Evidence-based medicine as Bayesian decision-making

Deborah Ashby1,* and Adrian F. M. Smith2

1Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, U.K.
2Office of the Principal, Queen Mary University of London, Mile End Road, London E1 4NS, U.K.

SUMMARY

We review two recent trends: the emergence of evidence-based medicine and the growing use of Bayesian statistics in medical applications. Evidence-based medicine requires an integrated assessment of the available evidence, and associated uncertainty, but there is also an emphasis on decision-making, for individual patients, or at other points in the health-care system. This demands consideration of the values and costs associated with potential outcomes. We argue that the natural statistical framework for evidence-based medicine is a Bayesian approach to decision-making that incorporates an integrated summary of the available evidence and associated uncertainty with assessment of utilities. We outline a practical agenda for further development.

Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

In the 30 years since the MSc course in Medical Statistics began at the London School of Hygiene and Tropical Medicine, both medicine and statistics have developed rapidly. Within medicine, one major recent trend is that of evidence-based medicine [1]. Over the same period, the application of Bayesian statistics has also increased, in particular in health research [2]. Both evidence-based medicine and Bayesian statistics integrate evidence in a systematic way, for decision-making purposes. At this conference, it seems opportune to examine these developments and how deep their parallels run.

Both for intellectual coherence and practical insight, there is much gain in assessing evidence-based medicine from the perspective of Bayesian decision-making. We review the recent take-up of Bayesian ideas in medicine and outline a systematic agenda for evidence-based medicine.

Section 2 describes the elements of evidence-based medicine. Section 3 introduces Bayesian statistics, reviews its applications in medicine for summarizing evidence, and addressing decision-making. In Section 4 we identify key issues arising from our review, and suggest an agenda for future development.
2. EVIDENCE-BASED MEDICINE

2.1. Background

A scan of medical journals over the last ten years reveals an explosion of interest in ‘evidence-based medicine’. Leading proponents Sackett et al. [1] have attempted to define it. Italics are ours.

‘Evidence-based-medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients’, taking into account ‘individual patients’ predicaments, rights, and preferences [using] best evidence from clinically relevant research’.

One area that has received a lot of attention is the collation of evidence. The randomized controlled trial is regarded as the gold standard for evaluation of new therapies [3]. The Cochrane Collaboration (http://www.cochrane.org) aims to collate systematically all randomized controlled trials through regularly updated systematic reviews to provide the best current evidence on different therapies. These are available in the Cochrane Library [4], disseminated through CD-ROM’s and via the Internet. A science of systematic reviews has evolved [5].

A more systematic approach is also evident in the development of clinical guidelines for different disease areas. These guidelines are most credible if they are explicit about their evidence base and about the values of various outcomes [6].

The pharmaceutical industry contributes to evidence-based medicine: in order to obtain marketing authorization for a product they have to provide regulatory authorities with structured evidence to demonstrate the safety, quality and efficacy of the product, justifying the indications, the appropriate dosages and any contra-indications. The application of similar ideas to other areas is known as ‘health technology assessment’. There is increasing interest in the relative merits of different treatment strategies, particularly with increasing financial accountability. In the U.S. health insurers exercise control over what treatments can be used and in the U.K. hospitals are developing approved lists of medicines. The National Institute for Clinical Excellence has been set up to appraise the relative merits of differing therapies [7]. This will be based on the available evidence, as are current licensing decisions, but also on cost-utility estimates [8]. This is already provoking vigorous comment [9], as it involves values and utilities, rather than just evidence of clinical safety and efficacy. An explicit aim is ‘the involvement of patients and the public in the decision-making process’.

These wider developments have in common with Sackett’s description of evidence-based medicine the need to take the best available evidence, and combine it with judgements about risks, benefits and costs to make decisions that will benefit the health of patients.

There exists a formalism, so-called Bayesian statistical decision theory, for such decision-making [10,11], which has the following general structure. An identified decision-maker is confronted with a set of possible actions, sometimes referred to as options or decisions. Each possible action has associated with it a set of possible consequences, one of which will obtain – but it is not known which.

The standard paradigm of rational decision-making now requires two forms of quantifications: first, the utility, sometimes referred to as values, typically constructed from cost-benefit trade-offs that attaches to each consequence; secondly, the probability of each consequence obtaining subsequent to a selected course of action. An index of (expected) utility for each possible action
is calculated by summing probability×utility over all the consequences that could follow from that action. The preferred action will be the one that has maximum expected utility.

