USC Dornsife Center for Economic and Social Research

USC Schaeffer Leonard D. Schaeffer Center for Health Policy & Economics

Cost Effectiveness Analysis With and Without Stochastic Uncertainty

Joel W. Hay

Paper No: 2015-005

CESR-SCHAEFFER WORKING PAPER SERIES

cesr.usc.edu **healthpolicy.usc.edu**

COST EFFECTIVENESS ANALYSIS WITH AND WITHOUT STOCHASTIC UNCERTAINTY

Draft: 5-1-15

Joel W. Hay, PhD
Professor, Pharmaceutical Economics & Policy
School of Pharmacy and
Schaeffer Center for Health Policy and Economics
University of Southern California
jhay@usc.edu

Preliminary versions presented at:

University of Tennessee Health Center, 2011
USC Schaeffer Center for Health Policy and Economics, 2011, 2012
UCLA, 2012
American Society of Health Economists, 2012
National University of Singapore, 2012
University of New Mexico, 2013
US Centers for Disease Control and Prevention, 2014
University of Washington Pharmaceutical Outcomes Research and Policy Program, 2015

COST EFFECTIVENESS ANALYSIS UNDER STOCHASTIC UNCERTAINTY

Objective: To evaluate the optimal allocation of health care treatments when treatment costs and outcomes are uncertain.

Background: In the context of perfectly certain health care costs and outcomes, Weinstein and others have demonstrated that optimal treatment allocation requires a comparison of all potential incremental cost effectiveness ratios (ICERs) against a decision-maker willingness-to-pay (WTP) ratio (cutoff threshold) using either linear programming methods, treatment 'prioritization' rules or incremental net monetary benefits (NMB) assessment. Thus far, stochastic treatment allocation evaluation has focused primarily on two-treatment comparisons where the uncertainty in the increment cost effectiveness ratio (ICER) is plotted on the cost-QALY plane and the stochastic properties are characterized in terms of WTP confidence intervals, incremental NMB plots, or cost effectiveness acceptability curves (CEAC). The decision rules and the characterization of global optimal treatment allocation under uncertainty has not been previously fully elucidated.

Results: With uncertain costs and outcomes, using NMB methods, it is demonstrated that if the decision-maker is only concerned about maximizing expected QALYs for a given budget (or in the dual optimization program, minimizing the expected cost of providing a fixed number of expected QALYs to the beneficiary population) all of the standard 'perfect certainty' results about optimal treatment allocation carry through. However, to the extent that the decision-maker is concerned about treatment cost and outcome variability (e.g., the decision-maker is risk averse) the optimal treatment allocation will necessarily involve a portfolio of treatment strategies involving tradeoffs between treatment risks and benefits. Unless all treatments are perfectly positively correlated, there will nearly always be some gain by using a mixed treatment strategy rather than characterizing certain treatments as 'dominant' or 'dominated.' The mathematics of modern portfolio theory apply directly to this problem and demonstrate 1:(Separation Theorem)—the optimal treatment portfolio efficiency frontier is independent of the decision-makers willingness-to-pay for health care treatments; 2: (Market Beta Analysis) the decision as to whether a new treatment should be added to the optimal treatment portfolio depends on the tradeoff between the new treatment incremental NMB and its covariance with the current optimal treatment portfolio risk.

Conclusions: Decision-makers using ICER methods to optimally allocate health care treatments under stochastic uncertainty will consider not only the NMB expected return to treatment but also treatment cost and outcome variability. This leads to the more realistic result that a portfolio of mixed treatment strategies will generally be preferable to one where only those treatments with the highest expected returns are implemented. This finding has substantial implications for formulary decision-making, comparative effectiveness research and the interpretation and characterization of stochastic ICERs.

Introduction

Weinstein and others have developed a theoretical method for allocation of health care resource allocation that, assuming all relevant parameters are known, has been shown to maximize health care outcomes (typically measured as QALYs-Quality Adjusted Life Years) subject to a health care planner's fixed budget.[1,2,3,4,5,6] This Cost Effectiveness Analysis Standard Model has been recommended by the U.S. Public Health Service Task Force on Cost Effectiveness in Health and Medicine[1] and adopted by many international health care payer authorities, including most notably the U.K. National Institute for Health and Clinical Excellence.[7]

It has been further shown that such an approach is amenable to solving for the unique optimal health care treatment allocation decision using three alternative calculation methods that produce equivalent optimal solutions; linear programming,[8,9] an iterative prioritization and editing binary treatment comparison algorithm approach,[10,11] and an approach that maximizes net monetary (or net health) benefits across feasible treatment comparisons.[12,13] In all cases, these solutions require that for every planner fixed health care budget C, there will be a unique critical ratio R which defines the planner's threshold willingness to pay for health (e.g., dollars per QALY). The planner will reject those treatments that have an incremental cost effectiveness ratio (ICER) above R and accept those non-dominated treatments that have ICERs below R.[13] While it is not necessary to repeat the optimality proof, a graphic representation helps to fix one's understanding (See Appendix Fig. 1 and the accompanying explanation).

Assume that there are I illnesses (i=1,2,…,I) with N patients distributed across these I illnesses (n_1, n_2, …, n_I) such that N = $\sum_i n_i$. Each illness i has j(i) possible treatments T_{ij} = (1,2,…, J(i)) with all T_{ij} assumed to possess constant returns to scale. Each treatment T_{ij} produces q_{ij} QALYs and costs c_{ij} ; $q_{ij}, c_{ij} \geq 0$ for all i,j. All diseases have a no treatment option $q_{i0} = c_{i0} = 0$ for all i.

It is straightforward that the optimal ranking of treatments doesn't change with the number of patients n_i in each disease category i.* Moreover, any distinction between diseases is not material for establishing the properties of the optimal treatment allocation. Many authors have referred to treatment alternatives as falling into "disease clusters" depending on whether or not the treatments are mutually exclusive or not.[10,13] For example, the use of chemotherapy or radiation treatment in metastatic pancreatic cancer might represent mutually exclusive alternatives, while a combination chemotherapy/radiation treatment might be a third treatment alternative for that disease. However, whatever treatments for pancreatic cancer are available, they are not mutually exclusive with treatments for hypertension. Given these observations we can abstract from the number of patients and diseases and simplify the analysis to consider a set of treatments, k, $k \in (1, 2, 3,\ldots, K)$ across all patients and diseases that may or may not be mutually exclusive.

