Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

A COMPARISON OF SCREENING METHODS FOR COLORECTAL CANCER USING SIMULATION MODELING

Ali Tafazzoli
Stephen D. Roberts

Department of Industrial Engineering
Campus Box 7906
North Carolina State University
Raleigh, NC 27695-7906, U.S.A.

Reid M. Ness
Robert S. Dittus

Department of Medicine
Vanderbilt University Medical Center
VA Tennessee Valley Healthcare System
Nashville, TN 37232-8300, U.S.A.

ABSTRACT

We used a discrete-event simulation model of the natural history of Colorectal Cancer (CRC) to do a cost-effectiveness analysis comparing the latest CRC screening strategies recommended by the American Gastroenterological Association (AGA) and the newest screening modalities for which clinical efficacy has been established. Cost-effectiveness was based on discounted costs and quality-adjusted life-years. A probabilistic sensitivity analysis examined the uncertainty in important parameter estimates. Considering all populations (average and high risk), annual Fecal Occult Blood Test (FOBT), Sigmoidoscopy every five years and annual FOBT, and Colonoscopy every ten years were the three strategies that demonstrated a greater than 50% probability of not being dominated in probabilistic sensitivity analysis. Depending on the maximum acceptable marginal cost-effectiveness value, any of these procedures have a high likelihood of becoming preferred (most effective strategy given a specific cost limit per Quality-Adjusted Life-Year (QALY) saved).

1 INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer death in the United States. This cancer has a very long asymptomatic phase and most patients are not aware of its presence until it enters an advanced stage when the survival rate is very low. Evidence from several studies suggests that screening for colorectal cancer and precancerous adenomatous polyps can reduce the incidence of CRC and related mortality. There are a number of screening strategies designed for this purpose based on existing screening methods, which vary considerably in their performance characteristics and cost. However, it is not possible to conduct clinical trials of all possible screening strategies for CRC. Simulation models offer an alternative means to evaluate and compare screening strategies.

We employ the Vanderbilt/NC State (V/NCS) discrete-event simulation model of the natural history of CRC (Cubbage 2004). We built a screening structure onto the model to perform deterministic and probabilistic cost-effectiveness analysis on the screening strategies recommended by the AGA's clinical guidelines (Winawer et al. 2003). The enhanced V/NCS model is capable of modeling homogeneous cohorts over time. Any modeled cohort includes a certain number of patients with specific birth year, race, gender, and family history of colorectal neoplasia. The model can simulate the effects of various screening, surveillance and treatment interventions to lessen CRC morbidity and mortality in a specific cohort over time. This model also computes the discounted costs and quality-adjusted life-years (QALYs) that are the main outcomes measures of this analysis.

1.1 CRC Screening and Surveillance Strategy

CRC screening strategies include an initial screening test followed by colonoscopy for positive tests. Strategies can be categorized into two groups depending on whether the initial test is endoscopic or non-endoscopic. Endoscopic tests provide direct views of the lining of the colon. The two types of endoscopy employed in CRC screening are sigmoidoscopy (evaluation of the colon distal to the splenic flexure) and colonoscopy (evaluation of the entire colon).

Non-endoscopic methods rely on consequences of adenomas or cancers present in the colon. One type of non-endoscopic tests, such as the fecal occult blood test (FOBT) and fecal DNA test, indirectly check the colon through examination of the stool. A second type uses non-endoscopic imaging tests, such as virtual colonoscopy and double contrast barium enema (DCBE) which apply radiographic techniques to indirectly image the lining of colon.

The accuracy of each screening test is judged by its sensitivity and specificity. A test with high sensitivity means that the test is nearly always positive when the dis-

ease is present. A test with high specificity means that the test is nearly always negative when the disease is absent. The usefulness of a test – predictive values of positive and negative tests - is dependent not only on the sensitivity and specificity of the test but the prevalence of disease in the screening population.

Screening tests are administered to asymptomatic individuals who are at risk for developing colorectal cancer. The tests are used to detect colorectal adenomas (the precursor of CRC) or cancer. Since the tests are not perfectly sensitive or specific, and since the prevalence of colorectal cancer is low in a general screening population, screening tests can generate many false positive and false negative results.

CRC surveillance refers to the follow-up examination of the colon in patients with a prior history of colorectal adenomas or cancer. Because a history of colorectal adenomas or cancer is a risk factor for future colonic neoplasia, surveillance regimens are more aggressive than standard screening strategies, meaning the frequency of the screening test is increased.

For any test to be effective, patients must be willing to have the test performed. CRC screening compliance is defined as a criterion that expresses an individual person's potential willingness over his/her lifetime to adhere to a screening strategy.

1.2 Clinical Guidelines

The most recently updated guideline for colorectal cancer screening and surveillance strategies was released in 2003 by the American Gastroenterological Association (Winawer et al. 2003). Based on new evidence and experienced judgment, clinical guidelines include different recommendations for people at average and increased risk.

The well-established approach to examining effectiveness of a medical intervention is the clinical trial. However, clinical trials are expensive, require a long length of time to establish results and cannot feasibly examine a large number of alternative strategies. A simulation model, however, can examine the cost-effectiveness of many alternative strategies in a timely fashion. In addition, a simulation model can explore strategies that might not be possible to include in a clinical trial as a result of current medical theories and practice paradigms that might be faulty, yet firmly established.. This paper is the first effort to compare all the CRC screening strategies recommended by the latest AGA's clinical guidelines and the most recently developed screening modalities.