2.2. Examples

So far this discussion has been in general terms. We now examine particular examples. For each we will identify the decision-maker(s), the possible actions, the uncertain consequences, the possible sources of evidence, and the kinds of utility assessments required.

Example 1. A couple presents wishing to try for a baby but concerned about the possibility of handicap. *Inter alia*, their doctor raises the question of whether to take folic acid supplementation in order to try and prevent a neural tube defect (NTD, also known as spina bifida or anencephaly).

As presented, the decision-maker is the couple. The possible actions are to take or not to take folic acid supplementation. The uncertain consequences are whether or not the pregnancy will result in a foetus with an NTD. The possible sources of evidence come from routine data to give the risk of NTD, and from randomized controlled trials to give the risk reduction associated from folic acid supplementation. The kinds of utility assessments required are the couple’s views on the seriousness of NTD, and the costs both financial, and in terms of side-effects. This example will be analysed formally in Section 3.

A second perspective here is that of public health policy. The decision-maker (in the U.K.) might be the Minister for Health acting on the advice of the Chief Medical Officer. The possible actions are whether or not to recommend on a population basis folic acid supplementation routinely peri-conceptually. The uncertain consequences are the fall in the incidence of pregnancies with NTDs. The possible sources of evidence regarding uncertain consequences are the same as in the case of the individual woman as decision-maker. The utility assessments involve the costs to the NHS budget of routine prescription of folic acid, and the costs of termination of pregnancy or care of people born with NTDs as well as more intangible aspects such as the desirability of reducing handicap.

Example 2. A patient is diagnosed with oesophageal cancer. He is advised that until recently routine treatment has been surgery, but a new suggestion is to precede the surgery by a course of several weeks of chemotherapy.

The decision-maker is the patient, who may effectively delegate to his doctor. The possible actions are to undergo surgery, or to opt for the combination treatment. The uncertain consequences are the length of his survival, and side-effects (such as severe nausea), and the delay in completion of treatment. The possible sources of evidence relating to his expected survival come from routine data such as cancer registries, and relating to the additional benefit of combined treatments from clinical trials. The utility assessment required is the patients trade-offs between extra survival, side-effects and time spent undergoing treatment.

An alternative perspective is that of the scientists running the trial to provide the required evidence. We consider the situation partway through such a trial in which patients are being randomized to the two regimens. The decision-maker is now the Data Monitoring Committee. The possible actions are to continue the trial or to stop it. The uncertain consequences are the effects on survival/quality-of-life of the patients who may enter it, and the information resulting from the trial, with associated uncertainty, that will be available to inform the decisions of future patients.
The possible sources of evidence are the data from the current trial, and any other available trials. The utility assessment required is the trade-off between the current expected utilities of outcomes for patients to be randomized and the value of more precise information.

Example 3. A 60-year-old woman presents for a routine health check. She has nothing obviously wrong with her, but wishes to reduce her future risk of heart disease and stroke. We consider just two options: that of giving up smoking, and of taking HRT long-term.

The decision-maker is the woman. The possible actions are to continue as before, or do one or both of giving up smoking, and embarking on long-term HRT. The uncertain (long-term) consequences are the risk of a heart attack or stroke, but also, for the smoking decision, the risk of lung cancer and, for the HRT decision, the risk of osteoporosis and associated fractures, and breast cancer. The uncertain (short-term) consequences are the pleasures associated with smoking, and the beneficial effects on skin and risk of unwanted bleeding that might arise from HRT. The possible sources of evidence for the effects of smoking come from epidemiological studies, and for the effects of HRT currently come from epidemiological studies, but will be informed by long-term randomized trials currently under way. The utility assessment required is the woman’s trade-offs between the short-term and long-term consequences. We will not be exploring smoking cessation or HRT from an individual perspective formally in this paper. This example is included to demonstrate the complexity of decision-making. However, we will look at stroke prevention policies from the health providers’ standpoint of choosing between screening and ‘watchful waiting’ in Section 3.

We contend that structuring thinking about problems in terms of the formal paradigm is always beneficial, even when the complete process of quantification and calculation to compare options is not followed through. The very process of systematically thinking through the component elements of the problem provides both a disciplined framework for individual analysis and also a common focus which can often prove invaluable as an aid to communication among interested parties. In specific examples the required quantification can be difficult – perhaps contentious – but even when precise quantifications may be difficult, the framework can provide significant qualitative and even quantitative insights by performing ‘what if’ calculations, or sensitivity analyses around a working set of quantifications. For example, such analyses might reveal that a particular action is preferred over quite wide ranges of probability and utility assessments, thus enabling a number of interested parties who do not agree precisely on assessments to agree, nevertheless, on the same preferred course of action.