The optimal allocation of treatments across diseases is characterized by ranking all treatments by their incremental cost effectiveness ratios (ICERs: $[c_i - c_j]/[q_i - q_j]$) from lowest to highest ICER, eliminating all treatments that are dominated (directly or through extended dominance), reordering all the non-dominated treatments by ICER rankings and repeating these steps until no re-orderings or re-rankings are possible. Given these final rankings, the planner then allocates all patients with disease 1 to non-dominated treatment 1, all patients with disease 2 to non-dominated treatment 2, etc. until the planner's fixed budget is exhausted at non-dominated treatment r.[10,13] The ICER at treatment r (relative to r-1) will reveal the planner's willingness to pay for health R as a function of the budget C.

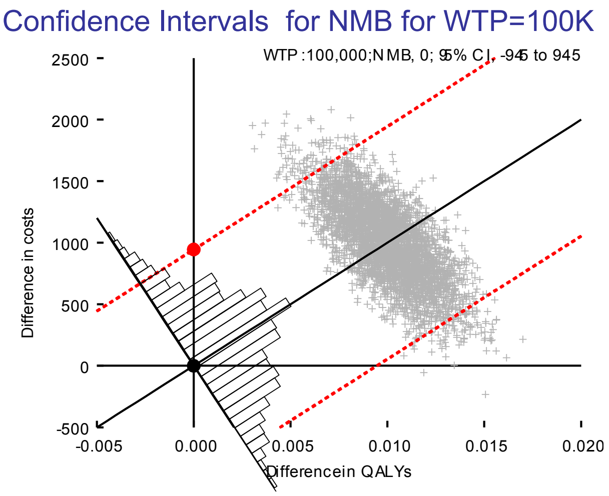
* For a proof take the Laska et al.[13] optimality condition, Equation 5, and multiply each term by the number of patients in each disease cluster.

Net Monetary Benefits

Net Monetary Benefit (NMB) is defined on the cost-effectiveness plane for a treatment T_i as $B_i = R*q_i - c_i$ where R is the "willingness-to-pay" for health (e.g., in dollars per QALY). Net Health Benefit (NHB) is defined as $Q_i = q_i - H*c_i$, where H = 1/R is the "willingness-to-pay" for health in QALYs per dollar. Health economists tend to think in terms of R rather than H – it it not very intuitive to talk about .00001 QALYs per dollar.

As shown in Figure 1, the Net Monetary Benefits associated with specific treatments i can be depicted on the Cost Effectiveness plane by a family of lines, all with a slope equal to R. Each line represents a single value of Net Monetary Benefit (NMB), which equals the negative of the point where the line intercepts the Y-axis (because at the intercept, q=0, thus R*Q=0, with NMB = -c). For the line with slope R passing through the origin, NMB = 0. Given R, NMB lines below and to the right of the NMB=0 line have positive net monetary benefits (i.e., acceptable cost-effectiveness ratios). Lines above and to the left of the NMB = 0 line have negative net monetary benefits. Lines increase in value as we travel southeasterly down the plane.

Figure 1. Net Monetary Benefits Lines with 95% Confidence



Optimal Treatment Allocation Without Uncertainty

Laska et al.[13] demonstrated the optimality proof for the cost effectiveness standard model eloquently and compactly using the net monetary benefits approach. They prove that the optimal treatment allocation maximizes the sum of net monetary benefits $B^* = \sum_i B_i$ across all possible allocations to the set of feasible non-dominated treatments. As they and others have shown, with perfect certainty for all parameters in the Cost Effectiveness Standard Model, the optimal treatment allocation consists of every disease category or disease cluster (containing mutually exclusive treatments) being treated by either only one dominant treatment (if $i < r$) , zero treatments (if $i > r$), or if i happens to be the marginal disease cluster with rank r, by at most two treatments (a linear combination of treatment r and treatment r-1).[†]

While this solution may characterize optimality under the Cost Effectiveness Standard Model, it is far from satisfying as a characterization of any system of health care treatment allocation in the real world. There are very few disease categories or clusters that are actually treated with only one, zero or at most two treatments. Moreover, it is unlikely that any real world P&T Formulary Committee or Treatment Guidelines Development Committee would ever commit to such a decision approach regardless of the quality and pedigree of the underlying cost effectiveness model. The primary reason for this is that no planner believes that any cost effectiveness model is fully accurate and reliable. There are unobservables (certainly ex ante) about heterogeneous patient treatment effects, adverse events and cost effectiveness model parameters that lead real world decision makers to be skeptical that any cost effectiveness model (or set of model results) can be so accurate that they would always correctly choose the single (or at most two) treatment(s) for each disease cluster with the highest net monetary benefit.

Treatment Allocation Under Uncertainty

[†] We continue to maintain the ranking of non-dominated treatments in the feasible choice set, i = 1,2,…, r-1, r. We assume purely for mathematical convenience that there are no exact ties for B_i and B_{i-1}. Thus, we assume that for every treatment there is some (possibly infinitesimally) small value $\epsilon > 0$ such that $B_i - B_{i-1} < \epsilon$ for all i.

There are many sources of stochastic and parameter uncertainty for the cost and QALY estimates in the Cost Effectiveness Standard Model. These can be formally characterized as:[11]

- Ex ante expectations differ from ex post outcomes (forecasting error)
- Individual patient and treatment variability (sampling error)
- Modeling parameter or model structural error (2nd Order uncertainty)

Patient and treatment heterogeneity and variability would typically cause the ex post results to vary around positive mean cost and QALY forecasts. Similarly, second order modeling error would lead to biased and variable forecasts that would diverge from the realized results but most likely would still result in positive realized costs and QALYs. However it is certainly plausible that in some cases the modeling errors would be large enough to exceed the forecast expected costs and QALYs in either the positive or negative directions.

An example of extreme negative forecasting error for QALYs was the treatment of hemophiliacs during 1980s before clotting factor could be tested for HIV.[14] At the time ex ante QALYs of clotting factor treatment were predicted to be positive. Ex post, clotting factor treatment wiped out an entire generation of hemophiliacs because clotting factor was contaminated with HIV and the realized QALYs were substantially negative (i.e. the forecasting error was larger than the ex ante projected benefits of treatment).

