A Bayesian Approach to Aid in Formulary Decision Making: Incorporating Institution-Specific Cost-Effectiveness Data with Clinical Trial Results

Shelby D. Reed, PhD, Peter W. Dillingham, MS, Andrew H. Briggs, DPhil, David L. Veenstra, PhD, PharmD, Sean D. Sullivan, PhD

Pharmacy and therapeutics committees commonly cite a lack of generalizability as a reason for not incorporating cost-effectiveness information into decision making. To address this concern, many committees undertake site-specific economic evaluations, which are often limited by small sample sizes and nonrandomized designs. We show how 2 complementary approaches were used to minimize these limitations in an economic evaluation of abciximab at 1 institution. Using a propensity score methodology, we selected patients who did not receive abciximab for the comparison cohort. Then, we adopted a Bayesian, hierarchical, random-effects model to integrate site-specific and clinical trial data. We applied the posterior distributions of effectiveness with local cost data in a traditional decision-analytic model. In 74% of the simulations, abciximab was cost-effective at 1 institution at the $50,000 per life year saved threshold, assuming a 50:50 split of patients undergoing coronary stenting and angioplasty. Among patients undergoing coronary stenting, the cost-effectiveness ratio of the addition of abciximab was at or below the $50,000 per life year saved threshold in 66.0% of the simulations. ***Key words:*** *Bayesian analysis; cost-effectiveness; decision making; formulary decision making; Monte Carlo simulation.*

Randomized clinical trials (RCTs) are the widely accepted gold-standard design for conducting clinical efficacy research. In recent years, cost-effectiveness analyses (CEAs) increasingly have been conducted alongside RCTs to take advantage of their methodological rigor.[1,2] However, decision makers in health care frequently state concerns that patients, providers, practice patterns, and costs are different at their respective institutions. Thus, they claim that they cannot directly apply the results of CEAs based on data from RCTs when deciding on the local adoption of an intervention. In response to these issues, institutions or managed care plans may undertake site-specific or plan-specific economic evaluations to assess whether an intervention is cost-effective in their settings.

Received 27 June 2001 from the Pharmaceutical Outcomes Research and Policy Program, University of Washington (SDR, DLV, SDS); the Center for Clinical and Genetic Economics, Duke Clinical Research Institute, Duke University (SDR); the Center for AIDS Research, Harborview Hospital, University of Washington (PWD); and the Health Economics Research Centre, Institute of Health Sciences, University of Oxford (AHB). Abstract presented at the 22nd annual meeting of the Society for Medical Decision Making, 2000. Financial support for this study was provided by the authors' institutions. The authors thank the 2 anonymous reviewers for their questions, comments, and advice. Their effort contributed greatly to the final version of this article. Revision accepted for publication 28 January 2003.

Address correspondence and reprint requests to Shelby D. Reed, PhD, Center for Clinical and Genetic Economics, Duke Clinical Research Institute, Duke University, P.O. Box 17969, Durham, NC 27715; telephone: (919) 668-8991; fax: (919) 668-7124; e-mail: reed0034@mc.duke.edu.

However, there are 2 major problems that arise in many local CEAs: an inability to randomize and insufficient statistical power. If the treatment in question has already been deemed to be safe and efficacious by the US Food and Drug Administration, institutional review boards will often not approve randomized designs to study the effectiveness of the drug locally, even when the population of patients is argued to differ from the patients enrolled in the RCTs. Therefore, analysts must rely on nonrandomized designs to evaluate the effectiveness of the drug as used in their institutions. With regard to statistical power, a sufficient sample size for detecting a statistical difference for CEAs is usually considerably larger than that required for health outcomes.[3] Furthermore, it is likely that local institutions will be treating fewer patients than were recruited to the RCTs, raising concern over the power to detect true differences, even in the clinical effect of treatment.[4] In these cases, differences in health outcomes (the denominator of the cost-effectiveness ratio) are not statistically different from zero, and the calculation of confidence intervals for incremental cost-effectiveness ratios (ICERs) is problematic.[5–7]

STUDY AIM

The aim of this article is to provide a real-world example of how the use of 2 complementary and innovative approaches can minimize limitations associated with relatively small, nonrandomized, site-specific CEAs. First, we used a propensity score methodology to select appropriate patients for a local comparison cohort. Second, we adopted a Bayesian statistical approach to derive institution-specific cost-effectiveness estimates on the basis of a combination of local and RCT data.

We expand on a previously conducted CEA of abciximab, a drug used in conjunction with percutaneous coronary revascularization procedures to prevent ischemic complications such as death, myocardial infarction (MI), and subsequent revascularization.[8] In 1995, the pharmacy and therapeutics committee at the University of Maryland Medical System (UMMS) added abciximab to the formulary on a provisional basis, under the premise that its formal inclusion would be reevaluated at the conclusion of the CEA. At that time, only 1 trial, the EPIC trial, had been published regarding the efficacy of abciximab.[9] The trial was conducted on high-risk patients undergoing balloon angioplasty, but many of the patients at UMMS were receiving abciximab in conjunction with coronary stenting, and many patients at UMMS would not have met the trial's inclusion criteria because they were at a lower risk for ischemic events. Therefore, on introduction of abciximab, many questions existed about its real-world effectiveness outside of clinical trial settings. While the local study was being conducted, the results of EPILOG and EPISTENT were published, thereby minimizing concerns regarding the effectiveness of the drug in lower risk patients and in those undergoing coronary stenting. However, it was still unknown whether patients receiving abciximab outside of clinical trial settings were experiencing similar outcomes to those in the trials. Since that time, much more information supporting the effectiveness of abciximab has been published. Therefore, the analyses described herein should serve more as a historical example of how these methods might have been used rather than an attempt to quantify the cost-effectiveness of abciximab in contemporary practice.

METHODS

Construction of Comparison Cohort

In nonrandomized study designs, differences between treated and nontreated patients are likely to produce biased estimates of treatment effects and cost differences. To minimize these biases in an analysis of the UMMS data, we used a propensity score methodology to identify patients not treated with abciximab who were similar to abciximab-treated patients on a wide range of observed covariates.[10] Matching using propensity scores has been shown to reduce bias to a greater extent than other matching techniques when a large number of covariates are considered.[11] Further, estimated propensity scores are superior to true propensity scores at eliminating bias because the method also removes biases that occur because of chance.[11–14] We prefer this matching technique over multivariable regression because it allows one to determine whether there is sufficient overlap between the treatment groups in terms of covariates associated with treatment.[15] If the 2 cohorts are dissimilar, adjusting for differences in a multivariable framework is likely to provide biased results because computations are based on linear (or other higher order) extrapolations of the covariates.

