-effects, compliance, safety, quality-of-life, health service delivery issues and resource use. The outcomes of a guidelines process are graded treatment recommendations which may reflect some or all of these attributes.
- The health economist, together with other group facilitators, has a responsibility of providing a rigorous exploration of treatment attributes with the available evidence. A general understanding of other disciplines (statistics, epidemiology and health services research) is essential alongside a training in economic evaluation methodology.
- Simple and transparent presentations, which permit exploration with different values, are most likely to be helpful to the guideline group and subsequent users of the guideline.
- Each attribute of treatment is assessed in turn on its own merits including bounds of uncertainty. Over-precision should be challenged and all uncertainties should be explored which are appropriate to the data or expressed in the group.
- Careful presentation and full discussion in the guidelines group is essential for an understanding of the attributes of treatment to evolve into a view of overall clinical importance. Data, although rigorously analysed are being used to put a treatment in a broad 'ball-park' with respect to its various attributes (e.g. safe, acceptable, effective, deliverable, cost-neutral).
- The importance attached to each attribute of treatment remains the responsibility of the guideline group as a whole and recommendations must be transparent and credible to the target audience. The process of aggregating up to an overall recommendation may be facilitated by a summary table of attributes, and presenting summary cost-effectiveness estimates when appropriate. Summary ratios need careful explanation and interpretation.
- A systematic review of previously conducted economic analyses relating to a treatment may provide useful background to the health economist but may have limited (or no) direct use in the guidelines process. If rigorously conducted, the guidelines process is likely to produce an understanding and evaluation of treatment inadequately reflected in any one published analysis.
- The scope for conducting traditional cost-effectiveness models may be limited or unhelpful in some therapeutic areas, especially where the various attributes of treatments contain conflicting messages. Where modelling is appropriate, clinicians appear more responsive to simple and transparent models, than to complex 'black box' methods requiring greater assumptions and extrapolation.
- Grading of recommendations of cost-effectiveness is in its infancy. However, the grade of recommendation should reflect not only precision and susceptibility to bias of data but also generalisability.
- As with clinical effectiveness, it is acceptable to say 'we just don't know yet' in a recommendation about cost-effectiveness of treatment and that more precise data is required.

NICEly DOES IT

Evidence-based cost-effectiveness guidelines are a fledging science with methodological development ongoing. However, it must be this 'science' that NICE is seen to communicate to the NHS, without modification for political ends. Cost containment guidelines are likely to have no currency with the medical profession. Credible guidelines stand the best chance of achieving an appropriate use of resources. For example, a recent guideline considering the use of Donepezil Hydrochloride for the treatment of Alzheimer's Disease concluded on the basis of the available phase II and III trials that there was currently insufficient evidence to recommend its use, or to continue secondary care initiated prescribing. [54]

For products approaching the licensing stage, NICE plans to require access to trial data four months before product launch. Appraisal will follow a Development and Evaluation Committee (DEC) style procedure using a Cost per Quality-Adjusted Life-Year framework (Cost/QALY) rather than a guideline approach. The major problem with cost-utility studies has always been the quality and handling of limited data. [24] How will NICE achieve a robust assessment of new technologies on the basis of trial data intended for licensing? In internal deliberations the most heroic assumptions will be required and as regards the external audience of practising clinicians, a cost-utility framework does not contain the right information to advise treatment decisions.

A simplified, evidence-based guideline process may be possible when considering close analogue and 'me-too' products at the near licensing stage. In four months, little more will be possible. To assess new pharmaceuticals and procedures in the context of the epidemiology of the disease, other available treatments and health service delivery issues may place insurmountable strain upon the appraisal process. An annual target of 30 to 50 appraisals and guidelines is evidently ambitious. The consequent danger is that the appraisal process will have to cut corners in the very circumstances where a full guideline is most needed. Tampering with the elements of the evidence-based clinical guideline process may threaten their validity and consequent authority in clinical decision-making. [27, 55] Where longer term data are required, a continuing-research period, either following or delaying the NHS reimbursement decision, is proposed - although how this will work is as yet unknown. A full guideline process would ensure the right questions are asked of any further research. NICE should focus on evaluating medicines in robust and physical terms rather than calculating assumption-laden QALYs. If the science wins then the guidelines process may become an implicit but effective fourth hurdle for pharmaceuticals in the UK.