Much detailed development and promulgation of case-studies is required to fully demonstrate the unifying contribution of the Bayesian decision-making perspective; the main purpose of this article is to argue that focusing on the decision structure at the heart of specific problems brings into sharp relief the forms of quantification that have to be made, and hence develops a critical awareness of the nature and quality of evidence required, even in cases where one does not currently feel comfortable in proceeding to a full implementation of the formalism.

2.3. Discussion

From clinically rather different examples, various themes emerge. There is a need for information, which can typically be obtained from some combination of routine data, observational studies and experimental studies such as randomized controlled trials. The information from these sources will have associated uncertainty. There are further judgements to be made about the applicability of evidence to particular clinical situations. The subsequent decisions depend both on the evidence
and on the utilities of the decision-makers. We now consider where the responsibility for the
decision-making lies, and the extent to which it is desirable to formalize the process.

The options available for a particular patient are contingent on decisions made at other levels
in the health system. For example, the decision to offer a service such as breast cancer screening
is a national one. This is currently offered routinely to women aged 50 to 65 in the U.K. The
lower limit rests on the relatively low incidence of breast cancer for the under-50s, but the upper
limit can be interpreted as society’s utility for ‘saving’ years of life in middle-age rather than the
elderly. A more formal rationalization of this decision might be in terms of ‘expected life years
saved’, which could then be weighed against costs.

The availability of a particular medicine as an option for an individual patient rests on a se-
ries of decisions: pharmaceutical company development approval by regulatory authorities, and a
willingness to pay for it by health care providers. Evidence about the drug’s efficacy and safety
will feature in all these decisions, but for the pharmaceutical company the utility of the drug will
largely be measured in terms of its ability to generate money for shareholders. The regulatory au-
thority will give a licence once there is evidence of safety and efficacy and favourable risk-benefit
for at least some patients. The health care provider will want cost-effectiveness, so comparisons
between competing alternatives are uppermost.

There are decisions to be made about collecting evidence about a treatment, for example, whether
to run a randomized trial. Once a trial is running, further decisions need to be taken about whether
to stop the trial early if the results indicate a clear preference for one treatment over another. Here
there is a complex balance of the utilities of future patients due to be randomized, who might
prefer to get the treatment thought likely to be better, versus a wider scientific interest where more
precise evidence of a treatment’s benefit is desirable. For trials where there may be reason to stop
early, responsibility for the decision is often delegated to a data monitoring committee.

There is a large literature on trading quality of life against survival, especially in diseases such as
cancer, with developments such as the quality adjusted life year (QUALY) [12]. In drug regulation
the concept of risk-benefit is essentially a utility trade-off, although it is rarely formally addressed.
In the evidence-based medicine literature, a new measure has emerged, of numbers needed to treat
(the NNT) [13]. This measures the number of patients who need to take a treatment in order for
one patient to benefit. It is claimed to have a role both in individual treatment decisions and in
decisions about the merits of offering treatments to groups of patients.

Finally there is more explicit acknowledgement of the generic need for a decision-making ap-
proach. A recent review of the Cochrane library said ‘good clinical decisions will often depend on
preferences for possible outcomes and attitudes to risk which will vary considerably among users
of evidence’ [4].

Given the increasing emphasis on decision-making, we argue that use of a more formal approach
can help clarify the process. It identifies the main elements that need to be considered, and, when
appropriate, gives a coherent means of combining those elements to reach a decision.

3. BAYESIAN STATISTICS AND ITS APPLICATIONS IN MEDICAL RESEARCH

3.1. Elements of Bayesian statistics

The logical foundation for handling the uncertainty component of the decision-making struc-
ture described above is Bayesian statistics. The basic concept of Bayesian statistics is very
straightforward. In any given decision-making situation, there are relevant quantities or outcomes we have observed and recorded, and other relevant quantities or outcomes we have not yet observed and recorded — and are therefore uncertain about. In order to make rational decisions, we need to weigh up all relevant uncertainties and provide some form of quantification. Bayesian statistics provides this quantification in the form of probability statements, having taken into account all relevant evidence provided by the quantities and outcomes we have observed and recorded. Textbook accounts of various aspects of Bayesian statistics are provided, for example, by O’Hagan [14], Lee [15] and Bernardo and Smith [16]. A wide-ranging collection of applications of Bayesian statistics in the biomedical field can be found in the volume edited by Berry and Stangl [2].