We estimated propensity scores using logistic regression to predict treatment with abciximab ($A = 1$). Independent variables consisted of patients' demographic characteristics (gender, age [< or ≥ 65 years]), previous procedures (bypass surgery, percutaneous transluminal coronary angioplasty [PTCA], stent), cardiovascular risk factors (diabetes mellitus, hyperlipidemia, hypertension, and congestive heart failure), procedure-related variables (indication for procedure,

PTCA v. stenting), angiographic characteristics (American College of Cardiology and American Heart Association morphology scores,[16] multivessel disease, presence of intracoronary thrombus), and level of care (intensive care unit or cardiac care unit). We used the propensity score matching technique described by Rosenbaum and Rubin[17] to construct the untreated cohort. For each abciximab-treated patient, a patient who was not treated with abciximab with the closest (higher or lower) propensity score was chosen for inclusion in the comparison cohort. The propensity score for an individual patient (i = 1, 2, 3, . . . n) was defined as the conditional probability of being treated with abciximab ($A_I = 1$), given a patient's vector of observed covariates, x_I:

$$e(x_i) = pr(A_i = 1 \mid X_i = x_i),$$

where the treatment (A_I) and the covariates (X_i) are assumed to be independent:

$$pr(A_1 = a_1, a_2, a_3, \ldots, A_N | X_1 = x_1, x_2, x_3, \ldots, X_N) = \prod_{i=1}^{N} e(x_i)^{a_i} \{1 - e(x_i)\}^{1-a_i}.$$

To assess the capacity of the propensity score methodology in constructing 2 comparable cohorts, descriptive statistics were computed, and covariates were compared using chi-square tests for dichotomous variables and t tests for continuous variables.

Estimation of Treatment Effects

The results of 3 large RCTs of abciximab have been reported. The EPIC trial was the 1st large-scale RCT to assess the efficacy of abciximab in high-risk patients undergoing balloon angioplasty (PTCA) or atherectomy.[9] This study was followed by a 2nd large RCT, EPILOG, to determine if abciximab would also be efficacious in a lower risk population.[18] Both studies used a similar primary endpoint: a composite endpoint of death, nonfatal MI, and the need for urgent revascularization at 30 days and 6 months of followup. Both studies demonstrated a beneficial effect of abciximab (35% and 54% relative reductions in the composite endpoint). The 3rd large trial was undertaken to assess the efficacy of abciximab in a broad range of patients undergoing coronary stenting.[19] Despite previous trials showing an approximate 20% to 30% relative reduction in ischemic events with stenting in comparison to balloon angioplasty,[20,21] the addition of abciximab to stenting was superior to both stenting alone and angioplasty plus abciximab (51% and 23% relative reductions in the composite endpoint).

We examined the effect of abciximab on the incidence of death and nonfatal MI at 6 months of followup. Although event rates would have been greater, the effect of abciximab on the rate of revascularization procedures was not considered as a health outcome to avoid double counting in the ICER and because there is insufficient evidence to model how the need for a subsequent revascularization procedure affects life expectancy. The effect of reduced revascularization rates was accounted for in the calculation of the numerator (total costs at 6 months).

Bayesian Model

The Bayesian paradigm is well suited to informing decision makers about the cost-effectiveness of medical therapies. It can be designed to incorporate multiple sources of information to predict the probability that a medical therapy is cost-effective. Our approach to integrate data from the clinical trials with our local data was to employ a Bayesian, hierarchical, random-effects model. A similar approach has been advocated for conducting meta-analyses for clinical trials.[22,23] We chose to use a random-effects specification because it allows one to model a probability for the outcome of each individual data source whereby the joint distribution is independent of the order in which the data sources are analyzed. The random-effects approach is based on 2 assumptions. First, the clinical trial data and the local data are exchangeable.[24,25] This simply means that it is not possible to know beforehand which trials (or local data) will have high or low event rates. The 2nd assumption is that there is a distribution that represents all possible treatment effects across various settings.[24] Thus, we assumed that event rates for death and nonfatal MI (within procedures) are similar across RCTs and at UMMS. This assumption can be considered as a middle ground between the extreme assumptions of "equivalence" and "complete independence."[25] The model is considered hierarchical in that there are 3 levels of statistical relationships being modeled: the individual patient, the individual study, and the population of possible studies. Specifically, the hierarchical model assumes that patients in each study are exchangeable in that they all are assumed to have the same probability of having an event (without other information). At the level of the individual studies, each study has certain characteristics that make it distinct from the other studies. Each study can be thought of as having its own distribution of possible results, but only 1 draw from that distribution is observed. Finally,

there is a distribution of results from all possible trials of the drug that we are attempting to model.

The hierarchical model was used to estimate posterior probabilities for death and MI at 6 months for 4 patient subgroups: patients undergoing PTCA (or atherectomy) with or without abciximab and patients undergoing stenting with or without abciximab. All Bayesian analyses were conducted using WinBUGS,[26] a publicly available Windows-based software package. The code for the base case model is included in the Appendix.

The number of events (death or MI), r_i, for the ith study for each of I studies is modeled as a binomial variable:

$$r_i \sim \text{binomial}(p_i, n_i),$$

where n_i is the number of patients in the ith study, and p_i is the probability of a death or MI in the ith study. The event probability p_i is modeled as

$$\text{logit}(p_i) \sim \text{normal}(\text{logit}[\theta], \sigma^2),$$

where θ is the probability of an adverse event (death or nonfatal MI) among all possible trials, and σ^2 is the variance in event probabilities between different trials on an inverse logit scale.

For this hierarchical model, the parameters (logit[θ] and σ^2) used to specify the distribution of local event probabilities (p_i) also need to be assigned distributions, referred to as hyperpriors. Hyperprior distributions, like all prior distributions, can be assigned to be very flexible or noninformative, so that even very few data dominate the posterior distribution. Because the number of studies in this example was small, the wide range of values allowed by noninformative hyperpriors for logit(θ) and for σ^2 led to unrealistic posterior distributions that included event probabilities up to 100%. To minimize the impact of these extreme estimates on the results, the hyperprior distributions for logit(θ) and σ^2 were assigned values that still allowed the posterior distribution to take on a wide range of event probabilities, but not so high that they were considered unrealistic. Event probabilities as high as 15% for death and 31% for nonfatal MI were assumed to be very unlikely and were chosen as conservative upper limits in the base case. No lower limits were assigned. Additionally, σ^2 was modeled as $1/\sigma^2 \sim \gamma(2, 2)$, which still allows for a reasonably wide range in variance from trial to trial while restricting the probability of extreme variance, which would not have been realistic. This eliminated the possibility of very large differences in event probabilities between clinical trials, an assumption impossible to test without more trials. However, our limited data suggest this to be a reasonable assumption.