Ex post treatment costs could certainly be higher than ex ante expectations. They could be lower (e.g. fatal adverse reaction). Ex post treatment costs could even be negative because of forecasting error. For example patients could receive long term statin treatment for heart disease and later find that statins also reduce their risk of dementia. Another example of negative ex post costs would be stronger than expected herd immunity associated with a vaccination program (as was seen with the Haemophilus influenza type B vaccine in the 1990's).[15]

While a lot of progress has been made on characterizing the effects of uncertainty on cost effectiveness model results, to this point it has focused almost exclusively on characterizing statistical confidence intervals around ICER ratios for binary treatment comparisons.[16,17 ,18] This is certainly a non-trivial issue, since the statistical properties of ratios of stochastic variables are complicated. As Glick et al. detail,[11] depending on the statistical properties of the cost and QALY data, ICER statistical confidence intervals (CIs) can be either well-defined, undefinable, or there are situations where either the ICER point estimate exceeds the CI lower limit which exceeds the CI upper limit or the CI lower limit exceeds the CI upper limit which exceeds the ICER point estimate. It was in fact these strange properties of the ICER statistical confidence intervals that led researchers to turn to net monetary benefits assessments as a more tractable statistical alternative. Net monetary benefits measures are linear in cost and QALYs and do not have these unusual statistical properties.[12]

Optimal Treatment Allocation Under Uncertainty

While the effects of stochastic uncertainty on binary treatment comparisons can be characterized in various ways, including ICER confidence intervals, net monetary benefit confidence intervals and cost effectiveness acceptability curves,[11] only O'Brien and Sculpher have contemplated the effects of treatment uncertainty on overall treatment decision making.[19] The characterization of stochastic uncertainty on the optimal treatment allocation in the Cost Effectiveness Standard Model has not yet even been characterized.

Assume every treatment is stochastic due to 1st and 2nd-order uncertainty. For every treatment T_i, treatment QALY q_i and treatment cost c_i:

$q_i = E(q_i) + \varepsilon_{qi}$; $c_i = E(c_i) + \varepsilon_{ci}$

$E(\varepsilon_{qi}) = E(\varepsilon_{ci}) = 0$, $\mathrm{Var}(\varepsilon_{qi}) = \sigma_{qi}$, $\mathrm{Var}(\varepsilon_{ci}) = \sigma_{ci}$

$E(\varepsilon_{qi}\varepsilon_{ci}) = \rho\, \sigma_{qi}\, \sigma_{ci} \qquad (-1 \leq \rho \leq 1)$

We define the stochastic net monetary benefits for treatment T_i i as:

$B_i = Rq_i - c_i = R[E(q_i) + \varepsilon_{qi}] - E(c_i) - \varepsilon_{ci} = E(B_i) + \varepsilon_{Bi}$

$E(\varepsilon_{Bi}) = 0$, $\mathrm{Var}(\varepsilon_{Bi}) = \sigma_{Bi}$,

Furthermore we assume that across all diseases and disease clusters there are multiple treatments T_j with: $B_j = E(B_j) + \varepsilon_{Bj} \quad i \neq j$ and that there are positive definite covariance matrices Ω_B, Ω_q, Ω_c that define the correlations between all possible treatment net monetary benefits, QALYs and costs.

If we are only interested in maximizing expected treatment benefits, then the same solution that characterizes the optimal treatment allocation in the Cost Effectiveness Standard Model with perfect certainty still prevails. However this is difficult or impossible to demonstrate using either the linear programming approach, or the prioritization/editing algorithm approach, since both of these approaches suffer parallel problems to the potentially undefined or unbounded stochastic ICER ratios that occur with stochastic binary ICER ratios.

The proof of optimality under uncertainty is an extension of the Laska et al.[13] proof of optimality for the set of feasible treatments that maximizes the planner's net monetary benefits without uncertainty. If $\Sigma_i B^*$ is an optimal allocation of treatments without uncertainty then $E(\Sigma_i B^*) = \Sigma_i B^* + E(\Sigma_i \varepsilon_{Bi}) = \Sigma_i B^*$ will also be the optimal allocation of treatments with uncertainty as long as the planner wants to maximize expected net monetary benefits.

However in a world of stochastic uncertainty this optimal solution makes little sense. For example suppose one treatment has $E(B_i) = 10{,}000$ with a standard deviation of 100,000 and another treatment in the same disease cluster has $E(B_j) = 9{,}999$ with a standard deviation deviation of 0. Why would any planner ever prefer B_i over B_j? In an uncertain world, the planner should also consider treatment riskiness, not just treatment expected benefits.

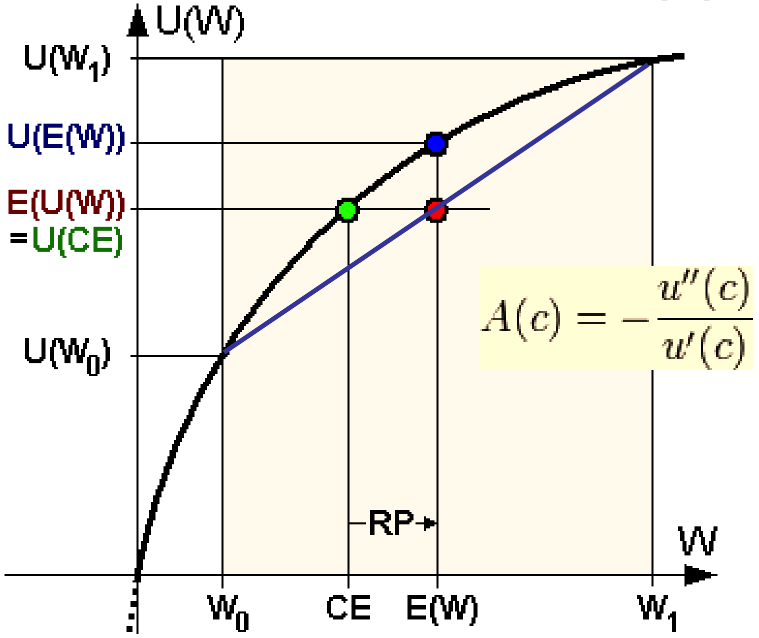
There are many examples demonstrating that the practice of medicine is highly risk averse including:

- Physician Practice Maxim "FIRST, DO NO HARM"
- FDA asymmetric treatment of errors in drug and medical device approvals and withdrawals
- Other medical care regulatory agencies and procedures
- The extensive certification and licensure procedures for every medical profession and every medical product and service
- Practitioners tend to be conservative -- they adopt new treatments slowly

Risk aversion is a well-known component of decision-making for consumers, insurance purchasers and all manner of economic agents. As shown in Figure 2, the basic concept is that people are willing to give up a specific amount of income (the risk premium) to insure that they receive a certain wealth (or health) compared to the situation where they could have larger expected wealth (health) with some additional risk.

