Article:
Claxton, K. orcid.org/0000-0003-2002-4694, Neumann, P.J., Araki, S. et al. (1 more author) (2001) Bayesian value-of-infomation analysis: an application to a policy model of Alzheimer's disease. International Journal of Technology Assessment in Health Care. pp. 38-55. ISSN: 0266-4623

https://doi.org/10.1017/S0266462301104058

International Journal of Technology Assessment in Health Care, **17:1** (2001), 38–55.

BAYESIAN VALUE-OF-INFORMATION ANALYSIS

An Application to a Policy Model of Alzheimer's Disease

Karl Claxton
University of York

Peter J. Neumann
Sally Araki
Milton C. Weinstein
Harvard Center for Risk Analysis

Abstract

A framework is presented that distinguishes the conceptually separate decisions of which treatment strategy is optimal from the question of whether more information is required to inform this choice in the future. The authors argue that the choice of treatment strategy should be based on expected utility, and the only valid reason to characterize the uncertainty surrounding outcomes of interest is to establish the value of acquiring additional information. A Bayesian decision theoretic approach is demonstrated through a probabilistic analysis of a published policy model of Alzheimer's disease. The expected value of perfect information is estimated for the decision to adopt a new pharmaceutical for the population of patients with Alzheimer's disease in the United States. This provides an upper bound on the value of additional research. The value of information is also estimated for each of the model inputs. This analysis can focus future research by identifying those parameters where more precise estimates would be most valuable and indicating whether an experimental design would be required. We also discuss how this type of analysis can also be used to design experimental research efficiently (identifying optimal sample size and optimal sample allocation) based on the marginal cost and marginal benefit of sample information. Value-of-information analysis can provide a measure of the expected payoff from proposed research, which can be used to set priorities in research and development. It can also inform an efficient regulatory framework for new healthcare technologies: an analysis of the value of information would define when a claim for a new technology should be deemed substantiated and when evidence should be considered competent and reliable when it is not cost-effective to gather any more information.

Bayesian decision theory provides a valuable framework for healthcare technology assessment that distinguishes the conceptually separate decision of whether a new technology should be adopted from the question of whether more research is required to inform this

The writers thank two anonymous referees, Bryan Luce, and Tina Shih for comments on previous drafts. Karl Claxton would also like to acknowledge the support provided by the Harkness Fellowships of the Commonwealth Fund of New York.

choice in the future (5). This type of approach is demonstrated by applying it to a published policy model of Alzheimer's disease (28). A probabilistic analysis of this model is conducted by assigning prior distributions to characterize the uncertainty surrounding model inputs, and the decision uncertainty is represented in the form of a cost-effectiveness acceptability curve. However, Bayesian decision theory suggests that the choice of treatment strategy should be based on expected utility, and the only valid reason to characterize the uncertainty surrounding outcomes of interest is to establish the value of acquiring additional information. The expected value of perfect information is estimated for the decision to adopt a new pharmaceutical for the population of patients with Alzheimer's disease in the United States. This provides an upper bound on the value of additional research. The value of information is also estimated for each of the model inputs so that future research may focus on those parameters where more precise estimates would be most valuable.

The value of conducting additional research to inform particular clinical decision problems and of acquiring more precise estimates of particular inputs used in a cost-effectiveness analysis of healthcare technologies is of general interest. It has implications for the design, conduct, and interpretation of research, as well as the more general policy issue of setting priorities in clinical research and development. These issues are also at the heart of the current international debate about the appropriate regulation of new healthcare technologies. In the United States this debate has been focused in part on the implementation of the Food and Drug Administration Modernization Act (16).

The FDA Modernization Act amended the standard for health economic claims from "substantial evidence [typically demonstrated] by two adequate and well controlled clinical trials" (15;27) to "competent and reliable scientific evidence." What constitutes competent and reliable evidence is not clear in the legislation, but Bayesian decision theory and an analysis of the value of information can be used to asses whether input data for cost-effectiveness analysis are "competent and reliable" and help determine whether an economic claim for a new technology is "sufficiently substantiated."

It cannot be efficient to demand the same standard of evidence in all circumstances and across all technologies irrespective of any evidence already available, the size of the patient population that could benefit from the new technology, and the costs of gathering more information (4). These issues seem to be recognized in the more recent U.S. legislation (16;17), which uses a definition of competent and reliable evidence from the Federal Trade Commission's standards (13):

> . . . a reasonable basis [for a claim of cost-effectiveness] depends . . . on a number of factors relevant to the benefits and costs of substantiating a particular claim. These factors include: the type of product, the consequences of a false claim, the benefits of a truthful claim, the costs of developing substantiation for the claim. . . . (16)

This standard of evidence requires explicit consideration of the marginal benefits and costs of acquiring additional information, but no method for estimating these costs and benefits has been suggested. The approach outlined in this paper provides a framework that can define a claim as "substantiated" and evidence as "competent and reliable" such that it is not efficient to gather any more information.

Bayesian value-of-information analysis is a useful analytic framework for both analysts in designing and conducting research and for policy makers in considering research priorities and the appropriate regulation of new technologies. We demonstrate the benefits and the practicality of this approach by applying it to a published policy model of Alzheimer's disease.

METHODOLOGIC BACKGROUND

Recently a Bayesian decision theoretic framework for the evaluation of healthcare programs has been presented (5;6;8). This analytic approach has a firm grounding in statistical decision theory (30;32;33;38), and has been used in other areas of research including engineering (20) and environmental risk assessment (19;43). The approach suggests that the choice between mutually exclusive programs should be distinguished from the conceptually separate question of whether more information should be acquired to inform this decision in the future. Within this framework the choice between programs should be based on expected utility, and the most important reason to consider the uncertainties surrounding the outcome of interest is to establish the value of acquiring additional information by conducting further research.

Information is valuable because it reduces the expected costs of uncertainty surrounding a clinical decision. The expected costs of uncertainty are determined by the probability that a treatment decision based on existing (prior) information will be wrong and by the consequences if the wrong decision is made (loss function). The expected costs of uncertainty can also be interpreted as the expected value of perfect information (EVPI), since perfect information (an infinite sample) can eliminate the possibility of making the wrong decision. It is also the maximum a decision maker should be willing to pay for additional evidence to inform this decision in the future (6;43). If the EVPI exceeds the expected costs of additional research, then it is potentially cost-effective to acquire more information by conducting additional research (the maximum benefits exceed the costs of further investigation). It is also possible to consider the value of information associated with reducing the uncertainty surrounding each of the parameters in a cost-effectiveness analysis of alternative strategies of patient management. This analysis can focus research priorities by identifying those parameters where more precise estimates would be most valuable and, in some circumstances, indicating which endpoint should be included in further experimental research.