The mathematical result that ensures that the required probabilities hang together in a logical way is Bayes theorem [17]. In a medical decision-making context, the evidence takes the form of numerical data and the unknown of interest is a quantity, \( \theta \), say, which appears as the parameter in a statistical model of data variability. Bayes theorem takes the form

\[
p(\theta | \text{data}) \propto p(\text{data} | \theta) \times p(\theta)
\]

with uncertainties expressed via probability density or mass functions for the relevant quantities. Here, \( p(\theta) \) specifies the prior distribution of uncertainty for the unknown parameter and \( p(\text{data} | \theta) \), the likelihood function, specifies the statistical model of data variability given the parameters. These two elements are formally combined in Bayes theorem to give \( p(\theta | \text{data}) \), the posterior distribution of uncertainty for \( \theta \) given the evidence provided by the data.

3.1.1. Example 1 continued. The basic ideas of Bayesian decision-making can be illustrated with reference to a simple decision: that facing the woman in example 1 about whether to take folic acid supplementation. We have already identified the structure of the problem; now we need to consider both the evidence and utilities in more detail.

For women who have already had an NTD pregnancy, our best evidence comes from a randomized controlled trial [18]: 21 out of 602 women randomized to placebo and 6 out of 593 randomized to folic acid supplementation had pregnancies with an NTD. The trial was explicitly designed so that the results could be extrapolated to all women, based on carefully considered scientific reasoning. On this assumption, statistically we can take the odds ratio from the trial, but need to turn to epidemiologic evidence for the baseline rate. Using modelling of birth prevalence of NTD, recorded terminations for NTD and allowing for underreporting of these terminations, Morris and Wald [19] estimate that the incidence in England and Wales was about 3.3 per 1000 in the early 1980s before folic acid supplementation became widespread. There is no formal elicitation of utilities, but it must be reasonable to assume that most couples would prefer a baby to be healthy rather than have an NTD. Views vary considerably over whether a baby with an NTD is better than no baby, as is evidenced by different attitudes to termination for this condition. Side-effects and inconvenience are negligible, and will be considered non-existent here, and costs of supplementation are low.

To formalize the evidence, let \( p_0 \) be the chance of having an NTD pregnancy without taking folic acid supplementation, \( p_1 \) be the chance of having an NTD pregnancy with folic acid supplementation.

The units associated with utilities on life and health are arbitrary, so designate a healthy baby as having utility 1, no baby as 0, and then we can describe the utility of a baby with NTD as \( k \) (say, with \( 0 < k < 1 \)).
A value for \( k \) of 1 would indicate indifference to the existence of an NTD, a value of 0 might be appropriate for a woman who would choose to terminate this pregnancy on those grounds, and values in between express a preference for a healthy baby over one with NTD. We denote costs (whether financial, inconvenience and psychological) of folic acid supplementation by \( c \).

Let \( U_0 \) be the utility associated with the action not to take folic acid, and \( U_1 \) the utility associated with the action of taking folic acid.

In formal terms, assuming, for initial illustrative purposes, \( p_0, p_1 \) to be known we have

\[
U_0 = p_0 \times k + (1-p_0) \times 1 = 1-p_0(1-k) \tag{1}
\]

\[
U_1 = -c + p_1 \times k + (1-p_1) \times 1 = -c + 1-p_1(1-k) \tag{2}
\]

The woman should therefore choose folic acid supplementation if \( U_1 > U_0 \). We note that with a little algebraic rearrangement, this is equivalent to choosing folic acid if

\[
\frac{1-k}{c} > \frac{1}{(p_0 - p_1)} \tag{3}
\]

Now, \( 1/(p_0 - p_1) \) is the number needed to treat, or NNT, a measure of the number of patients who need to take a treatment in order for one patient to benefit relative to the comparator [13] (see Section 2.3). However, ‘NNTs are only one element of decision-making and need to be integrated with patient preferences, and local constraints and conditions’ [20].

By rewriting (3) as

\[
\frac{1-k}{c} > \text{NNT}
\]

formal decision analysis immediately makes clear the very direct relationship between the extra utility of the treatment compared to placebo \((1-k)\) and the cost of treatment \((c)\). Once any two are specified, the third can be imputed. Although in this example we do not consider side-effects, if they are included, the formal decision analysis brings includes the analogous measure often referred to as number needed to harm (NNH) [20].