Figure 2: Utility with Risk Aversion

Absolute Risk Aversion A(c)



It has been argued that if the health care planner treats a large enough population (e.g., a national health insurance program) and has a long enough perspective, then since they are concerned with maximizing expected population health, individual variability in treatment isn't an important decision attribute, even in a stochastic world. However, this view ignores the uncertainties in the cost effectiveness models that the planners use themselves, as well as unknown treatment variability and heterogeneity. Since the decisions for medical treatment are made by physicians for each patient, one at a time, it would seem that risk aversion should be an important characteristic in the planner's utility associated with alternative treatments.

Suppose $\bar{B}_i > \bar{B}_j$. We still might want some patients to use $\bar{B}_j$. If $\sigma_i > \sigma_j$ $\bar{B}j$ has less risk. Suppose treatments i and j have correlation ρ ($-1 < \rho < 1$), if we combine the two treatments into a treatment portfolio P with weight α ($0 < \alpha < 1$),

$E(B_P) = \alpha \bar{B}_i + (1- \alpha) \bar{B}_j$.

$\mathrm{Var}(B_P) = \alpha^2 \sigma_i^2 + (1- \alpha)^2 \sigma_j^2 + 2\alpha(1 - \alpha) \sigma_i \sigma_j\rho$

For example, suppose $\overline{B}_i = \overline{B}_j$ and $\sigma_i = \sigma_j$, $\rho = 0$ (independent treatments). Chose $\alpha = .5$;

$E(B_P) = .5\ \overline{B}_i + (1- .5)\ \overline{B}_j = \overline{B}_i$

$\mathrm{Var}(B_P) = .5^2\ \sigma_i^2 + (.5)^2\sigma_j^2 + [\ 2\ \alpha\ (1 - \alpha)\ \sigma_i\ \sigma_j\ \rho\ ] = .5\ \sigma_i^2$

In this case, diversification has cut treatment portfolio risk in half.

Suppose $\overline{B}_i = \overline{B}_j$ and $\sigma_i = \sigma_j$, $\rho = -1$ (perfectly negatively correlated treatments). Chose $\alpha = .5$;

$E(B_P) = .5\ \overline{B}_i + (1- .5)\ \overline{B}_j = \overline{B}_i$

$\mathrm{Var}(B_P) = .5^2\ \sigma_i^2 + (.5)^2\sigma_j^2 + [\ 2\ \alpha\ (1 - \alpha)\ \sigma_i\ \sigma_j\ \rho\ ] = 0$

In this case, diversification has eliminated treatment portfolio risk.

Only if all feasible treatments are perfectly positively correlated with each other does it make sense to ignore treatment portfolio diversification (in that case choose just the treatment with the lowest standard deviation). Otherwise, optimal treatment portfolio selection will depend on both the planner's level of risk aversion and the potential risk diversification available in feasible treatment portfolios.

Treatment Portfolio Theory

The theory of optimal treatment allocation under uncertainty is mathematically equivalent to Modern Portfolio Theory (MPT) which was pioneered by Harry Markowitz in 1952.[20] The underlying assumptions of portfolio theory are actually even better-suited to stochastic health care treatments than to financial assets. Treatment NMB returns are assumed to be jointly normally distributed or otherwise elliptically distributed. This implies that one can completely characterize treatment returns solely in terms of means and variances.[‡]

The treatment portfolio expected return is the proportion-weighted combination of the constituent returns $E(B_p) = \Sigma_i\ \alpha_i\ \overline{B}_i$ where α_i is the proportion of the total budget C assigned to treatment i ($0 \leq \alpha_i \leq 1$). Letting α represent the vector ($\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_I$), then the portfolio risk is $\alpha'\Omega_B\alpha$.

For any given level of expected NMB return, the planner will want to minimize risk. The planner's optimization program is thus:

Choose the vector α {α_i} to minimize the quadratic Lagrangian:

$$\alpha'\Omega\alpha$$

$$\text{subject to}\quad E(B)\alpha = \mu$$

Thus one first solves the cost effectiveness model without uncertainty to get willingness to pay R. Then given R one can find what allocation of budget α to treatments minimizes risk subject to achieving a given expected NMB return.

An alternative formulation of the optimization problem is to choose α to minimize

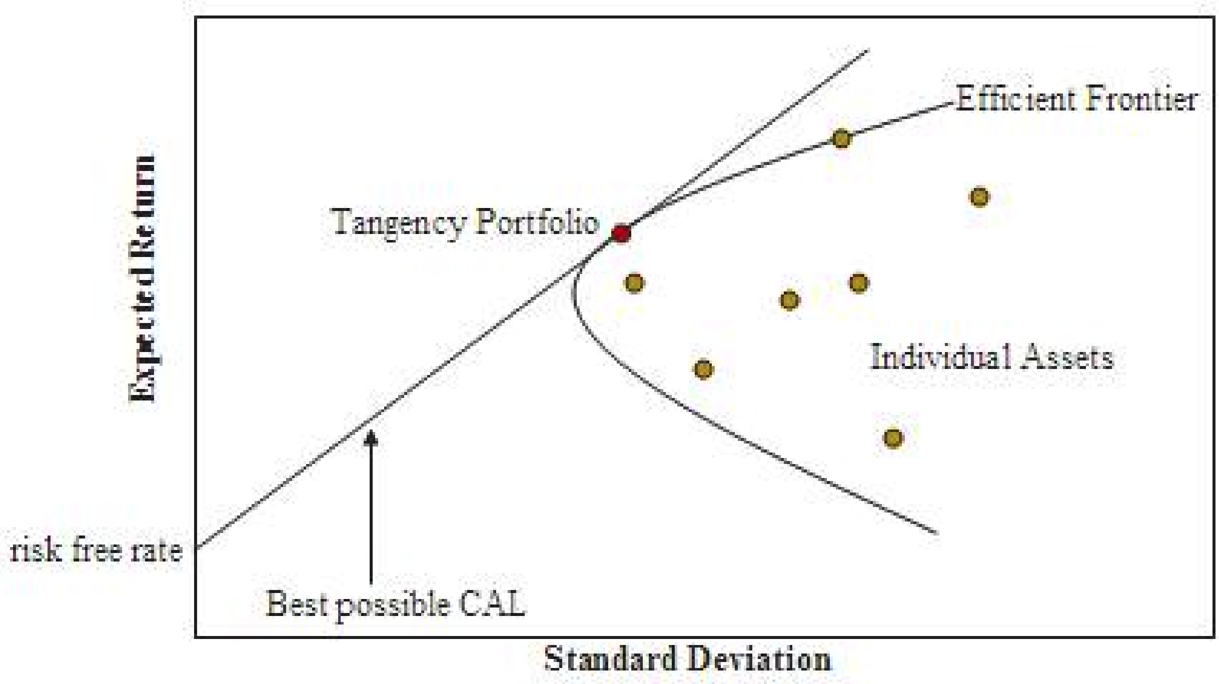
$$\alpha'\Omega\alpha\ -\ \theta E(B)\alpha$$

where θ reflects the planner's risk aversion propensity and is proportional to the planner's risk aversion coefficient $A(c) = -\ (\partial^2 U(c)/\ \partial c)/\ (\partial U(c)/\ \partial c)$.