2 BASIC DESIGN OF THE MODEL

The model has been constructed using a general object-oriented, open-source, discrete-event simulation platform implemented in the Microsoft .NET environment. The model has been programmed using managed C++.NET and VB.NET to be efficient in the Windows operating system. The simulation platform on which the CRC model is constructed provides a recent random number generator (L'Ecuyer et al. 2002) and uses inverse transform generators to attach a unique source of randomness to each random variable in the model. Consequently identical populations of individuals have "common random numbers" that reduce the sampling error throughout the simulation and statistically enhance the comparison of any given set of screening scenarios. A modeling scenario is defined as the unique set of parameters developed to inform a specific simulation. The simulation base provides for the automatic definition of entities, the creation and management of one or more event calendars, the processing of actions, and the collection of statistics.

The simulation model is constructed in this .NET object-oriented environment so that base objects can be used to define higher-level objects and specify their behavior without modifying lower-level objects. The individual patients and their adenomas are objects constructed from basic objects. In a dynamic modeling environment like CRC, not only data change, but also the modeling structure has to be frequently revised to accept new ideas, alternative uses, and improved calibration. An object-oriented simulation provides a general and convenient platform for implementing these changes.

In developing this model we used the inheritance feature to modularize the discrete-event simulation structure. Modules are defined in two separate levels. The lower level includes the general modules needed for a simulation program that define basic classes like the event calendar, simulation event, simulation entity, and other classes required for collecting observation-based and time-persistent statistics. The higher level consists of the modules related to the CRC natural history along with the added screening and surveillance procedures.

Model input has been separated from model execution through the use of an Access database. The database allows the definition of new variables, including random variables and the specification of their distributions to be added or modified without changing the simulation model. The database also stores scenarios for later retrieval and retains all of the output from a given scenario for future review.

2.1 Objects

To employ the object-oriented modeling capability, the V/NCS CRC model defines new "higher-level" objects and creates their interactions. The central object is an individual person, which is instantiated from the person class that is derived from the entity class. In each replication of the simulation, a person object is created and it flows in its own pathway among events. Each person has unique per-

sonal characteristics, which are sampled from the cohort variables that are to be studied.

The other important class of objects associated with each person in this model is the collection of colorectal neoplasms for that person. Whenever an adenoma is created, it is assigned characteristics that determine if and when it will become CRC. Once formed, a CRC can progress through various stages and lead to patient death. Each neoplasm has a separate event calendar that stores all the future events associated with it. This design is a key to efficient simulation of screening, since removing a neoplasm eliminates all the events related to that neoplasm so they can be removed from the event calendar. The total simulation efficiency gained from this modeling structure is considerable.

The scenario is another object that is instantiated from the scenario class only once at the beginning of the simulation run to define the characteristics of the cohort to be studied and the screening intervention that is to be performed. Parameters for each random variate generator are assigned based on scenario definitions. Parameters for each scenario are stored and retrieved from the database, which also stores the calibrated variables for the natural history simulation.

2.2 Events

The base simulation objects include events, which are managed in event calendars. The base simulation also removes events, advances time, and calls event actions. When a person object is created, several procedures are executed within its new object initialization. The model is constructed by employing an event graph. The event graph portrays all events in the simulation and, when relevant, what conditions are required for an event to be scheduled. If no conditions are stated, the event is always scheduled. Events in the CRC simulation are documented by their description, predecessor events (events that must take place prior to this event occurring), next events scheduled, and statistics collected.

3 MODEL VERIFICATION

The main purpose of the verification was to ensure that the screening and surveillance models worked properly. The screening model is programmed in Microsoft Visual Studio.NET 2003, which provides a very strong development environment for constructing, compiling, testing and executing code. Taking advantage of these tools and other techniques, two methods were employed to verify the model: (1) stepping through the screening source code line by-line and routine-by-routine to verify the general program flow, while watching the key variables, and (2) creating and examining an output trace file.

3.1 General Flow Verification

General flow analysis was the primary step used to verify the general program flow and insure the correct execution of the code and the behavior of the event calendars. Several people were processed through each of the screening strategies using the Visual Studio.NET "debug" feature to watch the processing of the simulation from within the code. One intention was to insure that the correct distributions were being retrieved from the database and sampled properly. The next step was to look at the program flow and make sure that people were moving through the screening process correctly and following the expected decisions at each event.

The most comprehensive verification occurred when the Visual Studio debugger was used to check the key variables, which were discounted cost and QALYs. The reason was that, validation of these two values is critical to the cost-effectiveness analysis. Therefore exact verification was done to check whether cost and utility are collected and discounted properly. All discounting calculations were also performed by hand to check their computational correctness.

3.2 Trace Output Analysis

Trace statements have been added to the code. They were placed at every event so that they reveal the event time and the action the code is taking relative to that event. The trace statements are written to an external text file whose content is then examined for discrepancies. Since the general flow of the model has usually been verified, the trace provides a summary of experience that is more easily seen as a whole, such as time between screening events and response to cancer or adenoma diagnosis. Looking over trace output was a tremendous help in verifying correct implementation of screening and surveillance scheduling.

