The Cochrane lecture

The best and the enemy of the good: randomised controlled trials, uncertainty, and assessing the role of patient choice in medical decision making

Klim McPherson

Abstract
This lecture aimed to create a bridge to span the conceptual and ideological gap between randomised controlled trials and systematic observational comparisons and to reduce unwanted and unproductive polarisation. The argument, simply put, is that since randomisation alone eliminates the selection effect of therapeutic decision making, anything short of randomisation to attribute cause to consequent outcome is a waste of time. If observational comparison does have any significant part in evaluating medical outcomes, there is a grave danger of “the best”, to paraphrase Voltaire, becoming “the enemy of the good”. The first section aims to emphasise the advantages of randomised controlled trials. Then the nature of an essential precondition – medical uncertainty – is discussed in terms of its extent and effect. Next, the role of patient choice in medical decision making is considered, both when outcomes can safely be attributed to treatment choice and when they cannot. There may be many important situations in which choice itself affects outcome and this could mean that random comparisons give biased estimates of true therapeutic effects. In the penultimate section, the implications of this possibility both for randomised controlled trials and for outcome research is pursued and lastly there are some simple recommendations for reliable outcome research.

Archie Cochrane’s great contribution to our discipline in 1972 was to describe a most important problem in health care and to posit a satisfactory solution. In his biography, he says of his work on Effectiveness and Efficiency:

“I decided to concentrate on one simple idea – the value of the randomised controlled trials in improving the NHS – and to keep the book short and simple.”

He did so in a manner which rings as loud and clear today as it did then. Nowadays, both the problem and solution are very familiar to us all. The problem was that much of the health care provided was unevaluated, and therefore possibly of no benefit, and the solution – randomised controlled trials – were the only secure way of knowing the truth of the matter.

Cochrane’s essential idea was that the reliable assessment of effectiveness was the only key to scientific health care, but such methods as he advocated had been disparaged by the medical establishment – paradoxically because they seemed to be less scientific than the more traditional basic methods of scientific evaluation. A problem of methodological imperialism which is with us still, but in different guise?

He argued forcibly that only when the effectiveness of a treatment or preventive measure had been established and quantified, could the issues of efficiency and equity be discussed intelligently. But in the absence of reliable information about effectiveness, the pursuit of optimal health policies was bound to be permanently elusive.

Since then Cochrane’s arguments have been largely and impressively accepted by health professionals. It is probable that without them we would not have all the welcome recent developments designed specifically to provide a more effective health care system. I am referring particularly to the NHS Research and Development initiative and the systematic concentration on issues of effectiveness and efficiency in the direction of research and health policy.

This is why there is a Cochrane lecture at the Annual Scientific Meeting of the UK Society of Social Medicine and I am honoured to be giving the lecture this year. I cannot do justice to my predecessors, who have contributed significantly and importantly to the work spearheaded by Archie Cochrane, but I hope to make some methodological points on the question of evaluation of health care, many of which I have made before. Largely, I want to try to divert any polarisation of methodologies which may, I believe, lead to an unnecessary prolongation of medical uncertainties. The issues are too important in my view.
(1) The value of randomised controlled trials

Since the publication of *Effectiveness and Efficiency* there have been several attempts to minimise or falsify the importance and relevance of Cochrane's message. Each succeeds only in accentuating its importance. It is impossible nowadays to deny the requirements for serious research on outcomes, but the gap between true knowledge of outcomes and the need to know and to understand, and hence to meet these requirements, remains large.

What was most important in 1972, and still is for us now, is how to bridge that gap. Last year Iain Chalmers gave a masterful demonstration of the practical extrapolation from Cochrane himself to the Cochrane Centre, clearly a central part of the "outcomes movement". This evolved from the National Perinatal Epidemiology Unit and took on the work of Iain, Richard Peto, Rory Collins, and others in summing the results of randomised trials of treatment. There is no question in my mind that these efforts in understanding the effectiveness of treatments, and hence helping people choose between options, are of enormous value. For me, the most exciting has been the work on breast cancer treatment.