Markowitz showed that the optimal solution lies on an quadratic efficiency frontier known as the "Markowitz Bullet" (Figure 3). This efficiency frontier consists of combinations of individual treatments that minimize portfolio risk for any given expected NMB return.

[‡] MPT has generated at least 6 Nobel Prizes in Economics and also generated several financial crashes, including 2008—present.

Figure 3. Characterization of the Optimal Minimum Risk Treatment Portfolio



CAL stands for Capital Allocation Line

If there is a risk free alternative for the available funds (e.g., invest in government bonds rather than health care treatments) then the budget line between the tangency portfolio and the risk-free investment in Figure 3 represents the planner's feasible budget line for treatment options. As has been shown in financial portfolio theory, the optimal portfolio of stochastic treatments does not depend on the planner's risk aversion or the planner's expected NMB return requirements. There is a single "mutual fund" portfolio of treatments that will be then mixed with the no-treatment (risk free securities) option to minimize the planner's total risk for any given level of expected return. This result is known at the Portfolio Separation Theorem.[21] While the optimal "mutual fund" portfolio of health care treatments does depend on R and the risk, return and correlation properties of the individual treatments, it doesn't reflect the planner's own utility trade-off between risk and return. Thus given the parameters of the system any planner would chose the same health care treatment portfolio, regardless of their risk versus reward preferences.

This separation theorem implies that by empirically determining the treatment risk-return efficiency frontier we can gain better insights into what treatments actually belong in the optimal health care treatment portfolio, regardless of the planners risk preferences. In fact, in a stochastic world, there is no reason to expect that any specific health care treatment is always dominated (either directly or through extended dominance). It is quite possible that treatments that are considered dominated in a perfectly certain cost effectiveness framework would still have the valuable characteristics of independence or negative correlation with other treatments to enter into the optimal portfolio of minimum risk treatments given the expected NMB returns. This further suggests that formulary decision makers that make sharp distinctions on copayment status for different treatments in a given disease cluster might want to consider some of these risk and correlation issues more carefully.

It is also possible to apply the results of MPT to the question of whether a new health care treatment A should be added to optimal treatment portfolio. Treatment A should be added only if:

$[E(B_A) - E(B_{RiskFree})] / \sigma_{AP} > [E(B_P) - E(B_{RiskFree})] / \sigma_{PP}$

Or if:

$[E(B_A) - E(B_{RiskFree})] > [E(B_P)- E(B_{RiskFree})]\sigma_{AP}/\sigma_{PP}$

σ_{AP}/σ_{PP} is referred to as the beta for treatment A. It is the coefficient of a linear regression of the Net Monetary Benefits of Treatment A on the Net Monetary Benefits of the entire treatment portfolio. To the extent that these covariances and expected NMB treatment returns can be determined empirically, as with financial assets, MPT theory thus gives additional insights on where biomedical researchers should turn to generate new treatments that are most likely to improve the returns to health care investments.

Discussion

Optimization of health care treatment allocation without uncertainty defines a unique relationship between the global budget C and the health care planner's willingness to pay for QALYs, R. Ironically the main reason for the development of the Cost Effectiveness Standard Model was because clinicians were strongly opposed to using cost benefit analysis in health care. They found it highly objectionable to assign dollar values to human life, and different dollar values to different health states. As it turns out the planner's threshold willingness to pay for QALYs, R, is now a fundamental component of the optimal solution, and certainly can't be avoided in health care treatment allocation. There is no reason not to close the circle and move back to cost benefit analysis, as economists have been advocating for decades.

Optimization with uncertainty requires the health care decision maker to formally consider how much risk they are willing to take. It also requires diversification of the treatment portfolio, which is much more realistic than the optimization result without uncertainty where for each disease cluster there are only zero, one, or at most two, possible treatments in the optimal treatment mix. It further eliminates the sharp distinction between dominant and dominated treatments, since a dominated therapy may still have value based on its risk and portfolio correlation profile.

By focusing only on stochastic properties of ICER and NMB differences rather than the risk characteristics of overall treatment portfolio selection, researchers to date have ignored a key aspect of health care treatment allocation -- Is the expected net benefit of a new treatment worth its risk? Clearly the information requirements needed to actually apply these MPT concepts to real world health care decision making are quite daunting. As electronic health records become increasingly more widely-used these data problems may become more manageable for managed care plans and government payers. In any case the difficulty in developing empirical applications of NMB portfolio theory doesn't negate the need to assess the theoretical bases for how health care treatments should be allocated under uncertainty.

Appendix:

Define:

R = Societal cut-off willingness to pay for QALYs; $R \geq 0$.

C = Societal budget for health care; $C \geq 0$.

Let C denote the planner's budget or the total cost of its resource base, and let N_i be the total number of patients having illness i. Then the Weinsteinian planner's decision problem—and the core problem of global CEA—is:

$$(1)\quad \left[\begin{array}{l} \text{Choose the } n_{ij} \geq 0,\ j = 1,2,\ldots,J_i;\ i = 1,2,\ldots,I \\[2ex] \text{To maximize } Q = \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J_i} q_{ij}n_{ij} \\[2ex] \text{Subject to the I+1 constraints } \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J_i} c_{ij}n_{ij} \leq C;\ \sum\limits_{j=1}^{J_i} n_{ij} = N_i,\ i = 1,2,\ldots,I. \end{array}\right.$$

Moreover, for all i,j $\quad c_{ij} \geq 0\ q_{ij} \geq 0$

In the optimal solution to 1, the shadow price $\pi(C)$ of the total cost constraint is therefore the reciprocal of a money price (i.e., it is expressed in units of health benefits per unit of money). Inasmuch as the health care planner is Weinsteinian, we assume that the total cost constraint is always binding at the optimum. Thus in the sense of Dorfman et al.[22] $1/\pi(C)$ is the money "price" of obtaining an optimal marginal increase in total health benefits by relaxing the total cost constraint by a given marginal money amount. It is, however, more readily understood as a marginal cost. To see that, observe that the inverse function of Q(C), $Q^{-1}(C) = C(Q)$, and the partial derivative $\frac{\partial C(Q)}{\partial Q} = \left[\frac{\partial Q(C)}{\partial C}\right]^{-1}$ both exist by the Inverse Function Theorem from classical mathematical analysis (e.g., Rudin [1976, p. 221]). As a consequence, $\frac{\partial C(Q)}{\partial Q} = \frac{1}{\pi(C)}$ by statement (iv) above, so that $\frac{1}{\pi(C)}$ is interpreted as the marginal cost of producing health benefits at the optimum.