However, observing an EVPI greater than the cost of additional research provides only the necessary but not the sufficient condition for deciding to acquire more experimental information (for example, conducting a clinical trial). It is necessary to estimate the benefits of sampling, or the expected value of sample information (EVSI) for the patient population, and the cost of sample information, including the additional treatment and reporting cost. The difference between the EVSI and sampling cost is the expected net benefits of sampling (ENBS), or the societal pay-off to proposed research. An estimate of the ENBS for every feasible allocation of each sample size is required to identify the optimal allocation of trial entrants (where ENBS reaches a maximum for a given sample size) (5). The optimal sample size for the trial is where ENBS reaches a maximum (given optimal sample allocation). If the maximum ENBS is greater than the fixed costs of the research, then it will be efficient to conduct further research at this technically efficient scale and design. Although estimates of ENBS are not presented here, the application of this type of analysis to Alzheimer's disease is discussed later.

AN APPLICATION OF BAYESIAN VALUE-OF-INFORMATION ANALYSIS

The benefits and practicality of taking a Bayesian decision theoretic approach to the value of information is demonstrated by applying it to a published policy model of Alzheimer's disease (28). A probabilistic analysis of this model demonstrates that value-of-information analysis (VOI) can inform important policy issues such as setting research priorities, establishing technically efficient research design, and informing an efficient regulatory framework.

A Policy Model of Alzheimer's Disease

The purpose of the original (deterministic) model was to evaluate the impact of a new pharmaceutical (donepezil) on the costs and outcomes of mild to moderate Alzheimer's disease (AD). Information about the efficacy of this new drug was provided by the results of a placebo-controlled double-blind clinical trial (35). However, the follow-up period of the trial was only 24 weeks, and economic and quality of life data were not collected. To inform the policy decision of whether to adopt this new technology, an assessment of the costs and health outcomes for a general population of Alzheimer's patients must be made over a longer period. A state transition model (Markov process) was used to characterize the progression of AD through different disease states and care settings (39). States in the model included three disease states (mild, moderate, and severe) defined by scores on the Clinical Demential Rating Scale (CDR) (25); death; and two care settings (community and nursing home). Alzheimer's disease is a chronic and progressive illness, and as patients enter a more severe disease state the probability of moving from community to nursing home care increases. All patients started in either the mild or moderate disease state (here we consider those starting in the mild state). The cycle length in the model was 6 weeks with a time horizon of 6, 12, and 18 months (see Figure 1 for an illustration of the states and possible transitions).

The underlying disease progression (transition probabilities) was derived using data from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD): a longitudinal data base of 1,145 AD patients (26). Transitions to the two care settings in each disease state were also based on the CERAD data. Health state utilities were assigned to each of the seven states using the Health Utilities Index Mark II (HUI:2) and the HUI:2 multi-attribute utility function (29;44). These measures of health-related quality of life were based on a cross-sectional study of 679 AD patient and caregiver pairs (29). Direct medical, nonmedical, and indirect costs were based on a previously published analysis (34). In this probabilistic analysis of the model, costs were based on an analysis of the cross-sectional study of patients and caregivers (21). The effectiveness of the new drug was modeled as relative risk ratios that were estimated using data from the 24-week trial and a Cox proportional hazards regression model. The natural history transition probabilities, between the mild and moderate health states (based on CERAD data), were converted into hazard rates. The estimated relative risk ratios were applied to the natural history hazard rate to provide hazard rates on treatment. These rates were converted back to treatment transition probabilities.

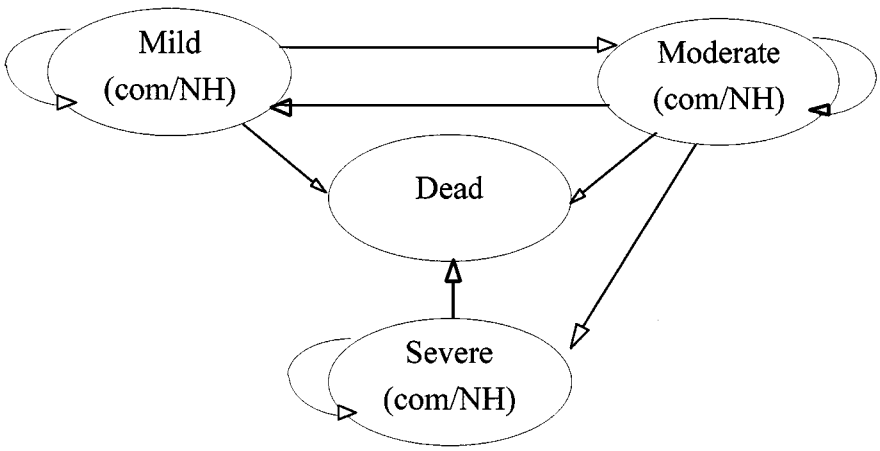

Figure 1. A Markov model of disease progression.

The authors conducted a series of univariate sensitivity analyses on key parameters to account for the considerable uncertainties in the model, including alternative scenarios of the duration of efficacy of the new drug beyond the 24-week trial follow-up. The cost-effectiveness of the new drug was sensitive to duration of drug effect, and the authors emphasize the need for more precise and direct estimates of key parameters in future research. Neumann et al. (28) provides a more detailed discussion of the model, data, and results.

The Policy Issues

This application poses two important and conceptually separate policy questions: a) given existing information should the new drug treatment be adopted; and b) should more information be acquired (by conducting additional research) to inform this choice in the future?

The first policy question is straightforward. That is, when faced with a choice between mutually exclusive strategies of patient management, we should simply choose the strategy with the highest expected utility. Inference and the distribution of expected utility is irrelevant to treatment choice (5). Indeed, we demonstrate below that the traditional rules of inference will impose substantial costs on patients with AD in the United States that can be measured in terms of resources or health improvement forgone.

The distribution of expected utility may be irrelevant to treatment choice, but it is very relevant to the second policy question: should more information be acquired to inform this treatment choice in the future? We can phrase this second policy decision in terms of the FDA Modernization Act: is the economic claim for this new pharmaceutical substantiated and can the evidence be regarded as competent and reliable? The framework presented below can help answer these general questions. We can also address a number of specific questions in the AD example, including:

- Is additional research in AD potentially cost-effective?
- Are the estimates of the AD model inputs adequate?
- For which model inputs would more precise estimates be most valuable?
- Is experimental design required for subsequent research?
- If so, which endpoints should be included in any future clinical trial?
- What is the optimal follow-up period?
- What is the optimal sample size?
- How should trial entrants be allocated between the arms of the trial?
- What is the value of this proposed research?

In short, the key questions are: what is a technically efficient research design and how should research and development resources be allocated?