4 MODEL VALIDATION

To validate the ability of the V/NCS model to simulate screening strategies over a specific period of time, a comparison of the Minnesota Colon Cancer Control Study was made to a simulation of that population. The Minnesota study was performed to compare the effectiveness of annual and biennial FOBT screening strategies in reducing CRC mortality to no screening program (Mandel et al. 1993). This trial was chosen because it is a well-known study within the medical literature and other models have used it for validation purposes since it provides comprehensive output not found in other randomized trials.

The Minnesota study was initiated by recruiting 46,551 men and women volunteers 50 to 80 years of age from 1975 through 1977 and randomly assigning them to three groups: annual screening, biennial screening and con-

trol (no Screening). The participants in these three groups were studied until 1991 when that study ended. During this interval the CRC incidence and mortality were accumulated for all three groups. The final report of this randomized trial revealed a 33% decrease in CRC mortality within the annual FOBT screening group in comparison with the control group.

The enhanced V/NCS model was designed to have the capability to replicate the Minnesota study. The most difficult part of this process was to generate a population with the same characteristics based on age, gender and family history in simulation model as was observed in Minnesota trial.

The results generated from the V/NCS model are shown in Table 1. The CRC mortality in the control and annual groups of the simulation model were observed to be very close to what happened in actual study as it is shown in Table 1. All simulation outcomes (CRC incidence and mortality) were somewhat lower than the actual trial. This may reflect a lower risk for CRC in the modeled population compared to the actual population in the trial. However, the percent decrease in CRC mortality, comparing the actual and simulated outcomes, was identical (~33%). This result adds substantial validation to the V/NCS model.

Table 1: Modeled versus Actual CRC Outcomes for the Minnesota Study

	CRC Incidence		CRC Mortality	
Strategy	Actual	Simulation	Actual	Simulation
Control	356	331	121	116
Annual Screening	323	294	82	81

Figure 1 compares the cumulative CRC mortality during the study years between the actual trial and simulated model. It is easily seen from these curves that both the annual and control group's mortality curves in the simulated model are similar to the actual trial.

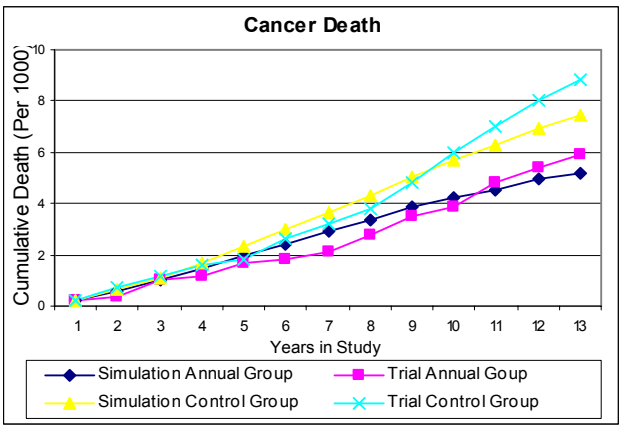

Figure 1: Comparison of Actual and Simulated CRC Mortality for the Minnesota Study

5 DETERMINISTIC COST-EFFECTIVENESS ANALYSIS

Cost-effectiveness analysis (CEA) is considered the most appropriate method of comparing preventive health services from an economic point of view (Gold et al. 1996). The basic purpose of CEA is to assess the cost of health-care resources dedicated to a health-care intervention relative to the health-care benefits that are produced by that same intervention.

The central measure in CEA of health-care interventions is the cost-effectiveness ratio. It is the difference in costs of two alternatives divided by the difference in their effectiveness. In other words, the CE ratio is essentially the incremental price of obtaining a unit health effect from a given health intervention when compared with an alternative. In case the intervention under study is both less costly and more effective than the alternative, it is said to dominate the alternative and there is no need to calculate a CE ratio. Therefore, CE analysis is performed only in circumstances where the intervention is both more costly and more effective than the alternative. Interventions that have a relatively low ratio would typically have high priority for resources.

CEA can also show tradeoffs involved in choosing among various interventions. In this situation the CE ratios should compare each intervention to the next most effective option, after eliminating options that are dominated to obtain incremental cost-effectiveness ratios. Similarly, options can be ruled out by extended dominance, i.e., when a linear combination of other options can produce greater benefits at lower costs.

We performed a deterministic cost-effectiveness analysis to quantitatively compare the expected outcomes of competing screening alternatives when the best estimates of model parameters were applied (base-case analysis).

5.1 Model Assumptions

The analysis was done separately on eight populations that can be defined according to various gender, race and family history combinations (see Table 2).

Table 2: Defined Populations

White Male NF*	White Male F**	White Female NF	White Female F
Black Male NF	Black Male F	Black Female NF	Black Female F

*NF: No Family History
**F: Family History

Each of these eight groups has its own response to different screening strategies, since all of the factors used in population definition (gender, race and family history) meaningfully influence the CRC development risk.

The screening strategies studied on average risk (no family history) populations were the ones recommended in the AGA's clinical guidelines. The capabilities of two newly proposed screening strategies, virtual colonoscopy and fecal DNA tests, were also examined to compare them with current strategies. Specifically, the strategies are summarized as:

- FOBT every year
- DCBE every five years
- Sigmoidoscopy every five years
- FOBT every year and sigmoidoscopy every 5 years (Sig & FOBT strategy)
- Colonoscopy every 10 years (Colon 10)
- Virtual Colonoscopy every 10 years (VColon)
- Fecal DNA every 3 years.