We postulate that for each illness i and each R there exists a C and that the function C(R) is unique everywhere above some minimal threshold R^o, monotonic above R^o with dC/dR > 0 and possesses an inverse function $R= C(R)^{-1}$ that is monotonic with dR/dC >0 above R^o.

Define: Bj as the Net Monetary Benefit (NMB) associated with treatment j.[11,12]

$B_j = q_jR – c_j$

j = 1,2,…J

We further define illness-specific NMB $B_{ik} = q_{ik}R - c_{ik}$ for each R, for all feasible treatments k, and for each illness i, i= 1,2,…I. However this additional notation is not necessary since optimal treatment allocation rules hold either within a single illness or across illnesses.

Lemma 1A; If $B_k \geq B_j$ = then treatment k is cost effective relative to treatment j given R.

Sequential proofs for:

1) $q_k > q_j$;

2) $q_k = q_j$;

3) $q_k < q_j$:

1) suppose $q_k > q_j$:

$B_k \geq B_j \Rightarrow q_kR - c_k \geq q_jR - c_j \Rightarrow R(q_k - q_j) \geq (c_k - c_j) \Rightarrow R \geq (c_k - c_j)/(q_k - q_j)$ QED

2) suppose $q_k = q_j$:

$B_k \geq B_j \Rightarrow q_kR - c_k \geq q_jR - c_j \Rightarrow R(q_k - q_j) \geq (c_k - c_j) \Rightarrow Rx0 \geq (c_k - c_j) \Rightarrow c_j \geq c_k$; and since $q_k = q_j$ QED

3) suppose $q_k < q_j$:

$B_k \geq B_j \Rightarrow q_kR - c_k \geq q_jR - c_j \Rightarrow R(q_k - q_j) \geq (c_k - c_j) \Rightarrow c_j > c_k$

Moreover this means;

$R(q_j - q_k) \leq (c_j - c_k) \Rightarrow R \leq (c_j - c_k)/(q_j - q_k)$ QED

Lemma 1B: If $B_k > B_j$ = then treatment k is strictly cost effective relative to treatment j given R.

Proof; Same as for Lemma 1A, replacing the loose inequality signs with strict inequality signs.

Define:

$B_0 = 0$ = NMB of doing nothing.

Lemma 2; No treatment with an NMB < 0 is ever cost effective.

Proof: Follows from Lemma 1B with treatment k being the no treatment option.

Definition: A treatment k for an illness is dominated if there exists a subset of treatments for the illness not containing treatment k, call it **D**, and a set of positive weights α_j such that

(2.1) all $\alpha_j > 0$, $\sum_{j \text{ in } \mathscr{D}} \alpha_j = 1$, $c_k \geq \sum_{j \text{ in } \mathscr{D}} \alpha_j c_j$, and $q_k \leq \sum_{j \text{ in } \mathscr{D}} \alpha_j q_j$,

and the treatments in **D** for which (2.1) holds are said to dominate treatment k. Perhaps more understandably (2.1) can also be written

(2.1a) all $\alpha_j > 0$, $\sum_{j \text{ in } \mathscr{D}} \alpha_j = 1$, $c_k n_k \geq \sum_{j \text{ in } \mathscr{D}} c_j (\alpha_j n_k)$, and $q_k n_k \leq \sum_{j \text{ in } \mathscr{D}} q_j (\alpha_j n_k)$.

Definition: A treatment k for an illness is strictly dominated if at least one of the two inequalities $c_k \geq \sum_{j \text{ in } \mathscr{D}} \alpha_j c_j$ and $q_k \leq \sum_{j \text{ in } \mathscr{D}} \alpha_j q_j$ in (1) is strict.

Lemma 3A; If treatment k is dominated by (composite) treatment j then $B_k \leq B_j$.
Moreover, If treatment k is (strictly) dominated by (composite) treatment j then $B_k < B_j$.

Proof: Let treatment j be the composite treatment with weights all $\alpha_j > 0$, $\sum_{j \text{ in } \mathscr{D}} \alpha_j = 1$,

If treatment k is dominated by treatment j, then by (2.1):

$c_k \geq \sum_{j \text{ in } \mathscr{D}} \alpha_j c_j$ and $q_k \leq \sum_{j \text{ in } \mathscr{D}} \alpha_j q_j$

$B_k = Rq_k - c_k$

$B_j = R[\sum_{j \text{ in } \mathscr{D}} \alpha_j q_j] - \sum_{j \text{ in } \mathscr{D}} \alpha_j c_j$

$Bj - Bk = R[\sum_{j \text{ in } \mathscr{D}} \alpha_j q_j] - \sum_{j \text{ in } \mathscr{D}} \alpha_j c_j - [Rq_k - c_k] = R\{[\sum_{j \text{ in } \mathscr{D}} \alpha_j q_j] - q_k\} - \{\sum_{j \text{ in } \mathscr{D}} \alpha_j c_j - c_k\}$.

However, by (2.1) the first of the curly bracketed terms is non-negative and the second curly bracketed term is non-positive so that the difference is non-negative.

Proof of the strict dominance case follows similarly.

Lemma 3B; If (composite) treatment k is dominated by treatment j then $B_k \leq B_j$.

Moreover, If treatment k is (strictly) dominated by (composite) treatment j then $B_k < B_j$.

Proof: Parallel to the proof for Lemma 3A.

Lemma 3C; If (composite) treatment k is dominated by (composite) treatment j then $B_k \leq B_j$.

Moreover, If (composite) treatment k is (strictly) dominated by (composite) treatment j then $B_k < B_j$.

Proof; Follows from Lemmas 3A and 3B.

