WHAT COULD BE NICER THAN NICE?

Alan Williams FBA

Research OHE Consulting
Office of Health Economics

Annual Lecture 2004

www.ohe.org
ISBN 1 899040 03 X
August 2004

WHAT COULD BE NICER THAN **NICE**?

Alan Williams FBA
Centre for Health Economics
University of York, England

OHE Annual Lecture 2004

The Office of Health Economics
12 Whitehall London SW1A 2DY
Telephone: +44 (0)20 7930 9203
www.ohe.org

About the Author

Alan Williams has been a key person in the development of health economics as an intellectual discipline. His influence as a teacher and researcher from his base in York has been profound, both in the UK and internationally. His pioneering work on the measurement and valuation of health greatly strengthened the case for adopting the Quality Adjusted Life Year as the best outcome measure to use in health care priority-setting. His recent research activities have concentrated on the integration of equity considerations into cost-effectiveness analysis. In the 21st year of his "retirement", he claims to be well past his "best before" date, but being elected a Senior Fellow of the British Academy in 2002 (the first health economist to ever receive this honour) seems to indicate that others think that there may yet be more to come.

Acknowledgements

I am very grateful to all those who devoted some of their valuable time an energy to ensuring that I understood better the impact of NICE upon their lives, including those within NICE who struggled hard to eradicate my misconceptions about it. It was a great educational experience for me, and although not all of their wise comments show up in this lecture, I have stored them all away and they may surface in recognisable form one of these days. If I have still got hold of the wrong end of the stick here and there, please be charitable and put it down to ignorance rather than to stupidity, because ignorance is potentially remediable!

Office of Health Economics

The Office of Health Economics (OHE) was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and by revenue from sales of its publications, consultancy and commissioned research.

What could be nicer than NICE?

The National Institute for Clinical Excellence is the closest anyone has yet come to fulfilling the economist's dream of how priority-setting in health care should be conducted. It is transparent, evidence-based, seeks to balance efficiency with equity, and uses a cost-per-QALY benchmark as the focus for its decision-making. What more could anyone ask for? Well, experience has taught me that it is not uncommon for an-economist's-dream-come-true to be seen as a nightmare by everyone else. After all what practical person, outside the rarified world of academic health economics, would waste time and energy pursuing the elusive will-o'-the-wisp of creating for the NHS *a comprehensive framework for healthcare prioritisation, underpinned by an explicit set of ethical and rational values to allow the relative costs and benefits of different areas of NHS spending to be comparatively assessed in an informed way?* The answer is...the House of Commons Select Committee on Health, whose words are those italicised above. And NICE would not have been born had not some influential person also had that vision in mind.

But it is an extraordinarily difficult task, and NICE is still in its infancy, and on a very steep learning curve, because as a pioneering institution it has no well-trodden path to follow. It would be amazing if it managed to get everything right straight away. So my purpose here is to argue the case for NICE to move, in the next 5 years, in particular directions which I will map out shortly.

A turbulent infancy?

NICE was born into a very turbulent political, intellectual and cultural environment. Traditional modes of thought were being challenged inside each of the disciplines on which NICE draws, not to mention the additional tensions that arise when different disciplines have to work together to resolve problems that they have hitherto seen as their own territory. Inside **medicine** there was (and still is) a sharp clash of cultures between those wanting to make it more evidence-based (and less reliant upon clinical opinion) and those who wish to maintain a high level of clinical freedom in order to be able to respond to the

infinite variety of the human condition. Inside **statistics** the clash is between the hypothesis-testing mode of the conventional frequentist model and the decision-support mode of the Bayesians. Inside **economics** the clash is between the welfarists whose approach focuses on the maximisation of individual utilities and the non-welfarists whose approach focuses on the maximisation of population health (but also taking distributional issues explicitly into account). In the world of **finance** there is the clash between those who want the state to run public services directly, and those who want to harness the profit-seeking innovatory drives of the private sector for the public good. Inside **philosophy** the clash is between the utilitarians (including rule-utilitarians) who provide the philosophical underpinnings for conventional welfarist economics, and the Rawlsians who challenge that culture and want us, on grounds of fairness, to focus attention especially on the worse off groups in society. In **political ideology** the clash is between the libertarians promoting freedom of choice and decentralised decision-making as epitomised in well-functioning markets, and egalitarians of various kinds who accept that freedom of choice often has to be severely constrained in order to reduce inequities in society. And in the field of **public administration** there is the continuing clash between those who would like to take important social institutions such as the NHS out of the rough-and-tumble of party-politics, and those who regard all appointed bodies as lacking democratic legitimacy and therefore constituting a non-accountable technocracy usurping the power of the elected representatives.