For the high risk population (with family history), along with the above mentioned strategies a colonoscopy test every 5 years (Colon 5) was also studied as it was recommended in the guidelines.

Cost effectiveness ratios for each screening strategy indicates the average cost a person incurs for following that intervention to gain a quality adjusted year of life. Therefore to compute this ratio, cost and QALYs results of a screening strategy should be compared to the cost and QALYs of the situation where no screening is present. For this reason all the populations were also simulated under a no-screening strategy.

5.2 Population Size

After observing several simulation outputs, it was determined that the mean estimates of QALYs had a much smaller range among the different screening strategies then the cost values. The magnitude of the difference in QALYs was in the thousandths.

Therefore to obtain a valid result, the variance of QALYs should be lowered to an extent that their confidence intervals do not overlap when comparing any two strategies. By achieving the desired amount of precision on QALYs, the necessary precision for cost would certainly be provided (Al, Van Hout, and Michel 1998).

Applying the concept of "common random numbers" to V/NCS model, it was possible to simulate different screening strategies on a uniform population with the same characteristics. Therefore differences in mean performances, such as cost and QALYs, between various screening strategies are estimated more precisely than absolute mean performances of an individual strategy (Kelton, Sadowski, and Sturrock 2004).

To estimate the proper model size, first a sample population of size 20,000 was simulated under Colon 10 and Sig & FOBT strategies and the 3% discounted QALYs for each person were computed. The 95% confidence interval on the average difference in QALYs were computed as following:

$$\bar{y} \pm t_{1-\alpha/2,n-1}\frac{s_y}{\sqrt{n}} = -.00667 \pm 1.96\frac{0.977}{\sqrt{20000}} = [-0.0202, 0.0068]\,.$$

The half-width of the difference in QALYs for a population of size 20,000 was about $H_0 = 0.0135$ years. A margin of error of ±0.005 years (less than 2 days) with 95% confidence interval was selected to satisfy our required precision. The total population necessary to be simulated to achieve the assigned precision was therefore computed as follows:

$$n \cong n_0\frac{H_0^2}{H_{desired}^2} = 20000\frac{0.0135^2}{0.005^2} = 145,800\,.$$

Based on this output, a population size of 150,000 was assigned for each replication of the screening strategies.

5.3 Cost-Effectiveness Results

Considering all the assumptions mentioned above, discounted cost and QALYs were computed for different populations and variations of the interventions.

Although all of the CRC screening strategies examined here provide higher QALYs (extended life expectancy) than "no screening", some of these strategies are more cost-effective as indicated by a lower cost per life-year saved than for the alternatives. In order to identify these strategies, a cost-effectiveness analysis was implemented.

First, all of the different populations were examined to determine whether any strategies were simply dominated by other strategies having lower costs and greater effectiveness (QALYs). Second, the principle of extended dominance (i.e., whether linear combinations of other strategies can produce greater benefit at lower cost) was applied to all strategies. Third, among the non-dominated screening strategies, the incremental cost-effectiveness ratios (ICER) were calculated by comparing each strategy to the next more costly and more effective intervention. This process produces an "efficiency frontier" of increasingly more costly and more effective strategies. Finally, the results of analysis for each population are presented on a cost-effectiveness plane, together with the efficiency frontier line for non-dominated strategies within a population (Gold et al. 1996).

The CE plane is a two-dimensional space with the x-axis being the average difference (treatment – control) in effectiveness (ΔE) and the y-axis being the average difference in cost (ΔC). The axes are unbounded from positive to negative infinity, and the origin represents the control group because scales are in difference form.

The cost-effectiveness calculations are only illustrated for "White Male NF" population in this paper. The necessary tables and graphs for the rest of populations are presented in the thesis by Tafazzoli (2004). The following table shows the cost-effectiveness analysis for the "White Male NF" population.

Table 3: ICER of Different Screening Strategies

Rank	Strategy	Lifetime Cost 3% Discounted	QALYs 3% Discounted	ICER ($ per QALY Gained)
	No Screening	1003	15.8125	…
1	FOBT	1274	15.8421	9,165
	DCBE	1548	15.8443	Dominated
2	SIG	1591	15.8536	27,436
	SIG & FOBT	1728	15.8569	Dominated
	VColon	1936	15.8388	Dominated
3	Colon 10	1989	15.8643	37,062
	Fecal DNA	2333	15.8312	Dominated

Closer examination of this table reveals that DCBE, Sig & FOBT, VColon and Fecal DNA strategies can be eliminated by either simple or extended dominance. For example, since the Fecal DNA strategy has both higher cost and less effectiveness in comparison with other strategies, it can be rejected by the simple dominance principle.

An ICER is also computed in Table 3 for each non-dominated strategy. For instance, the ICER of the sigmoidoscopy strategy in comparison with FOBT strategy is as follows:

$$ICER_{(FOBT\text{-}Sig)} = \frac{Cost_{(Sig)} - Cost_{(FOBT)}}{QALYs_{(Sig)} - QALYs_{(FOBT)}} = \frac{1591 - 1274}{15.8536 - 15.8421} = 27{,}436 \ \$/QALY.$$

Comparing each screening strategy to the no-screening strategy, the following CE plane was developed (Figure 2). In this figure the x-axis shows the average difference in cost between each screening strategy and the no-screening strategy while the y-axis displays the average difference in QALYs. Note that the origin represents the no-screening strategy.

