Breast Cancer Screening

nationality, body weight, diet, age at menarche, age at first pregnancy and several other factors. Even the grim reaper's very own transition, healthy-to-dead, can be influenced by smoking, obesity, environmental and occupational risks, family history and so on.

A truly individual-level model, then, needs to formulate all the various transition intensities and sojourn densities in terms of all known covariates, a truly formidable task. One can quite understand why the paper avoids it. But even more is needed. An individual's utilities or costs will take account of the inconvenience of repeated mammographies, the considerable anxiety of a biopsy following a false positive, and the near-disaster of needlessly receiving the full "curative" treatment. More social costs like increased length of income-generating and tax-paying life following successful treatment, greater demand on the pensions/old age benefits system and even increased costs of other medical treatments in later life, should all be built in as they impinge on an individual both directly and indirectly via the national taxation system. Life is far more complicated than the QALY-per-dollar view would like to make it.

This leads me to a general question. What should be the relation between the "right" utility function to be used at a collective level and the utilities of the individuals that make up that collection? There is probably a vast literature on this (as on so many other things) that I am completely unaware of, but I cannot at present see any objection to the proposition that the collective utility should be simply the sum of the individual ones. I'd appreciate references or thoughts about this.

One final comment on a related theme. We are fed almost daily with statistics about the benefits or otherwise of a whole host of public health issues, all at a collective level. A national programme of screening would save X thousand lives per year, environmental tobacco smoke causes Y thousand deaths per year, for example. As the paper shows, there is an amazing contrast between the apparently-huge collective benefits of breast cancer screening and the barely detectable benefit to an individual, and I suspect the same applies in many other cases. Can we not use our influence as statisticians to have matters clearly stated at individual level? How else can we hope to get a properly risk-aware public?

I thank Parmigiani for a stimulating, if not stimulating, paper and encourage him and the rest of the Berry family to go on with ever-more-practical modelling.

DAVID DRAPER (University of Bath, UK)

This is a fine paper on a methodological topic—applied Bayesian decision theory—which is already important and which will only grow in prominence in the years to come, as attention shifts in Bayesian work away from computational distractions and toward more realistic modeling. (In my view, any research-funding czars out there who may be reading this could do a lot worse than take the following five topics as the core of their high-priority agenda for the next 10 years.)

(i) Elicitation of scientifically defensible informative priors on parameters—we almost always know more than nothing, and it is time to stop being afraid of what non-Bayesians might say if we were to use what we know intelligently (i.e., in a manner driven by sensitivity analysis and external validation);

(ii) Elicitation of decision-defensible utility functions (from single and multiple points of view—more on this below);

(iii) Bayesian nonparametrics and cross-validation (e.g., Walker et al. 1998; Draper 1998), to deal with model uncertainty in ways that avoid cheating by using the data twice;

(iv) Development of user-friendly graphical and numerical environments (a kind of Meta-MCMC, you might say) to make comprehensive sensitivity analysis easier; and

(v) Improvements in computational efficiency, e.g., IID, or nearly-IID, posterior simulation (e.g., Walker 1998), to overcome the grudgingly slow character of some of our present MCMC implementations in complex problems.
The present paper makes progress on at least two of these topics. The author is also to be commended for learning so much of the science of breast cancer and its treatment—in an academic statistical world filled with tenure concerns and my-formula's-longer-than-yours posturing this is highly refreshing (Fisher's biography by his daughter (Box 1978) is worth re-reading on these grounds every five years or so, when you have begun to forget how well gambles like his Rothamsted choice can turn out, both personally and for science).

My main comments about the paper fall into four categories, as follows.

(1) **Bayesian model choice should be made decision-theoretically.** For a number of years prior to the meeting at which this paper was presented, there has been a fair amount of discussion about Bayesian model selection, and the topic continued to receive attention at the meeting. Some of this work has been sensible (e.g., Key et al. 1998), much of it less so (e.g., Atkin 1991; O'Hagan 1995; Berger and Pericchi 1996; Bayarri and Berger 1998). A leading symptom of membership in the latter category is the burgeoning of ad hoc choices that appear to be required when the problem has been mis-formulated. As Key et al. (1998) have noted, one of the nice things about Bayesian is that there is only one legitimate way forward in any given situation: maximization of expected utility based on probabilities updated with Bayes' Theorem. Ad-hockey in an attemptedly Bayesian solution to a problem is nature's way of telling you that you have not thought hard enough about problem formulation. (Yes, you may well have uncertainty about your utility, but there is plenty of room in the framework for that.)