And if all this internal conflict within disciplines were not enough, just think of the clashes ***between*** disciplines which have to be added to all this, for instance the clash between the individual perspective of clinical practice and the population perspective of epidemiological research, or between the wide-ranging optimisation perspective of economics and the bounded rationality (or muddling through) perspective of public administration.

It was into this dangerous intellectual jungle that the infant NICE was born, and it was inevitable that NICE would become a focal point for many of these tensions. And because NICE is a decision-making body, and not a debating society, it cannot sit on the

fence... it has to make up its mind and take a stance. And with its commendable commitment to being as transparent as possible, it lays itself open to intense scrutiny in everything it says and does, more so indeed than any governmental decision-making body that I know anything about.

NICE's role in the world

What then is NICE's role in the world of health care priority-setting, which is the role I am going to concentrate on? From my conversations with people over the past few months I have come to the conclusion that this is wrongly perceived by many people. A common perception of its much publicised activities in the field of Technology Appraisal appears to be that NICE is a quasi-judicial body, putting people and their practices on trial, seeking evidence from the defendants which is then considered by the prosecution in a highly adversarial context, and eventually placed before a jury (the Appraisals Committee), with a right of reply and an appeal system to which the principles of natural justice should apply. I find this a bit weird, because it implies that some kind of wrong-doing is the starting point for the process, whereas this is patently not so. So what kind of judgmental process is it that NICE is engaged in? I think the legal dictum that is the key to understanding NICE's role in the world is ***CAVEAT EMPTOR – Let the buyer beware!*** The NHS is under continual pressure from the health care industry, from health care professionals, and from particular patient groups, to invest its money in this that or the other. I regard all of these people as trying to "sell" something to the NHS. Like all sellers, they puff their wares, by which I mean that they present them in the most favourable possible light. In the world of the ordinary consumer we have the Consumers' Association and "Which?" reports to help guide us through this minefield, and we rely on them to examine the sellers' claims thoroughly. We then have to make up our own minds what to do in the light of this information. In the world of health care this is surely the role that NICE is playing...it is advising the buyer (the NHS in one or more of its various manifestations) on the properties of each of the various wares referred to it for analysis, and how they rate on a value-for-money scale.

But NICE goes one step further than this with its Technology Appraisals, and ***tells*** the NHS budget-holders that if the relevant clinical decision-makers want to follow NICE's positive recommendations the money must be found to enable them to do so. This mandatory element vis-a-vis those controlling the purse-strings is not present in the relationship between the Consumers' Association and the readers of "Which?" reports. I will return to the mandating issue later, but for now I simply want to emphasise that in my view NICE should **not** be seen as a quasi-judicial body dispensing ***justice***, but as a ***watchdog*** protecting the interests of the general public by dispensing well-founded advice about value-for-money to potential purchasers within the NHS, who are spending large chunks of the taxpayers' money on health care, and whose decisions we want to be as well-informed as possible.

NICE and priority-setting

Against that background, the two principal activities of NICE upon which I want to concentrate are the Technology Appraisals (TAs) and the Clinical Guidelines (CGs). They have developed in rather different ways. Put crudely, the TAs have been captured by the economists, and the CGs by the clinicians, and consequently they manifest two very different dominant cultures.

Technology Appraisals

The methodological underpinnings for TAs derive from non-welfarist health economics, and are driven by the objective of maximising population health subject to a budget constraint. This leads to the use of a decision criterion formulated in terms of the extra costs that would have to be incurred in order to bring about a unit improvement in health. The unit of *health* required for priority setting across different diseases cannot be a disease specific one, but has to be a generic one, and, since it needs to reflect the values of all of the people whose health might be affected by the decisions that are being made, it has to reflect the values of the entire citizenry of England and Wales. And since people value both length of life and quality of life, in principle the preferred outcome measure is the Quality-adjusted-life-year (or QALY), which embodies both of these benefits. The

measurement of gains in health-related quality-of-life is however a rather complex task, to which I have devoted a large part of my professional life over the past 20 years, and I am pleased to note that in practice most of the QALY-based evidence submitted to NICE is based on the EuroQol group's EQ5D measure and its associated valuation matrix, which I helped to develop, and which was designed specifically for NICE's purposes even before NICE was conceived! There's academic foresight for you!!!!