The ability of any guidelines to effect a substantial change in clinical practice remains a research issue. The decision-making environment will, in any case, change in England due to the introduction of clinical governance: a package of activities to ensure clinicians receive and act upon the best evidence concerning treatment, corporately monitor their performance and report to those with overall responsibility.

The impact of implementing guidelines

Substantial use is to be made of audit methodologies in informing the clinical governance process. To this end, NICE ‘will develop a range of audit methodologies that can be adapted for local use to support the guidance it produces’. Actual monitoring will come under the ‘National Framework for Assessing Performance’ initiative, led by the new Commission for Health Improvement (CHI). Although it sounds straightforward, there are a number of major informational challenges in implementing this model. [56] It will prove unrealistic to follow all clinical activities: instead tracer conditions will need to be identified, although these may vary from year to year. Information may be relatively easy to obtain for some activities: use of drugs in primary care, use of procedures in secondary care, and even some simple outcome data (e.g. 30-day operative mortality). However, appropriate use of these interventions (i.e. whether the right patient received the right intervention at the right time) may prove considerably more difficult and require linkage across different patient records.

Simply handing NICE guidelines to clinicians would not be expected, on current evidence, to achieve substantial appropriate changes in health care. Clinical governance is intended to create a pincer against unacceptable variations in performance, and the government will consider introducing new powers for NICE and CHI if problems persist. This reflects a belief that acceptable and unacceptable variations in practice are generally separable. The evidence for the appropriate use of many treatments provided by the NHS is inadequate and it is unclear how far such a strategy can be pursued. The government may need to consider the dual issues of measurability and enforceability implicit in its’ reforms. Clinical governance as described signals a shift away from clinical freedom and towards protocol-driven health care. Guideline methodology has been developed to improve the scientific basis of decision-making; guidelines have never been intended for mechanistic use or to form the basis of legal argument about defensible health care.

REFERENCES

1. NHS Executive. A First Class Service: Quality in the new NHS. HMSO, 1998.
2. Eccles M, Freemantle N, Mason JM. for the North of England Evidence-based Guideline Development Project. Evidence-based clinical practice guideline: ACE-inhibitors in the primary care management of adults with symptomatic heart failure. *British Medical Journal*, 1998; 316: 1369-75.
3. Eccles M, Freemantle N, Mason JM. for the North of England Evidence-based Guideline Development Project. Evidence-based clinical practice guideline: aspirin for the secondary prophylaxis of vascular disease in primary care. *British Medical Journal*, 1998; 316: 1303-9.
4. Eccles MP, Clapp Z, Grimshaw JM, Adams PC, Higgins B, Purves I, Russell IT. North of England evidence-based guideline development project: Methods of guideline development. *British Medical Journal* 1996;312:760-1.
5. North of England Asthma Guideline Development Project. Summary version of evidence-based guideline for the primary care management of asthma in adults. *British Medical Journal* 1996;312:762-6.
6. North of England Stable Angina Guideline Development Group. Summary version of evidence-based guideline for the primary care management of stable angina in adults. *British Medical Journal* 1996;312:827-832.
7. Eddy DM. A manual for assessing health practices and designing practice policies: the explicit approach. American College of Physicians, Philadelphia, 1992.
8. Institute of Medicine. *Guidelines for clinical practice. From development to use.* (eds Field MJ & Lohr KN). National Academy Press, Washington, 1992.
9. Williams A. How should information of cost-effectiveness influence practice? In: *Outcomes in clinical practice*; ed: Delamothe T. London, BMJ, 1995.
10. Grimshaw J, Russell I. Achieving health gain through clinical guidelines. I: Developing scientifically valid guidelines. *Quality in Health Care*, 1993; 2, 243-8.
11. Feder G, Grimshaw J. Guideline development. In: Royal College of General Practitioners. The development and implementation of clinical guidelines: report of the Clinical Guidelines Working Group. Report from Practice 26. Royal College of General Practitioners, London, 1995.
12. Grimshaw JM, Eccles M, Russell IT (1995). Developing clinically valid practice guidelines. *Journal of Effective Clinical Practice*, 1995;1:37-48.