Developing a Probabilistic Model

Before the value of additional information can be established, the existing (prior) information must be identified, characterized, and incorporated in the model. Characterizing existing information and the current uncertainties surrounding this decision problem can be done by assigning prior distributions to each of the model inputs. These distributions are of prior means and represent the current second order uncertainty (i.e., the distribution of the mean) surrounding the estimates of each of these inputs (3;42;44). Once existing or prior information has been characterized in this way, these distributions can be propagated through the model using Monte Carlo simulation (3;11;42;44). The output of this simulation

provides a distribution of the prior incremental net benefit of the new drug, which can be expressed in either health outcome (η) or monetary terms (μ) (5;6;41):

$$\mu = \lambda \cdot (U_2 - U_1) - (C_2 - C_1) \quad \text{or} \quad (\lambda.U_2 - C_2) - (\lambda \cdot U_1 - C_1) \tag{1a}$$

$$\eta = (U_2 - U_1) - 1/\lambda \cdot (C_2 - C_1) \quad \text{or} \quad (U_2 - 1/\lambda \cdot C_2) - (U_1 - 1/\lambda \cdot C_1) \tag{1b}$$

where U_2 and C_2 are the health outcome and costs associated with the new drug treatment, respectively, and U_1 and C_1 are the outcomes and costs associated with current practice (in this case, no treatment). Consistent with the original model we take a societal perspective, so the monetary valuation of health outcome (λ) can be interpreted as the marginal societal willingness to pay for an improvement in health outcome (the budget constraint is endogenous) (18;31). The value of λ is then the normative choice of a social decision maker, and although an analyst may not know with certainty which value of λ will be selected, it will not be uncertain to the decision maker at the time the choice is made. Therefore, it is not unreasonable to regard λ as a constant and conduct analysis conditional on a range of values of λ. It is then the task of societal decision makers to make a normative choice of which value of λ is acceptable. In this analysis we use $50,000 per quality-adjusted life-year as a central value but conduct analysis on values ranging from $1,000 to $100,000.

Characterizing Prior Information

The characterization of prior information surrounding model inputs in the probabilistic model is summarized in Table 1. The baseline transition probabilities were characterized as beta distributions, which seems appropriate for two reasons: the beta distribution takes values between 0 and 1, and its parameters (α, β) represent the number of "successes" and "failures" that were directly available from the CERAD database (1;3).[1] The prior mean health state utility for each state in the model was characterized as normally distributed with standard deviations based on standard errors from the data in the cros-sectional study (29).[2] Data for direct and indirect costs for each health state were reported only as means and standard errors. We characterized their prior distribution as lognormal (with standard deviations equal to the reported standard errors) because it has some useful characteristics for modeling costs data (it cannot take values less than zero and it is positively skewed) (11). For similar reasons we characterized the relative risk ratios applied to the mild to moderate and moderate to mild transition as lognormal based on the reported mean and confidence intervals from the Cox proportional hazards regression (28). Dropout or discontinuation

Table 1. Characterizing Prior Information

Parameter	Prior distribution	Source
Baseline transition probabilities	Beta	CERAD longitudinal data (n = 1,745, 1,320)
Health state utilities (HUI.2, n = 191, 55)	Normal	Cross-sectional study (HUI.2, n = 191, 55)
Direct costs	Lognormal	Cross-sectional study (n = 191, 10)
Indirect costs	Lognormal	Cross-sectional study (n = 191, 10)
Relative risk ratios Mild to moderate Moderate to mild	Lognormal (0.5, SD = 0.188) (2.65, SD = 1.56)	24-week double-blind placebo-controlled clinical trial
Efficacy duration	Lognormal (78 weeks, SD = 47.3)	Panel of clinical experts (n = 13, assumed not independent)
Dropout rate	Lognormal (0.04, SD = 0.0128)	Open label follow-up

rates were characterized as lognormal distributions with mean and standard deviations based on the estimates used in the original model and limited evidence from an open label follow-up study (36). Additional utilization and prices associated with the new drug treatment (conditional on disease state) are regarded as constants because any differences in utilization and price across decision makers/settings represent variability rather than second order uncertainty (see Thompson and Graham [44] for a discussion of this distinction and the dangers of conflating the two).

The duration of drug effect was found to be the most important uncertainty by Neumann et al. (28), but clearly some assessment of effect beyond the 24-week follow-up is required when making the policy decision of whether to adopt the new treatment. The evidence from an open label follow-up study suggested that there may be a longer term effect, but it was of limited use due to the absence of a control group (36). Neumann et al. (28) conducted a survey of clinical experts and elicited judgments about the expected efficacy duration from 13 respondents. A lognormal distribution with mean duration of 78 weeks and standard deviation of 47.3 weeks were fitted to these data.[3] This approach assumes that responses were not independent: all responses were observations from the same (common) prior distribution. This is the most conservative interpretation of the information provided by these judgments because it implies that the second and subsequent responses do not provide additional information. The alternative view would be to assume that the judgments were independent, in which case they would be exchangeable with sample information (the variance of the prior would be much smaller). This interpretation was rejected as far too optimistic.

The Results of Probabilistic Analysis

Monte Carlo simulation (10,000 iterations) was used to propagate these prior distributions through the model. The prior distribution of incremental net benefit is illustrated in Figure 2A for time horizons of 24 weeks and 210 weeks for patients starting in the mild/community state and using $\lambda = \$50{,}000$. As the time horizon is extended, the expected incremental net benefit increases and the new treatment becomes cost-effective ($\mu_0 > 0$) when costs and outcomes are considered beyond 54 weeks. However, the uncertainty surrounding

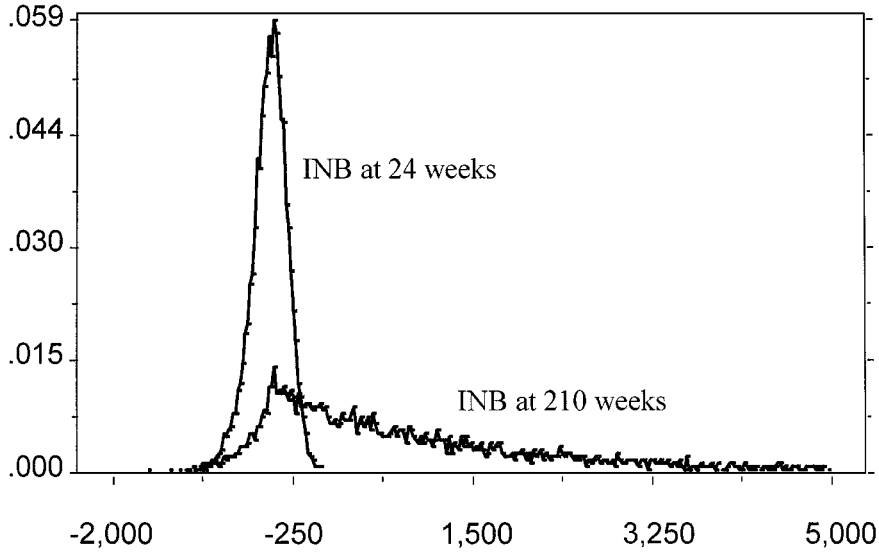

Figure 2A. Prior distribution of incremental net benefit.

these estimates of net benefit is also increasing as the model extrapolated beyond the trial period. So although the mean incremental net benefit is positive at 210 weeks ($1,220), the uncertainty surrounding this estimate is substantial. It is also clear from Figure 2A that evaluating this new treatment at 24 weeks would seriously underestimate the expected health outcomes and overestimate costs (and uncertainty) in this chronic disease because the benefits of an effective treatment can be expected to accumulate over a much longer period of time.