On the cost side, since NICE is advising the NHS, it concentrates attention primarily on costs falling on the NHS, with some ambivalence about the inclusion of other costs. This ambivalence is understandable when you consider for instance the issues surrounding the possible inclusion of productivity gains or losses in a Technology Appraisal. One obvious danger is that they are likely to be included when they favour a technology but not otherwise, and especially when a technology is close to the borderline and needs all the help it can get. But the inclusion of productivity gains in such circumstances would favour treatments affecting the working population (since we only count productivity gains and losses when we can pick them up in a market, and ignore the value of non-marketed services provided within the household). This then becomes problematic if you hold an egalitarian viewpoint about the value of a QALY being independent of somebody's economic worth.

NICE's benchmark or threshold

But ranking technologies according to their cost-per-QALY score still does not get us to our destination, because we still lack a cut-off point beyond which we say "this far but no further", and anything clocking in with a cost-per-QALY higher than this threshold is deemed not to be cost-effective, and will be rejected. NICE should have been given such a benchmark by Ministers when it was set up, but it wasn't. Nature abhors a vacuum, and NICE needed a benchmark, so it exercised its collective wisdom and came up with one, and it is roughly £30k per QALY, plus or minus £5k depending on the specific circumstances of the case. It is at this point that NICE's boldness in the transparency business deserves a cheer, but only one cheer! The second cheer that it might have earned has to be

withheld because for a while it denied it had any benchmark, and it only softened its position when it emerged from analyses of its decisions that it did have some kind of benchmark zone. The third cheer still has to be withheld because NICE still insists that this benchmark zone has nothing to do with affordability or with the rationing of health care, despite the fact that such considerations are the only justification there is for having a benchmark! NICE attempts to sustain this position by saying that NICE does not determine the NHS's budget, and it is those that do that determine affordability. But although that is true, NICE does not follow through the logic of that argument and let the threshold be determined by the budget holders.

When establishing a cut-off point it is necessary ***either*** to make a judgement about the intrinsic worth of a QALY and adjust the budget accordingly, ***or*** to make a judgement about how much money we can afford to spend on the NHS and let the value of the marginal QALY emerge from the priority setting activities of the budget-holders. In a properly integrated system, by choosing the one you determine the other, and they **are** brought into equilibrium. But there is no reason to suppose that they are brought into equilibrium in the circumstances we are facing today. This is because the budgets of purchasers, which determine the affordability of QALYs for the NHS, are not set explicitly so as to equalise QALYs at the margin, and the purchasers do not act explicitly so as to maximise QALYs. Consequently it is extremely likely that the "shadow price" of a QALY (i.e. the implicit value of a QALY as determined by the most cost-effective intervention that each purchaser just cannot afford to buy) will vary from purchaser to purchaser. And it is widely believed that this "shadow price" is much lower than the NICE benchmark of £30k. I think a major effort should be made to find out whether this belief is well-founded, though at the same time it might be useful to identify instances where budget-holders are (perhaps reluctantly) funding interventions which are well beyond the current NICE benchmark.

I too think that this £30k figure is far too high, though I am no more able than anyone else to prove that my figure is the right one. In the absence of such proof I suggest applying a bit of common sense so as to help get things into perspective. Currently in the UK we have about

£18k worth of real resources to provide each year for the all the needs of the average citizen (food, shelter, transport, education, etc), yet the current NICE benchmark means that it is willing to commit nearly twice that sum on medical care alone to provide one person with an extra year of healthy life (to sustain which they will need the food, shelter, transport, etc as well!) It is clear that we ***could*** do that at the margin for a few people without imposing great hardships on the bulk of the population, but we could not do it for many. For example, it seems from some recently reported research that on average we spend 5 times as much on medical care for people in the final 3 months of their lives as we spend on them per quarter during the rest of their lives, a fact which itself should generate some careful thought about the balance between costs and benefits. But NICE's benchmark figure is operating at more than 20 times the average healthcare expenditure per person ***per annum***, which seems to me to be rather excessive. So I would set the bench work no higher than £18k. This is only my opinion, of course, but then £30K is only ***their*** opinion, and I think my opinion has one great advantage over theirs, in that it is entirely unbiased, whereas it is obvious that a high threshold makes NICE more popular with the "sellers" by allowing more things through the net. But NICE is supposed to be on the side of the "buyers", who instead of being free to accept or reject NICE's advice, find themselves mandated to implement it. Something needs to give…