13. Williams A. Cost-benefit analysis: bastard science or insidious poison of the body politic? *Journal of Public Economics*, 1972; 1: 199-216.
14. Sugden R, Williams A. The principles of practical cost-benefit analysis. Oxford University Press, Oxford, 1978.
15. Alban A. Economic appraisal: what is the use? Paper presented at the Third Nordic Health Economists' Study Group, 1982. (Available from the Danish Hospital Institute, Copenhagen.)
16. Davies LM, Coyle D, Drummond MF and the EC Network. Current status of economic appraisal of health technology in the European Community: report of the Network. *Social Science and Medicine*, 1994; 38(12) 1601-1608.
17. Ross J. The use of economic evaluation in health care: Australian decision-makers' perceptions. *Health Policy*, 1995;31: 103-110.
18. Drummond M, Cooke J and Walley T. Economic evaluation under managed competition: evidence from the UK. *Social Science and Medicine*, 1997; 45: 583-595.
19. Lee JT, Sanchez LA. Interpretation of 'cost-effective' and soundness of economic evaluations in the pharmacy literature. *Am J Hosp Pharm*, 1991; 48: 2622-7.
20. Adams ME, McCall NT, Gray DT et al. Economic Analysis in randomized controlled trials. *Medical Care*, 1992; 30: 231-243.
21. Udvarhelyi IS, Colditz GA, Rai A, Epstein AM. Cost-effectiveness and cost-benefit analysis in the medical literature: are methods being used correctly? *Annals of Internal Medicine*, 1992; 116: 238-244.
22. Gerard, K. Setting priorities in the new NHS: can purchasers use cost-utility information? *Health Policy* 1993; 25: 109-125.
23. Mason JM, Drummond MF. The DH register of cost-effectiveness studies: content and quality. *Health Trends*, 1995a; 27: 50-56.
24. Mason J, Drummond M. Reporting guidelines for economic studies. *Health Economics*, 1995b; 4: 85-94.
25. Buxton MJ. Problems in the economic appraisal of new health technology: the evaluation of heart transplants in the UK. In MF Drummond (ed) Economic appraisal of health technology in the European Community. Oxford, Oxford Medical Publications, 1987.

26. Eccles M, Freemantle N, Mason JM. for the North of England Evidence-based Guideline Development Project. Evidence-based Guideline for the use of non-steroidal anti-inflammatory drugs (NSAIDs) versus basic analgesia in the treatment of pain believed to be due to degenerative arthritis. *British Medical Journal*, 1998;317:526-30.
27. Eccles M, Freemantle N, Mason JM for the North of England Guidelines Development Group. The choice of antidepressants for depression in primary care. *Family Practice* (In press).
28. Eccles M, Freemantle N, Mason JM. Methods of developing guidelines for efficient drug use in primary care: North of England Evidence-based Guidelines Development Project. *British Medical Journal*, 1998; 316: 1232-5.
29. [AHCPR] US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. *Acute pain management: operative or medical procedures and trauma*. Rockville, MD: Agency for Health Care Policy and Research Publications, 1992.
30. SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med*, 1991;325:293-302.
31. Mooney GH, Drummond MF. Essentials of Heath Economics: Part 1(continued) - What is economics? *British Medical Journal*, 1982; 285:1024-5.
32. Bogle SM, Harris CM. Measuring prescribing: the shortcomings of the item. *British Medical Journal*, 1994;308:637-40.
33. World Health Collaborating Centre for Drug Statistics Methodology. *Anatomical therapeutic chemical classification index*. 1992, Oslo: World Health Organisation.
34. *British National Formulary Number 32*. London: British Medical Association and Royal Pharmaceutical Society of Great Britain, 1996.
35. Narang R, Swedberg K, Cleland JGF. What is the ideal study design for evaluation of treatment for heart failure? Insights from trials assessing the effect of ACE-inhibitors on exercise capacity. *European Heart Journal*, 1996;17:120-34.
36. Rogers WJ, Johnstone DE, Yusuf S, Weiner DH, Gallagher P, Bittner VA, Ahn S, Schron E, Shumaker SA, Sheffield LT. Quality of life among 5,025 patients with left ventricular dysfunction randomized between placebo and enalapril: The Studies of Left Ventricular Dysfunction. *J Am Coll Cardiol*, 1994;23:393-400.
37. Pouleur H. ACE-inhibitors in the treatment of clinical heart failure. *Basic Research in Cardiology*, 1993;88 (suppl 1) : 203-9.