We see that when \( p_0, p_1 \) are known, decision-making is straightforward for specified \( c \) and \( k \), and that the formal analysis allows sensitivity analysis for \( c \) and \( k \).

We consider a couple with a previous NTD pregnancy. Using direct frequency estimates based on information from the trial, \( p_0 = 0.0349 \) and \( p_1 = 0.0101 \), so that NNT = 40.3. The couple would use folic acid if

\[
\frac{1-k}{c} > 40.3
\]

Values are very difficult to assign, but if a healthy baby were ‘worth’ a million pounds, and the ‘costs’ of folic acid of the order of £10, then on the utility scale \( c \) is approximately \( 10^{-5} \) and the woman should only use folic acid if \( 1-k \) were greater than \( 4 \times 10^{-4} \). This implies that even for women who value an NTD only slightly less than a healthy baby, using folic acid is still the rational choice. This conclusion would be unchanged even if \( c \) were changed by one or two orders of magnitude.

For a woman without a previous history of NTD, the relevant routine data estimate provides a value of \( p_0 = 0.0033 \). The point estimate of the odds ratio from the trial is 0.28. Assuming this applies to a woman with no history of NTD implies that \( p_1 = 0.0009 \). This then gives an NNT of 416.7. Using the same analysis as previously shows that a woman would use folic acid if \( 1-k \)
were greater than $4 \times 10^{-3}$. Broadly this gives the same picture as before, but if folic acid becomes more expensive, perhaps, $10^{-4}$, when translated onto the utility scale, then a woman might only use folic acid if her $k$ were less than 0.96, and if it becomes very expensive, say $10^{-3}$, then she would only choose it if her $k$ were less than 0.6.

A criticism of this analysis is the direct extrapolation from high risk to low risk women, which could be addressed by putting a prior distribution on the ratio of the odds ratios in the two groups. If this is has an expected value of one, the analysis stays unchanged, but if our prior belief about the odds ratio for this group was that it was, say, less than that for a high risk woman, the NNT required for use would increase, and if it were higher the NNT would decrease.

This example has illustrated what is meant by utilities, and how decision-making might change with varying costs. In particular, this can be expressed in terms of the NNT, which is already familiar to many working in evidence-based medicine. We have treated the estimates of $p_0$ and $p_1$ as if they were known. To acknowledge the uncertainty, a full Bayesian analysis is possible. However, in this example, utilities are linear in $p_0$ and $p_1$, so that we can take expectations and replace $p_0$ and $p_1$ by their expected values. Under vague priors, expectations closely approximate the maximum likelihood estimates, leaving the analysis essentially unchanged. The strength of the Bayesian approach comes, in this example, if we do not wish to rely entirely on the data, but wish to assume an informative prior distribution.

3.1.2. Example 2 continued. We now return to the patient with oesophageal cancer, deciding whether to take chemotherapy prior to surgery. He will need to use his own utilities for survival and quality of life. Spiegelhalter and Smith [21] investigated subjects’ attitudes to life expectancy, and found them to be highly non-linear, so a formal analysis would not reduce to replacing point estimates with expected values. Considering the Data Monitoring Committee’s perspective is even more complex. Tan and Smith [22] discuss utility functions that include the patients’ utilities and a term for scientific benefit. They propose the latter might, for example, be quadratic, to reflect knowledge about the true effect. Again expected utilities under uncertainty could only be calculated through a full Bayesian analysis.

Cox [23] and Cox and Hinkley [24] argue that in moving from making inferences to making decisions, it is necessary to include prior information and the losses arising from wrong decisions, and that a Bayes decision rule gives the means of making the optimum decision. We have shown, through our examples, how this applies to making decisions in evidence-based medicine. In line with the review/tutorial approach taken in this article, our examples have been comparatively straightforward. However, we should stress that the full power of Bayesian methodology is seen in more complex situations in (i) incorporating non-vague prior evidence, (ii) providing a coherent logic (marginal probability) for extracting uncertainty about specific parameters of interest from multi-parameter models with many nuisance parameters, (iii) providing a computing technology (simulation, in particular, Markov chain Monte Carlo methods) for calculating and displaying full relevant uncertainties for even very complex models and for calculating expected utilities when non-linear utilities are involved (as in utility for remaining years of life), (iv) communicating and summarizing results in the form of probabilities, which is an essential prerequisite for moving into the decision-making domain. The next section is therefore a guide to a wide range of applications where the references provided make clear the full power and scope of Bayesian statistical methodology.
3.2. Applications to summarizing the evidence

We first review those applications using Bayes theorem where emphasis is on summarizing evidence. When discussing what constitutes evidence, Sackett [1] included studies of diagnostic tests and prognostic markers, and the efficacy and safety of therapeutic, rehabilitative and preventive regimens. We look at these in turn.