Clinical Guidelines

But now let me leave TA's for a while and look at NICE's role in the development of clinical guidelines. These two streams of work have developed in very different ways, and for perfectly understandable reasons. Technology appraisals are narrowly focused upon a tightly defined problem for which there is believed to be enough good quality evidence to justify the application of the excellent analytical skills that NICE can bring to bear upon it. Clinical guidelines, on the other hand, focus on a broad stream of activity following patients through a series of episodes of care and offering advice both to practitioners and to patients on how best to proceed at each critical juncture. Although there will be good evidence at some points, there will be virtually none at others, so the gaps will be filled by offering advice based on expert opinion. Unlike the technology appraisals,

adoption of the recommended clinical guidelines has no mandatory element in it.

The first question that arose in my mind over this was why NICE was into clinical guidelines at all, since this seems to be essentially a matter for the Royal Colleges to sort out. But I soon saw that the reason for NICE being a driver here is once more the need to apply cost-effectiveness thinking. The classical purpose of a clinical guideline has been to create decision rules which maximise the likelihood that the patient will emerge from the process with the best achievable level of health. For this purpose, at each decision node it will be important to choose the path which maximises effectiveness, without worrying about costs. But NICE's purpose is broader. It wants decision rules created which maximise the chances that ***the entire population*** will emerge from the process with the best achievable health. This means taking into account the sacrifices imposed on others by devoting scarce resources to the subset of NHS patients who are the subject of the particular guideline. These sacrifices are proxied by the costs incurred at each decision-node, which is why NICE's guidelines have to be created according to cost-effectiveness criteria, and not just based on effectiveness criteria.

But having justified NICE's role in the guideline process, my next problem was how to relate the Clinical Guidelines to the Technology Appraisals. I understand that NICE would like to integrate the two, and see them as different ways of looking at the same underlying set of decision problems, it then becoming a matter of judgement whether it would be more productive to approach such problems as a Technology Appraisal or as a Clinical Guideline. I think that that is an excellent idea, but one fraught with difficulties, both managerial and intellectual. The managerial ones I must leave to others, but the intellectual ones I will now examine more closely.

The way guidelines are "scoped" at present is by assembling a group of people who are knowledgeable in the territory and getting them to lay out the problems that arise during a complicated path of diagnosis and treatment and follow-up, and turn these problems into a list of questions that need to be answered. Then a remit is drawn up for a guideline development group, and for those bits of the remit where

enough evidence exists to make a systematic review worthwhile it will be undertaken. At some point economists will get engaged in this activity, and not just as cost-accountants, but as people with an interest and expertise in problem definition and outcome measurement as well. But I think that guideline development needs to be strengthened from the outset by injecting into the process a strong dose of decision-analytic expertise, so as to ensure that the whole territory is mapped out in a systematic way, rather than leaving the creation of a comprehensive flowchart until later, when all the bits and pieces on which we have more information have been sorted out. I find it useful to visualize the development of clinical guidelines as a form of mapmaking, since doing so takes me back to my favourite subject at school... cartography! This has led me to picture a clinical guideline as a map of how to get from A to Z as quickly and safely as possible. When doing so, I imagine that A is a set of presenting symptoms, and Z is a set of endpoints, reflecting health outcomes, with the patient/travellers having clear preferences about where they do and do not wish to finish up!

Map-making

When creating such a map you need to know something about traffic volumes and possible routes, so that you can separate the highways from the byways, and you need to know the points at which congestion occurs, the accident blackspots, and the routes to avoid in bad weather. But you also need to know something about the characteristics of the travellers, such as whether they are relying on public transport, or travelling by car or on a bicycle, or are making their way on foot. You also need to know how risk averse they are in terms of the trade-off between time, convenience and safety. To do this we need not only a large-scale map of what to do at particularly tricky junctions, but also a small-scale map of the entire system covering all of the relevant highways and byways, and estimating the traffic flows along each. In technical terms this would mean formulating the decision-problem according to decision-analytic principles, distinguishing choices from chance events, and ensuring that everyone who enters the system is followed through and accounted for somewhere or other, including those who might otherwise disappear through departing from the recommended highway.