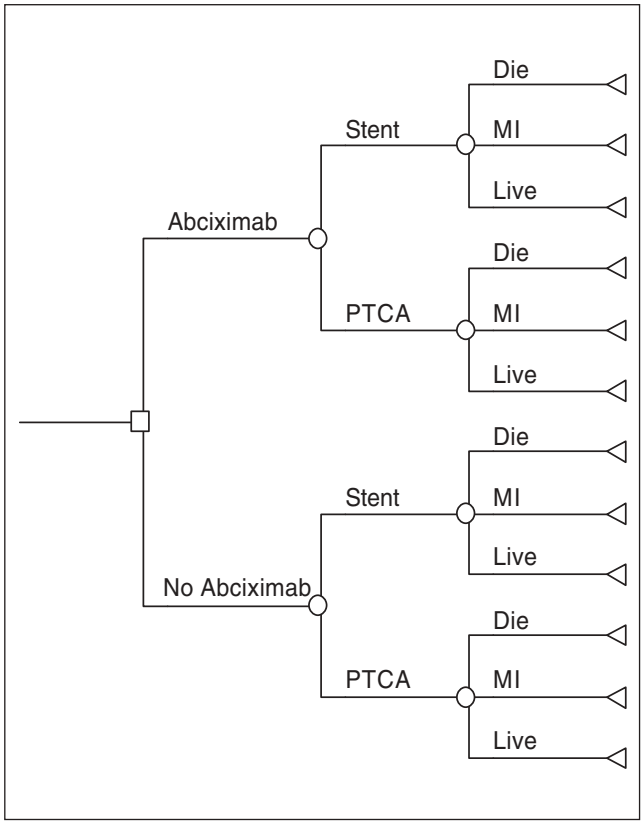

Figure 1 Decision tree.
Note: MI = myocardial infarction; PTCA = percutaneous transluminal coronary angioplasty.

Cost-Effectiveness Model

A decision-analytic model was constructed in Microsoft Excel to estimate the cost per life year saved (LYS) among patients treated with abciximab. The decision model is shown in Figure 1. The Bayesian random-effects model was used to compute posterior distributions for the incidence of events in each treatment and procedure group. Means, medians, and 95% credible intervals for each of the posterior distributions were computed on the basis of 2000 iterations after a "burn-in" of 5000 iterations. Excluding iterations from the burn-in period is necessary to avoid biased results that would occur from using iterations from the Markov chain Monte Carlo (MCMC) sampling algorithm before it had converged to the equilibrium distribution. The posterior distributions consisting of 2000 iterations

were integrated into the final branches of the decision tree for subsequent Monte Carlo simulation.

From a clinical perspective, some patients have coronary lesions that are suitable only for coronary angioplasty, and some patients have coronary lesions that are suitable for either angioplasty or stenting.[27] Thus, 4 comparisons of various treatment combinations were relevant:

- abciximab v. no abciximab (with a 50:50 split between PTCA and stenting),
- abciximab and angioplasty v. angioplasty alone,
- abciximab and stenting v. stenting alone, and
- abciximab and angioplasty v. stenting alone.

We assumed that treatment with abciximab or stenting would not incrementally affect life expectancy beyond the 1st 6 months of followup. Therefore, analogous estimates of life expectancy for each outcome were assigned for each treatment group. Patients who were event free during the 6-month followup period were assumed to have an average, undiscounted life expectancy of 20 years, and those who experienced nonfatal MIs were assumed to have an average, undiscounted life expectancy of 15 years.[28] For the final analyses, life expectancies were discounted at 5% per year.

Cost estimates were based on cumulative 6-month charge data collected at UMMS for patients treated with and without abciximab. Charges were converted to costs using cost-to-charge ratios and were converted to 1998 US dollars.[8] We assumed that costs after the 6-month followup period would be equivalent for each of the treatments. Therefore, we simply applied local, mean 6-month costs to each of the outcomes in the decision tree. We used Monte Carlo simulation to estimate the expected values for each treatment option. Distributions for the resultant ICERs were used to estimate the overall, joint uncertainty in the model. One thousand simulations were run for each of the 4 treatment comparisons. For the base case model, we constructed a scatterplot of the differences in costs and LYS to visually display variability in the results. We also report the proportion of ICERs estimated at or below $50,000/LYS and present cost-acceptability curves to display the results.

Although coronary stenting is not an option for all patients undergoing percutaneous revascularization, the large majority are candidates for either PTCA or stenting. Treating the 4 treatment strategies as mutually exclusive allowed us to choose the most cost-effective strategy across a range of threshold values that indicate the maximum amount of money a decision maker would pay to gain a statistical life year. To do this, we computed average net benefits for each strategy[29] for each simulation across a range of threshold values. For each simulation at each threshold value, 1 strategy is optimal (greatest net benefits). We plotted the proportion of times each strategy was considered optimal. The resulting curve is intuitively the same as a cost-acceptability curve.

Sensitivity Analysis

Because we were starting from the perspective of conducting an institution-specific CEA and we were interested in the relative value of considering additional data sources, we implemented the random-effects model beginning with data only from UMMS. Then, we consecutively added information provided by the trials in order of publication. For clarity, we report only the results of trials involving PTCA with or without abciximab. After each incremental addition to the model, we calculated 95% credible intervals and constructed cost-acceptability curves.

We also explored the effect of different specifications for the prior distribution using sensitivity analysis. First, we ran all the models without an upper limit for the event rate so that the MCMC procedure could sample from all points corresponding to 0% to 100% event probabilities. Then, we halved the upper limit of the probability of death (7.6%) and nonfatal MI (16.1%) to examine the effects of using a less conservative upper limit for the area of the distribution from which the MCMC procedure in WinBUGS could sample. Additional sensitivity analyses were undertaken to explore the effect of varying the precision of hyperpriors for the event probability θ and other parameters necessary to specify the model, such as the shape and scale of the γ distribution that contributes to the amount of variance generated by the model.

Finally, patients in the clinical trials were routinely monitored for increases in CPK-MB levels to detect non–Q wave MIs, whereas MIs in the UMMS cohort were dependent on physician reports of MIs documented in patients' medical record. Because patients in routine practice may have been less rigorously monitored for non–Q wave MIs than patients in clinical trials, we believed that we may have underestimated the incidence of MIs. If the MIs reported in patients' medical records primarily reflected Q wave MIs, we believed that we may have underestimated the number of MIs in the UMMS cohort by about 400%, given the ratio of Q wave MIs and non–Q wave MIs reported in the clinical trials. Therefore, we inflated the incidence of

Table 1 Comparison of Abciximab-Treated and Abciximab-Untreated Cohorts at the University of Maryland Medical System