The efficiency frontier in this figure is given by the lines joining the no-screening, FOBT, Sigmoidoscopy and Colon 10 strategies. DCBE and Sig & FOBT strategies are internal to this frontier, indicating that they also can be ruled out through the principle of extended dominance (i.e., a linear combination of FOBT and sigmoidoscopy strategies would strongly dominate the DCBE strategy). The slope of the frontier at any point reflects incremental cost-effectiveness, i.e., the additional cost at which additional QALYs can be purchased.

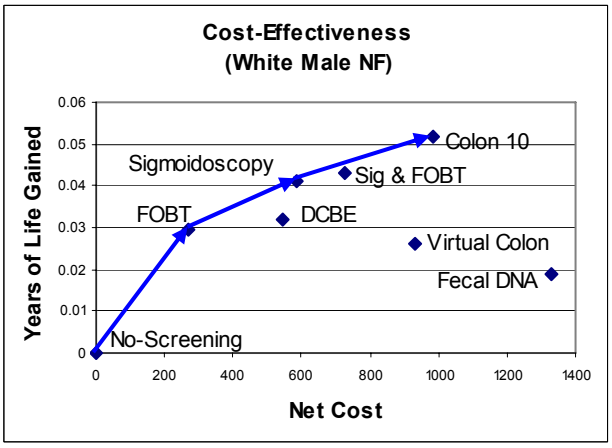

Figure 2: CE Plane and Frontier Line

Overall, this analysis showed that three strategies (FOBT, Sigmoidoscopy, Colonoscopy) may be preferred depending on the maximum acceptable marginal cost-effectiveness. The Incremental CE ratios for these strategies ranged from $9,165 per QALY gained to $37,062.

Reviewing the results of this analysis a few questions may arise, such as "How robust are these conclusions and how much does the uncertainty in parameters influenced these results?" or "Is it consistent for medical decision makers to simply rule out the dominated strategies?".

Taking a careful look at the cost-effectiveness planes and the efficiency frontiers, it was perceived that in several situations the dominated strategies might have formed a part of the frontier line if their cost was decreased or their effectiveness increased. For instance, the Sig & FOBT strategy in Figure 2, was very close to the frontier line and it wouldn't be wise to reject this strategy knowing that uncertainty exists in the perceived values of many cost and effectiveness parameters of this model.

6 PROBABILISTIC ANALYSIS

Confidence intervals for cost effectiveness ratios are a valid approach to addressing uncertainty in cost-effectiveness analysis. However, as a ratio statistic, the solution to confidence interval estimation is not straightforward (Briggs, O'Brien, and Blackhouse 2002). Two main approaches have been proposed for the purpose of estimating the confidence interval for incremental CE ratios: the parametric method introduced by Fieller (Fieller 1954) and the nonparametric approach of bootstrapping (Fenn, McGuire, and Backhouse 1996).

In this paper the bootstrapping method was applied in the context of the probabilistic sensitivity analysis (PSA) approach to generate the confidence intervals. The main

objective of doing a probabilistic sensitivity analysis using the V/NCS model was to examine the impact of uncertainty in a set of input parameters on the cost-effectiveness ratios of various screening strategies.

6.1 Candidate Parameters

In the V/NCS model there are three types of model parameters that can be subjected to sensitivity analysis:

1. Structural parameters used in the creation of the natural history model (e.g. the distribution of the lifetime CRC risk within a population).
2. Cancer control strategy parameters (e.g. colonoscopy test characteristics).
3. Economic parameters used in the analysis (e.g. cancer care costs).

All the structural parameters are intimately linked to the calibration of the natural history model. Although easily accomplished, alternation of these parameters would result in a new CRC natural history model that would need to be calibrated again. Therefore structural parameters were not included in the PSA.

Strategic and economic parameters are good candidates for PSA since most often their values are not known with certainty. A single point estimate of these parameters is usually given along with their range of variation (95% intervals) which can be used to do a one-way or multi-way sensitivity analysis.

In PSA, instead of assigning an estimated value to a strategic or economic parameter, it is assumed that each uncertain parameter can best be described by a probability distribution. This description entails assigning a prior probability distribution to each of these input parameters. These prior distributions reflect all available information and prior beliefs about the parameters' true values. So, for example, if cost data are available for the colonoscopy test, the mean and standard deviation of this sample could be used to define a proper distribution. Essentially this is a Bayesian approach, with the model parameters being treated as random variables (Briggs, O'Brien, and Blackhouse 2002).

Using data from literature along with recommendations given by the V/NCS model's expert panel, the prior distributions were developed. If more data become available on these parameters, the likelihood function of these parameters can be used to derive a better posterior distribution applying the Bayesian method.

6.1.1 Cost Parameters

In probabilistic sensitivity analysis (PSA), different cost variables were sampled from different Beta distributions. A Beta distribution was appropriate for cost variables because there existed information about a minimum value, a maximum value, and most likely value for these variables. The standard deviation of the distribution was estimated from the assumption that it was one-sixth of the range. The main issue was that according to the expert panel opinion different costs for a typical person are positively correlated. The expert panel decided to use 0.5 as the value of correlation coefficients between all pairs of cost variables. This choice was made because 0.5 shows the minimum amount of bias towards existence or absence of correlation and the expert panel did not have any reason to be biased towards any of them.