The increasing use of randomised comparisons in judging effectiveness is wholly appropriate and must go much further.

**IMPEDEMENTS**

I want to leave aside, for the moment, the extent of the importance of uncertainty in health care decision making, and examine briefly part of the existing resistance to the use of randomised controlled trials. Cochrane discussed impediments to the wider acceptance of these in practice. There are three basic reasons for not carrying out randomised controlled trials where uncertainty exists:

1. Ethical objections
2. Lack of objective outcome measures
3. The resistance of those "threatened" by randomised controlled trials.

**THE ARGUMENTS**

We must be clear about the basic arguments, because in advocating greater, more systematic, and rigorous use of observational methods in judging efficacy (as I shall be doing later) it is very easy to associate oneself with a camp one does not wish to be associated with. There have been many exponents of the view, encapsulated by Feinstein, that "Cause-effect can be evaluated observationally" is scientific heresy only if two cogent scientific realities are ignored - the first, that some things are not amenable to experimental intervention, and the second, that randomised controlled trials are too expensive, too difficult, or too controversial. Whether or not such notions derive from feeling threatened is of secondary importance only, but they are common.

Of course trials are expensive and difficult, but from a cost effective point of view they provide hard unbiased evidence about the relative effectiveness of common, important, and expensive treatments. It is absolutely no part of my argument to disparage randomised comparisons at all - least of all because they are too expensive or complicated. The end justifies the means most of the time. To accept these kind of arguments is really to suggest that some alternative is generally better and that somehow it does not matter, because it is inevitable, that most treatments will not be evaluated by randomised comparisons.

Feinstein and others (Colin Dollery, for example, in his Rock Carling lecture) assert that most treatments will be assessed by a mixture of insight, theory, acumen, and uncontrolled observation. Of course large effects can be reliably detected like that, but from now on I will be discussing uncertainties over small but important effect differences in common medical practices which are poorly understood and are simply not amenable to evaluation in this way.

There is a more extreme current idea - which is just plain wrong - and this is that the use of databases which record everything about patients can easily, as a consequence of recording everything, be used to adjust for all confounders and then compare treatments as if there had been no selection. This has to be dismissed as ridiculous. Unknown and unsuspected confounding is mostly important and always elusive.

The great idea of Fisher and Bradford Hill, so ably developed by many, including Archie Cochrane, is one of the few panaceas of our time. In the context of medical interventions, the essence of the argument has been repeatedly put by David Byer. He says that "Epidemiology is an essential discipline ... for assessing the importance of exposures ... there is a disassociation between the reason for exposure and the outcome. On the other hand, in medicine, the doctor chooses the therapy precisely to affect the outcome and for no other reason". Archie's view was most succinctly put, and poignant nonetheless: "Observational evidence is clearly better than opinion but it is thoroughly unsatisfactory".

**UNCERTAINTY, SUPPLIER-INDUCED DEMAND, AND CONSUMER PREFERENCES**

Let me restate the basic premise in case subsequent remarks should be misunderstood. Randomised comparisons provide the most rigorous assessment of the therapeutic effect of particular interventions. The problems they are set to help solve are, however, formidable, for decision making in medicine is rarely quite as simple as choosing the best treatment for each objective diagnosis, even when it is known. In my opinion the dominant issues in health care now consist of: firstly, understanding the extent of true uncertainty and, secondly, relating this to the quality and costs of care. Then, since medical care is simultaneously and consequently emerging from an era of paternalism and medical domination, all decision making now has explicitly and
increasingly to take account of two important concepts. These are the role of supplier-induced demand and, secondly, how to accommodate consumer preferences into decisions. These things were less important in Cochrane’s day, and it seems to me they bring with them new emphases and responsibilities, both in the method and the mechanism of understanding treatment outcomes. Of course, the scientific/biological methods of medicine and its evaluation will always remain dominant, but the implications of these other considerations must at least be properly assessed.

(2) The extent of true uncertainty in medicine
I would like to propose something I have often suggested before which will complement the existing method of setting research priorities — this is the routine monitoring of practice variation and examination of plausible causes.