Characteristic	Abciximab-Treated Cohort (n = 100)	Comparison Cohort (n = 100)	p Value[a]
Mean age (y)	57.2	57.6	0.803[b]
Gender (% male)	72.0	73.0	0.874
Risk factors (%)			
Congestive heart failure	9.0	9.0	1.000
Diabetes mellitus	16.0	14.0	0.692
Hyperlipidemia	55.0	58.0	0.669
Hypertension	48.0	54.0	0.396
Indication (%)			
Acute MI	12.0	13.0	0.831
Other indication	3.0	3.0	1.000[c]
Post-MI angina	51.0	45.0	0.396
Stable angina	2.0	1.0	1.000[c]
Unstable angina	31.0	40.0	0.184
History of previous (%)			
CABG	9.0	8.0	0.800
PTCA/stent	10.0	13.0	0.506
Intracoronary thrombus (%)	38.0	35.0	0.659
ACC/AHA morphology[d] (%)			
A	12.0	10.0	0.651
B1	24.0	27.0	0.626
B2	17.0	19.0	0.713
C	47.0	44.0	0.670
Coronary stenting (%)	50.0	51.0	0.888
Admitted to ICU/CCU (%)	9.0	5.0	0.326

Note: MI = myocardial infarction; CABG = coronary artery bypass graft; PTCA = percutaneous transluminal coronary angioplasty; ACC = American College of Cardiology; AHA = American Heart Association; ICU = intensive care unit; CCU = cardiac care unit.

a. p value based on Pearson chi-square.

b. p value based on Student's t test.

c. p value based on Fisher's exact test.

d. Scoring system based on morphologic characteristics of the coronary lesion used to predict ischemic events. "A" indicates a low-risk lesion, and "C" indicates a high-risk lesion.

MIs at UMMS 4-fold to examine the impact on the ICERs.

RESULTS

The 100 abciximab-treated patients did not statistically differ with regard to any of the observed covariates compared to the 100 matched patients not treated with abciximab (Table 1). In Table 2, patients treated at UMMS are stratified by type of percutaneous revascularization procedure to facilitate comparisons with patients enrolled in the RCTs. Patients treated at UMMS were similar to patients treated in the RCTs in terms of gender, age, hyperlipidemia, and hypertension (Table 2). However, patients at UMMS were slightly less likely to have diabetes. Also, patients who underwent PTCA at UMMS were less likely to have had previous bypass surgery, reflecting more current practice patterns in which patients with previous coronary artery bypass grafts are more likely to undergo coronary stenting.[30]

Descriptive statistics for the posterior distributions for death and nonfatal MI at 6 months of followup are shown in Tables 3 and 4 for each of the RCTs and UMMS. The means of the posterior distributions were greater than the medians in all cases, indicating a right-skewed distribution. The skewness in the posterior distributions was in part due to random MCMC sampling near the upper limit of θ for the event probabilities specified in the base case model. The credible intervals for all the distributions were relatively wide.

On the basis of the Monte Carlo simulations of the decision model, patients treated with stenting plus abciximab had the highest estimated life expectancy, 13.88 discounted life years, and the highest estimated costs, $15,379 (Table 5). Angioplasty alone was shown to have the lowest estimated life expectancy, 13.33 discounted life years. However, the stent-only option was the least costly, $11,018, on average. On the basis of point estimates, the angioplasty-only option was dominated by stent-only, being less effective and more costly.

The marginal gain in life years was 0.26 (3.1 months) with the addition of abciximab to percutaneous intervention (assuming a 50:50 split between PTCA and stenting) in comparison to percutaneous intervention without abciximab. A scatterplot displaying the joint distribution of incremental costs and effectiveness is shown in Figure 2. The ICER was $13,475/LYS. In 83.4% of the simulations, the addition of abciximab was more effective than percutaneous revascularization without the drug, and in 74% of the simulations, the ICER was estimated at or below $50,000/LYS (Fig. 3). Among patients exclusively undergoing PTCA and coronary stenting, the marginal gains in life years were about 3.7 months and 3.0 months, respectively, with the addition of abciximab. These gains translated to ICERs of approximately $8,000/LYS with PTCA and $17,500/LYS with stenting. When comparing PTCA plus abciximab to PTCA alone, 74.1% of the ICERs generated were $50,000/LYS or less. When comparing coronary stenting with abciximab with stenting alone,

Table 2 Characteristics of Patients from Randomized Clinical Trials and Local Institution

	Study				
	PTCA			Stent	
Characteristic	EPIC (*N* = 2099)	EPILOG (*N* = 2792)	UMMS PTCA (*n* = 100)	EPISTENT (*N* = 2399)	UMMS Stent (*n* = 100)
Male (%)	72	72	74	75	71
Average age (y)	61	60	55	59	60
Risk factors (%)					
Diabetes	24	23	14	20	16
Hypertension	55	NR	49	53	53
Hyperlipidemia	57	NR	58	NR	55
Previous CABG (%)	15	13	4	9	13
Procedures (%)					
PTCA	95	95[a]	87	98[b]	0
Atherectomy	5	6	9	0	0
Rotablator	0	0	4	0	0
Stent	0	14	0	100[c]	100
Previous MI	57	NR	58	51	62

Note: PTCA = percutaneous transluminal coronary angioplasty; UMMS = University of Maryland Medical System; NR = not reported; CABG = coronary artery bypass graft; MI = myocardial infarction.
a. Not mutually exclusive.
b. In the PTCA arm of the trial, 2.3% crossed over to stent.
c. In the stent arm of the trial.

Table 3 Base Case: Incidence of Death at 6 Months

PTCA and abciximab				PTCA only		
EPIC (*n* = 708) 3.11%	EPILOG (*N* = 935) 1.07%	EPISTENT (*n* = 796) 1.76%	UMMS (*n* = 50) 2.00%	EPIC (*n* = 696) 3.45%	EPILOG (*n* = 939) 1.70%	UMMS (*n* = 49) 4.08%
Posterior distribution Mean = 1.98% (95% CI = 0.57%–5.00%) Median = 1.72%				Posterior distribution Mean = 3.18% (95% CI = 0.67%–9.19%) Median= 2.63%		

Stent and abciximab		Stent only	
EPISTENT (*n* = 794) 0.50%	UMMS (*n* = 50) 0%	EPISTENT (*n* = 809) 1.24%	UMMS (*n* = 51) 1.96%
Posterior distribution Mean = 0.60% (95% CI = 0.02%–2.99%) Median = 0.32%		Posterior distribution Mean = 1.79% (95% CI = 0.20%–7.18%) median = 1.24%	

Note: PTCA = percutaneous transluminal coronary angioplasty; UMMS = University of Maryland Medical System; CI = confidence interval.