3.2.1. Diagnostic tests. Evidence about diagnostic tests involves estimating the sensitivity and specificity. Their performance in practice depends on the prevalence of the condition. This is a straightforward application of Bayes theorem. Murray [25] describes the assessment of stress thallium-201 scintography as an indicator of coronary heart disease.

3.2.2. Single clinical trials. The landmark paper by Spiegelhalter et al. [26] brings together much of the work on Bayesian approaches to randomized controlled trials. The framework used is mathematically simple and is a good place for the novice applied Bayesian to start before moving on to more complex applications.

The paper considers sources for prior beliefs for the parameter for the treatments comparison. They argue for a standard ‘community of priors’. The reference prior assumes minimal prior information. Clinical priors represent the views of well-informed individuals. A sceptical prior is structured with a mean of no difference, and only a small probability of a clinically important difference, and an enthusiastic prior has a mean centred on an important difference, and only a small probability of no difference (or worse). The sources of evidence for informative priors can be from other randomized trials or subjective clinical opinion. The sceptical prior is particularly useful for data monitoring. With the exception of data monitoring, Spiegelhalter et al. [26] explicitly do not consider trials as a decision-making framework in themselves. They therefore do not consider the incorporation of utility assessments.

Others have looked at more complex analysis of clinical trials. Hughes [27] provides methods of analysis using a logistic regression model. For survival data Dellaportas and Smith [28] considers Bayesian analysis of Cox’s proportional hazard’s model, and Abrams et al. [29] model survival data using a Weibull distribution. Although the computation for these is more complex, the essential approach is that of Spiegelhalter et al. [26]. More complex issues such as the analysis of cross-over studies [30], factorial studies [31,32] and subset analysis [33] are all further extensions which lead to useful summaries of various aspects of evidence.

3.2.3. Epidemiological applications. Evidence about prognostic markers for development of disease typically comes from epidemiological studies, and prognostic markers relevant to outcomes following diagnosis come from observational studies of patients, which are methodologically very similar. Typically the analysis from such observational studies are complex, largely because of the number of covariates. Probably for this reason, Bayesian applications in epidemiology had to wait for the recent explosion in computer power, but are now appearing in growing numbers. Breslow [34] cites examples from cancer epidemiology, and Ashby and Hutton [35] reviewed a range of applications covering routine data, case-control and cohort studies. In epidemiology, the focus is on description of patterns and the understanding of inter-relationships, and so unsurprisingly, in general, Bayesian analyses aim at producing posterior distributions of effects of interest.
3.2.4. Systematic reviews and meta-analysis. A systematic review has a complete methodology for finding studies pertinent to a particular question and extracting the relevant data. When appropriate, a meta-analysis is often used as a formal summary of the evidence from the studies. Although current software available, for example, to those involved in the Cochrane Collaboration, typically uses classical methodology, there is growing interest in Bayesian meta-analysis. In essence, meta-analysis uses a hierarchical model, which fits very naturally within a Bayesian framework.

Smith et al. [36] developed a Bayesian random-effects model, incorporating prior beliefs on the likely magnitude of the random effects. A recurrent issue in meta-analysis is heterogeneity between studies. Thompson et al. [37] have developed Bayesian models incorporating study level covariates. Daniels and Hughes [38] have used Bayesian meta-analysis to model the relationship between surrogate markers and clinical outcomes in HIV and AIDS studies. These models develop the ideas for the analysis of single observational studies or randomized trials, and provide a coherent framework for summarizing evidence from all relevant trials.

In summary, there has been an explosion of interest in Bayesian applications in medicine, reflecting a growing appreciation of the factors listed at the end of Section 3.1 and the availability of related software, such as the BUGS system [39]. For a systematic review see Spiegelhalter et al. [40].

3.3. Applications to decision-making

We now return to what we regard as the really fundamental question: that of how to make decisions based on evidence. This section is a guide to a wide range of applications where the references provided make clear the full power and scope of Bayesian decision-making methodology.