The uncertainty surrounding the cost-effectiveness of donepezil at 4 years for a range of values of λ can be represented as an acceptability curve in Figure 2B (2). The probability that donepezil is cost-effective ($p[\mu_0 \geq 0]$) is .6796 ($\mu_0 = \$1{,}220$) when $\lambda = \$50{,}000$ in Figure 2B. However, the uncertainty surrounding μ_0 is substantial (standard deviation (σ_0) of $2,168) and the error probability ($\alpha = 1 - p[\mu_0 \geq 0]$) that the new treatment is not cost-effective is .3204, which is greater than the conventional benchmarks of .05 or .025 used in both Bayesian inference (40) and traditional frequentist statistics. Equivalently μ_0 is within a Bayesian range of equivalence (40), or the lower 95% confidence limit includes zero in a frequentist framework.

According to the rules of inference (whether Bayesian or frequentist), the apparent cost-effectiveness of the new drug treatment is not statistically significant, we cannot reject the null hypothesis, and the result is indeterminate. In these circumstances the new treatment will be rejected in favor of current practice. However, these rules lead to the rejection of the alternative with the highest probability of being optimal: the probability that the new treatment will provide greater net benefits than current practice (based on the information currently available) is $1 - \alpha = .6798$ in Figure 2B. Failure to adopt the new treatment simply because the difference in net benefit is not regarded as statistically significant will impose unnecessary costs. For an individual AD patient, these costs can be valued at $1,220 (the additional net benefit forgone) or 0.0244 quality-adjusted life-years (QALYs) forgone. The costs imposed on the U.S. population of current and future AD patients over the effective lifetime of this new technology can be valued at $1,064 million or 21,279 QALYs forgone (estimates of the effective U.S. AD patient population are discussed later).

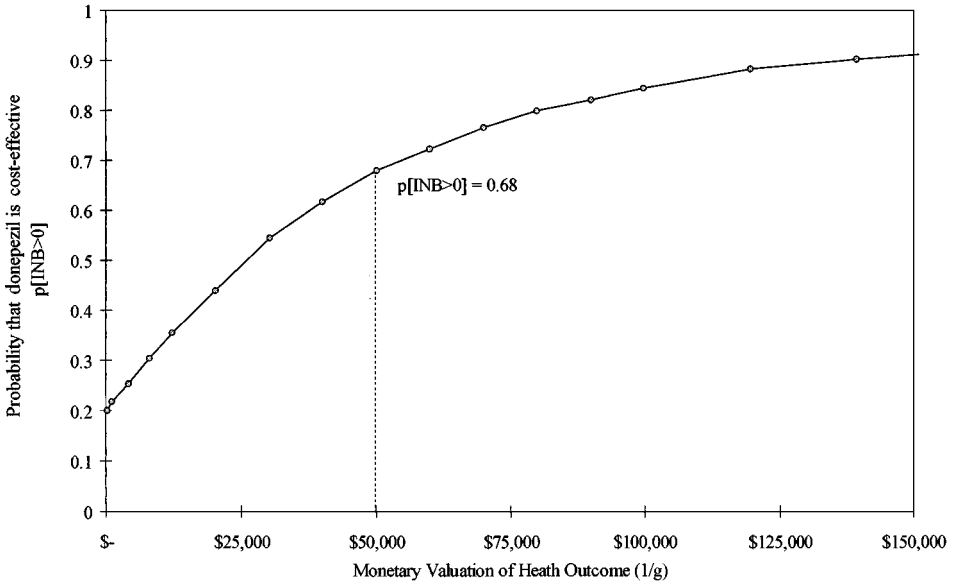

Figure 2B. Acceptability curve for donepezil.

If the societal objective is to maximize health gain subject to a budget constraint, then (in the absence of substantial sunk costs or irreversibilities)[4] alternative strategies of patient management should be selected based on expected net benefit, irrespective of whether any differences are regarded as statistically significant or fall outside a Bayesian range of equivalence.[5] This is because one of these two mutually exclusive alternatives must be chosen and this decision simply cannot be deferred. The opportunity costs of failing to make the correct decision based on expected net benefit are symmetrical, and the historical accident that dictates which of the alternatives is regarded as current practice is irrelevant. The measure of net benefit used in this analysis is based on a particular objective (or social welfare function) that may be judged inappropriate. If the decision maker has other legitimate concerns (for example, equity issues, a concern for rare but catastrophic events, or preferences toward risk), then the measure of net benefit may be regarded as incomplete. However, these arguments (which imply a different social welfare function) could be incorporated in the analysis by amending the measure of outcome, and they do not change the fundamental point that inference is irrelevant to treatment choice (5).

THE DECISION TO ACQUIRE MORE INFORMATION

Although the distribution of net benefit may not be relevant to the choice between treatment strategies for AD, it is relevant to the decision of whether to collect more information to inform treatment choice now and in the future. This decision theoretic approach distinguishes the simultaneous but conceptually separate steps of deciding which treatment should be adopted, given existing (prior) information, from the question of whether more information should be acquired.

Information is valuable because it reduces the expected costs of uncertainty surrounding clinical decisions. The expected cost of this uncertainty will be determined by the probability that a decision based on expected net benefit will be wrong (Figure 2A) and the size of the opportunity loss if the wrong decision is made (these costs include resource savings and health outcome forgone). The expected cost of uncertainty surrounding the treatment decision when it is based on existing (prior) information can also be interpreted as the expected value of perfect information (EVPI), since perfect information (an infinite sample) would eliminate the possibility of making the wrong decision (20;32;33;36;38).[6]

$$\text{EVPI} = \lambda \cdot \sigma_0 \cdot \text{L}(\text{D}_0), \quad \text{where} \quad \text{D}_0 = \frac{|\eta_0 - \eta_\text{b}|}{\sigma_0} \tag{2}$$

$\text{L}(\text{D}_0)$ = unit normal loss integral for standardized distance D_0, η_0 = prior mean incremental net benefit of the new treatment (in health outcome), and η_b = point of indifference between the two alternatives ($\eta_\text{b} = 0$).

The probability that a treatment decision based on prior expected net benefits will be wrong is determined by the distance η_0 from η_b and the uncertainty surrounding η_0, which is measured by the prior standard deviation (σ_0) and represents the amount and quality of prior information available. The opportunity losses if the wrong decision is made is simply the difference in net benefit between what would have been the optimal treatment choice and the choice actually made based on prior information ($|\eta_0 - \eta_\text{b}|$). These losses can be expressed in money terms as $\lambda \cdot |\eta_0 - \eta_\text{b}|$, so the slope of the loss function is simply λ or the monetary value placed on opportunity losses when they occur.