66.0% of the simulations were estimated at or below $50,000/LYS. The other relevant analysis was the addition of abciximab to PTCA in comparison to coronary stenting. The incremental difference in effectiveness between the strategies was very small (0.005 LYS), resulting in a large ICER ($604,000/LYS) and a relatively low percentage of simulations (40.0%) in which PTCA

Table 4 Base Case: Incidence of MI at 6 Months

PTCA and abciximab				PTCA only		
EPIC	EPILOG	EPISTENT	UMMS	EPIC	EPILOG	UMMS
(*n* = 708)	(*n* = 935)	(*n* = 796)	(*n* = 50)	(*n* = 696)	(*n* = 939)	(*n* = 49
6.90%	5.03%	6.53%	4.00%	10.49%	9.90%	2.04%
Posterior distribution				Posterior distribution		
Mean = 6.08% (95% CI = 2.06%–3.79%)				mean = 8.99% (95% CI = 2.16%–21.78%)		
Median = 5.47%				Median= 7.80%		

Stent and abciximab		Stent only	
EPISTENT	UMMS	EPISTENT	UMMS
(*n* = 794)	(*n* = 50)	(*n* = 809)	(*n* = 51)
5.16%	2.00%	10.26%	1.96%
Posterior distribution		Posterior distribution	
Mean = 4.53% (95% CI = 0.42%–16.89%)		Mean = 6.76% (95% C I= 5.30%–21.76%)	
Median = 3.35%		Median= 5.30%	

Note: MI = myocardial infarction; PTCA = percutaneous transluminal coronary angioplasty; UMMS = University of Maryland Medical System; CI = confidence interval.

Table 5 Baseline Estimates

Treatment/ Procedure Groups	Average Cost[a]	Discounted Life Years
Abciximab and stent	$15,379	13.88
Abciximab and PTCA	$13,898	13.64
Stent only	$11,018	13.63
PTCA only	$11,432	13.33
Abciximab with 50:50	$14,639	13.76
No abciximab with 50:50	$11,145	13.50

PTCA = percutaneous transluminal coronary angioplasty.
a. In 1998 US dollars.

plus angioplasty compared to stenting would be considered cost-effective at a $50,000/LYS threshold.

The results of these models also allow for the consideration of which of the 4 strategies is optimal for patients who are candidates for either PTCA or stenting. When examining the 4 revascularization strategies, the PTCA option was dominated on the basis of point estimates because it was less effective and more costly than stenting. Using the net benefits approach to examine each strategy separately (not incrementally) revealed that at threshold values up to $28,000/LYS, stent only was considered to be the optimal strategy (Fig. 4). Beyond that point, stent plus abciximab was the optimal strategy in the majority of simulations, followed by stent only. The strategies involving PTCA were optimal in < 10% of simulations across all threshold values.

Sensitivity Analysis

Increasing the number of data sources included in the random-effects model led to increased precision in the 6-month event probabilities. When using only data from patients who underwent angioplasty plus abciximab at UMMS, the 95% credible interval for the incidence of MI ranged from 0.1% to 22.5%, an absolute difference of 22.4 percentage points (Fig. 5). With the addition of data from the EPIC trial, the credible interval decreased to a range of 19.7 percentage points. The width of the interval further decreased with the addition of data from EPILOG to 14.9 percentage points, and with EPISTENT, the interval decreased to 11.7 percentage points. The increased precision in event rates translated to an increase in the number if simulations with PTCA plus abciximab were more cost-effective relative to PTCA alone. At the $50,000/LYS threshold, the proportion of simulations that showed that PTCA plus abciximab was cost-effective was 59.2% when using data from UMMS. With the addition of data from the 3 clinical trials, this proportion rose to 69.6%.

Sensitivity analyses revealed that the posterior distributions were sensitive to changes in the upper limit for the event probability (θ), particularly for the incidence of MI. When we did not specify an upper limit for θ, the estimated probability of events increased for all treatment strategies, and the credible intervals were wider than in the base case (Table 6). Conversely, when halving the upper limit of θ, the estimated probability of events decreased for all treatment strategies, and the

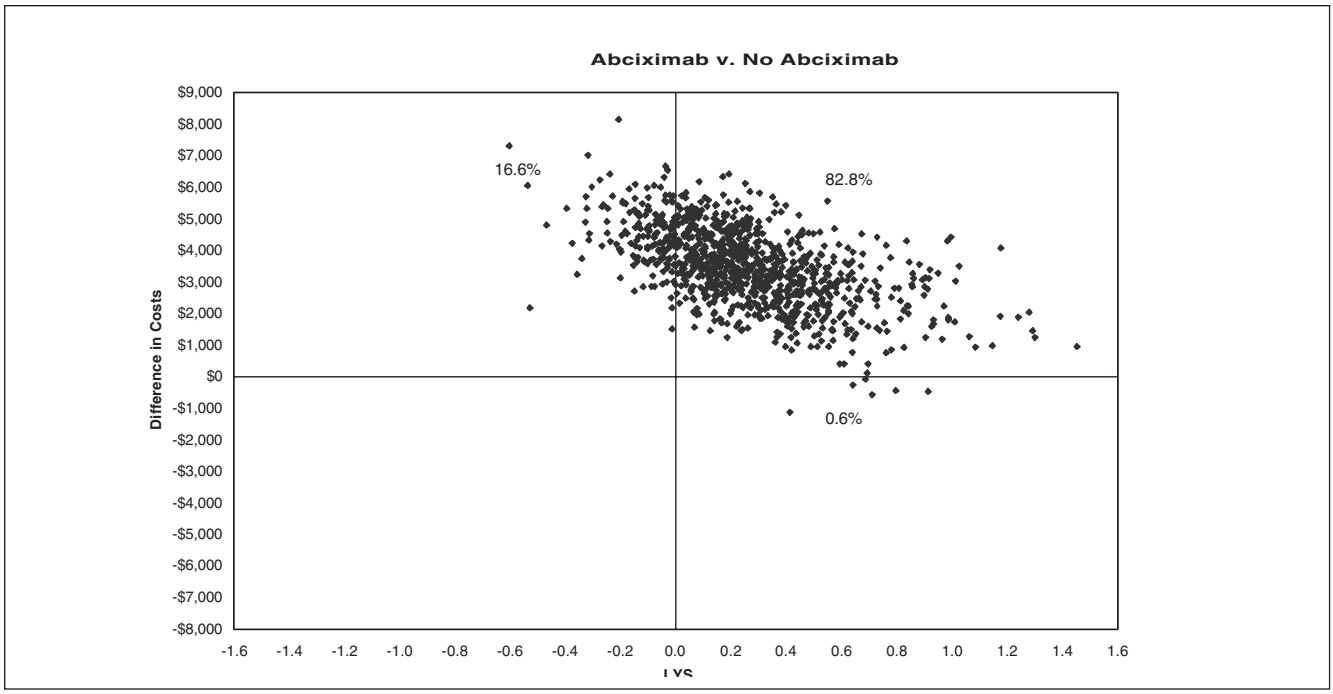

Figure 2 Scatterplot of simulations on cost-effectiveness plane.
Note: LYS = life year saved.