It seems to me that clinical guideline development is very analogous to traffic engineering, and it does not take a very vivid imagination to identify some interesting parallels. The first of these is obviously the similarity between traffic congestion and waiting lists, both of which slow up people's journeys between A and Z! A second parallel is the desire to reduce accident rates, which in the world of health care has its counterpart in the reduction of peri-operative mortality and complication rates. But even the overall objectives of each system have a close parallel, with traffic engineers try to make road journeys quicker and safer and healthcare engineers trying to maximize QALYs, all within the budget available.

Dangerous though it is to push analogies too far, I want to persist with traffic engineering a bit longer in order to examine what might happen when the results of a Technology Appraisal are dropped into a Clinical Guidelines route map. The way things are set up at present, the guideline must incorporate whatever recommendations emerged from any pre-existing technology appraisal that falls within its scope. This is partly for the obvious reason of maintaining consistency between the two streams of activity, but it also reflects the fact that (in principle at least) the technology appraisal will have utilised the very best evidence and applied the most rigorous methodology, and so should carry much more weight than anything that will be being done at the guideline development level.

Suppose the Technology Appraisal had been about the optimal phasing of traffic lights at a particularly tricky road junction through which a lot of traffic has to pass. A great deal of data would have been collected about traffic flows, separating buses from the other vehicles, and as a result the road layout had been changed to minimise the time that vehicles had to wait, and a new set of rules had been generated about which vehicles in which lanes should have priority. These rules have now to be incorporated in the guidelines. But suppose that when scoping the technology appraisal, pedestrians had been excluded, because there was little or no data about them and so the traffic engineers' models ignored them. As a result, it could turn out that at busy times pedestrians cannot get across the junction safely at all. But the Clinical Guidelines do have to take pedestrians into account, even though the Technology Appraisal ignored them, and so in the guidelines

it might be necessary to modify the rules emerging from the technology appraisal. Having superior evidence and a superior methodology is not the end of the story, the relevant decision problem has to be brought into the picture too. Which makes it even more important for the clinical guideline to be based on a really thorough decision analytic framework so that this is made abundantly clear.

The benchmark once more

My final point about guidelines concerns the cost-per-QALY threshold, which seems rarely applied in clinical guidelines, perhaps because it is rare to be able to make a cost-per-QALY calculation in the first place. This is primarily because the unit of effectiveness that is usually applied in guidelines is not the QALY but whatever clinical effectiveness measure is conventional in each particular field. This was fine when the effectiveness of treatments in that disease field did not have to be compared with the effectiveness of treatments in any other disease field, but that is no longer the case. Just as NICE needs QALYs in order to make comparative judgements across diseases in the Technology Appraisal field, so it needs them in the field of cost-effective clinical guidelines, and for consistency it needs to apply the same threshold to both enterprises. I am told by people in the guideline business that applying a £30k per QALY benchmark to a guideline would be so generous that it would be rare for a guideline ever to fall foul of it. As a consequence I understand that there has been some discussion about applying a lower benchmark to Clinical Guidelines than to Technology Appraisals. That would be a totally illogical thing to do, since both are part of the same decision process. But there is an obvious solution to this dilemma… how about applying the same benchmark figure of £18k to both?

So my main criticisms of guidelines as they are developed at present is that insufficient weight is given to getting the decision tree properly articulated, that they should start with a flowchart setting out the traffic flows in some detail, and that they need to go whole-heartedly down the cost-effectiveness route, using the same cost-per-QALY threshold as is used for the Technology Appraisals. This means that both cost estimates and QALY estimates need to be generated within the guideline development process.

"An impossible task" did I hear someone mutter? If creating such a map is indeed horrendously complicated, it is because reality is horrendously complicated, but if the traffic analysts can do it, surely the health care analysts can do it too! Indeed, the more complex the reality is, the more dangerous it is to rely on intuitive short-cuts rather than careful analysis. Fundamentally what I am asking for is simply explicitness and transparency, two of NICE's guiding principles. I want all omissions and short-cuts documented carefully so that we can see what has been ignored for the sake of simplicity or because of gaps in our knowledge. One day we may need to go back and do better, so we shall need to know where to focus our attention.