Proposal 1: Routine monitoring of practice variation and examination of plausible causes
In the determination of health care outcome research priorities, several criteria are obviously foremost. These are set out in table 1. This is clearly an essential list but of course tends to emphasise the glamorous, the interesting, the pet obsessions, and some general hierarchy of perceived medical scientific status, as well as the important. The problem may also be that health professionals have wholly different assessment priorities from the well informed consumers. If so, this would be a form of supplier-induced demand that is essentially unstudied in the health services literature.

SMALL AREA VARIATION AND UNCERTAINTY
The true extent of clinical uncertainty can often be measured by the amount of small area variations in hospital admission rates (fig 1), on the assumption that it is usually safe to assume that important endogenous factors are stochastically constant. This is the “small area variation” argument10 glibly put. Thus, we can take away random variation and are left with a crude measure of uncertainty.11 As an index of uncertainty, this is not a measure we are used to. To measure priorities for outcome research there is a growing tendency to ask people what the major areas are ripe for technology assessment.

Variations can give a crude index of existing uncertainties as measured by practice variation. Such a measure may also include genuine preferences for particular outcomes. But what it does not include is bias in the assessment of priorities based on possibly illegitimate, subjective (and possibly self-serving) hierarchies of relative importance. In this way a complement to the assessment of priorities for outcome research should, I think, be utterly routine. If we do this, we discover the by-now routine differences between countries in several very common reasons for hospital admission and that the amount of variation between small areas corresponds sensibly with what we actually know about uncertainties. Examples are shown in table 2.12

ETHICS OF UNCERTAINTY
From the point of view of the appropriate place of clinical trials in the evaluation of treatment we now come across a major obstacle of principle. The formal method for assessing

---

Table 1: Basic criteria for assessment priorities in outcome research

<table>
<thead>
<tr>
<th>Fracture of femur</th>
<th>Fracture of forearm</th>
<th>Fracture of knee</th>
<th>Lower back injury</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardised admission ratio (log scale)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.00</td>
<td>1.50</td>
<td>1.00</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Figure 1: Small area variations in standardised hospital admission ratios for various disorders.

Table 2: Magnitude of systematic variation (in ascending order) for selected causes of admission among 30 hospital market areas in Maine: 1980-82

<table>
<thead>
<tr>
<th>Variation</th>
<th>Medical</th>
<th>Surgical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low: 1.0 fold range</td>
<td>Acute myocardial infarction</td>
<td>Inguinal hernia repair</td>
</tr>
<tr>
<td>Moderate: 2.0 fold range</td>
<td>Gastrointestinal haemorrhage</td>
<td>Appendecotomy</td>
</tr>
<tr>
<td>High: 3.0 fold range</td>
<td>Cerebrovascular accident</td>
<td>Major bowel surgery</td>
</tr>
</tbody>
</table>

priorities for research rests on experts expressing their own uncertainties. When these uncertainties are expressed and agreed (that is, are explicit) the ethical basis for randomisation is widely accepted. When uncertainties are observed by epidemiological methods, as the most plausible explanation for small area variation, the uncertainties are implied only, and may be an inadequate basis for ethical randomisation, at least as far as the responsible clinician is concerned.

**WIDE VARIATION**
The most important finding from small area variation studies is that most hospital admissions are associated with variation as large or larger than for hysterectomy (table 3). Of course some of this variation may also be a measure of preferences. It is not difficult to imagine systematic differences in preferences between small areas with the current state of both the knowledge of outcomes and of information provision. Hence, most admissions are likely to be associated with more uncertainty than is commonly expressed or admitted, both from the clinician and the consumer, as to which is the appropriate treatment. Certainly, the most widely varying procedures do come up in explicit hierarchies of uncertainty but the common causes of hospital admission which vary importantly rarely do. From the point of view of establishing the appropriate role of randomised comparisons this might represent a temporary obstacle only, since the uncertainties are implied and not yet explicit. If they are legitimate, nonetheless, they must eventually become explicit.