To apply the concept of correlation, the capability of generating random variates from multivariate distributions with a given correlation matrix was added to the code.

6.1.2 Screening Characteristics Parameters

Sensitivity and specificity parameters of all screening modalities along with complication parameters associated with them were also considered in the probabilistic analysis. Using literature available on characteristics of screening tests, the expert panel suggested a range of variation for each of these parameters, as well as a most likely value.

All these parameters provide the model a true or false response. As in the case of complication parameters, they either show a complication such as hemorrhage, after a screening test or not. Therefore they can be considered as independent Bernoulli trials leading to a binomial form of the data likelihood. With such data, it is natural to use the proportion of the true responses as the estimate of the corresponding probability in the model.

Fortunately, Bayesian methods provide a straightforward method for assigning prior knowledge to the parameter of the binomial likelihood. The Beta distribution is a continuous distribution on the interval 0-1 and is a conjugate family for the binomial likelihood. Hence, the ease of updating the Beta prior to a Beta posterior when supplied with additional data, is one of the main advantages of using a Beta prior for the parameter of the binomial distribution (Gelman et al. 2004).

For this reason prior Beta distributions were fitted to these parameters based on existing statistics. The advantage of this distribution is that by varying its two shape parameters, a wide variety of possible shapes to the distribution over the interval can be obtained.

Based on recommendations of the expert panel, a correlation coefficient of +1 was used between all the pairs of sensitivity variables of a screening test, since they are fully correlated. Also between the specificity and each of the sensitivity parameters of the screening tests, a correlation coefficient of -1 was specified. The reason for this specification is that specificity and sensitivity of a test inversely affect each other and as one decreases the other will certainly increase.

6.1.3 Utility Parameters

Unfortunately, except for a point estimate, very little is known regarding the exact behavior of CRC utility parameters. Based on what information was available, the expert panel recommended a skewed Beta distribution for these parameters. The maximum utility was set to the age-utility the person would have in the absence of cancer (when the person is healthy). The minimum was obtained by multiplying the estimated utility value for the health state by the maximum utility value. The mode is then the estimated utility value for the health state. The standard deviation of the Beta distribution was assumed to be one sixth of the utility range $STDEV = (Max - Min)/6$. Thus, the Beta distribution is highly skewed toward the maximum value, since the minimum value usually was very close to the mode.

This distribution of the utilities is apparently similar to what happens in reality, based on the opinion of the expert panel. As an example, the procedure for computing these statistics for utility of a 54-64 years old person, diagnosed with regional low cancer is demonstrated. The age-utility for this person without cancer is 0.92, and the estimated utility after this cancer is 0.59. Therefore the Maximum should be equal to 0.92. Minimum is 0.59*0.92 = 0.54 and STDEV is equal to $(0.92 - 0.54)/6 = 0.063$.

6.1.4 Compliance Parameters

For a screening modality there are different levels of compliance such as never-compliant, one-time compliant, etc. For a particular modality, a typical person may have one of these compliance levels with a specific probability. Therefore, in general, if there are k levels of compliance, a multinomial distribution with parameters $p_1, p_2, \ldots, p_k$ is a reasonable distribution to model the compliance level of a person, where p_i is the probability that the person has compliance level i.

In probabilistic sensitivity analysis, it was decided to assign probability distributions to parameters of this multinomial distribution, i.e. $p_1, p_2, \ldots, p_k$. It should be noted that since the sum of these probabilities must be equal to 1, they are not independent and as a result, it is not possible to assign independent distributions to each of them. In Bayesian theory there is a straightforward conjugate family for the multinomial distribution, namely the Dirichlet distribution.

Drawing samples from a Dirichlet distribution is a simple procedure. Each sample is produced using k independent samples from Gamma distributions with parameters which depend upon the parameters of the Dirichlet (Devroye 1986). This generator was added to the code.

It should be mentioned that the parameters for each Beta distribution were derived using the VIM program (Roberts 2004).This program allows users to fit a desired distribution to certain statistical characteristics such as mode and standard deviation. For instance Figure 3 shows the output reported by the VIM program, after being provided the following information for the statistical characteristics of the "Diagnostic Colonoscopy cost": Mode = $661, STDEV = $59, Min = $ 529, Max = $1,058.

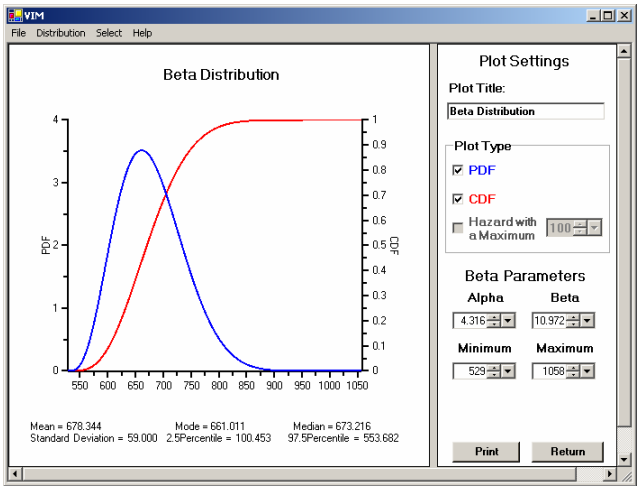

Figure 3: Screenshot of VIM Program Output

Figure 3 shows a Beta (4.316, 10.972) distribution for the studied parameter.