The applications described above have essentially been based on a prior-likelihood-posterior analysis, but recurrent themes accompanying discussions of these works are decision-making and incorporating utilities. In his article, ‘Biostatistics and Bayes’, Breslow [34] argues that where scientific data is used for decision-making or regulatory purposes, the introduction of prior beliefs is natural and unavoidable, so Bayesian methods have promise. However, neither he nor his discussants follow this through to argue for a structured approach to the actual decision-making. Rather, this is left to the somewhat anonymous decision-makers. However, in the discussion of Spiegelhalter et al. [26], several discussants argue for a full decision theoretic approach. Spiegelhalter et al. explicitly reject this approach, partly because it is too difficult, but also because an individual trial’s results do not map explicitly to a particular decision-making context. We have argued that identifying the decision-makers in particular contexts is a necessary part of any formal decision-making process.

Arguments for the use of Bayesian methods for decision-making are beginning to appear, although not necessarily with explicit consideration of utilities. Other Bayesian medical applications have been explicitly cast in a decision-making framework. We shall review these applications and arguments in some detail.

One point worth making is that many discussions, which appear only to invoke probabilities and make no specific reference to utilities, may, nevertheless, be full-blooded decision analyses. This happens whenever there are two possible consequences flowing from an action and an implicit or explicit utility structure of so-called 0-1 form; utility 0 for the ‘bad’ consequence, utility 1 for the ‘good’ consequence. A simple calculation reveals that the expected utility of the action reduces to the probability of the good consequence obtaining. Relevant weighing up of expected
utility becomes, in this case, a comparison of probabilities of outcomes – but is, none the less, an application of the decision-making formalism. An example of this is in data-monitoring, discussed below.

3.3.1. Decisions at different levels. Spiegelhalter and Smith [21] discuss the different levels of decision-making necessary within a health system, within an overall government structure: for the individual patient; prioritizing between patients; prioritizing between different areas of medicine; and prioritizing health within the national budget. They focus on the first two as being the domain of clinical decision-making, although they argue in general terms for decision analysis to be applied at all levels.

3.3.2. Decision-making for patients: diagnosis. One of the key skills of the individual clinician is in diagnosis. De Dombal et al. [41] discuss the use of Bayesian analysis to assist diagnosis and decision-making in patients presenting with acute abdominal pain. This combines information on the age and sex of the patient with clinical features to give relative probabilities of common causes of abdominal pain. In trials, its use compares well with experienced clinicians, but perhaps more importantly, helps raise the performance of inexperienced clinicians to that of their senior colleagues. There is no explicit consideration of the utility of different decisions, but in their analysis of de Dombal’s data, Crichton et al. [42] do explicitly incorporate the utilities of correct and incorrect diagnosis.

3.3.3. Decision-making for individual patients: interventions. After diagnosis, the next decision is often how to treat. Many clinical trials collect evidence on this, but the question of how to incorporate such evidence in the decision is less formalized. As we illustrated, for folic acid and NTDs the choice depends on utilities. However, for realistic applications, utilities must be assessed. Spiegelhalter and Smith [21] consider patients making stark choices, such as whether to undergo a risky operation. By considering hypothetical gambles, they elicit utilities for length of life and for various health states.

3.3.4. Decision-making about studies: whether to start. An important decision is whether to start a trial. Lilford and Jackson [43] consider ethics of randomization. They distinguish between the individual clinician, whose first priority is the well-being of the individual patient, and that of the ‘third-party payer’ whose responsibility is more general, and who can therefore restrict access to new technologies whilst evidence of scientific benefit accrues. This in turn enables the individual clinician to randomize. They argue that a Bayesian approach taking account of prior beliefs formalizes this. Berry et al. [44] do explicitly consider a decision-making approach is described for a sequential vaccine trial, where they consider the total patient horizon, both inside and outside the trial. Chaloner and Verdinelli [45] reviewed the general literature on Bayesian experimental design and argued experimental design should be viewed as a decision problem, and solved by maximizing expected utility. This is applicable to clinical trials; see, for example, Parmigiani and Berry [46].

3.3.5. Decision-making about studies: data monitoring. A challenging area of decision-making is data monitoring of clinical trials. Interim analysis of accruing information is necessary in order to monitor for unexpectedly large treatment effects and for excess toxicity. Those responsible for data monitoring must balance the interests of individual patients about to enter the trial with
the scientific value to future patients of the information which will accrue from the study. A simple Bayesian approach to data-monitoring formalizes the problem by using a sceptical prior to represent the position at the inception of the trial and then demanding a high posterior probability of a clinically important difference when the data are analysed using the sceptical prior [26]. Fayers et al. [47] discuss the monitoring of a trial of treatments for oesophageal cancer using this approach. There is no explicit consideration of utilities, but the specification of a ‘clinically worthwhile difference’ and the probability of exceeding this difference that must obtain before the trial is stopped act as surrogate utilities (see our earlier remark about 0-1 implicit utility structures). Qian et al. [48] use a similar approach to reanalyse a trial of non-small-cell lung cancer. Tan and Smith [22] consider utility functions for a trial that explicitly balance scientific interests with the benefits and risks to patients in the trial. Although they develop these ideas in order to make decisions about sample size and allocation of patients to treatments, the ideas are relevant for data monitoring.