**ADDITIONAL REASONS**
To add to Archie Cochrane’s possible reasons for not doing clinical trials there is now an important new fourth, and incidentally, a fifth, that is anyway quite well understood, and I won’t dwell on it. These reasons are as follows:

(4) Uncertainty is not well enough established
(5) Reluctance to compare institutions or carers.

**APPROPRIATE CONTEXT**
Much is going on along these lines but there is uncertainty that the appropriate context is always made clear. Thus, users of databases might well be disparaged by others, just as cost effectiveness studies might be ignored. The role of all of these methods has to be considered strictly in relation to the nature and the extent of true current uncertainties. Cost effectiveness studies, for example, are useless with unreliable information about effectiveness, if the plausible range includes zero or harmful effects. Likewise database analyses are a waste of time (except in circumstances discussed below) when large, unconfounded, randomised comparisons are available.

That is why organisations like the PORT’s in America have been useful, because they enable, in principle, the true research priorities for a particular outcome question to be properly evaluated, unconstrained by data or disciplinary constraints. Then these objective priorities point to and release these constraints where possible. At about the same time as Effectiveness and Efficiency was published, a seminal experience I had in America produced a book called Costs, Risk and Benefits of Surgery, which used many of the kind of, essentially ancillary, methodologies enumerated above. Archie Cochrane was asked to review it, of course, and his comment was: “An excellent, courageous, pioneering book. I read it with profit and pleasure.” So clearly Archie was also unambiguously not in favour of letting the “best become the enemy of the good”.

---

**Table 3** Proportion of hospital admissions categorised by characteristic variation between small areas (data on 400,000 hospital admissions in Maine in the 1980s)

<table>
<thead>
<tr>
<th>Amount of variation</th>
<th>Typical admission</th>
<th>% Admissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Hernia repair</td>
<td>1</td>
</tr>
<tr>
<td>Moderate</td>
<td>Appendicectomy</td>
<td>9</td>
</tr>
<tr>
<td>High</td>
<td>Hysterectomy</td>
<td>42</td>
</tr>
<tr>
<td>Very high</td>
<td>Disc removal</td>
<td>32</td>
</tr>
<tr>
<td>Extremely high</td>
<td>Tonsillectomy</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

suggested by Feinstein et al, and as a consequence, the explicit uncertainties are low but the implied uncertainties are very high. The criteria vary, the use rates vary, the costs vary, and presumably the outcomes vary. But nowhere did treatment for benign hypertrophy of the prostate appear on the upper echelons of research priorities—simply because most people thought they knew what they were doing.

Firstly, variation studies indicated a problem of implied uncertainties. Secondly, basic decision analysis, with comprehensive literature reviews, strongly suggested that the cited liberal reason for intervening surgically in patients with minimal symptoms—to prevent the condition worsening and then suffering a less than optimal outcome for surgery—was just not justified. The need for more reliable data was obvious but a clinical trial could not be contemplated, simply because the level of uncertainty among urologists was not admissible.

Databases of complete follow up of all operations indicated quite remarkable discrepancies in mortality for surgery than had ever been published from case series in prestigious institutions. An unsuspected finding emerged concerning the new non-invasive operation (the development of which was partly why randomised comparison was not deemed to be indicated), suggesting a systematically higher mortality in patients who underwent transurethral resection than open operation (table 4).

Such evidence about relative efficacy and safety as this provides is, of course, wholly compromised by the possibility of selection; the worse prognosis patients being offered the more benign treatment while similar patients are refused an open operation. Several studies since this publication have tried to adjust for confounding by taking account of case mix, and the evidence is inconclusive. The latest, by Alvan Feinstein et al, seems to suggest case mix as the explanation but the study is far too small to refute an independent effect of the surgery.

ESTABLISHMENT VIEWS
Most interesting are the responses of the medical/urological establishment to these results. In the USA a meeting of the executive council of the American Urological Association, convened shortly after the Roos study was published, emerged with the following statement, which explicitly acknowledges the uncertainty associated with the choice of open surgery and transurethral surgery.

"Though the mechanisms of a direct causal relationship are not apparent at this time, the weight of statistical evidence suggests that a randomized prospective study of international scope involving large numbers of patients, is indicated."