6.2 Probabilistic Sensitivity Analysis

Having specified distributions for all the relevant parameters of the model, the probabilistic analysis was done by randomly sampling from each of the parameter distributions and calculating the discounted expected cost and discounted expected QALYs for that combination of parameter values. This process formed a single replication of the model results for a specific screening strategy within a population. A total of 100 replications were performed to examine the distributions of the resulting costs and outcomes for each screening strategy.

Figure 4 shows the results of these 100 replications plus the base-case (i.e., the deterministic cost-effectiveness analysis) on the CE plane for the "White Male NF" population, together with the efficiency frontier of the base-case (Briggs, Blackhouse, and O'Brien 2002). For the other seven populations this CE plane is provided in the thesis done by Tafazzoli (2004).

Figure 4 shows how it may not be possible to rule out Sig & FOBT and DCBE strategies, since they potentially form part of the frontier in many replications. However, it is not possible to gain a clear view from Figure 4 as to how often these two strategies form part of the frontier because there can be substantial correlation between simulations plotted in this figure.

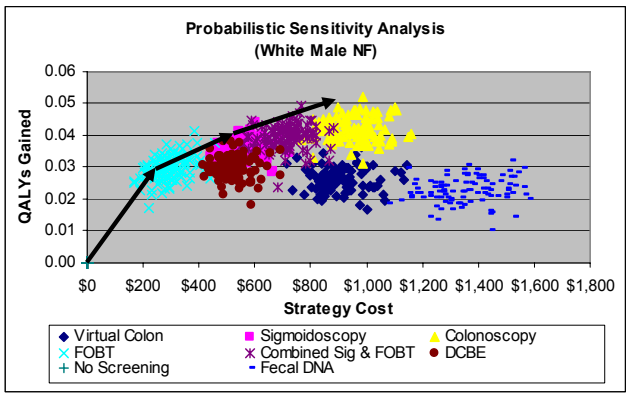

Figure 4: CE Plane for 100 Replications of Screening Strategies

The above figure also shows the empirical distribution of cost-effectiveness for different screening strategies. Confidence limits for each strategy were obtained by selecting the 3rd and 97th percentile of the 100 replicates (which excludes 2 (or 2.5%) of the observations from either end of the empirical distribution). Table 4 shows the lower and upper 95% confidence limits for cost-effectiveness of different strategies relative to "no screening".

Table 4: CI for CE Ratios of Different Screening Strategies

Strategy	Lower Limit	Upper Limit
FOBT	$7032/QALY	$12,919/QALY
Sigmoidoscopy	$13,082/QALY	$19,144/QALY
DCBE	$13,514/QALY	$23,240/QALY
SIG & FOBT	$17,649/QALY	$22,430/QALY
Colon 10	$18,862/QALY	$28,576/QALY
VColon	$24,668/QALY	$53,484/QALY
Fecal DNA	$38,727/QALY	$92,517/QALY

As a result, 95% of the estimated joint density falls within the wedge on the cost-effectiveness plane defined by the confidence limits. These wedges are shown for FOBT, DCBE and Colonoscopy strategies in Figure 5.

These confidence limits also support the conjecture that DCBE and Sig & FOBT strategies should not be ruled out from screening choices, since their confidence intervals overlap with confidence intervals of the strategies that formed the frontier line in the base-case analysis (i.e., Colonoscopy and Sigmoidoscopy strategies). Although these results are valuable, they do not recognize the inherent variability in the CE ratio.

As the next step of the analysis we figured out how often screening strategies form parts of the frontier line in the 100 replications. Table 5 shows these proportions for "No Family History" populations.

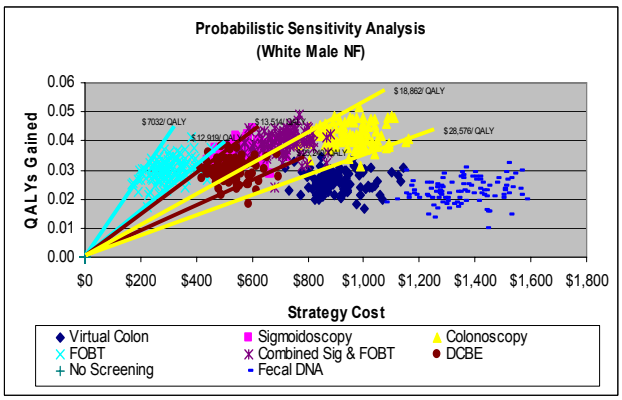

Figure 5: CI of FOBT (Blue wedge), DCBE (Brown wedge), and Colonoscopy (Yellow Wedge) Strategies

Table 5: Percentage of Time Each Screening Strategy Formed Part of the Frontier Line

Strategy	% formed part of frontier line
FOBT	100%
DCBE	7%
Sigmoidoscopy	46%
SIG & FOBT	74%
VColon	0%
Colon 10	75%
Fecal DNA	0%

It is clear from Table 5 that VColon and fecal DNA strategies can be ruled out, since they appeared zero times on the frontier. By contrast, it turned out that DCBE formed part of the frontier in 7% of simulations and Sig & FOBT also formed part of the frontier in 74% of simulations. However, even knowing the proportion of times a strategy forms part of the efficient frontier, it is not clear how this result could be interpreted by medical decision makers.