Equity and variable thresholds

Next let me turn to equity considerations and the role of variable thresholds. I accept the need for some flexibility, but in the interests of explicitness, transparency and consistency, the circumstances in which this discretion is exercised, its justifying principle, and the weight given to it on each occasion, all need to be stated. In this way NICE will slowly build up its own case law, which will eventually need to be codified into a supplementary set of rules known to all and approved of by the citizenry.

As I understand it, at present this variation in the benchmark may occur for two broad reasons, the first of which is that it is believed that the evidence base is not only unreliable but actually biased in one direction or the other, and for some reason or other this cannot be handled during the technical appraisal of the evidence, so it falls into the lap of the Appraisal Committee instead. This seems a rather inappropriate use of benchmark flexibility, being rather like attaching a risk premium to the discount rate in an investment appraisal, when what should be happening is a careful analysis of the risks of being wrong and where they come from, and trying to remedy the defects at source rather than in this rather blunderbuss fashion. But perhaps the underlying fear is that once something is allowed to proceed, even with explicit restrictions over its use, it is likely to get out of control and it will be very difficult to retrench, so in such circumstances it would be best to err on the side of caution and apply a lower threshold so as to keep matters in check for a while.

The second broad class of reasons that might justify variable thresholds concerns equity considerations, and, in particular, the view that some patients are more or less deserving than average. As many of you will know, I am a firm believer in such discrimination, having argued for some years that an extra QALY going to someone who is otherwise unlikely to enjoy a fair innings in terms of lifetime health should be more highly valued than an extra QALY going to someone like me, who has obviously enjoyed more than his fair share of good health. So I suppose that at a personal level I should be relieved that NICE's Citizen's Council were not persuaded that an acceptable equity argument could be constructed on a fair innings basis, though I did my best to persuade them otherwise. And I have to accept their judgement, since they are certainly better able to represent the views of the population of England and Wales than I am. Incidentally, they are also of the opinion that we should not be swayed in our judgements by the suspicion that a person's current health problems might have been partly due to their own foolishness in not heeding well-publicised evidence about healthy and unhealthy life styles. For the Citizens' Council, Charity clearly outweighs Desert!

But I understand that Appraisal Committee members ***have been*** swayed by at least two arguments about desert... one relating to the end-of-life, and the other to sufferers from certain diseases. Giving priority to people at the end-of-life seems already to be embedded in the system, though if it is basically a caring gesture it would seem to me to more appropriately applied to financing end-of-life palliative care (which if successful in reducing and pain and relieving anxiety would earn a high QALY count anyway), rather than applying it to heroic surgery or distressing drug regimens. In principle this seems fine, but it needs to have clear limits set, since giving high priority to people at the end-of-life means giving low priority to those at the beginning or in the middle of their lives, who will as a consequence be denied treatments which would have generated more health for them than will be provided for those at the end-of-life.

But when it comes to the second category of cases favoured by the Appraisal Committee on grounds of desert, I have strong misgivings. I cannot see the moral significance associated with being a sufferer from one disease rather than another (once we have ruled out whether or not

it is their own fault). If I am in a particularly bad state with particularly poor prospects because of cancer, or because of a rare disease for which there is no other effective treatment, why does that justify regarding some other patient as less worthy, despite the fact that they are in the same situation as I am, and with the same expected benefits from treatment, just because they have a different disease which happens to be quite common? If this priority-setting is in the interests of research, then the intervention should be seen as a trial and financed by the MRC in competition with other research, not financed out of the budget of the NHS and at the expense of proven therapy.

In the equity business, I think the NICE Appraisal Committee needs to go down the route that was being followed by the Oxford Priorities Panel when faced with this situation, when it established by case law a set of principles which eventually made it possible to apply such variations consistently and in a fairly routine manner. It may take 4 or 5 years to get there, but without explicitness and transparency right now, it will continue to look like ad hoc fudging based on special pleading, an accusation that NICE must be able to refute if it is to sustain its reputation for even-handedness.

Something NICER than NICE?