In my view this is the most rational response to these data, because the question is clearly one of attribution which can only be resolved by a randomised controlled trial. In contrast, an editorial in the British Medical Journal

---

**Table 4** One year age standardised mortality after prostatectomy in three large databases, 1963-85

<table>
<thead>
<tr>
<th>No of procedures</th>
<th>Mortality (%)</th>
<th>Transurethral resection</th>
<th>Open prostatectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>3600</td>
<td>7.6</td>
<td>5.7</td>
</tr>
<tr>
<td>Manitoba</td>
<td>12000</td>
<td>6.0</td>
<td>4.1</td>
</tr>
<tr>
<td>Oxford</td>
<td>5000</td>
<td>10.5</td>
<td>7.6</td>
</tr>
</tbody>
</table>


---

Figure 2. Crude rates for prostatectomy in various countries in the 1980s.

Figure 3. Adjusted relative risks of 5 year mortality after transurethral and open prostatectomy in various published studies.
does not budge an inch in the direction of recognising any uncertainties whatsoever:

"The fact that a recent review has suggested that TURP may not be as good as urologists have claimed does not detract from the important advantages of this technique in treating benign prostatic hyperplasia".

A wonderful quote which manages to dismiss the finding completely by calling it a "review" and then attributing a mere suggestion to that review. The important "advantages" of transurethral prostatectomy suggested by the data from a complete enumeration of all prostatectomy patients in three databases are an excess in one year mortality of around 40%. Of course, since the data are inconclusive and inconvenient they can be ignored, while inconclusive and convenient, but almost all small, selected, uncontrolled case series, can be cited, essentially as established. This surely ought not to happen.

WHAT DOES THE DIFFERENCE MEAN?

One is, of course, left wondering whether the observed difference in the databases is attributable to the operation type, for there are many plausible explanations for the transurethral prostatectomy really having a higher risk of mortality. These are mostly concerned with hypothermia associated with unheated irrigation fluid, for instance, or the nature of the irrigation fluid itself. The results of the only randomised study\(^2\) is interesting because it is consistent with an increased mortality. The five year mortality was 21% in the transurethral prostatectomy group and only 6% in the open prostatectomy group, however, the numbers were very small, some 40 in the transurethral prostatectomy group and 30 in the other.

A question we are left with is the extent to which consumers of this kind of health care would, given all the evidence outlined above, actually choose a transurethral prostatectomy in preference to an open operation.

(4) The role of patient choices in decision making

In deciding upon an appropriate medical or surgical treatment the choice obviously ought to be determined by the probabilities of particular outcomes associated with each choice, combined with the individual preferences for those outcomes. Where there is uncertainty about the true size of these probabilities, the procedure is complicated, not least because of the demonstrated placebo effect of medical certainty. That is to say, the paternalistic doctor may actually be doing his (gender chosen advisedly) patients more good than one who expresses the uncertainties honestly. So perhaps the above editorial in the *British Medical Journal*\(^3\) is dedicated only to making patients with benign hypertrophy of the prostate better - who would otherwise not improve if they suspected that the treatment could possibly have a significant downside. I will pursue some of the implications of this below.

<table>
<thead>
<tr>
<th>Table 5: Common conditions for which patient preferences may be dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign conditions of the uterus</td>
</tr>
<tr>
<td>Angina pectoris</td>
</tr>
<tr>
<td>Cataracts</td>
</tr>
<tr>
<td>Benign hypertrophy of the prostate</td>
</tr>
<tr>
<td>Gall stones</td>
</tr>
<tr>
<td>Menopausal symptoms</td>
</tr>
<tr>
<td>Breast cancer</td>
</tr>
</tbody>
</table>

PATIENT PREFERENCE

Meanwhile an important point receives little attention in the effectiveness literature; that for some decisions patient preferences are dominant while for others they are of no practical consequence, except for extreme cases of religious bigotry. This depends of course on the probabilities of different outcomes associated with each choice and the nature of the outcomes. For some decisions the probability of death associated with one choice is unity, while for another choice it is less than one in a thousand – which of course is no choice. However, much in medicine is dominated by choice and hence it is hoped by communicating the probabilities associated with different choices. Some common examples are shown in table 5.