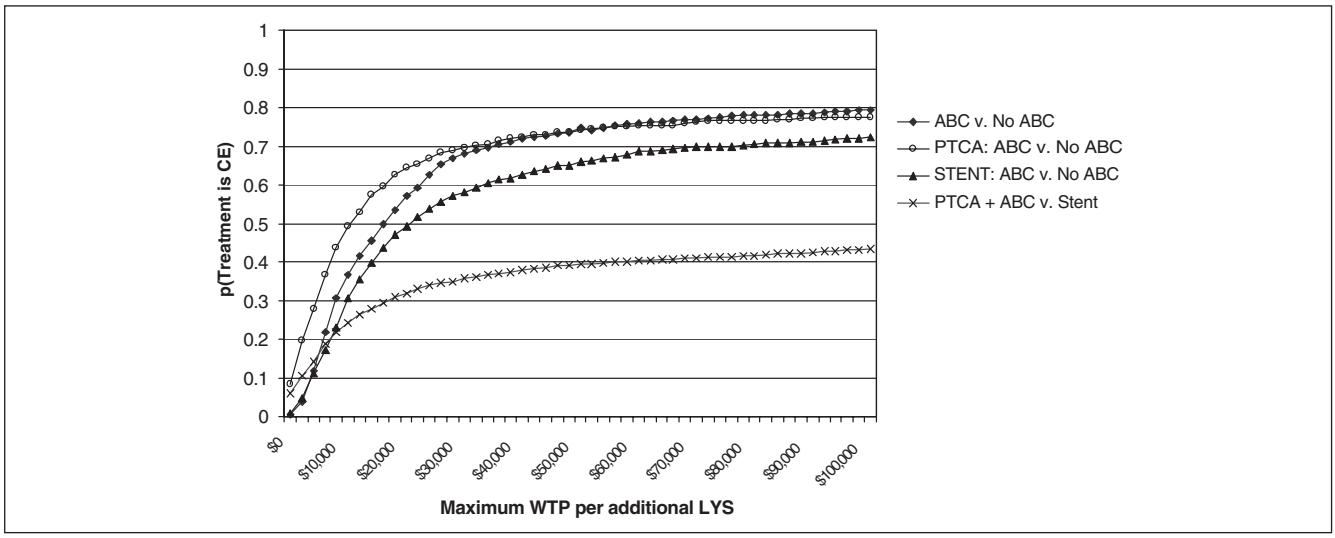

Figure 3 Base case: cost-acceptability curves.
Note: ABC = abciximab; CE = cost-effective; LYS = life year saved; PTCA = percutaneous transluminal coronary angioplasty; WTP = willingness to pay.

credible intervals were more narrow than in the base case. Varying the specifications for the precision of θ and the shape and scale of $1/\sigma^2$ had only negligible effects on the results. Further sensitivity analyses revealed that artificially increasing the incidence of MI at UMMS to account for potential differences in reporting of MIs increased the estimated posterior average by approximately 1 to 3 percentage points (Table 6).

Although altering model parameters and event rates at UMMS led to changes in the posterior distributions, these adjustments had almost no effect on the relative cost-effectiveness of the treatment strategies, because the parameter specifications and data assumptions were consistently applied across treatment strategies. The proportion of simulations in which the cost-effectiveness ratio was < $50,000/LYS was 73.7% when

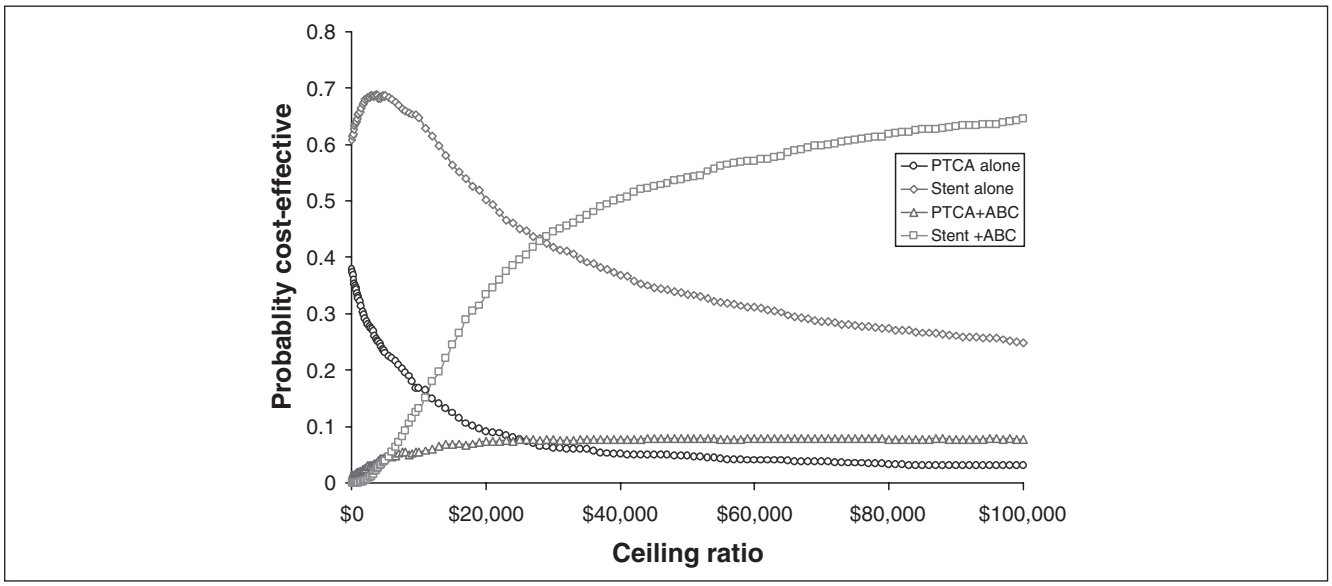

Figure 4 Proportion of simulations with greatest net benefit across all strategies.
ABC = abciximab; PTCA = percutaneous transluminal coronary angioplasty.