38. Netten A and Dennett J. *Units costs of health and social care*. Personal Social Services Research Unit (PSSRU), University of Canterbury, 1997.
39. CIPFA/HFM. *The Health Service Financial Database 1997* (CD-ROM). Chartered Institute of Public Finance and Accountancy, London, 1997.
40. McMurray J, Hart W, Rhodes G. An evaluation of the cost of heart failure to the National Health Service in the UK. *Br J Med Econ* 1993;6:99-110.
41. Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clinical Trials*, 1989;10:31-56.
42. Rochan PA, Gurwitz JH, Simms RW, Fortin PR, Felson DT, Minaker KL, Chalmers TC. A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Int Med* 1994;154:157-63.
43. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Treatment of knee osteoarthritis: relationship of clinical features of joint inflammation to the response to a nonsteroidal antiinflammatory drug or pure analgesic. *Journal of Rheumatology*, 1992;19(12):1950-4.
44. Williams HJ, Ward JR, Egger MJ, Neuner R, Brooks RH, Clegg DO, Field EH, Skosey JL, Alarcon GS, Willkens RF, et al. Comparison of naproxen and acetaminophen in a two-year study of treatment of osteoarthritis of the knee. *Arthritis & Rheumatism*, 1993;36(9):1196-206.
45. Parr G, Darekar B, Fletcher A, Bulpitt CJ. Joint pain and quality of life; results of a randomized trial. *Br J Clin Pharm*, 1989;27:235-42.
46. Silverstein FE, Graham DY, Senior JR, Davies HW et al. Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs: a randomized, double-blind, placebo-controlled trial. *Ann Int Med*, 1995;123:241-9.
47. Langman MJS, Weil J, Wainwright P, Lawson DH, Rawlins MD, Logan RFA et al. Risks of bleeding peptic ulcer associated with individual non-steroidal anti-inflammatory drugs. *Lancet*, 1994;343:1075-8.
48. Henry D, Lim LL-Y, Garcia Rodriguez LA, Perez Gutthann S, Carson JL, Griffin M et al. Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta analysis. *British Medical Journal*, 1996;312:1563-6.
49. Koch M, Dezi A, Ferrario F, Capurso L. Prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal mucosal injury: a meta analysis of randomized controlled clinical trials. *Arch Intern Med*, 1996;156:2321-32.

50. National Health Service in Scotland Information and Statistics Division. *Scottish health service costs, year ended 31 March 1996*. Edinburgh, Department of Health., 1996.
51. *Hospital Episode Statistics 1994-95*. London: Department of Health, 1995.
52. *Health and Personal Social Services Statistics for England*. Government Statistical Service. London: Department of Health, 1996.
53. Hawkey CJ, Swannell AJ, Eriksson S, Walan A, Lofberg I, Taure E, et al. Benefits of omeprazole over misoprostol in healing NSAID-associated ulcers. *Gut*, 1996;38(Suppl 1):T155.
54. Eccles M, Clarke J, Livingstone M, Freemantle N, Mason JM. for the North of England Evidence-based Guideline Development Project. Evidence-based clinical practice guideline: The primary care management of dementia. *British Medical Journal*, 1998;317:802-8.
55. Cluzeau F, Littlejohns P, Grimshaw JM, Feder GS. National survey of UK clinical guidelines for the management of coronary heart disease, lung and breast cancer, asthma and depression. *Journal of Clinical Effectiveness*, 1997; 2: 120 - 123.
56. Scally G, Donaldson LJ. Clinical governance and the drive for quality improvement in the new NHS in England. *BMJ* 1998;317:61-65.