THRESHOLD QUESTIONS

Most often the nature of the therapeutic choice is not about which procedure, but is with the timing and whether to do something or not; it is a threshold question, which dominates much of the variation in observed rates (see fig 4). It is simple to demonstrate that quite small differences in the threshold of signs or symptoms at which intervention happens on average can be responsible for dramatic differences in the rates of intervention. By the nature of the problem these differences in treatment policy are seldom assessed unbiasedly by randomised comparison; partly because the uncertainties are not sufficiently stark and partly because consumer choice is (often illegitimately\(^4\) cited as the driving force behind the decision. But the cost differences are often enormous.

The nature of the choices can be illuminated by using the methodologies in section (3), and sometimes it can be seen that the medical certainties are based on implausible theory. Rarely can the demonstration be sufficiently convincing to cause appropriate randomised comparison, however. As was shown above if vested interests do not want to believe plaus-
Table 6 Mortality of patients on placebo in double blind trials

<table>
<thead>
<tr>
<th>Patient adherence</th>
<th>Trial</th>
<th>Coronary drug project research group̈</th>
<th>β Blocker Heart Attack Trial 1y mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5y mortality</td>
<td>1y mortality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Crude</td>
<td>Adjusted</td>
</tr>
<tr>
<td>Poor</td>
<td>28 2</td>
<td>25-8</td>
<td>7 0</td>
</tr>
<tr>
<td>Good</td>
<td>15 1</td>
<td>16-4</td>
<td>3 0</td>
</tr>
</tbody>
</table>

P < 0.00001

suboptimal. These preferences may all be compatible with current medical knowledge which may be subject to very wide uncertainties (see above) nonetheless. Supplier induced demand clearly has strong theoretical, if not empirical, support. 24

RELIABLE INFORMATION
So the solution is, of course, to evaluate the outcomes to provide reliable information from which to make these decisions. In the short term this remains a forlorn hope, for reasons discussed above. Meanwhile, to test the above hypothesis directly, Wennberg et al20 have produced interactive video discs designed to inform patients about the extent of current scientific knowledge and uncertainty associated with particular clinical decisions. We are evaluating this technology in the UK using randomised methodology.

The questions we are trying to answer are the following. Firstly, do patients like having extra information and the chance to participate in the decision making process? Secondly, do the clinical decisions change as a consequence of providing this information? Thirdly, do patients do better, with respect to symptoms, survival, or quality of life, as a consequence? First indications in the USA are that the decisions made with more information are very different from those made without and that patients are in favour of the extra information. (See fig 5.)

EFFECT OF CHOICE ON OUTCOME
A more fascinating aspect still of the role of patient preference in clinical decision making comes from studying the effect that choice itself has on therapeutic outcome. This is my third proposal.

Proposal 3: Investigate areas systematically where treatment preferences affect outcome importantly
I will cite two important examples of observations from clinical trials where the indirect evidence strongly suggests that choice itself affects hard outcomes such as survival. 23 29

Table 6 shows a summary of the results of double blind clinical trials in which the outcome among patients given placebo is compared according to whether patients actually took the prescribed pills. Choosing to take the placebo pills is then the treatment comparison in circumstances when the patients did not know that they were on placebo. Both trials, after adjusting for all measured possible confounders, gave a strong hint that choice itself can dramatically effect important measures of outcome. There are many other examples of this phenomenon in the published reports and the mechanisms could, of course, be various, a synergistic combination of biology, immunology and pathology, with psychology much of which is quite well understood already.
The best and the enemy of the good

PLACEBO EFFECT
Professor Dave Sackett has cited five examples in the literature of the placebo taker faring better if they took the drug seriously. In the original hypertension trial in Virginia a 30% reduction in mortality was observed among compliant placebo takers (p = 0.08). Similar results have been found in weight loss and in alcoholism, schizophrenia, and in cancer patients. Hence not only are patient preferences intrinsically important in offering participation in affairs which intimately affect their lives, but also perhaps because choice itself is sometimes apparently of therapeutic benefit. This is important because such a possibility may have important methodological implications. Either patient self selection affects outcome because of confounding with prognostic variables and hence randomised comparisons are essential or, alternatively, choice or control affects outcome and hence unbiased randomised comparisons are biased (downwards).