EVPI for the Choice Between Strategies

The EVPI in equation 2 is the maximum value that can be placed on acquiring additional information to inform treatment choice for an individual AD patient. Figure 3A illustrates

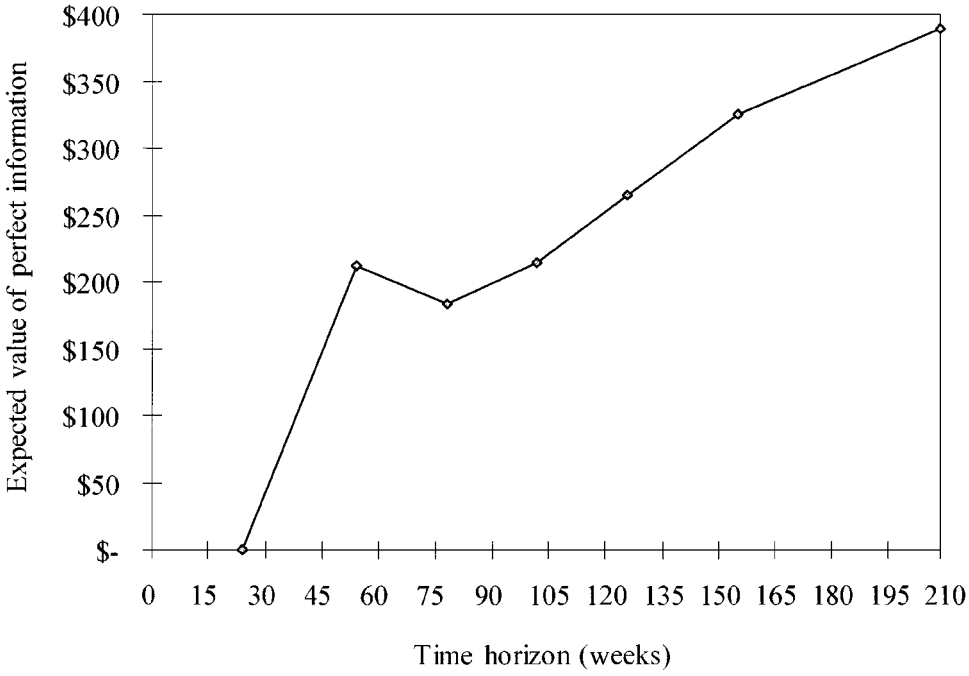

Figure 3A. EVPI for an individual patient.

the relationship between the EVPI and the time horizon when $\lambda = \$50{,}000$. Evaluating this problem at a time horizon of only 24 weeks will seriously underestimate the EVPI (as well as η_0) because the prior decision is to choose current practice ($\eta_0 < 0$) and the uncertainty is relatively small. As the time horizon is extended, η_0 increases and becomes positive after 54 weeks, and the uncertainty surrounding η_0 increases. The probability of making the wrong decision (and the EVPI) unambiguously increases as the time horizon is extended from 24 to 54 weeks because σ_0 increases and $|\eta_0 - \eta_b|$ falls. Beyond 54 weeks σ_0 increases but $|\eta_0 - \eta_b|$ also increases, and the impact on EVPI is ambiguous.

The information generated by research is nonrival and has public good characteristics (10;37). Once it is produced, it can be used to inform the treatment decisions for all eligible patients at no additional cost. The EVPI for the population of current and future AD patients over the effective lifetime of this new technology (T) can be established based on estimates of the incidence of AD patients (I) in each period (t) discounted at rate r.

$$\text{Population EVPI} = \text{EVPI} \cdot \sum_{t=1}^{T} \frac{I_t}{(1+r)^t} \tag{3}$$

The discounted effective U.S. population of AD patients was estimated to be 872,087. This estimate was based on reported incidence rates of AD by age and gender, U.S. census projections by age and gender, and an effective lifetime of the new technology ranging from 2 to 8 years.[7] It is worth noting that the effective population (and therefore EVPI) will be finite if either T is finite and/or $r > 0$.

The EVPI for the U.S. population of AD patients over a range of values of λ is illustrated in Figure 3B. At a time horizon of 210 weeks and $\lambda = \$50{,}000$, the EVPI is $339 million. This represents the maximum value of acquiring additional information and suggests that proposed research will be potentially cost-effective (the fixed costs of research are likely to

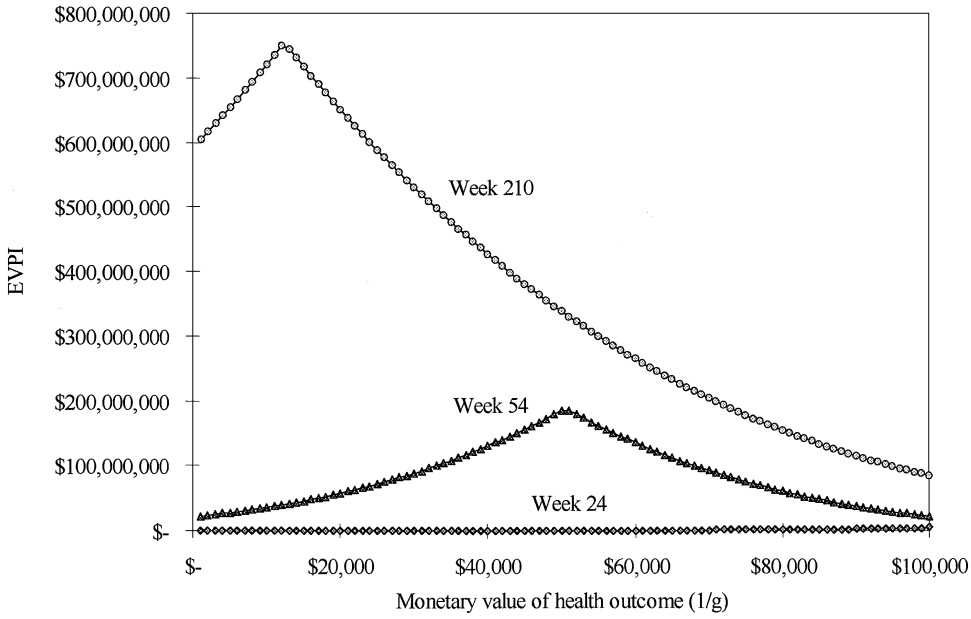

Figure 3B. EVPI for treatment choice (US population).

be less than the EVPI). The value of information is closely related to the monetary valuation of health outcome. At 210 weeks, as λ increases the EVPI falls (when $\lambda > \$12{,}000$). This is because $\eta_0 > 0$, so as λ increases $|\eta_0 - \eta_b|$ also increases, and the probability of making the wrong decision falls. However, the value placed on opportunity losses when they occur is increasing. In this case the former offsets the latter[8] (EVPI falls). At a time horizon of 54 weeks, the EVPI reaches a maximum at $\lambda = \$51{,}000$, where $\eta_0 = \eta_b$. At this point the decision maker is indifferent between the treatment alternatives, and the probability that a decision based on η_0 will be wrong reaches a maximum. At a time horizon of only 24 weeks, the new treatment is not cost-effective even at $\lambda = \$100{,}000$. The EVPI is underestimated and only rises to $4.6 million at $\lambda = \$100{,}000$.

The results in Figure 3B demonstrate that the explicit monetary valuation of health outcome is an essential and unavoidable issue in the decision to acquire additional information to inform treatment choice. It cannot be avoided because any decisions that are made about further research implicity assign such a value. The results also demonstrate that evaluating the treatment at 24 weeks will underestimate the value of additional information as well as underestimating health outcome and overestimating costs (also see Figure 2A).