Consider only $R > 0$, $C > 0$;

Lemma 4; If, given R, $B_{ij} > B_{ik}$, then for any $R' > R$ where B'_{ij} is defined to be $R'q_{ij} – c_{ij}$ and B'_{ik} is defined to be $R'q_{ik} - c_{ik}$

$B'_{ij} > B'_{ik}$.

Proof: Let $R’ = R(1+e)$ for any $e > 0$

$B_{ij} > B_{ik}$ ➔ $Rq_{ij} – c_{ij} > Rq_{ik} - c_{ik}$ ➔ $R(1+e)q_{ij} – (1+e)c_{ij} > R(1+e)q_{ik} - (1+e)c_{ik}$

➔ $R’q_{ij} – (1+e)c_{ij} > R’q_{ik} - (1+e)c_{ik}$

➔ $R’ > (1+e) [c_{ij} - c_{ik}] / [q_{ij} - q_{ik}] > [c_{ij} - c_{ik}] / [q_{ij} - q_{ik}]$

➔ $R > [c_{ij} - c_{ik}] / [q_{ij} - q_{ik}] > \{1/ (1+e)\} [c_{ij} - c_{ik}] / [q_{ij} - q_{ik}]$

Now suppose $B'_{ij} \le B'_{ik}$.

$B'_{ij} \le B'_{ik} \Rightarrow R(1+e)q_{ij} - c_{ij} \le R(1+e)q_{ik} - c_{ik} \Rightarrow R \le \{1/(1+e)\} [c_{ij} - c_{ik}] / [q_{ij} - q_{ik}]$

But this contradicts the last line above, therefore $B'_{ij} > B'_{ik}$

Lemma 4; If $n^* = (n_{1j(1)^*}, n_{2j(2)^*}, \ldots n_{Ij(I)^*})$ is the optimal allocation of the n_{ij} patients across all treatments and illnesses that solved the linear programming problem A.1, then define $B_{i^*j^*}$ to be the (composite) Net Monetary Benefit associated with the optimal assignment of patients to all treatments j within each illness i that solves the linear programming problem A.1.

Then the total Net Monetary Benefit $\Sigma n_{ij}{}^* B_{i^*j^*}$ will be maximized at n*, subject to the constraints of (A.1)

Proof; Let n_{ii^*} be the (composite) allocation of all n_{ii} patients across all treatments j for illness i. Suppose there were some other feasible (composite) allocation n(k) of the n_{ij} patients across all treatments j for each illness i, $\Sigma n_{ij}(k)B_{ij,}$ such that $\Sigma n_{ij}(k)B_{ij} = R[\sum_{j \text{ in } \mathscr{D}} n_{ij}\alpha_j q_{ij}] - \sum_{j \text{ in } \mathscr{D}} n_{ij}\alpha_j c_{ij}$

and $\Sigma n_{ij}(k)B_{ij} > \Sigma n^* B_{ij^*} \Rightarrow R(q_{1k} - q_{1j^*}) > (c_{1k} - c_{1j^*})$ (5.1)

However, since a solution to A.1 exists (n*) we know that $R(q_{1k} - q_{1j^*}) \le 0$ for any feasible q_{1k} or else q_{1j^*} would not provide the highest feasible q to the n_1 patients. We also know from the solution to the dual program to A.1 that $0 \le (c_{1k} - c_{1j^*})$ for any feasible c_{1k} or else c_{1j^*} would not provide the lowest feasible c to the n_1 patients.

Thus $R(q_{1k} - q_{1j^*}) \le 0 \le (c_{1k} - c_{1j^*})$ (5.2)

But given that the solution n* to A.1 is unique, the only possibly solution to both (5.1) and 5.2) is $q_{1k} = q_{1j^*}$, $c_{1k} = c_{1j^*}$.

Since $B_{i*j*} > B_{i*k}$ would similarly be true for all illnesses i, i = 1,2,…I, it follows from Lemma 3 that

ΣB_{i*j*} will be maximized at n*, subject to the constraints of (A.1).

Optimal Treatment Solution to Weinsteinian Planner Decision Problem A.1

Given R, Define B_{i*j*} to be the maximum achievable net monetary benefit for illness i across all available T_{ij} treatments j for illness i. For each illness i choose treatment i*j* such that $B_{i*j*} = \text{Max}\,(B_{ij})$ over all treatments j for illness i. Treatment T_{i*j*} can be a pure treatment or a composite treatment (linear combination of several treatments $\alpha_{ij}T_{ij}$ with $0 \leq \alpha_{ij} \leq 1$, $\Sigma\, \alpha_{ij} = 1$).

Since all treatments have constant returns to scale, one then ranks the B_{i*j*} over all illnesses i, and all treatments within illness i, j(i). One then ranks the B_{i*j*} sequentially:

$B_{i*j*} \geq B_{i*j+1*} \geq \ldots \geq B_{i*j(i)*} \geq B_{i+1*1*} \geq B_{i+1*2*} \geq \ldots B_{i+1*j(i)*} \geq B_{i+2*1*} \geq \ldots \geq B_{I*j(I)*}$