In the coronary drug project result reported above the idea that residual confounding can explain the difference in mortality is implausible, except that all the variables used to adjust were, of course, those measured at baseline, and possibly are thus compromised as potent confounders.

RANDOMISATION AND BIAS
The idea therefore that randomisation itself can give rise to biased results about outcome comes as shock to many people. But clearly when treatments are allocated randomly patients are deprived of choice and if choice and control can be of therapeutic benefit then a randomised comparison might provide a biased downwards estimate of the effectiveness of treatment. We are familiar with the idea that efficacy might give an euphoric estimate of effectiveness, because things that matter might change outside a clinical trial for the worse. I am now suggesting a possible countervailing influence where things might get better outside randomised comparisons. This would be true if only if choice or control itself affects outcome. If true, another reason for not generalising altogether from randomised trial is that they might thus be systematically biased. Hence a sixth reason for not wanting always to attribute outcomes by using randomised controlled trials is:

(6) Choice or preference or control affects outcome

There are, of course, many plausible mechanisms by which such an effect could be important. The literature on the physiology of the placebo effect is dominant here as well as the evidence, for instance, implicating social or professional control in the aetiology of coronary heart disease.

The role of choice in evaluating outcomes

Clearly the role of choice or control needs to be systematically studied in the following context explicitly. It is important here to distinguish between preferences for treatments because of their known effects and the role of choice in itself determining outcome to some degree. They will, in general, be inextricably confused because I suspect that choice only affects outcome via belief in particular attributable consequences of that choice. But here I am trying to emphasise the latter phenomenon as part of the therapeutic process itself. Areas in which patient preference may have an important role are as follows:

(1) Randomised controlled trials
(2) Therapeutic efficacy
(3) General priority setting
(4) Particular therapeutic choices

PREFERENCE TRIALS
One way to investigate this more closely in randomised trials is to mount preference trials along with prospective randomised trials to test for important therapeutic and biological effects and simultaneously any possible synergy with treatment choice: an example is shown in table 7. The explicit idea here is to compare the outcome in patients who have chosen a particular treatment with those who have been randomised to the same treatment. Obviously, there will be total confounding between choice and selection on known or unknown prognostic indicators. At least we can measure the potency of known prognostic indicators. But the alternative to not trying to investigate this phenomenon is, it seems to me, to deny its existence, and there is too much of that going on already. It seems quite plausible that for some conditions this will be important and for others utterly irrelevant, but it would be good to know which.

The challenges we are faced with are not only to understand about outcomes and their determinants, which is formidable enough, but also to assess routinely knowledge on the quality of life and functioning from a patient’s point of view. Most importantly, the main challenge is to create a decision making environment which is capable of optimally using reliable information on and data about patient outcomes.

(6) Conclusions

The evidence for common uncertainties in clinical practice is overwhelming and, clearly, important car ing away at key parts of it is the dominant component of Archie Cochrane’s legacy. I have sought to draw attention to some of the potential problems in priority setting and in the evaluation of treatments, where

Table 7 Preference trials – a way of investigating choice

<table>
<thead>
<tr>
<th></th>
<th>Measure therapeutic effect of choice, if any: for example randomised controlled trial of patients who have seen interactive video disc versus normal care</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wensberg proposal: randomised controlled trial of preferences with information versus randomised controlled trial with information – for example surgery, drugs, watchful waiting</td>
</tr>
</tbody>
</table>
consumer input is important and possibly neglected. I believe that these considerations are important methodologically as well as substantively and their methodological implications are poorly investigated as yet.