The population EVPI for the choice between these strategies measures the maximum possible payoff to additional research. It provides a first hurdle for proposed research (6), or in the terms of FDAMA, further research will not be efficient. A claim can be regarded as "substantiated" and evidence as "competent and reliable" if the EVPI is less than the fixed costs of additional research. Alternatively, further research is potentially cost-effective and a claim may not be substantiated if the EVPI exceeds the fixed costs of research. In this example, the EVPI does exceed the fixed cost of further research and the economic claim for the pharmaceutical may not be regarded as substantiated. However, observing EVPI greater than the fixed costs of research is only a necessary but not sufficient condition for demanding more information. The sufficient condition requires estimates of the marginal benefits and marginal costs of additional information (see Discussion).

EVPI for Model Inputs

The analysis described above established the EVPI for the clinical decision problem as a whole; however, it is also useful to consider the value of information associated with each of the uncertain parameters in the model. This type of analysis is the VOI equivalent to conditional probabilistic analysis (3;11;14) where the Monte Carlo simulation is run holding the parameter of interest constant at its expected value.[9] For example, in the AD model the EVPI associated with efficacy duration ($EVPI_{ED}$) is the difference between the EVPI when all parameters are allowed to vary and the EVPI conditional on efficacy duration taking its expected value ($EVPI|_{ED=78\,weeks}$).

$$EVPI_{ED} = EVPI - EVPI|_{ED=78\,weeks} \quad \textbf{(4)}$$

This analysis can help to focus research priorities in AD by identifying those model inputs where more precise estimates would be most valuable (where EVPI is high). In some circumstances this can indicate whether an experimental or an observational study may be required. Those parameters that are vulnerable to selection bias, such as measures of efficacy and the duration of efficacy, will require experimental design. However, other inputs, such as health state costs and health state utilities, may not be so vulnerable (particularly if they are conditional on clinical events within the model), and a clinical trial may not be required. This type of analysis can also start to address the questions of whether an experimental design will require a longer follow-up than previous clinical trials and which endpoints may be worth considering.

The EVPI associated with the AD model inputs at 210 weeks ($\lambda = \$50{,}000$) are illustrated in Figure 4. These results suggest that longer follow-up may be worthwhile because the EVPI associated with efficacy duration (ED) and with the relative risk ratio (RRR) beyond 24 weeks, conditional on efficacy being durable (RRR > 24), are substantial ($270 million and $93 million, respectively). However, a more precise estimate of efficacy within the

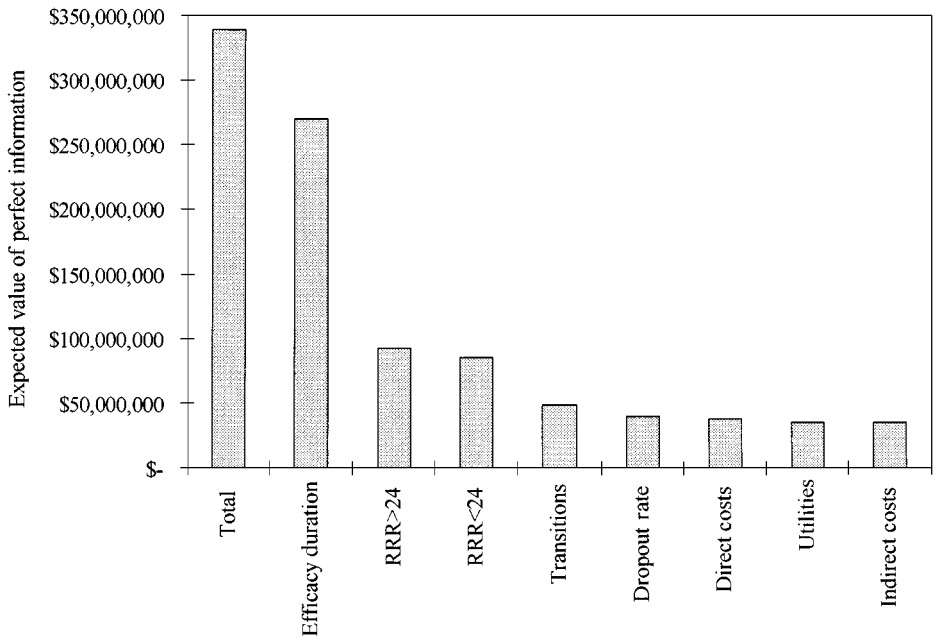

Figure 4. EVPI for model inputs (210 weeks).

existing trial period (RRR < 24) is still valuable ($84 million). Each of these inputs, as well as the dropout rate ($39 million), are particularly vulnerable to selection bias so either experimental design, or possibly econometric solutions to selection bias in observational data (23;24), would be required.

The EVPIs associated with the other groups of inputs[10] can be obtained by holding each of the inputs constant at their expected value. These include baseline transitions, direct costs, utilities, and indirect costs. The EVPIs for these groups of inputs are much lower ($49 million, $38 million, $36 million, and $35 million, respectively). However, at the margin this does not necessarily mean that additional research about these parameters should take a lower priority, because the costs of including these endpoint in further experimental research, both in terms of resources and the health gains forgone of those patients enrolled and awaiting the results of the research, may also be lower. Similarly, the costs of obtaining additional information about efficacy and duration may be high, so that high EVPI may not necessarily indicate high priority at the margin.

Despite this, these estimates do give some indication of the value of including economic as well as clinical endpoints in a proposed trial and the value of different design options. For example, if the fixed costs of including health state utilities in a trial exceed the EVPI, then it will not be efficient to include them as endpoints. It is also important to note that the value of information for a group of model inputs is not the simple addition of the value of information of each separately since this excludes the joint effects within the model. The EVPI associated with costs, utilities, and baseline transitions jointly was established by running the simulation while holding them all constant at their expected values. The EVPI or the maximum value of including these other endpoints was $64.4 million, which is substantially less than the sum of the EVPI for each of these parameters.

Although observing an EVPI for a model input that is greater than the fixed costs of research is only a necessary condition for demanding more information, it does start to address the question: are the model inputs adequate? In the terms of FDAMA, the evidence for a model input can be regarded as competent and reliable (or adequate) if the EVPI is less than the fixed costs of additional research. However, if the EVPI is greater than the fixed costs of research, then the evidence may not be regarded as competent and reliable (or inadequate) once the marginal benefits and costs of acquiring a more precise estimate have been established.

DISCUSSION

The framework presented above distinguishes the question of which strategy of patient management should be chosen, given existing information, from the conceptually separate decision of whether more information should be acquired to inform this choice in the future. The rules of classical statistical inference and their Bayesian counterparts (ranges of equivalence) appear to be inconsistent with the objectives of a coherent healthcare system, impose unnecessary costs, and could be rejected in favor of maximizing expected utility (net benefit) and establishing the value of information. Others have made similar arguments:

> It must be recognised that clinical trials are not there for inference but to reach a decision, and the omission of their raison d'etre is serious. In the long term utility is realistic and indeed necessary. . . . It is only by using expected utility that we can be sure that our actions fit together sensibly. I suspect that the procedure of continuing with a trial until a tail area probability in the posterior is small is just as incoherent as a belief based on the tail area p-value. Or if it is coherent, it implies an inept utility, such as one taking only values of 0 and 1 (22).