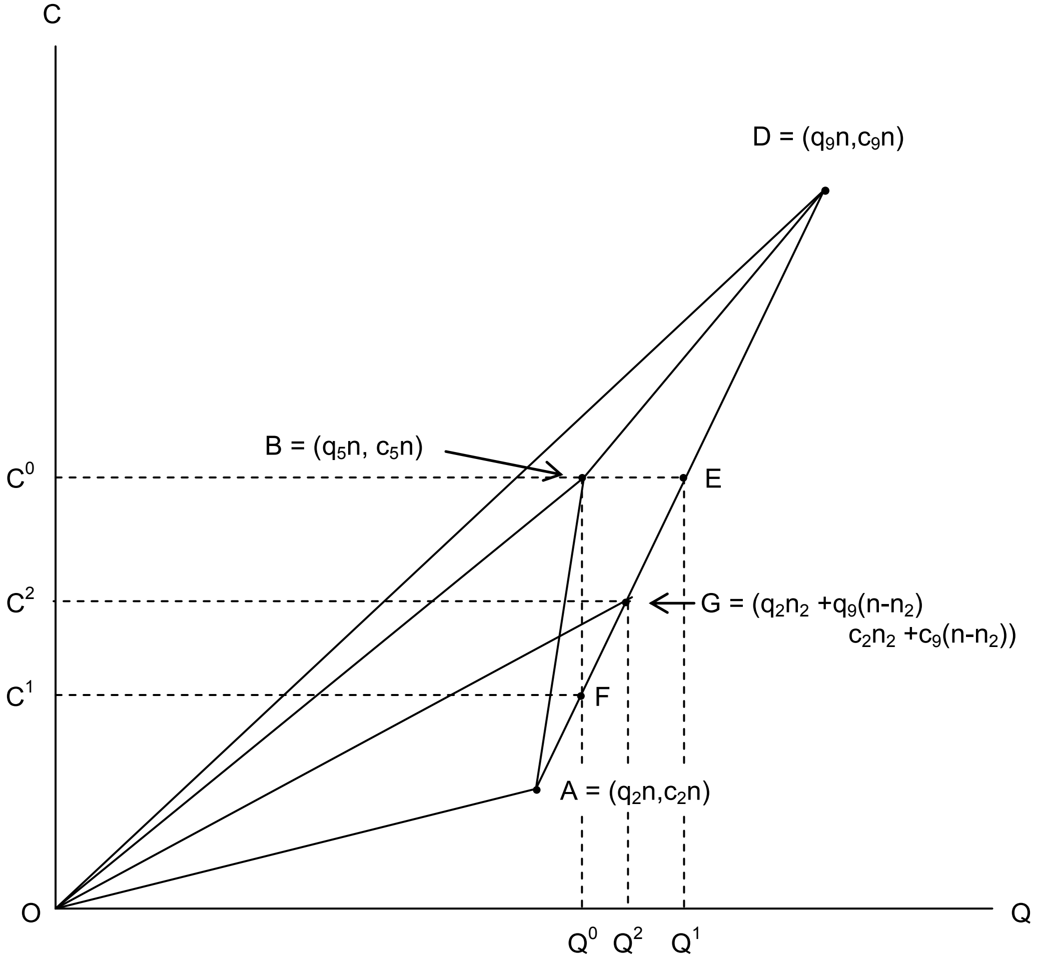
allocates the n_{ii} patients for each illness i to treatments 1*1*, 1*2*,1*j(1)*, 2*1*,2*2*,2*j(2)*,…,I*1*,I*2*,…I*j(I) where 1 is the illness with the highest NMB B_{1*1*}, 2 is the illness with the second-highest NMB B_{2*1*}, ….,I is the illness with the lowest NMB B_{I*J*} until one runs out of budget, C, given R, or B_{i*j*} becomes negative for the next disease treatment ranking. As long as one has unspent budget C remaining given R, one can raise R slightly and recompute the optimal treatment allocation B(R+ε) or use some other numerical computation algorithm to converge to R* = R(C).

Proof: Follows from Lemmas 1 - 4.

We postulate that it is computationally easier to solve this allocation problem for each R and iterate to the solution for each C using treatment NMBs, than to solve either the linear programming problem A.1 or the treatment prioritizing and editing algorithm.

Figure 1. Graphical Representation of Optimal Solution

Consider a single illness, i, with nine possible treatments for up to n patients.



We know that the polygon OAD traces out the feasible production possibility frontier, since any point interior to OAD, if feasible, can be improved on by a point on the frontier with the same cost and greater q (e.g., compare B to E). Now consider some planner willingness to pay threshold level R with budget C^2. Suppose R had the same slope as the ray OG. Then any interior point p on the C-Q plane to the southeast of the ray OG (e.g., point F) would be dominated by a feasible point on the production possibility frontier OAD with the same cost (on the Y-axis) and higher NMB (since the line with slope R though such a frontier point would intersect the Y-axis at a more negative point than the line with slope R though p). Furthermore any point on the production possibility frontier OAD below G leaves some unspent budget at C^2.

References

[1] Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

[2] Sloan, FS. Ed. Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. Cambridge University Press, New York, NY. 1995.

[3] Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia: WB Saunders, 1980.

[4] Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Eng J Med 1977; 296: 716-721.

[5] Weinstein MC, Zeckhauser R. Critical ratios and efficient allocation. Public Economics 1973; 2: 147-157.

[6] Torrance GW, Thomas WH, Sacket DL. A utility maximization model for evaluation of health care programs. Health Services Research 1972; 7: 118–133.

[7] Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Third edition: Oxford: Oxford University Press; 2005.

[8] Johannesson M, Weinstein, MC. On the decision rules of cost-effectiveness analysis. Journal of Health Economics 1993; 12: 459–467.

[9] Stinnett AA, Paltiel AD. Mathematical programming for the efficient allocation of health care resources. Journal of Health Economics 1996; 15: 641-653.

[10] Karlsson G, Johannesson M. The decision rules of cost-effectiveness analysis. Pharmacoeconomics 1996; 9: 113-120.

[11] Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. Oxford University Press, New York, NY, 2007.

[12] Stinnett AA, John Mullahy J. Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. Med Decis Making 1998;18 suppl:S68-S80.

[13] Laska GM, Meisnera M, Siegela C, Stinnett AA. Ratio-Based And Net Benefit-Based Approaches To Health Care Resource Allocation: Proofs of Optimality and Equivalence. Health Econ. 8: 171–174 (1999).

[14] Hay JW, Ernst R, Kessler G. Cost-effectiveness analysis of alternative factor VIII products in treatment of hemophilia A, Haemophilia, 1999, Vol. 5, pp. 191-202.

[15] Hay JW. and Daum, R. Cost-benefit Analysis of Haemophilus influenzae Type B Prevention: Conjugate Vaccination at Eighteen Months of Age, Ped Inf Dis J, April 1990, Vol. 9, pp. 246-252.

[16] O'Brien BJ, Briggs AH. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. Statistical Methods in Medical Research 2002; 11: 455-468.

[17] Glick HA, Briggs AH, Polsky D. Quantifying stochastic uncertainty and presenting results of cost-effectiveness analyses. Expert Rev. Pharmacoeconomics Outcomes Res. 1(1), 25–36 (2001).

[18] Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Econ 1997;6:243-252.

[19] O'Brien BJ, Sculpher MJ. Building Uncertainty into Cost-Effectiveness Rankings: Portfolio Risk-Return Tradeoffs and Implications for Decision Rules. Medical Care 38(5), May 2000, pp 460-468.

[20] Markowitz, H.M. (March 1952). "Portfolio Selection". The Journal of Finance 7 (1): 77–91.

[21] Merton, Robert. An analytic derivation of the efficient portfolio frontier. Journal of Financial and Quantitative Analysis, 1972; 7, 1851-1872.

[22] Dorfman R, Samuelson PA, Solow RM. Linear Programming and Economic Analysis. New York: McGraw-Hill Book Company. 1958.