The true role of uncertainty in medicine has yet to be properly investigated also, and this is partly because it fits uneasily into the dominant medical paradigm. Much of what happens in medical school would change, much of what happens in clinical practice would change, much that happens in media coverage of medicine would change, and perhaps much of what happens in the evaluation of outcome would change, if the true nature and extent of uncertainty in medicine were readily acknowledged. Perhaps effectiveness itself would change too, but until we know when and how, and in which direction, there seems little point in not investigating whether it would. Certainly the politics of health care planning would change.

STYLE DIFFERENCES

My final proposal is again something I have often suggested which is to take greater advantage of a manifestation of uncertainty, differences in practice style.

Proposal 4: Practice style differences exploited in natural experiments of outcome, where randomisation is not feasible

Systematic practice style differences are common and they exist because clinicians adopt different criteria for intervening, largely as a consequence of assuming that informal, theoretical, and anecdotal methods of assessment of outcome are good enough. Since they exist it always seems an enormous waste not to take advantage of these natural experiments where treatment at the margin is not chosen, as David Byer would have described things, entirely by objective measures of prognosis but much more as a random consequence of history. Figure 6 shows variations in the use of prostatectomy in English health regions in the 1970s.36 These variations in practice style have yielded no information whatsoever about their consequences for the patients, and there must be some.

OBSERVATIONAL METHODS

I believe that some of the observational methodology proposed is, in principle, cheap and could be routine. It is not so cheap now because one has to develop mechanisms not already in existence to capture all cases and to allow useful follow up of outcome as well as to attract the collaboration of clinicians. It cannot replace randomised controlled trials except in circumstances where the attribution of cause with respect to outcome is unambiguous, and this will sometimes happen. We will have to get better at determining the causes of observed differences by knowing about the potency of plausible confounders.4 What I am proposing is systematic study of enormous numbers of cases so that even rare and long term outcomes can be systematically compared.

CRITERION FOR RANDOMISED COMPARISON

Randomised comparisons are only practically feasible when the major primary outcome(s) of interest need to be unambiguously attributed to treatment choice and cannot be attributed without them. We are often concerned with secondary outcomes (like long term rare side effects) here and nobody is going to randomise patients when the primary outcome comparisons (functional status at 30 days for example) are essentially, or even approximately, understood. What is often required is a clear cohort collected at a time when there are uncertainties about the primary questions of outcome attribution and hence when the cohort is relatively uncontaminated by biased clinical selection on the basis of prognosis.

An example is the collaboration with the Royal College of Surgeons in their audit of all cases of prostatectomy for benign hypertrophy of the prostate. For the first time, and essentially incited by the other observational work on prostatectomy already quoted, they are studying outcome on some 4000 consecutive cases. The hypotheses being investigated are to do with the characteristics of patients who might benefit from this operation from among those now being offered the operation. Testing transurethral prostatectomy against open surgery remains a long way off I suspect. Clearly large numbers are not in themselves enough, but it emerges (conditioned somewhat by possible response bias) that fascinating differences, and in some cases fascinating similarities, in outcomes themselves raise important hypotheses to be tested individually in the future.

POSTAL ASSESSMENTS

However, what emerges clearly from this study, and another I have recently completed
in collaboration with Nick Black and others, is that patients who have received treatment are very willing to complete, even detailed, postal questionnaires about the state of their health. It seems slightly scandalous to waste the good will evidenced in table 8 in obtaining information which could prove to be extremely valuable. Record linkage has often been implemented to obtain follow up information on outcome, but it has obvious limitations.

Proposal 5: (Postal) outcome assessments by patients become routine, with and without treatment

Thus, if collected in a uniform and systematic manner, basic information not now available in any analysable form could usefully inform outcome research. This must eventually include, of course, patients with symptoms not deemed appropriate for intervention.

I conclude with a summary of the main points I have tried to make which encapsulate a hierarchy of methodologies designed simply to establish the effectiveness of health interventions. These main points are:

1. Monitor practice variations.
2. Enhance multi-disciplinary generation of appropriate uncertainties.
3. Determine where treatment preferences affect outcome.
4. Undertake natural experiments of practice style.
5. Routine long term postal follow up for important questions can be valuable.
6. Randomised trials for small, important, possibly confounded effects are essential.