The possibility of abandoning inference (based on either a frequentist or Bayesian view of probability) and taking a decision theoretic approach has been discussed for some time but has often been rejected because:

> . . .the consequences of any particular course of action are so uncertain that they make the meaningful specification of utilities rather speculative. . . . (40,360)

It is certainly true that, when attempting to fully characterize the uncertainties in a decision problem when there is a lack (or absence) of good quality data for key clinical events, speculation and judgment are inevitable. There are a number of possible responses to this situation: ignore the events for which evidence of an acceptable quality is unavailable (in which case the analysis will be partial and biased); only analyze decision problems where complete and good quality evidence has already been produced (in which case research will focus on relatively simple questions where we already have solutions); or address complex and uncertain problems in an explicit way, using the best evidence when it is available, but accept speculation and judgment when it is not. The choice is not between speculation or evidence but between methods that expose the lack of evidence and make judgments and speculation explicit or those that leave the judgments and speculation for individuals to make implicity and possibly inconsistently. Making judgment and speculation explicit has a number of advantages, because only then are the key uncertainties exposed to debate, alternative formulation, and an analysis that can indicate where more evidence should be acquired through further research.

For example, in the AD model there is no experimental evidence for efficacy duration beyond 24 weeks, and clinical judgments were used to form a diffuse prior distribution for this key parameter. An alternative is to assume (as the FDA implicity does for purposes of allowing promotional claims) that the drug becomes totally ineffective after 24 weeks. This assumption is no more grounded in experimental evidence, and many clinical experts would argue that this would not be credible (we have shown previously that it will underestimate outcomes and overestimate costs). We extrapolated from the trial using an explicit model of the disease process, combining informed speculation about possible efficacy duration. The considerable uncertainty, due to the poor quality of evidence on efficacy duration, was incorporated in the analysis by assuming that the clinical judgments were not independent (which generated a diffuse prior). The results showed that the value of acquiring further information about this uncertain but key parameter is substantial ($EVPI_{ED} = \$270$ million) and that additional experimental research is potentially cost-effective.

The same type of analytical framework can be used to establish the expected benefits and the costs of sample information. It is then possible to answer questions such as whether an additional clinical trial required before an economic claim for donepezil can be substantiated. If so, should an economic evaluation be conducted alongside the new trial, and what is the optimal follow-up, sample size, and patient allocation?

The societal payoff to proposed research, or the expected net benefits of sampling (ENBS), is simply the difference between the expected benefits and expected cost of sampling (5;6). If the ENBS > 0 for any sample size, then further experimental research will be efficient and an economic claim for donepezil cannot be regarded as substantiated until additional research has been conducted.[11] The optimal sample size for this proposed research will be where ENBS reaches a maximum, given that the trial entrants are allocated optimally between the two arms of the trial (5;7).

This framework can also be used to decide which endpoints should be included in a proposed trial. For example, excluding direct and indirect costs and heath state utilities will mean that even very large samples cannot resolve all the uncertainty surrounding η, and the expected benefits of sampling will be reduced. However, the fixed and marginal reporting

costs of sampling will also be reduced, and economic evaluation alongside a trial will be efficient if the former offsets the latter. In some circumstances large and simple clinical trials may well be efficient, but for other clinical decision problems, trials with economic endpoints will be required (9;12).

A similar approach can also be used to establish the optimal follow-up for any future trial. For example, a proposed trial with a 36-week follow-up cannot fully resolve the uncertainties about effectiveness and cost beyond 36 weeks (even when sample size is very large), so the expected benefits of sampling will always be lower than with a longer follow-up period. However, the fixed and marginal costs of research will also be lower, and the societal payoff may increase or fall. The technically efficient follow-up for a trial will be where the ENBS reaches a maximum over a range of possible follow-up periods.

This discussion suggests that technically efficient design can be established based on estimates of ENBS for each combination of alternative endpoints, follow-up periods, sample size, and sample allocation.

CONCLUSIONS

Once a Bayesian decision theoretic approach is adopted, the two conceptually separate policy questions (whether the new drug treatment should be adopted and whether more information should be acquired) can be addressed. We can phrase this second policy question in terms of the FDA Modernization Act: is the economic claim for this new pharmaceutical substantiated and can the evidence be regarded as competent and reliable? The analysis presented above shows that this framework can inform these general policy issues and address a number of specific questions in AD research, including:

- Is additional research in AD potentially cost-effective?
- Are the estimates of the AD model inputs adequate?
- For which model inputs would more precise estimates be most valuable?
- Is experimental design required for subsequent research?
- If so, which endpoints should be included in any future clinical trial?
- What is the optimal follow-up period?
- What is the optimal sample size?
- How should trial entrants be allocated between the arms of the trial?
- What is the value of this proposed research?

In short, this approach can establish technically efficient research design and provide a societal value of proposed research that can be used to allocate research and development resources efficiently. This type of analysis also informs an efficient regulatory framework for new healthcare technologies: an analysis of the value of information defines a claim for a new technology as substantiated and evidence as competent and reliable when it is not efficient to gather any more information.

NOTES

[1] Multiple transitions were restructured into a number of conditional probabilities so that each set of transitions is a series of binary events. This ensures that probabilities less than zero or greater than one cannot occur during simulation.

[2] Although we did have access to the individual observations, which were not normally distributed, here we are only concerned with the second order uncertainty (the distribution of the mean), so it is not unreasonable to use the normal distribution when sample size ranges from 55 to 191.

[3] Conditional on efficacy being durable, the relative risk ratios for the 24-week trial were applied to the baseline transition probabilities using a prior distribution based on the standard errors from the Cox regression. If efficacy is not durable, then baseline transitions are used. This is only one way to incorporate efficacy duration, and the uncertainties introduced as the results of the trial are extrapolated.

[4] If the adoption decision will result in sunk costs or irreversibilities, then maximizing expected net benefit can be amended. Either the cost that will be sunk can be compared to the benefits of adoption (5) or option prices can be used to adjust the estimates of net benefit. However, these issues do not lead back to traditional rules of inference.

[5] There may be other arguments in a societal decision maker's utility function, such as equity, access, and concern for catastrophic events. The appropriate response would be to incorporate these arguments in a measure of net benefit rather than use tradition rules of inference.

[6] This parametric approach requires the prior net benefit to be normally distributed (32;33). Although EVPI can be established using a nonparametric approach directly from the Monte Carlo simulation (14;43;44), we take the parametric approach for ease of exposition and so that the marginal benefits of sampling can be more easily considered in the Discussion.

[7] Where the probability of obsolescence rises from 2 to 8 years, the effective lifetime of the new technology is uncertain, but it is only expectation of T that is relevant because it is only the expected value of information (not its distribution) that is important for policy decisions.

[8] This need not be the case, and EVPI may rise or fall with λ, depending on the strength of prior information.