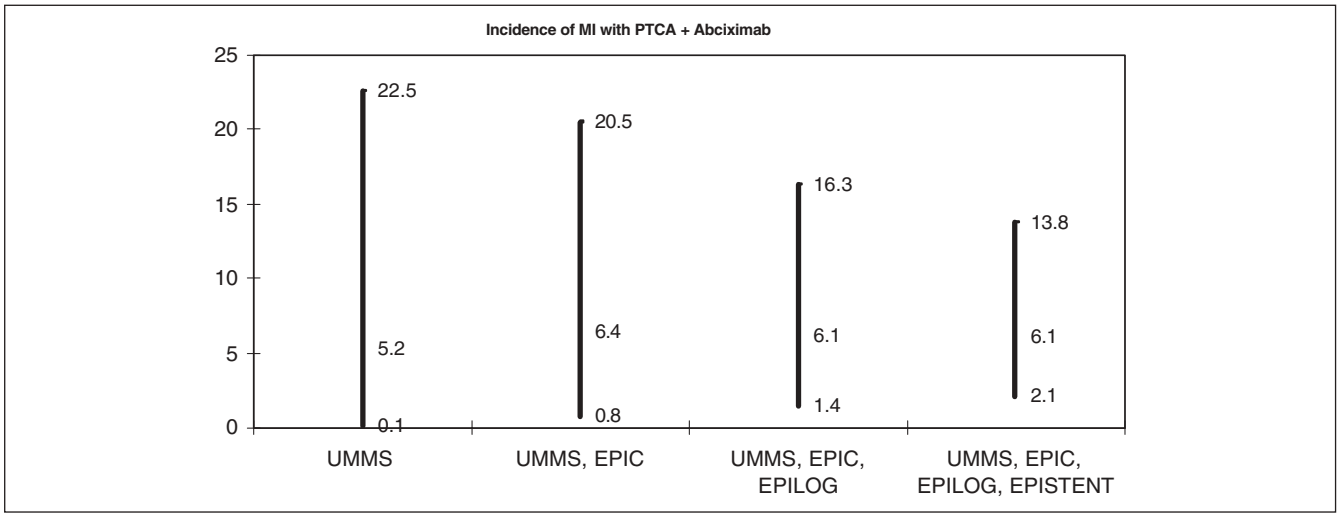

Figure 5 Increasing precision of event rates with additional data.
Note: MI = myocardial infarction; PTCA = percutaneous transluminal coronary angioplasty; UMMS = University of Maryland Medical System.

no upper limit for the event probabilities (θ) were specified, 71.5% when the upper limit was halved, and 72.8% when the number of MIs at UMMS was quadrupled.

DISCUSSION

We have shown how the use of propensity scores can be coupled with a Bayesian statistical approach to address 2 common limitations of site-specific CEAs: a lack of randomization and an insufficient sample size. To accomplish this, we used a Bayesian random-effects model to incorporate site-specific data with clinical trial results. We then applied the resulting posterior distributions with local cost data in a traditional decision-analytic model. The approach provides a transparent means of integrating data from multiple sources and produced more precise estimates of cost-effectiveness, thereby providing locally relevant information to formulary decision makers.

Table 6 Sensitivity Analysis: Incidence of Myocardial Infarction (MI)

Treatment Strategy	Baseline	Half Upper Limit	No Upper Limit	Four Times Rate of MI at UMMS
PTCA only	8.99%, median = 7.80% (2.16%–21.78%)	7.96%, median = 7.63% (2.13%–14.84%)	9.33%, median = 7.98% (2.11-24.21%)	10.28%, median = 9.26% (2.95-22.45%)
Stent only	6.76%, median = 5.30% (0.53%–21.76%)	5.59%, median = 4.81% (0.56%–14.14%)	7.13%, median = 5.34% (0.59%–25.83%)	9.8%, median = 8.48% (1.49%–25.47%)
PTCA and abciximab	6.08%, median = 5.47% (2.06%–13.79%)	5.92%, median = 5.45% (2.23%–12.42%)	6.11%, median = 5.50% (2.06%–13.84%)	8.08%, median = 7.31% (3.01%–18.01%)
Stent and abciximab	4.53%, median = 3.35% (0.42%–16.89%)	4.12%, median = 3.33% (0.08%–12.15%)	4.85%, median = 3.42% (0.43%–18.64%)	7.19%, median = 5.82% (1.10%–21.80%)

UMMS = University of Maryland Medical System; PTCA = percutaneous transluminal coronary angioplasty.

On the basis of our model, our results indicate that the addition of abciximab is cost-effective at UMMS in 74% of the simulated runs, assuming a maximum willingness to pay of $50,000/LYS and that approximately 50% of patients undergo coronary stenting and 50% undergo angioplasty. Sensitivity analyses revealed that this estimate was robust to changes in the specifications and assumptions used in the Bayesian model. Nevertheless, a wide range of cost-effectiveness ratios were consistent with the data because there was insufficient evidence to conclude that abciximab-treated patients have a longer life expectancy with a high degree of certainty.

Incorporating evidence from large RCTs decreased uncertainty in the estimates of cost-effectiveness relative to the site-specific economic evaluation. In the base case analysis, the proportion of simulations in which abciximab was shown to be cost-effective at the $50,000/LYS threshold was just 74%. Even if one were willing to spend any amount of money for a gain in life expectancy, this number would increase to only 83.4%, corresponding with the point at which the threshold value on the cost-acceptability curve approaches infinity. However, from a Bayesian decision theoretical perspective, arbitrary cutoff values for decision making (such as the commonly employed 5% level of significance) are irrelevant to decision makers who must make a choice between 1 intervention and another.[31] Instead, information on uncertainty should be combined with information about a decision maker's loss function, which should include the implications of making a wrong decision given the number of potential patients affected and the total number of dollars involved. Uncertainty then enters the equation only in the extent to which reducing uncertainty adds value to the decision maker.[32]

Our study has several implications for improving the usefulness of information provided to "real-world" decision makers. First, by considering the combination of data from clinical trials with local data, it compels one to directly compare the characteristics of patients from both sources and to explicitly address observed differences in those characteristics. In cases in which only minor differences are found, such as in our example, it may be appropriate to integrate data from both sources to obtain more precise estimates of cost-effectiveness. Another benefit of this exercise that may occur when minor differences are found is that it can quell local concerns about the generalizability of the trial data. However, care must be employed when inferring similarities in treatment effects from similarities in observed covariates. Although no statistically significant differences were found in the observed characteristics between the UMMS data and that from the RCTs (see Table 1), Altman[33] emphasized that even nonsignificant differences in prognostic factors can lead to influences on the observed outcome.

The approach outlined here can be adapted to assess the likely importance of collecting local data in the first place, given the extent and availability of randomized evidence. First, it can allow one to simulate to what degree patient outcomes would have to differ between local patients and clinical trial patients to have an impact on the results. Second, our approach could also be used to determine how many patients would be needed in a local evaluation to exert an influence on the clinical trial results. Although our application of these methods was limited in that it involved a treatment with a low incidence of clinical events and one in which large numbers of patients were enrolled in clinical trials, the methods we describe are generalizable to other clinical situations in which the incidence of events may differ

more substantially between patient groups and in which there may be fewer patients enrolled in clinical trials.

Others have also explored the use of a Bayesian paradigm to increase the generalizability of CEAs. Rittenhouse[34,35] suggested using a Bayesian approach to account for biases in patient selection in RCTs and other factors such as the detection of subclinical disease identified through protocol-induced testing. Although our analysis shares the same conceptual basis, ours differs in that it is the first to use empirical data from a real-world setting to demonstrate how one can improve the usefulness of site-specific evaluations by incorporating clinical trial data when appropriate.