In the case of model selection it would seem self-evident that to choose a model you have to say to what purpose the model will be put, for how else will you know whether your model is good enough? Specifying this purpose demands a decision-theoretic basis for model choice. To take two examples, (1) if you are going to choose which of several ways to behave in the future, then the model has to be good enough to reliably aid you in choosing the best behavior; or (2) if you wish to make a scientific summary of what is known, then—remembering that the hallmark of good science is good prediction—the model has to be good enough to make sufficiently accurate predictions of observable outcomes (in which the dimensions along which accuracy is to be monitored are driven by what is scientifically relevant).

Draper and Fouskas (1998) give one example of this view in action, demonstrating that variable selection in generalized linear models should often be governed by the principle that the final model should only contain variables that predict well enough given how much they cost to collect. The present paper gives another example: Q How should a choice be made among the modeling assumptions in Section 4.3, which support the estimation of the transition probabilities $q_{ij}$? A Choose the approach that leads to the best-calibrated predictions of observables not used in the model fitting process (e.g., the empirical part of Fig. 3). Evidence of this kind was not presented in the paper but must surely have served as the basis of model choices documented here or elsewhere.

(2) **Risk stratification is crucial for patient-level decisions.** The author notes at the beginning of Section 5 that "the additional life-expectancy for screening women in their forties over control is [estimated to be] 5.3 days", a shockingly small number at face value. However, this sort of summary—which is useful for many forms of policy analysis, where aggregate benefits drive the decisions—is not interesting (and may in fact be harmful) to individual women facing choices about how often to get mammograms, because it is not sufficiently conditional on their known risk factors. The average of 5.3 days per woman includes a lot of zeros that are irrelevant to someone at high risk. Summaries stratified by risk factors
such as genetic predisposition (with attendant uncertainty bands, some of which will be large because of sparse data) are crucial to individual choice-making.

(3) Resolving conflicts in multiple-level utility functions. Speaking of the different levels at which information can be useful or useless, a fascinating topic on which the author does not touch—and one that gives people who don’t like decision theory ammunition, on the ground that failure to come to grips with it is fatally unrealistic—is that of possible conflicts between the utility functions of the multiple actors in the health care drama: patient, doctor, hospital, government/society. As an ordinary citizen thinking utterly selfishly, I place a pretty low value on spending my tax money on fixing your health problems (unless I know you, perhaps), but then when I get sick I want the system to spend millions on me if necessary. The hospital I prefer turns out not to have had my direct welfare in mind when it decides to buy an expensive scanner that is of no use to me with my illness, thereby (given cost constraints) failing to buy a different piece of equipment that would materially have helped me. The government has the unenviable task of figuring out how much money should optimally be spent on the monitoring of health care quality, and how best to spend that money (e.g., Draper 1996). And so on: as admirable as the work in this paper is, broader issues raised by these multiple utility functions—such as whether some of the breast cancer screening money would better be spent on prenatal nutrition for low-income mothers—are left unaddressed. Does the author believe that this sort of thing can be resolved in practice with a kind of meta-level utility analysis, or are there new methodological issues here?

(4) Individual values within (not versus) prevention trials. Finally (and on a topic related to multiple utility), the author notes in Section 5 that “It is difficult, and almost never done, to measure individual’s values, or even quality-related health outcomes, in the context of a large prevention trial”. But this is crucial, for the reasons discussed in item (3) above. This creates another potential use of decision theory: how best to allocate resources in a prevention trial (allowing, e.g., numbers of subjects versus richness of outcomes measured to compete) to maximize the relevant information collected.

REPLY TO THE DISCUSSION
The discussants raise important and broad issues that are not addressed adequately in the paper. I am pleased to have the opportunity to second their insightful comments and to add some further remarks. I would like to touch on the role of policy models, their relation to Bayesian decision theory, and their interplay with clinical decision making.

Policy Models. “Policy models are intended to guide the choices of persons and organizations that affect the aggregate allocations of resources to health care problems. Although it is difficult to identify any single policymaker in the United States who can alter the aggregate effect of the millions (or billions) of individual clinical decisions, there are many potential users of policy models: payers, providers, state and local health departments, the National Institutes of Health, professional organizations, hospitals and producers of medical devices, among others.” (Weinstein, 1989). Weinstein’s own seminal work on cardiovascular disease modeling demonstrated that it is useful to build comprehensive decision-support models that can serve as the basis for multiple decision problems and multiple decision makers.

In complex health problems, the fragmentation of scientific investigations, the emphasis on testing and risk-factor identification, and the rarity of thorough reporting of uncertainties, make it very hard to make good decisions by simply examining the scientific literature. Additional modeling is required for informing decision makers; for framing information in a way that can be used for decision analysis (by representing uncertainties using probability); and for developing optimizations and cost-effectiveness calculations based on plausible utilities. Reliable