A question that taxes data monitoring committees is the extent to which information from other trials and even observational studies should be incorporated formally into analysis for data monitoring. Should such an analysis be desirable, all of the methods described above extend naturally to include other evidence, either through incorporation of such evidence into the prior distribution or by hierarchical models.

3.3.6. Decision-making for pharmaceutical companies. Decision-making within a pharmaceutical company can benefit from Bayesian approaches. Whitehead and Brunier [49] consider the problem of determining dose, using prior beliefs and an explicit gain function. Although in principle this may encompass financial and ethical gains and losses, Whitehead and Brunier restrict their application to a gain in information. Regan and Senn [50] discuss some of the practical aspects involved when a major pharmaceutical company wanted to enhance its decision-making capability, in order to improve productivity.

3.3.7. Public policy decision-making. Lilford and Braunholtz [51] have discussed the statistical basis of public policy, using as an example the increased risk of venous thrombosis with third generation oral contraceptive pills. In a similar vein, Burton et al. [52] consider how to choose between haemodialysis (HD) and continuous ambulatory peritoneal dialysis (CAPD) for end-stage renal disease on the basis of inconclusive evidence.

Parmigiani et al. [53] give a general overview of decision-making in clinical recommendations development, and show how the methodology has been applied to stroke prevention policy, with explicit consideration of the natural history of cerebrovascular disease, models for prevention strategies, patient preferences and costs.

4. THE FUTURE AGENDA

We have shown that evidence-based medicine leads inexorably to consideration of how best to make decisions, and that a full Bayesian analysis gives a coherent framework for this. We have illustrated this with some of the applied work that is now emerging; see Lilford et al. [54] and Matchar et al. [55] for a similar perspective. Given this argument, what are the issues that we, as medical statisticians, need to address?
Following the structure already proposed for deconstructing problems, we need to establish who are the decision-makers, what are the possible actions, what are the uncertain consequences, what are the possible sources of evidence, and what are the kinds of utility assessment required.

The final decision-makers are often individual patients, or in preventive medicine, healthy individuals, but acting within constraints set by other decision-makers such as drug regulators, health authorities and so on. There are important questions about where decision-making ought to lie, and where it actually does. The role of the statistician is to ask questions to clarify the position though wider political issues will certainly emerge from such analysis.

The specification of actions may be straightforward, for example, to take a treatment or not. However, often there will be a complex choice of combinations of different treatment modalities, at differing doses, and for different time periods. In such situations, deconstructing the decisions systematically may be a valuable exercise in sorting out where serious uncertainty still remains. This in turn can inform future research. The role of the statistician is to work with clinical colleagues to define a set of options amenable to analysis that still inform clinical situations.

The specification of consequences is familiar to statisticians. We often start a conversation about a new study by asking ‘What are the primary outcome measures?’ However, deeper consideration of a decision-making context forces us to take seriously outcomes that may seem almost irrelevant from a medical perspective but are very legitimately part of the wider utility context.

The possible sources of evidence are home territory for statisticians, ranging from clinical opinion through routine data, epidemiological studies, randomized trials to systematic reviews. Nothing in this paper dispenses with the need for the most rigorous attention to study design, sources of bias, good measurement technique and all the other aspects of collecting evidence that medical statisticians have worked on. Consideration of the decision-making contexts in which evidence is likely to be used will in turn clarify issues such as the best ways of presenting information from a trial.

Finally the assessment of utilities is an area to which the statistical community has given very little thought. We need to work with, for example, psychologists and health economists to determine utilities of decision-makers. How to determine the trade-offs between various components and perhaps to document how utilities change over time (for example with age) has many statistical aspects.

Evidence-based medicine promises to provide decision-making based on sound evidence, considered in the light of patients’ priorities. Bayesian decision-making provides both an intellectual and practical framework for its achievement.

ACKNOWLEDGEMENT

We thank the referee for thoughtful comments on a previous version of this paper.

REFERENCES