My final topic goes beyond NICE's current remit and perhaps calls for a sister organisation to be set up rather than loading still more responsibilities onto NICE. Since I see the scoping of Technology Appraisals as often so narrow that their relevance is severely limited, and that of Clinical Guidelines as often so broad that their analytical underpinnings are very fragile, I have been trying to locate some middle ground that might enable us to get the best of both worlds. I am not confident that I have succeeded, but I am going to offer you my ideas anyway just in case you share my long felt want to expose the core business of the NHS to the same cost-effectiveness scrutiny as is applied to anything new!

As far as I can make out, some 300 interventions account for about 90% of the clinical work done by the NHS (as measured by the costs incurred). NICE currently embarks on about 30 new appraisals each

year, so at that rate it would take 10 years to work systematically and comprehensively through those 300 interventions, at which point it would doubtless be time to start all over again. In any one year the 30 chosen interventions could be spread thinly across all medical specialties, and each specialty would be expected to convene a small expert group to provide evidence about the costs and effectiveness of the selected intervention.

Since at the first stage we are talking about back-of-the-envelope stuff, I would be content on the cost side initially to use reference costs. But things are rather more difficult on the effectiveness side, since there I want to make estimates of QALY gains, which I suspect will only rarely be possible from existing data sources. Low in scientific status though it is, I am willing to use ***expert opinion*** to fill the gap, by getting a panel of expert clinicians to estimate length and quality-of-life profiles for patients, as my clinical friends did for me in 1985 when I was trying to estimate the QALY gains from CABG compared with drug therapy for patients with coronary artery disease. At that time I was using the Rosser Index for the quality-of-life estimates, but now I would use EQ5D, which would have the added advantage that it is the most commonly used method in the clinical trials that inform Technology Appraisals. For that reason too I would also like to see this same "expert opinion" approach using EQ5D to fill the gaps in the outcome measurement process in clinical guideline development, then all three processes could slowly converge on a common approach and generate a shared data base. It would not surprise me if the opinion of the experts erred on the optimistic side when it came to estimating the therapeutic benefit of the interventions which they themselves practice, but I can live with that at this preliminary stage. I expect the result of this process will be much what it was in 1985, namely that a consensus is eventually reached about the likely risks and benefits for different sub-groups of patients as defined by relevant clinical characteristics. At that stage we should be able to generate a rough cost-per-QALY ratio for each patient sub-group, and with luck this will be well below my benchmark of £18k for all of them, so we need go no further. But it is very likely that some of the sub-groups will have cost-per-QALY ratios higher than this, and they will then be designated as "on probation" and enter the second stage.

During this second, probationary, phase, simple monitoring of costs and outcomes will be required of a large enough representative sample of treated patients in the relevant category to be able to check whether the suspicion that it falls beyond the cost-effectiveness boundary is justified or not. If not, well and good, but if so, we enter a third phase, when negotiation takes place as to whether this treatment is really appropriate for this class of patient, or whether some alternative might be better. If this cannot be resolved, then it is time for this intervention to be referred to NICE for a full scale evaluation, possibly preceded by the setting up of a proper trial so that it is only permissible to continue treating such patients as part of such a trial.

Whether NICE is the best body to conduct this cost-effectiveness "triage" system I am not sure, but the sooner somebody starts doing it the more likely it is that the NHS will be able to survive the pressures it will surely be under soon when the current generous but unsustainable rate of resource growth (at more than 7% per annum in real terms) comes to an end. Perhaps it calls for the setting up of a sister body which might be called the National Institute for Cost-Effectiveness Reviews, which has the acronymic virtue of definitely being NICER than NICE! But if NICE would like to absorb this into its portfolio of activities that is fine by me. So long as someone does it and gets going with it soon, I shall rest content.

Epilogue

(which I hope will not prove to be indistinguishable from "Epitaph") It may well be that that final proposal of mine will turn out to be the only one that ends up unambiguously NICER than NICE. I am sure that each of my other suggestions will be met with howls of protest from someone or other. Nevertheless, here is my suggested action plan once more:

Change the QALY benchmark zone from around £30k to around £18k

All departures from the standard benchmark should be fully documented

Get Clinical Guidelines more formally based on decision-analytic principles and more susceptible to cost-effectiveness thinking

Embark on a comprehensive cost-effectiveness audit of the main existing clinical activities of the NHS

I remain naively (and even defiantly) optimistic about the power of argument to improve the world, though it is often a very slow process, and since my hourglass is running out, I am perhaps over-optimistic in hoping that some of this will happen during my lifetime. But that is in your hands, dear reader, not mine.