If the shadow price for an extra QALY (the maximum willingness to pay or "ceiling ratio") were known, it would be possible to choose between all of the screening strategies, not just identify those that form the efficient frontier. Therefore, conditional upon knowing the ceiling ratio, in each replication there is only one strategy of choice from eight screening strategies under evaluation. The proportion of times that an intervention is the strategy of choice from the 100 replications of each population gives the strength of evidence in favor of that strategy.

7 CONCLUSIONS

This analysis added insight into the comparison of alternative screening strategies by explicitly considering the uncertainty in cost-effectiveness. Such uncertainty can be added into the simulation model as described in this work.

Overall, considering the CEA for each of the populations (average and high risk), FOBT, Sig & FOBT, and Colonoscopy were three screening strategies recommended by the AGA's clinical guidelines that demonstrated a greater than 50% probability of not being dominated in PSA. Depending on the maximum acceptable marginal cost-effectiveness value, each of these procedures have a high likelihood of being the preferred strategy.

REFERENCES

Al, M. J., B. A. van Hout, B. C. Michel and F.F. Rutten. 1998,. Sample size calculation in economic evaluations. *Health Economics* 7: 327-35.

Briggs, A. H., G Blackhouse, and B. J. O'Brien. 2002. Probabilistic analysis of cost-effectiveness models. *Medical Decision Making* 22: 290-308.

Briggs, A. H., B. J. O'Brien, and G Blackhouse. 2002. Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual Review of Public Health* 23: 377-401.

Cubbage, D. 2004. Simulation of colorectal cancer: The natural history of disease. Master thesis, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available via <www.lib.ncsu.edu/theses/available/etd-05282004-134141/unrestricted/etd.pdf> [accessed March 30, 2005].

Devroye, L. 1986. *Non-uniform random variate generation*. Springer-Verlag, New York.

Fenn, P., A. McGuire, M. Backhouse, and D. Jones. 1996. Modeling programme costs in economic evaluation. *Journal of Health Economics* 15: 115-25.

Fieller, E. C. 1954. Some problems in interval estimation. *Journal of the Royal Statistical Society* 16: 175-83.

Gelman, A., J. B. Carlin, H. S. Stern, and B. D. Rubin. 2004. *Bayesian data analysis*. Chapman and Hall.

Gold, M. R., J. E. Seigel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-effectiveness in health and medicine*. New York, NY: Oxford University Press.

Kelton, W. D., R. P. Sadowski, and D. T. Sturrock. 2004. *Simulation with Arena*. 3d ed., Series in Industrial Engineering and Management Science. Boston: McGraw-Hill Higher Education.

L'Ecuyer, P. L, R. Simard, J. E Chen, and D. W Kelton. 2002. An object-oriented random-number package with many long streams and substreams. *Operations Research* 50, (6): 1073-1075.

Lothgren, M. and M. Zethraues. 2000. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Economics* 9: 623-30.

Mandel, J. S., J. H. Bond, T. R. Church et al. 1993. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study [published erratum appears in N Engl J Med 1993 Aug 26;329(9):672]. *New England Journal of Medicine* 328: 1365-71.

Roberts, S. D. 2004. Simulation input without data. *Proceedings of 6th Annual Simulation Solutions Conference '04*, March 16, 2004, Orlando, Florida.

Tafazzoli, A. 2004. A comparison of screening methods for colorectal cancer. Master thesis, Department of Industrial Engineering, North Carolina State University, Raleigh, North Carolina. Available via <www.lib.ncsu.edu/theses/available/etd-11232004-160814/unrestricted/etd.pdf> [accessed March 30, 2005].

Winawer, S., R. Fletcher, D. K. Rex, et al. 2003. Colorectal cancer screening and surveillance: Clinical guidelines and rationale -- Update based on new evidence. *Gastroenterology* 124: 544-60.

AUTHOR BIOGRAPHIES

ALI TAFAZZOLI is a Ph.D. student in the Department of Industrial Engineering at North Carolina State University. His research interests include simulation modeling and analysis. This paper reports part of the work conducted for his master thesis. He is a member of IIE. His e-mail address is <atafazz@ncsu.edu>.

STEPHEN D. ROBERTS is a professor in the Department of Industrial Engineering at North Carolina State University. He has served as Proceedings Editor and Program Chair for WSC. His e-mail address is <roberts@ncsu.edu>, and his web page is <www.ie.ncsu.edu/roberts/>.

REID M. NESS, MD, MPH is an Assistant Professor of Medicine at Vanderbilt University and a member of the Geriatric Research, Education and Clinical Center in the VA Tennessee Valley Healthcare System. He has published previously concerning the cost-effectiveness of colorectal cancer screening. His e-mail address is <reid.ness@vanderbilt.edu>.

ROBERT S. DITTUS, MD, MPH is the Albert and Bernard Werthan Professor of Medicine and Director of the Center for Health Services Research and the Geriatric Research, Education and Clinical Center at Vanderbilt University and the VA Tennessee Valley Healthcare System. His e-mail address is <robert.dittus@vanderbilt.edu>.