[9] We should only be concerned with the *expected* value of information because a societal decision maker should be risk-neutral with respect to the payoff from additional research. Therefore, if the relationship between a parameter is not markedly nonlinear, it may be reasonable to conduct analysis conditional on the expected value of the parameter.

[10] We could also establish VOI for the costs, utilities, and baseline transition probabilities of each heath state.

[11] There is tension between the societal benefits of research (expressed as ENBS) and ethical concerns for those enrolled in the trial. We do not attempt to incorporate important and legitimate concerns for those enrolled, which is the responsibility of ethics and data monitoring committees. However, the ENBS does provide an estimate of the opportunity costs to society (collective ethics) of stopping a trial or failing to approve a trial on individual ethical grounds. Establishing the opportunity cost of holding ethical concerns can help to achieve some consistency in the inevitable trade-off between collective and individual ethics.

REFERENCES

1. Berry DA, Stangl DK. Bayesian methods in health related research. In DA Berry, DK Stangl (eds.). *Bayesian biostatistics*. New York:Marcel Dekker Inc; 1996.
2. Briggs AH. Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ*. 1999;8: 253-261.
3. Chrichfield GC, Willard KE. Probabilistic analysis of decision trees using Monte Carlo simulation. *Med Decis Making*. 1986;6:85-92.
4. Claxton K. Bayesian approaches to the value of information: Implications for the regulation of new pharmaceuticals. *Health Econ*. 1999;8:269-274.
5. Claxton K. The irrelevance of inference: A decision making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999;18:341-364.
6. Claxton K, Posnett J. An economic approach to clinical trial design and research priority setting. *Health Econ*. 1996;5:513-524.
7. Claxton K, Thompson K. *A dynamic programming approach to efficient clinical trial design. Centre for Health Economics Technical Paper No. 13*. 1999.
8. Claxton K, Walker S, Lacey L. A utility based approach to decision making with multiple disease state process. *J Royal Stat Soc A*. In press.
9. Collins R, Peto R, Gray R, Parish S. Large scale randomised evidence: Trials and overviews. In: D Weatherall, JG Ledingham, DA Warrell (eds.). *Oxford textbook of medicine, vol. 1*. Oxford: Oxford University Press; 1996.

10. Culyer AJ. *Economics and public policy: NHS research and development as a public good. Centre for Health Economics Discussion Paper*. 1999, 163.
11. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. *Med Decis Making*. 1985;5:157-177.
12. Drummond MF, Davies L. Economic analysis alongside clinical trials: Revisiting the methodological issues. *Int J Technol Assess Health Care*. 1991;7:561-573.
13. Federal Trade Commission. Policy statement on advertising substantiation. *49 Federal Register* 30999 (1984).
14. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Making*. 1998;18:95-109.
15. Food and Drug Administration Division of Drug Marketing, Advertising, and Communications. Principles for the review of pharmacoeconomic promotion, 1995, draft guidelines.
16. Food and Drug Administration Modernization and Accountability Act of 1997. 1997, Section 114.
17. Food and Drug Administration Prescription Drug User Fee Act of 1997. *Bureau of National Affairs Health Policy Report*. 1997;5:1443-1444.
18. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *J Health Econ*. 1997;16:1-31.
19. Hammitt JK, Cave JAK. *Research planning for food safety: A value of information approach*. RAND Publication Series, 1991.
20. Howard RA. Information value theory. *IEEE transactions on systems science and cybernetics*. 1966;SSC-2:122-126.
21. Leon J, Cheng CK, Neumann, PJ. Alzheimer's disease care: Costs and potential savings. *Health Affairs*. 1988;17:206-216.
22. Lindley DV. Discussion of the paper by Spiegelhalter, Freedman, and Parmar. *J Royal Stat Soc A*. 1994;157:393.
23. Manning R, Claxton K. Experimental and econometric solutions to selection bias. Rotterdam: International Health Economists Association; 1999.
24. McClellan M, Newhouse JP. The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach. *Journal of Econometrics*. 1997;15:39-64.
25. Morris JC. The clinical dementia rating (CDR): Current version and scoring rules. *Neurology*. 1993;43:2412-2424.
26. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD), part 1: Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989;39:1159-1165.
27. Neumann PJ, Claxton K, Weinstein MC. The FDA's regulation of health economic information: Issues and options for Section 114 of FDAMA. *Health Affairs*. In press.
28. Neumann PJ, Hermann RC, Kuntz KM, et al. Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. *Neurology*. 1999;52:1138-1145.
29. Neumann PJ, Kuntz KM, Leon J. Health utilities in Alzheimer's disease: A cross sectional study of patients and caregivers. *Med Care*. 1999;37:27-32.
30. Pratt J, Raiffa H, Schlaifer R. *Statistical decision theory*. Cambridge, MA: MIT Press; 1995.
31. Pratt J, Zeckhauser RJ. Willingness to pay and the distribution of risk and wealth. *Journal of Political Economy*. 1996;104;747-763.
32. Raiffa H. *Decision analysis: Introductory lectures on choices under uncertainty*. New York: Addison-Wesley; 1968.
33. Raiffa H, Schlaifer R. *Probability and statistics for business decisions*. New York: McGraw-Hill; 1959.
34. Rice DP, Fox PJ, Max W, et al. The economic burden of Alzheimer's disease care. *Health Affairs*. 1993;2:164-176.
35. Rogers SL, Farlow MR, Mohs R, Friedhoff LT. A 24-week, double blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. *Neurology*. 1998;50:136-145.
36. Roger SL, Friedhoff LT. Long-term efficacy and safety of donepezil in the treatment of Alzheimer's disease: An interim analysis of the results of a US multicentre open label extension study. *Eur Neuropsychopharmacol*. 1998;8:67-75.

37. Samuelson PA. The pure theory of public good expenditure. *Review of Economics and Statistics*. 1954;36:387-389.
38. Schlaifer, R. Introduction to statistics for business decisions. New York: McGraw-Hill; 1961.
39. Sonnenberg FA, Beck JR. Markov models in medical decision making: A practical guide. *Med Decisi Making*. 1993;13:322-338.
40. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomised trials. *J Royal Stat Soc A*. 1994;157:357-416.
41. Stinnett AA, Paltiel D. Estimating C/E ratios under second order uncertainty: The mean ratio versus the ratio of the means. *Med Decis Making*. 1997;17:483-489.
42. Thompson KM, Burmaster DE, Crouch EAC. Monte Carlo techniques for quantitative uncertainty analysis in public health risk assessments. *Risk Anal*. 1992;1:253-263.
43. Thompson KM, Evans JS. The value of improved national exposure information for perchloroethylene (perc): A case study for dry cleaners. *Risk Anal*. 1997;17:253-271.
44. Thompson KM, Graham JD. Going beyond the single number: Using probabilistic risk assessment to improve risk management. *Human and Ecological Risk Assessment*. 1996;2;1008-1034.
45. Torrance GW, Feeny DH, Furlong WJ, et al. Multi-attribute preference functions for a comprehensive health status classification systems: Health Utilities Index Mark 2. *Med Care*. 1996;24: 702-722.