Making decisions in a health care system early in the life cycle of a new technology is challenging. Current approaches to performing cost-effectiveness studies often do not provide decision makers with the information or flexibility they may require. Our study suggests that a Bayesian approach can provide useful information for decision makers.

APPENDIX
Code for WinBUGS Model

```
model
{
   # For each of N trials, model the following:
   for( i in 1 : N ) {
      # b[i] = logit(p[i]), mu, tau defined below
      b[i] ~ dnorm(mu,tau)

      # p[i] = pr(death in ith trial)
      logit(p[i]) <- b[i]

      # r[i] = No. of deaths, n[i] = No. of patients in ith trial
      r[i] ~ dbin(p[i],n[i])
}

   # Hyperpriors:
   # mu = logit(θ), taken from a flat, wide normal distribution that cannot take on
   # values that equate to a greater than 15% risk of death.
   mu ~ dnorm(-3.7,1.0E-5)I(,-1.7)

   # tau = 1 / σ^2; this distribution prevents extremely large values for
   # σ^2, while still allowing for a fairly wide range of realistic values.
   tau ~ dgamma(2,2)

   # Logical nodes of interest: Population mean probability of death theta and
   # population standard deviation (on an inverse logit scale) sigma
   theta <- exp(mu) / (1 + exp(mu))
   sigma <- 1 / sqrt(tau)
}
```

REFERENCES

1. Powe NR, Griffiths RI. The clinical-economic trial: promise, problems, and challenges. Control Clin Trials. 1995;16:377–94.
2. Schulman KA, Glick H, Buxton M, et al. The economic evaluation of the FIRST study: design of a prospective analysis alongside a multinational phase III clinical trial. Control Clin Trial. 1996;17:304–15.
3. O'Brien BJ, Drummond MF, LaBelle RJ, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. Med Care. 1994;32:150–63.
4. Frieman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, type II error and sample size in the randomized control trial: survey of 71 "negative" trials. N Engl J Med. 1978;299:690–4.

5. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios. Health Econ. 1997;6:229–42.
6. Cook JR, Heyse JF. Use of an angular transformation for ratio estimation in cost-effectiveness analysis. Statist Med. 2000;19:2989–3003.
7. Heitjan DF, Moskowitz AJ, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. Health Econ. 1999;8:191–201.
8. Reed SD, Mullins CD, Magder LS. Cost-effectiveness of abciximab during routine medical practice. Pharmacoeconomics. 2000;18:265–74.
9. EPIC Investigators. Use of a monoclonal antibody directed against the platelet glycoprotein IIb/IIIa receptor in high-risk coronary angioplasty. N Engl J Med. 1994;330:956–61.
10. Rosenbaum PR, Rubin DB. The central role of propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
11. Gu XS, Rosenbaum PR. Comparison of multivariate matching methods: structures, distances and algorithims. J Comput Graph Stat. 1993;2:405–20.
12. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc. 1984;79:516–24.
13. Rosenbaum PR. Conditional permutation tests and the propensity score in observational studies. J Am Stat Assoc. 1984;79:565–74.
14. Robins JM, Mark SD, Newey WK. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics. 1992;48:479–95.
15. Little RJ, Rubin DB. Causal effects in clinical and epidemiologic studies via potential outcomes: concepts and analytical approaches. Ann Rev Public Health. 2000;21:121–45.
16. Ellis SG, Vandormael MG, Cowley MJ, et al. Coronary morphologic and clinical determinants of procedural outcome with angioplasty for multivessel coronary disease: implications for patient selection. Circulation. 1990;82:1193–1202.
17. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am Statistician. 1985;39:33–8.
18. EPILOG Investigators. Platelet glycoprotein IIb/IIIa receptor blockade and low-dose heparin during percutaneous coronary revascularization. N Engl J Med. 1997;336:1689–96.
19. EPISTENT Investigators. Randomised placebo-controlled and balloon-angioplasty-controlled trial to assess safety of coronary stenting with use of platelet glycoprotein-IIb/IIIa blockade. Lancet. 1998;352:87–92.
20. Serruys PW, De Jaegere P, Kiemeneij F, et al. A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. N Engl J Med. 1994;331:489–95.
21. Fischman DL, Leon MB, Baim DS, et al. A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. N Engl J Med. 1994;331:496–501.
22. Stangl DK, Berry DA. Meta-analysis: past and present challenges. In: Stangl DK, Berry DA, editors. Meta-Analysis in Medicine and Health Policy. New York: Marcel Dekker; 2000. p. 1–28.
23. Brophy JM, Lawrence J, Rouleau JL. β-blockers in congestive heart failure: a Bayesian meta-analysis. Ann Intern Med. 2001;134:550–60.
24. Gelman A, Carlin JB, Stern H, Rubin DB. Bayesian Data Analysis. New York: Chapman and Hall; 1995.
25. Smith TC, Spiegelhalter DJ, Parmar MKB. Bayesian meta-analysis of randomized trials using graphical models and BUGS. In: Berry DA, Stangl DK, editors. Bayesian Biostatistics. New York: Marcel Dekker; 2000. p. 411–27.
26. Spiegelhalter DJ, Thomas A, Best NG, Gilks WR. WinBUGS: Bayesian Inference Using Gibbs Sampling, Version 1.2. Cambridge, UK: MRC Biostatistics Unit; 1999.
27. Eeckhout E, Wijns W, Meier B, Goy JJ. Indications for intracoronary stent placement: the European view. Working Group on Coronary Circulation of the European Society of Cardiology. Eur Heart J. 1999;20:1014–9.
28. Boersma H, van der Vlugt MJ, Arnold AER, Deckers JW, Simoons ML. Estimated gain in life expectancy: a simple tool to select optimal reperfusion treatment in individual patients with evolving myocardial infarction. Eur Heart J. 1996;17:64–75.
29. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18(suppl 2):S65–80.
30. Savage MP, Douglas JS, Fischman DL, et al. Stent placement compared with balloon angioplasty for obstructed coronary bypass grafts. Saphenous Vein De Novo Trial Investigators. N Engl J Med. 1997;337:740–7.
31. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18:341–64.
32. Claxton K, Neumann PJ, Araki S, Weinstein MC. Bayesian value-of-information analysis: an application to a policy model of Alzheimer's disease. Int J Technol Assess Health Care. 2001;17:38–55.
33. Altman DG. Comparability of randomized groups. *Statistician*. 1985;34:125–36.
34. Rittenhouse BE. The relevance of searching for effects under a clinical trial lamppost: a key issue. Med Decis Making. 1995;15:348–57.
35. Rittenhouse BE. Exorcising protocol-induced spirits: making the clinical trial relevant for economics. Med Decis Making. 1997; 17:331–9.